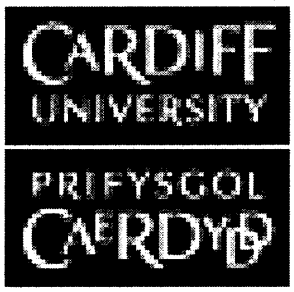


Evaluating Dynamic General Equilibrium Models

By

Konstantinos Theodoridis

A Thesis Submitted in Fulfilment of the Requirements for the Degree of Doctor of Philosophy of Cardiff University

Cardiff Business School

Cardiff University

September 2006

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ..... (Konstantinos Theodoridis)

Date .....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated.

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ..... (Konstantinos Theodoridis)

Date .....

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... (Konstantinos Theodoridis)

Date .....

ABSTRACT

In this thesis we introduce a new bootstrap method for testing structural DSGE models according to their dynamic performance. The method maintains a separation between the structural (non-linear) model as the null hypothesis and its dynamic time series representation. The model's errors are discovered and used for bootstrapping (after whitening); the resulting pseudo-samples are used to discover the sampling distribution of the dynamic time series model. The test then consists of discovering whether the parameters of the time-series model estimated on the actual data lie within some confidence interval of this distribution. A test statistic for the parameters taken as a whole is developed (the M-metric, a Wald statistic).

To the memory of my grandfather, Konstantinos Theodoridis, who, heroically, served in the battle for an independent and a democratic Greece.

ACKNOWLEGMENTS

I wish to express my sincere gratitude to all those who have assisted me whilst working on this project. Special thanks, undeniably, go to my supervisor Professor Patrick Minford for his guidance and for his useful advices and suggestions. I am also grateful to Professor Halbert White for his remarkable comments given in such generous and unpretentious manner.

My colleagues, starting from the youngest one, Miss Le Vo Phuong Mai, Mr Eric Scheffel, Mr Gareth Evans, Mrs Yanhui Zhu, Miss Mei Chuang, Dr Subarna Pal, Dr David Meenagh, Dr Ruthira Naraidoo and Dr Francesco Perugini, that kept in touch deserve undoubtedly an acknowledge for their encouragement and cheerfulness along the way. I would like to thank my friends, Dr Stauros Avramidi, Dr Penny Bazou and Dr Leonidas Vachtsevanos for always being a source of optimism whenever I was about to lose it. I could not forget my friend, Ioannis Trikoukis, for his unselfish, generous and, continuously, provided support.

I reserved my final acknowledgment for my family and particularly my parents Anastasios and Vera Theorodoridi for all the sacrifices that they have done for me. My father's guidance and my mother's moral support were always sources of strength in achieving my goals and their encouragement and financial support was a prerequisite for my success throughout my life. Special thanks go to my sister, Anatoli Kanellidou Theodoridou, for always standing by me and supporting my decisions and to my gorgeous niece, Eleutheria Kanellidou, for her love.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. PREFATORY REMARKS

This paper proposes a new way of testing macroeconomic models in terms of their dynamic performance. It is of course by no means entirely new. In recent years proponents of Real Business Cycle models have proposed that models be tested dynamically by comparing their simulated moments with the moments in the data. They have disregarded the ability of these models to fit the data in static terms, something that other economists have generally regarded as important. Instead they have argued that a model is a pure abstraction that can be calibrated based on other (often micro) evidence (indeed some proponents have even calibrated models to fit dynamically as well as possible.).

We take no position here on whether models should be calibrated or estimated to fit the data statically as closely as possible. Our testing procedure takes as a primary assertion that some model is 'true' (as a working hypothesis) or at least to be treated as the null hypothesis (i.e. the hypothesis being provisionally regarded as holding). How this model is arrived at is not considered. From this point on however our method departs from RBC procedure in key ways:

1. we propose that the data restrict the error processes used to produce the model's simulated behavior. A model implies a division of the observed data into 'model prediction' and 'error'. The implied error is rather like the 'fundamental' in a test of capital market efficiency: one may 'make up' a world in which some error process drives this capital market's price efficiently, but if that error process bears no relation to the actual profits or dividend process from which the price is derived then the test is hardly convincing. Similarly for example in an RBC model if the productivity process assumed in simulation bears no relation to the actual productivity process in the data as implied by the model, then the simulated moments cannot be those of the relevant world of the sample data. Thus we propose that actual implied errors be used in a bootstrap to generate simulated model behaviour.

2. We recognize the nonlinearity inherent in macroeconomic models, and hence that any data representation process (DRP) cannot be considered a 'reduced form' of the structural model (SM). Rather if the SM holds the DRP is some

approximation of the SM's implications, rather than a closed-form transformation of the SM. It follows that to find the standard errors for the DRP's parameters under the assumption that the SM holds we need to simulate the SM; we cannot use the estimated DRP standard errors as these may be biased by the particular approximation involved in the DRP.

3. Nevertheless the SM may be able to supply restrictions on the DRP. This is especially relevant in considering whether the impulse responses implied by the DRP are consistent with the SM, as in general the DRP cannot supply identification restrictions for the current innovations.

4. By the use of bootstrapping we are able to derive ad hoc small sample distributions, rather than rely on asymptotic ones which are inappropriate in general given the small size of economic data samples.

In what follows we compare and contrast our proposed methods with existing ones; we give extended examples of their application; we consider a variety of theoretical and practical issues that arise with their use. The plan of the thesis is:

- by way of extended illustration apply these methods to a model of the UK economy that has been in use for over twenty years and also to apply them to a panel sample of postwar growth data across nearly 100 countries (chapter 1)
- review the existing methods that are comparable (chapter 2)
- explain the methods proposed (chapter 3)
- consider the theoretical fundamentals of the approach (chapters 4)
- conclude with some general remarks about the strengths and weaknesses of the approach.

2. AN APPROACH TO TESTING MACRO MODELS WITH APPLICATIONS

2.1 Introductory remarks

Let us suppose a model has been estimated appropriately and that its parameter values have not been rejected by the data, in the usual sense that they exceed twice their standard errors. Such a test of a model is rather limited; many rival models can be fitted to data acceptably in this way. Furthermore tests of models against each other through non-nested or encompassing tests are often found to be inconclusive. The reason is easy to see: different theories can be pushed to explain the same facts by structuring the unexplained errors differently.

For this reason economists have tried to test models in other ways. They have in particular considered whether the models could:

1. predict within and outside sample, based on lagged data.
2. generate effects of shocks (impulse response functions) like those seen in the data
3. generate simulated moments like those in the data
4. generate VAR and other representations like those estimated on the data.

In chapter 2 we survey the methods that have been used in these four categories in some detail. But we may note some problems that have been encountered with them:

- the difficulty of forecasting tests (which measure success or low standard error in forecasting) is that errors in forecasting are inherent in any model because shocks cannot be forecasted. Thus for example in a world of large shocks model errors should be large; a model with a low forecasting error would be misspecified. Or, another example, if there is structural change, then a model should forecast badly beyond a structural break. Thus any test needs to be adjusted for these problems.

- identifying the impulse response functions in the data is problematic because such identifying restrictions can only come from structural models. Thus the models being tested must be used to specify the identifying restrictions, which alters the test since there is now no unambiguous data-given function with which to compare each model. We will argue later for a way of testing the impulse response functions implied by structural models but it differs from such a comparison.

- data moments are apparently theory-free; that is one may calculate them without assuming any structural theory. The main problem faced is the choice of filter; but this in practice is of great importance. Not filtering implies that trended data is dominated by its trend; but filtering by differencing or common filters like HP or Band-Pass may produce quite different values.

- VAR and other time-series representations face a similar problem. There are many ways in which one can choose to represent a set of time-series; and for them too a decision must be made on filtering.

We suggest that a way forward is, in line with Popper's original ideas, to set up as the 'null' (strictly, the 'working') hypothesis the structural model (SM) to be tested and derive from this null, as if it is the true model, its predictions about such features of the data as moments, impulse responses and time series properties. We would then test whether the model's predicted features cohered statistically with those of the data. At all possible points in this procedure we impose the restrictions embodied in the SM.

A further suggestion is that we should use all available data in our tests. Thus the SM implies what the structural errors are and their values are in general supplied by the actual data; thus testing against the data should allow for the data's restrictions under the SM on the errors being used in generating the SM's simulated performance.

A further, related, suggestion is that we respect the small sample nature of our tests by using bootstrapping rather than asymptotic distributions which will in general mislead in small samples. Bootstrapping under small sample conditions can be carried out by using the actual data on the errors just mentioned.

In this chapter we pursue an extended example of such a procedure using the Liverpool Model of the UK on UK data since the 1970s; we also consider an application of a growth model to postwar panel data on 76 countries. First, however, we consider the methods that are currently or have recently been in use. They divide, following Canova (2005), into four groups:

1. Watson's R^2 measure

2. The Generalized Method of Moments (GMM) and related approaches using the distribution of the actual data

3. The Simulated Method of Moments (SMM) and related approaches using the distribution on the simulated data.

4. Bayesian methods.

1. Watson's R^2 statistic is not really a test statistic at all. Watson proposes a measure of the 'fit' of a model against the data in terms of the error process that needs to be added to the model in order for the model's autocovariance function to match exactly that of the data. The size of this vector error process is used to measure the extent to which the model fails to capture the data. The problem with this measure is that it has no distribution and thus cannot generate a test statistic.

2. Under the GMM method a model is linearized and fitted to the data by the Generalized Method of Moments. The error distributions on the model's parameters create an implied distribution of the difference of the model's moments from the data moments. This distribution is χ^2. The method relies throughout on asymptotic distributions. It also requires that the model is linear and that the data are made stationary by some filter. Recently Christiano has pointed out that in small samples the asymptotic distributions are badly misleading.

3. Under the SMM the parameters that are regarded as 'free' , ie available to be varied according to the data, are moved until the vector of VAR coefficients implied by the model (which can be analytically calculated) is as close as possible to that from the data. The difference between these two vectors generates a χ^2 statistic whose distribution is based on the simulated behavior of the model. This statistic is again asymptotic. There are a number of variants of this approach based on other features produced by the model and the data- eg spectral measures and moments.

4. Bayesian methods strictly do not test a model but rather regard the initial model as some sort of approximation to the true model; thus the question is how far this approximation can be improved by using the process of estimation and comparison of the model with the data to modify the prior model. At no point does the researcher consider rejection of the model. Rather one model variant may be compared with another (as 'nested' hypotheses) and one chosen

as giving a greater improvement than the other. Consequently the approach must be considered as quite different from that of testing being developed here, where the idea is to test quite different theories- which cannot be nested- against the data. Plainly if economics reached the stage where no such testing was necessary and the only issue remaining was to improve an agreed general model, then testing in our sense would no longer be relevant and instead the Bayesian agenda could naturally replace it. Nevertheless even then a new uncommitted researcher might wish to test the consensus and for that a test would be required in our sense.

A fuller account of all these methods is in Chapter 3. But in sum, we suggest that the relevant class of existing tests are those in 2) and 3). These tests have points in common with what we propose. In effect what we are suggesting incrementally is the use of bootstrapping to overcome small sample problems; the use of the original nonlinear model rather than linearization in simulation; and the use of implied errors from the data and model structure, rather than assumed error processes. With these tools it should be possible to make clear probabilistic statements about models.

Our paper is organized as follows. In the first section we elaborate on the nature of the test we propose . In the second section we apply our test to a particular model, the Liverpool Model of the UK, as an initial experiment; the reason for choosing this model is simply that we have extensive experience with using it and have FIML estimates of the parameters and error processes. In the third section we apply our test to a model of endogenous growth and taxation, using postwar panel data on a large number of countries. In the last section we draw some tentative conclusions from these experiments and discuss further work, some of which is already in hand.

2.2 *Our proposed test*

The basic test: comparing the reduced form parameters with the distribution based on the structural model

In this section we explain our proposed method for testing macro models by using the bootstrap; we think of the method as an extension of that used by Real Business Cycle (RBC) modellers. Consider a) a data set: say the UK economy from 1979 to 2004; b) a structural theory of its behavior stating that it behaves according to certain parameters whose values have been estimated by some means. Notice that in the course of estimation various tests of the model will probably have been performed- notably whether the parameters are statistically significant, that is are sufficiently large to reject the hypothesis that they are zero. But let us assume that

the model has either passed such tests or is in any case asserted to be a reasonable null hypothesis. Thus we consider models that have already been estimated or calibrated, by whatever means; methods for estimating or calibrating models are not discussed in this paper. Our starting point is the assertions of some modellers to the effect that their model is to be regarded as 'the truth', that is as a null hypothesis. We shall use 'truth' and 'true' to mean that the modellers assert that their model is to be taken as an assertion of how the economy in question works. Such an assertion is made in order for testing to be done of a concrete hypothesis. Thus the 'truth' is the same as the economist's null (strictly, working) hypothesis.

Now consider the idea for testing models already in use by RBC modellers. This idea has been to reorganize the facts in the form of correlations and cross-correlations at different leads and lags. In this form the facts are metamorphosed into 'stylized facts' which describe the data in a relevant way but one that is entirely theory-free. We can call this relevant way the 'time-series behavior' of a set of variables, meaning by this the description of the variables in terms of their 'trends' and their 'cycles', common and individual, that is their correlations (and inter-correlations) over time. Then the RBC proposal has been that the testing of models may be done by comparing the correlations they imply in simulation with the actual ones in the data. This sort of test is quite different from the testing of parameters (typically against zero) that may be done in estimating a model. The model could pass such a test for every parameter and yet not be a model which implies good dynamic behavior like that of the economy. Vice versa, a model could fail every such test and yet produce behavior that resembles the dynamic behavior of the economy.

The method we propose here takes this RBC idea but suggests two changes. First, instead of using the theoretical error distributions postulated by the model to compute the model's correlations with the facts, we suggest using the actual error distributions implied by the data to compute them, using the bootstrap. Second, instead of using correlations we suggest using a more general time-series representation of the data such as a VAR[1].

Let us consider these two proposals in turn.

1. If a model is asserted to be true, then it implies a set of (structural) errors between the theory and the data. Since the model is true, the implied errors are also true. Hence under the null hypothesis the errors driving the model are

[1] We also considered using a VECM and carried out a number of experiments with it. In the end we abandoned it. The reason was that the LVP model, our null hypothesis, implies that there should be no cointegrating vectors within our group of 5 endogenous variables: cointegration within the LVP model requires the presence of other (exogenous) variables . Hence in a restricted reduced form the 'cointegrating vectors' we 'found' were spurious statistical artefacts and were invalid.

these implied ones. To test the null we use these errors to see what behavior in conjunction with the model they imply. Under the bootstrap we can preserve these sample error distributions as the basis for our stochastic draws.

It should also be noted that if a model contains exogenous variables then the modeler must decide whether these are stochastic or non-stochastic. If stochastic then they too must be modeled as some sort of stochastic process and their errors included in the stochastic simulations.

To summarize, a model is an assertion of a set of relationships and of which processes are to be regarded as stochastic, which non-stochastic. Given the model, the data over a sample period then define the actual (true) errors for the stochastic parts as well as the non-stochastic parts.

2. we would like to find a general description of the data that emphasizes its trend and cyclical features in a reasonably summary way, that is with a set of parameters that can be compared with the implications of the theory we wish to test. However, the potential size of the matrix of correlations and cross-correlations at different lags becomes very large very quickly; this is not therefore a compact description of the data. However a VAR is a potentially complete but parsimonious representation of the data in time-series form. We may add to this a GARCH process in the error. The VAR/GARCH representation is not the only possibility; we discuss others below. But it is widely used and convenient; our main point is that a more general representation of the facts than the correlation matrices can provide us with a more comprehensive test. The more we can restrict this representation of the data in conformity with the implications of the structural model, the clearer a test we can perform of the latter as the null hypothesis. However since in general the structural models we consider are nonlinear they do not have a linear reduced form. None of the data representations will be derivable from the structural model: they are all some sort of approximation, on which we may possibly be able to impose some restrictions from the structural model.

Thus in summary the null defines the model and the structural 'error' data and it also may restrict the reduced form in possible ways. We can then estimate the reduced form on the data and compare it with the bootstrapped predictions of the structural model. Our suggested test is thus designed to reject the null hypothesis of the model at some specified confidence level (say 95%). Under the null hypothesis the model stochastic error processes are defined within the sample period. We may then ask whether this model and these processes

could have generated the facts as captured by the VAR estimated over the same sample period. Thus under the null hypothesis, the sampling variation would be given by bootstraps of the stochastic errors with the model; these bootstraps would provide us with a large number of pseudo-samples on which we can estimate the same VAR to establish the sampling distribution for the VAR parameters[2].

We believe that this test provides a statistical method for rejecting the null of a particular model, in the sense that it answers the question 'could the facts of the economy have been produced by this model?'

A test for multi-variate parameter distributions

In practice a VAR or other data representation has many parameters; a test that relies on the single-parameter confidence intervals cannot reliably be used as a test of all parameters taken together. We have accordingly developed a method for estimating the joint distribution of the parameters. We begin by making the heuristic assumption (in the sense that we can test it at a later stage and either correct it or check it for robustness) that the parameters of the times series model follow a multi-variate normal distribution; this allows us to estimate the likelihood value of each parameter combination, and from this we can rank them all in order of their likelihood. The bootstraps then give us the frequency of each combination and its percentile ranking relative to the highest-likelihood combination (i.e. the mean of the bootstrap parameter values). We call the M-metric. If we choose a 95% confidence interval then the actual data should generate a parameter combination whose M-metric is less than 95% for the model not to be rejected.

If we could assume that the parameter combinations followed a multi-variate normal distribution these M-metric values could be translated back into a normal likelihood value. (If we found from the bootstrap distribution that that the distribution was not multi-variate normal but some other one, then we could repeat the operation under this other distribution and translate the M-metric back into a likelihood value from this one.) This becomes useful for comparing two models since we can think of the M-metric as ranking their likelihood (that is to say, if Model A has a higher metric than Model B, then under Model A the data is less likely to

[2] It might be thought that one could compare the 'reduced form' distribution of the data representation's parameters with the distribution generated by the structural model. However because the 'reduced form' is not derived from the structural model but is rather some approximation, its distribution is also not derived from the structural model; therefore we do not know what relation the two distributions should have to each other.

be observed than under Model B; or alternatively, given the data, the likelihood of Model A is less than that of Model B). We would prefer to use Model B in this case.

Testing the model's impulse responses against those in the reduced form

We also test the impulse responses implied by the SM against those in the data. We may note that Christiano, Eichenbaum, and Evans (1999):(CEE) have been for some time operating a methodology which is designed to do just this. However we believe it is flawed by problems of identification. CEE set up a calibrated or estimated theoretical structure and generate its impulse response function to particular shocks. This is then compared with the impulse response function obtained from a VAR estimated on the actual data but one that is restricted (or 'identified') in a certain manner that is only weakly related to the theoretical structure. Thus the idea is to identify the VAR by imposing on it restrictions regarded as universally accepted by economists, non-controversial as it were, so that this VAR could be regarded as an acceptable representation of the facts. The theory can then be tested in respect of its ability to fit selected aspects of this representation that are of interest. For example CEE look at impulse response functions to a monetary policy shock. The theory can then be regarded if it matches this function as being a satisfactory representation for the purpose of monetary policy analysis.

However, the identification scheme being used for the VAR can hardly fail to be controversial and more to the point will in general be inconsistent with the structural model under test. For example CEE identify the VAR so that a money shock has 'hump-shaped' response functions on both output and prices; this is a prior assumption emerging from some loose theory they happen to hold. Yet if the particular SM under test is taken as the null it may well not imply such a pattern of response. The SM might well be 'rejected' ('accepted') by such a representation when it would not be if the theory's own restrictions are applied in identification.

An alternative approach, which we use below, is to compare the impulse responses of the model with those of the unrestricted VAR when *the model's restrictions* are used to identify the relevant shocks to the VAR. This is done by establishing a mapping from the model shock involved to the model-implied shocks for the variables in the VAR; these shocks are then input into the VAR and the resulting movement in those variables plotted; the 95% confidence bounds for this are found from the VAR bootstraps.

2.3 Our experiments on the Liverpool Model of the UK:

2.3.1 Our procedure in summary

We look at the LVP model over the period from 1979 to 2004. As we know the monetary regime during this whole period has varied substantially. During the early 1980s there were money supply targets but the money measures targeted switched periodically; from the late 1980s there was 'DM shadowing', a sort of fixed exchange rate regime, followed by the ERM proper from 1990-92. Finally there was a switch to inflation targeting from end-1992, which continues to this day.

Plainly in these circumstances we should not reasonably test the model against whole period correlations or VARs. We looked at data on correlations for a sample starting in 1970 and treating the model as one single-regime model over this period. However it turns out that it is hard to interpret the comparison between model-generated and data correlations, because different filters give quite different results and in addition there are many different variables and sets of cross-correlations between them. The general problem is that there is no unambiguous test of the model across this plethora of information, even if we could take seriously the idea that the whole period was governed by the model under one regime. So we turn at once to a general data representation form which we may use to test the model in a way that allows for regime shifts.

2.3.2 The LVP Model with three separate regimes from 1979-2003:

Thus we take seriously the null hypothesis separation of monetary regimes. We bootstrap each regime separately to obtain pseudo-samples from 1979-2003 that fully embody the different regimes. We then impose these different regimes as far as we can on the reduced form structure.

Plainly therefore our structural model indicates some sort of shifting in the data representation model. We tried various data representations but the most successful empirically was a VAR where there is Markov switching of variances only. What is encouraging in this experiment is that the three states picked out do correspond to a high degree with the a priori regime allocations. Thus state 1 occurs largely from 1979-84, suggesting it is the 'money targeting' regime. State 3 occurs mainly in the mid-1990s and then from the late 1990s to 2003, suggesting it is the 'inflation targeting' regime. State 2 according to our null hypothesis is the 'exchange rate targeting' regime which should occur from 1986-1992. According to the data determination here it occurs in 1986, 1992 and again in 1997; but not in 1987-1991 which the data allocate to money targeting. This last regime also intrudes into the

inflation targeting era, in 1998. Of course our regime allocation is far from cast-iron, especially in the period from 1986-1992 when considerable confusion reigned in policy (the prime minister and Chancellor publicly disagreed on whether the UK was targeting the exchange rate or the money supply).

The model fares well in terms of the 95% confidence intervals for individual coefficients. Thus, with the marginal exception of the variance of inflation in state 2, it picks up the whole variance-covariance matrix in the three switching states (Table B.3 in Appendix B.1). For the VAR parameters it rejects 5 out of 30 at the 95% level, with one of them only borderline (Table B.2). Thus taking all the parameters of the Markov-switching model, it rejects 5 out of 75 (Tables B.4 and B.5). The M-metric comes out at 94.8% which means that the data (just) do not reject the model.

We now go on to try a data representation where we assume different VARs in the different regimes; we move from a Markov- to a deterministic-switching description. In order to keep as many degrees of freedom as possible, we treat the sample as a whole and in effect introduce dummies for each regime, testing in turn whether there is evidence in the data of shifts across regimes in constants, VAR parameters and the VAR error var-covar matrix.

Results of Imposing 3 Regimes on VAR

The first regime here is the monetary targeting regime, ending at 1985Q4. This is followed by the Fixed exchange regime until 1992Q3, at which point the inflation targeting regime begins and continues until the end of the sample in 2003. In this specification all the parameters- means, autoregressive, and var-covariance values- are allowed to vary, in effect through a set of dummy shift parameters. Below, in the Appendix B.2, we show first the actual versus fitted values from the data (Figure B.2); and then the parameters compared with their 95% confidence limits from the bootstraps (Tables B.6, B.7, B.8 and B.9).

When we compare the number of parameters rejected at 95% confidence with the total, it is clear that this model, restricted as it is by the fixed switching of regimes, is a lot less successful with individual parameters than the one where we allowed Markov-switching. Its M-metric is 98.8%; this is worse than the Markov representation and the model is thus rejected by the data when represented in this form. What this suggests is that the model requires adjustment; while it is borderline not rejected by the data when represented by a Markov-switching process this non-rejection is not robust to how the data is represented, since when the data is represented by a fixed-switch point VAR it is rejected at the 95% level. Quite

how the model should be adjusted is a matter for future research; but the fact that rejection seems to be associated with fixed or moving switch points seems to suggest exploring different regime switch points in the model's application.

2.3.3 Impulse responses

Further information on the model's strengths and weaknesses can be obtained from comparing its impulse responses with those in the data. In what follows we show the impulse responses generated by a supply shock (a rise in the employers' tax rate on workers, BO) and by a nominal shock (MTEM: this is a fall in the money supply under the Money targeting regime; a fall in foreign prices under Exchange rate targeting; and a temporary rise in interest rates under the Inflation targeting regime). The response of the VAR when the VAR shocks are identified via the effect of the structural model (so that in effect their first period effect is the same as that of the LVP) is shown by the solid line; and the 95% confidence interval for the VAR effect coming from the model bootstraps is shown by the two dotted lines. If the model is correct, we should observe that the VAR effects lie within the 95% interval. (The LVP response itself is not strictly relevant and not therefore shown). All the responses shown are for the differences of all the variables, not their levels.

Notice that the 95% confidence intervals around the impulse responses come from the same VAR parameter confidence intervals we have already used to test the model against the data as summarized in the VAR data representations; the M-metric tested the model as a whole. There is no further test of the model to be found in the impulse responses; these are shown below merely as a source of information about how well the model fits in this particular dimension.

A: Money Targeting Regime (1979-85)

Here we see that all the VAR effects lie within the 95% interval except for unemployment which is somewhat outside for both the supply and nominal shocks (Figures B.3 and B.3 in Appendix B.3; dotted red lines are the boundaries of the distribution, discontinuous black line is the actual response, blue solid line is the response of the structural model).

B: Exchange Rate Regime (1986-92)

In this regime all the VAR effects lie within the interval (Figures B.5 and B.6).

C: Inflation Targeting Regime (1992-2003)

In this regime the VAR effects lie outside the 95% interval for several variables under both shocks. Under the supply shock inflation, the exchange rate and unemployment lie outside; under the nominal shock the exchange rate and unemployment lie outside (Figures B.7 and B.8).

D. Comments:

What we find here is that, with some exceptions, the model is close to capturing the impulse responses. A handful of responses, virtually all confined to the inflation targeting regime, lie outside the model's 95% limits. It is possible that these faulty responses are connected with the model's rejection by the data, as discussed above. However, the fact that the model captures most impulse responses suggests that the model is not too far from acceptance by the data. On the assumption that the model should thus be taken seriously, we may notice that none of the impulse responses have a 'hump shape' (in levels too, if we integrate these impulses, this would also be true, as can easily be verified.). Such a shape would be the product of particular identifying restrictions on the VAR innovations but such restrictions do not hold with the LVP model. Thus we have no reason here, given that the data is not too far from accepting the LVP, to impose other identifying restrictions on the VAR than those implied by the LVP. It is of course possible that some other model implying hump-shaped responses if put as the null hypothesis could also fit the data- even possibly better than the LVP. But we have been unable to check on this so far; and the key point made here is that one is not compelled by the data to assume hump-shaped impulse responses. Rather we note that the LVP model responses, largely accepted by the data, do not have them.

2.3.4 A Second Example: Growth and Taxation

In a recent paper, Minford, Meenagh, and Wang (2006) use the method set out here to test the effect of taxation on growth. This is of interest because it tests between several models and so extends the method here beyond just one model. As indicated above, we may use a model's M-metric as an indicator of the likelihood of the data given that model; thus we may compare the M-metrics of different models and rank them accordingly by likelihood. In the paper Minford et all use annual panel data on 76 countries from 1970 to the present day. They set up a standard model of endogenous growth for a small open economy and are able to derive two structural equations from it, one for the production function allowing for the effect

of the business-tax-modified incentive to innovate on productivity growth (in this the effect on growth of a 0.1 cut in the business tax was 0.14% p.a.- a coefficient of -0.014) and another for labor supply with general tax as an incentive variable. These two structural equations are estimated on the panel data and the implied errors extracted; these errors are modeled as time-series (AR) processes. The two tax variables are modeled in the same way. The innovations in these time series processes are then treated as a shock vector; this vector is a 76-country vector with thirty-odd years' values. The vector is resampled with replacement, re-input into the time series processes and then into the structural model to generate a large number of pseudo-samples. These constitute the 'sampling world of the model', that is to say the samples that the model could have produced with different random shock draws.

Just as we used a VAR to describe the data in a 'reduced form' for the LVP model, so here Minford et al use a data description in terms of the 'reduced form' effect of tax on growth plus time and country dummies. They find that only business tax is significant and so the equation is cut back to one with growth regressed on the level of business tax, plus dummies. The resulting descriptive equation is:

$$\Delta \log y_t = \alpha_1(\tau_t + \pi_t')$$

Using panel data with fixed effects may not be the most efficient model to run. Estimating the model with random effects will give a more efficient estimator (the reason for this is that the estimator saves degrees of freedom by not using the fixed country dummies but instead using the regression with fixed country dummies with a weight, to correct the regression with time dummies only). The results of the estimation are shown in the following table.

Tab. 2.1: **Estimation Output**

Estimation Methods	*Estimate*	*Standard Error*
Fixed Effect (Country and Time Dummy)	-0.043	0.014
Random Effect	-0.05	0.027

To test whether we should use the fixed or random effects model we run a Hausman test, the results from this test were

From Table 2.2 we find that we can use either fixed or random effects in the actual data sample without serious risk of inconsistency. This is of interest in that it tells us that the effect of business tax in the descriptive regression is well-determined and highly significant.

Tab. 2.2: **Hausman Test**

Coefficient	*Fixed*	*Random*	*Difference*	*Standar Error*	χ^2	*Prob. Value*
a_1	-0.043	-0.05	-0.007	0.022	0.1	0.751

We now turn to the bootstrap distribution of the business tax parameter. While the Hausman test indicates the random effects estimator is probably consistent for the actual data sample we cannot be sure it will be consistent for any sample, let alone the bootstrap samples. Consistency is essential for these samples since each sample provides us with a 'central' estimate of the parameter in that sample; if it is not central then our distribution estimates will be biased. Furthermore we do not care about the efficiency of the estimator because we will not be using the standard error of the estimates; we use only the estimates to construct the bootstrap distribution (and implied standard error) of the 'reduced form' coefficient. Thus we use the estimator that is sure to be consistent, namely that with dummies for country and time. On this basis, the results were as follows:

Tab. 2.3: **Bootstrap Results for Model with Estimated Tax Effects**

Coefficient	*Actual*	*Lower Bound*	*Upper Bound*	*M-metric*
a_1	-0.05	-0.054	0.017	90.8%

The basic model is thus accepted- notice the M-metric is redundant here when only one coefficient is of interest. However, it might be that a model in which there is no tax effect at all but otherwise the same, would do just as well. If so the acceptance of the basic model is hardly persuasive that taxes matter for growth. To test this Minford et al set up the same model but this time with a zero effect of taxes; the model was reestimated with this constraint, the new errors extracted and the bootstrap exercise repeated. The results were:

Tab. 2.4: **Bootstrap Results for Model with Zero Tax Effects**

Coefficient	*Actual*	*Lower Bound*	*Upper Bound*	*M-metric*
a_1	-0.05	-0.04	0.03	98.1%

We see that this alternative model is rejected, with an M-metric of 98.1%. Thus the basic model is accepted by the data at the 95% level, whereas the alternative model with no tax effect is rejected.

So far the basic model was being tested from the zero side, so to speak- to see

whether it dominates a no-tax-effect model. It is also of interest to test it from the other side: to see whether a business tax effect higher than freely estimated would satisfy the data description. So Minford et al also reestimated the model imposing an increased coefficient on business tax and retrieving the implied new errors. They used two cases, one in which they set the coefficient to -0.02 and another in which they set the coefficient to -0.04. The results for the -0.02 case are shown in Table and the -0.04 case in Table 2.6.

Tab. 2.5: **Coefficient on Business Tax set to** -0.02

Coefficient	*Actual*	*Lower Bound*	*Upper Bound*	*M-metric*
a_1	-0.05	-0.06	0.012	82.1%

Tab. 2.6: **Coefficient on Business Tax set to** -0.04

Coefficient	*Actual*	*Lower Bound*	*Upper Bound*	*M-metric*
a_1	-0.05	-0.062	0.011	81.7%

What is interesting about this is that there is an improvement in the model's performance vis-a-vis the data description as the model's business tax effect is raised. Thus if it is raised in absolute size by two standard errors to -0.02 (from the estimated -0.014) the M-metric falls from 90.8% to 82.1%. However the improvement stops from here on. If it is raised further to -0.04 it improves barely at all; this must be because it induces errors in the model whose variation is correlated with the tax and offset its effect on the distribution of a_1. Hence the data estimation of the model itself combined with the data description tell us that a business tax parameter of between -0.014 and -0.02 is the most compatible with the data.

A discussion of the empirical results

There are a number of aspects of this illustration that are of interest. We may start by discussing the 'conventional' way of testing the model using the standard reduced form approach. Thus we note that the model implication- viz that the level of business tax and the rate of change of general tax both affect growth- meets a mixed reception. The business tax effect alone is fairly significant against the usual zero alternative; the general tax effect is not. We concluded from this that the data description should not include the general tax effect as it does not contribute to explaining growth. We might also have concluded that there was evidence of a business tax effect. However as we have argued above this 'reduced form' evidence is

not a persuasive test for two reasons. First, the error terms in the reduced form will include omitted nonlinear effects of tax on growth that can bias the 'reduced form' coefficient. Second, other models in which tax plays no part could also generate this 'reduced form' result.

So we reviewed next the evidence from the bootstrapping method, where instead of the confidence intervals generated by the 'reduced form' we look at those produced by bootstrapping the structural model. We found here that the model was accepted by the data description and furthermore that an alternative model with no tax effects was rejected by it and thus also dominated in likelihood by the basic model. In fact a model with a higher business tax coefficient of -0.02 is more likely viewed from its fit with the data description equation (though less likely viewed as a direct estimate from the production function).

What is also striking is the insight afforded by the bootstrapping procedure into the biases in the 'reduced form' coefficients under the null hypothesis. Thus we know from simulating the model for a shock to the business tax rate that growth (in steady state) increases by 0.14% to 0.2% for every 0.1 (ie 10 percentage point) fall in the business tax rate under the model. However the 'reduced form' coefficients give a value for this business tax effect that is up to three and a half times as big. This indicates a huge amount of bias in the 'reduced form' coefficients; these values bear little relation to what the model would produce as the simulated effect. The model when bootstrapped reveals that the correlation of the tax shocks with the errors creates massive bias in the 'reduced form' estimates. To put it in concrete terms, for example when the business tax rate is cut this causes a rise in consumption and labor supply as well as in productivity growth; the former two create an independent source of output increase over and above the steady state increase; this association raises the estimated effect of a business tax cut on growth.

A last point of interest is that we were unable to change the structural coefficients on tax upwards beyond a certain point, even though the 'reduced form' results would have been better fitted by a large business tax coefficient, ceteris paribus. What we found was that the data forced the structural model errors to offset the effect of raising the business tax coefficient beyond a certain point. Had we kept the freedom to 'make up' the structural errors we would have been able to fit the 'reduced form' results easily. But because we forced the structural model to fit the data through the implied errors used in the bootstrapping, the fitting of the 'reduced form' was constrained. It is as if our results can only emerge satisfactorily if they can go through two mincers, each of a different shape; a structural mincer and a 'reduced form' mincer; only if the model can force its way through both are its results to be

believed.

2.3.5 Conclusions

In this chapter we have described a new method for testing macro models and illustrated its use. First we looked at the LVP model across three monetary regimes: floating with monetary targets, exchange rate targeting (implemented by fixing the exchange rate), and floating with inflation targeting. We represented it in the data by two types of regime-dependent VARs, one with Markov-switching of variances, the other with fixed point regime switching. The model was accepted by the data on the first representation but not on the second, using a measure of overall fit, the 'M-metric'. Together with the evidence from the impulse responses where most were within bootstrap confidence limits but some, especially under inflation targeting, were not, this performance is suggestive of ways to improve the model.

Our second illustrates the test method on panel data to do with growth and taxation. Here we actually compared two models, one with tax effects and the other where tax has no effect. Here we were able to use the M-metric not merely as the criterion of rejection but also as a means of ranking the two models. It ranked the tax-effect model above that with no tax effects.

Our final conclusion is that these methods are both practicable and provide a powerful tool to test and so improve model performance; we hope to use them in future work on other models.

3. LITERATURE REVIEW

This chapter reviews the existing methods of evaluating Dynamic Stochastic General Equilibrium (DSGE) models. These methods are divided into two main categories; what we may call structural econometric methods and methods for dealing with calibrated models. This separation is founded on the basis of the null hypothesis. In the first category the structural model is treated as the true Data Generation Process (DGP), in economic terms, the structural model is true and its basic set-up is not questioned; the issue is simply one of fitting it to the data as accurately as possible. This consideration allows the use of formal econometric methods to address questions regarding the validity of the null hypothesis; if these questions are answered satisfactorily then no further questioning of the model is required. On the other hand calibration researchers argue that DSGE models are heavily misspecified and, therefore, should not be treated as true. They can only explain certain features of the real world and, therefore, their evaluation should be concentrated on those features which they have been created to explain. According to Canova (1994) the differences between the two approaches are linked to the question the two approaches ask. Scientists from the first category start from the condition that their model is true and they are asking *"how false is it?"* while the calibration people consider their model as false and they are investigating *"how true is it?"*.

3.1 Structural Econometric Methods

Canova (1994) argues that the origin of this approach goes back fifty years ago to the methodology proposed by Haavelmo (1944). It would be useful at this stage to introduce the functional *state space* form of a linearized around a nonstochastic steady state solved DSGE model.

$$y_t = Ax_t \qquad \text{The Measurement Equation} \tag{3.1}$$

$$x_t = Bx_{t-1} + v_t \qquad \text{The State Equation} \tag{3.2}$$

(3.2) is the law of motion of the economy that describes the evolution of the state vector $x_t \in \mathbb{R}^{dx}$ while (3.1) relates the state vector of the model with the observable

variables $y_t \in \mathbb{R}^{dy}$ and $v_t \in \mathbb{R}^{dv}$ are the stochastic processes that govern the system. DSGE modellers from both categories emphasize on the tractability of the model and this induces the *stochastic singularity* issue; the number of the shocks in the model is less than the number of the observable variables ($dv < dy$)[1]. When this is the case a term u_t, called *measurement error or unobservable shock*, is added to equation (3.1) to complete the probability structure of the model[2]

$$y_t = Ax_t + u_t \tag{3.3}$$

3.1.1 Maximum Likelihood Estimation / Kalman Filter Algorithm

A structural model written in the state space form (equations 3.2 and 3.1 or 3.3 when the latter is singular) can easily be estimated by using Kalman Filter Algorithm an important building block in the prediction error decomposition of the likelihood (for a detailed discussion see Sargent and Hansen, 2004; Canova, 2005, Chapter 9, Chapter 6, respectively).

Given this estimate of the structural vector $\theta = \left(vec(A)', vec(B)', vech(\Sigma_v)', vech(\Sigma_u)'\right)'$ the validity of the null hypothesis is examined by testing restrictions implied by the theory, or by general goodness of fit tests, or by comparing the fit of two nested models. Additionally, when the stochastic vector is augmented with some measurement errors Canova (2005) argues that the comparison of the size of the estimated standard errors of the structural shocks and of the measurement errors could be used as a misspecification indicator.

Measurement errors are serially and contemporaneously uncorrelated and their introduction does not alter the dynamic of the model. Ireland (2004) goes a step further and allows these errors to follow a Vector Autoregressive process (VAR(1)). This action obviously increases the fit of his model; for instance when conditions ensuring the errors are serially and contemporaneously uncorrelated are imposed the likelihood value falls dramatically and the Likelihood Ratio statistic reject these constraints. However, the new model, called hybrid model, can no longer be considered as structural.

There are various difficulties with using these methods (for discussion see Sargent (1979); Altug (1989); McGrattan, Rogerson, and Wright (1997) and Canova (2005, Chapter 6).) However, the principal one from our viewpoint here is that the methods

[1] However, this may not be an issue these days, for instance, Smets and Wouters (2003) built a model with ten structural shocks.

[2] See Bierens (2005) for an alternative treatment of the singularity issue.

do not examine the 'dynamic performance' of the model, in the sense of its ability to reproduce the correlation patterns in the data, unless one can be completely sure that the structural model is well specified, this ability needs to be examined as a test of the model. It is this that is the focus of calibration methods, to which we now turn.

3.2 Calibration Methods

Evaluation methods for calibrated models are reviewed here. These methods attempt to assess the extent to which a model (usually calibrated but potentially also, at least partially, estimated) can reproduce the dynamic features of the data. Equations (3.1) and (3.2) are carried over in this section. In contrast to the preceding approach, no uncertainty is allowed to θ and the model deterministically links the endogenous variables to the parameters and exogenous stochastic process (see equation (3.1)). It is, either, the sampling variability of the exogenous stochastic processes (v_t) in the classical framework, or, the cross section variability of θ in the Bayesian framework, or, both that allows them to define a distance between the model and the data. Canova (2005) categorizes all these methods into four main groups on the basis of the nature of the mean used to measure this distance:

a . Approaches based on R^2-type measure;

b . Approaches based on the sampling variability of the actual data;

c . Approaches based on the sampling variability of the simulated data;

d . Approaches which use the sampling variability of both actual and simulated data.

For convenience this scheme is maintained here.

3.2.1 R^2-type Measures

This subsection mainly reviews the method developed by Watson (1993).

Watson's Method

Watson's working assumption is that the economic model is viewed as an approximation to the stochastic processes generating the actual data or in statistical sense the model is not true. Goodness-of-fit (R^2-type) measures are developed to assess this approximation. The core of these measures is the amount of the error needed

to be added to the data generated by the model so that the autocovariance implied by the model plus the error match the autocovariance of the observed data.

The economic model describes the evolution of a vector series say x_t while its empirical counterpart is denoted by y_t. ACF_y is the autocovariance function for y_t, which is unknown but it can be estimated from the data either nonparametrically or parametrically by using a time series representation say a VAR or a *Vector Moving Average* (VMA) process. Differences between the estimator $\widehat{ACF_y}$ and ACF_y arise solely from sampling error. Now if $ACF_y = ACF_x$ sampling error also accounts for the differences between $\widehat{ACF_y}$ and ACF_x. Standard probability measures indicate if any discrepancy between arises solely from the sampling error. However, this test would be meaningful if the economic model is the actual DGP under the null, which is not the case in the present context.

His measure of fit is based on the size of the stochastic error required to reconcile the autocovariance function of x_t with that of y_t. To be precise, u_t denotes the error vector required to make the autocovariances of $x_t + u_t$ equal to the autocovariances of y_t. If the variance of u_t is large, then the discrepancy between ACF_y and ACF_x is large and vice versa. Loosely speaking, you could think this as the error term in a regression model in which the set of regressors is interpreted as the economic model. The economic model might be viewed as good approximation to the data if the error term is small (i.e. the R^2 of the regression is large) and vice versa.

Error's autocovariance is defined as $AFC_u = ACF_y + ACF_x - ACF_{yx} - ACF_{xy}$ where ACF_y is unknown but can be estimated from the data while ACF_x is completely determined by the model. DSGE models are linearized around a nonstochastic steady state and expressed as low order VAR and VMA process from which ACF_x can be readily calculated (i.e. a standard RBC model with a unique technology shock is expressed as $x_t = a(L)\varepsilon_t$ where $a(L)$ is an $n \times 1$ matrix polynomial in L and ε_t is the unique normally zero mean distributed error).

However, ACF_{yx} is unknown because the joint probability between x_t and y_t is unknown. At this point an assumption regarding the ACF_{yx} is required. The existing literature provides two assumptions regarding this issue. They are strongly related with the way whereby data is collected or expectations are formed. The researcher could either set $ACF_{yx} = ACF_x$, which implies that x_t and u_t are are uncorrelated (econometric assumption), or $ACF_{yx} = ACF_y$, which would imply that u_t could be interpreted as signal extraction error, with y_t be an optimal estimate of the unobserved "signal" x_t (signal extraction assumption). Obviously, neither of these assumptions seems to fit in the present framework, the error is not the result of imprecise measurement or a forecast/signal extraction error, it represents

the approximation or abstraction error in the economic model.

Watson's way to define the autocovariance between y_t and x_t is to choose that ACF_{yx}, which minimize the variance of u_t. This means that the error process is chosen in that way in order to make the model as close to the data as possible. As Watson (1993) points out this arises from the fact that, independently of any assumption made regarding ACF_{yx}, if the lower bound of the variance of u_t is large then the model fits the data poorly and vice versa. Therefore, the bound is calculated by choosing ACF_{yx} to minimize the variance of u_t subject to the constraint that ACF_{yx} is positive semidefinite (a condition that ensure that the variance will be meaningful). His first example is used to illustrate this process. Say that y_t, x_t and u_t are scalar and for simplicity it assumed that they are serially uncorrelated random variables. The problem is to choose σ_{yx} to minimize σ_u^2 subject to the constraint that σ_{xy} is positive semidefinite $\left(\min \sigma_u^2 = \sigma_y^2 + \sigma_x^2 - 2\sigma_{yx} \; s.t. \; |\sigma_{yx}| \leq \sigma_y \sigma_x\right)$ and the solution is $\sigma_{yx} = \sigma_y \sigma_x$, which implies that x_t and y_t are perfectly correlated with $x_t = \frac{\sigma_x}{\sigma_y} y_t$ or $x_t = \gamma y_t$ $\left(E(x_t y_t) = \frac{\sigma_x}{\sigma_y} E(y_t y_t) \Longrightarrow \sigma_{yx} = \frac{\sigma_x}{\sigma_y}\sigma_y^2 = \sigma_y \sigma_x\right)$. Watson (1993) uses the last equation as another measure to assess the fit of the model. This equation shows how to calculate fitted values of x_t given the data. Then fitted data is plotted against the actual and this plot offers a way to judge whether or not the former captures growth or cyclical components of the latter. Once an expression about ACF_{xy} is obtained the following R^2-type measure could be used:

$$R_i^2(\omega) = \frac{ACF_u(\omega)_{ii}}{ACF_y(\omega)_{ii}} \qquad \tilde{R}_i^2(\omega) = \frac{\int_{[\omega_1,\omega_2]} ACF_u(\omega)_{ii}}{\int [\omega_1, \omega_2] ACF_y(\omega)_{ii}} \tag{3.4}$$

$R_i^2(\omega)$ measures the variance of the i-th component of the error relative to the variance of the i-th component of the data at frequency ω. The latter could be seen as the lower bound for the distance of model from the data, frequency by frequency. $\tilde{R}_i^2(\omega)$ does the same thing, however, over a band of frequencies.

As the author notices this method ignores nonlinearities and variation in conditional second and higher moments. Canova (2005) also reports two other shortcoming of Watson's procedure. He notices that it is not clear why one should concentrate only on the best possible fit, he suggests to use both the best and the worst fit and if the range is narrow and $1 - R^2$ of the worst outcome small, one could conclude that the model is satisfactory. The second drawback is related with the fact that this method does not provide information that may be useful in respecifying the model. R^2 could be low for a variety of reasons, the variances of the shocks in the data may be high or the dynamics of the model and of the data are different or the process for the states has large AR coefficients. Obviously, it makes a lot of difference whether

it is the first or the last of these causes that makes R^2 low.

3.2.2 Evaluation Methods based on the Sampling Variability

Two methods are studied in this section. One treats the structural model as true and construct a measure of fit on the basis of the estimation variability while the second does not make the same assumption and introduce a R^2-type measure based on the data variability.

Generalized Method of Moments

Calibration techniques were criticized for their assumption that structural parameters are known with certainty. Responding to this criticism Christiano and Eichenbaum (1992) develop a methodology, which evaluates the fit of an DSGE model, when some uncertainty is assigned to them. They use formal econometric methods to estimate the vector of the structural parameters $\left(\hat{\theta}_T\right)$, the only unknown element in their model, and use the sampling variability implied by their estimates to evaluate the fit of the model. It is known from asymptotic theory that under the usual regularity conditions $\hat{\theta}_T$ is consistent and asymptotically normally distributed $\left(\hat{\theta}_T \xrightarrow{P} \theta^o, \sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \sim^A N(0, \Sigma_\theta)\right)$. Another very useful asymptotic result, which is the cornerstone of this procedure, shows that a vector function of $\hat{\theta}_T$, $g\left(\hat{\theta}_T\right)$, is also normally distributed such as $\sqrt{T}\left(g\left(\hat{\theta}_T\right) - g\left(\theta^o\right)\right) \sim^A N\left(0, \nabla_\theta g\left(\hat{\theta}_T\right) \Sigma_\theta \nabla_\theta g\left(\hat{\theta}_T\right)'\right)$ (see Serfling, 1980).

Their working hypothesis is that the world is generated by their model (the DGP) so that the moments generated by the model would be identical to the actual moments under an infinite sample since sampling variability would be eliminated; but would differ under a finite sample because of sampling variability both directly and because it would cause the parameters to be estimated with error. The authors formulate a testing procedure to investigate whether these differences in the moments arise because of sampling variability rather than from model misspecification. Initially the model is linearized around a steady state growth path and then θ is estimated by using Hansen (1982) *Generalized Method of Moment* (GMM)[3]. Using the authors' notation, we let Ψ_1 denote the structural parameters. Thus given the estimate of this, $\hat{\Psi}_{1,T}$, and of the assumed stochastic distributions, the significance of the difference of the model's moments from those of the data, Ψ_2, could be examined by a Wald J- type statistic (Canova (2005) calls this statistic *economic* Wald

[3] For a detailed discussion regarding this method see Hamilton (1994, Chapter 14) or Canova (2005, Chapter 5).

statistic).

Given Ψ_1 the model implies particular values for Ψ_2. Denote these values by $g(\Psi_1)$. The test asks whether or not the differences between $g(\Psi_1)$ and Ψ_2 are equal to zero ($F(\Psi) = g(\Psi_1) - \Psi_2$ where $\Psi = [\Psi_1, \Psi_2]$). For instance and without loss of generality, assume that the structural model is given by $y_t = ax_t + u_t$, $x_t = px_{t-1} + v_t$ (where $|p| \leq 1$ and $E(x_t u_t) = 0$); in this case $\Psi_1 = (a, p, \sigma_u^2, \sigma_v^2)$, $\Psi_2 = \sigma_y$, $g(\Psi_1) = \frac{a^2}{1-p^2}\sigma_v^2 + \sigma_u^2$ and $F(\Psi) = \left(\frac{a^2}{1-p^2}\sigma_v^2 + \sigma_u^2\right) - \sigma_y^2$. A second order Taylor expansion of $F(\Psi)$ about Ψ^0 gives the J-statistic as $J = F\left(\hat{\Psi}\right)' Var\left(F\left(\hat{\Psi}\right)\right) F\left(\hat{\Psi}\right)$ whose main element is the variance-covariance matrix of the estimated parameters $Var\left(F\left(\hat{\Psi}\right)\right) = \frac{\partial F(\hat{\Psi})}{\partial \hat{\Psi}}' \Sigma_{\hat{\Psi}} \frac{\partial F(\hat{\Psi})}{\partial \hat{\Psi}}$. Under the null hypothesis this statistic is distributed as a chi square with degrees of freedom equal to the number of the moments tested ($\chi^2(\dim(\Psi_2))$). Thus in the simple example above the statistic follows a $\chi^2(1)$ distribution.

This distribution is based on asymptotic theory. This implies that this process would work properly if the sample size was big enough. However, big samples are a luxury that, usually, we do not have in macroeconomics. This implies the existence of a small sample bias which distorts the inferences obtained from the above procedure. The evidence[4] suggests that estimates of the parameters and of the standard errors are biased in small samples; this means that the J-statistic should be used with caution in small samples.

The use of GMM methods and the J-statistic require stationary series. This implies that either growth rates (differencing) or some filter should be used in order to ensure that the above moments exist. However, the use of a filter is not innocuous, Christiano and den Haan (1996) found that *Hodrick-Prescott* (HP) filtering induces large and persistent serial correlation in the residuals of the orthogonality conditions and this creates problems in the estimation of the spectral density at frequency zero[5] [6].

Apart from Christiano and Eichenbaum (1992), Burnside, Eichenbaum, and Re-

[4] Issues related with the sources of this bias, its direction and its importance are explained by Canova (2005, Section 5.4.1). He provides a list of studies examined the small sample properties of GMM estimators.

[5] This matrix is one of the main elements of the parameters' variance covariance matrix, which is estimated by using HACC techniques at zero frequency. The authors conducting a Monte Carlo study found that when the series have been HP filtered that matrix is downward biased and this affects the coverage probabilities that seem to be very far from the nominal values in this case. For more details see Christiano and den Haan (1996, Section 5).

[6] Christiano and den Haan (1996) applied a Monte Carlo study for the GMM and for the *economic* Wald statistic and they note: *"The results are disappointing. The asymptotic theory appears to provide a poor approximation in finite samples, particularly when the data have been HP filtered."*.

belo (1993) and F'eve and Langot (1994) have also used this method to evaluate their DSGE models.

Resampling Methods

The method introduced by Diebold, Ohanian, and Berkowitz (1998) is reviewed in this section. They consider their work as an extension of Watson (1993) procedure and they assess models by comparing model's spectra $S_x(\omega, \hat{\theta}, v_t)$ to data's spectra $S_y(\omega)$. This method allows for graphical comparison between $S_x(\omega, \theta, v_t)$ and $S_y(\omega)$ over ω, so we could, easily, see along which frequencies the model performs well and along which frequencies it performs poorly. For instance, *Real Business Cycle* (RBC) scientists would concentrate their attention on the business cycle frequencies.

$S_x(\omega, \hat{\theta}, v_t)$ is calculated either analytically or numerically by simulating long series of data from the model. Therefore, there is no sampling variability because the DGP is known, however, this is not the case for the actual data. An estimate for the latter $\widehat{S_y}(\omega)$ is obtained by using well established techniques and its sampling variability is created through the *Cholesky factor bootstrap* algorithm.

Bootstrap distributions are valid frequency by frequency, however, one often wants to assess the sampling variability of the the spectral density function over a band of frequencies, such as the business cycle frequencies. In this case a set of $(1-a)\%$ confidence intervals constructed for each of n ordinates will not achieve $(1-a)\%$ joint coverage probability. This is resolved by using the *Bonferroni* method to approximate the desired coverage level, by assign $(1-a/n)\%$ coverage to each ordinate since the resulting tunnel has coverage of at least $(1-a)\%$.

A measure of fit described by the authors, which can also be used for estimation of θ is the following:

$$L(\theta, v_t) = \int_{\omega_1}^{\omega_2} L^* \left(\widehat{S_y}(\omega), S_x(\omega, \theta, v_t) \right) W(\omega) d\omega \tag{3.5}$$

where in practice, the integral is replaced with a sum over frequencies $\omega_j = 2\pi j/T$, $j = 1, ..., T/2-1$, L^* measures the distance between $\widehat{S_y}(\omega)$ and $S_x(\omega, \theta, v_t)$ and it is usually of quadratic form, $tr\left(D'(\omega_j;\theta)D(\omega_j;\theta)\right)$, where $D(\omega_j;\theta) = S_x(\omega, \theta, v_t) - \widehat{S_y}(\omega)$ and, finally, an estimate of the structural vector is given by $\hat{\theta} = \arg\max \sum_j tr\left(D'(\omega_j;\theta)D(\omega_j;\theta)\right)$.

Macroeconomic data, particularly, interest rate and exchange rate time series display nonlinear dynamic structure which cannot be captured by the suggested resampling technique; a weakness noticed by the authors.

3.2.3 Evaluation Methods based on the Sampling Variability of the Simulated Data

We study methods that construct measures of fit based on the sampling variability of the simulated data. In a classical framework this variability arises from the variability of the stochastic processes of the model while in a Bayesian framework this, normally, arises from the parameter uncertainty.

Calibration as Testing

This method provides a simple way to judge the distance between population moments or statistics of a simulated macroeconomic model, $S_x(\theta, v_t)$, and their observed counterparts, $\widehat{S_{yT}}$, where $\widehat{S_{yT}} \xrightarrow{P} S_y$. Gregory and Smith (1991, 1993) show that a wide class of measures of fit could be obtained through the randomization of the stochastic processes of the model, v_t. An estimate of the distance between $S_x(\theta, v_t)$ and $\widehat{S_{yT}}$ could be obtained by using either asymptotic or probabilistic, *Monte Carlo*, criteria.

To be precise, sequence of errors, $\left\{\left\{v_t^j\right\}_{t=1}^{T}\right\}_{j=1}^{J}$, are drawn from the hypothetical distribution and used for the calculation of $\left\{S_x^j(\theta, v_t^j)\right\}_{j=1}^{J}$. The latter sequence is ordered numerically and we check if $\widehat{S_{yT}}$ lies inside a prespecified range of the simulated distribution or we calculate the number of the replications for which $S_x^j(\theta, v_t^j) < \widehat{S_{yT}}$. According to Gregory and Smith (1991) this gives the *size* of the calibration test. If a model is a poor approximation to the data, the simulated distribution of $S_x^j(\theta, v_t^j)$ will be far away from the distribution in the data and extreme statistics will be obtained.

Since $S_x(\theta, v_t)$ is a random variable we know from the *Central Limit Theorem* that

$$\sqrt{T}\frac{S_x(\theta, v_t) - E\left(S_x(\theta, v_t)\right)}{var\left(S_x(\theta, v_t)\right)} \xrightarrow{D} N(0,1)$$

and this offers an alternative testing procedure. For each replication j we test whether or not $\sqrt{T}\frac{S_x(\theta,v_t)-E(S_x(\theta,v_t))}{var(S_x(\theta,v_t))}$ is statistically different from $N(0,1)$ and we construct a dummy variable d_j that takes value one if this is the case and zero otherwise and, finally, $J^{-1}\sum_{j=1}^{J} d_j$ gives the percentage of times the model is rejected.

Apart from Gregory and Smith (1991, 1993), Soderlind (1994) and Cogley and Nason (1994) have also used this method to evaluate their DSGE models where $S_x(\theta, v_t)$ denotes spectral densities and semi-structural responses, respectively.

Canova (1994, 1995) and Maffezzoli (2000) augment the above analysis by allowing parameter uncertainty. This additional source of sampling variability is a maneuver to the criticism that θ is known with certainty; an assumption made by

calibrationists. The density from which θ is drawn is a smooth histogram of all the estimates (both time series and cross-sectional) existing in the literature.

Simulated Quasi-Maximum Likelihood

This technique, introduced by Smith (1993), is more an estimation procedure for nonlinear DSGE models with its own measure of fit rather than an evaluation process. θ is selected in such way so the density of the simulated data reconciles with the density of the actual data, $f(x_t(\theta)/x_{t-1}(\theta), ..., x_{t-p}(\theta), \gamma) = f(y_t/y_{t-1}, ..., y_{t-p}, \gamma)$. A VAR with *identically independently distributed* (i.i.d.) residuals is chosen to approximate the true conditional density because of its computational advantages.

The mapping from θ to γ (VAR coefficients on lagged variables and the parameters of the covariance matrix) is highly nonlinear and it is calculated by simulation. To be precise, we know that for a large series of the simulated data

$$\hat{\gamma}_T(\theta) = \arg\max \sum_{s=1}^{\tau T} \log f(x_s(\theta)/x_{s-1}(\theta), ..., x_{s-p}(\theta), \gamma(\theta)) \xrightarrow{P} \gamma(\theta)$$

and under the null hypothesis there exists a θ^o such that $\gamma^o = \gamma(\theta^o)$, which is the limit of the sequences of the VAR(p) estimate using historical data, $\hat{\gamma}_T$. Therefore, the null hypothesis could be tested on the basis of the $dim(\gamma) - dim(\theta)$ over-identifying restrictions imposed by the structural model on γ

$$\mathcal{W} \equiv T(1+\tau^{-1})^{-1} (\hat{\gamma}_T - \hat{\gamma}_T(\theta))' \Sigma_{\hat{\gamma}_T}^{-1} (\hat{\gamma}_T - \hat{\gamma}_T(\theta)) \tag{3.6}$$

where $\Sigma_{\hat{\gamma}_T}$ is the Quasi Maximum Likelihood covariance matrix of

$$\hat{\gamma}_T = \arg\max \sum_{t=1}^{T} \log f(y_t/y_{t-1}, ..., y_{t-p}, \gamma) \xrightarrow{P} \gamma^o$$

and $\mathcal{W}$'s asymptotic distribution is nonstandard[7].

3.2.4 *Evaluation Methods Based on the Variability of both Actual and Simulated Data*

Evaluation procedures studied in this section construct measures of fit based on sampling and simulation variability.

[7] $\mathcal{W}$ follows a χ^2 distribution with degrees of freedom equal to $dim(\gamma) - dim(\theta)$ when the *Information Matrix Equality* holds (see White, 1994, Chapter 11).

Resampling Methods

Canova and De Nicolo (1995) evaluate the performance of a DSGE model by examining the degree of the overlap between the bootstrapped distribution of the statistic of interested, say S_y, using actual data and the simulated distribution of the statistic conditional on the structural model, $S_x(\theta, v_t)$. In this exercise, simulated and actual data are treated symmetrically, for instance, one can either ask the probability $S_x(\theta, v_t)$ falls within the $q\%$ bootstrap contour of the S_y or the probability that S_y falls within the $q\%$ contour of the distribution of $S_x(\theta, v_t)$.

Few words are necessary to be said about the way whereby the simulated distribution of $S_x(\theta, v_t)$ is obtained. Given the assumption that the joint density of θ is just the product of the marginals of each component of θ, $f(\theta/I_t) = \prod_{i=1}^{dim(\theta)} f(\theta_i/I_t)$, simple bootstrap techniques are used to derive the empirical distribution of θ. For instance the technological parameters are obtained by bootstrapping 1000 times the residuals of a regression of the consumption series on a constant. At each replication they generate a new consumption series and collect values of the mean, the standard deviation and the AR(1) coefficient which are used to construct the corresponding structural parameter. Finally, the simulated distribution of $S_x(\theta, v_t)$ is derived by drawing randomly with replacement θ from $f(\theta/I_t)$ and simulate the model.

Corradi and Swanson (2005) assess the performance of l DSGE models, simultaneously, based on how close is the joint distribution of the simulated data of model i, $F^i(u;\theta^i)$ (*the cumulative density function, c.d.f.*), to the joint distribution of the historical data, $F^o(u;\theta^o)$. According to their null hypothesis no model can provide a better approximation than model 1, the *benchmark* model. This comparison is done using a *Kolmogorov/squared error sense*-type statistic and a distributional generalization of White's (2000) *reality check*, which assesses whether at least one of the alternative models provides a more *accurate* approximation to the true cumulative distribution than does the benchmark model.

$$H_0 : \max_{i=2,\ldots,\, l} \int_U E\left(\left(F^o(u;\theta^o) - F^1(u;\theta^1)\right)^2 - \left(F^o(u;\theta^o) - F^i(u;\theta^i)\right)^2\right)\phi(u)du \leq 0 \quad (3.7)$$

$$H_A : \max_{i=2,\ldots,\, l} \int_U E\left(\left(F^o(u;\theta^o) - F^1(u;\theta^1)\right)^2 - \left(F^o(u;\theta^o) - F^i(u;\theta^i)\right)^2\right)\phi(u)du > 0 \quad (3.8)$$

The relevant statistic is $\sqrt{T}Z_{T,S}$ where $Z_{T,S} = \max_{i=2,...,I} \int_U Z_{i,T,S}(u)\phi(u)du$ with

$$Z_{i,T,S}(u) = \frac{1}{T}\sum_{t=1}^{T}\left(1\{y_t \leq u\} - \frac{1}{S}\sum_{s=1}^{S}\left(1\left\{x_s^1\left(\hat{\theta}_T^1\right) \leq u\right\}\right)\right)^2 - \frac{1}{T}\sum_{t=1}^{T}\left(1\{y_t \leq u\} - \frac{1}{S}\sum_{s=1}^{S}\left(1\left\{x_s^i\left(\hat{\theta}_T^i\right) \leq u\right\}\right)\right)^2 \quad (3.9)$$

where $1\{y_t \leq u\}$ is a indicator function and S is the length of the simulated data. The limiting distribution of the statistics is a functional over a *Gaussian* process with a covariance kernel that reflects the contribution of the structural parameter estimation error. This implies that the latter distribution is not free of nuisance parameters and critical values cannot be tabulated, however, they introduce *Moving Block Bootstrap* (MBB) procedures to overcome this problem. For instance, the bootstrapped data is produced by resampling b blocks of length m of the actual data. The structural model is fitted to the new data in order to get the bootstrapped estimate of the structural vector which is used for the new simulation of the model. If the simulation error does not vanish then the simulated data is resampling again. At each replication $\sqrt{T}Z^*_{T,S}$ is calculated and the null hypothesis is rejected when $\sqrt{T}Z_{T,S}$ is greater than the $(1-a)$th-quantile of the distribution of $\sqrt{T}Z^*_{T,S}$.

Bayesian Calibration Methods

Bayesian DSGE economists have developed their own procedures to evaluate structural models and we review these methods in this subsection. Since these are not very widespread it would be useful for the continuation of this discussion to highlight some of the key features, which are common to all of them. Measures of fit created in this category are based on the simulation and sampling variability, which arise from the uncertainty assigned on the parameters of the structural model (θ) and on the coefficients of the parametric representation of the data (γ), respectively. Since any DSGE model in the form of equations (3.1) and (3.2) (*state space representation*) can always be written as a theory restricted VAR(1), $y_t = AB(B'B)^{-1}B'y_{t-1} + Av_t$[8], a natural choice to describe the data would be a VAR(p), $y_t = \Gamma_1 y_{t-1} + ... + \Gamma_p y_{t-p} + u_t$ or using the companion matrix,

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \cdots & \Gamma_p \\ I_{dy} & 0_{dy} & \cdots & 0_{dy} \\ \vdots & \ddots & \cdots & \vdots \\ 0_{dy} & 0_{dy} & \cdots & 0_{dy} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} u_t \\ 0_{dy} \\ \vdots \\ 0_{dy} \end{bmatrix} \equiv Y_t = \Gamma Y_{t-1} + U_t$$

[8] Equation (3.2) is substituted into equation (3.1) and the *generalized inverse matrix* is used.

with $y_t = \left[I_{dy} \;\; 0_{dy} \;\; \cdots \;\; 0_{dy} \right] Y_t \equiv y_t = \Pi Y_t$. Therefore, any comparison between the model and the data can be viewed a test for the validity of the theoretical restrictions. The second moments of y_t can be easily shown that are functions of $\gamma \equiv (vec(\Gamma_1)', ..., vec(\Gamma_p)', vech(\Sigma_U)')'$, for instance $\Sigma_Y = \Gamma \Sigma_Y \Gamma' + \Sigma_U$ or $vec(\Sigma_Y) = \left(I_{(dyp)^2} - \Gamma \otimes \Gamma\right)^{-1} vec(\Sigma_U)$ and the autocorrelation function, due to *Yule-Walker equations* (see Lutkepohl, 1993, Section 2.1.4b, page:23), is a first order difference equation with solution $\Psi_Y(j) = \Gamma^J \Sigma_Y$.

The uncertainty assigned on, say, γ is illustrated by the *prior* distribution, $g(\gamma) = \prod_{i=1}^{d\gamma} g(\gamma_i)$, which, in a significant level, is subjective. This is an assumption for which Bayesian methods are criticized a lot. This happens because the posterior distribution of γ, $g(\gamma/y_t)$, which is the product of the prior and the likelihood of y_t, $g(\gamma/y_t) = g(\gamma)L(y_t/\gamma)$, is heavily influenced by this *subjective* choice when either the likelihood is very flat or the sample size is small. This simply says that in these two cases the posterior estimate will not be much different from the prior.

Most of the methods studied in this subsection use a noninformative prior for γ that preserves the stationarity condition *i.e.* the eigenvalues of Γ are less than one in absolute terms. A noninformative prior and a normal likelihood for y_t generate a *Normal-Wishart* closed form posterior distribution for γ. Therefore, the posterior distributions of Σ_Y and $\Psi(j)$ are obtained by drawing randomly γ from the above posterior and Σ_Y and $\Psi_Y(j)$ are calculated for each draw.

DeJong, Ingram, and Whiteman (1996) also make a *subjective* assumption regarding $g(\theta)$ and they derive the simulated distribution of Σ_X and $\Psi_X(j)$ by randomly drawing θ from $g(\theta)$, solving the model and calculating Σ_X and $\Psi_X(j)$. Let $g(S_{Y,\,i})$ be the data based distribution of the element i of S_Y and let $\tilde{g}(S_{X,\,i})$ be the model based distribution for $S_{X,\,i}$ then a measure of fit is given by the overlap between the two distributions

$$CIC_i = \frac{1}{1-\rho} \int_{\rho/2}^{1-\rho/2} \tilde{g}(S_{X,\,i}) d(S_{X,\,i}) \tag{3.10}$$

which is called *Confidence Interval Criterion* (CIC), where $1-\rho = \int_{\rho/2}^{1-\rho/2} g(S_{Y,\,i}) d(S_{Y,\,i})$ and $0 \leq CIC_i \leq \frac{1}{1-\rho}$. CIC_i indicates the relative proportions of $g(S_{Y,\,i})$ and $\tilde{g}(S_{X,\,i})$, which lie in the interval $[\rho/2 \; 1-\rho/2]$. Values of CIC_i close to zero indicate poor fit of the structural model, $\tilde{g}(S_{X,\,i})$ displays little overlap with $g(S_{Y,\,i})$, while, like a *rule of thumb*, the bigger the CIC_i the better the structural model. DeJong, Ingram, and Whiteman (1996) introduce also a second measure of fit

$$d_i = \frac{(E\tilde{g}(S_{X,\,i}) - Eg(S_{Y,\,i}))}{(var\; g(S_{Y,\,i}))^{1/2}} \tag{3.11}$$

which shows how different is the location of $\tilde{g}(S_{X,\,i})$ from that of $g(S_{Y,\,i})$. This is analogous to t-statistic and large values of this measure indicate that the fit of the model is poor. The drawback of this method is that it compares the posterior of the data with the predictive density of the model. The posterior of the model requires *Bayesian Maximum Likelihood* (BML) techniques and, more importantly, *Monte Carlo Markov Chain* (MCMC) integration techniques. We will see below that the same authors improve their method and eliminate this asymmetry.

Geweke (1999) suggests another way to overcome the above mentioned asymmetry. His advice is to view a DSGE model as a representation for the *population* moments of observable functions of the data, not their *sample* counterparts. This framework allows comparisons across models without using the likelihood of the data. To be precise, let $S^{\infty}_{Y,1} \equiv E(S_Y/\theta_1, \mathcal{M}_1)$, $S^{\infty}_{Y,2} \equiv E(S_Y/\theta_2, \mathcal{M}_2)$, $p(S^{\infty}_{Y,1}/\mathcal{M}_1)$ and $p(S^{\infty}_{Y,2}/\mathcal{M}_2)$ be the *population* moments of observable of a subset of the data and their densities for two alternative DSGE models $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively. The introduction of a third, econometric, model $\mathcal{M}_3$ to bridge population and sample statistics is necessary given that DSGE models have no interpretation for the observables and $p(S^{\infty}_Y/y_t, \mathcal{M}_3)$ denotes the posterior of S^{∞}_Y. From *Assumption 1* of Geweke (1999, page:20), if S^{∞}_Y is known in the context of $\mathcal{M}_3$ then $\mathcal{M}_1$ and $\mathcal{M}_2$ have nothing further to say about y ($p(y_t/S^{\infty}_Y, \mathcal{M}_1, \mathcal{M}_3) = p(y_t/S^{\infty}_Y, \mathcal{M}_2, \mathcal{M}_3) = p(y_t/S^{\infty}_Y, \mathcal{M}_3)$), we get result *Result 1* (see Geweke, 1999, page:20) $p(\mathcal{M}_1/S^{\infty}_Y, y_t, \mathcal{M}_3) = p(\mathcal{M}_1/S^{\infty}_Y, \mathcal{M}_3)$ and $p(\mathcal{M}_2/S^{\infty}_Y, y_t, \mathcal{M}_3) = p(\mathcal{M}_2/S^{\infty}_Y, \mathcal{M}_3)$ which means that if we knew S^{∞}_Y we could draw conclusion about $\mathcal{M}_1$ and $\mathcal{M}_2$ without even collecting data y_t. In fact we do not know S^{∞}_Y but we can use $\mathcal{M}_3$ to get sample version of it. However, *Assumption 2* of Geweke (1999), $p(S^{\infty}_Y/\mathcal{M}_3) \propto const, p(S^{\infty}_Y/\mathcal{M}_1, \mathcal{M}_3) = p(S^{\infty}_Y/\mathcal{M}_1)$ and $p(S^{\infty}_Y/\mathcal{M}_2, \mathcal{M}_3) = p(S^{\infty}_Y/\mathcal{M}_2)$, restricts $\mathcal{M}_3$ to say nothing about S^{∞}_Y either absolutely or relative to $\mathcal{M}_1$ and $\mathcal{M}_2$. The latter assumption leads to the following result, $p(\mathcal{M}_1/y_t, \mathcal{M}_3) \propto p(\mathcal{M}_1/\mathcal{M}_3) \int p(S^{\infty}_Y/\mathcal{M}_1) p(S^{\infty}_Y/y_t, \mathcal{M}_3) d(S^{\infty}_Y)$ that is the product of a scale factor, the prior of the model, and the convolution of the density for S^{∞}_Y implied by the DSGE model $\mathcal{M}_1$ and of its posterior obtained by using the empirical model $\mathcal{M}_3$ and the data y_t. From this a measure of fit can be constructed

$$\frac{p(\mathcal{M}_1/y_t, \mathcal{M}_3)}{p(\mathcal{M}_2/y_t, \mathcal{M}_3)} = \frac{p(\mathcal{M}_1/\mathcal{M}_3) \int p(S^{\infty}_Y/\mathcal{M}_1) p(S^{\infty}_Y/y_t, \mathcal{M}_3) d(S^{\infty}_Y)}{p(\mathcal{M}_2/\mathcal{M}_3) \int p(S^{\infty}_Y/\mathcal{M}_2) p(S^{\infty}_Y/y_t, \mathcal{M}_3) d(S^{\infty}_Y)} \tag{3.12}$$

which is called the *posterior odds ratio*, which is the product of the *prior odds ratio* and the *Bayes factor*. Loosely, speaking the greater the overlap between the two densities say $p(S^{\infty}_Y/\mathcal{M}_1)$ and $p(S^{\infty}_Y/y_t, \mathcal{M}_3)$ the grater the *Bayes factor* in favor of model $\mathcal{M}_1$.

Bayesian Maximum Likelihood Methods

DeJong, Ingram, and Whiteman (2000) solve the asymmetry, which rises in their earlier work, by deriving the posterior distribution of θ, $p(\theta/y_t, \mathcal{M}_1) = p(\theta)L(y_t/\theta, \mathcal{M}_1)$. Given $p(\theta/y_t, \mathcal{M}_1)$, they construct the predictive distributions of the model (out of sample forecasts from one to forty quarters) and compare them, through graphical methods, with those produced by a *Bayesian* VAR (BVAR) model with *Minnesota* Priors[9]. However, $p(\theta/y_t, \mathcal{M}_1)$ is analytically intractable and numerical methods, importance sampling (see Geweke, 1989), are used to analyze it. Given an importance sample and the system (3.1) and (3.2), predictive distributions are easily calculated by dynamically simulating the transition equation. It should be mentioned that a similar evaluation analysis is also undertaken by Smets and Wouters (2003).

Schorfheide (2000) constructs a method of evaluating two DSGE models by using a *benchmark* model and a *Loss Function.* This *benchmark* is a mixture of the DSGE models ($\mathcal{M}_1$, $\mathcal{M}_2$) and a reference, econometric model ($\mathcal{M}_3$), which is introduced to cope with the potential misspecification of the DSGE models. This augmentation is necessary in order the posterior distribution of the vector of some population characteristics, S_Y^{∞}, to be obtained. The posterior model probabilities, $\pi_{i,T} = \frac{\pi_{i,0}p(y_t/\mathcal{M}_i)}{\sum_{i=1}^{3}\pi_{i,0}p(y_t/\mathcal{M}_i)}$[10], indicate the mixture weights of $\mathcal{M}_1$, $\mathcal{M}_2$ and $\mathcal{M}_3$ and also their statistical fit. It should be noticed that these probabilities are obtained only after the parameter's posterior distribution, $p(\theta_i/y_t, \mathcal{M}_i)$ is obtained and this is done by using a *random walk Metropolis* algorithm. Given $p(\theta_i/y_t, \mathcal{M}_i)$, the posterior distribution of S_Y^{∞}, which is function of the structural parameter, is obtained, $p(S_Y^{\infty,i}/y_t, \mathcal{M}_i)$. In the next step of this procedure *Loss Functions*, $L\left(S_Y^{\infty}, \widetilde{S_Y^{\infty}}\right)$, are introduced to assess the ability of the ability of the DSGE models to replicate patterns observed in the reality. These functions are initially used to get an optimal predictor of $S_Y^{\infty,i}$, $\widehat{S_Y^{\infty,i}} = \arg\max_{\widetilde{S_Y^{\infty,i}} \in \mathbb{R}^{ds}} \int L\left(S_Y^{\infty}, \widetilde{S_Y^{\infty}}\right) p(S_Y^{\infty,i}/y_t, \mathcal{M}_i) d\left(S_Y^{\infty}\right)$ and given this estimate and the loss function we can evaluate the expected loss implied by $\widehat{S_Y^{\infty,i}}$, $R(\widehat{S_Y^{\infty,i}}/y_t) = \int L\left(S_Y^{\infty}, \widehat{S_Y^{\infty}}\right) p(S_Y^{\infty}/y_t) d\left(S_Y^{\infty}\right)$, where $p(S_Y^{\infty}/y_t) = \sum_{i=1}^{3}\pi_{i,T}p(S_Y^{\infty,i}/y_t, \mathcal{M}_i)$. As it is indicated by the author the posterior risk $R(\widehat{S_Y^{\infty,i}}/y_t)$ provides an absolute measure of how well model $\mathcal{M}_i$ predicts the populations characteristics S_Y^{∞}, which can be used for model comparisons.

Del Negro and Schorfheide (2004) construct a very neat measure of fit, λ, which assesses the plausibility of the restrictions imposed by the model on the data. The

[9] BVARs with *Minnesota* Priors are considered to have a remarkable forecasting performance

[10] $\pi_{i,0}$ are the prior probabilities for all models, where $i = 1, 2, 3$

cornerstone of their procedure is a model called DSGE-VAR(λ), which is an unrestricted VAR with priors coming from the structural model. To be precise, these priors are obtained by generating dummy observations from the DSGE model and adding them to the actual data. The ratio of dummy over actual observations, λ, measures the weight of the prior relative to the sample. After a proper factorization of the posterior distribution of the parameters of the DSGE-VAR(λ) model, $p(\gamma, \theta/y_t) = p(\gamma/y_t, \theta)p(\theta/y_t)$, it can be shown that the posterior mean of γ is a function of λ. For instance, let a VAR(p), $y_t = A_1 y_{t-1} + ... + A_p y_{t-p} + u_t$, be written as $Y = X\Phi + U$ where $\Phi = [A_1, ..., A_p]$ and $x_t = [1, y'_{t-1}, ..., y_{t-p}]'$ and let $\Sigma^*_{xx}(\theta) = E_\theta[x_t x'_t]$ and $\Sigma^*_{yx}(\theta) = E_\theta[y_t x'_t]$ be the second moments generated by the dummy observations. Del Negro and Schorfheide show that the posterior mean of Φ is $\tilde{\Phi}(\theta) = (\lambda T\Sigma^*_{xx}(\theta) + X'X)^{-1}(\lambda T\Sigma^*_{yx}(\theta) + X'Y)$ and the posterior mean of Σ_U is $\tilde{\Sigma}_U(\theta) = \frac{1}{T}\left[(\lambda T\Sigma^*_{yy}(\theta) + Y'Y) - (\lambda T\Sigma^*_{yx}(\theta) + Y'X)(\lambda T\Sigma^*_{xx}(\theta) + X'X)^{-1}(\lambda T\Sigma^*_{xy}(\theta) + X'Y)\right]$. Obviously for $\lambda = 0$ the posterior mean of Φ conditional on θ is given by the OLS estimator while for large values of λ the posterior mean of Φ will stay close to the restrictions implied by the DSGE model. A data-driven procedure, which is the maximization of the marginal data density $\int p_\lambda(y_t/\theta)p(\theta)d(\theta)$ with respect to λ over some grid $\Lambda = \{\lambda_1, ..., \lambda_q\}$, is suggested to determine an appropriate value for λ. Given this value, $\hat{\lambda}$, the DSGE-VAR($\hat{\lambda}$) is compared with unrestricted VARs. Similar analysis has also be undertaken by Del Negro and Schorfheide (2005) and Del Negro, Schorfheide, Smets, and Wouters (2004).

Fernandez-Villaverde and Rubio-Ramirez (2004) do not really introduce a new evaluation procedure. They assess DSGE models in a standard *Bayesian* way (posterior odds ratio, BVAR), however, the structural model needs not being linearized now. In a earlier work Fernandez-Villaverde and Rubio-Ramirez (2002) have developed an MCMC algorithm that allows to construct the posterior distribution of the parameter vector of a nonlinear structural model (this is done by using *Sequential Monte Carlo filtering* in order to evaluate the likelihood of a nonlinear DSGE model), $p(\theta/y_t)$. Given the latter distribution any evaluation exercise studied in this subsection could be contacted.

4. EVALUATING MACROECONOMIC MODELS: A DETAILED DISCUSSION

In this chapter we go through the steps in our method in more detail, reviewing particular issues that arise at each stage. In the first part the proposed evaluation procedure is decomposed into three simple steps. Initially, it is illustrated how abstraction error is obtained using perturbation solution methods. In the second step the probability density function (p.d.f.) of the above errors, pseudo data and statistic of interest is derived though bootstrapping schemes and simulation of the structural model respectively. The final step is the one from which the inference regarding the structural model is drawn. Given the existence of macroeconomic softwares like Dynare[1], or, for more computationally advanced users, Sims' solution package[2] this decomposition makes clear that the proposed method can be applied by researchers with minimal computational skills.

The second part of this chapter is again divided into two sections. In the first we illustrate how the true but unknown data generating process (DGP, that is here the process that describes the time series behavior of the data) may be approximated. In the second we show how methods studied in the last chapter, which construct measures of fit based on sampling or simulation variability or both, can be encompassed by our method. We can illustrate these two points from our example in chapter 1 when we tested the LVP model. In it, the policy regime varies over the sample period. Hence, the DGP chosen to describe the data over this period should be able to capture this feature. In our exercise three alternative time series models were used whose parameter vectors are, in the first case, time dependent and, for the other two cases, regime dependent both in a stochastic and deterministic way. We then go on to show how the time series model, which provides closed form expressions for the population moments, can be compared with the structural model in a way that encompasses the existing methods; we do this using both time and frequency domain techniques.

The implementation of the Wald type statistic, discussed in the first chapter, is described in the final part. Its relation with the existing methods and its sensitivity

[1] See http://www.cepremap.cnrs.fr/dynare/
[2] See http://sims.princeton.edu/yftp/gensys2/

regarding the choice of the time series model is investigated there. The inference about the structural model does not rely on asymptotic theory but on the, numerically approximated, small sample distribution of the above statistic. The introduced methodology relies heavily on the properties of the Quasi Maximum Likelihood Estimator (QMLE) whose variance-covariance matrix is calculated numerically. We discuss the use of various section closed form expressions for the score vector and the Hessian matrix of an Intervention[3] VAR(p) and for the scores of a Markov Switching VAR(p) are provided. These expressions eliminate numerical instabilities and significantly reduce the approximation error.

4.1 The Proposed Procedure

4.1.1 A Detailed Description of the Assessment Process

The set of equilibrium conditions of a wide set of DSGE models in macroeconomics form a system of nonlinear rational expectations equations. For instance, the equilibrium conditions of the simple neoclassical growth model are given by

$$c_t^{-\varphi} = \beta E_t c_{t+1}^{-\varphi}\left[A_{t+1}k_{t+1}^{a-1} + 1 - \delta\right] \tag{4.1}$$

$$c_t + k_{t+1} = A_t k_t^a + (1-\delta)k_t \tag{4.2}$$

$$\ln A_{t+1} = \gamma \ln A_t + \sigma\epsilon_{t+1} \tag{4.3}$$

which can be rewritten as

$$\Psi\left(s_{t+1}, s_t, \sigma\varepsilon_{t+1}\right) + \Lambda\sigma\eta_{t+1} \equiv \begin{bmatrix} s_{1,t}^{-\varphi} - \beta\left(s_{1,t+1}^{-\varphi} + \sigma\eta_{t+1}\right)\left[ae^{s_{3,t+1}}s_{2,t+1}^{a-1} + 1 - \delta\right] \\ s_{1,t} + s_{2,t} - e^{s_{3,t}}s_{2,t}^a - (1-\delta)\,s_{2,t} \\ s_{3,t+1} - \gamma s_{3,t} - \sigma\varepsilon_{t+1} \end{bmatrix} = 0 \tag{4.4}$$

Following Kim, Kim, Schaumburg, and Sims (2005) we do not discriminate between *predetermined*, $x_t \in \mathbb{R}^{dx}$ $\left(x_t = [k_t, \ln A_t]' \text{ in this example }\right)$, and *non-predetermined*, $y_t \in \mathbb{R}^{dy}$ ($y_t = c_t$ in this example), variables, $s_t = [y_t,\ x_t]'$.

One of the most important features of the evaluation process proposed here is that it tries to preserve the nonlinear properties of the structural model. For this purpose Perturbation Solution Methods, second order approximations to the policy function, are used. This is one of the two most popular nonlinear ways of solving DSGE models; the other is Projection Methods. Further research, which investigates how Projection Methods .

[3] See below

Step 1: Determination of the approximation or abstraction error of the economic model conditional on the actual data.

In this step the structural model is solved using a second order approximation to the policy function and the error implied by the theory conditional on the historical data is calculated, $y_t - y_t(\theta) = \varepsilon_t$. To be precise, the solution of equation (4.4) has the following form

$$s_t = \Phi(s_{t-1}, \varepsilon_t; \theta) \tag{4.5}$$

A second order *Taylor Expansion* applied on (4.5) around a deterministic steady state, $(\bar{s}, 0)$, gives

$$\underset{ds\times 1}{\tilde{s}_t} = \underset{ds\times(ds+d\varepsilon)}{\left[\underset{ds\times ds}{\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)} \quad \underset{ds\times d\varepsilon}{\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)}\right]}\underset{(ds+d\varepsilon)\times 1}{\begin{bmatrix}\tilde{s}_{t-1}\\ \varepsilon_t\end{bmatrix}} + \frac{1}{2}$$

$$\underset{ds\times((ds+d\varepsilon)ds)}{\left(\underset{1\times(ds+d\varepsilon)}{\begin{bmatrix}\tilde{s}'_{t-1} & \varepsilon'_t\end{bmatrix}}\otimes I_{ds}\right)}\underset{((ds+d\varepsilon)ds)\times(ds+d\varepsilon)}{\begin{bmatrix}\underset{ds^2\times(ds+d\varepsilon)}{\begin{bmatrix}\underset{ds^2\times ds}{\nabla_s\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)} & \underset{ds^2\times d\varepsilon}{\nabla_\varepsilon\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)}\end{bmatrix}}\\ \underset{(ds\times d\varepsilon)\times(ds+d\varepsilon)}{\begin{bmatrix}\underset{(ds\times d\varepsilon)\times ds}{\nabla_\varepsilon\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)} & \underset{(ds\times d\varepsilon)\times d\varepsilon}{\nabla_\varepsilon\left(vec\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)}\end{bmatrix}}\end{bmatrix}}$$

$$\underset{(ds+d\varepsilon)\times 1}{\begin{bmatrix}\tilde{s}_{t-1}\\ \varepsilon_t\end{bmatrix}} \tag{4.6}$$

Variables and derivatives, which are expressed as deviations or evaluated, respectively, at the steady state are denoted with $\sim$. The use of the *vec* operator allows us to avoid the use of both *tensor* notation, usually employed in this literature and matrices with more than two arrays; the matrix of the second derivatives of $f : \mathbb{R}^{dx} \to \mathbb{R}^{dy}$ has three dimensions, namely, $dy \times dx \times dx$. I believe the way whereby (4.6) is derived provides a *comparative advantage* to *tensor* notation, which will be

shown immediately. (4.6) is rewritten as

$$\begin{aligned}\tilde{s}_t &= \left[\begin{array}{cc} \nabla_s\Phi\left(\tilde{s}_{t-1},\varepsilon_t;\theta\right) & \nabla_\varepsilon\Phi\left(\tilde{s}_{t-1},\varepsilon_t;\theta\right)\end{array}\right]\left[\begin{array}{c}\tilde{s}_{t-1}\\ \varepsilon_t\end{array}\right]\\ &+\frac{1}{2}\left(\left[\begin{array}{cc} s'_{t-1} & \varepsilon'_t\end{array}\right]\otimes\left(\left(\left[\begin{array}{cc} s'_{t-1} & \varepsilon'_t\end{array}\right]\otimes I_{ds}\right)\right)\right)\left[\begin{array}{c} vec\left(\nabla_s\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)\\ vec\left(\nabla_\varepsilon\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)\\ vec\left(\nabla_s\left(vec\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)\\ vec\left(\nabla_\varepsilon\left(vec\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)\end{array}\right]\end{aligned}$$

or,

$$\begin{aligned}\tilde{s}_t &= \left[\begin{array}{cc} \nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right) & \nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\end{array}\right]\left[\begin{array}{c}\tilde{s}_{t-1}\\ \varepsilon_t\end{array}\right]\\ &+\frac{1}{2}\left(\left[\begin{array}{cccc} \underset{1\times(ds\times ds)}{\left(\tilde{s}'_{t-1}\otimes\tilde{s}'_{t-1}\right)} & \underset{1\times(ds\times d\varepsilon)}{\left(\tilde{s}'_{t-1}\otimes\varepsilon'_t\right)} & \underset{1\times(ds\times d\varepsilon)}{\left(\varepsilon'_t\otimes\tilde{s}'_{t-1}\right)} & \underset{1\times(d\varepsilon\times d\varepsilon)}{\left(\varepsilon'_t\otimes\varepsilon'_t\right)}\end{array}\right]\otimes I_{ds}\right)\\ &\left[\begin{array}{c} \underset{ds^3\times 1}{vec\left(\nabla_s\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)}\\ \underset{((ds\times d\varepsilon)ds)\times 1}{vec\left(\nabla_\varepsilon\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)}\\ \underset{((ds\times d\varepsilon)ds)\times 1}{vec\left(\nabla_s\left(vec\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)}\\ \underset{((ds\times d\varepsilon)d\varepsilon)\times 1}{vec\left(\nabla_\varepsilon\left(vec\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)}\end{array}\right]\end{aligned}\tag{4.7}$$

From (4.7) we get

$$\begin{aligned}\tilde{s}_t &= \nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\tilde{s}_{t-1}+\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\varepsilon_t\\ &+\frac{1}{2}\left(\nabla_s\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)'\left(\tilde{s}_{t-1}\otimes\tilde{s}_{t-1}\right)\\ &+\left(\nabla_\varepsilon\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)'\left(\tilde{s}_{t-1}\otimes\varepsilon_t\right)\\ &+\frac{1}{2}\nabla_\varepsilon\left(vec\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\left(\varepsilon_t\otimes\varepsilon_t\right)\\ &= \nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\tilde{s}_{t-1}+\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\varepsilon_t\\ &+\frac{1}{2}\left(\nabla_s\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)'\left(\tilde{s}_{t-1}\otimes\tilde{s}_{t-1}\right)\\ &+\frac{1}{2}\nabla_\varepsilon\left(vec\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\left(\varepsilon_t\otimes\varepsilon_t\right)\\ &+\left(\nabla_\varepsilon\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)'\left(\tilde{s}_{t-1}\otimes I_{d\varepsilon}\right)\left(1\otimes\varepsilon_t\right)\end{aligned}$$

$$
\begin{aligned}
&= \nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)s_{t-1}+\frac{1}{2}\left(\nabla_s\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)'\left(\tilde{s}_{t-1}\otimes\tilde{s}_{t-1}\right)\\
&\quad+\frac{1}{2}\nabla_\varepsilon\left(vec\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\left(\varepsilon_t\otimes\varepsilon_t\right)+\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\varepsilon_t\\
&\quad+\left(\nabla_\varepsilon\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)'\left(\tilde{s}_{t-1}\otimes I_{d\varepsilon}\right)\varepsilon_t\\
&= \nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\tilde{s}_{t-1}+\frac{1}{2}\left(\nabla_s\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)'\left(\tilde{s}_{t-1}\otimes\tilde{s}_{t-1}\right)\\
&\quad+\frac{1}{2}\nabla_\varepsilon\left(vec\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)vec\left(\varepsilon_t\varepsilon_t'\right)\\
&\quad+\left(\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)+\left(\nabla_\varepsilon\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)'\left(\tilde{s}_{t-1}\otimes I_{d\varepsilon}\right)\right)\varepsilon_t\\
&= \nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\tilde{s}_{t-1}+\frac{1}{2}\left(\nabla_s\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)'\left(\tilde{s}_{t-1}\otimes\tilde{s}_{t-1}\right)\\
&\quad+\left[\begin{array}{cc}\frac{1}{2}\nabla_\varepsilon\left(vec\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right) & \left(\begin{array}{c}\nabla_\varepsilon\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\\+\left(\nabla_\varepsilon\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)'\left(\tilde{s}_{t-1}\otimes I_{d\varepsilon}\right)\end{array}\right)\end{array}\right]\\
&\quad\left[\begin{array}{c}vec\left(\varepsilon_t\varepsilon_t'\right)\\ \varepsilon_t\end{array}\right]\\
&= \nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\tilde{s}_{t-1}+\frac{1}{2}\left(\nabla_s\left(vec\nabla_s\tilde{\Phi}\left(s_{t-1},\varepsilon_t;\theta\right)\right)\right)'\left(\tilde{s}_{t-1}\otimes\tilde{s}_{t-1}\right)\\
&\quad+\Lambda\left[\begin{array}{c}\varepsilon_t\otimes\varepsilon_t\\ \varepsilon_t\end{array}\right] \qquad (4.8)
\end{aligned}
$$

Initially, it should be emphasized that the normality assumption, usually, adopted regarding the distribution of $\tilde{s}_t$ does not seem acceptable. Clearly, from equation (4.8) $\tilde{s}_t$, conditional on $\tilde{s}_{t-1}$, follows a *non-central chi-squared* distribution; a characteristic also highlighted by Sungbae (2005, page:15). Equation (4.8) is very important for the method proposed here for two reasons. Given the historical data the *abstraction* theory error, ε_t, can be derived, while, in the next step, given the empirical distribution of the latter, F_ε, the distribution of $\tilde{s}_t$, F_s, conditional on the structural model is obtained. Now, $\tilde{y}_{t+1}$ is just a selection of $\tilde{s}_{t+1}$, for simplicity and without loss of generality let's assume that $y_{t+1}=\tilde{s}_{t+1}$ and by equation (4.8), we

get

$$\Lambda \begin{bmatrix} \varepsilon_t \otimes \varepsilon_t \\ \varepsilon_t \end{bmatrix} = y_{t+1} - \begin{bmatrix} \nabla_s \tilde{\Phi}(s_{t-1}, \varepsilon_t; \theta)\, \tilde{s}_{t-1} \\ +\frac{1}{2}\left(\nabla_s \left(vec \nabla_s \tilde{\Phi}(s_{t-1}, \varepsilon_t; \theta)\right)\right)' (\tilde{s}_{t-1} \otimes \tilde{s}_{t-1}) \end{bmatrix}$$

$$\begin{bmatrix} \varepsilon_t \\ 1 \end{bmatrix} \otimes \varepsilon_t = \Lambda^{+} \left[y_{t+1} - \begin{bmatrix} \nabla_s \tilde{\Phi}(s_{t-1}, \varepsilon_t; \theta)\, \tilde{s}_{t-1} \\ +\frac{1}{2}\left(\nabla_s \left(vec \nabla_s \tilde{\Phi}(s_{t-1}, \varepsilon_t; \theta)\right)\right)' (\tilde{s}_{t-1} \otimes \tilde{s}_{t-1}) \end{bmatrix} \right]$$

$$\left(\begin{bmatrix} \underset{(1 \times d_e)}{\mathbf{0}} & 1 \end{bmatrix} \otimes I_{d\varepsilon} \right) \left(\begin{bmatrix} \varepsilon_t \\ 1 \end{bmatrix} \otimes \varepsilon_t \right) = \left(\begin{bmatrix} \underset{(1 \times d_e)}{\mathbf{0}} & 1 \end{bmatrix} \otimes I_{d\varepsilon} \right)$$

$$\Lambda^{+} \left[y_{t+1} - \begin{bmatrix} \nabla_s \tilde{\Phi}(s_{t-1}, \varepsilon_t; \theta)\, \tilde{s}_{t-1} \\ +\frac{1}{2}\left(\nabla_s \left(vec \nabla_s \tilde{\Phi}(s_{t-1}, \varepsilon_t; \theta)\right)\right)' (\tilde{s}_{t-1} \otimes \tilde{s}_{t-1}) \end{bmatrix} \right]$$

$$\varepsilon_t = \left(\begin{bmatrix} \underset{(1 \times d_e)}{\mathbf{0}} & 1 \end{bmatrix} \otimes I_{d\varepsilon} \right)$$

$$\Lambda^{+} \left[y_{t+1} - \begin{bmatrix} \nabla_s \tilde{\Phi}(s_{t-1}, \varepsilon_t; \theta)\, \tilde{s}_{t-1} \\ +\frac{1}{2}\left(\nabla_s \left(vec \nabla_s \tilde{\Phi}(s_{t-1}, \varepsilon_t; \theta)\right)\right)' (\tilde{s}_{t-1} \otimes \tilde{s}_{t-1}) \end{bmatrix} \right]$$

The crucial assumption here is that ε_t has the same dimension as $\tilde{y}_t$, $dy = d\varepsilon$, this eliminates the stochastic singularity issue mentioned in the last chapter. Additional, the null hypothesis is imposed and $+$ denotes the *generalized inverse* matrix, again mentioned in the last chapter.

These are two of the points where the proposed process departs from existing evaluation methods. First, instead of using the theoretical errors drawn from the hypothetical distribution to assess the performance of the structural model, the proposed process uses the errors implied by the actual data. Second, the nonlinearity of the DSGE model is preserved. This will be more obvious in the following steps.

Step 2: Approximation of the distributions of interest conditional on the structural model or conditional on the null hypothesis.

According to the null hypothesis $\{\varepsilon_t\}_{t=1}^{T}$ should not contain any structure, in other words they should be *identically independently distributed* (i.i.d). In order to ensure this $\{\varepsilon_t\}_{t=1}^{T}$ are properly *filtered.* Given the *filtered* errors their empirical distribution, F_ε, is readily estimated by using standard resampling techniques for i.i.d data. Let the bootstrapped errors be denoted by $\left\{\left\{\varepsilon_t^{*,j}\right\}_{t=1}^{T}\right\}_{j=1}^{J}$. Given equation (4.8) the empirical distribution of $\tilde{y}_t$ conditional of the structural model $\mathcal{M}$, $f(\tilde{y}_t/\mathcal{M})$, can be approximated through simulation. Let $\left\{\left\{\tilde{y}_t^{*,j}\right\}_{t=1}^{T}\right\}_{j=1}^{J}$ stands for the new pseudo data sets. Although resampling techniques are used to approximate F_ε, during the simulation of the structural model the hypothetical variance-covariance matrix of

ε_t, Σ_ε, is used. This oddity could be justified by the fact that Σ_ε is part of the hypothetical world $\mathcal{M}$, strictly speaking part of the structural parameter vector θ, and, therefore, an important determinant of ε_t.

Given $\left\{\left\{\tilde{y}_t^{*,j}\right\}_{t=1}^{T}\right\}_{j=1}^{J}$ the distribution of any statistic or population moments of interest, $S_{\tilde{y}}$, conditional on $\mathcal{M}$, $f\left(S_{\tilde{y}}/\mathcal{M}\right)$, is easily derived. We cam clarify this point through an example. It is well known that the conditional distribution of the actual data is well approximated by a VAR with i.i.d. residuals. Using the companion matrix a VAR(p), $\tilde{y}_t = A_1\tilde{y}_{t-1} + ... + A_p\tilde{y}_{t-p} + u_t$, can be written as VAR(1), $\tilde{Y}_t = A_1\tilde{Y}_{t-1} + U_t \equiv \begin{bmatrix} \tilde{y}_t \\ \tilde{y}_{t-1} \\ \vdots \\ \tilde{y}_{t-p+1} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 & \cdots & A_p \\ I_{dy} & 0_{dy} & \cdots & 0_{dy} \\ \vdots & \ddots & \cdots & \vdots \\ 0_{dy} & 0_{dy} & \cdots & 0_{dy} \end{bmatrix} \begin{bmatrix} \tilde{y}_{t-1} \\ \tilde{y}_{t-2} \\ \vdots \\ \tilde{y}_{t-p} \end{bmatrix} + \begin{bmatrix} u_t \\ 0_{dy} \\ \vdots \\ 0_{dy} \end{bmatrix}$ and the second moments are given by $\Sigma_Y = A\Sigma_Y A' + \Sigma_U$ or $vec(\Sigma_Y) = \left(I_{(dyp)^2} - A \otimes A\right)^{-1} vec(\Sigma_U)$ and $\Psi_Y(i) = A^i\Sigma_Y$. Given $\left\{\left\{\tilde{y}_t^{*,j}\right\}_{t=1}^{T}\right\}_{j=1}^{J}$, $f\left(\Sigma_Y/\mathcal{M}\right)$ and $f\left(\Psi_Y(i)/\mathcal{M}\right)$ are derived by fitting at each $\left\{\tilde{y}_t^{*,j}\right\}_{t=1}^{T}$ a VAR(p) and calculating $\Sigma_Y^{*,j}$ and $\Psi_Y(i)^{*,j}$ by using the above given equations.

The benefit of following this process is that we can establish $f\left(S_{\tilde{y}}/\mathcal{M}\right)$ for any time series model, $\mathcal{T}$, without requiring the mapping from the structural, θ, to reduced from, γ, parameters to be established analytically, $\theta : \mathcal{Q} \rightarrow \gamma$. For non-homogeneous historical data $\mathcal{T}$ should be able to capture these nonlinearities which implies that $\mathcal{T}$ is probably nonlinear and the only way to establish $\mathcal{Q}$ is via simulation (this idea is similar to the one introduced by Smith, 1993). For instance, if interest or exchange rates are elements of $\tilde{y}_t$ then a linear VAR(p) is not sufficient to capture the higher order dynamics, usually, displayed by these series. This requires $\mathcal{T}$ being a VAR(p) with a time varying covariance matrix for its error vector. In this cases $\mathcal{Q}$ is derived by fitting $\mathcal{T}$ into $\left\{\tilde{y}_t^{*,j}\right\}_{t=1}^{T}$ and then $\gamma_T^{*,j} = \arg\max_{\gamma^{*,j}\in\Gamma} L^{\mathcal{T}}\left(\tilde{y}_t^{*,j}/\gamma^{*,j},\ \mathcal{M}\right)$ is used for the calculation of $S_{\tilde{y}}^{*,j}$, where $L^{\mathcal{T}}$ is the log-likelihood using $\mathcal{T}$.

Step 3: Inference Step.

We show in this step that given $\left\{S_{\tilde{y}}^{*,j}\right\}_{j=1}^{J}$ almost all methods, except the *Bayesian* one, studied in the last chapter can be encompassed by the one proposed here.

We start with the methods that use the sample variability to assess the perfor-

mance of the structural model such as the one proposed by Diebold, Ohanian, and Berkowitz (1998) and reviewed in the last chapter. It was mentioned in the preceding step that the time series model $\mathcal{T}$ is selected with the aim of summarizing the dynamic structure of the actual data satisfactorily and standard econometric tests, like the Information Matrix Equality test (see White, 1994, Chapter 11), could be used to investigate if this holds. In this case the residuals of T are i.i.d and standard resampling schemes are applied to establish the sampling distribution of $\tilde{S}_{\tilde{y}}^{*,j}$ after $\tilde{\gamma}_T^{*,j} = \arg\max_{\tilde{\gamma}^{*,j}\in\Gamma} L^{\mathcal{T}}\left(\tilde{y}_t^{*,j}/\gamma^{*,j}\right)$ is obtained. Then we investigate whether or not $\frac{1}{J}\sum_{j=1}^{J} S_{\tilde{y}}^{*,j}$ lies within a prespecified confidence interval, say 95%, of $f\left(\tilde{S}_{\tilde{y}}/\mathcal{T}\right)$.

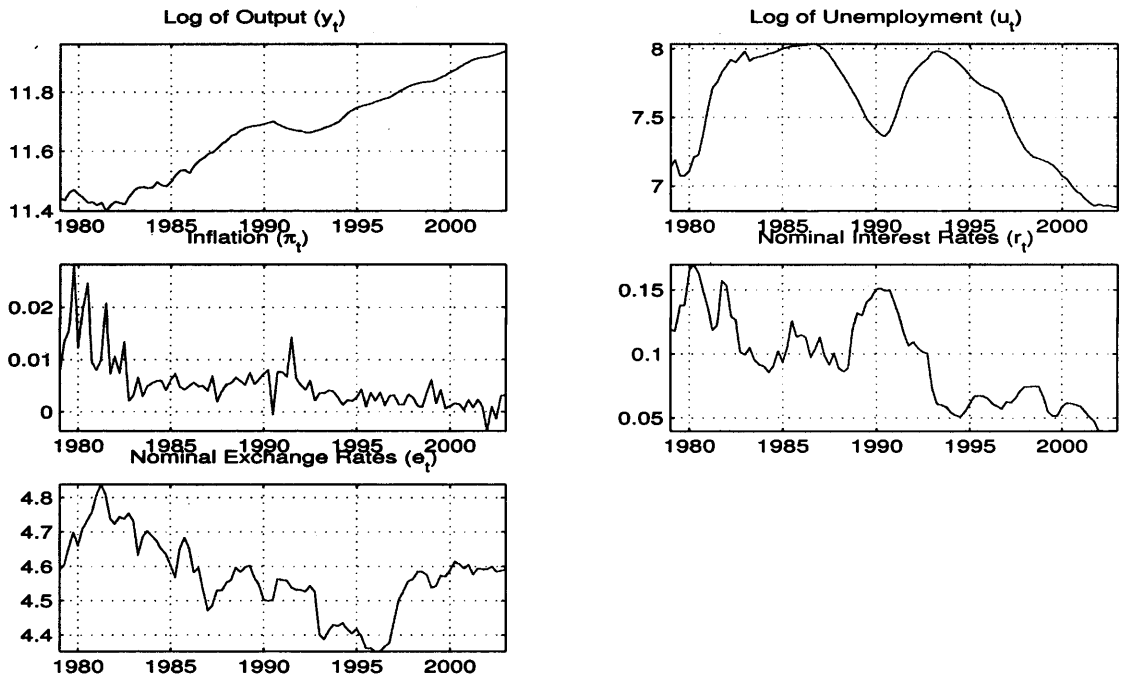
In a similar manner it can be shown that the proposed method encompasses those which construct measures of fit based on the simulation variability. Given $\mathcal{T}$ and $\hat{\gamma}_T = \arg\max_{\gamma\in\Gamma} L^{\mathcal{T}}(y_t\ ;\gamma)$, $\tilde{S}_{\tilde{y}}$ can be readily calculated and, now, we investigate whether or not $\tilde{S}_{\tilde{y}}$ lies within a prespecified interval of $f\left(S_{\tilde{y}}/\mathcal{M}\right)$ given by *step 2.* We can also see from this that the methods of the third category, measures of fit based on both simulation and sampling variability, are also encompassed. Examples for all these three cases are given in the following section.

In contrast to these methods, reviewed in the last chapter, the one suggested here respects the nonlinearity of the structural model and therefore the nonlinearities in the relationships describing the data. Structural nonlinearities may be important for explaining features of the historical data and, therefore, should be tested. On the other hand nonlinearities within the actual data processes are equally important and should not be ignored. For instance, consider a sequence of time series model, $\{\mathcal{T}_i\}_{i=1}^{K}$, which starts from a very simple linear time series model $\mathcal{T}_1$, say a VAR(p), and it goes up to a highly nonlinear model for which the *Information Matrix Equality* cannot be rejected $\mathcal{T}_K$, say a VAR(p) whose parameters are state dependent and which follows a *Markov Switching* process. Applying the proposed method sequentially, starting from $\mathcal{T}_1$ up to $\mathcal{T}_K$, we could identify what characteristics of the real world the structural model, $\mathcal{M}$, fails to explain. This type of information may be useful to the researchers who are interested to improve the fit of their models.

4.1.2 Some Exercises

In this section we illustrate from the Liverpool Model exercise (see chapter 1) how the evaluation methods existing so far in the literature could be encompassed by the one proposed here. The Liverpool Model of UK economy and UK data between 1979 and 2003 for output (GDP at factor cost), unemployment, inflation (Retail Price Index), nominal interest rates (Nominal deposit interest rate with local authorities, 3 month) and nominal exchange rates (Trade-weighted exchange rate)

Fig. 4.1: **Series Representation**



have been used in this exercise (see Minford, Meenagh, and Webb, 2004, for a detailed description of the model and the data used here).

The policy regime followed during the sample period was not constant. It switched, at 1986, from a monetary targeting to exchange rate targeting regime and, finally, from 1992 to an inflation targeting regime. Therefore, time series representations, $\mathcal{T}$, able to capture these changes should be used. Three different $\mathcal{T}$s were used in this study, the first allowing the parameter set γ to vary over the time, in the second model γ follows a Markov process and, finally, γ changes at some prespecified points. The length of $y_t \in R^5$ is large and very quickly the number of the estimated parameters becomes huge while, on the other hand, the macroeconomic data is limited. Due to this data limitation the *Information Matrix Equality Test* (see White, 1994, Chapter 11) cannot be applied and the choice of the most plausible $\mathcal{T}$ is founded on the basis of some diagnostic tests on the estimated residuals, selection criteria and computational simplicity.

The parameters of the first time series model follow a VAR(1) process

$$\begin{aligned} \Delta x_t &= c_t + A_{1,t}\Delta x_{t-1} + u_t = \begin{bmatrix} c_t & A_{1,t} \end{bmatrix} \begin{bmatrix} 1 \\ \Delta x_{t-1} \end{bmatrix} + u_t \qquad (4.9) \\ &= B_t z_t + u_t = (z_t' \otimes I_{dx})\,\beta_t + u_t \\ \beta_t &= v + F\beta_{t-1} + v_t \qquad (4.10) \end{aligned}$$

where $u_t \sim N(0, \Sigma_{u,t})$, $v_t \sim N(0, \Sigma_v)$, $E(u_t v_t') = 0$. $u_t = \Sigma_{u,t}^{-1/2} w_t$ is conditional heteroskedastic aiming to capture clusters in volatility, usually displayed by interest and exchange rate series or, in economic terms, asymmetries in agents' risk preferences toward policy regime changes. The estimation process of this model explores the *special* structure of its *Information Matrix*; it is *block diagonal.* In the first step equations (4.9) and (4.10) are estimated by using *Kalman Filter* techniques (see Lutkepohl, 1993; Hamilton, 1994, Chapter 13) while in the second step a *Dynamic Conditional Correlation Multivariate GARCH* is estimated

$$\Sigma_{u,t} = D_t R_t D_t \qquad (4.11)$$

where D_t is a diagonal matrix of time varying standard deviations for univariate GARCH models and $R_t = Q_t^{*-1} Q_t\, Q_t^{*-1}$ is the time varying correlation matrix, where $Q_t = (1 - \alpha - g)\,\bar{Q} + a\left(u_{t-1}u_{t-1}'\right) + gQ_{t-1}$ and Q_t^* is a diagonal matrix composed by the square root of the diagonal elements of Q_t (see Engel and Sheppard, 2001). Since the number of the estimated parameters quickly becomes large F and Σ_v are constrained to be *diagonal* matrices. The fit of this model is illustrated by the following diagrams, which present the evolution of the mean and the normalized *power spectrum*,

$$\frac{2\pi \mathbf{s}_t^x(\omega)}{\int_{-\pi}^{\pi} \mathbf{s}_t^x(\omega)} \qquad (4.12)$$

where

$$\mathbf{s}_t^x = (2\pi)^{-1}(I_{dx} - A_{1,\tau}e^{-i\omega})^{-1}\Sigma_{u,\tau}(I_{dx} - A_{1,\tau}'e^{i\omega})^{-1} \qquad (4.13)$$

$A_{1,\tau}$ and $\Sigma_{u,\tau}$ denoted the smoothed estimates of $A_{1,t}$ and $\Sigma_{u,t}$, respectively. Equation (4.12), which is used by Cogley, Morozov, and Sargent (2003) as inflation persistence measure, would be equal to one for a white noise process.

Figures 4.2 to 4.2 show actual and predicted mean values by regime for all five selected variables, and also the normalized power spectrum. Let us analyze the behavior of the inflation change $\Delta\pi_t$ in Figure 4.4 (the analysis of the rest of the series follows similarly). From the pattern of the *fitted mean* we can see significant variation during the *Monetary Targeting Regime*, which dies away in the second

Fig. 4.2: **TVP-VAR(1)/DCC(1,1)-GARCH(1,1):** Δy_t

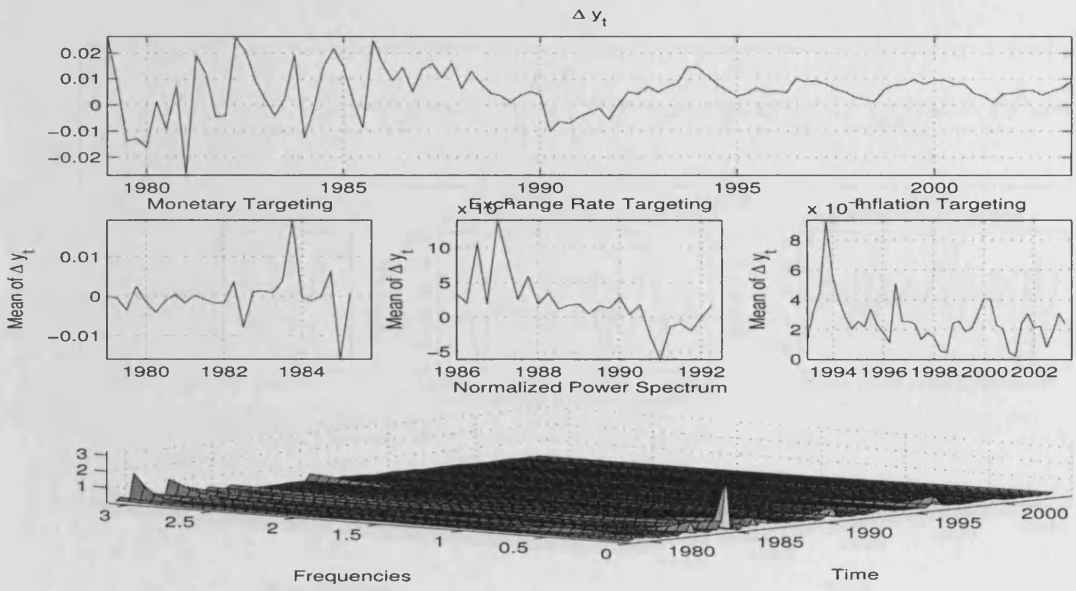


Fig. 4.3: **TVP-VAR(1)/DCC(1,1)-GARCH(1,1):** Δu_t

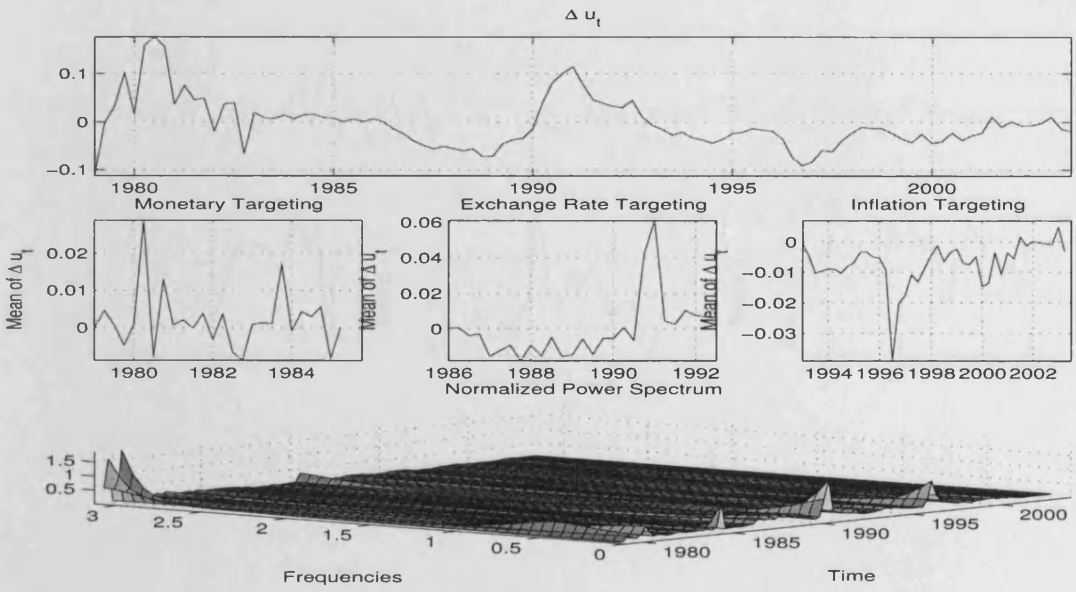


Fig. 4.4: **TVP-VAR(1)/DCC(1,1)-GARCH(1,1):** $\Delta\pi_t$

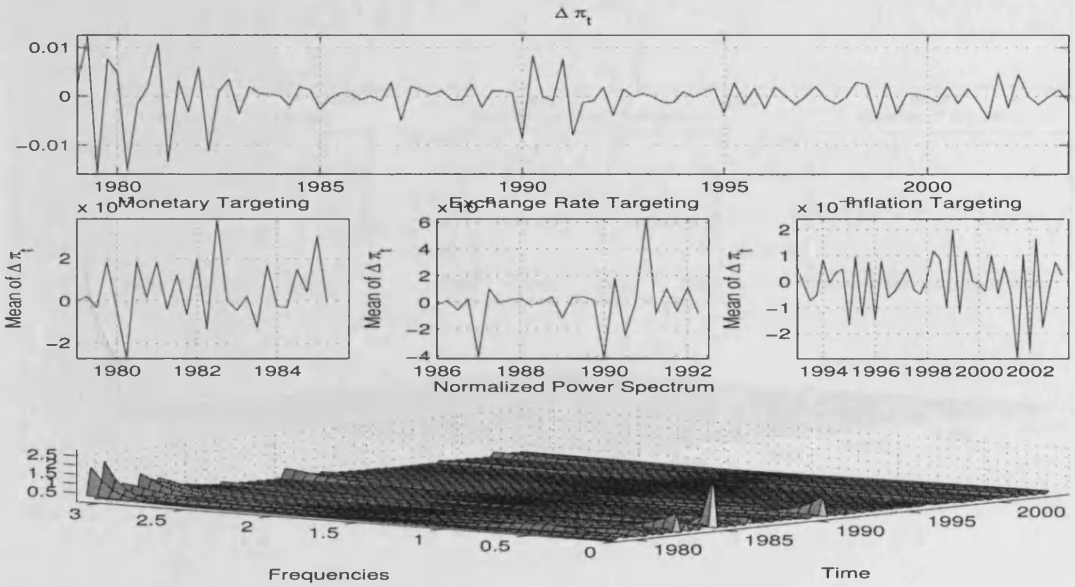


Fig. 4.5: **TVP-VAR(1)/DCC(1,1)-GARCH(1,1):** Δr_t

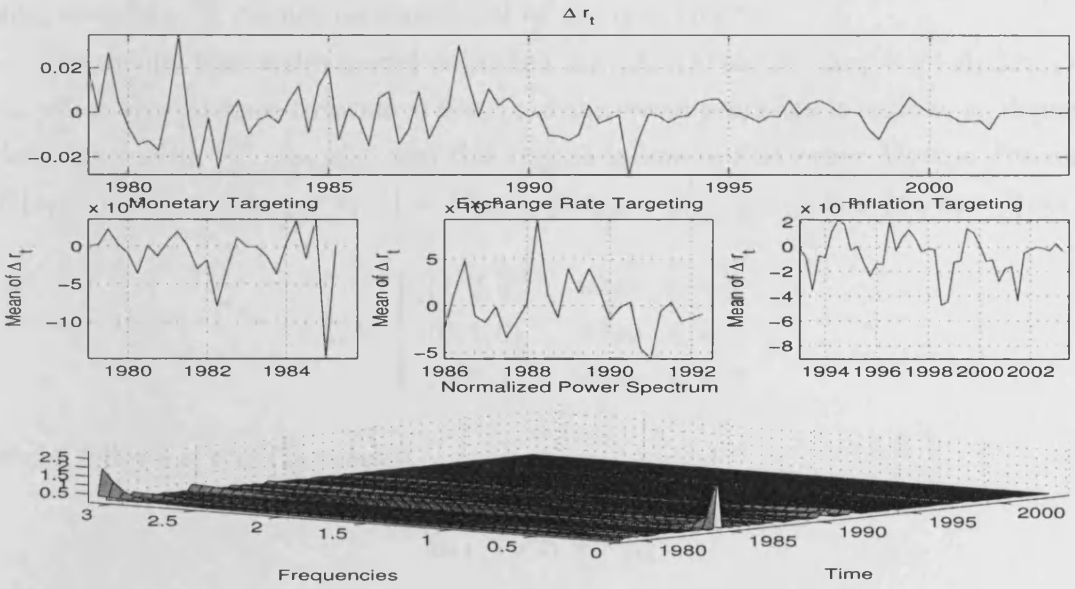
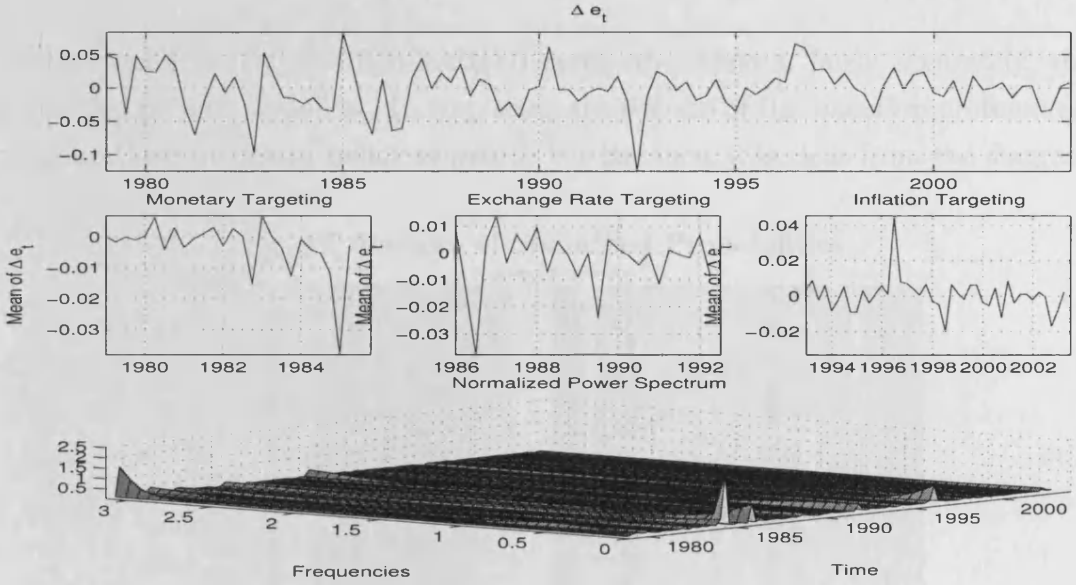


Fig. 4.6: **TVP-VAR(1)/DCC(1,1)-GARCH(1,1):** Δe_t



period and it is further reduced after 1992. This pattern seems plausible, however, we know that the adoption, by the *monetary authorities*, the *Inflation Targeting* policy would induce π_t and, consequently, $\Delta\pi_t$ to be a *white noise* process. This implies that after 1992 the *normalized power spectrum* would be flat and equally to one, however, the latter is flat but significantly lower than one, which reflects negative autocorrelation both in sort and long horizons. This is a contradiction and, therefore, $\mathcal{T}_1$ cannot be considered as the true DGP.

The second time series model estimated is a VAR(1) model, $\Delta x_t = c + A_1 \Delta x_{t-1} + u_t$, where the variance-covariance matrix of the errors processes is regime, s_t, dependent, $u_t \sim N(0, \sum_{i=1}^{3} \Sigma_{u,s}\xi_t)$, and this regime follows a first order *Markov Process*, $P\{s_t = j/s_{t-1} = i, s_{t-1} = k, ...\} = P\{s_t = j/s_{t-1} = i\} = p_{ij}$. ξ_t is a *Markov Chain*

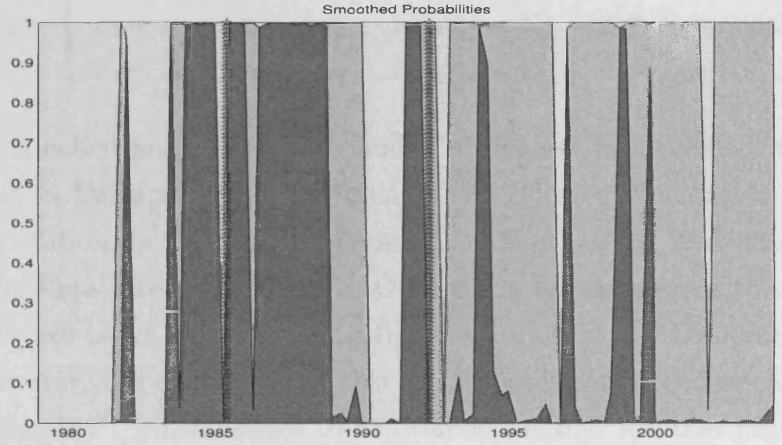
$$\xi_t = \begin{cases} (1,0,0)' & \text{when } s_t = 1 \\ (0,1,0)' & \text{when } s_t = 2 \\ (0,0,1)' & \text{when } s_t = 3 \end{cases}$$

which follows a VAR(1) process

$$\xi_{t+1} = P\xi_t + w_{t+1}$$

where $P = \begin{bmatrix} p_{11} & p_{21} & p_{31} \\ p_{12} & p_{22} & p_{32} \\ p_{13} & p_{23} & p_{33} \end{bmatrix}$ is the matrix of the transition probabilities. This model (which we call MSH(3)-VAR(1)) seems to estimate, fairly accurately, the policy regime turning points. To be precise, the pattern of the *smoothed probabilities* picks up these monetary policy switches. For instance, it is clear from the diagram

Fig. 4.7: **Pattern of Smoothed Probabilities**



that the area after 1992 (second, from the LHS, vertical red line) is dominated by the existence of one state (light green area), which could be viewed as the *inflation targeting regime* introduced after that date. While the *monetary targeting regime* (yellow area) ends few quarters earlier than it is expected to end (1985Q4, fist, from LHS, vertical red line). Table B.1 illustrates some of the properties of the estimated residuals. The *Diagnostic Tests* applied on the residuals of $\mathcal{T}_2$ are *Portmanteau* tests for autocorrelation known as Q and *Adjusted-Q* statistics, a multivariate ARCH-LM test, which investigates for the existence of conditional heteroscedasticity (for both statistics see Lutkepohl and Kratzig, 1994, Section 3.5.2), while, for multivariate unconditional heteroscedasticity the statistic introduced by Doornik (1996) is used. Clearly, these results do not support the view that $\mathcal{T}_2$ is a proper estimate of the true

Tab. 4.1: **Diagnostic Tests/MSH(3)-VAR(1)**

Statistics	*Chi-Square*	*DF*	*P-values*
Conditional Heteroskedasticity	1101.1	900	0
Heteroskedasticity	266.5	150	0
Q- Statistics lags (12)	385.6	275	0
Adj-Q- Statistics lags (12)	410.95	275	0

DGP. The fact that this time series representation correctly identifies the known structural breaks is not borne out by Table B.1, which indicate the presence of important misspecification. However, this model is going to be used below.

The third, and final, time series model, $\mathcal{T}_3$, estimated here is a VAR(1) whose autoregressive and variance-covariance parameters vary across regimes in a deterministic way.

$$\Delta x_t = \begin{cases} c_1 + A_{1,1}\Delta x_{t-1} + \epsilon_t & E(\epsilon_t\epsilon_t') = \Sigma_1 \quad t \leq T_1 = 1986 \\ c_2 + A_{1,2}\Delta x_{t-1} + \epsilon_t & E(\epsilon_t\epsilon_t') = \Sigma_2 \quad t \leq T_2 = 1992 \\ c_3 + A_{1,3}\Delta x_{t-1} + \epsilon_t & E(\epsilon_t\epsilon_t') = \Sigma_3 \quad t > T_3 = 1992 \end{cases}$$

This model is called an Intervention model; a detailed discussion of these models can be found in Lutkepohl (1993, Section 12.4). This representation implies that a particular stationary DGP is in operation until period T_1 and another process generates the data after that period. One of the advantages of this data representation process is its simplicity both from an intuitive and computational point of view. However, the estimation of this model poses some theoretical difficulties, such as consistency[4], which cannot to be overcome easily. Another important characteristic of the Intervention model used above is that after an intervention the moments of the process do not reach a fixed new level immediately but only asymptotically. For instance, let's assume that the *intervention* appears at $t = 4$ solving backwards up to $t = 3$ we get $\Delta x_3 = \sum_{i=0}^{2} A_{1,1}^i c_1 + \sum_{i=0}^{2} A_{1,1}^i \epsilon_{t-i} + A_{1,1}^3 \Delta x_0$, for $t = 4$, $\Delta x_4 = c_2 + A_{1,2}\Delta x_3 + \epsilon_4$ and, for $t > 4$, $\Delta x_t = \sum_{i=0}^{t-1} A_{1,2}^i c_2 + \sum_{i=0}^{t-1} A_{1,2}^i \epsilon_{t-i} + A_{1,2}^t \left(\sum_{i=0}^{2} A_{1,1}^i c_1 + \sum_{i=0}^{2} A_{1,1}^i \epsilon_{t-i} + A_{1,1}^3 \Delta x_0\right)$. The last term of the latter expression vanishes as $t \to \infty$ due to stationarity assumption (the maximum *eigenvalue* of $A_{1,2}$ less than one in absolute terms). This assumption, $t \to \infty$, may make sense if there is only one *intervention* but this is not the case here. However, this behavior may be quite plausible in practice because a system may react slowly to an *intervention* (see Lutkepohl, 1993, Section 12.4.1, page:409).

The Table 4.2 illustrates the properties of $\mathcal{T}_3$'s standardized residuals which do not seem to contain any structure. Additionally, Table 4.3, which presents the values of various selection (*Akaike, Schwartz* and *Hannan-Quinn*) criteria, indicates higher fit for $\mathcal{T}_3$ than for $\mathcal{T}_1$ and $\mathcal{T}_2$. Also, $\mathcal{T}_3$ provides some computational advantages which are important for the exercises carried in this section. For all these reasons $\mathcal{T}_3$ is used for the rest of this section.

[4] The estimates of the middle regimes are functions of their sample sizes which do not tend to infinity. A convenient but not very plausible assumption requires T_i, for $i = 1, 2$, being fractions of T and they tend to infinity as$T \to \infty$.

Tab. 4.2: **Diagnostic Tests**

Statistics	*Chi-Square*	*DF*	*P-values*
Conditional Heteroskedasticity	203.45	225	0.846
Heteroskedasticity	162.75	150	0.225
Q- Statistics lags (18)	383.1	425	0.928
Adj-Q- Statistics lags (18)	469.3	425	0.068

Tab. 4.3: **Selection Criteria**

Information Criteria	*MSH(3)-VAR(1)*	*TVP-VAR(1)-DCC(1,1)-GARCH(1,1)*	*Intervention VAR(1)*
AIC	-31.657	-31.357	-37.888
SIC	-28.140	-30.315	-34.349
HQ	-30.233	-30.936	-36.456

Since $\mathcal{T}_3$ is just a description of the data, its parameter vector, γ, does not convey any economic interpretation and so in this exercise $\hat{\gamma}$ is not reported. However, the *maximum eigenvalue* of $\left\{\hat{A}_{1,i}\right\}_{i=1}^{3}$ and the *trace* of the Variance-Covariance matrices of ϵ_t, $\{tr(\Sigma_\epsilon^i)\}_{i=1}^{3}$, are calculated (Table 4.4). The *maximum eigenvalue* and the *trace* indicate the persistence and the uncertainty, respectively, that the whole vector, $x_t = (y_t', u_t', \pi_t', r_t', e_t')'$, displays over *policy regimes*. Two characteristics are worth mention from this table, first, the remarkable persistence that the vector of 5 variables y_t shows during the middle regime and, second, that the size of uncertainty falls steadily.

Tab. 4.4: **Regime Features**

Features/Regimes	*Monetary*	*Exchange Rate*	*Inflation*
Maximum Eigenvalue	0.451	0.933	0.68
Trace of the Covariance Matrix	0.0043	0.0012	0.0007

Since ϵ_t does not indicate any dependence or heterogeneity, bootstrapped samples of x_t, $\left\{x_t^{*,j}\right\}_{j=1}^{J}$, are easily generated by using *parametric* resampling techniques. Pseudo samples of ϵ_t, $\left\{\epsilon_t^{*,j}\right\}_{j=1}^{J}$, are produced by drawing randomly with replacement from $\hat{\epsilon}_t$ and then $\mathcal{T}_3$ is used recursively to produce $x_t^{*,j}$ given the initial condition x_0. Next the sampling distribution of various population moment or other statistics, S_x, are established, $f(S_x/x_t)$, and the question asked here is whether or not moments or statistics of the data produced by *Liverpool* Model's ($\mathcal{M}$), $\tilde{S}_x$, lies within 95% of $f(S_x/x_t)$. S_x in our case is a vector of population moments both in time and

frequency domain, $vech(\Sigma_x)$ and $diag(\Sigma_x(\omega))$ where $\omega = [0, \pi]$, respectively.

The next three tables indicate the performance of $\mathcal{M}$ over these three periods. The second column of these tables show the moments produced by using $\mathcal{T}_3$ and data simulated by $\mathcal{M}$. The third and fourth column illustrate the lower and the upper, respectively, bounds of the bootstrapped distribution and, finally, the fifth column contains the bootstrapped probability values. Values outside the empirical CI are denoted by **.

Tab. 4.5: **Monetary Targeting/Sampling Variability**

Second Moments	*$\mathcal{M}$'s Moments*	*$\mathcal{T}$'s Lower Bound*	*$\mathcal{T}$'s Upper Bound*	*Bootstrapped Prob-Values*
$\sigma_{\Delta y \Delta y}$	0.0005	0.0001	0.0019	0.686
$\sigma_{\Delta u \Delta y}$	-0.0027	-0.0044	0.0000	0.123
$\sigma_{\Delta \pi \Delta y}$	-0.0001	-0.0001	0.0006	0.102
$\sigma_{\Delta r \Delta y}$	-0.0001	-0.0003	0.0006	0.220
$\sigma_{\Delta e \Delta y}$	0.0002	-0.0011	0.0003	0.968
$\sigma_{\Delta u \Delta u}$	0.0337	0.0010	0.0231	0.989**
$\sigma_{\Delta \pi \Delta u}$	0.0001	-0.0026	0.0000	0.998**
$\sigma_{\Delta r \Delta u}$	0.0023	-0.0025	0.0009	0.997**
$\sigma_{\Delta e \Delta u}$	-0.0002	-0.0006	0.0058	0.084
$\sigma_{\Delta \pi \Delta \pi}$	0.0002	0.0000	0.0007	0.647
$\sigma_{\Delta r \Delta \pi}$	0.0000	0.0000	0.0007	0.149
$\sigma_{\Delta e \Delta \pi}$	-0.0001	-0.0006	0.0002	0.570
$\sigma_{\Delta r \Delta r}$	0.0012	0.0001	0.0020	0.923
$\sigma_{\Delta e \Delta r}$	0.0001	-0.0004	0.0013	0.478
$\sigma_{\Delta e \Delta e}$	0.0037	0.0002	0.0036	0.978**

Before presenting the results it should be emphasized that the length of $x_t \in \mathbb{R}^5$ is rather large for a standard DSGE model comparison exercise. dx implies a significant number of moments, $\frac{5(5+1)}{2} = 15$, per policy regime which is not usually met in this literature. However, even in this case $\mathcal{M}$ performs remarkable well. To be precise, for the first and second regimes only two among fifteen moments lie outside the empirical CI while in the final period this number increases to four, however, some of the key moments are included in this number. To be precise, from tables 4.5 and 4.6 it could be inferred that during the first period $\mathcal{M}$ has some difficulties in explaining the variance of Δu_t and its cross moment with $\Delta \pi_t$, while, during the Exchange Rate Targeting regime it is the variance of Δr_t and its covariance with $\Delta \pi_t$ that worsen the outstanding performance of $\mathcal{M}$. In the third regime $\mathcal{M}$ seems to capture almost all the cross moments of x_t, however, the variance of $\Delta y_t, \Delta \pi_t, \Delta r_t$, and Δe_t lie outside the bootstrapped CI.

Tab. 4.6: **Exchange Rate Targeting/Sampling Variability**

Second Moments	$\mathcal{M}$*'s Moments*	$\mathcal{T}$*'s Lower Bound*	$\mathcal{T}$*'s Upper Bound*	*Bootstrapped Prob-Values*
$\sigma_{\Delta y \Delta y}$	0.00275	0.0001	0.0255	0.876
$\sigma_{\Delta u \Delta y}$	-0.01466	-0.0539	0.0005	0.142
$\sigma_{\Delta \pi \Delta y}$	0.00159	-0.0002	0.0030	0.961
$\sigma_{\Delta r \Delta y}$	0.00793	-0.0004	0.0096	0.971
$\sigma_{\Delta e \Delta y}$	0.00000	-0.0006	0.0130	0.184
$\sigma_{\Delta u \Delta u}$	0.09276	0.0022	1.3973	0.828
$\sigma_{\Delta \pi \Delta u}$	-0.00830	-0.0083	0.0036	0.050
$\sigma_{\Delta r \Delta u}$	-0.03763	-0.0352	0.0044	0.049
$\sigma_{\Delta e \Delta u}$	0.00001	-0.0301	0.0203	0.595
$\sigma_{\Delta \pi \Delta \pi}$	0.00132	0.0000	0.0011	0.976**
$\sigma_{\Delta r \Delta \pi}$	0.00503	-0.0002	0.0013	0.988**
$\sigma_{\Delta e \Delta \pi}$	0.00000	-0.0009	0.0014	0.710
$\sigma_{\Delta r \Delta r}$	0.04070	0.0001	0.0106	0.987**
$\sigma_{\Delta e \Delta r}$	-0.00001	-0.0007	0.0067	0.218
$\sigma_{\Delta e \Delta e}$	0.00000	0.0005	0.0288	0**

Tab. 4.7: **Inflation Targeting /Sampling Variability**

Second Moments	$\mathcal{M}$*'s Moments*	$\mathcal{T}$*'s Lower Bound*	$\mathcal{T}$*'s Upper Bound*	*Bootstrapped Prob-Values*
$\sigma_{\Delta y \Delta y}$	0.0003	0.0006	0.0120	0.005**
$\sigma_{\Delta u \Delta y}$	-0.0017	-0.0273	0.0005	0.838
$\sigma_{\Delta \pi \Delta y}$	0.0000	-0.0005	0.0007	0.448
$\sigma_{\Delta r \Delta y}$	0.0001	0.0002	0.0065	0.008**
$\sigma_{\Delta e \Delta y}$	0.0002	-0.0090	0.0077	0.642
$\sigma_{\Delta u \Delta u}$	0.0169	0.0069	0.1387	0.298
$\sigma_{\Delta \pi \Delta u}$	-0.0001	-0.0017	0.0024	0.435
$\sigma_{\Delta r \Delta u}$	-0.0011	-0.0145	0.0012	0.746
$\sigma_{\Delta e \Delta u}$	-0.0010	-0.0533	0.0106	0.815
$\sigma_{\Delta \pi \Delta \pi}$	0.0009	0.0001	0.0004	0.998**
$\sigma_{\Delta r \Delta \pi}$	-0.0002	-0.0002	0.0006	0.030
$\sigma_{\Delta e \Delta \pi}$	-0.0003	-0.0018	0.0013	0.384
$\sigma_{\Delta r \Delta r}$	0.0055	0.0002	0.0039	0.989**
$\sigma_{\Delta e \Delta r}$	0.0010	-0.0069	0.0030	0.895
$\sigma_{\Delta e \Delta e}$	0.0022	0.0037	0.0785	0.009**

Fig. 4.8: **Monetary Targeting/Sampling Variability**

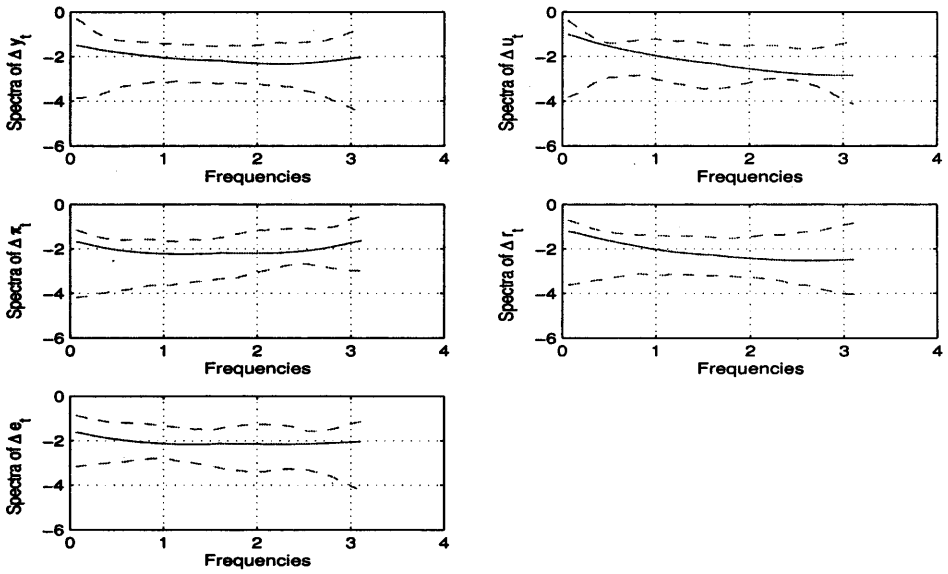


Fig. 4.9: **Exchange Rate Targeting/Sampling Variability**

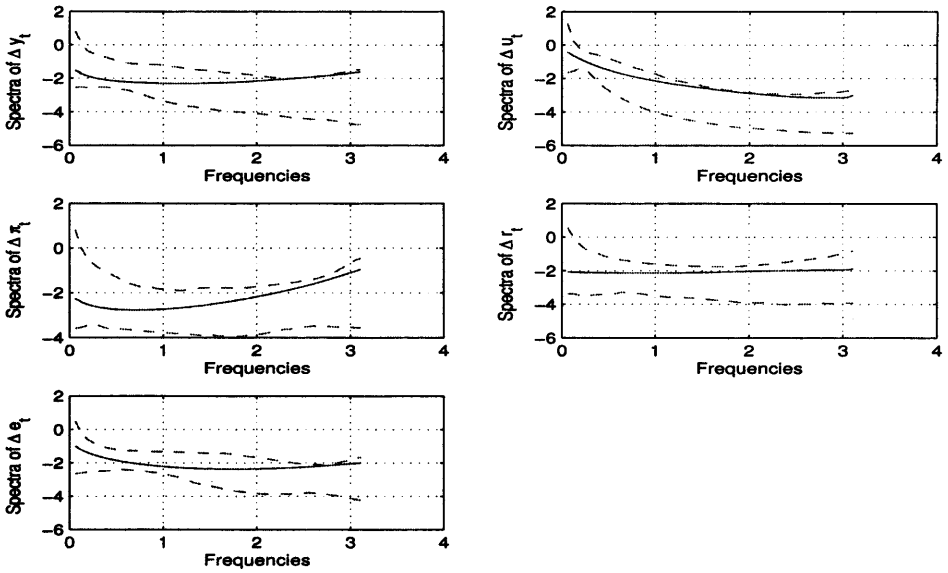
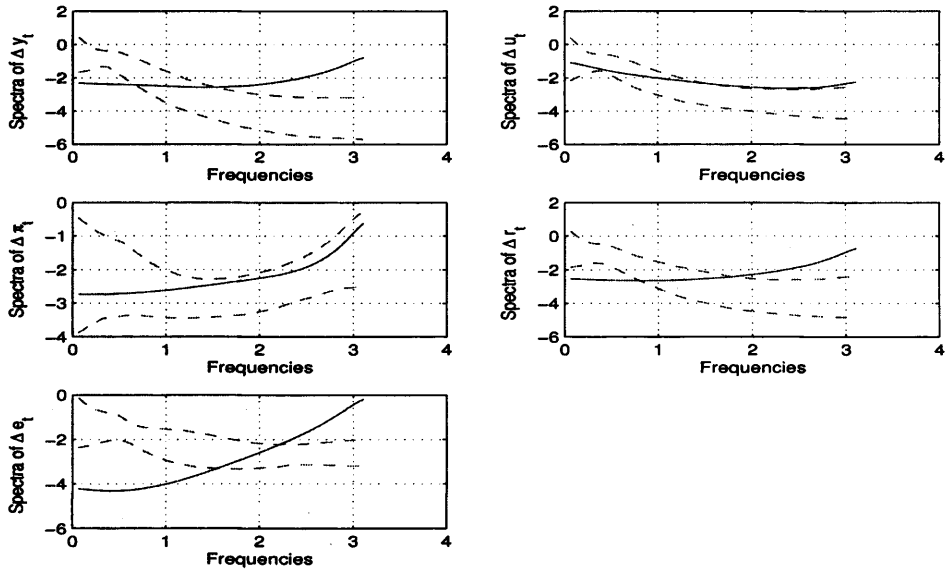


Fig. 4.10: **Inflation Targeting/Sampling Variability**



This good performance is also confirmed when *frequency domain* measures, such as the *normalized power spectrum* discussed above, are used.

$$\mathbf{g}_s^x(\omega) = \frac{2\pi \mathbf{s}_s^x(\omega)}{\int_{-\pi}^{\pi} \mathbf{s}_s^x(\omega)} \tag{4.14}$$

where now

$$\mathbf{s}_s^x = (2\pi)^{-1}(I_{dx} - A_{1,s}e^{-i\omega})^{-1}\Sigma_{u,s}(I_{dx} - A_{1,s}'e^{i\omega})^{-1} \tag{4.15}$$

However, in the figures below the log of equation (4.14) is plotted and this is done only for illustration purposes. Cogley, Morozov, and Sargent (2003) argue that equation (4.14) measures more autocorrelation rather than auto-covariance. Therefore, figures 4.8, 4.9 and 4.10 provide an optical way to assess the performance of $\mathcal{M}$ in terms of the autocorrelation. These figures indicate that $\mathcal{M}$, correctly, estimate the size of *persistence* in the economy, the estimated equation (4.14) using data simulated by $\mathcal{M}$ (blue solid line) lies within the 95% of the bootstrapped distribution, $f(\mathbf{g}_s^x(\omega)/\mathcal{T}_3, x_t)$ (red solid lines). This may not being totally true for the last regime, however, even in this case the size of persistence of $\Delta\pi_t$ and Δu_t series produced by $\mathcal{M}$ lies within the bootstrapped CI.

So far it was shown how methods that construct sampling variability measures can be encompassed by the one proposed in this section. This is also true for the methods that use simulation variability to assess the performance of the structural

Tab. 4.8: **Monetary Targeting/Simulation Variability**

Second Moments	$\mathcal{T}$*'s Moments*	$\mathcal{M}$*'s Lower Bound*	$\mathcal{M}$*'s Upper Bound*	*Bootstrapped Prob-Values*
$\sigma_{\Delta y \Delta y}$	0.0089	0.0009	0.0321	0.252
$\sigma_{\Delta u \Delta y}$	-0.0009	-0.0061	0.0070	0.468
$\sigma_{\Delta \pi \Delta y}$	0.0134	-0.0078	0.0224	0.059
$\sigma_{\Delta r \Delta y}$	-0.0002	-0.0347	0.0079	0.423
$\sigma_{\Delta e \Delta y}$	-0.0006	-0.0091	0.0082	0.658
$\sigma_{\Delta u \Delta u}$	0.0026	0.0027	0.0210	0.978*
$\sigma_{\Delta \pi \Delta u}$	-0.0014	-0.0071	0.0082	0.817
$\sigma_{\Delta r \Delta u}$	-0.0003	-0.0049	0.0109	0.767
$\sigma_{\Delta e \Delta u}$	0.0004	-0.0023	0.0064	0.510
$\sigma_{\Delta \pi \Delta \pi}$	0.0327	0.0006	0.0462	0.045
$\sigma_{\Delta r \Delta \pi}$	-0.0012	-0.0336	0.0062	0.557
$\sigma_{\Delta e \Delta \pi}$	-0.0013	-0.0097	0.0081	0.782
$\sigma_{\Delta r \Delta r}$	0.0008	0.0008	0.0528	0.983*
$\sigma_{\Delta e \Delta r}$	0.0001	-0.0053	0.0132	0.535
$\sigma_{\Delta e \Delta e}$	0.0015	0.0016	0.0163	0.985*

Tab. 4.9: **Exchange Rate Targeting/Simulation Variability**

Second Moments	$\mathcal{T}$*'s Moments*	$\mathcal{M}$*'s Lower Bound*	$\mathcal{M}$*'s Upper Bound*	*Bootstrapped Prob-Values*
$\sigma_{\Delta y \Delta y}$	0.0026	0.0008	0.1451	0.161
$\sigma_{\Delta u \Delta y}$	-0.0013	-0.0044	0.0478	0.166
$\sigma_{\Delta \pi \Delta y}$	0.0005	-0.0299	0.0284	0.584
$\sigma_{\Delta r \Delta y}$	-0.0001	-0.0081	0.0067	0.490
$\sigma_{\Delta e \Delta y}$	-0.0001	-0.3487	0.2117	0.447
$\sigma_{\Delta u \Delta u}$	0.0024	0.0019	0.0350	0.069
$\sigma_{\Delta \pi \Delta u}$	-0.0026	-0.0076	0.0099	0.140
$\sigma_{\Delta r \Delta u}$	0.0008	-0.0032	0.0027	0.750
$\sigma_{\Delta e \Delta u}$	0.0006	-0.1763	0.1823	0.594
$\sigma_{\Delta \pi \Delta \pi}$	0.0048	0.0009	0.0608	0.510
$\sigma_{\Delta r \Delta \pi}$	-0.0017	-0.0037	0.0033	0.084
$\sigma_{\Delta e \Delta \pi}$	-0.0008	-0.1380	0.1173	0.555
$\sigma_{\Delta r \Delta r}$	0.0006	0.0031	0.0200	0.000**
$\sigma_{\Delta e \Delta r}$	0.0002	-0.0253	0.0436	0.510
$\sigma_{\Delta e \Delta e}$	0.0009	0.0015	35.2037	0.012**

Tab. 4.10: **Inflation Rate Targeting/Simulation Variability**

Second Moments	$\mathcal{T}$*'s Moments*	$\mathcal{M}$*'s Lower Bound*	$\mathcal{M}$*'s Upper Bound*	*Bootstrapped Prob-Values*
$\sigma_{\Delta y \Delta y}$	0.0012	0.0013	0.0778	0.017**
$\sigma_{\Delta u \Delta y}$	-0.0002	-0.0047	0.0151	0.542
$\sigma_{\Delta \pi \Delta y}$	0.0001	-0.0112	0.0064	0.658
$\sigma_{\Delta r \Delta y}$	-0.0008	-0.0156	0.0183	0.443
$\sigma_{\Delta e \Delta y}$	0.0001	-0.0239	0.0257	0.443
$\sigma_{\Delta u \Delta u}$	0.0011	0.0027	0.0215	0.000**
$\sigma_{\Delta \pi \Delta u}$	-0.0004	-0.0031	0.0018	0.530
$\sigma_{\Delta r \Delta u}$	0.0031	-0.0047	0.0036	0.955
$\sigma_{\Delta e \Delta u}$	-0.0004	-0.0041	0.0071	0.191
$\sigma_{\Delta \pi \Delta \pi}$	0.0002	0.0030	0.0108	0.000**
$\sigma_{\Delta r \Delta \pi}$	-0.0011	-0.0037	0.0127	0.228
$\sigma_{\Delta e \Delta \pi}$	0.0001	-0.0113	-0.0001	0.975
$\sigma_{\Delta r \Delta r}$	0.0101	0.0014	0.0525	0.408
$\sigma_{\Delta e \Delta r}$	-0.0009	-0.0282	0.0094	0.757
$\sigma_{\Delta e \Delta e}$	0.0006	0.0076	0.0384	0.000**

model. It was explained above that given equation (4.8) and the actual data, x_t, the simulated distribution of $\tilde{S}_x$ conditional on the actual data, $f\left(\tilde{S}_x/\mathcal{M}, \mathcal{T}_3, x_t\right)$, is readily derived. Given $f\left(\tilde{S}_x/\mathcal{M}, \mathcal{T}_3, x_t\right)$, the question is reversed and it is investigated whether or not S_x lies within 95% of the latter distribution. Again the same population moments, both in time and frequency domain, are used for this purpose.

When the exercise is conducted in this way the picture regarding the performance of $\mathcal{M}$ does not change. From tables 4.8 and 4.9 we can see that the number of moments that lie outside the boundaries remains the same. However, this time, they are the variances of Δr_t and Δe_t that cause the problem in both regimes. Similarly, in the last period the cross moments of x_t fall inside the 95% *simulated* CI but not the variances of its components (see table 4.10).

In terms of the spectral domain analysis the size of persistence of the historical x_t always lies into within the 95% of the simulated distribution of $\tilde{\mathbf{g}}_x$, $f\left(\tilde{\mathbf{g}}_x/\mathcal{M}, \mathcal{T}_3, x_t\right)$, for the first and second policy regime (figures 4.11 and 4.12). However, in the Inflation Targeting period there is now a departure compared to the preceding case. This time they are the sizes of persistence of the historical Δy_t and Δu_t, which are captured by $f\left(\tilde{\mathbf{g}}_x/\mathcal{M}, \mathcal{T}_3, x_t\right)$, and not the size of $\Delta \pi_t$ (figure 4.13).

It was shown in this section how the two of the four *calibration* methods of evaluating DSGE models studied in the last chapter could be encompassed by the proposed one. It was mentioned above that $x_t \in \mathbb{R}^{dx}$ implies $\frac{dx(dx+1)}{2}$ moments, how-

Fig. 4.11: **Monetary Targeting/Simulation Variability**

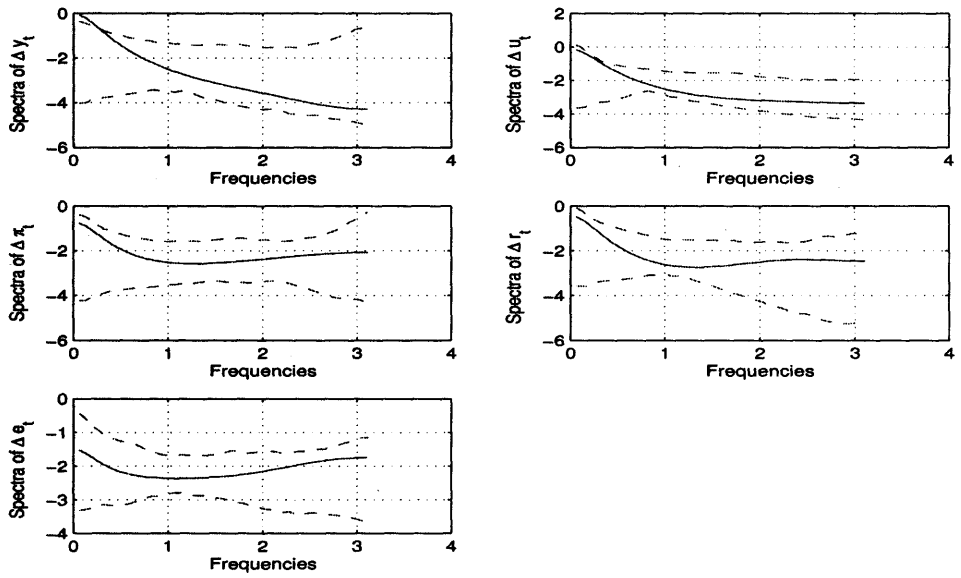


Fig. 4.12: **Exchange Rate Targeting/Simulation Variability**

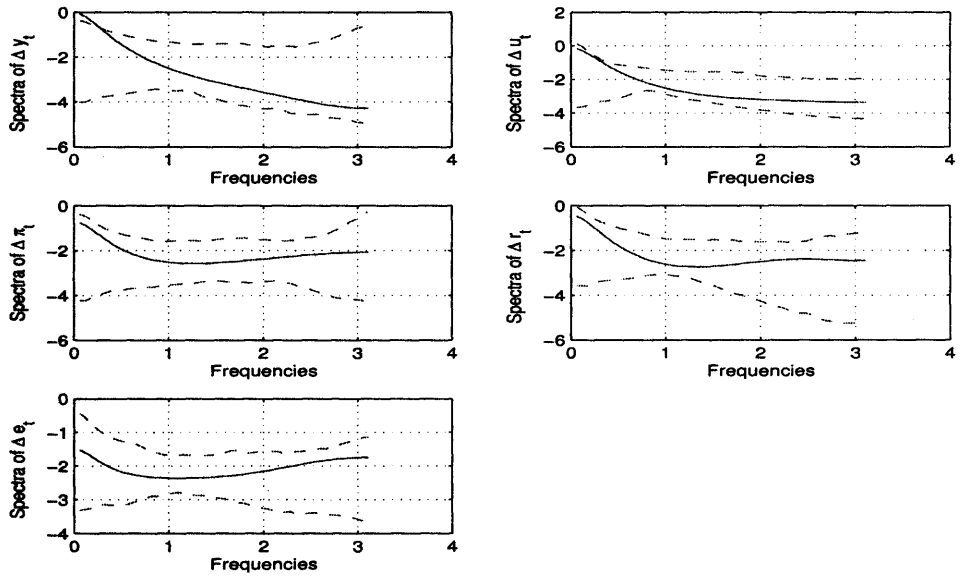
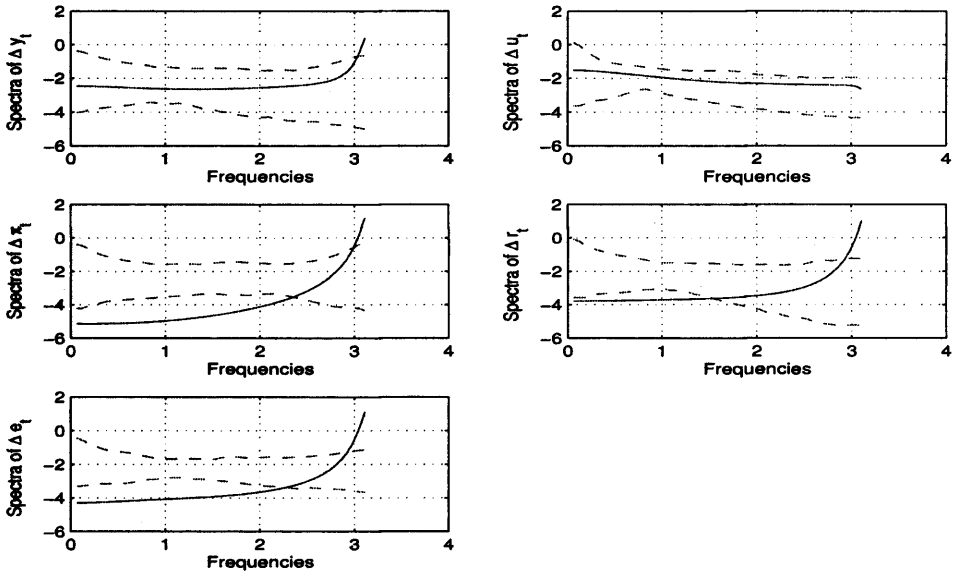


Fig. 4.13: **Inflation Targeting/Simulation Variability**



ever, the assessment of the structural model was done of the basis of each moment individually and not on the basis of $\frac{dx(dx+1)}{2}$ simultaneously. Certainly, we know from the *Cramer-Wold Theorem* that a vector random sequence converges to a random vector if and only if each every sequence of the vector converges (see Davidson, 1994, Theorem 25.5, page:405), which probably means that S_x does not converge to $\tilde{S}_x$ and vice versa. In this section a *Wald* type statistic is described which allows for multivariate comparison.

4.1.3 The Proposed Statistic or M-metric

The above exercises are based on functions of the estimated parameter vector $\hat{\gamma}$ of the time series representation $\mathcal{T}$, namely, $\zeta(\hat{\gamma})$. A first order Taylor expansion around $\tilde{\gamma}$ gives $\zeta(\hat{\gamma}) = \zeta(\tilde{\gamma}) + \nabla\gamma\zeta(\tilde{\gamma})(\hat{\gamma} - \tilde{\gamma})$ and from here standard asymptotic results can be used to show that $\sqrt{T}(\zeta(\hat{\gamma}) - \zeta(\tilde{\gamma})) \sim N\left(0, \nabla\gamma\zeta(\tilde{\gamma})\Sigma_{\tilde{\gamma}}\nabla\gamma\zeta(\tilde{\gamma})'\right)$ (see Serfling, 1980, page:122-124).

However, $\zeta(\tilde{\gamma})$, $\nabla\gamma\zeta(\tilde{\gamma})$ and $\Sigma_{\tilde{\gamma}}$ are unknown probability limits which are, usually, replaced by consistent estimates. Due to the theory developed in the last chapter these limits can be recovered through stochastic simulation and we denoted them by $\zeta(\tilde{\gamma}/\mathcal{M}, \mathcal{T}, x_t)$, $\nabla\gamma\zeta(\tilde{\gamma}/\mathcal{M}, \mathcal{T}, x_t)$ and $\Sigma_{\tilde{\gamma}/\mathcal{M},\mathcal{T},x_t}$ respectively. Given these limits there are various ways to test the hypothesis $H_0 : \zeta(\hat{\gamma}) = \zeta(\tilde{\gamma}/\mathcal{M}, \mathcal{T}, x_t)$.

1. *Asymptotic Theory*: Under the null hypothesis

$$\sqrt{T}\left(\zeta\left(\hat{\gamma}\right)-\zeta\left(\tilde{\gamma}/\mathcal{M},\mathcal{T},x_t\right)\right)\sim N\left(0,\nabla\gamma\zeta\left(\tilde{\gamma}/\mathcal{M},\mathcal{T},x_t\right)\Sigma_{\tilde{\gamma}}\nabla\gamma\zeta\left(\tilde{\gamma}/\mathcal{M},\mathcal{T},x_t\right)'\right)$$

which implies that if the value of the following expression

$$\mathcal{W}_{\tilde{\gamma}}\equiv T\left(\zeta\left(\hat{\gamma}\right)-\zeta\left(\tilde{\gamma}/\mathcal{M},\mathcal{T},x_t\right)\right)'H_{\tilde{\gamma}}^{-1}\left(\zeta\left(\hat{\gamma}\right)-\zeta\left(\tilde{\gamma}/\mathcal{M},\mathcal{T},x_t\right)\right) \tag{4.16}$$

exceeds the value of the χ^2 distribution with degrees of freedom equal to the dimension of $\zeta\left(\hat{\gamma}\right)$ and significance level α then the null hypothesis is rejected.

2. *Small Sample Distribution*: The small sample distribution of expression 4.16 conditional on the the structural model and the actual data, $f\left(\mathcal{W}_{\tilde{\gamma}}/\mathcal{M},\mathcal{T},x_t\right)$, is readily derived. Once the above mentioned limits have been calculated the simulated data used for these calculation could be used again to approximate the distribution of expression 4.16. If $\mathcal{W}_{\tilde{\gamma}}$ exceeds the $1-\alpha$ values of $f\left(\mathcal{W}_{\tilde{\gamma}}/\mathcal{M},\mathcal{T},x_t\right)$ then the null hypothesis is rejected.

The second way has been used in this section and $\zeta\left(\hat{\gamma}\right)\equiv\hat{\gamma}$, therefore, the null is $H_0:\hat{\gamma}=\tilde{\gamma}\left(\mathcal{M}\right)$. Two time series representation have been used in this exercise, the *Intervention*(3)-VAR(1), $\mathcal{T}_3$, and the MSH(3)-VAR(1), $\mathcal{T}_2$.

The analysis relies on the estimation properties of the QML estimator. The latter estimator under very weak assumptions is consistent and normally distributed, $\hat{\gamma}_T\sim N\left(\tilde{\gamma},H_{\tilde{\gamma}}^{-1}D_{\tilde{\gamma}}H_{\tilde{\gamma}}^{-1}\right)$ where $H_{\tilde{\gamma}}\equiv\frac{1}{S}\sum_{s=1}^{S}\nabla_{\tilde{\gamma}}^2\log f\left(y_s/\tilde{\gamma},\mathcal{M}\right)$ and $D_{\tilde{\gamma}}\equiv var\Big(\frac{1}{\sqrt{S}}\sum_{s=1}^{S}\nabla_{\tilde{\gamma}}\log f\left(y_s/\tilde{\gamma},\mathcal{M}\right)\Big)$. I am providing here closed form expressions for both the score vector and the Hessian matrix of the log-Likelihood of $\mathcal{T}_3$, which could be used for $\mathcal{T}_2$ as well, however, under very restrictive assumptions. To be precise, $\mathcal{T}_3$ can be written as

$$y=\sum_{s=1}^{S}\left(\Xi_s\bar{X}\otimes I_{dy}\right)\beta_s+u \tag{4.17}$$

where $u\sim N\left(0,\sum_{s=1}^{S}\Xi_s\otimes\Sigma_s\right)$, $\Xi_s=diag\left(\xi_s\right)$ and $\bar{X}=\left[\mathbf{1}_T,Y_{-1},..,Y_{-p}\right]$ and the

log-likelihood in this case is written as

$$\begin{aligned} L_{\cdot} &= -\frac{dyT}{2}\log(2\pi) - \frac{1}{2}\sum_{s=1}^{S}\left\{ \begin{array}{c} \log|\Xi_s \otimes \Sigma_s| + \left(y - \left(\Xi_s\bar{X} \otimes I_{dy}\right)\beta_s\right)'\left(\Xi_s \otimes \Sigma_s^{-1}\right) \\ \left(y - \left(\Xi_s\bar{X} \otimes I_{dy}\right)\beta_s\right) \end{array} \right\} \\ &= -\frac{dyT}{2}\log(2\pi) - \frac{1}{2}\sum_{s=1}^{S}\left\{ \begin{array}{c} \log\left\{|\Xi_s|^{dy}|\Sigma_s|^{T_s}\right\} + \left(y - \left(\Xi_s\bar{X} \otimes I_{dy}\right)\beta_s\right)' \\ \left(\Xi_s' \otimes I_{dy}\right)\left(I_T \otimes \Sigma_s^{-1}\right)\left(\Xi_s \otimes I_{dy}\right) \\ \left(y - \left(\Xi_s\bar{X} \otimes I_{dy}\right)\beta_s\right) \end{array} \right\} \\ &= -\frac{dyT}{2}\log(2\pi) - \frac{dy}{2}\sum_{s=1}^{S}\log|\Xi_s| - \frac{1}{2}\sum_{s=1}^{S}\left\{T_s\log|\Sigma_s| + u_s'\left(I_T \otimes \Sigma_s^{-1}\right)u_s\right\} \end{aligned} \quad (4.18)$$

All the matrix derivatives derived below rely on matrix calculus rules developed by Lutkepohl (1993, Appendix A.13). The first block of the score vector is given by

$$\begin{aligned} \frac{\partial L}{\partial \beta_s} &= -\frac{1}{2}(-2)\left(\left(\Xi_s\bar{X} \otimes I_{dy}\right)'\left(\Xi_s' \otimes I_{dy}\right)\left(I_T \otimes \Sigma_s^{-1}\right)\left(\Xi_s \otimes I_{dy}\right)\left(y - \left(\Xi_s\bar{X} \otimes I_{dy}\right)\beta_s\right)\right) \\ &= \left(\Xi_s\bar{X} \otimes I_{dy}\right)'\left(\Xi_s \otimes \Sigma_s^{-1}\right)\left(y - \left(\Xi_s\bar{X} \otimes I_{dy}\right)\beta_s\right) \end{aligned} \quad (4.19)$$

while the first derivatives with respect to the error variances of each regime are given by

$$\begin{aligned} \frac{\partial L}{\partial \Sigma_s} &= \frac{\partial\left(-\frac{T_s}{2}\log|\Sigma_s| - \frac{1}{2}trace\left\{u_s'\Sigma_s^{-1}u_s\right\}\right)}{\partial \Sigma_s} \\ &= -\frac{T_s}{2}\Sigma_s^{-1} + \frac{1}{2}\left(\Sigma_s^{-1}U_sU_s'\Sigma_s^{-1}\right) \end{aligned} \quad (4.20)$$

The Hessian matrix in this case has the following form

$$H_{\tilde{\gamma}} = \begin{bmatrix} \begin{bmatrix} \frac{\partial^2 L}{\partial\beta_1\partial\beta_1'} & 0 & \cdots & 0 \\ 0 & \frac{\partial^2 L}{\partial\beta_2\partial\beta_2'} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial^2 L}{\partial\beta_S\partial\beta_S'} \end{bmatrix} & \begin{bmatrix} \left(\frac{\partial^2 L}{\partial\sigma_1\partial\beta_1'}\right)' & 0 & \cdots & 0 \\ 0 & \left(\frac{\partial^2 L}{\partial\sigma_2\partial\beta_2'}\right)' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\frac{\partial^2 L}{\partial\sigma_S\partial\beta_S'}\right)' \end{bmatrix} \\ \begin{bmatrix} \frac{\partial^2 L}{\partial\sigma_1\partial\beta_1'} & 0 & \cdots & 0 \\ 0 & \frac{\partial^2 L}{\partial\sigma_2\partial\beta_2'} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial^2 L}{\partial\sigma_S\partial\beta_S'} \end{bmatrix} & \begin{bmatrix} \frac{\partial^2 L}{\partial\sigma_1\partial\sigma_1'} & 0 & \cdots & 0 \\ 0 & \frac{\partial^2 L}{\partial\sigma_2\partial\sigma_2'} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial^2 L}{\partial\sigma_S\partial\sigma_S'} \end{bmatrix} \end{bmatrix} \quad (4.21)$$

where

$$\frac{\partial^2 L}{\partial\beta_s\partial\beta_s'} = -\left(\Xi_s\bar{X} \otimes I_{dy}\right)'\left(\Xi_s \otimes \Sigma_s^{-1}\right)\left(\Xi_s\bar{X} \otimes I_{dy}\right) \quad (4.22)$$

Tab. 4.11: **DGP: INTERVENTION(3)-VAR(1)**

CI	*Simulated Statistic*	*Actual Statistic*
94%	1.21	
95%	1.243	
96%	1.325	
97%	1.574	
98%	1.699	
99%	***1.965***	***1.9598 (98.76%)***
100%	5.918	

$$\begin{aligned}\frac{\partial^2 L}{\partial\sigma_s\partial\beta_s'} &= \frac{1}{2}\left(\Sigma_s^{-1}\otimes\Sigma_s^{-1}\right)\frac{\partial vec\left(U_sU_s'\right)}{\partial\beta_s'} \\ &= \frac{1}{2}\left(\Sigma_s^{-1}\otimes\Sigma_s^{-1}\right)\left\{\left(U_s\otimes I_{dy}\right)\left(-\left(\bar{X}\otimes I_{dy}\right)\right)+\left(I_{dy}\otimes U_s\right)\mathbf{K}_{T,dy}\left(-\left(\bar{X}\otimes I_{dy}\right)\right)\right\} \\ &= -\frac{1}{2}\left(\Sigma_s^{-1}\otimes\Sigma_s^{-1}\right)\left\{\left(U_s\bar{X}\otimes I_{dy}\right)+\left(I_{dy}\otimes U_s\right)\mathbf{K}_{T,dy}\left(\bar{X}\otimes I_{dy}\right)\right\} \qquad (4.23)\end{aligned}$$

and, finally,

$$\frac{\partial^2 L}{\partial\sigma_s\partial\sigma_s'}=\frac{T_s}{2}\left(\Sigma_s^{-1}\otimes\Sigma_s^{-1}\right)-\frac{1}{2}\left(\Sigma_s^{-1}U_sU_s'\Sigma_s^{-1}\otimes\Sigma_s^{-1}\right)-\frac{1}{2}\left(\Sigma_s^{-1}\otimes\Sigma_s^{-1}U_sU_s'\Sigma_s^{-1}\right) \qquad (4.24)$$

Given the above expressions for the *score* vector and the *Hessian* matrix the derivation of $f\left(\mathcal{W}_{\tilde{\gamma}}/\mathcal{M},\mathcal{T},x_t\right)$ is something less than easy. Table 4.11 presents the results from this exercise using $\mathcal{T}_3$

Obviously, the hypothesis $H_0:\hat{\gamma}_T=\tilde{\gamma}$ is rejected at significance level, α, equal to 5% but not for $\alpha=1\%$, the value of $\mathcal{W}_{\tilde{\gamma}}$ is less than the value of the 99% percentile of $f\left(\mathcal{W}_{\tilde{\gamma}}/\mathcal{M},\mathcal{T},x_t\right)$. This result confirms the inference drawn regarding $\mathcal{M}$ in the last section. It was shown there that the structural model performs remarkably well in terms of the second moments when these are examined individually. This section reaches the same conclusion by using a *Wald* test statistic or M-metric. In simple terms this statistic says that the parameter vector γ obtained by using $\mathcal{T}$ and historical data is not statistically different from that obtained by using $\mathcal{T}$ and data simulated by the structural model conditional on the actual data.

The same exercise is repeated by using $\mathcal{T}_2$ this time. The estimation of this time series model is done by using the *Expectation Maximization* algorithm. This is a two steps procedure where in the first, *Expectation*, step the pattern of the smoothed probabilities, $\{\xi_{s,t}\}_{s=1}^{S}$, is estimated and in the second, *Maximization*, step these probabilities are used during the maximization of 4.18. The Hessian matrix in this

Tab. 4.12: **DGP: MSH(3)-VAR(1)/FULL**

CI	*Simulated Statistic*	*Actual Statistic*
94%	0.032	
95%	**0.0354**	***0.0344***
96%	0.0366	
97%	0.0389	
98%	0.0414	
99%	0.0507	
100%	0.3921	

case has the following form

$$H = \begin{bmatrix} \left[\frac{\partial^2 L}{\partial\beta\partial\beta'}\right] & \begin{bmatrix} \left(\frac{\partial^2 L}{\partial\sigma_1\partial\beta'}\right)' & 0 & \cdots & 0 \\ 0 & \left(\frac{\partial^2 L}{\partial\sigma_2\partial\beta'}\right)' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \left(\frac{\partial^2 L}{\partial\sigma_S\partial\beta'}\right)' \end{bmatrix} \\ \begin{bmatrix} \frac{\partial^2 L}{\partial\sigma_1\partial\beta'} & 0 & \cdots & 0 \\ 0 & \frac{\partial^2 L}{\partial\sigma_2\partial\beta'} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial^2 L}{\partial\sigma_S\partial\beta'} \end{bmatrix} & \begin{bmatrix} \frac{\partial^2 L}{\partial\sigma_1\partial\sigma_1'} & 0 & \cdots & 0 \\ 0 & \frac{\partial^2 L}{\partial\sigma_2\partial\sigma_2'} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial^2 L}{\partial\sigma_S\partial\sigma_S'} \end{bmatrix} \end{bmatrix} \tag{4.25}$$

However, the above expressions regarding the first and the second derivatives of γ cannot be used for the construction of 4.25 and this is because $\xi_{s,t}$ are also functions of γ. Therefore, closed form expressions both for the Score vector and the Hessian matrix cannot be derived and the latter have to be calculated numerically.

Table 4.12 summarizes the results from this exercise. In the present case the null hypothesis, $H_0 : \hat{\gamma}_T = \tilde{\gamma}$, cannot be rejected at significance level of 5%. The benefit from the latter exercise is twofold, it could be viewed as an additional indication regarding the good performance of $\mathcal{M}$ and, also, as a robust check of the sensitivity of results when $\mathcal{T}$ varies. The estimation of the above and the calculation of $\Sigma_{\tilde{\gamma}}$ is computationally cumbersome by itself and when this has to be done for a large number of simulated sets this task becomes even worse. From Krolzig (1997) we know that there are closed form expressions for γ (set equations (4.19) and (4.20) equal to zero and solve for β_s and σ_s respectively) and consequently for the Scores of (4.18), however, this is not true for the Hessian matrix.

An alternative approach, which decreases the computational effort substantially,

is to treat the pattern of the smoothed probabilities for each regime, Ξ_s, during the maximization of (4.18) as *known*. In this case the estimation problem becomes similar to the one studied above, *Intervention*(3)-VAR(1) ($\mathcal{T}_3$), because the *Maximization* step becomes independent of *Expectation* step. Certainly, this a very restrictive assumption and I was not able to think, so far, of a conceptual justification of treating the two steps as independent, however, this does not seem to make much difference in terms of the statistical inference. Table 4.13 indicates, again, that the null hypothesis cannot be rejected at significance level 5% even when the latter procedure is applied.

Tab. 4.13: **DGP: MSH(3)-VAR(1)**

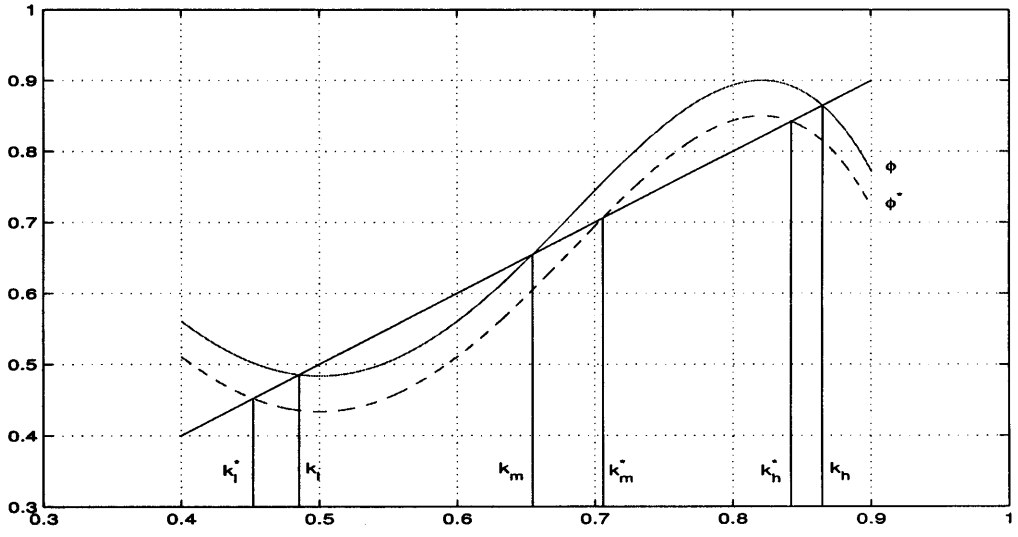
CI	*Simulated Statistic*	*Actual Statistic*
94%	35.842	
95%	**38.698**	**37.594**
96%	41.734	
97%	45.266	
98%	67.38	
99%	78.388	
100%	139.51	

5. MEASURE THEORY RESULTS

In a very recent paper Santos and Peralta-Alva (2005) provide a general framework of the simulation of stochastic dynamic models. They establish that the simulated moments from numerical approximations of model's equilibrium function (due to the nonlinearity of the structural model an explicit solution of the latter function is not available) converge to their exact values as the approximation errors of the computed solutions converge to zero. The above studied evaluation process defines a notion of distance in which a selected set of statistics is compared with its data counterparts. However, this exercise is only meaningful if the statistics computed from the numerical approximation are sufficiently close to the true ones.

In order the importance of their result to be fully understood their first example of an earlier version of the above paper is presented here. Let ϕ be the real-values

Fig. 5.1: **Numerical Approximation**



function shown in the figure, which has three interior steady states k_l, k_m and k_h and ϕ^* is another real valued function with its own steady states k_l^*, k_m^* and k_h^*. From the figure it is clear from the figure that steady states k_m and k_m^* are unstable, therefore, any initial condition over the interval (k_m, k_m^*) will converge to point k_h under ϕ but will converge to point k_l^* under ϕ^*. Consequently, a small perturbation on the function ϕ will not generally have good stability properties near the stationary

solution k_m. From these functions we can now construct a random dynamical system under the following *i.i.d* process. At each date, $t = 0, 1, 2, ...$, let the system move by function ϕ with probability $1/2$ and by ϕ^* with equal probability. The resulting random dynamical system has two ergodic invariant distributions whose supports are the intervals $[k_l^*, k_l]$ and $[k_h^*, k_h]$. Points over the interval $[k_m, k_m^*]$ will leave this domain with probability one and this interval no longer will contain a stationary solution. Therefore, by assigning a small stochastic perturbation to a deterministic dynamical system may lead to substantial discrepancies in the long run.

Their results entail that the moments computed from numerical simulations converge to the moments of the model's invariant distributions as the approximation errors of the computed solutions converge to zeros. However, their results apply only when one single sample path of arbitrary large length is simulated by the structural model and not for a large sample of paths (J) of length equal to that of the data sample (T), which is the common practice in macroeconomics. By using *Product Space Analysis* and similar to the above authors methodology we develop results that cover the latter case, actually, the former results could be viewed as a special case, $J = 1$, of the results introduced here.

The second issue is related with the measurability of the estimated parameters of a time series model ($\tilde{\gamma}$) fitted to the simulated data. The structural parameters (θ) used for the simulation are the estimated one ($\hat{\theta}$). The set of the time series parameters ($\mathbf{\Gamma}$) becomes a function of $\hat{\theta}$ and $\epsilon_{j,t}$ or, in other words, $\mathbf{\Gamma}(\hat{\theta}, \epsilon_{j,t})$ is a random set. Stinchcombe and White (1992) provide conditions whereby the extremum of a random function over a random set is itself a random object. However, these results cannot, directly, be applied here. We provide conditions when this holds in the present framework.

5.1 Accuracy of Numerical Simulations

5.1.1 Asymptotic Convergence Results

The aim of this subsection is not to provided new theoretical results regarding the limiting behavior of Markovian Economic models. The author's target is to supply those conditions so the theory has been developed all these years (see Futia, 1982; Stockey, Lucas, and Prescott, 1989; Hopenhayn and Prescott, 1992; Santos and Peralta-Alva, 2005) can be utilized here. This is done through the use of the product space analysis.

The core of a DSGE model is the equilibrium law of motion mainly written as

a dynamical system of equations $s_{j,t+1} = \phi\left((s_t, \epsilon_{t+1})_j\right)$[1],where $s \in \mathbb{R}^{ds}$ and $\epsilon \in \mathbb{R}^{d\epsilon}$. The observable or jump variables are function of the $s_{j,t}$, $y_{j,t} = \psi\left((s_t, \epsilon_t, v_t)_j\right)$ with $y \in \mathbb{R}^{dy}$ and $v \in \mathbb{R}^{dv}$. The only thing needed to say about the last equation is that y is a function of v only when the number of the stochastic processes is smaller than the observable variables ($dy > d\epsilon$). From the estimation of the structural model point of view this augmentation is necessary so the stochastic singularity is eliminated[2]. For the rest chapter the latter equation is not going to be used again and all the attention is concentrated on the law of motion equation. The following definition ideally summarizes the a random dynamical system.

Definition 5.1.1: A random dynamical system ($r.d.s$) is defined by the following data: (a) a state space $\left(\tilde{\mathbf{S}} = \mathsf{X}_{j=1}^{J}\mathbf{S}, \tilde{\mathcal{S}} = \otimes_{j=1}^{J}\mathcal{S}\right)$ is a separable metric space, and a Borel subset and metric subspace of a complete, separable metric space[3]; $\tilde{\mathcal{S}}$ is its Borel σ-field; (b) a measurable space $\left(\tilde{\mathbf{E}} = \mathsf{X}_{j=1}^{J}\mathbf{E}, \tilde{\mathcal{E}} \otimes_{j=1}^{J} \mathcal{E}, Q_E = \times_{j=1}^{J}\mu_\epsilon\right)$ of *events*; (c) a stochastic kernel $Q_E : \left(\mathsf{X}_{j=1}^{J}(\mathbf{S} \times \mathcal{E})\right) \rightarrow [0,1]$; thus $Q_E\left(\mathsf{X}_{j=1}^{J}(s_t, A)_j\right)$ is the probability of realizing event $A \in \tilde{\mathcal{E}}$ given that the current state is $s_{j,t} \in \tilde{\mathbf{S}}$; (d) a mapping $\phi : \left(\tilde{\mathbf{S}} \times \tilde{\mathbf{E}}\right) = \left(\mathsf{X}_{j=1}^{J}(\mathbf{S} \times \mathbf{E})\right) \rightarrow \tilde{\mathbf{S}}$; for each $s_{j,t} \in \tilde{\mathbf{S}}$ $\phi\left((s_t, \cdot)_j\right)$ is assumed to measurable while for each $\epsilon_{j,t} \in \tilde{\mathbf{E}}$ $\phi\left((\cdot, \epsilon_t)_j\right)$ is continuous.

The above definition is the one given by Futia (1982, Definition 5.1) adjusted in terms of the current analysis. In technical terms this means that product space notation has been used to account for the fact that multiple series of shocks are drawn from the hypothetical distribution. This is a characteristic that bequeaths the present definition with a significant level of flexibility and a taste regarding this property is given below. The above definition simple says that the state of the system evolves form one period to the next in the following way. If the state at j,t is $s_{j,t}$, an event $\epsilon_{j,t} \in \tilde{\mathcal{E}}$ is realized according to the probability law $Q_E\left(\mathsf{X}_{j=1}^{J}(s_t, \cdot)_j\right)$. The state at date $j, t+1$ is then $s_{j,t+1} = (\phi(s_t, \epsilon_{t+1})_j)$.

Random Dynamical Systems can generate very complex dynamics. To analyze the average behavior of sample paths, it is required to define the transition probability function. The following theorem shows how the function ϕ and the probability measure of $\epsilon_{j,t}$ can be used to construct the transition function for the $s_{j,t}$[4].

[1] Double index notation illustrates that the second procedure of simulation is going to be used

[2] For an alternative treatment of the stochastic singularity issue see Bierens (2005)

[3] Separability of the product metric space requires that each component is separable metric space. This is a very essential assumption for the product space analysis undertaken here (see Billingsley, 1968, page:20-21) and it will be maintained for the rest of the thesis

[4] The theorem is similar to the one introduced by Futia (1982, Theorem 5.2, page:398). However, the proof uses the analysis developed by Stockey, Lucas, and Prescott (1989, Theorem 8.9, page:234).

Theorem 5.1.1: Let $\left(\tilde{\mathbf{E}}, \tilde{\mathcal{E}}, \mu = \times_{j=1}^{J} \mu_{\epsilon}\right)$ be a product probability space and let $\left(\tilde{\mathbf{S}}, \tilde{\mathcal{S}}\right)$ be a product measurable space. Let $\phi : \mathsf{X}_{j=1}^{J}(\mathbf{S} \times \mathbf{E}) \to \mathsf{X}_{j=1}^{J}\mathbf{S}$ be a measurable function and define the correspondence $\Gamma : \otimes_{j=1}^{J}\mathcal{S} \to \mathsf{X}_{j=1}^{J}(\mathbf{S} \times \mathbf{E})$ to be the inverse of the ϕ, $\Gamma(A) = \left\{(s,\epsilon)_j \in \mathsf{X}_{j=1}^{J}(\mathbf{S} \times \mathbf{E}) : \phi\left((s,\epsilon)_j\right) \in A\right\}$ all $A \in \mathsf{X}_{j=1}^{J}\mathbf{S}$. Then $Q_E\left(\mathsf{X}_{j=1}^{J}(s,A)_j\right) = \mathsf{X}_{j=1}^{J}\mu_E\left([\Gamma(A)]_{s_j}\right)$ defines a transition probability function on $\left(\tilde{\mathbf{S}}, \tilde{\mathcal{S}}\right)$ ($[\Gamma(A)]_{s_j} = \left\{s_j \in \tilde{\mathbf{S}} : (s,\epsilon)_j \in \mathsf{X}_{j=1}^{J}(\mathbf{S} \times \mathbf{E}), \phi\left((s,\epsilon)_j\right) \in A\right\}$ is s_j-section of $[\Gamma(A)]$ see Definition A.0.1 on Appendix A).

Proof: First we must show that Q_E is well defined. To do this, it suffices to show that $[\Gamma(A)]_{s_j} \in \tilde{\mathcal{E}}$ for all $s_j \in \tilde{\mathbf{S}}$.The desired results follows from (i) the assumption made regarding ϕ (it is measurable), which implies that $\Gamma(A) \in \tilde{\mathcal{S}} \otimes \tilde{\mathcal{E}}$ and (ii) the theorem A.0.3 (see Appendix A or for more details Halmos (1950, Section 37)). Next we must show that for each $s_j \in \tilde{\mathbf{S}}$, $Q_E\left(\mathsf{X}_{j=1}^{J}(s,\cdot)_j\right)$ is a probability measure on $\tilde{\mathcal{S}}$. Fix $s_j \in \tilde{\mathbf{S}}$. Clearly, $Q_E\left(\mathsf{X}_{j=1}^{J}(s,\emptyset)_j\right) = \mu(\emptyset) = 0$ and $Q_E\left(\mathsf{X}_{j=1}^{J}(s,\tilde{\mathbf{S}})_j\right) = \mu(\tilde{\mathbf{S}}) = 1$. Also, for any disjoint sequence $\{A_i\}$ in $\tilde{\mathcal{S}}$ the sets $C_i = \Gamma(A_i)$ $i = 1, 2, ...,$ in $\tilde{\mathcal{S}} \otimes \tilde{\mathcal{E}}$ are disjoint [5]. Hence their s_j-section are also disjoint so $Q_E\left(\mathsf{X}_{j=1}^{J}(s, \bigcup_{i=1}^{\infty} A_i)_j\right) = \mathsf{X}_{j=1}^{J}\mu_E\left(\left[\bigcup_{i=1}^{\infty} C_i\right]_{s_j}\right) = \mu\left(\bigcup_{i=1}^{\infty}[C_i]_{s_j}\right) = \sum_{i=1}^{\infty}\mu\left([C_i]_{s_j}\right) = \sum_{i=1}^{\infty} Q_E\left(\mathsf{X}_{j=1}^{J}(s,A_i)_j\right)$. Therefore $Q_E\left(\mathsf{X}_{j=1}^{J}(s,\cdot)_j\right)$ is countably additive. Finally, we must show that for each $A \in \tilde{\mathcal{S}}$, $Q_E\left(\mathsf{X}_{j=1}^{J}(\cdot,A)_j\right)$ is $\tilde{\mathcal{S}}$-measurable function. Since ϕ is measurable then for each $A \in \tilde{\mathcal{S}}$ the set $C = \Gamma(A) = \phi^{-1}(A)$ is in $\tilde{\mathcal{S}} \otimes \tilde{\mathcal{E}}$, therefore, it suffices to show that the function $\mu(C_{s_j})$, viewed as function of s_j, is $\tilde{\mathcal{S}}$ measurable for all $C \in \tilde{\mathcal{S}} \otimes \tilde{\mathcal{E}}$. Let $\mathcal{G} \equiv \left\{C \in \tilde{\mathcal{S}} \otimes \tilde{\mathcal{E}} : \mu(C_{s_j}) \text{ is a measurable function of } s_j\right\}$. By the Monotone Class Lemma (see Lemma A.0.1 on Appendix A), it suffices to show that $\mathcal{G}$ contains all the finite unions of measurable rectangles and that $\mathcal{G}$ is a monotone class. First, let $C = A \times B$ where $A \in \tilde{\mathbf{S}}$ and $B \in \tilde{\mathbf{E}}$ with $\mu(C_{s_j}) = \mu(B)\mathbf{1}_{[s_j \in A]}$. Since A is a measurable set, $\mu(C_{s_j})$ is a measurable function of s_j because the latter is a *simple* function, which implies that $C \in \mathcal{G}$. From Theorem A.0.11 and the *inclusion-exclusion formula*

$$\mu\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n}\mu(A_j) - \sum_{j\neq k}\mu(A_j \cap A_k) + \sum_{j\neq k\neq l}\mu(A_j \cap A_k \cap A_l) - ... \pm \mu(A_1 \cap A_2 \cap ... \cap A_n) \tag{5.1}$$

,where the sign of the last term is negative if n is even and positive if n is odd and there are $2^n - 1$ in the sum in total, we know that if $G_1, ..., G_n$ are measurable rectangles, then $\bigcup_{i=1}^{n} G_i \in \mathcal{G}$. (5.1) is obtained by induction; the claim holds for

[5] There is always a way to write the set of all finite unions of measurable rectangles as the finite union of the set of all measurable rectangles (see Stockey, Lucas, and Prescott, 1989, Exercise 7.27, page:196)

$n = 1$. Suppose that it holds for $n - 1$.

$$\begin{aligned}
\mu\left[\left(\bigcup_{i=1}^{n} G_i\right)_{s_j}\right] &= \mu\left[\bigcup_{i=1}^{n}(G_i)_{s_j}\right] \\
&= \mu\left[\bigcup_{i=1}^{n-1}(G_i)_{s_j} \cup (G_n)_{s_j}\right] \\
&= \mu\left[\bigcup_{i=1}^{n-1}(G_i)_{s_j}\right] + \mu\left[(G_n)_{s_j}\right] - \mu\left[\left(\bigcup_{i=1}^{n-1}(G_i)_{s_j}\right) \cap (G_n)_{s_j}\right] \\
&= \mu\left[\bigcup_{i=1}^{n-1}(G_i)_{s_j}\right] + \mu\left[(G_n)_{s_j}\right] - \mu\left[\left(\bigcup_{i=1}^{n-1}(G_i \cap G_n)\right)\right] \qquad (5.2)
\end{aligned}$$

Finally, we will show that $\mathcal{G}$ is a monotone class. Let $G_1 \subseteq G_2 \subseteq \ldots$ be an increasing sequence of sets in $\mathcal{G}$, with $G = \bigcup_{i=1}^{\infty} G_i$. Note that this implies that the $\tilde{\mathbf{S}}$ sections $(G_1)_{s_j} \subseteq (G_2)_{s_j} \subseteq \ldots$ form an increasing sequence in $\tilde{\mathcal{E}}$. From theorem A.0.12[6] we know that μ is continues which implies that $\mu\left((G_i)_{s_j}\right) \rightarrow \mu\left((G)_{s_j}\right)$, which the pointwise limit of a sequence of measure functions, since by hypothesis $\mu\left((G_i)_{s_j}\right)$ is measurable function of s_j for each i. From theorem A.0.13 and the equality, $\liminf_i \mu\left((G_i)_{s_j}\right) = \limsup_i \mu\left((G_i)_{s_j}\right)$, we know that $\lim_i \mu\left((G_i)_{s_j}\right)$ is also a measurable function and G is in $\mathcal{G}$. We proceed in a similar way when $\{G_i\}$ is a nested decreasing sequence in $\mathcal{G}$. Hence, it is a monotone class. ■

Given the transition function and any initial probability measure μ_0 on $\tilde{\mathcal{S}}$ the evolution of future probabilities is determined by the the recursion law

$$\mu_{t+1}(A) = \int \cdots \int_S Q_E\left(\mathsf{X}_{j=1}^{J}(s, A)_j\right) \mu_t(ds_j) = \int_{\tilde{\mathbf{S}}} Q_E\left(\mathsf{X}_{j=1}^{J}(s, A)_j\right) \mu_t(ds_j) \qquad (5.3)$$

The necessity rises here is those conditions, which guarantee that equation 5.3 converges to an invariant or stationary distribution,

$$\mu^*(A) = \int \cdots \int_S Q_E\left(\mathsf{X}_{j=1}^{J}(s, A)_j\right) \mu^*(ds_j) = \int_{\tilde{\mathbf{S}}} Q_E\left(\mathsf{X}_{j=1}^{J}(s, A)_j\right) \mu^*(ds_j)$$

. Futia (1982) and Santos and Peralta-Alva (2005) advise that this could be easily done through the use of *Markov-Kakutani*'s theorem (see Appendix A Theorem A.0.6). In order to verify the conditions required by the latter theorem some additional notation should be introduced.

[6] This actually holds generally, not excluding the case $\mu\left((G)_{s_j}\right) = \infty$ see Davidson (1994, page:38)

Let $\mathbf{C}(\tilde{\mathbf{S}})$ be the space of all continuous real-valued functions f on $\tilde{\mathbf{S}}$. The integral $\int_{\tilde{\mathbf{S}}} f(s_j)\ \mu d(s_j)$ or expected value of f over μ will be denoted by $E(f)$ whenever the distribution μ is clear from the context. The weak topology is the coarsest topology such that every functional in the set $\left(\mu \rightarrow \int_{\tilde{\mathbf{S}}} f(s_j)\mu d(s_j), f \in \mathbf{C}(\tilde{\mathbf{S}})\right)$ is continuous. If we endow $\mathbf{C}(\tilde{\mathbf{S}})$ with the following norm $\parallel f \parallel = \sup_{s_j \in \tilde{\mathbf{S}}} |f(s_j)|$ then the latter is a Banach space. A transition function Q_E on $(\tilde{\mathbf{S}}, \tilde{\mathcal{S}})$ defines a Markov operator T from $\mathbf{C}(\tilde{\mathbf{S}})$ to itself, $(Tf) = \int f(s_{j,t+1}) Q_E \left(\times_{j=1}^{J}(s_t, ds_{t+1})_j\right)$ all $s_j \in \tilde{\mathbf{S}}$, and each operator T is associated with its adjoint T^* such that for $\langle f, \mu \rangle = \int_{\tilde{S}} f(s_j)\mu(ds_j)$ we have $\langle Tf, \mu \rangle = \langle f, T^*\mu \rangle$ for all f and μ $\left(\text{where } T^*\mu = \int_{\tilde{S}} Q_E \left(\times_{j=1}^{J}(s, A)_j\right) \mu(ds_j) \text{ all } A \in \tilde{\mathcal{S}}\right)$.

Theorem 5.1.2: Given the definition 5.1.1 there exists a probability measure μ^* such that $\mu^* = T\mu^*$.

Proof: The proof is established when the hypotheses of *Markov-Kakutani's* Theorm (see Appendix A Theorem A.0.6) are satisfied. Santos and Peralta-Alva (2005, Assumption 2) make a convenient assumption that shows that T maps $\mathbf{C}(\tilde{\mathbf{S}})$ into itself. On the other hand Futia (1982, Proposition 5.6, page:400) proves that the latter is true (see Appendix A Proposition A.0.1). Hence, operator T^* must be weakly continuous. Moreover, from Bertsekas and Shreve (1996, Proposition 7.22, page:130) we know that in the weak topology, $(\mathbf{C}(\tilde{\mathbf{S}})^*)$, the set of all probability measures in $\tilde{\mathcal{S}}$, $(P(S))$, is compact, consequently a convex set. Therefore, *Markov-Kakutani's* Theorem conditions are satisfied.■

The analysis for the rest of this section moves on the lines drawn by Santos and Peralta-Alva (2005). The assumptions adopted by these authors are not so hard to be checked in applications; a characteristic that makes the whole study attractive. In the present work ϕ is a real-matrix-valued function on $\mathbb{R}^{Tds \times J}$ and $\| \cdot \|$ is its matrix norm in $\mathbb{R}^{Tds \times J}$. One more thing, which is going to be used extensively in the rest of this section, is *Fubini's* Theorem (see Appendix A Theorem A.0.7 or, for more details, Halmos (1950, Chapter 7)). Equation ϕ could also be viewed as the mapping $\phi : \mathbf{S} \times \mathbf{E} \times \mathbf{J} \rightarrow \mathbf{S} \times \mathbf{J}$. This notation facilitates the rest analysis a lot. Following Santos and Peralta-Alva (2005) we define for any two real-matrix-valued functions the following metric

$$d(\phi, \hat{\phi}) = \max_{s \in \tilde{S}} \left[\int_{E \times J} \parallel \phi\left((s, \epsilon)_j\right) - \hat{\phi}\left((s, \epsilon)_j\right) \parallel Q_{EJ}(d\epsilon_j) \right] \tag{5.4}$$

Using Definition A.0.2 and *Fubinis's* Theorem we can rewrite equation 5.4 as

$$\begin{aligned} d(\phi,\hat{\phi}) &= \max_{s\in\tilde{S}}\left[\int_{E\times J} \parallel \phi\left((s,\epsilon)_j\right) - \hat{\phi}\left((s,\epsilon)_j\right) \parallel Q_{EJ}(d\epsilon_j)\right] \\ &= \max_{s\in\tilde{S}}\left[\int_{E\times J} \parallel \phi(s,\epsilon) - \hat{\phi}(s,\epsilon) \parallel_j Q_{EJ}(d\epsilon_j)\right] \\ &= \max_{s\in\tilde{S}}\left[\int_{E} \parallel \phi(s,\epsilon) - \hat{\phi}(s,\epsilon) \parallel Q_{E}(d\epsilon)\right] \end{aligned} \tag{5.5}$$

The last equality is obtained by setting $\int_J \parallel \phi(s,\epsilon) - \hat{\phi}(s,\epsilon) \parallel_j Q_J(dj) = \parallel \phi(s,\epsilon) - \hat{\phi}(s,\epsilon) \parallel$ and using *Fubinis's* Theorem. Equation 5.5 is identical to Equation (3.1) of Santos and Peralta-Alva (2005, page:1945). From the analysis so far it is clear that each ϕ_i defines the associated pair $(Q_{E,i}, T_i^*)$. While from Theorem 5.1.2 there always exist an invariant distribution $\mu_i^* = T_i^*\mu_i^*$. The aim of this section is to establish those conditions that guarantee that the moments of an invariant distribution of a sufficiently good numerical approximation must be arbitrarily close to the moments of some invariant distribution of the model. The following Theorem is in this direction.

Theorem 5.1.3: Let $\{\phi_i\}$ be a sequence of functions that converge to ϕ. Let $\{\mu_i^*\}$ be a sequence of probabilities on $\tilde{S}$ such that $\mu_i^* = T_i^*\mu_i^*$ for each i. Under the assumptions made regarding the compactness of $\tilde{S}$ and the weak topology $(\mathbf{C}(\tilde{\mathbf{S}})^*)$, if μ^* is a weak limit point of $\{\mu_i\}$, then $\mu^* = T\mu^*$.

Proof: Santos and Peralta-Alva (2005), basen on the analysis of Dubins and Freedman (1966, Theorem 3.4, page:839), mention that the Theorem will be established if we could show the continuity of the evaluation of the map $ev(\phi,\mu) = \phi\cdot\mu$.

An alternative way to define the topology of weak convergence is by using the following metric

$$d(\mu,v) = \sup_{\tilde{f}\in\tilde{\mathcal{L}}}\left\{\left|\int_{\mathbf{S}\times\mathbf{J}} \tilde{f}(s_j)\mu(ds_j) - \int_{\mathbf{S}\times\mathbf{J}} \tilde{f}(s_j)v(ds_j)\right|\right\} \tag{5.6}$$

where $\tilde{\mathcal{L}}$ is the set of Lipschitz functions on $\tilde{S}$. The following equation $\int_{\mathbf{S}\times\mathbf{J}} \tilde{f}(s_j)\mu(ds_j)$ can be rewritten as $\int_S \left(\int_J \tilde{f}(s,j)\mu_j(dj)\right)\mu_s(ds)$ (by definition). At this point we set $f(s)$ equals to $\int_J \tilde{f}(s,j)\mu_j(dj)$ then from *Fubini's* Theorem (see Appendix A Theorem A.0.7) we have $\int_{\mathbf{S}\times\mathbf{J}} \tilde{f}(s_j)\mu(ds_j) = \int_S f(s)d\mu_s(ds)$. Applying the same logic to

the second part of the right hand side (RHS) of equation 5.6 we get

$$\begin{aligned} d(\mu,v) &= \sup_{\tilde{f}\in\tilde{\mathcal{L}}}\left\{\left|\int_{\mathbf{S}\times\mathbf{J}}\tilde{f}(s_j)\mu(ds_j)-\int_{\mathbf{S}\times\mathbf{J}}\tilde{f}(s_j)v(ds_j)\right|\right\} \\ &= \sup_{f\in\mathcal{L}}\left\{\left|\int_{\mathbf{S}}f(s)\mu_s(ds)-\int_{\mathbf{S}}f(s)v_s(ds)\right|\right\} \end{aligned} \tag{5.7}$$

The last equality of the above equation is identical to equation (A.1) of Santos and Peralta-Alva (2005, proof of Theorem 2, page:1968). It should be noticed that $\mathcal{L}$ is, now, the set of Lipschitz functions on $\mathbf{S}$ with constant ζ and such that $-1 \leq f \leq 1$. From this point the proof is identical to the one given by Santos and Peralta-Alva (2005, proof of Theorem 2). For any two mappings $\phi = \int_J \phi_J(s,\epsilon)Q_J(d_j)$ and $\hat{\phi} = \int_J \hat{\phi}_J(s,\epsilon)Q_J(d_j)$ and any two measures μ_S and v_S we have[7]

$$\begin{aligned} &\left|\int_{\mathbf{S}} f(s)\left[\phi\cdot\mu_S\right]-\int_{\mathbf{S}} f(s)\left[\hat{\phi}\cdot v_S\right]\right| \\ &= \left|\int_{\mathbf{S}}\left[\int_E f\left(\phi(s,\epsilon)\right)Q_E(d\epsilon)\right]\mu_S(ds)-\int_{\mathbf{S}}\left[\int_{\mathbf{E}} f\left(\hat{\phi}(s,\epsilon)\right)Q_E(d\epsilon)\right]v_S(ds)\right| \\ &\leq \left|\int_{\mathbf{S}}\left[\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right)Q_E(d\epsilon)\right]\mu_S(ds)-\int_{\mathbf{S}}\left[\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right)Q_E(d\epsilon)\right]v_S(ds)\right| \\ &+ \left|\int_{\mathbf{S}}\left[\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right)Q_E(d\epsilon)\right]v_S(ds)-\int_{\mathbf{S}}\left[\int_{\mathbf{E}} f\left(\hat{\phi}(s,\epsilon)\right)Q_E(d\epsilon)\right]v_S(ds)\right| \\ &\leq \left|\int_{\mathbf{S}}\left[\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right)Q_E(d\epsilon)\right]\left[\mu_S(ds)-v_S(ds)\right]\right| + d(\phi,\hat{\phi}) \end{aligned} \tag{5.8}$$

The first inequality arise from the triangle inequality, while the second one from the fact that $f \in \mathcal{L}$. Then by 5.7 the Theorem will be established if we could show that for every arbitrary $\eta > 0$ there exists a weak neighborhood $V(\mu_S)$ of μ_S such that for all v_S in $V(\mu_S)$ and all $\int_E f\left(\phi(s,\epsilon)\right)Q_E(d\epsilon)$ in $\tilde{\mathcal{L}}$

$$\begin{aligned} &V_\eta\left(\mu_S : \int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right)Q_E(d\epsilon)\right) \\ &= \left\{v_S\in\mathcal{P}(\mathcal{S}) : \int_{\mathbf{S}}\left[\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right)Q_E(d\epsilon)\right]\left[\mu_S(ds)-v_S(ds)\right] < \eta\right\} \end{aligned} \tag{5.9}$$

The condition of *Arzela-Ascoli's* Theorem (see Appendix A, Theorem A.0.8) are satisfied, this means that the set $\mathcal{L}$ is compact. Hence,we can define a set of elements $\{f_i\}$ such that for every f in $\mathcal{L}$ an element f_i satisfies that $\sup|f-f_i| < \frac{\eta_0}{3}$.

[7] The subscript denotes the section

Additionally, by definition $\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)$ is continuous in $\mathbf{S}$. Hence, for every f_i there exists a weak neighborhood $V_i(\mu_S)$ such that for all v_S in $V_i(\mu_S)$

$$\begin{aligned} & V_\eta\left(\mu_S, k, \int_{\mathbf{E}} f_1\left(\phi(s,\epsilon)\right) Q_E(d\epsilon), ..., \int_{\mathbf{E}} f_k\left(\phi(s,\epsilon)\right) Q_E(d\epsilon), \frac{\eta_0}{3}\right) \\ &= \left\{ v_S \in \mathcal{P}(\mathcal{S}) : \int_{\mathbf{S}} \left[\int_{\mathbf{E}} f_i\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)\right] \left[\mu_S(ds) - v_S(ds)\right] < \frac{\eta_0}{3}, i = 1, ..., k \right\} \end{aligned} \tag{5.10}$$

By ranging over all the possible $f_1, ..., f_k$ and η_0, for each $k \in \mathbb{N}$, we define a collection of open neighborhoods of μ_S. The base collection $V_\eta\left(\mu_S, k, \int_E f_i\left(\phi(s,\epsilon)\right) Q_E(d\epsilon), \frac{\eta_0}{3}\right)$ defines a weak topology on $\mathcal{P}(\mathcal{S})$. Define now

$$\eta_0 = \eta - \left| \int_{\mathbf{S}} \left[\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)\right] \left[\mu_{0S}(ds) - v(ds)\right] \right| > 0$$

Then for any $v_S \in V_{\frac{\eta}{3}}\left(v_S : \int_{\mathbf{E}} f_i\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)\right)$

$$\begin{aligned} & \left| \int_{\mathbf{S}} \left[\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)\right] \left[\mu_S(ds) - v_S(ds)\right] \right| \\ &\leq \left| \int_{\mathbf{S}} \left[\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)\right] \mu_S(ds) - \int_{\mathbf{S}} \left[\int_{\mathbf{E}} f_i\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)\right] \mu_S(ds) \right| \\ &+ \left| \int_{\mathbf{S}} \left[\int_{\mathbf{E}} f_i\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)\right] \left[\mu_S(ds) - \mu_{0S}(ds)\right] \right| \\ &+ \left| \int_{\mathbf{S}} \left[\int_{\mathbf{E}} f_i\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)\right] \mu_{0S}(ds) - \int_{\mathbf{S}} \left[\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)\right] \mu_{0S}(ds) \right| \\ &+ \left| \int_{\mathbf{S}} \left[\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)\right] \left[\mu_{0S}(ds) - v_S(ds)\right] \right| \\ &< \frac{\eta_0}{3} + \frac{\eta_0}{3} + \frac{\eta_0}{3} + \left| \int_{\mathbf{S}} \left[\int_{\mathbf{E}} f\left(\phi(s,\epsilon)\right) Q_E(d\epsilon)\right] \left[\mu_{0S}(ds) - v(ds)\right] \right| \\ &= \eta \end{aligned} \tag{5.11}$$

.■

As it is described by Santos and Peralta-Alva (2005) the quintessence of the above theorem relies on the fact that the correspondence of the invariant distributions is closed (bilinear convergence of $T_i^* \mu_i^*$ to $T^* \mu*$ in the weak topology), and given the fact that the set of all probability measures in $\tilde{\mathcal{S}}$ is compact we can conclude that the correspondence of the invariant distributions is upper semicontinuous. Mainly, upper semicontinuity is established by imposing assumptions on the transition function Q_E

(see Stockey, Lucas, and Prescott, 1989, Section 12.4), however, the latter function is not handled easily. It is usually derived by the knowledge of ϕ and any assumption imposed on Q_{EJ} have to go through conditions applied on ϕ, which cannot be done straightforwardly.

5.1.2 A Robust Law of Large Numbers for DSGE models

Laws of Large Numbers *LLN* for Random Dynamical Systems are based on the ergodic theorem. The application of the latter requires the initial state $s_{j,0}$ must lies into an ergodic set. Mainly, this is guaranteed either by imposing a technical condition, known as *Hypothesis D* (Doob, 1953; Stockey, Lucas, and Prescott, 1989), which ensures that for every initial value $s_{j,0}$, dynamic system will enter one of its ergodic sets almost surely, or by assuming a unique invariant distribution (Breiman, 1960).

However, both assumptions seem to be weak in the applied work. For instance, both Santos and Peralta-Alva (2005, Section 3.2) and Stockey, Lucas, and Prescott (1989, Chapter 11) declare that *Hypothesis D* is difficult to be verified in economic applications. On the other hand uniqueness of the invariant distribution is viewed as strict assumption for numerical approximations. Even if conditions, which ensure uniqueness, would be imposed on the original model they would be hardly preserved during the discretization of the state space[8].

Santos and Peralta-Alva (2005, Theorem 3, page:1950) goes a step further and derives a *LLN* robust to the above mentioned weakness. The task of this subsection is to make the latter result compatible with our current setup. We define a new probability space that comprises all infinite sequences $\omega = (\epsilon_1, \epsilon_2, ...)$. Let $\Omega = \mathsf{X}_{t=1}^{\infty} E$ and $\mathcal{F}$ be the σ-filed in $\mathsf{X}_{t=1}^{\infty} E$ generated by the collection of all cylinders

$$\begin{bmatrix} A_{1,1} \times A_{1,2} \times ... \times A_{1,T} \times \mathsf{X}_{t=1}^{\infty} E \\ A_{2,1} \times A_{2,2} \times ... \times A_{2,T} \times \mathsf{X}_{t=1}^{\infty} E \\ \vdots \\ A_{J,1} \times A_{J,2} \times ... \times A_{J,T} \times \mathsf{X}_{t=1}^{\infty} E \end{bmatrix} \text{ where } A_{j,i} \in \mathbf{E} \text{ for } j = 1, .., J \text{ and } i = 1, ..., T.$$

A probability measure λ can be constructed over these finite dimensional sets as

$$\lambda \left\{ \omega_j : \begin{bmatrix} \epsilon_{1,1} \in A_{1,1} & \epsilon_{1,2} \in A_{1,2} & ... & \epsilon_{1,T} \in A_{1,T} \\ \vdots & \vdots & ... & \vdots \\ \epsilon_{J,1} \in A_{J,1} & \epsilon_{J,2} \in A_{J,2} & ... & \epsilon_{J,T} \in A_{J,T} \end{bmatrix} \right\} = \prod_{i=1}^{T} \prod_{j=1}^{J} Q_E \left(A_{j,i} \right) \quad (5.12)$$

From Caratheodory and Hahn extension Theorems (see Appendix A Theorem A.0.9

[8] An example regarding this failure is given by Santos and Peralta-Alva (2005, Section 3.2, page:1948)

and Theorem A.0.10, respectively) λ has a unique extension on $\mathcal{F}$, which allows us to define the following product probability space $(\Omega, \mathcal{F}, \lambda)$.

Theorem 5.1.4: Let f belongs to $\mathbf{C}(\tilde{\mathbf{S}})$. Then, given the Definition 5.1.1 the following limits exist for λ-almost all ω_j.

$$\lim_{T\to\infty}\left(\inf_{s_0\in\tilde{\mathbf{S}}}\left[\frac{1}{T}\frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J} f\left(s_{j,t}(s_0,\omega_j)\right)\right]\right) = E^{\min}(f) \tag{5.13}$$

$$\lim_{T\to\infty}\left(\sup_{s_0\in\tilde{\mathbf{S}}}\left[\frac{1}{T}\frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J} f\left(s_{j,t}(s_0,\omega_j)\right)\right]\right) = E^{\max}(f) \tag{5.14}$$

Proof: The proof is based on Crauel (2002, Chapter 6) and Santos and Peralta-Alva (2005) results[9]. Time t is allowed to range from $-\infty$ to ∞, which implies that $\omega_j = (..., \epsilon_{-t}, ..., \epsilon_{-1}, \epsilon_0, \epsilon_1, ..., \epsilon_t, ...)_j$. Similar to equation 5.12 we can define a new probability measure and a probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\lambda})$. Let $G_T = \sup_{s_0\in\tilde{\mathbf{S}}}\left[\sum_{t=1}^{T}\sum_{j=1}^{J} f\left(s_{j,t}(s_0,\omega_j)\right)\right]$ $= \sup_{s_0\in\tilde{\mathbf{S}}}\left[\sum_{t=1}^{T}\tilde{f}\left(s_t(s_0,\omega)\right)\right]$, where $\tilde{f}\left(s_t(s_0,\omega)\right) = \sum_{j=1}^{J} f\left(s_{j,t}(s_0,\omega_j)\right)$. In order to use the ergodic theorem of Kingman (1968) we need to show that G_T is a subadditive process, which is done through the use of the shift operator. For each integer k the k-shift operator is defined as $\varpi_k : \hat{\Omega} \to \hat{\Omega}$, $\varpi_k(..., \epsilon_{-t}, ..., \epsilon_{-1}, \epsilon_0, \epsilon_1, ..., \epsilon_t, ...)_j = (..., \epsilon_{-t+k}, ..., \epsilon_{-1+k}, \epsilon_k, \epsilon_1 + k, ..., \epsilon_t + k, ...)_j$. Easily, you could see that ϖ_k is bijective and measurable[10]. Crauel (2002, page:96) shows that $G_{T+k}(\omega_j) \leq G_T(\omega_j) + G_k(\omega_j)$ (for positive integers k and T) is derived through straightforward calculations. This means that $G_T(\omega_j)$ is subadditive process and from Kingman (1968) we know that

$$\lim_{T\to\infty}\frac{G_T(\omega_j)}{T} = H \tag{5.15}$$

for λ-almost all ω_j. By Kingman (1968) result equality 5.13 will be established if we can show that $\int_{\mathbf{S}\times\mathbf{J}} f(s_j)\mu^*(ds_j) = H$. For $\eta > 0$ and $t \in \mathbb{N}$ consider the set

$$H_N^{\eta} = \left\{(\omega_j, s_0) : \sum_{t=1}^{T}\sum_{j=1}^{J} f\left(s_{j,t}(s_0,\omega_j)\right) \geq H - \eta\right\}$$

. Due to continuity assumption (Proposition A.0.1) the latter set is measurable. A primitive result of the Measure Theory result indicates that there exist a measurable h_N function on $Proj_{\hat{\Omega}}(H_N^{\eta})$ such that $(\omega_j, h_N(\omega_j)) \in H_N^{\eta}$. Following Santos and Peralta-Alva (2005) the empirical distribution $\mu^T_{(s_0,\omega_j)}$ for any sample path

[9] The latter results could be viewed as an extension of the previous work.

[10] For full discussion of its properties see Crauel (2002, page:82-84)

$\left\{\{s_{j,t}(s_0,\omega_j)\}_{j=1}^{J}\right\}_{t=1}^{T}$ is defined as $\mu^T_{(s_0,\omega_j)}(s_{j,t}(s_0,\omega_j)) = \frac{1}{TJ}$ for every $s_{j,t}(s_0,\omega_j)$, for every $1 \leq t \leq T$ and $1 \leq j \leq J$. Then let $\mu^N = \int_{\hat{\Omega}} \mu^T_{(h_N(\omega_j),\omega_j)}\lambda(d\omega)$. By Corollary 6.13 of Crauel (2002, page:88) there exists a subsequence μ^{T_n} ($n \in \mathbb{N}$) converging to an invariant distribution μ^*. Upper semi-continuity of $\mu \to \int_S f(s)\mu d(s)$ together with the definition of μ^T gives

$$\begin{aligned}\int_S f(s)\mu^* d(s) &\geq \lim_{n\to\infty}\int_S f(s)\mu^{T_n} d(s) \\ &= \lim_{n\to\infty} E\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{f}\left(s_t(s_0,\omega)\right)\right) \\ &\geq H-\eta \end{aligned} \tag{5.16}$$

We set $R_f = \sup\left\{\int_S f(s)\mu d(s) : \mu(K)=1,\ \mu \text{ invariant for } \phi\right\}$[11], then $R_f \geq H-1$, say. Crauel (2002, page:97) proves that the set $\{\mu : \mu(K)=1, \mu\ \phi-\text{invariant}$, and $\int_S f(s)\mu d(s) \geq H-1\}$ is compact and nonvoid. Thus R_f is realized as the maximum instead of just a supremum, which implies, $\{\mu : \mu(K)=1, \mu\ \phi-$invariant, and $\int f d\mu = R_f\} \neq \emptyset$. Since extremal points of this compact convex set are extremal in the set $\{v : v, \phi-\text{invariant}\}$, hence ergodic, there exist an ergodic ϕ-invariant μ supported by K with $\int_S f(s)\mu d(s) = R_f$. From 5.16 we have

$$\int_S f(s)\mu d(s) = R_f \geq \int_S f(s)\mu^* d(s) \geq H-\eta$$

for $\eta > 0$ arbitrary gives $\int_S f(s)\mu d(s) = R_f = \int_S f(s)\mu^* d(s) = H$. ■

As it is discussed by Santos and Peralta-Alva (2005), the above Theorem ensures that $\frac{1}{T}\frac{1}{J}\sum_{t=1}^{T}\sum_{j=1}^{J} f\left(s_{j,t}(s_0,\omega_j)\right)$ approaches the interval $\left[E^{\min}(f) = \min_{\{\mu*:\mu*=T^*\mu^*\}} \int_S f(s)\mu^* d(s),\ E^{\max}(f) = \max_{\{\mu*:\mu*=T^*\mu^*\}}\int_S f(s)\mu^* d(s)\right]$[12] *uniformly* in s_0 for λ-almost all ω_j, and the bounds $E^{\min}(f)$ and $E^{\max}(f)$ are tight. Breiman (1960) theorem is obtained as corollary of the above Theorem (See Santos and Peralta-Alva, 2005, Corollary 3, page:1950) under the condition that there exist a unique invariant distribution μ^*. In this case

$$E^{\min}(f) = E^{\max}(f) = E(f)$$

[11] K is the compact support of the invariant measures for ϕ

[12] The integral $\int_S f(s)\mu^* d(s)$ is the iterated integral over the j-sections.

5.2 Extrema of Random Functions

The testing procedure described in the next chapters relies on the estimated parameter vector of a time series model (γ) fitted to the data simulated by a structural model whose (structural)parameter vector (θ) has been estimated.

$$\tilde{\gamma} = \arg\max_{\gamma\in\Gamma} L_R\left(y_t\left(\arg\max_{\theta\in\Theta} L_S(y_t,\theta)\right),\gamma\right)^{13} \tag{5.17}$$

It is clear from the above formula that $\tilde{\gamma}$ is a constrained estimate, to be precise, it would be the time series estimate implied by the structural model if the latter could be written as the former model.

We could view this as two stage estimation problem

$$L(y_t, y_t(\theta), \theta, \gamma) = \begin{bmatrix} L_S(y_t,\theta) \\ L_R\left(y_t(\hat{\theta}),\gamma\right) \end{bmatrix} \tag{5.18}$$

with $\nabla_{\theta,\gamma} L = 0$[14]. Since Quasi Maximum Likelihood (QML) estimation techniques are used in both stages, White (1994) notation and terminology will be used for the rest of this section. The complication, addressed in this section, is the fact that Γ is dependent of $\hat{\theta}$ and the realizations of the stochastic processes u_t and v_t. In other words Γ becomes random (Γ_T).

The properties (Consistency and Normality) of a two-stage QML estimator are well established (see White, 1994). However, these results cannot be directly used. We must show that the supremum over Γ_T is a itself a random or a measurable function. Fortunately, Stinchcombe and White (1992) - using *Analytic Set Theory* - and White (1994, Theorem 8.17, page:185) - verifying the conditions of the Theorem 4.5 of Debreu (1967) - show that this is true under some general conditions.

Given an unconstrained QML estimate (QMLE) $\hat{\gamma}$, an estimate of the structural vector $\hat{\theta}$ and the assumption that the mapping from the structural to time series parameter vector ($\delta : \theta \longrightarrow \gamma$) has closed form, a test of $\gamma = \delta(\theta)$ could be based on $\hat{\gamma} - \delta(\hat{\theta})$. However, the latter mapping can be established under very restrictive condition, for instance, for a linearized DSGE model and for a linear time series model e.g low order VAR or VARMA. However, similar to Smith (1993), this mapping can be produced via simulation and this is what we get form equation 5.17. In other words 5.17 could be viewed as $\max_{\gamma\in\Gamma} L_R(y_t,\hat{\theta},\gamma)$ *s.t.* $r_T(\hat{\theta},\gamma) = 0$.

Assumption 5.2.1: Let $(\Omega,\mathcal{F},\mu_\Omega)$, $(\tilde{E},\tilde{\mathcal{E}},\mu_{\tilde{E}})$ and $(\tilde{V},\tilde{\mathcal{V}},\mu_{\tilde{V}})$ be complete separable

[13] L_R is the likelihood of the time series model, while L_S is the likelihood of the structural model.
[14] $\nabla_{\theta,\gamma}$ is the matrix of the cross derivative of θ with respect of γ

probability spaces. Let Θ and Γ be two Souslin topological spaces, with respective Borel σ-fields $\mathcal{N}$ and $\mathcal{M}$. Let $r : \Theta \times \Gamma \to \mathbb{R}$ be a measurable with respect to the product σ-field $\mathcal{N} \otimes \mathcal{M}$, and let $L : \Omega \times E \times V \times \Theta \times \Gamma \to \mathbb{R}$ be measurable in $\Omega \times E \times V \times \Theta$ and continuous in Γ. Further, let $\zeta : \Omega \to \Theta$ be a measurable function and define the correspondence $S : \Omega \times E \times V \to \Gamma$ by $S(\omega, \epsilon, v) = \left\{\gamma \in \Gamma : r(\hat{\theta}, \gamma) = 0\right\}$. Define, $L^*(\omega, \epsilon, v) = \sup\left\{L(\omega, \epsilon, v, \hat{\theta}, \gamma) : \gamma \in S(\omega, \epsilon, v)\right\}$.

Theorem 5.2.1: Given assumption 5.2.1 there exists a measurable function $\tilde{\gamma}_T : \Omega \times E \times V \to \mathcal{M}$ and a set $A_T \in \mathcal{F} \otimes \mathcal{E} \otimes \mathcal{V}$ with $\mu_{\Omega \times E \times V}[A_T] = 1$ such that for all $(\omega, \epsilon, v) \in A_T$

$$L_R\left(\omega, \epsilon, v, \hat{\theta}(\omega, \epsilon, v), \tilde{\gamma}(\omega, \epsilon, v)\right) = \max_{\gamma \in S(\omega, \epsilon, v)} L_R\left(\omega, \epsilon, v, \hat{\theta}(\omega, \epsilon, v), \gamma\right)$$

Proof: By the assumption 5.2.1 the product probability space $(\Omega \times \tilde{E} \times \tilde{V}, \mathcal{F} \otimes \tilde{\mathcal{E}} \otimes \tilde{\mathcal{V}}, \mu_\Omega \times \mu_{\tilde{E}} \times \mu_{\tilde{V}})$ is complete. The rest of the proof follows the proof of Theorem 8.17 of White (1994, page:213). ■

Given the above result the properties of $\tilde{\gamma}$ are well established (see White, 1994, Section 8.2).

6. CONCLUSION

In this thesis we have introduced a new bootstrap method for testing structural DSGE models according to their dynamic performance. The method maintains a separation between the structural (non-linear) model as the null hypothesis and its dynamic time series representation. The model's errors are discovered and used for bootstrapping (after whitening); the resulting pseudo-samples are used to discover the sampling distribution of the dynamic time series model. The test then consists of discovering whether the parameters of the time-series model estimated on the actual data lie within some confidence interval of this distribution. A test statistic for the parameters taken as a whole is developed (the M-metric, a Wald statistic).

In contrast to the existed methods used for the evaluation of the DSGE models, only, the proposed one could be used for non-linear structural models. However, it should be emphasized that this method could not be used for the estimation of the latter model. This is a task that requires further research.

The thesis also discusses whether the simulations we use have the property that their approximation error converges to zero asymptotically. Here we follow the analysis of Santos and Peralta-Alva (2005) for the case of a single error. We establish that in the case of multiple errors the same convergence property holds. This underpins the perturbation methods we have used here for our stochastic simulations.

We demonstrated the use of these methods in two applications- one to the Liverpool Model of the UK over recent postwar data and one to postwar panel data for 76 countries. We tested the Liverpool Model and found that, according to which time series reprsentation of the data we used it was either marginally accepted at the 95% level or at the 99% level; another interesting finding was that when VAR impulse response functions were identified by the Liverpool Model they were not found to be hump-shaped but rather in gereal consistent with the new classical patterns of the Liverpool Model. On the panel data we were able to test between a model in which taxation affected growth negatively and one where there was no effect at all and we showed that the latter was rejected at 95% while the former was accepted.

APPENDIX

A. APPENDIX A

Definition A.0.1: Let $(\mathbf{X}, \mathcal{X})$, $(\mathbf{Y}, \mathcal{Y})$ be measurable spaces and $(\mathbf{Z}, \mathcal{Z})$ is their product. Let $\mathbf{E} \subseteq \mathbf{Z}$ and $x \in \mathbf{X}$. Then the *x-section of* $\mathbf{E}$ is the set (in $\mathbf{Y}$) $E_x = \{y \in \mathbf{Y} : (x, y) \in \mathbf{E}\}$. The *y-section* (a set in $\mathbf{X}$) is defined similarly.

Definition A.0.2: Let $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$, defined as in the above definition. Let $f : \mathbf{Z} \to \mathbb{R}$. Then the *x-section of* f is the function $f_x : \mathbf{Y} \to \mathbb{R}$ defined by $f_x(y) = f(x, y)$. The *y-section* of f, $f_y : \mathbf{X} \to \mathbb{R}$ is defined similarly.

Definition A.0.3: A *monotone class* is a nonempty collection $\mathcal{M}$ of sets such that $\mathcal{M}$ contains:

a. The union of every nested increasing sequence $A_1 \subseteq A_2 \subseteq ...$ of sets in $\mathcal{M}$.

b. The intersection of every nested decreasing sequence $A_1 \supseteq A_2 \supseteq ...$ of sets in $\mathcal{M}$.

Theorem A.0.2: Every section of a measurable set is a measurable set.

Proof: See Halmos (1950, Theorem A, page:141).■

Theorem A.0.3: Every section of a measurable function is a measurable function

Proof: See Halmos (1950, Theorem B, page:142). ■

Theorem A.0.4: Let $(\mathbf{S} \times \mathbf{J}, \mathcal{S} \otimes \mathcal{J})$ be a product measurable space:

a. If $\{A_n\}_{n=1}^{\infty}$ is an increasing sequence in $\mathcal{S} \otimes \mathcal{J}$ that is $A_n \subseteq A_{n+1} \subseteq \ldots$, all n,then $\mu(\bigcup_{n=1}^{\infty}) = \lim_{n\to\infty} \mu(A_n)$.

b. If $\{B_n\}_{n=1}^{\infty}$ is a decreasing sequence in $\mathcal{S} \otimes \mathcal{J}$, that is, if $B_{n+1} \subset B_n \subseteq \cdots$ and if $\mu(B_m) < \infty$ for some m, then $\mu(\bigcap_{n=1}^{\infty}) = \lim_{n\to\infty} \mu(B_n)$.

Proof: a. If $\mu(A_n) = \infty$ for any n, then the result is trivial. Suppose that $\mu(A_n) < \infty$, all n. Let $A_0 = \emptyset$. Then $\{(A_n \backslash A_{n-1})\}$ is a sequence of disjoint sets in $\mathcal{S} \otimes \mathcal{J}$, $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \{(A_n \backslash A_{n-1})\}$. Then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left[\bigcup_{n=1}^{\infty}(A_n \backslash A_{n-1})\right] = \sum_{n=1}^{\infty} \mu(A_n \backslash A_{n_1}) = \lim_{N\to\infty} \sum_{n=1}^{N} \mu(A_n \backslash A_{n_1} \text{(A.1)}$$

At this point we define $A_n \backslash A_{n-1} = \Phi = \Phi \cap A_{n-1}^c \in \mathcal{S} \otimes \mathcal{J}$ and by assumption $(A_n \subseteq A_{n-1})$ we have $A_{n-1} \cup \Phi = A_n$ and $A_{n-1} \cap \Phi = \emptyset$. Hence

$$\mu(A_{n-1}) + \mu(\Phi) = \mu(A_n) \tag{A.2}$$

as $\mu(\Phi) \geq 0$ we have $\mu(A_{n-1}) \leq \mu(A_n)$, since $\mu(A_n) < \infty$ for all n the difference $\mu(A_n) - \mu(A_{n-1})$ is well defined and

$$\mu(\Phi) = \mu(A_n \backslash A_{n-1}) = \mu(A_n) - \mu(A_{n-1}) \tag{A.3}$$

which implies that equation A.1 is equal to $\lim_{N\to\infty} \sum_{n-1}^{N} \left[\mu(A_n) - \mu(A_{n-1})\right] = \lim_{N\to\infty} \mu(A_n)$

b. Without loss of generality we assume that $\mu(B_1) < \infty$. Then it is sufficient to show that $\mu(B_1) - \mu\left(\bigcap_{n=1}^{\infty} B_n\right) = \mu(B_1) - \lim_{n\to\infty} \mu(B_n)$ using the same logic the last equality can be written as $\mu\left(B_1 \backslash \bigcap_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \left[\mu(B_1) - \mu(B_n)\right] = \lim_{n\to\infty} \mu(B_1 \backslash B_n)$. $\{B_1 \backslash B_n\}$ is an increasing sequence in $\mathcal{S} \otimes \mathcal{J}$ and $\bigcup_{n=1}^{\infty} (B_1 \backslash B_n) = B_1 \backslash \bigcap_{n=1}^{\infty} B_n$. Therefore,

$$\mu\left(B_1 \backslash \bigcap_{n=1}^{\infty} B_n\right) = \mu\left[\bigcup_{n=1}^{\infty} (B_1 \backslash B_n)\right] = \lim_{n\to\infty} \mu(B_1 \backslash B_n) = \lim_{n\to\infty} \mu(B_n)$$

.■

Theorem A.0.5: Let $(\mathbf{S} \times \mathbf{J}, \mathcal{S} \otimes \mathcal{J})$ be a measurable space and let $\{f_n\}$ be a sequence of $\mathcal{S} \otimes \mathcal{J}$ of measurable functions converging pointwise to f

$$\lim_{n\to\infty} f_n(s_j) = f(s_j) \qquad \text{all } s_j = (s, j) \in \mathbf{S} \times \mathbf{J} \tag{A.4}$$

Then f is also measurable.

Proof: We want to show that for any $\alpha \in \mathbb{R}$

$$A \equiv \left\{s_j = (s, j) \in \mathbf{S} \times \mathbf{J} | f(s_j) \leq \alpha\right\} \in \mathcal{S} \otimes \mathcal{J} \tag{A.5}$$

Fix $\alpha \in \mathbb{R}$ and for $k = 1, 2, ...$ and $n = 1, 2, ...$, let $A_{nk} \equiv \left\{s_j = (s, j) \in \mathbf{S} \times \mathbf{J} | f_n(s_j) \leq \alpha + \frac{1}{k}\right\}$, since f_n is measurable, $A_{nk} \in \mathcal{S} \otimes \mathcal{J}$, all n, k. Then, since $\mathcal{S} \otimes \mathcal{J}$ is closed under countable unions and intersections, it follows that $B_{Nk} = \bigcap_{n=N}^{\infty} A_{nk} \in \mathcal{S} \otimes \mathcal{J}$ all N, k; so that $B_k = \bigcup_{N=1}^{\infty} B_{Nk} \in \mathcal{S} \otimes \mathcal{J}$ all k; and finally, that $B = \bigcap_{k=1}^{\infty} B_k \in \mathcal{S} \otimes \mathcal{J}$. Hence, it is sufficient to show that $B = A$, where A is defined by (A.5). First we will show that $A \subseteq B$. Suppose that $s_j \in A$. Then $f(s_j) \leq \alpha$, and for each $k = 1, 2, ...$, it

follows from (A.4) that for some N_k, $n \geq N_k$ we have $f_n(s_j) \leq \alpha + \frac{1}{K}$ with $s_j \in A_{nk}$. Hence, for each k, we have $s_j \in B_{Nk}$ all $N > N_k$ so $s_j \in B_k$ and hence $s_j \in B$. Therefore, $A \subseteq B$. Next will show that $B \subseteq A$. Suppose that $s_j \in B$. Then $s \in B_k$ for each k. Hence for each k, $s_j \in B_{Nk}$ for some Nk. Hence, for each k we have $n > Nk$ such that $s_j \in A_{nk}$, which implies that $f(s_j) \leq \alpha$ so $s_j \in A$. Hence $B \subseteq A$. ■

Lemma A.0.1: Let S be a set and let $\mathcal{A}$ be an algebra of subsets of S. Then the monotone class $\mathcal{M}$ generated by $\mathcal{A}$ is the same as the σ-algebra $\mathcal{S}$ generated by $\mathcal{A}$.

Proof: See Stockey, Lucas, and Prescott (1989, Lemma 7.15, page:200). ■

Theorem A.0.6: Let $\mathbf{K}$ be a compact convex subset of a linear topological spaces $\mathcal{X}$. Let $\mathcal{F}$ be a commuting family of continuous linear mappincs which map $\mathbf{K}$ into itself. Then there exist a point p in $\mathbf{K}$ such that $Tp = p$ for each $T \in \mathcal{F}$.

Proof: See Dunford and Schwartz (1958, Theorem 9, page:456)

Proposition A.0.1: Suppose a random dynamical system is given. Assume tat the map $\tilde{S} \rightarrow ca(E \times J)$, where $ca(E \times J)$ is the set consists of the countably additive set functions, defined be $s_j \rightarrow Q_{EJ}(s_j, \cdot)$ is continuous when $ca(E \times J)$ is regarded as Banach Space in the total variation norm. if $f \in \mathbf{C}(\tilde{\mathbf{S}})$, it then follows that $Tf \in \mathbf{C}(\tilde{\mathbf{S}})$.

Proof: See Futia (1982, Proposition 5.6, page:400).■

Proposition A.0.2: If S is a compact metrizable space, then $P(S)$ (the set of all probability measures in S) is a compact metrizable space.

Proof: See Bertsekas and Shreve (1996, Proposition 7.22, page:130).■

Theorem A.0.7: If f is an integrable function on $\mathbf{S} \times \mathbf{J}$, then almost every section of f is integrable. If the functions h and g are defined by $h(s) = \int f(s_j) d\mu(j)$ and $g(j) = \int f(s_j) d\mu(s)$, then h and g are integrable and

$$\int f d\left(\mu(s) \times \mu(j)\right) = \int h d\mu(s) = \int g d\mu(j) \tag{A.6}$$

Proof: See Halmos (1950, Theorem C, page:148).■

Theorem A.0.8: Let (S, d) be a totally bounded metric space. A set $\mathbf{A} \subseteq \mathbf{C(S)}$ is relative compact under $d_U(f,g) = \sup_{x \in S} |f(x) - g(x)|$ if and only if it bounded and uniformly equicontinuous.

Proof: See Davidson (1994, Theorem 5.28, page:91). ■

Theorem A.0.9: Let $\mathbf{S}$ be a set, $\mathcal{A}$ an algebra of its subset and μ a measure on $\mathcal{A}$. Let $\mathcal{F}$ be the smallest σ-algebra containing $\mathcal{A}$. Then there exists a measure μ^* on $\mathcal{F}$ such that $\mu(A) = \mu^*(A)$, all $A \in \mathcal{A}$.

Theorem A.0.10: Let $\mathbf{S}$, $\mathcal{A}$, μ and $\mathcal{F}$ be specified as in Theorem A.0.9. If μ is σ-finite, then the extension is μ^* to $\mathcal{F}$ is unique.

Theorem A.0.11: For arbitrary $\mathcal{F}$-sets, A, B and $\{A_j,\ j \in \mathbb{N}\}$,

(i) $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$ (monotonicity);

(ii) $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$;

(iii) $\mu\left(\cup_j A_j\right) = \sum_j \mu\left(A_j\right)$ (countable subadditivity).

Proof: See Davidson (1994, Theorem 3.3, page:37). ■

Theorem A.0.12: A finite measure is continues.

Proof: See Davidson (1994, Theorem 3.4, page:38). ■

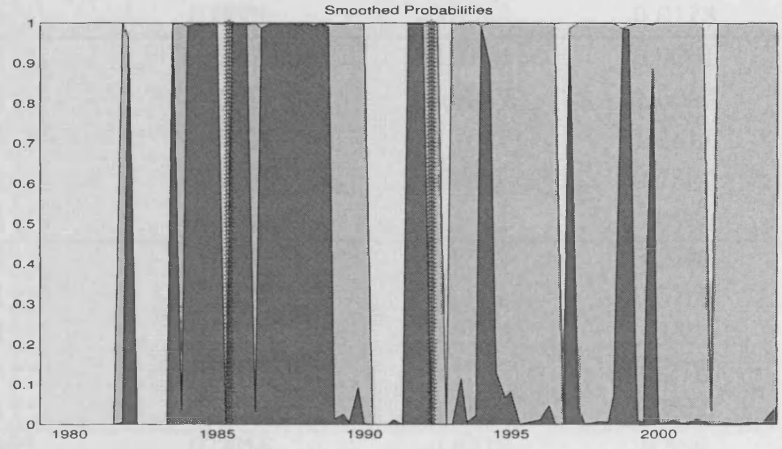
Theorem A.0.13: Let $\{f_n\}$ be a sequence of $\mathcal{F}/\mathcal{B}$-measurable functions. Then $\inf_n f_n$, $\sup_n f_n$, $\liminf_n f_n$ and $\limsup_n f_n$ are $\mathcal{F}/\mathcal{B}$-measurable.

Proof: See Davidson (1994, Theorem 3.26, page:52). ■

B. APPENDIX B

B.1 Results from the Markov Switching VAR(1) Exercise

Fig. B.1: **Pattern of Smoothed Probabilities**



Tab. B.1: **Diagnostic Tests/MSH(3)-VAR(1)**

Statistics	*Chi-Square*	*DF*	*P-values*
Conditional Heteroskedasticity	1101.1	900	0
Heteroskedasticity	266.5	150	0
Q- Statistics lags (12)	385.6	275	0
Adj-Q- Statistics lags (12)	410.95	275	0

Tab. B.2: **Distribution of the Autoregressive Parameters / MSH(3)-VAR(1)**

Parameters	*Actual Estimates*	*Lower Bound*	*Upper Bound*	*State*
$c_{\Delta y}$	-0.0003	-0.0321	0.0131	IN
$c_{\Delta u}$	-0.0142	-0.0347	0.0351	IN
$c_{\Delta \pi}$	0.0006	-0.0387	0.0173	IN
$c_{\Delta r}$	-0.0015	-0.0252	0.0116	IN
$c_{\Delta e}$	0.0065	-0.0343	0.0178	IN
$A_{\Delta y}^{\Delta y}$	-0.2377	-0.9734	0.9061	IN
$A_{\Delta u}^{\Delta y}$	2.0814	-1.5017	1.5007	OUT
$A_{\Delta \pi}^{\Delta y}$	0.5809	-1.1047	1.2411	IN
$A_{\Delta r}^{\Delta y}$	0.7502	-0.816	0.7364	OUT
$A_{\Delta e}^{\Delta y}$	-0.1129	-1.1481	0.9939	IN
$A_{\Delta y}^{\Delta u}$	0.0762	-0.2589	0.2895	IN
$A_{\Delta u}^{\Delta u}$	0.1478	-0.413	0.5709	IN
$A_{\Delta \pi}^{\Delta u}$	0.0699	-0.2935	0.3651	IN
$A_{\Delta r}^{\Delta u}$	0.1073	-0.1925	0.278	IN
$A_{\Delta e}^{\Delta u}$	0.2129	-0.4088	0.3353	IN
$A_{\Delta y}^{\Delta \pi}$	0.0656	-0.6312	0.589	IN
$A_{\Delta u}^{\Delta \pi}$	0.2451	-1.0365	1.0584	IN
$A_{\Delta \pi}^{\Delta \pi}$	0.0551	-0.6704	0.762	IN
$A_{\Delta r}^{\Delta \pi}$	0.099	-0.6169	0.3988	IN
$A_{\Delta e}^{\Delta \pi}$	0.2492	-0.863	0.7246	IN
$A_{\Delta y}^{\Delta r}$	0.6101	-0.2411	0.2166	OUT
$A_{\Delta u}^{\Delta r}$	-0.5829	-0.3497	0.4128	OUT
$A_{\Delta \pi}^{\Delta r}$	-0.0322	-0.3376	0.2661	IN
$A_{\Delta r}^{\Delta r}$	-0.2724	-0.2999	0.3293	IN
$A_{\Delta e}^{\Delta r}$	0.4387	-0.3468	0.3038	OUT
$A_{\Delta y}^{\Delta e}$	-0.061	-0.6441	0.4252	IN
$A_{\Delta u}^{\Delta e}$	-0.0043	-0.74	0.8845	IN
$A_{\Delta \pi}^{\Delta e}$	-0.1366	-0.5907	0.7798	IN
$A_{\Delta r}^{\Delta e}$	-0.1213	-0.4298	0.5003	IN
$A_{\Delta e}^{\Delta e}$	-0.1555	-0.6315	0.522	IN

Tab. B.3: **Distribution of the Covariance Parameters / MSH(3)-VAR(1)**

Covariances	*Actual estimates*	*Lower Bound*	*Upper Bound*	*State*
$\sigma^1_{\Delta y \Delta y}$	0.0015	0.0004	0.1648	IN
$\sigma^1_{\Delta u \Delta y}$	0.0034	-0.0528	0.0755	IN
$\sigma^1_{\Delta \pi \Delta y}$	0.0009	-0.0346	0.0643	IN
$\sigma^1_{\Delta r \Delta y}$	0.0013	-0.0297	0.0251	IN
$\sigma^1_{\Delta e \Delta y}$	0.0007	-0.0133	0.0517	IN
$\sigma^1_{\Delta u \Delta u}$	0.0123	0.003	0.7145	IN
$\sigma^1_{\Delta \pi \Delta u}$	0.0022	-0.0514	0.0594	IN
$\sigma^1_{\Delta r \Delta u}$	0.0028	-0.065	0.0864	IN
$\sigma^1_{\Delta e \Delta u}$	0.0029	-0.0892	0.0318	IN
$\sigma^1_{\Delta \pi \Delta \pi}$	0.001	0.0007	0.2563	IN
$\sigma^1_{\Delta r \Delta \pi}$	0.0008	-0.057	0.0262	IN
$\sigma^1_{\Delta e \Delta \pi}$	0.0004	-0.1038	0.0292	IN
$\sigma^1_{\Delta r \Delta r}$	0.0013	0.0006	0.293	IN
$\sigma^1_{\Delta e \Delta r}$	0.0004	-0.0265	0.0399	IN
$\sigma^1_{\Delta e \Delta e}$	0.0013	0.0005	0.2484	IN
$\sigma^2_{\Delta y \Delta y}$	0.0008	0.0003	0.1238	IN
$\sigma^2_{\Delta u \Delta y}$	0.001	-0.0299	0.0537	IN
$\sigma^2_{\Delta \pi \Delta y}$	0.0005	-0.0125	0.0297	IN
$\sigma^2_{\Delta r \Delta y}$	0.0006	-0.0301	0.0271	IN
$\sigma^2_{\Delta e \Delta y}$	0.0008	-0.0169	0.0382	IN
$\sigma^2_{\Delta u \Delta u}$	0.0022	0.0017	0.1966	IN
$\sigma^2_{\Delta \pi \Delta u}$	0.0003	-0.0405	0.0292	IN
$\sigma^2_{\Delta r \Delta u}$	0.0000	-0.0256	0.0955	IN
$\sigma^2_{\Delta e \Delta u}$	-0.0001	-0.0321	0.0224	IN
$\sigma^2_{\Delta \pi \Delta \pi}$	0.0004	0.0006	0.2037	OUT
$\sigma^2_{\Delta r \Delta \pi}$	0.0004	-0.0462	0.0196	IN
$\sigma^2_{\Delta e \Delta \pi}$	0.0003	-0.0367	0.0135	IN
$\sigma^2_{\Delta r \Delta r}$	0.0005	0.0005	0.187	IN
$\sigma^2_{\Delta e \Delta r}$	0.0007	-0.0308	0.0363	IN
$\sigma^2_{\Delta e \Delta e}$	0.0021	0.0006	0.1805	IN
$\sigma^3_{\Delta y \Delta y}$	0.0009	0.0003	0.1048	IN
$\sigma^3_{\Delta u \Delta y}$	0.0021	-0.0373	0.0258	IN
$\sigma^3_{\Delta \pi \Delta y}$	0.0007	-0.011	0.03	IN
$\sigma^3_{\Delta r \Delta y}$	0.0007	-0.0109	0.0157	IN
$\sigma^3_{\Delta e \Delta y}$	0.0005	-0.012	0.0196	IN
$\sigma^3_{\Delta u \Delta u}$	0.0063	0.002	0.1809	IN
$\sigma^3_{\Delta \pi \Delta u}$	0.002	-0.0406	0.0169	IN
$\sigma^3_{\Delta r \Delta u}$	0.0016	-0.0092	0.0477	IN
$\sigma^3_{\Delta e \Delta u}$	0.0015	-0.0221	0.0201	IN
$\sigma^3_{\Delta \pi \Delta \pi}$	0.0007	0.0006	0.1356	IN
$\sigma^3_{\Delta r \Delta \pi}$	0.0006	-0.0233	0.0058	IN
$\sigma^3_{\Delta e \Delta \pi}$	0.0005	-0.0382	0.0113	IN
$\sigma^3_{\Delta r \Delta r}$	0.0006	0.0005	0.0826	IN
$\sigma^3_{\Delta e \Delta r}$	0.0004	-0.0079	0.0195	IN
$\sigma^3_{\Delta e \Delta e}$	0.0005	0.0004	0.0852	IN

Tab. B.4: **The Number of the Autoregressive Parameters Rejected**

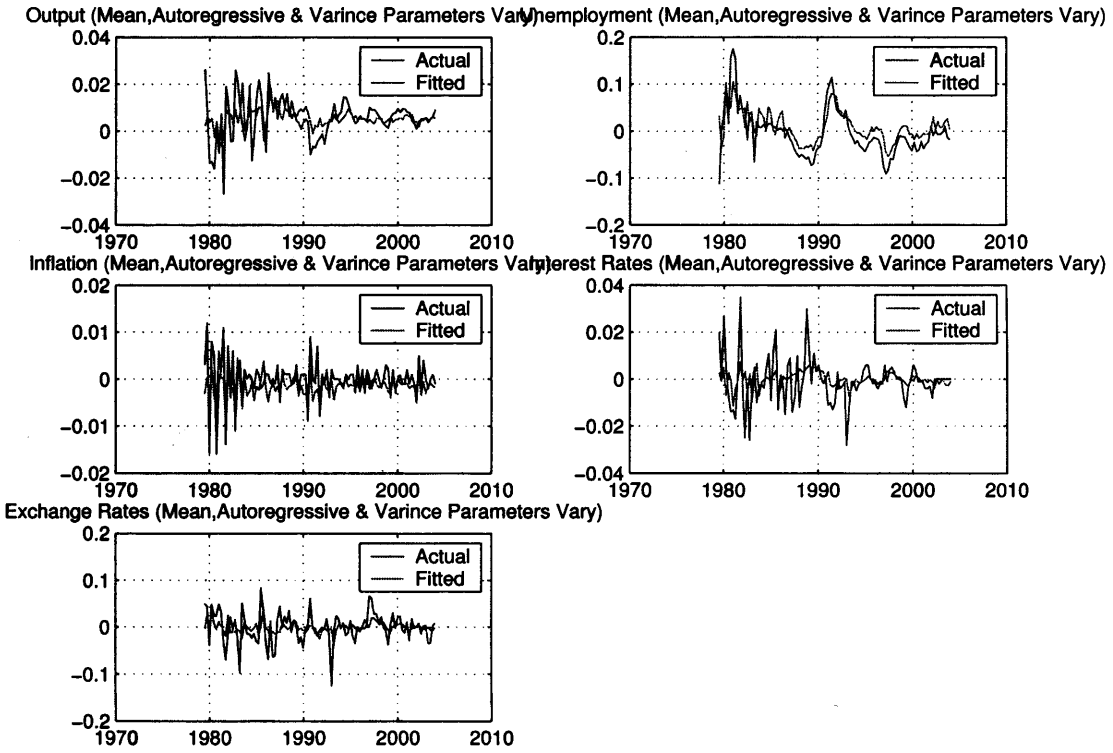
Confidence Range	*Number of Parameters Rejected*
1	0
0.995	1
0.99	1
0.985	3
0.98	3
0.975	4
0.97	4
0.965	4
0.96	4
0.955	5
0.95	5

Tab. B.5: **The Number of the Covariance Parameters Rejected**

Confidence Range	*Number of Parameters Rejected*
1	0
0.995	0
0.99	0
0.985	0
0.98	0
0.975	0
0.97	1
0.965	1
0.96	1
0.955	1

B.2 Results from the Intervention VAR(1) Exercise

Fig. B.2: **Intervention Mean Adjusted VAR(1) Model**



Tab. B.6: **Distribution of the Mean/ Intervention Mean Adjusted VAR(1)**

Mean	*Actual*	*Lower Bound*	*Upper Bound*	*Bootstrapped Prob Values*	*State*
$\mu^1_{\Delta y}$	0.0042	0.0004	0.0125	0.847	IN
$\mu^1_{\Delta u}$	0.0299	-0.1239	0.0311	0.032	IN
$\mu^1_{\Delta \pi}$	-0.0003	-0.0022	0.0015	0.418	IN
$\mu^1_{\Delta r}$	-0.0002	-0.0081	0.0102	0.433	IN
$\mu^1_{\Delta e}$	-0.0008	-0.0455	0.0043	0.059	IN
$\mu^2_{\Delta y}$	0.0088	0.0074	0.0225	0.958	IN
$\mu^2_{\Delta u}$	0.0290	-0.2046	-0.0207	0.000	OUT
$\mu^2_{\Delta \pi}$	-0.0003	-0.0055	-0.0028	0.000	OUT
$\mu^2_{\Delta r}$	-0.0017	-0.0011	0.0045	0.990	OUT
$\mu^2_{\Delta e}$	-0.0076	-0.0456	0.0074	0.253	IN
$\mu^3_{\Delta y}$	0.0117	0.0094	0.0188	0.879	IN
$\mu^3_{\Delta u}$	-0.0082	-0.1367	-0.0161	0.015	OUT
$\mu^3_{\Delta \pi}$	-0.0003	0.0005	0.0028	0.000	OUT
$\mu^3_{\Delta r}$	-0.0017	-0.0028	0.002	0.849	IN
$\mu^3_{\Delta e}$	-0.0016	-0.0281	0.0046	0.144	IN

Tab. B.7: **Distribution of the Parameters (Regime 1)**

Parameters	*Actual*	*Lower Bound*	*Upper Bound*	*Bootstrapped Prov. Values*	*State*
$A_{\Delta y}^{\Delta y}$	0.0128	-0.5577	0.4447	0.366	IN
$A_{\Delta y}^{\Delta u}$	-1.6212	-1.6854	1.0895	0.973	IN
$A_{\Delta y}^{\Delta \pi}$	0.1404	-0.3343	0.4907	0.356	IN
$A_{\Delta y}^{\Delta r}$	0.0385	-0.6769	0.7618	0.525	IN
$A_{\Delta y}^{\Delta e}$	0.007	-0.8879	1.0028	0.473	IN
$A_{\Delta u}^{\Delta y}$	-0.0455	-0.0839	0.226	0.923	IN
$A_{\Delta u}^{\Delta u}$	0.1606	-0.1373	0.7542	0.844	IN
$A_{\Delta u}^{\Delta \pi}$	0.0194	-0.1637	0.0577	0.116	IN
$A_{\Delta u}^{\Delta r}$	0.0165	-0.2276	0.2384	0.475	IN
$A_{\Delta u}^{\Delta e}$	-0.0033	-0.2867	0.2748	0.463	IN
$A_{\Delta \pi}^{\Delta y}$	0.1409	-0.5999	0.5317	0.267	IN
$A_{\Delta \pi}^{\Delta u}$	-2.9414	-1.4825	1.4928	0.000	OUT
$A_{\Delta \pi}^{\Delta \pi}$	-0.3617	-0.6956	0.1234	0.661	IN
$A_{\Delta \pi}^{\Delta r}$	0.6422	-0.7575	0.8289	0.045	IN
$A_{\Delta \pi}^{\Delta e}$	-1.1134	-1.1167	0.7214	0.975	OUT
$A_{\Delta r}^{\Delta y}$	-0.0536	-0.702	0.2467	0.287	IN
$A_{\Delta r}^{\Delta u}$	-0.0684	-0.8594	1.5914	0.802	IN
$A_{\Delta r}^{\Delta \pi}$	0.0666	-0.5257	0.1092	0.035	IN
$A_{\Delta r}^{\Delta r}$	0.0882	-0.8109	0.6828	0.451	IN
$A_{\Delta r}^{\Delta e}$	0.8146	-0.8272	0.5971	0.007	OUT
$A_{\Delta e}^{\Delta y}$	-0.1094	-0.3299	0.187	0.649	IN
$A_{\Delta e}^{\Delta u}$	0.0666	-0.6253	0.8872	0.676	IN
$A_{\Delta e}^{\Delta \pi}$	0.0257	-0.0461	0.3785	0.896	IN
$A_{\Delta e}^{\Delta r}$	-0.0581	-0.1803	0.4167	0.879	IN
$A_{\Delta e}^{\Delta e}$	0.2576	-0.4259	0.3434	0.050	IN
$\sigma_{\Delta y \Delta y}$	0.0002	0.0003	0.0012	0.000	OUT
$\sigma_{\Delta u \Delta y}$	-0.0001	-0.0029	-0.0005	0.000	OUT
$\sigma_{\Delta \pi \Delta y}$	0.0000	-0.0005	0.0000	0.035	IN
$\sigma_{\Delta r \Delta y}$	0.0000	-0.0004	0.0004	0.540	IN
$\sigma_{\Delta e \Delta y}$	0.0000	-0.0003	0.0010	0.809	IN
$\sigma_{\Delta u \Delta u}$	0.0025	0.0023	0.0112	0.973	IN
$\sigma_{\Delta \pi \Delta u}$	-0.0002	-0.0001	0.0013	0.980	OUT
$\sigma_{\Delta r \Delta u}$	-0.0002	-0.0011	0.0012	0.772	IN
$\sigma_{\Delta e \Delta u}$	0.0005	-0.0025	0.0014	0.153	IN
$\sigma_{\Delta \pi \Delta \pi}$	0.0000	0.0002	0.0009	0.000	OUT
$\sigma_{\Delta r \Delta \pi}$	0.0000	0.0000	0.0009	0.963	IN
$\sigma_{\Delta e \Delta \pi}$	0.0000	-0.0011	0.0001	0.059	IN
$\sigma_{\Delta r \Delta r}$	0.0002	0.0004	0.0032	0.000	OUT
$\sigma_{\Delta e \Delta r}$	-0.0001	-0.0012	0.0004	0.428	IN
$\sigma_{\Delta e \Delta e}$	0.0016	0.0011	0.0046	0.844	IN

Tab. B.8: **Distribution of the Parameters (Regime 2)**

Parameters	*Actual*	*Lower Bound*	*Upper Bound*	*Bootstrapped Prov. Values*	*State*
$A_{\Delta y}^{\Delta y}$	0.2194	-0.4603	0.2233	0.030	IN
$A_{\Delta y}^{\Delta u}$	-1.1006	-1.0583	1.0384	0.985	OUT
$A_{\Delta y}^{\Delta \pi}$	0.0888	-0.4245	0.3437	0.312	IN
$A_{\Delta y}^{\Delta r}$	-0.0236	-1.2564	1.1196	0.515	IN
$A_{\Delta y}^{\Delta e}$	0.1560	-0.5401	0.6296	0.337	IN
$A_{\Delta u}^{\Delta y}$	-0.0275	-0.0839	0.1553	0.728	IN
$A_{\Delta u}^{\Delta u}$	0.5950	0.1412	0.8215	0.500	IN
$A_{\Delta u}^{\Delta \pi}$	0.0086	-0.1607	0.0376	0.079	IN
$A_{\Delta u}^{\Delta r}$	-0.0401	-0.3664	0.3115	0.629	IN
$A_{\Delta u}^{\Delta e}$	-0.0087	-0.1498	0.2690	0.748	IN
$A_{\Delta \pi}^{\Delta y}$	0.2565	-0.3542	0.1614	0.003	OUT
$A_{\Delta \pi}^{\Delta u}$	-1.3935	-0.4574	0.9537	0.000	OUT
$A_{\Delta \pi}^{\Delta \pi}$	-0.4377	-0.7264	-0.1493	0.525	IN
$A_{\Delta \pi}^{\Delta r}$	0.2642	-1.1670	1.1637	0.300	IN
$A_{\Delta \pi}^{\Delta e}$	-0.9034	-0.3492	0.3443	0.000	OUT
$A_{\Delta r}^{\Delta y}$	-0.0715	-0.1031	0.0769	0.936	IN
$A_{\Delta r}^{\Delta u}$	0.0218	-0.2266	0.2782	0.460	IN
$A_{\Delta r}^{\Delta \pi}$	0.0619	-0.1540	0.0900	0.072	IN
$A_{\Delta r}^{\Delta r}$	0.0817	-0.4519	0.4811	0.300	IN
$A_{\Delta r}^{\Delta e}$	0.6310	-0.1487	0.1007	0.000	OUT
$A_{\Delta e}^{\Delta y}$	-0.0842	-0.3084	0.1680	0.443	IN
$A_{\Delta e}^{\Delta u}$	-0.0815	-0.6240	0.9357	0.884	IN
$A_{\Delta e}^{\Delta \pi}$	0.0272	-0.0163	0.3365	0.928	IN
$A_{\Delta e}^{\Delta r}$	0.0095	-0.2134	0.4592	0.723	IN
$A_{\Delta e}^{\Delta e}$	0.2444	-0.2999	0.4227	0.173	IN
$\sigma_{\Delta y \Delta y}$	0.0002	0.0007	0.0020	0.000	OUT
$\sigma_{\Delta u \Delta y}$	-0.0001	-0.0047	-0.0014	0.000	OUT
$\sigma_{\Delta \pi \Delta y}$	0.0000	-0.0007	0.0002	0.168	IN
$\sigma_{\Delta r \Delta y}$	0.0000	-0.0018	0.0012	0.478	IN
$\sigma_{\Delta e \Delta y}$	-0.0001	-0.0005	0.0010	0.715	IN
$\sigma_{\Delta u \Delta u}$	0.0033	0.0048	0.0161	0.000	OUT
$\sigma_{\Delta \pi \Delta u}$	-0.0002	-0.0005	0.0018	0.903	IN
$\sigma_{\Delta r \Delta u}$	-0.0003	-0.0037	0.0047	0.658	IN
$\sigma_{\Delta e \Delta u}$	0.0005	-0.0025	0.0020	0.290	IN
$\sigma_{\Delta \pi \Delta \pi}$	0.0000	0.0009	0.0023	0.000	OUT
$\sigma_{\Delta r \Delta \pi}$	0.0000	-0.0036	0.0025	0.342	IN
$\sigma_{\Delta e \Delta \pi}$	0.0000	-0.0015	-0.0002	0.005	OUT
$\sigma_{\Delta r \Delta r}$	0.0003	0.0072	0.0544	0.000	OUT
$\sigma_{\Delta e \Delta r}$	-0.0002	-0.0026	0.0037	0.748	IN
$\sigma_{\Delta e \Delta e}$	0.0029	0.0017	0.0054	0.458	IN

Tab. B.9: **Distribution of the Parameters (Regime 3)**

Parameters	*Actual*	*Lower Bound*	*Upper Bound*	*Bootstrapped Prov. Values*	*State*
$A_{\Delta y}^{\Delta y}$	0.3733	-0.3820	0.1778	0.007	OUT
$A_{\Delta y}^{\Delta u}$	-0.7354	-0.6356	0.7531	0.985	OUT
$A_{\Delta y}^{\Delta \pi}$	0.0561	-0.5523	0.4018	0.295	IN
$A_{\Delta y}^{\Delta r}$	-0.0125	-0.5889	0.5776	0.559	IN
$A_{\Delta y}^{\Delta e}$	-0.0093	-0.4048	0.6115	0.708	IN
$A_{\Delta u}^{\Delta y}$	-0.0088	-0.0794	0.1243	0.564	IN
$A_{\Delta u}^{\Delta u}$	0.6629	0.3482	0.8514	0.488	IN
$A_{\Delta u}^{\Delta \pi}$	0.0058	-0.1529	0.0786	0.230	IN
$A_{\Delta u}^{\Delta r}$	-0.0399	-0.1886	0.1735	0.703	IN
$A_{\Delta u}^{\Delta e}$	-0.0689	-0.1449	0.2478	0.901	IN
$A_{\Delta \pi}^{\Delta y}$	0.2715	-0.1531	0.0392	0.000	OUT
$A_{\Delta \pi}^{\Delta u}$	-1.0064	-0.0661	0.3944	0.000	OUT
$A_{\Delta \pi}^{\Delta \pi}$	-0.4491	-0.7150	-0.2804	0.300	IN
$A_{\Delta \pi}^{\Delta r}$	0.2122	-0.1375	0.1683	0.012	OUT
$A_{\Delta \pi}^{\Delta e}$	-1.1928	-0.0594	0.4041	0.000	OUT
$A_{\Delta r}^{\Delta y}$	-0.0099	-0.1433	0.0528	0.470	IN
$A_{\Delta r}^{\Delta u}$	0.0888	-0.1325	0.2723	0.131	IN
$A_{\Delta r}^{\Delta \pi}$	0.0370	-0.2078	0.0899	0.094	IN
$A_{\Delta r}^{\Delta r}$	0.1174	-0.2917	0.2477	0.097	IN
$A_{\Delta r}^{\Delta e}$	0.3020	-0.1228	0.2146	0.000	OUT
$A_{\Delta e}^{\Delta y}$	-0.0501	-0.2451	0.0995	0.267	IN
$A_{\Delta e}^{\Delta u}$	-0.1224	-0.3002	0.5749	0.956	IN
$A_{\Delta e}^{\Delta \pi}$	0.0128	0.2657	0.8698	0.000	OUT
$A_{\Delta e}^{\Delta r}$	0.0212	-0.1720	0.2328	0.542	IN
$A_{\Delta e}^{\Delta e}$	0.2147	-0.4580	0.1212	0.010	OUT
$\sigma_{\Delta y \Delta y}$	0.0002	0.0009	0.0028	0.000	OUT
$\sigma_{\Delta u \Delta y}$	-0.0001	-0.0060	-0.0020	0.000	OUT
$\sigma_{\Delta \pi \Delta y}$	0.0000	-0.0011	0.0006	0.364	IN
$\sigma_{\Delta r \Delta y}$	0.0000	-0.0010	0.0009	0.550	IN
$\sigma_{\Delta e \Delta y}$	0.0000	-0.0005	0.0013	0.696	IN
$\sigma_{\Delta u \Delta u}$	0.0023	0.0062	0.0154	0.000	OUT
$\sigma_{\Delta \pi \Delta u}$	-0.0001	-0.0018	0.0014	0.408	IN
$\sigma_{\Delta r \Delta u}$	-0.0002	-0.0026	0.0030	0.644	IN
$\sigma_{\Delta e \Delta u}$	0.0002	-0.0018	0.0021	0.579	IN
$\sigma_{\Delta \pi \Delta \pi}$	0.0000	0.0033	0.0076	0.000	OUT
$\sigma_{\Delta r \Delta \pi}$	0.0000	-0.0022	0.0018	0.364	IN
$\sigma_{\Delta e \Delta \pi}$	0.0000	-0.0062	-0.0022	0.000	OUT
$\sigma_{\Delta r \Delta r}$	0.0002	0.0047	0.0333	0.000	OUT
$\sigma_{\Delta e \Delta r}$	-0.0001	-0.0016	0.0026	0.767	IN
$\sigma_{\Delta e \Delta e}$	0.0021	0.0038	0.0091	0.000	OUT

Tab. B.10: **The Number of Parameters Rejected per Regime**

Targeting Regimes	*Means*	*Autoregressive*	*Var-Covariances*	*All*
Monetary	0/5	3/25	5/15	8/45
Exchange	3/5	5/25	6/15	14/45
Inflation	2/5	9/25	7/15	18/45
Total	**5/15**	**17/75**	**18/45**	**40/135**

B.3 Impulse Response Analysis

Fig. B.3: **Monetary Targeting Regime / Supply Shock**

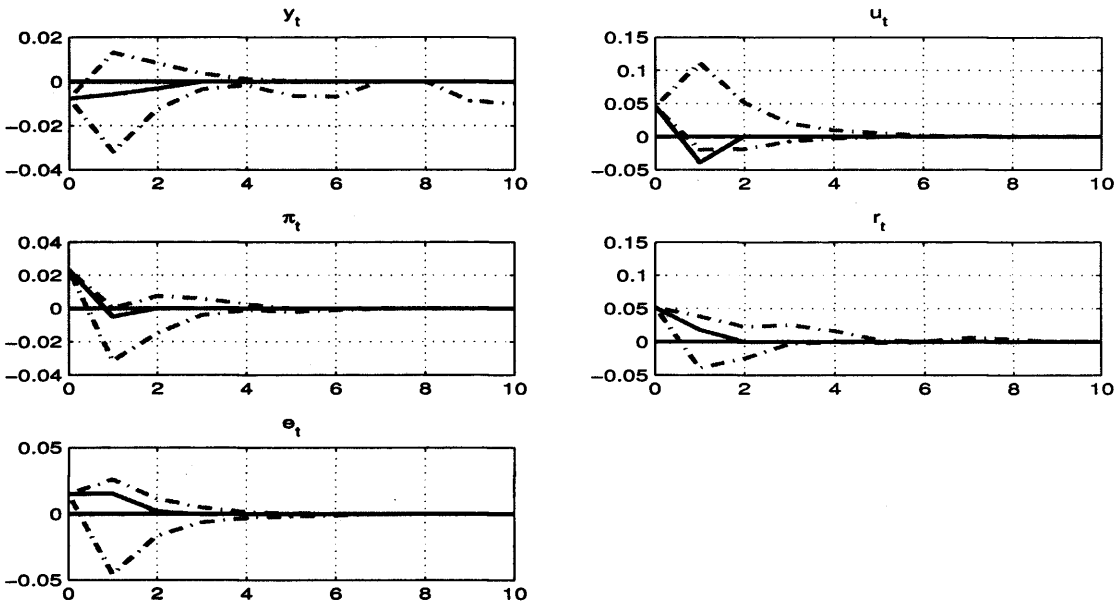


Fig. B.4: **Monetary Targeting Regime / Nominal Shock**

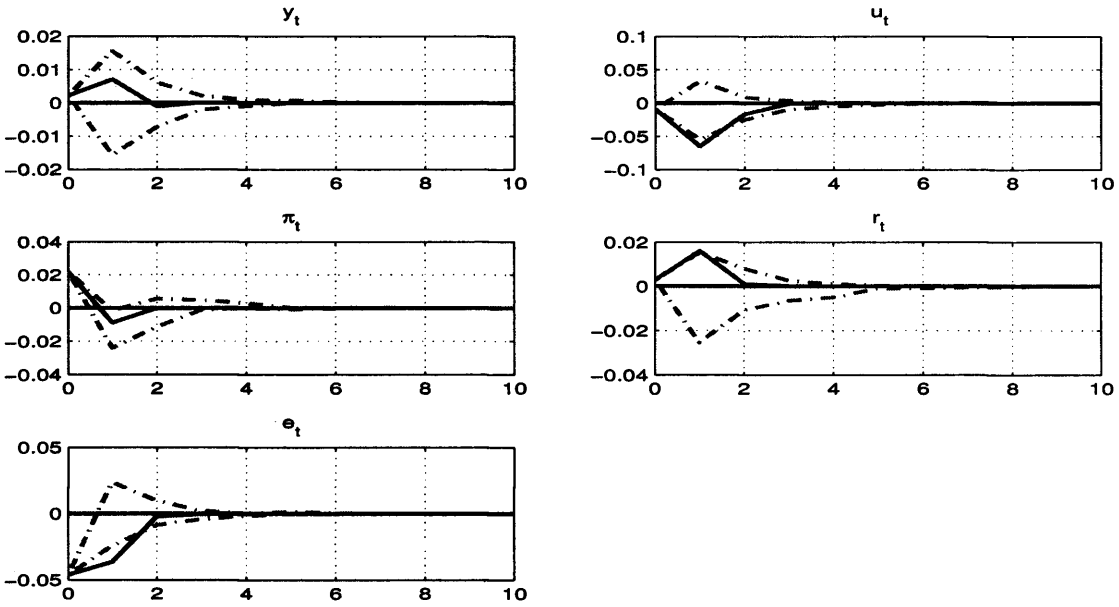


Fig. B.5: **Exchange Rate Targeting Regime / Supply Shock**

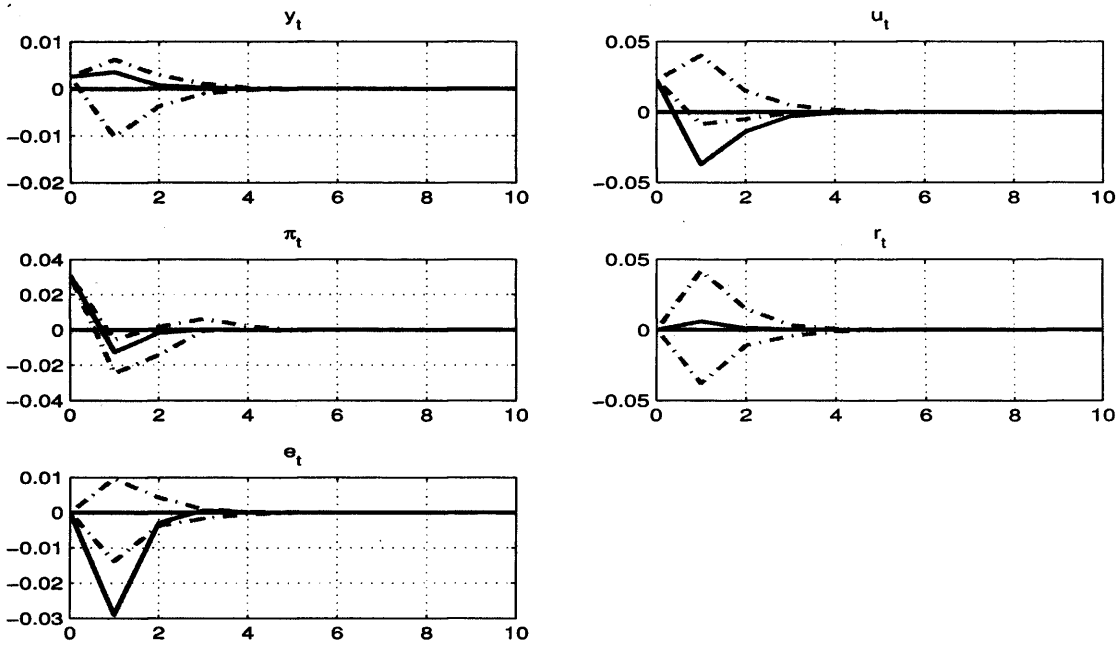


Fig. B.6: **Exchange Rate Targeting Regime / Nominal Shock**

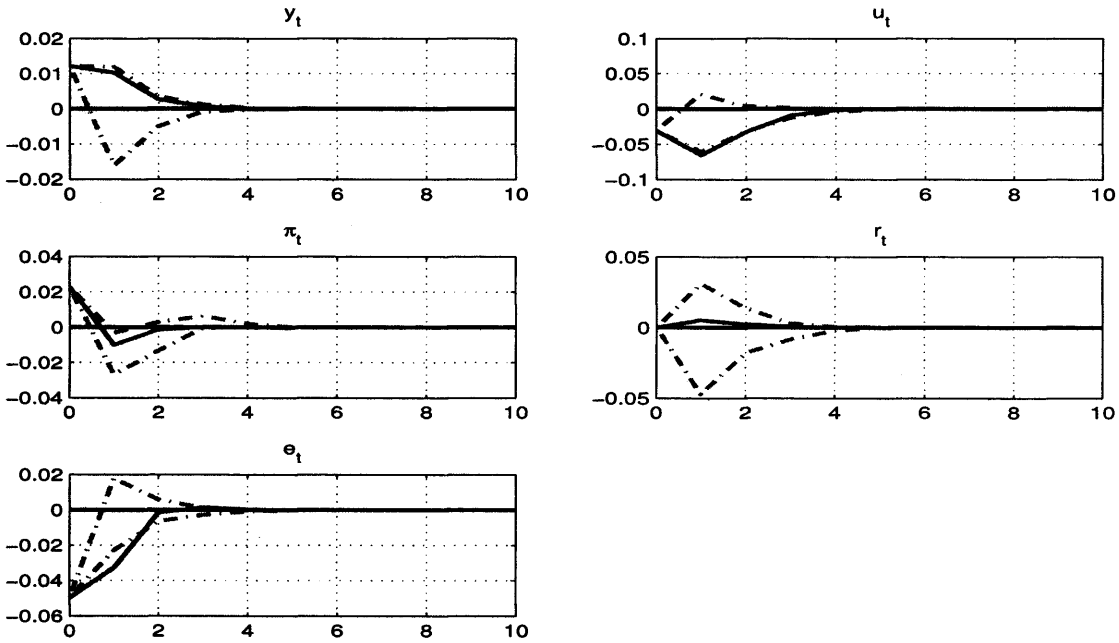


Fig. B.7: **Inflation Targeting Regime / Supply Shock**

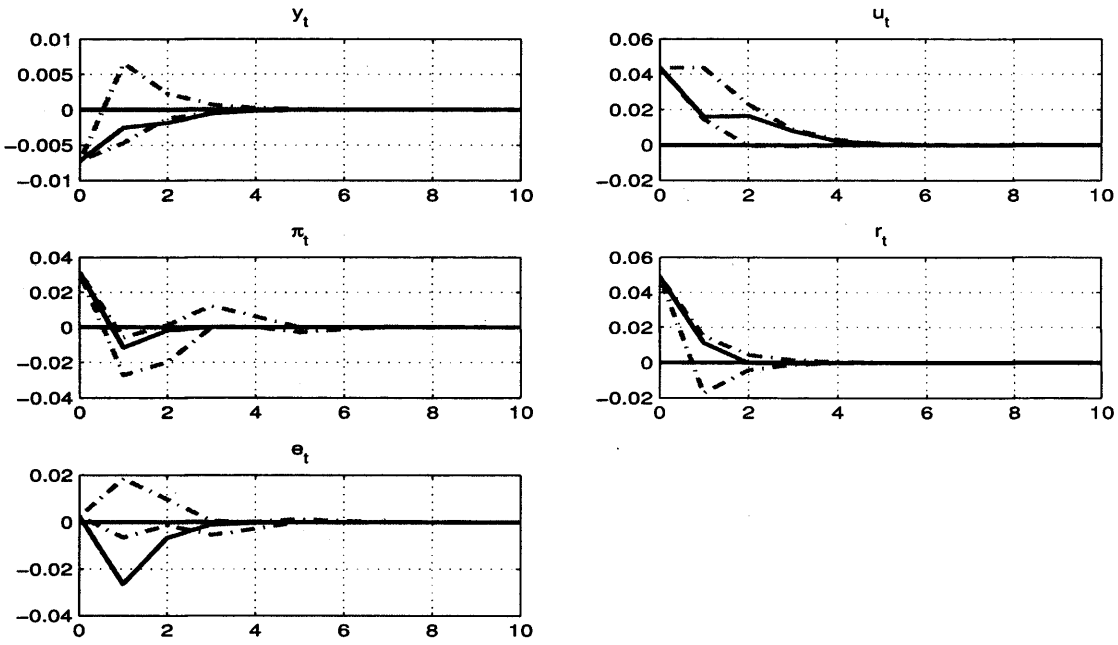
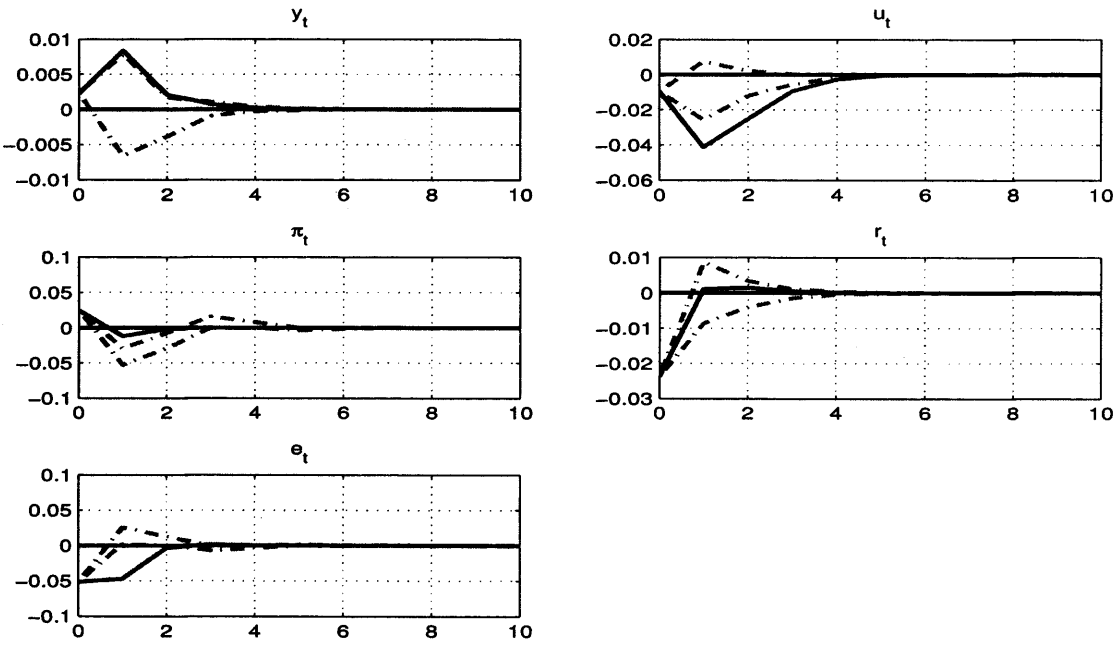


Fig. B.8: **Inflation Targeting Regime / Nominal Shock**



BIBLIOGRAPHY

ALTUG, S. (1989): "Time to build and aggregate fluctuations:some new evidences," *International Economic Review*, 30, 889–920.

BERTSEKAS, D., AND S. SHREVE (1996): *Stochastic Optimal Control: The Discrete Time Case.* Athena Scientific, Massachusetts.

BIERENS, H. (2005): "Econometric Analysis of Linearized Singular Dynamic Stochastic General Equilibrium Models," *Journal of Econometrics*, forthcoming.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures.* John Willey and Sons, New York.

BREIMAN, L. (1960): "The Strong Law of Large Numbers for a Class of Markov Chains," *The Annals of Mathematical Statistics*, 31, 801–803.

BURNSIDE, C., M. EICHENBAUM, AND S. REBELO (1993): "Labor Hoarding and the Business Cycle," *Journal of Political Economy*, 101, 245–273.

CANOVA, F. (1994): "Statistical Inference in Calibrated Models," *Journal of Applied Econometrics*, 9, S123–S144.

——— (1995): "Sensitivity Analysis and Model Evaluation in Simulated Dynamic General Equilibrium Economies," *International Economic Review*, 36, 477–501.

——— (2005): *Methods for Applied Macroeconomic Research.* Princeton University Press, Princeton.

CANOVA, F., AND G. DE NICOLO (1995): "The Equity Premium and the Risk Free Rate:A Cross Country, Cross Maturity Examination," *CEPR*, No.1119.

CHRISTIANO, L., AND W. DEN HAAN (1996): "Small Sample Properties of GMM for Business Cycle Analysis," *Journal of Business and Economic Statistics*, 14, 309–327.

CHRISTIANO, L., AND M. EICHENBAUM (1992): "Current Real Business Cycle Theories and Aggregate Labor Market Fluctuation," *American Economic Review*, 82, 430–450.

CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (1999): "Monetary Policy Shocks: What Have We Learned and to What End?," in *Handbook of Monetary Economics*, ed. by Taylor, and Woodford.

COGLEY, T., S. MOROZOV, AND T. SARGENT (2003): "Bayesian Fan Charts for U.K. Inflation: Forecasting and Sources of Uncertainty in an Evolving Monetary System," *Working Paper*, Center for Financial Studies, No. 2003/44.

COGLEY, T., AND J. NASON (1994): "Testing the Implications of Long-run Neutrality for Monetary Business Cycle Models," *Journal of Applied Econometrics*, 9, S37–S70.

CORRADI, V., AND N. SWANSON (2005): "Evaluation of Dynamic Stochastic General Equilibrium Models Based on Distributional Comparison of Simulated and Historical Data," *Journal of Econometrics*, forthcoming.

CRAUEL, H. (2002): *Random Probability Measures on Polish Spaces*. Taylor and Francis, London.

DAVIDSON, J. (1994): *Stochastic Limit Theory*. Oxford University Press, Oxford.

DEBREU, G. (1967): *Integration of Correspondences*. University of California Press, Berkley.

DEJONG, D., B. INGRAM, AND C. WHITEMAN (1996): "Beyond Calibration," *Journal of Business and Economic Statistics*, 14, 1–10.

——— (2000): "A Bayesian Approach to Dynamic Macroeconomics," *Journal of Econometrics*, 98, 203–223.

DEL NEGRO, M., AND F. SCHORFHEIDE (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45, 643–673.

——— (2005): "Monetary Policy Analysis with Potentially Misspecified Models," *Mimeo*, University of Pennsylvania.

DEL NEGRO, M., F. SCHORFHEIDE, F. SMETS, AND R. WOUTERS (2004): "On the Fit and Forecasting Performance of New Keynesian Models," *University of Pennsylvania*, Manuscript.

DIEBOLD, F., L. OHANIAN, AND J. BERKOWITZ (1998): "Dynamics General Equilibrium Economies: A Framework for Comparing Models and Data," *Review of Economic Studies*, 65, 433–451.

DOOB, L. (1953): *Stochastic Processes.* John Wiley and Sons, New York.

DOORNIK, J. (1996): "Testing Vector Error Autocorrelation and Heteroscedasticity," *Manuscript*, Nuffield College, Oxford.

DUBINS, L., AND D. FREEDMAN (1966): "Invariant Probability Measures for Certain Markov Processes," *The Annals of Mathematical Statistics*, 37, 837–848.

DUNFORD, N., AND J. SCHWARTZ (1958): *Linear Operators.* Interscience Publishers Inc., New York.

ENGEL, R., AND K. SHEPPARD (2001): "Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH," *Mimeo*, UCSD.

FERNANDEZ-VILLAVERDE, J., AND J. RUBIO-RAMIREZ (2002): "Estimating Nonlinear Dynamic Equilibrium Economies: A Likelihood Approach," *Mimeo*, University of Pennsylvania.

——— (2004): "Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach," *Journal of Econometrics*, 123, 153–185.

F'EVE, P., AND F. LANGOT (1994): "The RBC Models through Statistical Inference: An Application with French Data," *Journal of Applied Econometrics*, 9, S11–S37.

FUTIA, C. A. (1982): "Invariant Distributions and the Limiting Behavior of Markovian Economic Models," *Econometrica*, 50, 377–408.

GEWEKE, J. (1989): "Bayesian Inference in Econometric Models using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

——— (1999): "Computational Experiment and Reality," *University of Iowa*, manuscript.

GREGORY, A., AND G. SMITH (1991): "Calibration as Testing: Inference in Simulated Macro Models," *Journal of Business and Economic Statistics*, 9, 293–303.

——— (1993): "Calibration in Macroeconomics," in *Handbook of Statistics*, ed. by G. Maddala. vol. 11, Elsevier, St.Louis, Mo., 703-719.

HAAVELMO, G. (1944): "The Probability Approach in Econometrics," *Econometrica*, 12, Supplement.

HALMOS, P. (1950): *Measure Theory.* Van Nostrand, Princeton.

HAMILTON, J. (1994): *Time Series Analysis.* Princeton University Press, New York.

HANSEN, L. (1982): "Large Sample Properties of GMM Estimators," *Econometrica*, 50, 1029–1054.

HOPENHAYN, H., AND E. PRESCOTT (1992): "Stochastic Monotonicityand Stationary Distibutions for Dynamic Economies," *Econometrica*, 60, 1387–1406.

IRELAND, P. (2004): "A Method for Taking Models to the Data," *Journal of Economic Dynamics and Control*, 28, 1205–1226.

KIM, J., S. KIM, E. SCHAUMBURG, AND C. SIMS (2005): "Calculating and Using Second Order Accuarate Solutions of Discrete Time Dynamic Equilibrium Models," *Manuscript*, University of Princeton.

KINGMAN, J. (1968): "The Ergodic Theory of Subadditive Processes," *Journal of Royal Statistical Society: Series B*, 30, 499–510.

KROLZIG, H. (1997): "Markov-Switching Vector Autoregressions. Modelling, Statistical Inference and Application to Business Cycle Analysis," *Lecture Notes in Economics and Mathematical Systems*, Volume 454.

LUTKEPOHL, H. (1993): *Introduction to Multiple Time Series Analysis.* Springer, Berlin.

LUTKEPOHL, H., AND M. KRATZIG (1994): *Applied Time Series Econometrics.* Cambridge University Press, Cambridge.

MAFFEZZOLI, M. (2000): "Human Capital and International Business Cycles," *Review of Economic Dynamics*, 3, 137–165.

MCGRATTAN, E., R. ROGERSON, AND R. WRIGHT (1997): "An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy," *International Economic Review*, 38, 267–290.

MINFORD, P., D. MEENAGH, AND J. WANG (2006): "Testing a Simple Structural Model of Endogenous Growth," *University of Cardiff*, Working Paper.

MINFORD, P., D. MEENAGH, AND B. WEBB (2004): "Britain and EMU: Assessing the Costs in Macroeconomic Variability," *The World Economy*, 27, 301–358.

SANTOS, M., AND A. PERALTA-ALVA (2005): "Accuracy of Simulations for Stochastic Dynamic Models," *Econometrica*, 73, 1939–1976.

SARGENT, T. (1979): "A Note on Maximum Likelihood estimation of Rational Expectations Model of the Term Structure," *Journal of Monetary Economics*, 5, 133–143.

SARGENT, T., AND L. HANSEN (2004): *Recursive Models of Dynamic Linear Economies.* http://homepages.nyu.edu/ ts43/.

SCHORFHEIDE, F. (2000): "Loss function based evaluation of DSGE models," *Journal of Applied Econometrics*, 15, 645–670.

SERFLING, J. (1980): *Approximation Theorems of Mathematical Satistics.* John Wiley, New York.

SMETS, F., AND R. WOUTERS (2003): "An Estimated Stochastic DSGE model of the Euro Area," *Journal of the European Economic Association*, 1, 1123–1175.

SMITH, A. (1993): "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.

SODERLIND, P. (1994): "Cyclical Properties of a Real Business Cycle Model," *Journal of Applied Econometrics*, 9, S113–S122.

STINCHCOMBE, M., AND H. WHITE (1992): "Some Measurability Results for Extremea of Random Functions Over Random Sets," *Review of Economic Studies*, 59, 495–514.

STOCKEY, N., R. LUCAS, AND E. PRESCOTT (1989): *Recursive Methods of Economic Dynamics.* Harvard University Press, Cambridge.

SUNGBAE, A. (2005): "Bayesian Estimation of DSGE models: Lessons from Second-Order Approximations," *Mimeo*, University of Pennsylvania.

WATSON, M. (1993): "Measures of Fit for Calibrated Models," *Journal of Political Economy*, 101, 1011–1041.

WHITE, H. (1994): *Estimation, Inference and Specification Analysis.* Cambridge University Press, Cambridge.