ASSESSING THE FIT OF SIMULATED MULTIVARIATE DYNAMIC MODELS

Eva Ortega

Banco de España - Servicio de Estudios
Documento de Trabajo nº 9821

ASSESSING THE FIT OF SIMULATED MULTIVARIATE DYNAMIC MODELS

Eva Ortega *

* This paper is based on Chapter 3 of my dissertation at the European University Institute in Florence, Italy. Part of this work was conducted when I was visiting the Economics Department at UPenn, Philadelphia. I would like to thank Fabio Canova, Humberto López, Chiara Monfardini and Ángel Ubide for very helpful discussions, and Valentina Corradi, Frank Diebold and graduate students at UPenn for their useful comments and the warm hospitality.

Banco de España - Servicio de Estudios
Documento de Trabajo nº 9821

In publishing this series the Banco de España seeks to disseminate studies of interest that will help acquaint readers better with the Spanish economy.

The analyses, opinions and findings of these papers represent the views of their authors; they are not necessarily those of the Banco de España.

The Banco de España is disseminating some of its principal reports via INTERNET and INFOVÍA.

The respective WWW server addresses are: http://www.bde.es and http://www.bde.inf.

ISSN: 0213-2710
ISBN: 84-7793-635-8
Depósito legal: M. 41466-1998
Imprenta del Banco de España

ABSTRACT

This paper develops a formal framework based on multivariate spectral techniques for assessing the performance of multivariate dynamic models whose solution is approximated through simulation. The approach is especially suitable for models that focus on a particular frequency range, such as business cycle models.

An asymptotic test is presented for assessing how well a simulated model reproduces the dynamic properties of a vector of actual series. A further test is then derived to compare the relative performance of alternative model specifications with respect to the multivariate vector of interest. Monte Carlo evidence is provided to show the finite sample behavior of the tests. Both tests are found to have high power even in small sample sizes. As an application, we evaluate to which extent different versions of a two-country two-good International Real Business Cycle model can reproduce the interdependencies observed between the US and European real GDP at business cycle frequencies.

1 Introduction

The increasing complexity of the issues economists want to address has induced the wide use of multivariate dynamic models such as the stochastic dynamic general equilibrium models. However, their own complexity often implies the inability to obtain analytical solutions to these models and hence simulation techniques are used to approximate the equilibrium solution; i.e., simulate solution paths for the endogenous variables in terms of the exogenous variables and the parameters (see Marcet (1994) for a review of the application of simulation methods to economics). The empirical analysis of such models has to deal with the obvious possibility of error not only when selecting the specific functionals linking the endogenous to the exogenous variables of the model, and when parameterizing the model's structure and the distribution of the exogenous processes, but also when finding an approximate solution. Precisely because of the various possible sources of error, it is an important issue to assess adequately the ability of a simulated model to reproduce certain aspects observed in actual data (i.e. assess the fit of the model) as well as to compare the performance of alternative models.

In the calibration approach pioneered by Kydland and Prescott (1982), the assessment of the fit of a model often reduces to an informal subjective comparison of selected model and actual statistics [1]. Typically, a model is specified given a concrete question a researcher wants to study. The model is then solved (usually through an approximation method), the parameters are given fixed values and exogenous processes fixed distributions and data series for the variables of interest are generated by simulation of the model. The assessment of the performance of a model is typically reduced to a relatively subjective comparison of two reduced sets of summary statistics obtained from the simulated and the actual data. The model economy is considered a "good" approximation of the actual world if it can broadly reproduce the observed features of the series that it purports to model. The adequacy of a particular parameterization is typically checked through sensitivity analysis, which essentially consists of computing and comparing the same statistics for different parameterizations. Comparison of competing models very seldom takes place and when it does it is typically reduced to a similarly informal subjective comparison of selected statistics. These procedures, based on comparing the similarities between simulated and actual data are essentially ad-hoc, lack statistical foundations and ignore information that

[1] Economic models tend to be small and strongly theory based, and hence, as Pagan (1994) stresses, unlikely to obey the 'axiom of correct specification'. That is the idea underlying all calibration exercises, as made specific by Kydland and Prescott (1991): "..., no attempt is made (in the Calibration Approach) to determine the true model. All model economies are abstractions and are by definition false". In a calibration exercise, the researcher cares for the economic structure of the model (the desired properties of the model equilibrium and its tractability dictate the particular functional forms chosen and constrain parameters' values -see, e.g. King, Plosser and Rebelo (1990)-) rather than for capturing the joint distribution of actual data.

could be used for model evaluation purposes.

Recent research in applied macroeconomics and time series econometrics has suggested alternatives to such informal approach to assess the fit of a model. Canova and Ortega (1996) roughly classify those alternatives into four categories (see also Kim and Pagan (1995) for a review of recent methods for evaluating calibrated models). The first class uses the uncertainty in the simulated data to provide a measure of distance between the model and the data. In adopting this approach some take the parameters of the model as given and the exogenous processes as stochastic (such as Gregory and Smith (1991, 1993), Sõderlind (1994) or Cogley and Nason (1994)). Others formalize the uncertainty faced by the researcher when giving values to the parameters of the model and treat both parameters and exogenous processes as stochastic (see Canova (1994, 1995)), so that this possible source of misspecification is taken care of when assessing the fit of the model. They all rely on Monte Carlo methods to conduct inference under the assumption that the model is the correct DGP for the actual data.

The second approach uses the sampling variability of the actual data to propose a measure of distance between the model and the data, which sometimes comes from the estimation of model parameters using the GMM based approach of Christiano and Eichenbaum (1992) (see also Cecchetti, Lam and Mark (1993) or Fève and Langot (1994)). The multivariate frequency domain approach of Diebold, Ohanian and Berkowitz (1995) also belongs to this class. They assume that model statistics can be estimated without error simply by simulating very long time series from the model and hence use only the sampling variability of actual data statistics (evaluated with bootstrap algorithms) to propose goodness-of-fit criteria and to derive associated optimal parameter estimators. Although they deal with the issue of parameter uncertainty, their approach does not take into account the possible error induced when approximating the solution of the model.

The third approach is given by Watson (1993), which does not consider sampling variability of the actual data or uncertainty in simulated data, but instead focuses on the statistical properties of the discrepancy between the model and the actual data (which includes the approximation error of simulating the model instead of solving it analytically). He proposes a lower bound for a (1-R^2) measure of fit between the model and the actual data based on the minimization of the spectrum of that discrepancy. Watson's approach deals explicitly with the issue of evaluating models which are known to be an incorrect DGP for the actual data. However, his (1-R^2) measure of fit is evaluated informally and does not provide much insight into possible re-specifications that might improve the model.

Finally, another approach considers both the sampling variability of the actual data and the uncertainty in simulated data when evaluating the fit of the model, either allowing for variability in the parameters of the model while keeping the exogenous processes fixed, as DeJong, Ingram and Whiteman (1996), or allowing for both the parameters and the exogenous processes to vary, as Canova and De Nicolò (1995).

This paper contributes to this literature by proposing an alternative measure of fit to evaluate a multivariate dynamic model and to compare the performance of alternative model specifications. The evaluation methodology proposed here addresses two important issues highlighted in the literature in a unifying fashion.

Firstly, we explicitly acknowledge that the solution paths generated by the model for the variables of interest are only approximations to the true model solution. Some simulation techniques approximate model solutions with an arbitrary degree of accuracy but are so demanding in terms of either complexity or computer time that the most common position is to use faster but less accurate approximations (see Marcet (1994))[2]. Hence, it may not be so reasonable to assume that the approximation is the original model as is usual. Watson (1993) also recognises that there is an approximation error but, contrary to his approach, we take it into account when deriving a formal test of the distance between the model and the observed data.

Secondly, as in the last approach, our tests take into account both the sampling variability of actual data and the uncertainty in the simulated series. While not excluding the possibility of stochastic parameters in the model, the uncertainty we consider in the model derives from the fact that there exists an approximation error. As in Diebold, Ohanian and Berkowitz (1995), the measure of distance and tests presented in this paper evaluate how well the model matches the spectral density matrix of the actual data. However, we compare actual to simulated data by treating them as samples from their unknown DGP and hence both spectral density matrices are estimated with error. The required asymptotic theory is developed to test the hypothesis that the multivariate spectral density matrix of the model and the actual data (or of two models) are alike (either equal or differing to an arbitrary prespecified limit). DeJong, Ingram and Whiteman (1996) and Canova and De Nicolò (1995) suggest also measures of fit which are symmetric in the statistical treatment of model and actual data. The main differences between the methodology proposed in this paper and that of DeJong, Ingram and Whiteman (1996) and of Canova and De Nicolò (1995) are, first, that we estimate both sets of statistics in a classical way instead of using Bayesian methods and, second, that model and actual data (or another model) statistics are compared using asymptotic tests.

The rest of the paper is structured as follows. Section 2 proposes a formal *measure of fit* to evaluate models against actual data using multivariate frequency domain techniques. An

[2]Many tests have been proposed in the literature (see, for example, Den Haan and Marcet (1994)) for the accuracy of the numerical approximation to the solution of a model. Such tests are certainly helpful in selecting more accurate approximations for a given model, but as long as the approximation error exists it may affect the properties of the model and how well it reproduces the observed properties of the actual data.

asymptotic test is derived for the hypothesis that the distance between the spectral density matrices of simulated model series and actual data is zero or less than an arbitrary prespecified bound. It is especially suitable for assessing the performance of models that focus on the dynamic behavior of a set of key variables at a certain frequency range, such as business cycle models. In a similar fashion, Section 3 derives a formal test for the equivalence of competing models, possibly misspecified, or of alternative parameterizations of a same model (i.e. to perform global sensitivity analysis). It can be seen as a *comparison test* between different model specifications. The test is able to address the complicated and interesting issue of comparing misspecified models, testing whether they are similar to each other while being different from the actual DGP. Section 4 examines the finite sample properties of both tests via Monte Carlo experiments. The sensitivity of the tests proposed in this paper to the sample size and to the parameter structure is also studied.

Section 5 applies the *fit* test to alternative versions of an International Real Business Cycle based on Backus, Kehoe and Kydland (1993). We want to evaluate the effect of final goods trade, common shocks and spillovers across national disturbances on the macroeconomic interdependencies between countries. Assessment of the fit of these models focusses on how well they reproduce the bivariate spectral density matrix of the US and European GDPs at business cycle frequencies using the *fit* test. The spectral density matrices implied by each alternative model are compared using the *comparison* test proposed in Section 3. Section 6 summarizes and concludes.

2 A measure of distance between simulated and actual data

We are interested in comparing the dynamic properties of a multivariate vector of observed economic data with those generated from a simulated multivariate dynamic model.

Let y_t be the $N \times 1$ vector of actual data series and x_t be the $N \times 1$ vector of data simulated from the model, where $t = 1,\ldots,T$. We view the artificial series x_t in a similar way as the observed data y_t, as samples of an unknown DGP. The reason being that x_t is not typically generated through an analytical solution of the model but approximating that solution with some simulation algorithm. The DGP of the artificial data is unknown but is known to be sufficiently close to the model (see Marcet (1994)), so that the artificial series obtained can be used to evaluate the performance of the theoretical model as long as we take the possible approximation error into account.

In order to take into account all of the interactions between actual and artificial data, we use the joint spectral density matrix for the 2N×1 vector $z_t = [\, y_t' \;\; x_t' \,]'$. For each frequency $\omega \in [-\pi,\ \pi]$, let $f(\omega;\gamma) = \{f_{ij}(\omega;\gamma)\}$ denote the 2N×2N theoretical spectral density matrix of vector z_t in which the model series x_t are obtained using the parameter vector γ. The ij-th element represents the corresponding crosspectrum between a pair of variables, z_{it} and z_{jt}, i,j = 1,...,2N. $f(\omega;\gamma)$ is arranged as follows

$$f(\omega;\gamma) = \begin{pmatrix} f^{\mathbf{y}}(\omega) & f^{\mathbf{yx}}(\omega;\gamma) \\ f^{\mathbf{xy}}(\omega;\gamma) & f^{\mathbf{x}}(\omega;\gamma) \end{pmatrix}$$

where $f^{\mathbf{y}}(\omega)$ ($f^{\mathbf{x}}(\omega;\gamma)$) corresponds to the crosspectra between the actual (artificial) series and the other two submatrices correspond to crosspectra between pairs of actual and model series. Let $\hat{f}(\omega;\gamma) = \{\hat{f}_{ij}(\omega;\gamma)\}$ denote the estimated spectral density matrix, defined as

$$\hat{f}(\omega;\gamma) = \frac{1}{2\pi}\sum_{\tau=-T+1}^{T-1} k_M(\tau)\hat{\Gamma}(\tau;\gamma)e^{-i\omega\tau}$$

where $\hat{\Gamma}(\tau;\gamma)$ is the variance-covariance matrix estimate of vector z_t for lag τ, $k_M(\tau)$ is the lag window function and M is the lag/spectral window parameter. Under general conditions of stationarity of z_t, and standard assumptions on $k_M(\tau)$ and M (see the appendix for the assumptions and the properties of spectral estimators), $\hat{f}(\omega;\gamma)$ is a consistent and asymptotically unbiased estimator of $f(\omega;\gamma)$ with the following asymptotic distribution

$$\sqrt{\frac{\nu}{2}}\mathrm{vec}\hat{f}(\omega;\gamma) \overset{a}{\sim} \mathrm{CN}_{N^2}\left(\sqrt{\frac{\nu}{2}}\mathrm{vec}f(\omega;\gamma), \overline{f(\omega;\gamma)} \otimes f(\omega;\gamma)\right) \quad for\ \omega \neq 0, \pm\pi \tag{1}$$

where $\overset{a}{\sim}\mathrm{CN}_{N^2}$ indicates an asymptotic multivariate complex Normal distribution of dimension N^2, $\otimes$ denotes the kronecker product and ν is a constant called "equivalent degrees of freedom" of the spectral estimator and is defined as $\nu = \frac{2T}{M\int_{\tau=-\infty}^{+\infty} k_M^2(\tau)d\tau}$ (the value of ν for each lag window function is tabulated, see Priestley (1981)).

We define, for each frequency ω, the *theoretical distance* between actual and model spectral density matrices by:

$$D(\omega;\gamma) = \mathrm{S}\ \mathrm{vec}f(\omega;\gamma) = \mathrm{vec}f^{\mathbf{y}}(\omega) - \mathrm{vec}f^{\mathbf{x}}(\omega;\gamma) \tag{2}$$

where vec(.) is the column vectorization operator, and S is a $N^2 \times (2N)^2$ selection matrix that transforms $\mathrm{vec}f(\omega;\gamma)$ into the difference between the elements of its submatrices $f^{\mathbf{y}}$ and $f^{\mathbf{x}}$. Note that $D(\omega;\gamma)$ results in a $N^2 \times 1$ vector.

As an illustration of our measure of distance $D(\omega;\gamma)$, let y_t follow a bi-variate zero-mean VAR(1) process, $y_t = \Phi^y y_{t-1} + \epsilon_t$, t = 1,...,T, where Φ^y is a 2×2 parameter matrix whose eigenvalues lie all inside the unit circle, and ϵ_t is a bi-variate white noise (WN) vector, with $E[\epsilon_t\ \epsilon_t'] = \Omega_\epsilon$. Then, we can express y_t in its MA(∞) form

$$y_t = (I_2 - \Phi^y L)^{-1}\epsilon_t = \sum_{\tau=0}^{\infty}(\Phi^y)^\tau \epsilon_{t-\tau} = \sum_{\tau=0}^{\infty}(\Phi^y L)^\tau \epsilon_t = \sum_{\tau=0}^{\infty} A_\tau^y L^\tau \epsilon_t = A^y(L)\epsilon_t$$

and obtain the theoretical spectrum:

$$f^{\mathbf{y}}(\omega) = \frac{1}{2\pi}(I_2 - \Phi^y e^{-i\omega})^{-1}\Omega_\epsilon(I_2 - (\Phi^y)'e^{i\omega})^{-1} = A^y(e^{-i\omega})f^\epsilon(\omega)A^y(e^{i\omega})'$$

The simulated series generated from the model x_t being also covariance-stationary, they can be expressed in their MA form, $x_t = A^x(L)\ u_t$, where u_t is a bi-variate WN process as ϵ_t. Hence,

$$f^{\mathbf{x}}(\omega;\gamma) = A^x(e^{-i\omega})f^{\mathbf{u}}(\omega;\gamma)A^x(e^{i\omega})'$$

A plot of the row i, column j element of A_τ as a function of the lag τ is called the *impulse response function.* Hence, the measure of distance between model and actual data series $D(\omega;\gamma)$ we have defined as the difference between their spectral density matrices, can also be thought of as analogous to a measure of distance between the theoretical *impulse responses* of the model and the actual data,

$$D(\omega;\gamma) = \text{vec}f^{\mathbf{y}}(\omega) - \text{vec}f^{\mathbf{x}}(\omega;\gamma) = \text{vec}[A^y(e^{-i\omega})f^\epsilon(\omega)A^y(e^{i\omega})' - \tag{3}$$
$$A^x(e^{-i\omega})f^{\mathbf{u}}(\omega;\gamma)A^x(e^{i\omega})'] = \frac{1}{2\pi}\text{vec}[A^y(e^{-i\omega})\Omega_\epsilon A^y(e^{i\omega})' - A^x(e^{-i\omega})\Omega_u A^x(e^{i\omega})']$$

Since ϵ_t and u_t are WN processes, $f^\epsilon(\omega)$ and $f^{\mathbf{u}}(\omega;\gamma)$ are flat, and equal if $\Omega_\epsilon = \Omega_u$. In that case, $D(\omega;\gamma)$ measures the distance frequency by frequency between the "squared" theoretical impulse responses of the actual data and those of the model. If $\Omega_\epsilon \neq \Omega_u$, the distance between the two spectral density matrices takes into account the different covariance structures of actual and simulated data innovations i.e. ϵ_t and u_t.

Now we define the *estimated distance* between actual and model spectral density matrices by:

$$\hat{D}(\omega;\gamma) = S\text{vec}\hat{f}(\omega;\gamma) = \text{vec}\hat{f}^{\mathbf{y}}(\omega) - \text{vec}\hat{f}^{\mathbf{x}}(\omega;\gamma) \tag{4}$$

The asymptotic distribution and properties of $\hat{D}(\omega;\gamma)$ are derived from those of $\hat{f}(\omega;\gamma)$ (see the appendix):

(a) asymptotic complex Normal distribution

$$\sqrt{\frac{\nu}{2}}\Big(\hat{D}(\omega;\gamma) - D(\omega;\gamma)\Big) = \sqrt{\frac{\nu}{2}}S\text{vec}\Big(\hat{f}(\omega;\gamma) - f(\omega;\gamma)\Big) \stackrel{\cdot}{\sim} \text{CN}_{N^2}, \tag{5}$$

(b) asymptotic unbiasedness

$$\lim_{T\to\infty} E[\hat{D}(\omega;\gamma)] = \lim_{T\to\infty} E[S\text{vec}\hat{f}(\omega;\gamma)] = S\text{vec}f(\omega;\gamma) = D(\omega;\gamma) \tag{6}$$

(c)-(d) asymptotic variance-covariance structure

$$\Sigma_D(\omega;\gamma) = \lim_{T\to\infty} \text{var}\left[\sqrt{\frac{\nu}{2}}\hat{D}(\omega;\gamma)\right] = \lim_{T\to\infty} var[S\sqrt{\frac{\nu}{2}}\text{vec}\hat{f}(\omega;\gamma)] =$$
$$= S\overline{f(\omega;\gamma)} \otimes f(\omega;\gamma)S' \ , \quad \omega \neq 0, \pm\pi \tag{7}$$

Therefore, for $\omega \neq 0, \pm\pi$,

$$\sqrt{\frac{\nu}{2}}\hat{D}(\omega;\gamma) \overset{\cdot}{\sim} \text{CN}_{N^2}\Big(\sqrt{\frac{\nu}{2}}D(\omega;\gamma),\ \Sigma_D(\omega;\gamma)\Big) \tag{8}$$

Recall that we are interested in evaluating the performance of multivariate dynamic models that are, in general, solved by approximation through simulation techniques. It means that the model yields a multivariate vector series of same dimension as the actual data for each time it is simulated, and hence a $\hat{f}_h(\omega;\gamma)$ is estimated keeping y_t fixed and using x_{ht} at each replication, for h = 1,...,H. In practice, what we are interested in obtaining is the average across the H replications of the estimated distance $\hat{D}(\omega;\gamma) = \frac{1}{H}\sum_{h=1}^{H}\hat{D}_h(\omega;\gamma) = \frac{1}{H}\sum_{h=1}^{H} \text{S vec}\hat{f}_h(\omega;\gamma)$. Given that x_{ht} are iid, that average is the sample mean of iid random variables, where the sample size is H and the iid r.v. are $\text{vec}\hat{f}_h(\omega;\gamma)$ premultiplied by S, for h = 1,...,H. Hence, the sample mean across replications has same distribution and theoretical mean as each of its elements, $\text{vec}\hat{f}_h(\omega;\gamma)$, and a variance which is $\frac{1}{H}\text{var}(\text{vec}\hat{f}_h(\omega;\gamma))$. Hence, instead of (8), for H finite, the asymptotic distribution of $\hat{D}(\omega;\gamma)$ is:

$$\sqrt{H}\sqrt{\frac{\nu}{2}}\hat{D}(\omega;\gamma) \overset{\cdot}{\sim} \text{CN}_{N^2}\Big(\sqrt{H}\sqrt{\frac{\nu}{2}}D(\omega;\gamma),\ S\overline{f(\omega;\gamma)} \otimes f(\omega;\gamma)S'\Big), \tag{9}$$

In what follows, for simplicity, we will consider only the case of H=1. All results can be easily generalised to a generic H. However, it can be the case that for very short sample sizes where the convergence of $\hat{D}(\omega;\gamma)$ to its limiting distribution is not ensured and may be closer to a χ^2 distribution (see property (a) in the appendix), the actual number of $\hat{D}_h(\omega;\gamma)$ elements it aggregates (H) may matter, probably increasing the degrees of freedom of the χ^2 distribution. Obviously, this effect will only be noticeable for large H.

We present next a test as to whether the elements of $D(\omega;\gamma)$ are significantly different from zero for a particular single frequency, without having to deal with complex distributions, and extend it to the case for when we are interested in the significance of the distance of a model from the observed data over a particular frequency range.

2.1 Assessing the fit of a model

Testing for the significance of the distance of a model spectral density matrix from that of actual data for a given frequency ω means that we are testing the following null hypothesis:

$$H_0 : \Lambda D(\omega;\gamma) = 0$$

where Λ is an $N^2 \times N^2$ diagonal selection matrix. Such a matrix may have unequal diagonal elements so as to introduce weights in the measure of distance $D(\omega;\gamma)$ if we care about some relationship more than others. The $((i-1)\times N+j)$-th element of its diagonal represents the weight given to the relationship between the j-th and i-th elements in y_t and x_t, with i, j = 1,..., N.

In a sense, testing such null hypothesis is equivalent to formally testing whether the model under evaluation is representative of the class of models which have the following particular characteristic: the spectral density matrix of a subset of model series equal to that of the corresponding actual data series, for a given ω frequency.

To test H_0, we construct the following test statistic

$$fit(\omega;\gamma) = \left(\overline{\sqrt{\frac{\nu}{2}}\Lambda\hat{D}(\omega;\gamma)}\right)'\left(\Lambda\Sigma_D(\omega;\gamma)\Lambda'\right)^{-1}\left(\sqrt{\frac{\nu}{2}}\Lambda\hat{D}(\omega;\gamma)\right) \tag{10}$$

which is a real number because we are multiplying element by element the standarized estimated distance vector of interest $(\Lambda\Sigma_D(\omega;\gamma)\Lambda')^{-\frac{1}{2}}\sqrt{\frac{\nu}{2}}\Lambda\hat{D}(\omega;\gamma)$ by its conjugate. Given that $\hat{f}(\omega;\gamma)$ is a consistent estimator of $f(\omega;\gamma)$, $\Sigma_D(\omega;\gamma)$ can be replaced by $\hat{\Sigma}_D(\omega;\gamma) = S\overline{\hat{f}(\omega;\gamma)}\otimes\hat{f}(\omega;\gamma)S'$. Thus, the test statistic $fit(\omega;\gamma)$ becomes

$$fit(\omega;\gamma) = \left(\overline{\sqrt{\frac{\nu}{2}}\Lambda\hat{D}(\omega;\gamma)}\right)'\left(\Lambda\hat{\Sigma}_D(\omega;\gamma)\Lambda'\right)^{-1}\sqrt{\frac{\nu}{2}}\Lambda\hat{D}(\omega;\gamma)$$

Under H_0, (8) becomes $\sqrt{\frac{\nu}{2}}\hat{D}(\omega;\gamma) \stackrel{.}{\sim} CN_{N^2}\left(0,\Sigma_D(\omega;\gamma)\right)$ and, hence, the asymptotic distribution of the test statistic is

$$fit(\omega;\gamma) \;\stackrel{.}{\sim}\; \chi^2_{(N^2-Q)}\,, \quad \omega \neq 0, \pm\pi \tag{11}$$

where Q is the number of zero elements in the diagonal of Λ. H_0 will be rejected and the distance between the model and the actual data found significantly different from zero if $fit(\omega;\gamma)$ is greater than the critical value of a $\chi^2_{(N^2-Q)}$, for a selected significance level α.

When we suspect a model to be false, we may be more interested in testing whether its distance to the actual data is smaller than an arbitrary constant C rather than testing the above null of zero distance. Then, the relevant null hypothesis is

$$H_0 : \Lambda D(\omega;\gamma) \;\leq\; C$$

The only difference with respect to the test described above is that, under the null, the test statistic $fit(\omega;\gamma)$ has a non-central asymptotic $\chi^2_{(N^2-Q,\delta)}$ distribution, where δ is a non-centrality parameter of value

$$\delta = \left(\overline{\sqrt{\frac{\nu}{2}}(\Lambda D(\omega;\gamma)-C)}\right)'\left(\Lambda\Sigma_D(\omega;\gamma)\Lambda'\right)^{-1}\left(\sqrt{\frac{\nu}{2}}(\Lambda D(\omega;\gamma)-C)\right)$$

We may want to test the significance of the distance of the model to the actual data for a given set of L frequencies $[\omega_1, \omega_2]$, where $\omega_1, \omega_2 \neq 0, \pm\pi$, e.g. the frequencies associated with the business cycle, which is typically associated to those cycles whose periods lie within 2 and 8 years. Then, the null hypothesis is

$$H_0 : \Lambda D(\omega;\gamma) = 0, \quad \forall\omega \in [\omega_1, \omega_2]$$

Under H_0, the test statistic $fit(\omega;\gamma)$ becomes

$$\text{fit}([\omega_1,\omega_2];\gamma) = \sum_{\omega=\omega_1}^{\omega_2} \left(\sqrt{\frac{\nu}{2}}\Lambda\hat{D}(\omega;\gamma)\right)' \left(\Lambda\hat{\Sigma}_D(\omega;\gamma)\Lambda'\right)^{-1} \sqrt{\frac{\nu}{2}}\Lambda\hat{D}(\omega;\gamma) \tag{12}$$

which has a $\chi^2_{L(N^2-Q)}$ asymptotic distribution, since it is the sum of L independent $\chi^2_{(N^2-Q)}$ variates.

3 Comparing alternative models

An important characteristic of a model is its performance relative to other models in capturing a particular aspect of reality. In this section we develop a formal test for assessing whether *two different model specifications* display similar dynamic properties for a selected group of variables. If this is the case, they can be considered equally successful in capturing the dynamic properties observed in the actual data. How far or close to the actual data each of the models is can be assessed using the *fit* test presented in the previous section.

Let $x_t^i(\gamma_i)$ denote the N×1 vector obtained by simulating model m_i with the particular set of parameters γ_i. We are interested in taking into account all the interactions between the two alternative model specifications $(m_1;\gamma_1)$ and $(m_2;\gamma_2)$ we want to compare. Therefore, the relevant z_t vector whose spectral density matrix we want to estimate is the 2N×1 vector $z_t = [\ x_t^1(\gamma_1)'\ \ x_t^2(\gamma_2)']'$. Both $f(\omega;\gamma_1,\gamma_2)$ and $\hat{f}(\omega;\gamma_1,\gamma_2)$ are 2N×2N matrices.

Alternatively, we could also compare the relative success of the same model m_1 under two alternative parameterizations γ_1 and γ_2. Then, the relevant z_t vector would be $z_t = [\ x_t^1(\gamma_1)'\ \ x_t^1(\gamma_2)']'$. This can be regarded as a formalization of the sensitivity analysis the researcher may want to undertake over certain elements of the parameter vector γ. A more global sensitivity analysis could be performed with a modified version of the *fit* test, in which y_t is compared to an average of x_t^i over γ. If one considers all possible realistic values of the parameter space we would be in fact introducing a way to deal with parameter uncertainty, along the same lines as Canova (1994)-(1995), DeJong, Ingram and Whiteman (1996) and Canova and De Nicolò (1995).

In a similar way as in Section 2, we define for each frequency ω the *theoretical distance* between two alternative model specifications (m_1, γ_1) and (m_2, γ_2) by:

$$D(\omega;\gamma_1,\gamma_2) = Svecf(\omega;\gamma_1,\gamma_2) = vecf^2(\omega;\gamma_2) - vecf^1(\omega;\gamma_1) \tag{13}$$

where, as in section 2, S is a $N^2 \times (2N)^2$ selection matrix and $D(\omega;\gamma_1,\gamma_2)$ results in a $N^2 \times 1$ vector. Note that we compare the dynamic properties of two models independently of how close each of them is to the actual data. Had we included the actual data vector y_t, we would have defined $D(\omega;\gamma_1,\gamma_2)$ as the difference between the fit of each model, i.e. $D(\omega;\gamma_1,\gamma_2)$ would have been defined as $(vecf^2(\omega;\gamma_2) - vecf^y(\omega)) - (vecf^1(\omega;\gamma_2) - vecf^y(\omega))$, which is equal to (13).

Our definition of the distance between two alternative models allows us to test the null that two models have the same spectral density matrices (or submatrices if we are interested in the dynamic properties of only a subset of the variables included in the models) both in the case in which one of the models is the actual DGP and when neither of them is, i.e. when the comparison is made between misspecified models.

The *estimated distance* between the two alternative model specifications is defined as follows:

$$\hat{D}(\omega;\gamma_1,\gamma_2) = Svec\hat{f}(\omega;\gamma_1,\gamma_2) = vec\hat{f}^2(\omega;\gamma_2) - vec\hat{f}^1(\omega;\gamma_1) \tag{14}$$

and has similar asymptotic properties to $\hat{D}(\omega;\gamma)$. For $\omega \neq 0, \pm\pi$,

$$\sqrt{\frac{\nu}{2}}\hat{D}(\omega;\gamma_1,\gamma_2) \dot{\sim} \mathrm{CN}_{N^2}\Big(\sqrt{\frac{\nu}{2}}D(\omega;\gamma_1,\gamma_2),\ \Sigma_D(\omega;\gamma_1,\gamma_2) = S\overline{f(\omega;\gamma_1,\gamma_2)}\otimes f(\omega;\gamma_1,\gamma_2)S'\Big) \tag{15}$$

Note that we are assuming that both model specifications have been simulated the same number of times, H. As before, for values of $H \neq 1$, $\hat{D}(\omega;\gamma_1,\gamma_2)$ is replaced by $\sqrt{H}\hat{D}(\omega;\gamma_1,\gamma_2)$.

Testing for the equal performance of two model specifications with respect to the dynamic properties of a selected set of series requires testing whether the null hypothesis

$$H_0 : \Lambda D(\omega;\gamma_1,\gamma_2) = 0$$

at frequency ω, is accepted; again, Λ is a selection matrix.

In a similar fashion as the $fit(\omega;\gamma)$ test, we construct the following test statistic:

$$\mathrm{comp}(\omega;\gamma_1,\gamma_2) = \Big(\overline{\sqrt{\frac{\nu}{2}}\Lambda\hat{\mathrm{D}}(\omega;\gamma_1,\gamma_2)}\Big)'\Big(\Lambda\hat{\Sigma}_{\mathrm{D}}(\omega;\gamma_1,\gamma_2)\Lambda'\Big)^{-1}\sqrt{\frac{\nu}{2}}\Lambda\hat{\mathrm{D}}(\omega;\gamma_1,\gamma_2) \tag{16}$$

Under the null, and for $\omega \neq 0, \pm\pi$

$$comp(\omega;\gamma_1,\gamma_2) \dot{\sim} \chi^2_{(N^2-Q)} \tag{17}$$

H_0 will be rejected and the relative distance between the model specifications is significantly different from zero if $comp(\omega; \gamma_1, \gamma_2)$ is greater than the critical value of a $\chi^2_{(N^2-Q)}$, for a given significance level α, and where Q is the number of zero elements in the diagonal of Λ.

Here, too, we may want to test the significance of the distance between two alternative model specifications for a given set of L frequencies, $[\omega_1, \omega_2]$, where $\omega_1, \omega_2 \neq 0, \pm\pi$. Then the null hypothesis is

$$H_0 : \Lambda D(\omega; \gamma_1, \gamma_2) = 0, \quad \forall \omega \in [\omega_1, \omega_2].$$

To test such H_0, we use an aggregated version of $c\acute{o}mp(\omega; \gamma)$

$$\text{comp}([\omega_1, \omega_2]; \gamma_1, \gamma_2) = \sum_{\omega=\omega_1}^{\omega_2} \left(\sqrt{\frac{\nu}{2}} \Lambda \hat{\mathrm{D}}(\omega; \gamma_1, \gamma_2)\right)' \left(\Lambda \hat{\Sigma}_{\mathrm{D}}(\omega; \gamma_1, \gamma_2)\Lambda'\right)^{-1} \sqrt{\frac{\nu}{2}} \Lambda \hat{\mathrm{D}}(\omega; \gamma_1, \gamma_2) \quad (18)$$

which has a $\chi^2_{L(N^2-Q)}$ asymptotic distribution under H_0.

4 Performance of the tests: Monte Carlo evidence

In this section we present some Monte Carlo evidence to examine the finite sample properties of the two proposed tests, $fit([\omega_1, \omega_2]; \gamma)$, and $comp([\omega_1, \omega_2]; \gamma_1, \gamma_2)$ for the case of a bivariate model (N=2). Experiments have been conducted for the following sample sizes: T = 100, 200 and 500[3]. In all cases, we evaluate the performance of the bivariate model at business cycle frequencies and we define $[\omega_1, \omega_2]$ as the set of frequencies associated with cycles 8 to 2 years long, and all variables are given equal weight i.e. $\Lambda = I_{N^2} = I_4$.

The y_t series have been generated from a bivariate VAR(1) of the form:

$$y_t = \Phi^{y} y_{t-1} + \epsilon_t, \quad t = 1, ..., T$$

where Φ^{y} is a 2x2 parameter matrix and ϵ_t a bivariate white noise (WN) process. In particular, $\Phi^{y} = \begin{pmatrix} 0.7 & 0.1 \\ 0.2 & 0.6 \end{pmatrix}$. These concrete values generate a bi-variate covariance stationary process in which the two series are correlated to each other and both show the "typical spectral shape" Granger (1964) attributes to macroeconomic data series, i.e. most of the variability concentrated in the lower frequencies.

In order to see the sensitivity of the tests to the particular dynamic structure of the model, the x_t^i series have been simulated from three alternative models, i = 1, 2 and 3, of the form:

$$x_t^i = \Phi^{i} x_{t-1}^i + u_t^i, \quad t = 1, ..., T$$

[3]We have also experimented with T=1000 obtaining results very similar to those with T=500, therefore we do not report them. There are slight differences but all in the same lines as the changes observed when going from T=200 to T=500.

where Φ^i is a 2×2 parameter matrix and the residuals vector, u_t^i, is a bivariate WN process. The three models simulated to obtain the x_t^i vector series differ on their parameter structure.

- Model 1: VAR(1) with $\Phi^1 = \Phi^y$.
- Model 2: No spillovers: VAR(1) with $\Phi^2 = \begin{pmatrix} 0.7 & 0 \\ 0 & 0.6 \end{pmatrix}$.
- Model 3: No dependence structure: Bivariate WN process, $x_t^3 = u_t^3$. Hence, $\Phi^3 = 0_2$.

Error vectors, both u_t^i and ϵ_t, have been generated with the Matlab 4.2 random generator from a standard bivariate Normal distribution[4]. Since $\{u_t^i\}_{t=1}^T$ and $\{\epsilon_t\}_{t=1}^T$ have the same spectral density matrices, $\{x_t^i\}_{t=1}^T$ will have equal spectral density matrix to $\{y_t\}_{t=1}^T$ as long as the parameters of their respective DGPs are the same. 100 extra observations were generated for each error $\{\epsilon_t\}_{t=1}^T$ and residual $\{u_t\}_{t=1}^T$ vector sequences in order to avoid initial condition problems.

Before proceeding, we have to choose both the functional form of the lag/spectral window, $k_M(\tau)$, and the lag/spectral window parameter estimator, $\hat{M}$, so that they fulfill conditions (ii), (iii) and (iv) of spectral estimates as stated in the appendix. In this we follow Priestley (1981) and Andrews (1991) who show that the Quadratic Spectral window is optimal (see comment on selection of the appropriate lag/spectral window function in the appendix). We also follow Andrews (1991) in choosing an "automatic bandwidth estimator $\hat{M}$" which is a function of the data and asymptotically optimal under general conditions (see comment on the estimator for the lag/spectral window parameter, also in the appendix).

4.1 Fit Test

Table 1 reports the finite sample behavior of the fit test, when the null is that the model evaluated follows a VAR(1) with same parameter values as the actual DGP. The numbers displayed in Table 1 are the percentage rejection across 1000 Monte Carlo replications of the null hypothesis of zero distance between the model and the observed data for the tests $fit(\omega;\gamma)$ and $fit([\omega_1,\omega_2];\gamma)$. Under "Model i" (i = 1,2 and 3) we compute the test statistic comparing the spectral density matrix of $\{y_t\}_{t=1}^T$ to the one of $\{x_t^i\}_{t=1}^T$. Therefore, the numbers under "Model 1" measure the *size* of the fit test and the rest measure the *power* under different alternative hypotheses,i.e. x_t^i coming from Models 2 or 3.

[4]Except for the seed, to avoid the possibility that the random number series were exactly the same for ϵ_t than for u_t^i. Instead, the u_t^i series have been generated with the same random number generator, starting from the point where the ϵ_t series ended. This way, ϵ_t and u_t^i, i = 1,2 and 3, are independent r.v. and inference can be constructed on the distance between transformations of them.

The performance of each model, either correctly specified (Model 1) or not (models 2 and 3), with respect to the actual data is evaluated for one single frequency (ω_1 is the frequency associated with cycles of periodicity 8 years, and ω_2 with those of periodicity 2 years) and for the inclusive business cycle set of frequencies ($[\omega_1, \omega_2]$). The critical values used with $fit(\omega;\gamma)$ and $fit([\omega_1,\omega_2];\gamma)$ are the critical values of a $\chi^2_{(N^2)}$ and a $\chi^2_{(N^2L)}$ distribution, respectively, which correspond to the theoretical size indicated (either 5% or 10%). N=2 in all cases. The quantity L depends on the lag/spectral window parameter $\hat{M}$ we use in the estimation of the spectral density matrix as explained in the appendix. $\hat{M}$ depends on the length of the series, T, as well as on the parametric DGP z_t is supposed to follow. This guarantees that spectral estimates at different frequencies are independent (see comment on property (c) at the appendix). In particular, we estimate the crosspectra at frequencies distant $\frac{\pi}{\hat{M}}$ to each other. For a sample size T of 100, 200 and 500 observations, the estimation procedure followed here yields L=4, 4 and 5 frequency points, respectively.

Table 1 shows that the *size* of the test (panel "Model 1") is found smaller than its theoretical value even for sample size of 100 observations. This indicates that the empirical distribution of the test statistic is skewed relative to the theoretical one, more concentrated around values closer to zero. The *size* of the aggregate version of the test, $fit([\omega_1,\omega_2];\gamma)$, is poor for sample size T=100. The reason is simply that we are aggregating the small sample biases of the spectral estimates each single-frequency test statistic carries[5]. However, this effect disappears fast as the sample size increases.

One feature Table 1 shows is the low percentage rejection of models 2 and 3 at the frequency associated roughly to cycles of 2 years length i.e. ω_2. This should not be interpreted as a low power of the test but as the ability of the test to capture the fact that although two models may look different in the time domain, they may generate similar dynamics for a particular frequency range. In this case, with respect to Model 2, the two VAR models have very similar spectra except at the lower frequencies. For the particular DGP considered, ω_2 is outside these lower frequencies and therefore discrimination between different VAR models is difficult (low rejection frequency). A White Noise process (Model 3) has a flat spectral function which can have the same power density as a VAR for a particular frequency, with similar values as well in the neighborhood of that frequency. These two features can be more easily appreciated looking at Figure 1. Figure 1 shows the average, across the 1000 replications of the Monte Carlo experiments, of the $fit(\omega;\gamma)$ test-statistics for different sample sizes and for the whole

[5]This small sample bias has also been found in existing Monte Carlo studies of other kinds of estimation procedures, such as GMM (see Andersen and Sorensen (1995)) or Christiano and Den Haan (1995), who study the sensitivity of the small sample bias in GMM estimation of Business Cycle models to estimation tools used in this paper too, such as the choice of the lag/spectral window and of the bandwidth parameter M.

frequency range. All values are transformed in logs. The horizontal line corresponds to the critical value. The solid line under the critical value is the test-statistic for Model 1. The discontinuous and starred lines correspond to Models 2 and 3, respectively. It can be seen that the distance between Models 1 and 2 (the other VAR, dashed line) is not significant apart from the very low frequencies, and that the distance between the DGP and Model 3 (starred line) is not significant in the vicinity of frequency ω_2, for the reasons explained above.

The aggregate versions of the test, $fit([\omega_1, \omega_2]; \gamma)$, performs in general worse than the one-frequency version at ω_1 but substantially better than when evaluated at ω_2. Because of the possibility of a spurious coincidence of the spectra of two different models around a particular frequency (as noticed around ω_2 when evaluating Model 3), the *fit* test appears clearly more powerful when used to examine the performance of a model in a frequency band rather than when used to assess the fit of a model at ω_2.

It can be seen that the *fit* test manages to correctly rank the models according to the actual DGP: the rejection frequency of Model 1 (equal to the actual DGP) is lower than that of Model 2 (a VAR(1) as Model 1 but with different parameter structure), which in turn is rejected as similar to the actual DGP a lower number of times than Model 3 (bivariate WN) in almost all cases. The particular value of the test statistic (alternatively the associated p-value) can therefore be considered as a *ranking device*: the further away is that value from the fixed critical value (or the significance level) the worse is the performance of the Model with respect to the particular actual data set and for the selected frequencies. Given that we may be interested in evaluating models that almost surely differ from the true DGP, as the majority of multivariate dynamic models in many fields of economics, it is of great interest to evaluate how "poor" is their performance with respect to alternative models aimed at explaining the same relationships observed in the actual data. It is in these cases where the *fit* test can be found most useful as a ranking device, with respect to an arbitrary lower bound B (which we consider to be the best fit possible). In particular, B would be the critical value of the non-central χ^2 distribution we referred to in section 2, once we have fixed the arbitrary minimum distance we expect of the evaluated model to the actual data, i.e. once we have fixed C.

We have also performed some sensitivity analysis on the choice of parameter values for the DGP. If one is interested in capturing the cross-variable relationships observed in the actual data (e.g. interdependencies between the cyclical properties of the GDP of two countries) it is important to check whether our test is able to discriminate between models which give different degrees of interdependence between variables. In the VAR framework, a higher degree of cross-variable dependence can be translated into higher off-diagonal coefficients in the Φ matrix ("spillover" coefficients). Table 2 displays the corresponding Monte Carlo rejection frequencies when the actual data DGP has a different VAR structure:

I) No spillovers: $\Phi^{y} = \Phi^{2} = \begin{pmatrix} 0.7 & 0 \\ 0 & 0.6 \end{pmatrix}$

II) Larger spillovers: $\Phi^{y} = \begin{pmatrix} 0.7 & 0.1 \\ 0.4 & 0.6 \end{pmatrix}$

Figures 2 and 3 represent the log of the average $fit(\omega;\gamma)$ test statistic across the 1000 Monte Carlo replications for Case I and II, respectively.

For Case I the two series in both bivariate vectors y_t and x_t^1 follow independent AR(1) processes. Then, the theoretical power spectrum of the actual data series is more concentrated at the lower frequencies (hence more distinguishable from other models, yielding higher power for the test) while keeping the difficulty to discriminate a VAR from a WN at certain frequencies close to ω_2. However, as the DGP displays smaller interdependencies, $\hat{M}$ becomes smaller (only low order covariance estimates needed to be included in $\hat{f}(\omega;\gamma)$ to capture the DGP dynamics). Given that the number of independent frequency point estimates decreases the smaller $\hat{M}$ is, we obtain estimates of the spectral density matrices (and hence of the test statistics) for a smaller number of frequencies than when using other DGPs. Therefore comparisons between the empirical rejection frequencies obtained with other DGPs have to be made with care.

Under Case II interdependence between the series in both y_t and x_t^1 vectors increases. The spectral power of the the true DGP becomes less concentrated in the lower frequencies the higher the interdependence because it increases the relative magnitude of lagged covariances with respect to the contemporaneous one. As a result, the true spectra and crosspectra are less distinguishable from the WN Model (equal power spectra and crosspectra at every frequency). In general, the Monte Carlo experiments for this case show lower size and power than under Case I, especially for the shorter sample sizes considered, but the other features we have described are unchanged. However, there is an important effect on the automatic bandwidth estimator: $\hat{M}$ increases to capture this higher interdependence in the DGP (to include higher order covariances in $\hat{f}(\omega;\gamma)$) and hence the number of frequency point estimates increases. The difference in the number of frequencies estimated with respect to Case I suggests again that care should be taken in comparing the numbers in Table 2 with those in Table 1. More importantly, it confirms the influence of the characteristics of spectral density estimates (e.g. the effect of $\hat{M}$) in the small sample performance of the *fit* test relative to other aspects of the model evaluation methodology we proposed. Andersen and Sorensen (1995) reach similar conclusions. Also Christiano and den Haan (1995) who study the sensitivity of the small sample bias in GMM estimation of Business Cycle models to estimation tools used in this paper too, such as the choice of the lag/spectral window and of the bandwidth parameter M.

4.2 Comparison test

The comparison test has been introduced as a formal device to assess the difference between the dynamic properties of multivariate time series generated by alternative model specifications.

Table 3 displays the percentage rejection of the null of no difference between the spectral density matrix of the multivariate series of interest when they are simulated from model specification (m_i, γ_i), $\{x_t^i\}_{t=1}^T$, and from (m_j, γ_j), $\{x^j\}_{t=1}^T$, for i,j = 1,2,3. When i = j we report the *size* of the test and when i $\neq$ j, its *power*. The first two blocks of Table 3 (Case11 and Case33) compute the size of the test when the models are equal to the DGP of the actual data (both are Model 1) and when they are equal but both misspecified (both are Model 3). Note that the null hypothesis of the comparison test is that both models have the same spectral density matrix for the multivariate vector of interest, not that this spectral density matrix takes a particular array of values. Hence, the test can be applied equally to correctly specified or to misspecified models. The last two blocks compute the power of the test when the two models are VAR(1) but with different parameters (Case12: one is Model 1 and the other Model 2, and Case13: one is Model 1 and the other a bivariate WN –Model 3–).

Because the asymptotic distribution of the comparison test is the same as that of the fit test, the critical values used are the same as in subsection 4.1.

As in the case of the fit test, the *size* becomes smaller the larger the sample size. Note that the empirical size of the test in Case33 (both models WN) is smaller than in Case11 (both having the same VAR structure). Probably, the main reason for this difference is that the dynamic structure assumed for z_t in the test is more complicated under Case11 and therefore bound to induce higher small sample bias than when the DGP of the simulated series is a WN (Case33)[6]. Further research in terms of DGP parameter sensitivity would clarify this point.

Here again, it is very useful and intuitive to represent graphically the comparison test-statistic under different specifications. Similarly to Figure 1, Figure 4 shows the average value for the $comp(\omega;\gamma)$ test-statistic across the 1000 Monte Carlo replications, and for different sample sizes. As before, all values are transformed in logs. The horizontal lines correspond to the 90% and 95% critical value for the one-frequency test. The solid (starred) line under the critical value is the test statistic for the case in which Model i and and Model j are generated from the same DGP, both are Model 1 (Model 3). The discontinuous (dashdotted) line corresponds

[6]Figure 4 shows that the average values across Monte Carlo replications for the *comp* test statistic under Case11 and under Case33 are similar for all frequencies and clearly below the 90% and 95% critical values. Lower percentage rejection of the null under Case33 (hence, smaller size) than under Case11 means only that the *comp* test statistic is closer to its average at each replication under Case33, which suggests that the spectral density estimates used in the derivation of the *comp* test statistic are more precise in that case than under Case11.

to Case12 (Case13). We can see that the test has difficulties to discriminate between a WN and a VAR in the frequencies immediately higher than $\frac{\pi}{2}$; the *comp* test-statistic lies under the critical value line. This is the same type of effects observed in figure 1: a VAR(1) such as Model 1 can imply at some frequency the same power density as the flat spectrum of a WN (Model 3). As we found for the *fit* test, the performance of the *comp* test depends substantially on the characteristics of the spectral estimates, when applied to VAR models with high persistence and for business cycle frequencies (rather than lower ones).

Also as in the *fit* test case, it is remarkable how informative the value of the rejection frequency is about how different the model specifications compared are. The *comp* test is found for all sample sizes more powerful rejecting a WN model as equal to a VAR(1) (Case13) than two different VAR(1) as equal to each other (Case12).

5 An example

In this section we apply the *fit* and *comp* tests to versions of a standard International Real Business Cycle (IRBC) model to assess how they reproduce the interdependencies observed between the US and European business cycles over the period 1970Q1-1993Q3. We measure these interdependencies by estimating the bivariate spectral density matrix of the US and European real GDPs at business cycle frequencies (those associated to cycles 8 to 2 years long). We have used the Quadratic Spectral density window and the Andrews' optimal bandwidth parameter. Figure 5 plots the linearly detrended logs of the series (quarterly seasonally adjusted real GDP from OECD Main Economic Indicators). The information contained in the estimated spectral density matrix is rearranged in Figure 6, which shows the individual spectra of the two GDPs, the *phase* (whether there exists a lead or lag between the two series) and the *coherence* (the equivalent to the correlation in the frequency domain, also varying between 0 and 1) for all frequencies. The business cycle frequency range is indicated with vertical bars in the phase and coherence plots. All statistics are plotted with their corresponding asymptotic confidence intervals. In the case of the coherence, the lower (upper) horizontal line is the 95% (99%) critical value of the test H_0: coherence=0.

The European GDP is found substantially more volatile than the US one (higher values of the spectrum). The short sample size (95 observations) causes an imprecise estimate in both cases: the 95% asymptotic confidence interval bands are wide and the lower one is not distinguishable from 0 for all frequencies. A significant coherence is found between the two GDPs at all business cycle frequencies (and for most of them, even at a 99% confidence level), although none of the GDPs clearly lead the other one. The well known "locomotive role" of the US economy with respect to the European one is not clear for this period. Figure 5 reveals that this is so because the two GDPs where evolving very synchronized in the 70s and it is only in the 80s when the US GDP clearly leads the European one.

5.1 The models

The benchmark model economy is a standard two-country two-good International Real Business Cycle model. The possible sources of economic fluctuations in the model are stochastic shocks from both the demand and the supply side of the economy, either country-specific or common (i.e. contemporaneous cross-country correlation between shocks). Demand disturbances take the form of exogenous government expenditure shocks while supply disturbances are identified with technology shocks. The specific mechanisms of international transmission of shocks and fluctuations allowed in the model are trade in final goods and services and spillovers in the shocks processes.

Each country specializes in the production of a single differenciated good, in the lines of Backus, Kehoe and Kydland (1993). Each country is populated by a large number of utility maximizers infinitely-lived identical agents. The representative household sells the services of capital and labor, owns all the firms and receives all the profits. The goods produced by the firms will be purchased by the household in order to be consumed or invested. There are complete financial markets within countries and free mobility of physical and financial capital across countries. However, labor is immobile internationally. Variables are measured in per-capita terms. Each household in country h=1,2 has preferences given by the utility function

$$U = E_0 \sum_{t=0}^{\infty} \beta^t (1-\sigma)^{-1} (C_{ht+1}^{\theta} L_{ht+1}^{1-\theta})^{1-\sigma} \tag{19}$$

where C_{ht} is consumption at time t, L_{ht} is the share of time devoted to leisure, $0 < \theta < 1$ is the relative weight of consumption to leisure and σ is the risk aversion parameter. Leisure choices are constrained by:

$$0 \leq L_{ht} + H_{ht} \leq 1 \tag{20}$$

where the total endowment of time is normalized to 1 and H_{ht} is the share of time the representative household devotes to market activities.

There is a representative firm operating in each country that produces output with a constant returns-to-scale Cobb-Douglas production function

$$Y_{ht} = A_{ht} K_{ht}^{\alpha} (X_{ht} H_{ht})^{1-\alpha} \tag{21}$$

where K_{ht} is the capital input, α is the share of labor in GDP, and X_{ht} is labor-augmenting Harrod-neutral technological progress with a deterministic growth rate equal to θ_x-1, i.e. X_{ht} =

$\theta_x X_{h,t-1}$ with $\theta_x \geq 1$. X_{ht} represents permanent technical change while temporary changes in technology are represented by the variation in total factor productivity according to the following joint process

$$\begin{bmatrix} lnA_{1t} \\ lnA_{2t} \end{bmatrix} = \begin{bmatrix} \rho_{A1} & \nu_{12} \\ \nu_{21} & \rho_{A2} \end{bmatrix} \begin{bmatrix} lnA_{1t-1} \\ lnA_{2t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{bmatrix}, \quad \epsilon_t \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\epsilon 1}^2 & \psi \\ \psi & \sigma_{\epsilon 2}^2 \end{bmatrix} \right) \tag{22}$$

where ρ_{Ah} is the parameter that governs the persistence of the technology process within country h, ν_{hj} is the spillover parameter determining the speed at which changes in technology in country h are transmitted to country j, $\sigma_{\epsilon h}$ is the standard deviation of the stationary exogenous technology shocks in country i, ϵ_{ht}, and ψ represents the covariance between the innovations to technology processes in both countries, i.e. a common shock to both countries total factor productivities will be characterized by a high ψ, and the higher ψ the less country-specific is the shock.

Multiple goods are introduced by assuming that Y_{ht} can be either used domestically or exported

$$Y_{1t} = Y_{11t} + \frac{\Pi_2}{\Pi_1}\tilde{Y}_{12t}, \quad Y_{2t} = \frac{\Pi_1}{\Pi_2}\tilde{Y}_{21t} + Y_{22t}$$

where Π_h represents the size of each country, e.g. population, and the tilde indicates that the production of country j imported by country h (Y_{jht}) has been transformed into per capita terms of the importing country (country h). Imports ($\tilde{Y}_{jht}$) and home-produced home-consumed goods (Y_{hht}) are used in the production of a final good in each country, V_{ht}, according to the following constant elasticity of substitution technology (see Armington (1969)):

$$V_{1t} = (\omega_1 Y_{11t}^{1-\rho} + \omega_2 \tilde{Y}_{21t}^{1-\rho})^{\frac{1}{1-\rho}}, \quad V_{2t} = (\omega_1 Y_{22t}^{1-\rho} + \omega_2 \tilde{Y}_{12t}^{1-\rho})^{\frac{1}{1-\rho}} \tag{23}$$

where $\frac{1}{\rho}$ is the elasticity of substitution between domestic and foreign goods and ω_1 and ω_2 are parameters regulating the domestic and foreign content of GNP. This is a very convenient specification because it allows for cross-hauling (a situation in which a country imports and exports goods of the same category) and permits both countries to produce same categories of goods. The relative price of imports to exports (terms of trade) is given by

$$P_{ht} = \frac{\partial V_{ht}/\partial \tilde{Y}_{jht}}{\partial V_{ht}/\partial Y_{hht}} = \frac{\omega_2 \tilde{Y}_{jht}^{-\rho}}{\omega_1 Y_{hht}^{-\rho}}$$

where $\omega_1 = (1-\text{MS})^\rho$, $\omega_2 = \text{MS}^\rho$ and MS is the import share in output. A value of MS of zero would mean that there is no trade between the economies (autarky).

Firms accumulate capital goods according to the following law of motion

$$K_{h,t+1} = (1-\delta)K_{ht} + I_{ht} \tag{24}$$

where K_{ht} is the total stock of capital in country h, $0 < \delta_h < 1$ is the rate of depreciation of capital stock and I_{ht} is total investment in country h.

In addition to consumers and producers, each country is endowed with a government. The government consumes G_{ht}, taxes national output with a distortionary proportional income tax (τ_{ht}) and transfers back the remaining to domestic residents (TR_{ht}). Alternatively, the government can issue debt that will be repaid by increases in lump-sum taxes or decreases in transfers. The infinite horizon of this economy makes Ricardian equivalence hold: it is equivalent to finance government expenditure with taxes or with debt that will be compensated in the future with either more taxes or less transfers. The government flow budget constraint is given by

$$G_{ht} + TR_{ht} = \tau_{ht} Y_{ht} \tag{25}$$

which has to hold on a period by period basis. Government spending is assumed to follow an autoregressive stochastic process of the form

$$\begin{bmatrix} G_{1t} \\ G_{2t} \end{bmatrix} = \begin{bmatrix} \rho_{G1} & \nu_{G,12} \\ \nu_{G,21} & \rho_{G2} \end{bmatrix} \begin{bmatrix} G_{1t-1} \\ G_{2t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{G1t} \\ \epsilon_{G2t} \end{bmatrix}, \quad \epsilon_{Gt} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\epsilon_{G1}} & \psi_{G,12} \\ \psi_{G,21} & \sigma^2_{\epsilon_{G2}} \end{bmatrix} \right) \tag{26}$$

where ν_G controls for the spillover effect from government spending and $\psi_G \neq 0$ means that we allow for common fiscal policy shocks. ρ_{Gh} is the persistence parameter of the government spending process and $\sigma_{\epsilon_{Gh}}$ is the standard deviation of the innovation to the government spending process in country h.

There is frictionless international trade and capital markets are complete, which implies that individuals in the two countries can achieve both consumption smoothing (intertemporal transfer of consumption) and risk pooling (transfer of consumption across states of nature). The trade balance, or net exports, in country h is then given by $NX_{ht} = X_{ht} - M_{ht} = Y_{hjt} - Y_{jht}$.

Finally, the aggregate resources constraint for the traded goods in the world economy is

$$\Pi_1(C_{1t} + I_{1t} + G_{1t}) + \Pi_2(C_{2t} + I_{2t} + G_{2t}) = \Pi_1 V_{1t} + \Pi_2 V_{2t} \tag{27}$$

The equilibrium solution of the model can be obtained by deriving the sequences for the endogenous variables of the model that maximize (3.19) subject to (3.20)-(3.26) and given the initial endowment of capital. The complexity of the economic relations described in the model (highly nonlinear) causes, as in most RBC models, that an analytical solution which derives the endogenous variables in terms of the exogenous and the parameters cannot be obtained. We follow King, Plosser and Rebelo (1988) and use an Euler equation approximation approach that log-linearizes the set of first order conditions of the model, expressed in terms of "detrended" variables (all trending variables are transformed into ratios of the permanent technology change x_{ht}), around the steady state. Once the approximate solution is found, the parameters are given fixed values and solution paths for the variables of interest are simulated from the model drawing realizations from the exogenous processes' fixed distributions.

Parameter values are chosen from the IRBC literature, basically from Backus, Kehoe and Kydland (BKK) (1992)-(1993) and Ravn (1993). The representative agent's discount factor, β, is 0.9875; the coefficient of relative risk aversion, σ, is 2; the steady state share of time the household allocates to market activities, $\bar{H}$, is 20%.

On the production side of the economy, the output share of capital, α, is 0.36 and the quarterly depreciation rate, δ, is 0.025%. BKK (1992), Ravn (1993) and Reynolds (1993) have constructed several measures of aggregate Solow residuals for different measures of production factors and selected pairs of countries. They all obtain values of persistence (ρ_A) very close to one and positive correlations (ψ) across technology shocks to different countries. In addition, both BKK (1992) and Ravn (1993) find evidence of significant spillovers (ν_{hj}), in particular from the US to other countries, although Reynolds (1993) suggests that these spillovers may be quite low. We follow BKK (1992) whose estimate for the persistence parameter, ρ_A is 0.906, for the spillover parameter ν is 0.088 ($\nu_{12} = \nu_{21}$), for the standard deviation of the technology shock σ_ϵ is 0.00852 and for the cross-country correlation between shocks ψ is 0.258.

Parameters for the government sector are taken from Baxter and King (1993), Aiyagari, Christiano and Eichenbaum (1992) and King, Plosser and Rebelo (1988). We impose government budget balance in the steady state by assuming a constant tax rate (τ) equal to a constant government spending output share (sg) and no transfers. We have taken a value for τ and sg of 25%, which lays in between the one suggested by King, Plosser and Rebelo (1988) of 30% and that used by Baxter and King (1993) of 20% for the case of steady state balanced budget (Aiyagari, Christiano and Eichenbaum (1992) suggest a government spending share of 17.7%). We have imposed independent AR(1) processes for the two countries' government spending series (i.e. $\nu_{G,12}$=$\nu_{G,21}$=0 and $\psi_{G,12}$=$\psi_{G,21}$=0) and with equal parameter values for ρ_G and σ_G, taken from Aiyagari, Christiano and Eichenbaum (1992).

The steady state import share, MS is set to 15%, and the parameter governing the elasticity of substitution in the Armington aggregator, ρ, to 1.5. Finally, we assume that both countries have equal size, $\Pi = 1/2$.

By modifying certain key parameters which govern the international interdependencies included in the model, we derive four different model specifications:

(i) Autarky: No trade (MS=0), no spillovers ($\nu_{hj} = 0$) and uncorrelated shocks (no common shocks).

(ii) Trade Only: No spillovers ($\nu_{hj} = 0$) nor common shocks ($\psi = 0$) but trade in final goods and services is allowed between the two economies (MS $\neq$ 0).

(iii) Autarky with common shocks: No trade (MS=0), no spillovers ($\nu_{hj} = 0$) but contemporaneously correlated technology shocks ($\psi \neq 0$). There are common and country-specific shocks to closed economies.
(iv) Full interdependence: common and country-specific shocks transmitted through trade and through spillovers in the shock processes (ν_{hj}, MS and $\psi \neq 0$).

Table 4 summarizes the parameter vector under each model specification.

In order to characterize the international interdependencies generated by the four models, we perform the same multivariate spectral analysis of Figure 6 to the output series generated for the two countries under each model specification. Figures 7 and 8 plot the individual spectra, phases and coherences of the linearly detrended logs of the two output series simulated once under each of the four alternative model specifications. As with actual data, we obtain imprecise estimates of the output spectra (wide bands) under all specifications. However, it is clear that model spectra are of similar size for the two countries, contrary to what is observed in actual data (European GDP far more volatile than US one). All but the "Autarky with common shock" specification predict a lead of the GDP of one country over the other one, which we saw is not the case in actual data. However, the four models are able to capture the significant coherence (at 95% confidence level) between GDPs at business cycle frequencies, and predict values similar to the actual coherence. It is the "full interdependence" model the one displaying the higher coherence (significantly different from zero even at a 99% confidence level).

5.2 Assessment of the models

Assessment focusses on how well each model reproduces the dynamic relationship between both GDPs at business cycle frequencies. We have simulated the model 100 times (H=100) and estimated at each simulation the measure of distance $\hat{D}_h(\omega;\gamma)$ between the actual and simulated bi-variate spectral densities (for both countries' GDPs), where γ is the corresponding column in Table 4. The fit of the model is computed at each frequency based on the average estimated distance across simulations, $\hat{D}(\omega;\gamma)$.

We first apply the *fit* test to all four model specifications. The average fit across business cycle frequencies for each model is displayed in the diagonal of Table 5. The 90% and 95% critical values for the test statistics can be found in the bottom part of the table. They correspond to a χ^2 distribution with degrees of freedom 2^2*7, since there are 2 variables and the Andrews' optimal bandwidth estimator yields 7 frequencies in the business cycle frequency range. The standard IRBC two-country model with multiple goods has the best fit when common shocks are the only mechanism by which economies may move together (there are no spillovers across national shocks nor trade in final goods), but still is clearly rejected as the US-Europe DGP with respect to the comovements of the two countries' real GDPs at business cycle frequencies. Trade in final goods is clearly not the main mechanism by which fluctuations are transmitted among the US and European economies: our methodology assigns to the "trade only" model the worse fit. Figure 9 plots the *fit* test statistic for each model specification and for all frequencies. The horizontal lines are the 90% and 95% critical values of the one-frequency test (from a χ^2_4 distribution). It can be seen that all models are found similarly distant to the actual data and that it is the model using common shocks as the only explanation of international comovements the one reaching values of the fit test which get closer to the critical values at business cycle frequencies.

The off-diagonal figures in Table 5 are the $comp([\omega_1, \omega_2])$ test statistic applied to compare the performance (failure) of the four IRBC models two by two, for the business cycle frequency range, too. Their frequency by frequency values are plotted with the critical values in Figure 10. Any model is clearly rejected as equal to any other, with test statistics indicating in all cases that they are further to other models than to the actual data. Given that no model is accepted as equal to any other with respect to the comovements of the two countries' GDPs at business cycle frequencies, there is not much interpretation that can be done with respect to the *comp* test statistics. However, we observe that aggregating across business cycle frequencies suggests that the closer to each other are the "autarky" and the "full interdependence" models, while Figure 10 shows that the "full interdependence" model is closer to that including only "common shocks" at the lower frequencies. While acknowledging the caution needed as the previous remark suggests, these last observations could be interpreted in the following direction: higher frequency spectral density of GDPs may be generated by own country shocks (a feature shared by the "autarky" and "full interdependence" model specifications) while the lower frequency one may be more linked to common shocks (a feature shared between the "full interdependence" and the "common shocks" model specifications). The models more distant to each other are, logically, those with the best and worse fit: "common shocks" to "trade only" models.

This example borrowed from the IRBC literature allows us to illustrate how the *fit* test statistic proposed in this paper assesses the fit of a dynamic general equilibrium model and is able to produce a ranking of competing models. In particular, it has evidenced the importance of the existence of common shocks to explain the significant comovement observed between the US and European economies in 1970Q1-1993Q3 (high coherence between their real GDPs at business cycle frequencies). A simple comparison of the spectral density properties of the four models we have studied was able to conclude that all of them where quite insatisfactorily reproducing those of the actual data, but was not sufficient to discriminate across model specifications. Our model evaluation methodology confirms statistically the rejection of all models (fit statistics greater than their asymptotic critical value) and manages to identify that it is the one including common shocks as the only source of international comovements the one having least bad fit.

6 Summary and Conclusions

This paper develops a general formal framework for assessing the fit of multivariate dynamic models whose solution is approximated through simulation. The procedure is based on multivariate spectral techniques which are especially suitable for models that focus on a particular frequency range, such as business cycle models. The test we propose evaluates the distance between the spectral density matrices of the actual data and of data simulated from a model. Important features of the test are that it treats the sample of observed data and the simulated series symmetrically, and that it formally takes into account the fact that model series are simulated with an approximation error.

The necessary asymptotic theory is derived to test how well a simulated model reproduces the dynamic properties of actual data (*fit* test). Another asymptotic test is derived to compare the performance of alternative model specifications with respect to the multivariate vector of interest (*comparison* test). Monte Carlo evidence is provided showing the finite sample behavior of the tests. We find that both tests are very powerful against different alternatives even with small sample sizes. Sensitivity analysis shows the robustness of this result to alternative DGP specifications. However, the empirical distribution of the tests seems to be more concentrated around zero than the theoretical asymptotic distribution (which implies lower size in spite of the high power). Further research in terms of sensitivity of the methodology to the DGP structure and the sample size would clarify further this point. Confirming other studies, we also find that the small sample properties of our tests depend on the small sample characteristics of spectral estimators, in particular on the bandwidth parameter.

We have illustrated the use of the *fit* and *comparison* test statistics to assess a model by evaluating to which extent different versions of a two-country two-good International Real Business Cycle model can reproduce the interdependencies observed between the US and European real GDP at business cycle frequencies. Our model evaluation methodology confirms statistically the rejection of all models and manages to produce a ranking of competing models according to their fit, which could not be done in our case with simple inspection of the spectral densities of the actual and simulated data. The existence of common shocks is suggested as a relevant element to help explain the significant comovement observed between the US and European economies in 1970Q1-1993Q3 (high coherence between their real GDPs at business cycle frequencies).

References

[1] Andersen, T.G. and B.E. Sorensen (1995), "GMM Estimation of a Stochastic Volatility Model: a Monte Carlo Study", University of Copenhagen Discussion Paper 95-19.

[2] Andrews, D.W. (1991), "Heteroskedasticity and autocorrelation consistent covariance matrix estimation", *Econometrica*, vol.59, No.3, pp.817–858.

[3] Canova, F. (1994) " Statistical Inference in Calibrated Models", *Journal of Applied Econometrics*, 9, S123-S144.

[4] Canova, F. (1995) " Sensitivity Analysis and Model Evaluation in Simulated Dynamic General Equilibrium Economies", *International Economic Review*, 36, 477-501.

[5] Canova, F. and De Nicolò, G. (1995), " The Equity Premium and the Risk Free Rate: A Cross Country, Cross Maturity Examination", CEPR working paper 1119.

[6] Canova, F. and Ortega, E. (1996), "Testing Calibrated General Equilibrium Models", forthcoming in Mariano, R., Schuermann, T. and M. Weeks (eds.), *Simulation Based Inference in Econometrics: Methods and Applications*, Cambridge: Cambridge University Press.

[7] Cecchetti, S.G., Lam, P. and N. Mark (1993), "The Equity Premium and the Risk-free Rate: Matching the Moments", *Journal of Monetary Economics*, 31, 21-45.

[8] Christiano, L.J. and M. Eichenbaum (1992), " Current Business Cycle Theories and Aggregate Labor Market Fluctuations", *American Economic Review*, 82, 430-450.

[9] Christiano, L.J. and W. Den Haan (1995), "Small Sample Properties of GMM for Business Cycle Analysis", Federal Reserve Bank of Minneapolis Staff Reports 199.

[10] Cogley, T. and Nason, J.M. (1994), "Testing the Implications of Long-run Neutrality for Monetary Business Cycle Models", *Journal of Applied Econometrics*, 9, S37-S70.

[11] Diebold, F., Ohanian, L. and J. Berkowitz (1995), "Dynamic Equilibrium Economies: A Framework for Comparing Models and Data", NBER Technical Working Paper No.174.

[12] DeJong, D., Ingram, B. and C. Whiteman (1996), "A Bayesian Approach to Calibration", *Journal of Business and Economic Statistics*, 14, 1-9.

[13] Den Haan, W. and A. Marcet (1994), "Accuracy in Simulations", *Review of Economic Studies*, 61, 3-17.

[14] Fève, P. and Langot, F. (1994), "The RBC Models through Statistical Inference: An Application with French Data", *Journal of Applied Econometrics*, 9, S11-S37.

[15] Granger, C.W.J. (1964), *Spectral Analysis of Economic Time Series*, Princeton University Press.

[16] Gregory, A. and Smith, G. (1991), " Calibration as Testing: Inference in Simulated Macro Models", *Journal of Business and Economic Statistics*, 9(3), 293-303.

[17] Gregory, A. and Smith, G. (1993), " Calibration in Macroeconomics", in Maddala, G.S. (ed.), *Handbook of Statistics*, vol. 11, Amsterdam: North Holland.

[18] Hannan, E.J. (1970), *Multiple Time Series*, John Wiley and Sons.

[19] Kim, K. and Pagan, A.R. (1995), "The Econometric Analysis of Calibrated Macroeconomic Models", in Pesaran, H. and M. Wickens (eds.), *Handbook of Applied Econometrics*, John Wiley and Sons.

[20] King, R., Plosser, C. and S. Rebelo (1990), "Production, Growth and Business Cycles: Technical Appendix", mimeo, University of Rochester.

[21] Kydland, F. and Prescott, E. (1982), "Time to Build and Aggregate Fluctuations", *Econometrica*, 50, 1345-1370 .

[22] Kydland, F. and Prescott, E. (1991), "The Econometrics of the General Equilibrium Approach to Business Cycles", *The Scandinavian Journal of Economics*, 93(2), 161-178.

[23] Marcet, A. (1994), "Simulation Analysis of Stochastic Dynamic Models: Applications to Theory and Econometrics" in Sims, C. (ed.), *Advances in Econometrics. Sixth World Congress of the Econometric Society*, Cambridge: Cambridge University Press.

[24] Pagan, A. (1994), "Calibration and Econometric Research: An Overview", *Journal of Applied Econometrics*, 9, S1-S10.

[25] Priestley, M.B. (1981), *Spectral Analysis and Time Series*, Academic Press.

[26] Söderlind, P. (1994), "Cyclical Properties of a Real Business Cycle Model", *Journal of Applied Econometrics*, 9, S113-S122.

[27] Watson, M. (1993), "Measures of fit for calibrated models", *Journal of Political Economy*, vol.101, No.6.

7 Appendix 1: Asymptotic properties of spectral estimators.

Here we report a standard general theorem which determines the asymptotic distribution of spectral estimators and we discuss its properties. For more complete references see, for example, ch.IV and V in Hannan (1970), ch.6 and 9 in Priestley (1981), or Andrews (1991).

If the following conditions are satisfied:

(i) z_t is a zero mean multivariate random linear process satisfying

$$z_t = \sum_{\tau=-\infty}^{+\infty} A_\tau \epsilon_{t-\tau}, \tag{28}$$

where $\sum_{\tau=-\infty}^{+\infty} |A_\tau| < \infty$, and ϵ_t are iid processes with finite 4th order moments, and $\sum_{\tau=-\infty}^{+\infty} \|\sigma_{ij}(\tau)\| < \infty$, where $\sigma_{ij}(\tau)$ is the covariance for lag τ between z_{it} and z_{jt}, i,j = 1,..., N,

(ii) the lag window function $k_M(\tau)$ is a real valued continuous uniformly bounded function such that:

- $k_M(0) = 1$,
- $k_M(\tau) = k_M(-\tau)$, $\forall\tau$,
- $\int_{\tau=-\infty}^{+\infty} k_M^2(\tau) d\tau < \infty$ and,
- at each ω, the corresponding spectral window function is defined by

$$W_M(\omega) = \frac{1}{2\pi} \int_{\tau=-\infty}^{+\infty} k_M^2(\tau) e^{-i\tau\omega} d\tau \geq 0, \forall\omega. \tag{29}$$

(iii) the *lag/spectral window parameter* M is such that $M \to \infty$ as $T \to \infty$, and

(iv) M small relative to T, so that $M/\sqrt{T} \to 0$ as $M, T \to \infty$, i.e. M = $o(T^{1/2})$,

the spectral estimator:

$$\hat{f}_{ij}(\omega) = \frac{1}{2\pi} \sum_{\tau=-T+1}^{T-1} k_M(\tau) \hat{\sigma}_{ij}(\tau) e^{-i\omega\tau} \tag{30}$$

where $\hat{\sigma}_{ij}(\tau)$ is the covariance estimate for lag τ between z_i and z_j,

has the following properties:

(a)

$$\sqrt{\frac{T}{M}}\Big(\hat{f}_{ij}(\omega) - E[\hat{f}_{ij}(\omega)]\Big) \dot{\sim} \text{ Complex Normal,} \tag{31}$$

(b)

$$\lim_{T\to\infty} E[\hat{f}_{ij}(\omega)] = f_{ij}(\omega) \tag{32}$$

(c)

$$\lim_{T\to\infty}\Big(\frac{T}{M}cov\big[\hat{f}_{ij}(\omega_1), \hat{f}_{kl}(\omega_2)\big]\Big) = 0, \quad \omega_1 \neq \pm\omega_2, \tag{33}$$

(d)

$$\lim_{T\to\infty}\Big[\frac{T}{M}cov\big[\hat{f}_{ij}(\omega), \hat{f}_{kl}(\omega)\big]\Big) =$$

$$= \Big(\int_{\tau=-\infty}^{+\infty} k_M^2(\tau)d\tau\Big)f_{ik}(\omega)f_{jl}(\omega), \quad \omega \neq 0, \pm\pi$$

$$= 2\Big(\int_{\tau=-\infty}^{+\infty} k_M^2(\tau)d\tau\Big)\Big(f_{ik}(\omega)\overline{f_{jl}(\omega)} + f_{il}(\omega)f_{jk}(\omega)\Big), \quad \omega = 0, \pm\pi \tag{34}$$

where $\overline{f(\omega)}$ stands for the complex conjugate of $f(\omega)$.

Property (a) is the key to derive the **asymptotic distribution** of our measure of distance defined in (8). A standard result in spectral estimation is that the asymptotic distribution of $\nu\frac{\hat{f}_{ij}(\omega)}{f_{ij}(\omega)}$ can be approximated by χ^2_ν. ν is a constant called *equivalent degrees of freedom* of the spectral estimate and is defined as $\nu = \frac{2T}{M\int_{\tau=-\infty}^{+\infty} k_M^2(\tau)d\tau}$. The value of ν for each lag/spectral window function is tabulated (see Priestley (1981)), e.g. $\frac{5T}{3M}$ for the case of the Quadratic Spectral (QS) window, which is the one we will use in this paper.

However, for spectral estimates satisfying condition (iv), i.e. $M/T \to 0$ as $M, T \to \infty$, $\nu \to \infty$ and therefore the χ^2_ν distribution tends to a Normal distribution (see Priestley(1981), ch.6.4). Hence, $\hat{f}_{ij}(\omega)$ has an asymptotic Normal distribution, as in property (a).

Property (b) states the **asymptotic unbiasedness** of spectral estimates. This property holds because of assumption (iii): whichever the lag/spectral window function used, the asymptotic unbiasedness of spectral estimates is guaranteed as long as $M \to \infty$, or if M is a function of T (as it is generally the case, e.g. QS window) as long as $M \to \infty$ as $T \to \infty$.

Property (c) indicates the **independence of spectral estimates at different frequencies**. This property holds in general only when the separation between frequencies is greater than the bandwidth of the spectral window. In the case of the QS window, this requirement is fulfilled estimating the spectrum at frequencies distant $\frac{\pi}{M}$ to each other, which is the criterion we use in the Monte Carlo simulations of Section 3.4.

Property (d) derives the **asymptotic variance-covariance structure** of multivariate spectral estimates. The use of a spectral estimator with a lag/spectral window function as defined in assumption (ii) is introduced to overcome the asymptotic inconsistency of periodogram estimates. A lag/spectral window weights the sample covariances in the spectral estimator in (30) so as to reduce the contribution of distant lags/frequencies (and omits the lags/frequencies distant more than what the parameter M indicates), and thus the variance of spectral estimates is reduced. Assumption (iv), together with (iii), controls for the asymptotic properties of the lag/spectral window so that the variances and covariances of spectral estimates $\rightarrow 0$ as $T \rightarrow \infty$. These assumptions guarantee that property (d), together with (b), imply **consistency** of spectral estimates.

Regarding the condition required in assumption (iv), i.e. M $= o(T^{1/2})$, Andrews (1991) shows that optimal growth rates of M are typically less than $T^{1/2}$. He devices an automatic estimator for the **lag/spectral window parameter M,** $\hat{M}$, as a function of the sample size T and of the parametric DGP the z_t vector is supposed to follow, that is optimal under general conditions and is e.g. $O(T^{1/3})$ for the Bartlett window and $O(T^{1/5})$ for the Parzen, Tukey-Hanning and QS windows. His automatic estimator is used in the Monte Carlo experiments of section 3.4.

The **selection of the appropriate lag/spectral window function** is a controversial issue in the frequency domain literature. In this paper we want a lag/spectral window function that satisfies assumption (ii). Those conditions are satisfied by almost all standard "scale windows" $k(\frac{\tau}{M})$ used in practice, e.g. Bartlett, Parzen, Quadratic Spectral(QS) andTukey-Hanning windows.

Priestley (1981) shows that the best performing lag/spectral window, in terms of minimizing the relative Mean Square Error ($= \frac{variance^2(\omega)+bias^2(\omega)}{f^2(\omega)}$), is the Quadratic Spectral window. Andrews (1991) finds that, even when allowing for conditional heteroskedasticity and autocorrelation in the data process, the QS window is the best under a similar criterion. Therefore we use the QS window function in our simulation exercises: its functional form is (see Andrews(1991)):

$$k_{QS}(\frac{\tau}{M}) = \frac{25M^2}{12\pi^2\tau^2}\Big(\frac{sin(\frac{6\pi\tau}{5M})}{\frac{6\pi\tau}{5M}} - cos(\frac{6\pi\tau}{5M})\Big),\ \forall\tau. \tag{35}$$

Table 1: **Monte Carlo on the FIT TEST**

	Model 1 (size)			Model 2 (power)			Model 3 (power)		
	ω_1	ω_2	$[\omega_1,\omega_2]$	ω_1	ω_2	$[\omega_1,\omega_2]$	ω_1	ω_2	$[\omega_1,\omega_2]$
Theoretical size: 5%									
T=100	3.1%	3.6%	13.2%	10.8%	3%	19.1%	81.1%	6.4%	87.7%
T=200	2.5%	2.2%	4.3%	16%	1.8%	16.8%	99.5%	9.8%	99.5%
T=500	1.7%	1.7%	2.5%	45.6%	3.1%	34.2%	100%	34.1%	100%
Theoretical size: 10%									
T=100	13.1%	12.3%	24.4%	25.6%	11.2%	34.5%	94%	18.2%	94.3%
T=200	8.1%	6.9%	10.8%	30.1%	6.8%	27.5%	100%	28.6%	99.8%
T=500	4.9%	4.5%	6%	62.1%	6.9%	48.7%	100%	52.8%	100%

Notes. Actual data DGP is Model 1, a bivariate VAR(1) (see description in the text).

ω_1 (ω_2) is the frequency associated to cycles 8 years (2 years) long. $[\omega_1,\omega_2]$ aggregates the test statistics for all frequencies associated to cycles between 8 to 2 years long (Business Cycle frequencies).

The Monte Carlo standard deviation for these rejection frequency estimates is MCstd = $\sqrt{\frac{\alpha(1-\alpha)}{NREPL}}$, where α is the theoretical size and NREPL (=1000) the number of replications of the Monte Carlo experiment, i.e. MCstd = $\sqrt{\frac{0.05(1-0.05)}{1000}}$ = 0.69% for the first panel and MCstd = $\sqrt{\frac{0.1(1-0.1)}{1000}}$ = 0.95% for the second one.

Table 2: **Sensitivity of the FIT TEST to the parameter structure**

CaseI: No spillovers

	Model 2 (size)			Model 1 (power)			Model 3 (power)		
	ω_1	ω_2	$[\omega_1, \omega_2]$	ω_1	ω_2	$[\omega_1, \omega_2]$	ω_1	ω_2	$[\omega_1, \omega_2]$
Theoretical size: 5%									
T=100	17 2%	11.8%	28.1%	49%	9.9%	51.6%	99.7%	29.2%	99.7%
T=200	8.3%	5.5%	9.5%	59.8%	6.9%	50.9%	100%	49%	100%
T=500	3.7%	2 6%	4.6%	91.1%	5.2%	78.6%	100%	37.5%	100%
Theoretical size: 10%									
T=100	29.8%	23.5%	41.7%	64.9%	23.7%	63.2%	100%	45.5%	99.8%
T=200	16.4%	12.4%	17.1%	72.2%	15%	63.8%	100%	64.8%	100%
T=500	9.2%	6.7%	8.6%	95.8%	10.2%	86.7%	100%	54.6%	100%

CaseII: Larger spillovers

	actual DGP (size)			Model 2 (power)			Model 3 (power)		
	ω_1	ω_2	$[\omega_1, \omega_2]$	ω_1	ω_2	$[\omega_1, \omega_2]$	ω_1	ω_2	$[\omega_1, \omega_2]$
Theoretical size: 5%									
T=100	0.9%	0.7%	4.6%	14.4%	2.2%	30.3%	42.5%	3.6%	82.5%
T=200	0.7%	0.5%	1.7%	43 3%	1.8%	50.1%	85.4%	3.3%	99.3%
T=500	1.5%	1.1%	1%	96.3%	7.8%	97.3%	100%	21.3%	100%
Theoretical size: 10%									
T=100	5.5%	4.9%	12.2%	35.2%	8.5%	50.5%	71.3%	13.3%	93.1%
T=200	4.3%	2.7%	5%	67.4%	8.2%	65.1%	99.5%	12.8%	99.8%
T=500	3.6%	2.7%	2.4%	98.9%	17.2%	98.8%	100%	42.1%	100%

Notes: Actual data DGP is Model 2 in CaseI (No spillover: $\Phi^{\mathcal{Y}} = \begin{pmatrix} 0.7 & 0 \\ 0 & 0.6 \end{pmatrix}$) and a VAR(1) with parameter matrix $\Phi^{\mathcal{Y}} = \begin{pmatrix} 0.7 & 0.1 \\ 0.4 & 0.6 \end{pmatrix}$ in CaseII (same DGP was used to generate model series in the cases of the column called "Actual DGP").

ω_1 (ω_2) is the frequency associated to cycles 8 years (2 years) long. $[\omega_1, \omega_2]$ aggregates the test statistics for all frequencies associated to cycles between 8 to 2 years long (Business Cycle frequencies).

The Monte Carlo standard deviation for these rejection frequency estimates is MCstd $= \sqrt{\frac{\alpha(1-\alpha)}{NREPL}}$, where α is the theoretical size and NREPL (=1000) the number of replications of the Monte Carlo experiment, i.e. MCstd $= \sqrt{\frac{0.05(1-0.05)}{1000}} = 0.69\%$ for the first panel and MCstd $= \sqrt{\frac{0.1(1-0.1)}{1000}} = 0.95\%$ for the second one.

Table 3: **Monte Carlo on the COMPARISON TEST, H_0: $x^{m_i} = x^{m_j}$**

	Case11 (size)			Case33 (size, missp.)		
	ω_1	ω_2	$[\omega_1, \omega_2]$	ω_1	ω_2	$[\omega_1, \omega_2]$
Theoretical size: 5%						
T=100	4.6%	3.1%	11.6%	3.9%	4.4%	11.2%
T=200	2.6%	1.8%	3.4%	1.7%	2.2%	4.6%
T=500	2.1%	1.9%	2.1%	1.2%	1.2%	2.5%
Theoretical size: 10%						
T=100	13.8%	11%	23.2%	11.8%	12.9%	22.1%
T=200	7.7%	7.7%	9.3%	6.6%	7.6%	9.3%
T=500	4.5%	5.5%	5.1%	3.7%	5.1%	4.6%
	Case12 (power)			Case13 (power)		
	ω_1	ω_2	$[\omega_1, \omega_2]$	ω_1	ω_2	$[\omega_1, \omega_2]$
Theoretical size: 5%						
T=100	10.4%	2.6%	18.6%	79.4%	6.8%	88%
T=200	14.1%	3%	14.9%	99%	11.8%	99.1%
T=500	46.2%	2.5%	33%	100%	35.7%	100%
Theoretical size: 10%						
T=100	25.5%	10%	33.4%	94.3%	18.7%	94.7%
T=200	29.2%	7.9%	26.8%	100%	27.5%	99.9%
T=500	62.5%	6.3%	46.5%	100%	52.9%	100%

Notes: Case ij indicates that we are testing the null that Model i has the same spectral density matrix than Model j, i,j=1,2 and 3. Hence, Case11 and Case33 compute the size of the test under either correct specification of the two models (Case11: both are equal to the actual data DGP, Model 1) or misspecification (Case33: the two models compared are equal, generated with Model 3, but different to the actual data DGP, Model 1). Case12 and 13 compute the power of the test for two departures from the null.

ω_1 (ω_2) is the frequency associated to cycles 8 years (2 years) long. $[\omega_1, \omega_2]$ aggregates the test statistics for all frequencies associated to cycles 8 to 2 years long (Business Cycle frequencies).

The Monte Carlo standard deviation for these rejection frequency estimates is MCstd $= \sqrt{\frac{\alpha(1-\alpha)}{NREPL}}$, where α is the theoretical size and NREPL (=1000) the number of replications of the Monte Carlo experiment, i.e. MCstd $= \sqrt{\frac{0.05(1-0.05)}{1000}} = 0.69\%$ for the 5% size case and MCstd $= \sqrt{\frac{0.1(1-0.1)}{1000}} = 0.95\%$ for the 10% size case.

Table 4: **Parameter values for the IRBC models**

Parameter	Autarky	Trade only	Comm.shocks	Full Interdep.
Output Share of Labor (1-α)	0.64	0.64	0.64	0.64
Growth rate (θ_x)	1.004	1.004	1.004	1.004
Depreciation Rate (δ_K)	0.025	0.025	0.025	0.25
Discount Factor (β)	0.9875	0.9875	0.9875	0.9875
Steady State hours ($\bar{H}$)	0.20	0.20	0.20	0.20
Risk Aversion (σ)	2	2	2	2
Output Share of Gov.Sp. (sg)	0.25	0.25	0.25	0.25
Tax Rate (τ)	0.25	0.25	0.25	0.25
Persistence of Tech.Dist. (ρ_A)	0.9	0.9	0.9	0.9
Spillover across Tech.Dist. (ν_{hj})	0	0	0	0.088
Persistence of Gov.Sp.Dist. (ρ_G)	0.97	0.97	0.97	0.97
S.D. of Tech.shocks (σ_A)	0.00852	0.00852	0.00852	0.00852
Corr. of Tech.shocks (ψ)	0	0	0.258	0.258
S.D. of Gov.Sp.shocks (σ_G)	0.0036	0.0036	0.0036	0.0036
Imports share (MS)	0	0.15	0	0.15
Armington parameter (ρ)	1.5	1.5	1.5	1.5
Size of each country (Π)	0.5	0.5	0.5	0.5

Table 5: **Summary matrix of the fit at business cycle frequencies**

	Autarky	Trade only	Common shocks	Full Interdep.
Autarky	952.9			
Trade only	1612.8	1034.9		
Common shocks	1852.4	2700.6	772.4	
Full Interdep.	1015.2	1457.7	1187.9	932.6
CV 90%	37.9	37.9	37.9	37.9
CV 95%	41.3	41.3	41.3	41.3

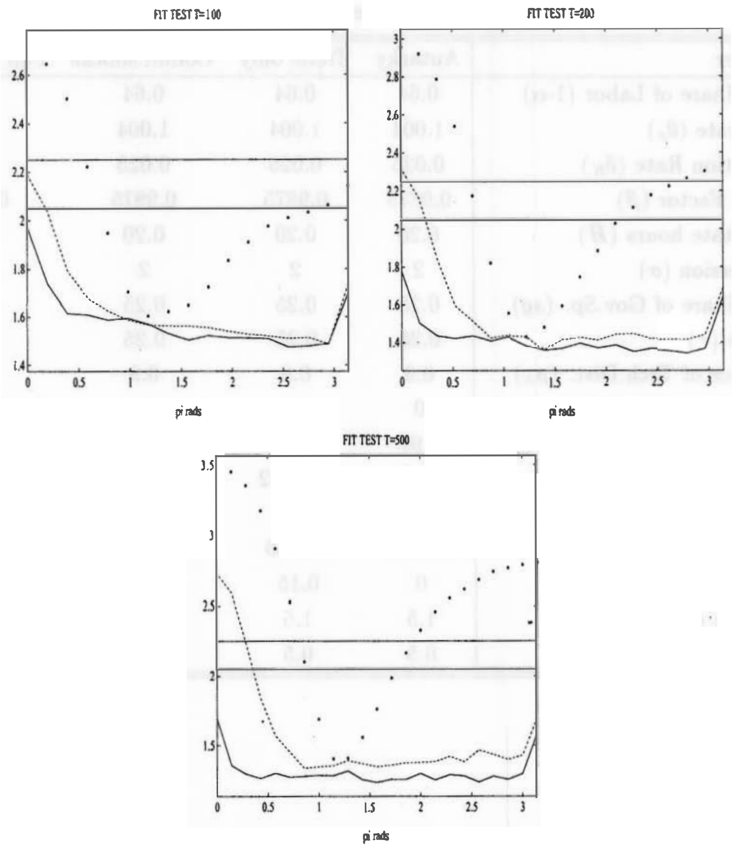

Figure 1: **Fit tests for different sample sizes**. Model 1 –, Model 2 - -, Model 3 *. The horizontal lines are the 90% and 95% critical values for the one-frequency test.

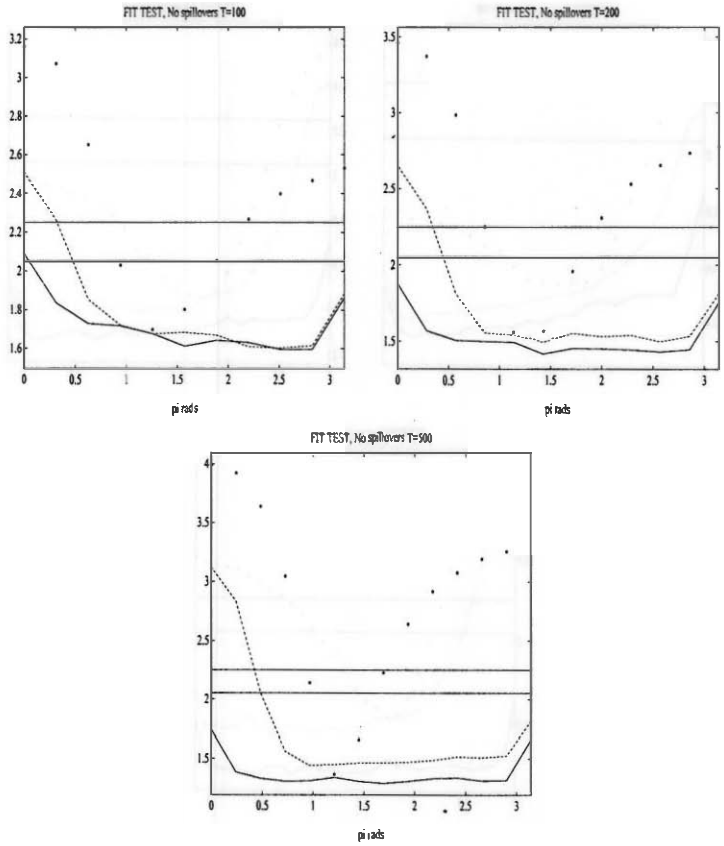

Figure 2: **Sensitivity analysis on the Fit test.** Actual data DGP with no spillover among variables (model 2). Model 2 –, Model 1 - -, Model 3 *. The horizontal lines are the 90% and 95% critical values for the one-frequency test.

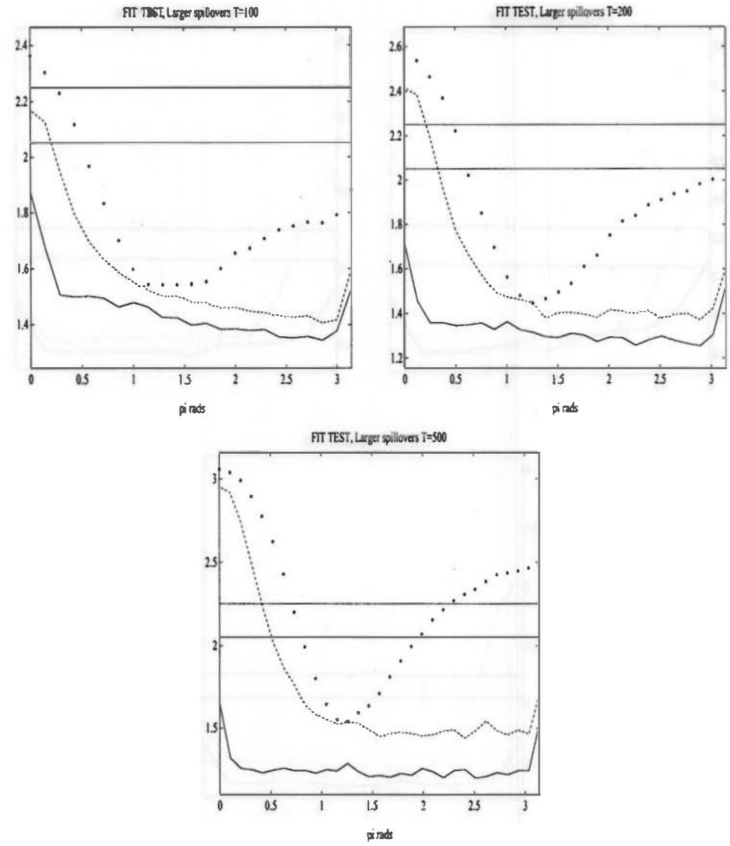

Figure 3: **Sensitivity analysis on the Fit test.** Actual data DGP with more spillover among variables. Model equal to actual data DGP –, Model 2 - -, Model 3 *. The horizontal lines are the 90% and 95% critical values for the one-frequency test.

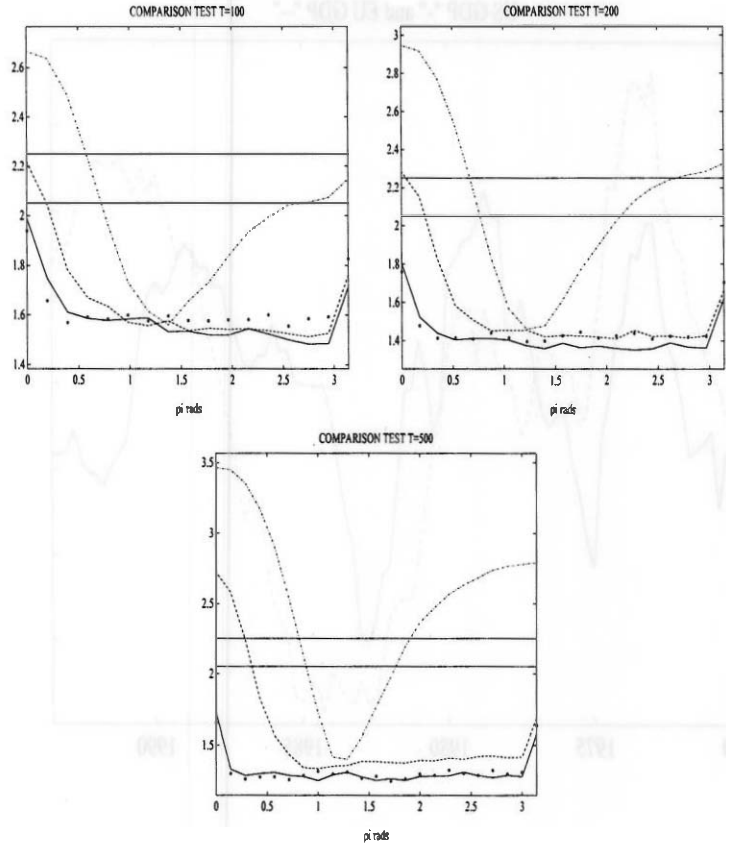

Figure 4: **Comparison tests for different sample sizes**. Case 11 –, case33 *, case 12 - -, case13 -.-. The horizontal lines are the 90% and 95% critical values for the one-frequency test.

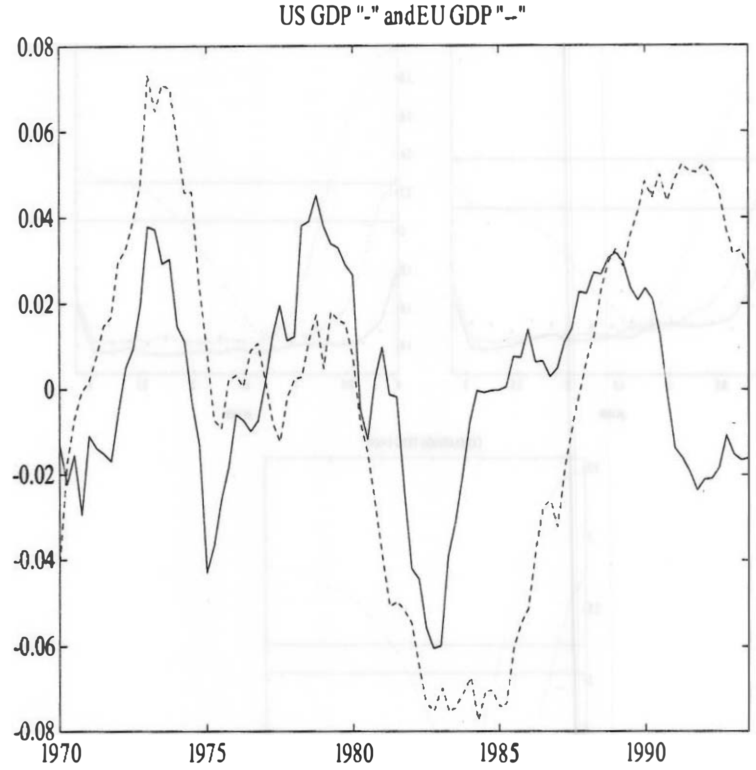

Figure 5: **US and European real GDPs, 1970Q1-1993Q3**. Linearly detrended logs of the series.

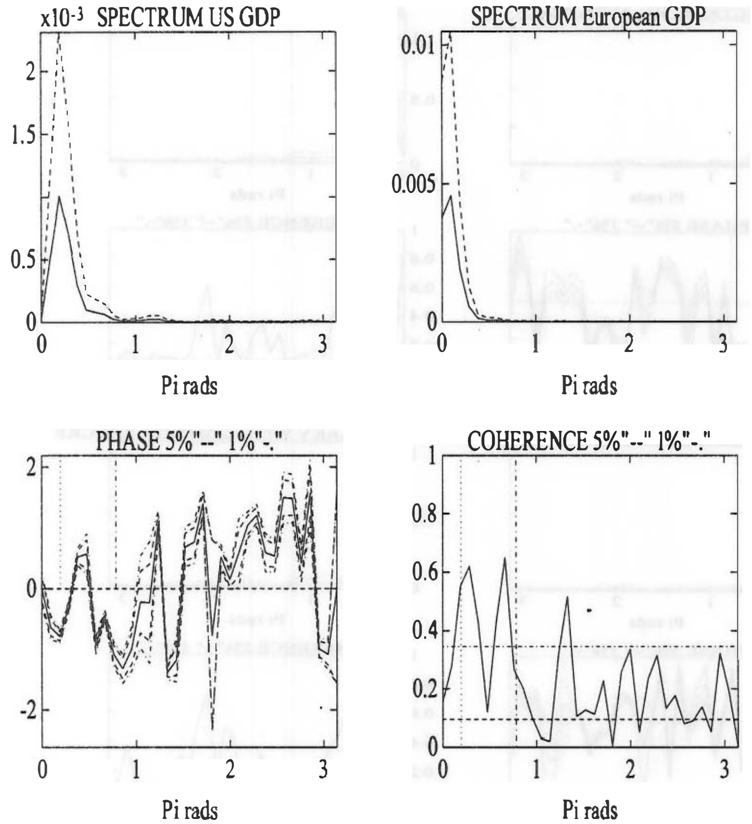

Figure 6: **Spectral properties of US and European real GDPs**. Individual spectra in the upper plots (with their 95% asymptotic confidence intervals). The lower plots display the phase (with its 95% and 99% asymptotic confidence intervals -5% and 1% significance levels-) and coherence (with the asymptotic critical values corresponding to the 5% and 1% significance levels, too), with the business cycle interval limited by the vertical lines.

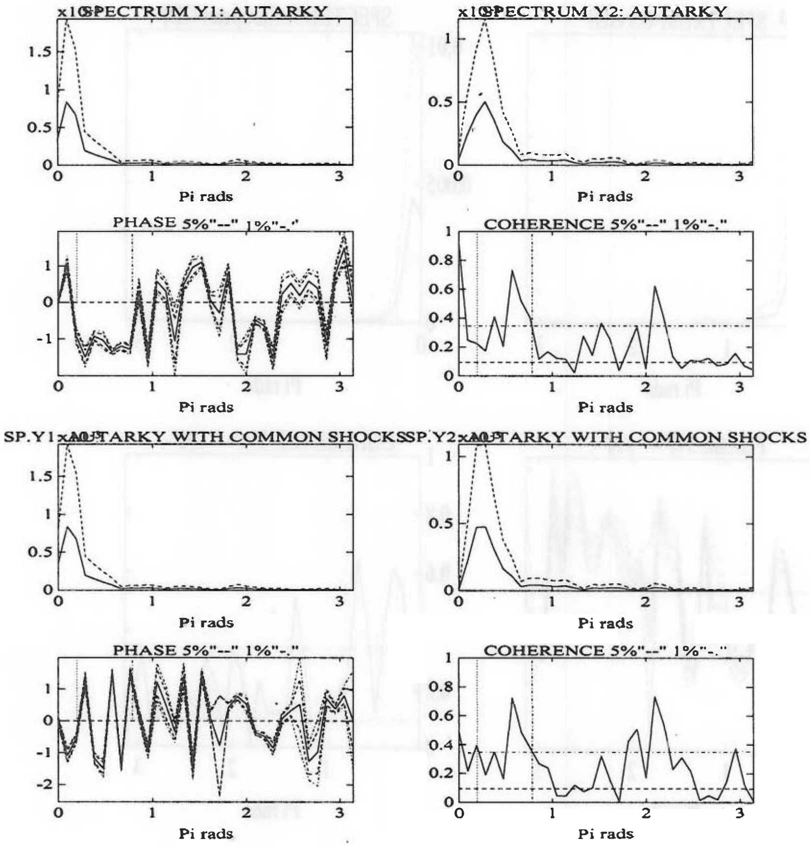

Figure 7: **Spectral properties of simulated output series for the two countries** under "Autarky" and "Autarky with common shocks" specifications of the two-country two-good International Real Business Cycle model.

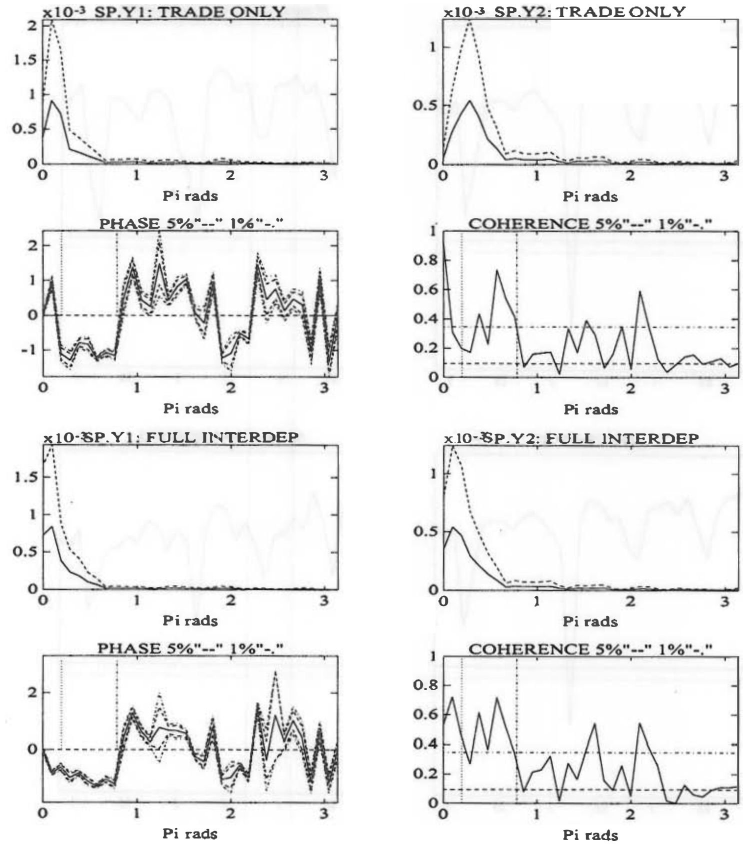

Figure 8: **Spectral properties of simulated output series for the two countries** under "Trade Only" and "Full Interdependence" specifications of the two-country two-good International Real Business Cycle model.

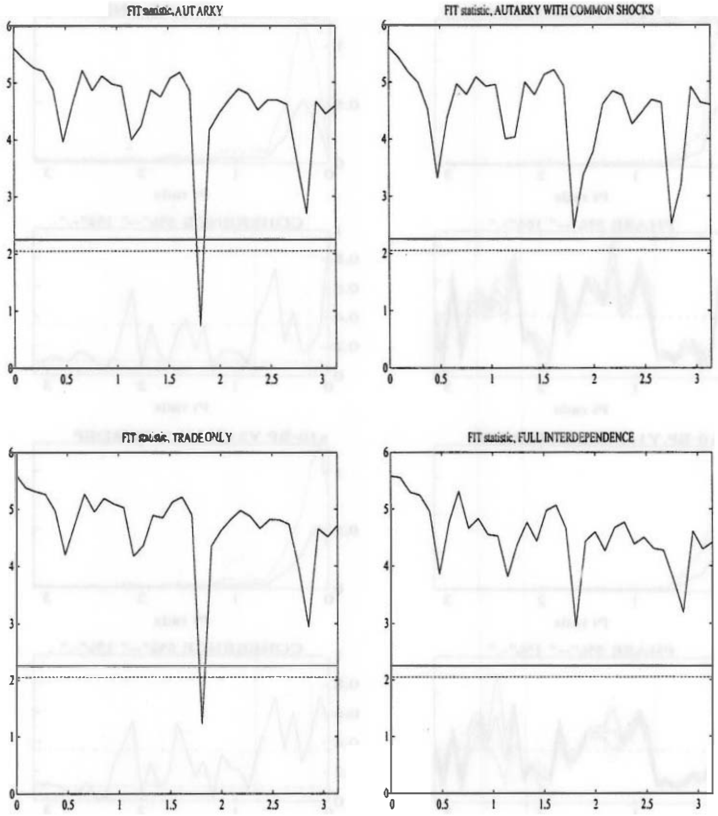

Figure 9: **Fit of the four IRBC models**. The horizontal lines are the 90% and 95% critical values for the one-frequency test.

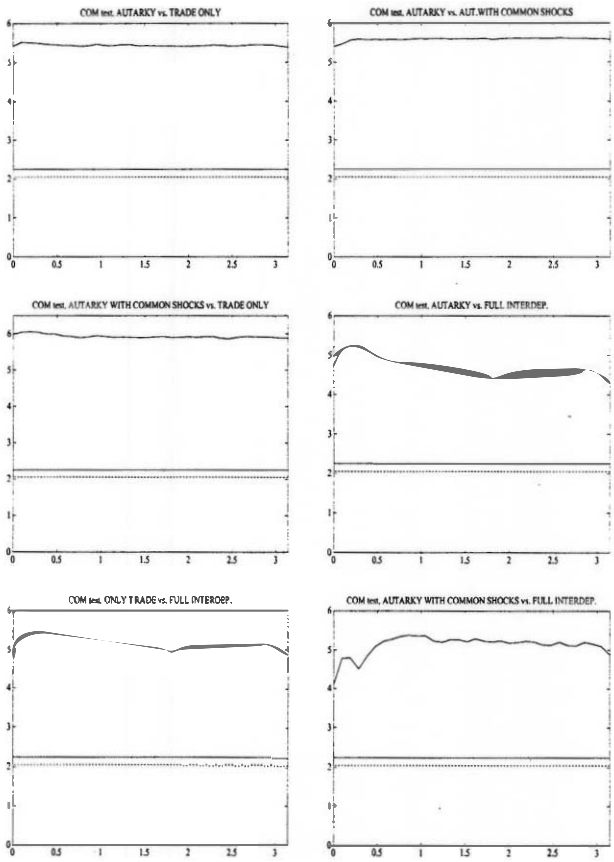

Figure 10: **Comparison of the four IRBC models two by two**. The horizontal lines are the 90% and 95% critical values for the one-frequency test.

WORKING PAPERS (1)

9525 **Aurora Alejano y Juan M.ª Peñalosa:** La integración financiera de la economía española: efectos sobre los mercados financieros y la política monetaria.

9526 **Ramón Gómez Salvador y Juan J. Doladó:** Creación y destrucción de empleo en España: un análisis descriptivo con datos de la CBBE.

9527 **Santiago Fernández de Lis y Javier Santillán:** Regímenes cambiarios e integración monetaria en Europa.

9528 **Gabriel Quirós:** Mercados financieros alemanes.

9529 **Juan Ayuso Huertas:** Is there a trade-off between exchange rate risk and interest rate risk? (The Spanish original of this publication has the same number.)

9530 **Fernando Restoy:** Determinantes de la curva de rendimientos: hipótesis expectacional y primas de riesgo.

9531 **Juan Ayuso and María Pérez Jurado:** Devaluations and depreciation expectations in the EMS.

9532 **Paul Schulstad and Ángel Serrat:** An Empirical Examination of a Multilateral Target Zone Model.

9601 **Juan Ayuso, Soledad Núñez and María Pérez-Jurado:** Volatility in Spanish financial markets: The recent experience.

9602 **Javier Andrés e Ignacio Hernando:** ¿Cómo afecta la inflación al crecimiento económico? Evidencia para los países de la OCDE.

9603 **Barbara Dluhosch:** On the fate of newcomers in the European Union: Lessons from the Spanish experience.

9604 **Santiago Fernández de Lis:** Classifications of Central Banks by Autonomy: A comparative analysis.

9605 **M.ª Cruz Manzano Frías y Sofía Galmés Belmonte:** Credit Institutions' Price Policies and Type of Customer: Impact on the Monetary Transmission Mechanism. (The Spanish original of this publication has the same number.)

9606 **Malte Krüger:** Speculation, Hedging and Intermediation in the Foreign Exchange Market.

9607 **Agustín Maravall:** Short-Term Analysis of Macroeconomic Time Series.

9608 **Agustín Maravall and Christophe Planas:** Estimation Error and the Specification of Unobserved Component Models.

9609 **Agustín Maravall:** Unobserved Components in Economic Time Series.

9610 **Matthew B. Canzoneri, Behzad Diba and Gwen Eudey:** Trends in European Productivity and Real Exchange Rates.

9611 **Francisco Alonso, Jorge Martínez Pagés y María Pérez Jurado:** Weighted Monetary Aggregates: an Empirical Approach. (The Spanish original of this publication has the same number.)

9612 **Agustín Maravall and Daniel Peña:** Missing Observations and Additive Outliers in Time Series Models.

9613 **Juan Ayuso and Juan L. Vega:** An empirical analysis of the peseta's exchange rate dynamics.

9614 **Juan Ayuso Huertas:** Un análisis empírico de los tipos de interés reales *ex-ante* en España.

9615 **Enrique Alberola Ila:** Optimal exchange rate targets and macroeconomic stabilization.

(1) Previously published Working Papers are listed in the Banco de España publications catalogue.

Queries should be addressed to: Banco de España
Sección de Publicaciones. Negociado de Distribución y Gestión
Telephone: 91 338 5180
Alcalá, 50. 28014 Madrid