UNIVERSITA' DEGLI STUDI ROMA TOR VERGATA
FACOLTA' DI ECONOMIA

PhD in Econometrics and Empirical Economics

Topics in Unobserved Components Models

Stefano Grassi

Supervisor:
prof.
Tommaso Proietti

Coordinator :
prof.
Franco Peracchi

XXII Cycle
Academic Year 2009 - 2010

Acknowledgments

This is the best opportunity I have to express formally my deep gratitude to Tommaso Proietti, my supervisor, for his constant guidance, suggestions, patience and more important to me our friendship. I really enjoyed our *time series* conversations and our non *time series* conversations. Without his help this work would have never been possible.

Then I would like to thank all the people in my department for their support and their friendship which made my studying in Rome the best period of my life. In particular I wish to thank Franco Peracchi for coordinating the Phd program.

I am thankful to several people for reading and commenting early drafts of these chapters. This group includes Alessandra Luati, Giuseppe Garofalo and my dear friend Raffaele Ciula.

Very special thanks to the VU University of Amsterdam for the warm hospitality during the period February - June 2009 and February 2010. Special thanks to Siem Jan Koopman for giving me the possibility to work with him and his group. I would like to thank Charles Bos, Michael Massman, Marius Ooms, Guilherme Moura, Drew Creal for the fruitful discussions about econometrics during lunch times and coffee breaks.

Finally I would like to thank my family an my girlfriend Lucia for the understanding during all these years and for silently accepting all my decisions.

Contents

Chapter 1

Introduction

This dissertation deals with structural time series models. Structural time series models refer to a class of parametric models that are specified directly in terms of unobserved components which capture essential features of the series, such as trend, cycle and seasonality. This dissertation is divided into three main parts: the first (chapter 2) deals with stochastic volatility models, the second (chapter 3) deals with dynamic factor models and the third (chapter 4) deals with Bayesian model selection. A general introduction to the state space methodology is provided in chapter 1.

1.1 The Linear State Space Form in Econometrics

The state space representation is the statistical framework for unobserved components models, which is made up of a *measurement equation*

$$y_t = Z_t\alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad t = 1, 2, \ldots, T, \tag{1.1}$$

and a *transition equation*

$$\alpha_{t+1} = T_t\alpha_t + R_t\eta_t, \quad \eta_t \sim N(0, Q_t). \tag{1.2}$$

The first relates the time series y_t to a $q \times 1$ vector of unobserved components or state vector, α_t, through Z_t that is an $1 \times q$ vector. The second is a dynamic linear model for the state α_t, taking the form of a first order vector autoregression. In this case α_t is a $q \times 1$ vector of unobserved states, T_t is a $q \times q$ matrix, R_t is a $q \times r$ matrix and η_t is a $r \times 1$ vector of random disturbances. The state space form of equation (1.1) and (1.2) can be easily extended to a multivariate setting, in which case y_t is a $N \times 1$ vector of time series, Z_t is a $N \times q$ matrix and ε_t is a $N \times 1$ vector of random disturbances, with $H_t = \mathrm{Var}(\varepsilon_t)$. Those matrices are related to a set of parameters, θ, which usually will have to be estimated, see section 1.6. We say that a state space model is *time invariant* if the system matrices are constant, that is $Z_t = Z$, $H_t = H$, $T_t = T$, and $R_t = R$, otherwise we call it *time varying* state space model.

In the analysis of economic time series α_t represents unobservable dynamic processes, such as seasonality or stochastic trends or core inflation and the NAIRU. In an aerospace mission, like the Apollo, the observation equation (1.1) describes radar observations y_t, disturbed by noise, on the state vector α_t (position, velocity, ...) of a spacecraft and the transition equation (1.2) is a linearized and discretized version of motion in space. In both cases given the observations y_t on-line estimation or *filtering* of α_t for $t = 1, 2, \ldots$, and prediction of α_s, $s > t$, are of primary interest.

Similar problems arose in on-line monitoring of patients (Smith and West, 1983) or in ecological processes (Frühwirth-Schnatter, 1994a).

The state space methodology offers many advantages with respect to autoregressive integrated moving average (ARIMA) models. They are very flexible and can easily deal with missing values and regression effects. In ARIMA framework the stochastic trends are removed in order to get a stationarity model, those trends have an economic meaning and they can be efficiently estimated using the Kalman filter. Finally they have an ARIMA representation, see Harvey (1989). The flexibility of this formulation created an active field of research that brought this methodology to the econometrics and economic literature. During the 70's and 80' the approach to the analysis of time series data has been dominated by the Box-Jenkins (1970) methodology. In time series analysis, the Box-Jenkins methodology, named after the statisticians George Box and Gwilym Jenkins, applies autoregressive moving average (ARMA) or ARIMA models to find the best fit to a given time series, in order to make, for example, forecasts. The original model uses an iterative three-stage modeling approach:

- *Model identification*: making sure that the variables are stationary using, for example, log first difference of the time series. Identifying seasonality in the dependent series (if any) and using a plot of the autocorrelation and partial autocorrelation functions to decide which autoregressive or moving average component should be used in the model.
- *Parameter estimation*: the most common methods use maximum likelihood estimation or nonlinear least-squares estimation.
- *Model checking*: testing if the estimated model conforms to the specifications of a stationary univariate process. In particular, the residuals should be independent with constant mean and variance.

One of the first contribution that showed the potentiality of state space models and the Kalman filter in this field was the paper of Harvey and Phillips (1979), where they use this formulation to calculate the likelihood of ARMA models. Engle and Watson (1981) showed how to use the Kalman filter to estimate the unobserved metropolitan wage rate for Los Angeles.

1.2 The Kalman filter

Theoretically the Kalman Filter, Kalman (1960) and Kalman and Bucy (1961), is a fundamental algorithm for the statistical treatment of state space models. Under the Gaussian assumption the Kalman filter delivers the minimum mean square estimator (MMSE) of the state vector conditional on past information $a_t = \mathrm{E}(\alpha_t|Y_{t-1})$ jointly with the mean square error matrix (MSE) give by $P_t = \mathrm{Var}(\alpha_t|Y_{t-1})$, where $Y_{t-1} = \{y_1, \ldots, y_{t-1}\}$.

Practically, it is certainly one of the greater discoveries in the history of statistical estimation theory and possibly the greatest discovery in the twentieth century. Stanley F. Schmidt is generally credited with developing the first implementation of the Kalman filter. It was during a visit of Kalman to the NASA Ames Research Center that he saw the applicability of his ideas to the problem of trajectory estimation for the Apollo program, leading to its incorporation in the Apollo navigation computer. Its most immediate applications have been for the control of complex dynamic systems such as continuous manufacturing processes, aircraft, ships, or spacecraft. For

these applications, it is not always possible or desirable to measure every variable that you want to control, and the Kalman filter provides a means for inferring the missing information from indirect (and noisy) measurements. The Kalman Filter is also used for predicting the likely future courses of dynamic systems that people are not likely to control, such as the flow of rivers during flood, the trajectories of celestial bodies, or the prices of traded commodities. The applications of Kalman filter encompasses many fields, but its use as a tool is almost exclusively for two purposes: *estimation* and *signal extraction*.

Starting from the state space formulation (1.1) and (1.2), the Kalman recursions can be written in the following form:

$$\begin{aligned} v_t &= y_t - Z_t a_t, & F_t &= Z_t P_t Z_t^{'} + \sigma_t^2, \\ K_t &= T_t P_t Z_t^{'} F_t^{-1}, & & \\ a_{t|t} &= T_t a_t + K_t v_t, & P_{t|t} &= P_t - K_t F_t K_t^{'}, \\ a_{t+1} &= T_t a_{t|t}, & P_{t+1} &= T_t P_{t|t} T_t^{'} + R_t Q_t R_t^{'}. \end{aligned} \tag{1.3}$$

The state space model and the Kalman filter recursion are completed by the specification of initial mean a_1 and variance P_1 concerning the distribution of α_1. This is an important issue when non-stationary components are present, see de Jong (1991). If the system is time invariant and α_t is stationary, the initial conditions are given by the unconditional mean and covariance matrix of the state vector, $\mathrm{E}(\alpha_1) = 0$ and $\mathrm{Var}(\alpha_1) = P$, satisfying the matrix equation $P = TPT^{'} + R_t Q_t R_t^{'}$.

The Kalman filter can easily handle missing values, in such case the vectors v_t and $a_{t|t}$ and the matrices K_t and F_t of equation (1.3) are set to missing for these values. Thus when the set of observations y_t, for $t = \upsilon, \ldots, \upsilon^* - 1$, are missing the Kalman updates become:

$$a_{t+1} = T_t a_t, \quad P_{t+1} = T_t P_t T_t^{'} + R_t Q_t R_t^{'}, \quad t = \upsilon, \ldots, \upsilon^* - 1. \tag{1.4}$$

1.3 Linear state space methods

Smoothing refers to the optimal estimation of unobserved components based on future observations. Define $p(\alpha_t|Y_l)$ as the marginal posterior distribution of the states at time t conditional on a sequence of data available until l. If $l < t$ then this process is known as *prediction*, and the Kalman filter can easily treat the difference $l - t$ as missing observations; if $l = t$ then it is commonly referred to as *filtering*; and if $l > t$ then one is conducting the process of smoothing. The smoothing problem is commonly segmented into three problems:

- *fixed interval smoothing*, where one is interested in calculating $p(\alpha_t|Y_T)$ for all time indices $t = 1, \ldots, T$;
- *fixed lag smoothing*, where one calculates $p(\alpha_t|Y_{t+L})$ where $L > 0$ is some fixed value and t varies;
- *fixed point smoothing*, where $p(\alpha_t|Y_L)$ is calculated for a fixed value t with $L > t$ increasing.

In contrast to filtering and prediction, different algorithms are available for smoothing. All of them involve at least an additional backward recursive pass through the data, after the forward pass of

the Kalman filter, required to store the quantities v_t, F_t, a_t and P_t for $t = 1, \ldots, T$. We report the state smoothing recursion that are carried out after the Kalman filter for $t = T, \ldots, 1$:

$$\begin{aligned} r_{t-1} &= Z_t' F_t^{-1} v_t + L_t' r_t, \qquad & N_{t-1} &= Z_t' F_t^{-1} Z_t + L_t N_t L_t, \\ \hat{\alpha}_t &= a_t + P_t r_{t-1}, & V_t &= P_t - P_t N_{t-1} P_t, \end{aligned} \tag{1.5}$$

where $L_t = T_t - K_t Z_t$, $r_T = 0$ and $N_T = 0$. The quantity $\hat{\alpha}$ represents the *smoothed estimate* and the matrix V_t is the associated mean square error matrix (MSE). It follows that the MSE matrix of the filtered estimator exceeds the MSE matrix of the smoothed estimator by a positive semidefinite matrix, see Harvey (1989).

For certain applications such as Bayesian analysis, we may require samples generated by a given state space model, a local level model for example, conditional on the observed time series $y_1, \ldots, y_t$. Simulation smoothing is a technique that provide this sample drawing the state variables (or innovations) in discrete time state space models from their conditional distribution given parameters and observations. Several methods for Gaussian simulation smoothing exist, most of which are based on the Kalman filter. Frühwirth-Schnatter (1994b) and Carter and Kohn (1994) independently developed methods for simulation smoothing based on the following decomposition:

$$p(\alpha_1, \ldots, \alpha_t | Y_T) = p(\alpha_t | Y_T) p(\alpha_{t-1} | Y_T, \alpha_t) \ldots p(\alpha_1 | Y_T, \alpha_2, \ldots, \alpha_t). \tag{1.6}$$

Other simulation smoothers have been proposed, for example de Jong and Shepard (1995) focus on sampling the disturbances and then the states. Durbin and Koopman (2002) proposed a new simulation smoother that is simpler then the other present in the literature.

1.4 Nonlinear and Non-Gaussian State Space Form in Econometrics

The Kalman filter is optimal in the important case when the equations are linear and the noises are independent, additive, and Gaussian as the model presented in (1.1) and (1.2). In this situation, the distributions of interest (filtering, predictive, or smoothing) are also Gaussian and the Kalman filter and smoother, as described in sections 1.2 and 1.3, can compute the conditional first and second moments exactly. Although it was originally derived for a linear problem, the Kalman filter is a basic algorithm for nonlinear problems. The linearity assumption is an important limitation in some contexts; a more general formulation is the nonlinear state space with non-additive disturbances, given by:

$$\begin{aligned} y_t &= G_t(\alpha_t, \varepsilon_t), \\ \alpha_t &= F_t(\alpha_{t-1}, \eta_t), \end{aligned} \tag{1.7}$$

where y_t is a vector of observations, α_t is the state vector, $G_t(\cdot)$ is a measurement function, $F_t(\cdot)$ is a system transition function, ε_t and η_t are noise vectors, and the subscript t denotes time index. The functions $G_t(\cdot)$ and $F_t(\cdot)$ can be nonlinear and the errors ε_t and η_t can be non-normally distributed. This formulation can be written in additive form by placing the non-additive disturbances in the state vector. In this framework the linear state space model is nothing more than a particular case.

The nonlinear state space models can be specified in the form of conditional distributions with the following transition and observation densities:

$$\begin{aligned} p(\alpha_{t+1}|\alpha_{1:t}, y_{1:t}) &= p(\alpha_{t+1}|\alpha_t, y_{1:t}) = p(\alpha_{t+1}|\alpha_t), \\ p(y_t|\alpha_{1:t}, y_{1:t-1}) &= p(y_t|\alpha_t, y_{1:t-1}) = p(y_t|\alpha_t). \end{aligned} \tag{1.8}$$

The notation $\alpha_{1:t}$ is used to denote a set of variables or observations between 1 and t, that is $\alpha_{1:t} = \{\alpha_1, \ldots, \alpha_t\}$. We use $p(\cdot)$ to represent a probability density of its arguments, thus $p(\cdot)$ can represent densities from different distributions in the same context. Finally $y_{1:t}$ denotes the information set up to time t, i.e. $y_{1:t} = \{y_1, \ldots, y_t\}$.

The oldest approach to handling nonlinear models is to use linearization and approximate Gaussianity. Schmidt (1966) introduced the idea of evaluating the partial derivatives at the filtered values of the state variables, this approach is generally called the *extended Kalman filter*. Unlike its linear counterpart, the *extended Kalman filter* in general is not an optimal estimator unless the measurement and the state equation are both linear, in this case is identical to the Kalman filter. Moreover in the *extended Kalman filter* the estimated covariance matrix tends to underestimate the true covariance matrix and therefore can become inconsistent without the addition of "stabilising noise". Finally when the transition and observation equation are highly nonlinear, the *extended Kalman filter* gives particularly poor performance because the mean and covariance are propagated through linearization of the underlying nonlinear model, see for a detailed discussion Orderud (2005).

The necessity to improve to the *extended Kalman filter* led to the development by Julier and Uhlmann (1997) of the *unscented Kalman filter* (UKF), also a nonlinear filter. The *unscented Kalman filter* uses a deterministic sampling technique known as the unscented transformation to pick a minimal set of sample points (called sigma points) around the mean. These sigma points are then propagated through the nonlinear function, from which the mean and covariance of the probability distributions are then recovered. The result is a filter which is more accurate than a first-order Taylor expansion of a nonlinear function as in the EKF.

Grewal ad Andrews (2001) underline that these *approximations* are still not adequate for all nonlinear problems. Thanks to the rapid development of computational power and Monte Carlo techniques, new computation-based methods such as particle filter have been developed for nonlinear and non-Gaussian state space models in the past decade. They are the sequential implementation of importance sampling and are often an alternative to the *extended Kalman filter* (EKF) or *unscented Kalman filter* (UKF) with the advantage that they can be made much more accurate.

1.5 Filtering Nonlinear and Non-Gaussian State Space Models: Sequential Importance Sampling

The basic idea of the Sequential Monte Carlo (SMC) methodology is the sequential sample of relevant probability distributions using the concept of importance sampling. In comparison with standard approximation methods, such as the popular *extended Kalman Filter*, they do not rely on any local linearization technique or any functional approximation, giving more precise results. The price to pay is that these methods are computationally very expensive.

The earliest applications of sequential Monte Carlo methods were in the area of growing polymers, see Hammersley and Morton (1954) and Rosenbluth and Rosenbluth (1956), and later they

expanded to other fields including physics and engineering. Sequential Monte Carlo methods found limited use in the past due to the very high computational complexity and the lack of adequate computing resources of that time. The fast advances of computer science during the 80's and the great potential of particle filters have made them recently a very active area of research. These methods are already used in real-time applications appearing in fields as diverse as chemical engineering, computer vision, financial econometrics, target tracking and robotics. Moreover these methods can be a powerful alternative to Markov chain Monte Carlo (MCMC) algorithms or they can be used to design very efficient MCMC schemes. Recent reviews and accounts of new developments on the subject can be found in, Doucet and Johansen (2008), Arulampalam et al. (2002), Doucet et al. (2001) and Doucet et al. (2000). The current interest in particle filtering for signal processing applications was brought on by Gordon et al. (1993) where they introduce the concept of resampling. A large portion of the theory on sequential signal processing is about the state and observation equations given in (1.7) that we report also here:

$$\begin{aligned} y_t &= G_t(\alpha_t, \varepsilon_t), \\ \alpha_t &= F_t(\alpha_{t-1}, \eta_t), \end{aligned} \tag{1.9}$$

where y_t is a vector of observations, α_t is a state vector, $G_t(\cdot)$ is a measurement function, $F_t(\cdot)$ is a system transition function, ε_t and η_t are noise vectors, and the subscript t denotes time index. The transition and the observation densities are given by

$$\begin{aligned} p(\alpha_{t+1}|\alpha_{1:t}, y_{1:t}) &= p(\alpha_{t+1}|\alpha_t, y_{1:t}) = p(\alpha_{t+1}|\alpha_t), \\ p(y_t|\alpha_{1:t}, y_{1:t-1}) &= p(y_t|\alpha_t, y_{1:t-1}) = p(y_t|\alpha_t). \end{aligned} \tag{1.10}$$

Equation (1.10) define a Bayesian model where $p(\alpha_{t+1}|\alpha_t)$ can be seen as the prior and $p(y_t|\alpha_t)$ as the likelihood, that is:

$$\begin{aligned} p(\alpha_{1:T}) &= \mu(\alpha_1) \prod_{k=2}^{T} f(\alpha_k|\alpha_{k-1}) \qquad \textit{Prior}, \\ p(y_{1:T}|\alpha_{1:T}) &= \prod_{k=1}^{T} g(y_k|\alpha_k) \qquad\qquad \textit{Likelihood}. \end{aligned} \tag{1.11}$$

In this context we can infer about $\alpha_{1:t}$ given a realization of the observations $y_{1:t}$:

$$p(\alpha_{1:t}|y_{1:t}) = \frac{p(\alpha_{1:t}, y_{1:t})}{p(y_{1:t})}, \tag{1.12}$$

where

$$\begin{aligned} p(\alpha_{1:t}, y_{1:t}) &= p(\alpha_{1:t}) p(y_{1:t}|\alpha_{1:t}), \\ p(y_{1:t}) &= \int p(\alpha_{1:t}, y_{1:t}) d\alpha_{1:t}, \end{aligned} \tag{1.13}$$

in general these distributions have no closed-form and numerical or simulation methods are necessary to evaluate them.

Given a sample $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_N\}$ generated from a density $\pi(\cdot)$, a generic problem of evaluating an integral

$$\mathrm{E}_\pi[h(\boldsymbol{\alpha})] = \int \pi(\alpha) h(\alpha) d\alpha, \tag{1.14}$$

is accomplished using the empirical average:

$$\bar{h}(\boldsymbol{\alpha}) = \frac{1}{N} \sum_{i=1}^{N} h(\alpha^{(i)}),$$

this is called Monte Carlo integration. The integral evaluation based on simulations from $\pi(\cdot)$ is not necessary optimal, see Theorem 3.12 in Robert and Casella (2004). The principal alternative, in order to evaluate (1.14), is not to direct sampling from $\pi(\cdot)$ but using importance sampling.

Importance sampling is a technique for approximating integrals of one probability distribution $\pi(\cdot)$ (target distribution) using a collection of samples from another distribution easy to simulate, $\mu(\cdot)$ (instrumental distribution). This method is based on the so called *importance sampling fundamental identity*:

$$E_\pi\left[h(\boldsymbol{\alpha})\right] = \int \pi(\alpha) h(\alpha) d\alpha = \int \mu(\alpha) \frac{\pi(\alpha)}{\mu(\alpha)} h(\alpha) d\alpha \tag{1.15}$$

where $w(\alpha) = \dfrac{\pi(\alpha)}{\mu(\alpha)}$ are the importance weights. As before using a sample from a given importance distribution $\mu(\cdot)$ the integral can be evaluated as follows:

$$\hat{h}(\boldsymbol{\alpha}) = \frac{1}{N} \sum_{i=1}^{N} w(\alpha^{(i)}) h(\alpha^{(i)}). \tag{1.16}$$

To obtain a weighted sample that resembles the target distribution $\pi(\cdot)$ it is essential to use a good importance distribution, $\mu(\cdot)$.

The sequential importance sampling (SIS) techniques, allows importance samples to be computed from a sequence of distributions defined on increasing spaces, this involves selecting an importance distribution which has the following structure:

$$\begin{aligned} \mu(\alpha_{1:t}) &= \mu_{t-1}(\alpha_{1:t-1}) \mu_t(\alpha_t | \alpha_{1:t-1}) \\ &= \mu_1(\alpha_1) \prod_{k=2}^{T} \mu_k(\alpha_k | \alpha_{1:k-1}), \end{aligned} \tag{1.17}$$

this means that to obtain at time t a collection of *particles* $\boldsymbol{\alpha}_t^{(i)} \sim \mu_t(\alpha_{1:t})$, where $i = 1, \ldots, N$, we sample $\boldsymbol{\alpha}_1^{(i)} \sim \mu_1(\alpha_1)$ and then recursively $\boldsymbol{\alpha}_t^{(i)} \sim \mu_t(\alpha_t | \boldsymbol{\alpha}_{1:t-1}^{(i)})$ at time $t = 2, \ldots, T$. The associated unnormalized weights are computed recursively using the following relation:

$$\begin{aligned} w_t(\alpha_{1:t}) &= \frac{\pi_t(\alpha_{1:t})}{\mu_t(\alpha_{1:t})} \\ &= \frac{\pi_{t-1}(\alpha_{1:t-1})}{\mu_{t-1}(\alpha_{1:t-1})} \frac{\pi_t(\alpha_{1:t})}{\pi_{t-1}(\alpha_{1:t-1}) \mu_t(\alpha_t | \alpha_{1:t-1})}, \end{aligned} \tag{1.18}$$

that can be rewritten more compactly:

$$\begin{aligned} w_t(\alpha_{1:t}) &= w_{t-1}(\alpha_{1:t-1})\gamma_t(\alpha_{1:t}) \\ &= w_1(\alpha_1)\prod_{k=2}^{T}\gamma_k(\alpha_{1:k}), \end{aligned} \tag{1.19}$$

where $\gamma_t(\alpha_{1:t})$ are the incremental importance weights give by the following relation:

$$\gamma(\alpha_{1:t}) = \frac{\pi_t(\alpha_{1:t})}{\pi_{t-1}(\alpha_{1:t-1})\mu_t(\alpha_t|\alpha_{1:t-1})}. \tag{1.20}$$

The SIS algorithm can be sketched as follows:

- At time $t = 1$:
  - Sample N particles from $\boldsymbol{\alpha}_1^{(i)} \sim \mu_1(\alpha_1)$;
  - Compute the weights $w_1(\boldsymbol{\alpha}_1^{(i)})$ and normalize them $W_1^{(i)} = \dfrac{w_1(\boldsymbol{\alpha}_1^{(i)})}{\sum_{j=1}^{N} w_1(\boldsymbol{\alpha}_1^{(j)})}$;
  - The filtering distribution at time 1 may be approximated by:

$$\hat{p}(\alpha_1) = \sum_{i=1}^{N} W_1^{(i)}\delta_{\boldsymbol{\alpha}_1^{(i)}}. \tag{1.21}$$

- At time $t > 2$ until $t = T$:
  - Sample N particles from $\boldsymbol{\alpha}_t^{(i)} \sim \mu_t(\alpha_t|\boldsymbol{\alpha}_{1:t-1}^{(i)})$;
  - Compute the incremental importance weights and normalize them:

$$\begin{aligned} w_t(\boldsymbol{\alpha}_{1:t}^{(i)}) &= w_{t-1}(\boldsymbol{\alpha}_{1:t-1}^{(i)})\gamma_t(\boldsymbol{\alpha}_{1:t}^{(i)}) && \textit{Incremental Weights}, \\ W_t^{(i)} &= \frac{w_t(\boldsymbol{\alpha}_{1:t}^{(i)})}{\sum_{j=1}^{N} w_t(\boldsymbol{\alpha}_{1:t}^{(j)})} && \textit{Normalized Weights}; \end{aligned} \tag{1.22}$$

  - The filtering distribution at time t may be approximated by:

$$\hat{p}(\alpha_t) = \sum_{i=1}^{N} W_t^{(i)}\delta_{\boldsymbol{\alpha}_t^{(i)}}. \tag{1.23}$$

The δ_x is the Dirac delta mass located at x.

The main problem of SIS algorithm is the phenomenon know as *weights degeneracy*, which brings to an increasing variance of the importance weights at every iteration. Consequently, the quality of the estimators will decrease as time progresses, see for example Kong et al. (1994). In other words the particles no longer provide an adequate description of the distributions of interest. This problem requires a mechanism for resetting the importance weights regularly to prevent this variance accumulation, concerning the filtering distribution, Gordon et al. (1993) developed the

resampling approach. The resampling step eliminates the particles with very low weights and replicates those with large weights, allowing just the *important* particles to contribute significantly to the approximation of the distributions of interest. A full account of resampling algorithms can be found in Liu (2001).

Denoting with $\left\{\frac{1}{N}, \bar{\boldsymbol{\alpha}}_t^{(i)}\right\}$ the collection of equally weighted resampled particles at time t, where the first element in parenthesis denotes the weight assigned to each particles and with $\boldsymbol{\alpha}_{t+1}^{(i)} \sim \mu_{t+1}(\alpha_{t+1}|\bar{\boldsymbol{\alpha}}_{1:t}^{(i)})$ the sampled particle at time $t+1$. Then $\left(\bar{\boldsymbol{\alpha}}_{1:t}^{(i)}, \boldsymbol{\alpha}_{t+1}^{(i)}\right)$ is approximately distributed according to $p(\alpha_{1:t})\mu_{t+1}(\alpha_{t+1}|\alpha_{1:t})$ where $p(\cdot)$ is the filtered distribution. A generic SIS algorithm can be sketched as follows:

- At time $t = 1$:
  - Sample N particles from $\boldsymbol{\alpha}_1^{(i)} \sim \mu_1(\alpha_1)$;
  - Compute the weights $w_1(\boldsymbol{\alpha}_1^{(i)})$ and normalize them $W_1^{(i)} = \dfrac{w_1(\boldsymbol{\alpha}_1^{(i)})}{\sum_{j=1}^N w_1(\boldsymbol{\alpha}_1^{(j)})}$;
  - Resample $\{W_1^{(i)}, \boldsymbol{\alpha}_1^{(i)}\}$ to obtain N equally-weighted particles $\left\{\frac{1}{N}, \bar{\boldsymbol{\alpha}}_1^{(i)}\right\}$;
  - The filtering distribution at time 1 may be approximated by:

$$\hat{p}(\alpha_1) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\bar{\boldsymbol{\alpha}}_1^{(i)}}. \tag{1.24}$$

- At time $t > 2$ until $t = T$:
  - Sample N particles from $\boldsymbol{\alpha}_t^{(i)} \sim \mu_t(\alpha_t|\bar{\boldsymbol{\alpha}}_{1:t-1}^{(i)})$ and set $\boldsymbol{\alpha}_{1:t}^{(i)} \leftarrow \left(\bar{\boldsymbol{\alpha}}_{1:t-1}^{(i)}, \boldsymbol{\alpha}_t^{(i)}\right)$;
  - Compute the weights and normalize them:

$$\begin{aligned} w_t(\boldsymbol{\alpha}_{1:t}^{(i)}) &= w_{t-1}(\boldsymbol{\alpha}_{1:t-1}^{(i)})\gamma_t(\boldsymbol{\alpha}_{1:t}^{(i)}) && \textit{Incremental Weights}, \\ W_t^{(i)} &= \frac{w_t(\boldsymbol{\alpha}_{1:t}^{(i)})}{\sum_{j=1}^N w_t(\boldsymbol{\alpha}_{1:t}^{(j)})} && \textit{Normalized Weights}; \end{aligned} \tag{1.25}$$

  - Resample $\{W_t^{(i)}, \boldsymbol{\alpha}_t^{(i)}\}$ to obtain N equally-weighted particles $\left\{\frac{1}{N}, \bar{\boldsymbol{\alpha}}_t^{(i)}\right\}$;
  - The filtering distribution at time t may be approximated by:

$$\hat{p}(\alpha_t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\bar{\boldsymbol{\alpha}}_t^{(i)}}. \tag{1.26}$$

Resampling the particles is not a problem provided that the most reasonable number of distinct particles are retained and that they lead to a good degree of sample diversity. The resampling procedure has negative effects on the Monte Carlo estimator because it increases its variance, to avoid this problem it is more sensible to resample only if strictly necessary. For this reason in the literature, it has been proposed to resample when the variance of the unnormalized weights is superior to a pre-specified threshold, this is assessed by looking to the Effective Sample Size (ESS), see Liu (2001) or Doucet and Johansen (2008) for a full account.

1.6 Estimation strategies for State Space models

All the system matrices, Z_t, H_t, T_t and R_t of a linear state space form, see equations (1.1) and (1.2) are related to a set of parameters θ. If these parameters are known then the application of the Kalman filter to the state space form (1.1) and (1.2) will deliver the optimal estimation of the unknown states. Unfortunately, in almost all real situations the parameters are unknown and have to be estimated. To estimate these parameters two main approaches are available, the Maximum likelihood estimation and the Bayesian approach. Section 1.6.1 introduce the maximum likelihood and provides an example of this estimation strategy for a linear state space model. Section 1.6.2 introduce the Markov Chain Monte Carlo and provides an example of this estimation strategy for a nonlinear state space model.

1.6.1 Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) is a popular statistical method used for fitting a statistical model to data, providing estimates for the model parameters. It searches the parameters values that make the data "more likely" than any other parameters values. The likelihood can be calculated by a routine application of the Kalman filter, even if the state vector is fully or partially diffuse. More precisely the likelihood function can be written in prediction error decomposition form, see Schweppe (1965), and apart from a constant term the log-likelihood is:

$$l(y_1, \ldots, y_T; \theta) = \sum_{t=1}^{T} l(y_t | Y_{t-1}; \theta) = -\frac{1}{2} \left(\sum_{t=1}^{T} \log |F_t| + \sum_{t=1}^{T} v_t' F_t^{-1} v_t \right), \tag{1.27}$$

where all the quantities are given in output by the Kalman filter. This log-likelihood can be maximized by means of iterative numerical procedures based on Newton's method, see Nocedal and Wright (1999) for an extensive treatment. Those numerical procedures require the evaluation of the first-order and the second-order partial derivatives called, respectively, score vector and the Hessian matrix. It turns out that the score vector is a fundamental quantity in the maximization steps. In case of small parameter dimension it is efficiently computed using numerical approximations. The latter are not feasible in the case of a large dimensional parameter vector, in these situations the derivation of the analytical score is compulsory, see Chapter 3.

A useful algorithm for Maximum Likelihood estimation, particularly in the early stages of the maximization procedure, is Expectation Maximization (EM) algorithm, introduced for general latent variable models in the seminal paper by Dempster and Laird (1977). The earlier EM methods for the state space mode were developed by Shumway and Stoffer (1982) and Watson and Engle (1983). The EM is an iterative method which alternates between performing an expectation (E) step by means of the Kalman filter and smoother, which computes the conditional expectation of the complete log-likelihood with respect to the posterior of the states α_t, and a maximization (M) step, which computes the parameters which maximize the complete log-likelihood computed by the E step. These parameters are then passed to the next E step. This procedures continues until the difference between the log-likelihood at step $i-1$ and i is less then a fixed tolerance value. Under fairly mild regularity conditions the EM algorithm converges to a local maximum of the log-likelihood function. Despite a fast convergence in the early stage of the maximization process, the EM algorithm then becomes much slower in convergence compared to direct maximization of the log-likelihood function.

Illustrative example: Dynamic Factor Models

We provide here a simple example that shows the usefulness of the Kalman filter and the state space methodology to represent and estimate dynamic factor models, that will be deeply discussed in Chapter 3.

A widely used method to analyze large quantities of data in the social sciences is factor analysis developed by psychologists for measurement of intelligence in the beginning of the twentieth century, see Burt (1909). In dynamic factor analysis large number of observed variables are described by few unobserved processes, called *common factors*. In recent years large dimensional factor models have become more and more popular in econometric and economic research too, for an extensive overview see Bai and Ng (2008).

The dynamic factor model can be cast in state space form and apply the Kalman filter for estimation and prediction and the smoother for extracting the latent factors. This approach has been around for some time (see Molenaar, 1985), but has been reintroduced by Kapetanios and Marcellino (2006), Doz et al. (2007) and Jungbacker and Koopman (2008). To convert the dynamic factor model into state space form one needs to know the number of factors, q and the lag order. It is assumed through that they are known a priory.

Consider a panel of N time series where we denote $y_{i,t}$ as the observation at time t in the i series, then the dynamic factor model is given by:

$$y_{i,t} = \Lambda_i^{'} f_t + \varepsilon_{i,t}, \qquad t = 1, \ldots, T, \qquad i = 1, \ldots, N, \tag{1.28}$$

where Λ_i is a $q \times 1$ vector of factor loadings, f_t is a $q \times 1$ vector of unobserved common factors, $\varepsilon_{i,t}$ is the *individual-specific noise*, T is the number of observations and N is the number of series.

The dynamic factor model given in (1.28) can be represented in matrix form as follow:

$$y_t = \Lambda f_t + \varepsilon_t, \quad t = 1, \ldots, T, \tag{1.29}$$

where $\Lambda = (\lambda_1, \ldots, \lambda_N)^{'}$ is a matrix of factor loadings, f_t is a $q \times 1$ vector of unknown factors. The common factors f_t are modelled as a stationary first order vector autoregressive ($VAR(1)$) process, and the error components ε_t are assumed to be multivariate Normal with $N \times N$ diagonal variance covariance matrix Σ_ε. A more general version of the factor model, that allows for some dynamics in the *individual-specific noise*, or some exogenous variables can be considered. The model (1.29) can be represented in linear state space form:

$$\begin{aligned} y_t &= Z\alpha_t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \Sigma_\varepsilon), \\ \alpha_{t+1} &= T\alpha_t + \eta_t \qquad \eta_t \sim N(0, \Sigma_\eta). \end{aligned} \tag{1.30}$$

To get the representation (1.30) consider that f_t can be rewritten as a linear combination of the unobserved states α_t using a suitable full rank selection matrix G, in particular:

$$f_t = G\alpha_t, \tag{1.31}$$

where α_t has the following representation:

$$\alpha_{t+1} = T\alpha_t + \eta_t, \qquad \eta_t \sim N(0, \Sigma_\eta), \tag{1.32}$$

and Σ_η is the variance covariance matrix of the states. In the case where the factors and the latent states follow a $VAR(1)$ process, the selection matrix G is just a unitary matrix. Finally

the Z matrix is given by the combination of the Λ and G, indeed $Z = \Lambda G$. The log-likelihood function, $l(y_1, \ldots, y_t; \theta)$, of model (1.30) can be evaluated via prediction error decomposition as in formula (1.27). The prediction error vector v_t and its variance matrix F_t are obtained from the Kalman filter (1.3) applied to the linear state space model (1.30). Then a Quasi-Newton BFGS algorithm can be adopted to maximize the log-likelihood function with respect to the parameter vector θ. Usually in dynamic factor analysis the dimension of θ is very high and computing the score vector numerically is not feasible. Fortunately the exact score vector for this class of model has been provided by Jungbacker and Koopman (2008). In the presence of a large amount of data and factors, the Kalman filter and smoother slow down enormously and are prone to numerical inaccuracy. Jungbacker and Koopman (2008) presented a new result that lead to a computationally efficient procedure for the estimation of the factors and the parameters by maximum likelihood.

1.6.2 Bayesian Estimation

This section provides a description of the main estimation strategies for the nonlinear and non-Gaussian models introduced in section 1.4. For this class of models purely analytical maximization technique is not available and the estimation is based on simulations. We have seen in section 1.6.1 that the Kalman filter is fundamental for maximum likelihood estimation, in the case of nonlinear and non-Gaussian state space models it would be natural to consider the filtering technique presented in section 1.5 for evaluating the likelihood. Unfortunately this approach is not feasible because the sequential importance sampling technique approximate, using a discrete set of points, the filtering distribution, bringing to an unsmooth likelihood that cannot be maximized using Newton's methods. To overcome this problem a lot of different solutions have been proposed in the literature: Andrieu et al. (2005) focused to an on-line EM algorithm, Chen and Liu (2000) proposed a mixture Kalman filter, Johansen et al. (2006) used sequential Monte Carlo methods, Doucet and Tadic (2003) proposed to approximate the numerical derivative with a suitable set of recursions. Despite those and many other contributions, we still do not have general and stable procedure to maximize the likelihood for these models. The Markov Chain Monte Carlo methods (MCMC) are a valuable alternative to estimate these models.

In Bayesian analysis we are interested in the joint posterior of the parameters and latent processes:

$$p(\alpha, \theta | Y_T) \propto p(\alpha) p(\theta) f(Y_T | \alpha, \theta)$$

where α are the unknown states, θ is a set of parameters, $p(\alpha)$ and $p(\theta)$ are the corresponding priors and $f(Y_T|\alpha, \theta)$ is the likelihood. In this case there is no conjugacy and no closed-form expression, therefore this distribution cannot be simulated. The essence of MCMC methods is to simulate from an ergodic Markov Chain that has the target distribution as the limiting distribution. In a complex case like the one presented before it is useful to split it into a set of easier problems where we can use the conjugacy. Proceeding in this way we can estimate for example sequentially the following quantities:

$$p(\alpha | \theta, Y_T)$$
$$p(\theta | \alpha, Y_T)$$

where it is possible to have closed-form solutions. This way to proceed is very similar to the Gibbs Sampler, more on this in the following. The MCMC methods have almost unlimited applicability,

even though their performances varies widely, depending on the complexity of the problem. They are mainly used to estimate nonlinear and non-Gaussian state space models like the stochastic volatility models, see Chapter 2.

MCMC methods, are algorithms that starting from a probability distribution (instrumental distribution) and using a Markov chain gives the desired distribution (target distribution) in output. The state of the chain after a large number of steps converges to the so called equilibrium distributions that represent the target distributions. The quality of the chain improves as a function of the number of steps, see Robert and Casella (2004). The original Metropolis algorithm was introduced by Metropolis et al. (1953) in a set up of optimization on a discrete state space, in connection with particle physics: the paper was published in the *Journal of Chemical Physics*. All the authors of this seminal paper were involved in the Los Alamos research laboratory during and after the second World War, the main research in that period was to evaluate the behavior of nuclear explosives. After the second World War new computational capabilities become available thanks to the work of von Neumann. Metropolis and Stanislaw Ulam can be considered the fathers of Monte Carlo methods, they ran the first Monte Carlo experiment on the MANIAC (*Mathematical Analyzer, Numerical Integration and Computer*) in 1948.

The Metropolis algorithm was later generalized by Hasting (1970) to statistical simulation. Although there are a large amount of papers that highlighted the usefulness of this algorithm before 1990, the starting point of an intensive use of Markov Chain Monte Carlo methods; by statistical community can be traced to the presentation of the *Gibbs Sampler* by Gelman and Smith (1990).

The Gibbs sampling or Gibbs sampler is an algorithm to generate a sequence of samples from the *joint* probability distribution of two or more random variables. Gibbs sampling is an example of a Markov chain Monte Carlo algorithm and is useful when the joint distribution is not known explicitly, but the conditional distribution of each variable is known. In this case a computationally difficult problem can be splitted in more parts that are easier to treat. Suppose that the parameters of interest are $\theta = (\theta_1, \ldots, \theta_p)$, the joint posterior distribution of θ is denoted by $(\theta|Y_T)$, where Y_T are our observations. The joint posterior $(\theta|Y_T)$ may be of high dimension and difficult to summarize. Suppose that is possible to define a set of conditional distributions given by:

$$\begin{aligned} &(\theta_1|\theta_2, \ldots, \theta_p, Y_T), \\ &(\theta_2|\theta_1, \ldots, \theta_p, Y_T), \\ &\ldots \\ &(\theta_p|\theta_1, \ldots, \theta_{p-1}, Y_T), \end{aligned} \tag{1.33}$$

where $(\theta_1|\theta_2, \ldots, \theta_p, Y_T)$ defines the conditional distribution of θ_1 given $(\theta_2, \ldots, \theta_p)$.
Formula (1.33) shows how the Gibbs sampler works. The Gibbs sampling algorithm generates an instance from the distribution of each variable in turn, conditional on the current values of the other variables, this is called one cycle of Gibbs sampling. The general algorithm with two variables can be sketched as follows:

(1) At iteration 0 choose a starting value the variables $\theta_1^{(0)}$ and $\theta_2^{(0)}$;

(2) At iteration j sample from the conditional distributions:

$$\begin{aligned} &(\theta_1^{(j)}|\theta_2^{(j-1)}, Y_T), \\ &(\theta_2^{(j)}|\theta_1^{(j)}, Y_T); \end{aligned} \tag{1.34}$$

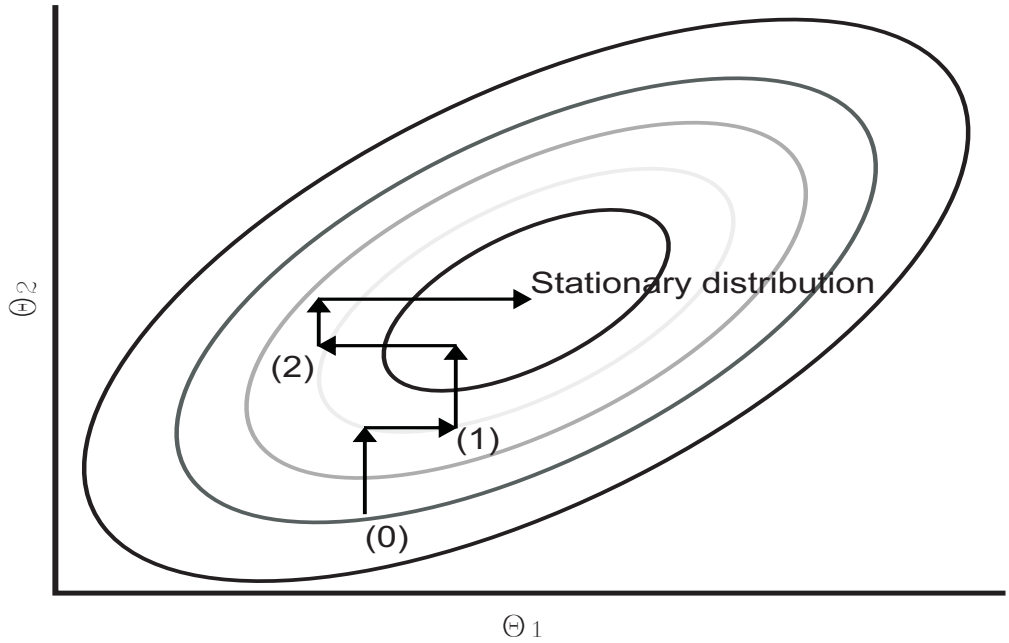


Figure 1.1: *The Gibbs sampling iterations for the parameters θ_1 and θ_2*

(3) Go to step 2 for M times, where M is a large number.

(4) After discarding the first K iterations as burn in period to eliminate the effect of the initial conditions, take the last $N-K$ iterations and infer about the moment of the true parameters.

The retained draws can be averaged to create estimates of posterior features of interest. Similar to Monte Carlo integration a weak law of large numbers can be invoked to say that, if $f(\cdot)$ is a function of interest and:

$$\hat{f}_{(N-K)} = \frac{1}{N} \sum_{s=K+1}^{N} g(\theta_i^s), \tag{1.35}$$

then $\hat{f}_{(N-K)}$ converges to $\mathrm{E}[f(\theta|Y_T)]$ as $N-K$ goes to infinity, see Cappé et al. (2007).

Figure 1.1 reports a graphical representation of the Gibbs sampling. It can be shown more rigorously that the sequence of samples constitutes a Markov chain, and under general conditions they converge to the target distribution called the stationary distribution of that Markov chain, that is nothing more than the joint posterior of $(\theta|Y_T)$. In situations where it is not convenient or possible to sample directly from the conditional distributions, one can use a Metropolis-Hasting algorithm inside the Gibbs bringing to the well known *Metropolis within Gibbs* algorithm.

The Metropolis-Hasting algorithm can draw samples from any probability distribution $f(\cdot)$ requiring a proposal density $Q(\cdot)$ easy to simulate and explicitly available. In Bayesian applications the constant of proportionality is often extremely difficult to compute, so the ability to generate a sample without knowing it is a major virtue of the algorithm. The algorithm associated with the target density $f(\cdot)$ and the proposal density $Q(\cdot)$ produces a Markov chain in which each state, x^{t+1}, depends only on the previous, x^t. The algorithm uses the proposal density $Q(x^{'}; x^t)$, which depends on the current state x^t, to generate a new proposed value $x^{'}$, if the proposed value is

accepted with acceptance probability

$$\alpha < A(x^{'}; x^{t}) = min \left\{ \frac{p(x^{'})Q(x^{t}; x^{'})}{p(x^{t})Q(x^{'}; x^{t})}, 1 \right\} \qquad \alpha \sim \boldsymbol{\mathcal{U}}(0, 1), \tag{1.36}$$

then the chain moves towards $x^{t+1} = x^{'}$ otherwise the current value of x_t is retained, $x^{t+1} = x^{t}$. The Metropolis-Hasting algorithm is very simple, but it requires careful design of the proposal distribution $Q(x^{t}; x^{'})$ that brings to different algorithms with different convergence properties, see Robert and Casella (2004) for a detailed discussions.

Illustrative example: Stochastic Volatility Models

Markov Chain Monte Carlo are widely used in econometrics and in particular in financial econometrics, where they are useful for the estimation of the stochastic volatility (SV) models, introduced by Taylor (1986). Since then the SV models have been extensively studied, see Shepard (2006) for a review. The second chapter of this dissertation deals with estimation, filtering and testing for this class of models and it seems useful to introduce the general idea of the estimation strategy.

The SV models assume that the underlying volatility is not constant but is time varying and in the general formulation the logarithm of the volatility is modelled as an AR(1) process. The corresponding model, known as $ARSV(1)$, is given by:

$$\begin{aligned} &y_t = \beta \epsilon_t e^{h_t/2}, && \epsilon_t \sim N(0, 1), \\ &h_{t+1} = \phi h_t + \eta_t, && \eta_t \sim N(0, \sigma_\eta^2), \\ &h_1 \sim N\left(0, \frac{\sigma^2}{1-\phi^2}\right), && \mathrm{Cov}(\epsilon_t, \eta_t) = 0, \quad t = 1, \ldots, T, \end{aligned} \tag{1.37}$$

where y_t is the observed series at time t and β is a scalar parameter. The properties of the $ARSV(1)$ models have been provided by Taylor (1994) and Shepard (1996) among other, see Broto and Ruiz (2004) for a review. The estimation of model (1.37) is difficult because the likelihood function is hard to evaluate. One of the main approach to the estimation of these models is the Quasi-Maximum Likelihood (QML) approach developed by Nelson (1988) and Harvey et al. (1994) based on linearizing the equation of formula (1.37):

$$\begin{aligned} &\log(y_t^2) = h_t + \xi_t, \\ &h_{t+1} = \phi h_t + \eta_t, && \eta_t \sim N(0, \sigma_\eta^2), \end{aligned} \tag{1.38}$$

where $\xi_t = \log(\epsilon_t^2) - E[\log(\epsilon_t^2)]$. The model (1.38) is a non-Gaussian linear state space model. Harvey et al. (1994) use the Kalman filter to estimate the model parameters $\theta = (\phi, \sigma_\eta^2)$ by maximizing the quasi log-likelihood:

$$\log L(y_1, \ldots, y_t, \theta) = -\frac{1}{2} \sum_{t=1}^{T} \log F_t - \frac{1}{2} \sum_{t=1}^{T} \frac{v_t^2}{F_t}, \tag{1.39}$$

where v_t is the one-step-ahead prediction error for the best linear estimator of $\log(y_t^2)$ and F_t is the associated mean square error. Although consistent and asymptotically normally distributed

as showed in Ruiz (1994), the QML procedure is inefficient and has poor small sample properties because does not rely on the exact likelihood of $\log(y_t^2)$.

The exact approach to inference of such model is based on MCMC and in particular on a *Metropolis within Gibbs* algorithm, see Chapter 2. This approach has been proposed by Jacquier et al. (1994) and Kim et al. (1998), later Bos and Shepard (2006) extended this methodology to more general settings.

For the model presented in (1.37) a generic MCMC algorithm can be sketched as follows:

(1) Initialize the volatility process h and the unknown parameters θ;

(2) Sample the latent stochastic volatility h, see Jacquier et al. (1994) and Kim et al. (1998);

(3) Sample every parameter from each corresponding posterior, if no closed-form solution is available use a Metropolis-Hasting;

(4) Go to 2 for many times.

The MCMC algorithm is not based on crude approximations of the model and provide a better inference even in small samples, the main problem is the computational burdensome of this method.

1.7 Model Selection

Starting with the seminal paper of Akaike (1974), which defined one of the first model selection criterion (the Akaike Information Criterion, AIC) there has been an enormous and growing, literature on the topic of model selection, from both a frequentist and Bayesian point of view. Model uncertainty and model selection deal with statistical models that are not defined precisely. The major focus of model selection methodologies is to examine the properties of each possible model and to insure that the "more" correct model is selected based on the available information set. The are many reasons for carry out a selection procedure and they can be summarized as follows:

- *Model construction*: when K different models are available the researcher wants to decide which one fits better the data at hand. It is not guaranteed that one of those K models is correct.
- *Model checking*: after a model or a family of models has been selected for various theoretical and practical reasons, one wants to know whether the data agrees with this type of models.
- *Model comparison*: when a few models are proposed because they fitted correctly other samples and one wonders which of these models best fits the current sample.

In the $ARIMA(p, d, q)$ framework, model choice involves the selection of the optimal order of the p, d, and q. Usually the selection procedure is carried out sequentially, starting from the biggest model the researcher impose restrictions on p, d and q, till the model with the lower AIC is reached.

Testing the state space models raises a number of issues, primarily because of the breakdown of standard regularity conditions, in this case the AIC criterion cannot be applied. This leads to a test that is more complicated, see Harvey (1989) and Harvey and Proietti (2005) for a review. In this dissertation we are interested in Bayesian model selection for linear, nonlinear and Non-Gaussian state space models and it turns out that it is more flexible than the classical approach.

The advantages of the Bayesian model selection respect to the classical tests are summarized among other in Koop (1992, 1994), Sims (1998), Sims and Uhlig (1991).

There are mainly two Bayesian approaches to deal with model selection. The first is the trans-dimensional Markov Chain Monte Carlo useful to obtain draws from the joint posterior density $p(M_k, \theta_1, \ldots, \theta_k | Y_T)$. The second relies to the marginal likelihood evaluation for each model at hand, $p(Y_T|M_k) = \int p(Y_T|M_k, \theta_k) p(\theta_k|M_k) d\theta_k$ for $k = 1, \ldots, K$. The marginal likelihood calculation is a challenging task and it can be evaluated (up to some numerical and approximation errors) using a simulation-based approach (e.g. Importance sampling and Bridge sampling) and the density ratios (e.g. Chib's estimators and Laplace approximation).

Suppose that we have a collection of k models for $k = 1, \ldots, K$, and each model has parameters $\theta_k \in \Theta_k$, in formula:

$$M_k = \{f(\cdot|\theta_k);\ \theta_k \in \Theta_k\},$$

where $f_k(\cdot|\theta_k)$ denotes the likelihood and Θ_k denotes the parameter space. For k models the whole parameter space is denoted by $(\mathbf{M}, \theta_1, \theta_2, \ldots, \theta_k)$ where $\mathbf{M}$ represents all the possible models. Each model is associated with a collection of priors on the parameters $\pi(\theta_k)$.

Reversible jumps algorithms introduced by Green (1995) are a class of algorithms that sample form the posterior distribution $p(\mathbf{M}, \theta_1, \ldots, \theta_k | Y_T)$ given by the following relation:

$$p(\mathbf{M}|\theta_1, \ldots, \theta_k, Y_T) \propto p(Y_T|\theta_1, \ldots, \theta_k) p(\theta_1, \ldots, \theta_k, \mathbf{M}),$$

where $p(\theta_1, \ldots, \theta_k, \mathbf{M}) = p(\theta_1, \ldots, \theta_k|\mathbf{M}) p(\mathbf{M})$ is the prior distribution of the parameters and the indices of these models.

Reversible jumps MCMC are based on a Markov Chain that lives on the state space

$$\mathbf{\Theta} = \bigcup_{k=1}^{K} (M_k \times \Theta_k), \tag{1.40}$$

and samples between different models, by making moves from the selected model $(M_k, \theta_k)^{(i)}$ at iteration i, to a new model $(M_j, \theta_j)^{(i+1)}$ at iteration $i+1$, these moves are called *jumps*. The Green (1995) algorithm is based on a transition kernel satisfying the following relation:

$$\int_A \int_B K(x, dy) \pi(x) dx = \int_B \int_A K(y, dx) \pi(y) dy, \tag{1.41}$$

for all $A, B \subset \mathbf{\Theta}$ and for some invariant density $\pi(\cdot)$. In Robert and Casella (2004) it is shown how to construct a proper reversible kernel for this problem. Typically the jumps are limited to moves from a model M_k to another one M_m that is close to the dimension of M_k using a sensible proposal $q_m(\cdot)$.

The reversible jump algorithm allows the movement to an enlarged parameter space Θ_{k+1} or to a reduced space Θ_k. It turns out that the movement from the model with k components to a model with $k+1$ although conceptually simple creates problems in the reverse move, indeed from a model with $k+1$ to a model with k components. The solution proposed by Green (1995) is to supplement each parameter space Θ_k and Θ_{k+1} with an artificial space in order to create a *bijection* between them, in this case the reverse move turns out to be natural.

Importance sampling is a simulation-based method useful to approximate the marginal likelihood given by:

$$p(Y_T|M_k) = \int p(Y_T|\theta_k)p(\theta_k)d\theta_k.$$

The marginal likelihood approximation, using the importance sampling, is obtained by the following relation:

$$p(Y_T|M_k) = \int \frac{p(Y_T|\theta_k)p(\theta_k)}{q(\theta_k)} q(\theta_k)d\theta_k, \tag{1.42}$$

where the $q(\theta_k)$ is the importance density. Given a sample $q(\theta_k^{(i)}) \sim q(\theta_k)$ for $i = 1, \ldots, N$, from the importance density the marginal likelihood can be estimated using

$$\tilde{p}(Y_T|M_k) = \frac{1}{N} \sum_{i=1}^{N} \frac{p(Y_T|\theta_k^{(i)})p(\theta_k^{(i)})}{q(\theta_k^{(i)})}. \tag{1.43}$$

The tail of the importance density has to be fat respect to the posterior density $p(\theta_k|Y_T)$, this requirement is not sufficient in some circumstances to guarantee a good approximation of the marginal likelihood.

The bridge sampling has been proposed by Meng and Wong (1996) for evaluating the ratios of normalizing constant and has been later extended to the evaluation of the marginal likelihood in a model selection framework by DiCiccio et al. (1997). The main problem with the importance sampling is the unstable behaviour in the case of unbounded ratio of the nonnormalized posterior density over the importance density. The bridge sampling generalizes the importance sampling combining *i.i.d.* samples from an importance density with the MCMC draws from the posterior density and it turns out to be much more stable.

Define the $q(\theta_k)$ as the importance density and $\alpha(\theta_k)$ a function satisfying:

$$C_\alpha = \int \alpha(\theta_k)p(\theta_k|Y_T, M_k)q(\theta_k)d\theta_k > 0. \tag{1.44}$$

The bridge sampling is based on the following equation:

$$\begin{aligned} 1 &= \frac{\int \alpha(\theta_k)p(\theta_k|Y_T, M_k)q(\theta_k)d\theta_k}{\int \alpha(\theta_k)q(\theta_k)p(\theta_k|Y_T, M_k)d\theta_k} \\ &= \frac{\mathrm{E}_q[\alpha(\theta_k)p(\theta_k|Y_T, M_k)]}{\mathrm{E}_p[\alpha(\theta_k)q(\theta_k)]} \end{aligned} \tag{1.45}$$

where $\mathrm{E}_p[\alpha(\theta_k)q(\theta_k)]$ is the expectation with respect to $p(\theta_k|Y_T, M_k)$. Substituting the relation $p(\theta|Y_T, M_k) = \hat{p}(\theta_k|Y_T, M_k)/p(Y_T|M_k)$ in equation (1.45) we get the main bridge sampling identity:

$$p(Y_T|M_k) = \frac{\mathrm{E}_q[\alpha(\theta_k)\hat{p}(\theta_k|y, M_k)]}{\mathrm{E}_p[\alpha(\theta_k)q(\theta_k)]}. \tag{1.46}$$

To estimate the equation (1.46) the expectations on the right-hand side are substituted by sample averages:

$$\tilde{p}(Y_T|M_k) = \frac{N^{-1}\sum_{i=1}^{N} \alpha(\theta_k^{(i)})\hat{p}(\theta_k^{(i)}|Y_T, M_k)}{M^{-1}\sum_{m=1}^{M} \alpha(\theta_k^{(m)})q(\theta_k^{(m)})}, \tag{1.47}$$

where the numerator is approximated using Markov Chain Monte Carlo draws, $\{\theta_k^{(i)}, i = 1, \ldots, N\}$ from $p(\theta_k|Y_T, M_k)$ and the denominator is the approximated using the *i.i.d.* draws, $\{\theta_k^{(m)}, m = 1, \ldots, M\}$ from the importance density $q(\theta_k)$. The bridge sampling encompasses different simulation-based methods depending by the function $\alpha(\theta_k)$, see Frühwirth-Schnatter (2006) and the reference therein. Finally the optimal choice of $\alpha(\theta_k)$ that minimizes the expected relative error of the estimator $\tilde{p}(Y_T|M_k)$ is provided by Meng and Wong (1996).

The posterior density ratio provides a easy way to compute the marginal likelihoods of k different models $p(Y_T|M_k)$. It uses a formal equivalence between the marginal likelihood and the normalizing constant appearing in the definition of the posterior distribution $p(\theta_k|Y_T, M_k)$. Chib (1995) and Chib and Jeliazkov (2001) are examples of this technique, see Chapter 2.

Laplace approximation is a widely used method of approximating the marginal likelihood. This is accomplished by substituting the $p(\theta_k|Y_T, M_k)$ with a local normal density $f_N(M_k; \tilde{\theta}_k, \Sigma)$ where $\tilde{\theta}_k$ is the posterior mode and Σ^{-1} is minus the Hessian matrix of the log posterior evaluated at the point $\tilde{\theta}_k$.

The log of the marginal likelihood can therefore be calculated by:

$$\log \hat{p}(Y_T|M_L) \approx \log p(Y_T|\tilde{\theta}_k) + \log p(\tilde{\theta}_k) + \frac{dim(\theta_k)}{2} \log(2\pi) + 0.5 \log |\Sigma|.$$

The accuracy of this method depends crucially on the local normal approximation that can be poor in some circumstances, moreover Σ^{-1} is often not available. Lewis and Raftery (1997) to overcome those problems proposed a generalization of Laplace approximation based on the posterior simulation output for the individual models, this is called Laplace-Metropolis estimator.

In the linear state space model we carry out a search strategy that given a limited number of predictors, searches over all possible models. The estimation of the unknown parameters and the model selection are carried out jointly, see Chapter 4. In case of nonlinear and Non-Gaussian state space models we use the Chib and Jeliazkov (2001) technique based mainly on MCMC and Particle Filter algorithms, see Chapter 2.

1.8 Thesis Outline

The remainder of this thesis is organized as follows. Chapter 2 deals with stochastic volatility models. The local level model with stochastic volatility, recently proposed for U.S. Inflation by Stock and Watson (2007) provides a simple yet sufficiently rich framework for characterizing the evolution of the main stylized facts concerning the U.S. inflation. The model decomposes inflation into a permanent component, evolving as a random walk, and a transitory component. The volatility of the disturbances driving both components is allowed to vary over time. The chapter provides a full Bayesian analysis of this model and readdresses some of the main issues that were raised by the literature concerning the evolution of persistence and predictability and the extent and timing of the great moderation. The assessment of various nested models of inflation volatility and systematic model selection provide strong evidence in favor of a model with heteroscedastic disturbances in the permanent component, whereas the transitory component has time invariant size. The main evidence is that the great moderation is over, and that volatility, persistence and predictability of inflation, underwent a turning point around 1995. During the last decade volatility and persistence have been increasing and predictability has been going down.

Chapter 3 implements a new maximum likelihood estimation method for high-dimensional dynamic multi-factor models in presence of large amount of missingness. The exact treatment of missing values with a reduction technique proposed by Jungbacker and Koopman (2008) enables the estimation of the factors and a large number of parameters in very fast and efficient way. This new methodology is used to analyze the evolution of the degree of global and regional interdependence over the period 1950-2007. The chapter also provides the decomposition of the aggregate output for 150 countries into factors that are (i) common across all countries, (ii) common across regional areas and (iii) specific to each countries. The chapter provides a systematic assessment of the estimation strategy and discusses the empirical evidence in the light of the previous literature. Finally it provides new results about global and regional convergence.

Chapter 4 provides a new Bayesian model selection technique (Frühwirth-Schnatter and Wagner, 2009) to characterising the nature of the trend in macroeconomic time series. It illustrates that the methodology known as stochastic model specification search can be quite successfully applied to discriminate between stochastic and deterministic trends, it then provides a more general version of the test that allows for autoregressive parameters. The validity of this procedure is shown using the same fourteen macroeconomic and financial series considered in the seminal paper by Nelson and Plosser (1982) and to an updated data set that spans the period between 1947 and 2009. The results show that the unit root hypothesis cannot be rejected in most of the series. Finally the chapter provides a study about the robustness of this technique and a Monte Carlo experiment.

Chapter 5 contains the general conclusions.

Chapter 2

Stochastic Volatility Models in US Inflation

2.1 Introduction

Inflation's volatility has attracted a great deal of attention recently; the interest has been sparked by the debate on the Great Moderation, that has been documented for real economic aggregates. Inflation stabilization is indeed a possible source of the reduction in the volatility of macroeconomic aggregates. The issue is also closely bound up with inflation persistence and predictability. In an influential paper Stock and Watson (2007), using a local level model with stochastic volatility, document that inflation is less volatile now than it was in the 1970s and early 1980s; moreover, persistence, which measure the long run effect of a shock, has declined, and predictability has increased.

There is still an ongoing debate about the statistical significance of inflation persistence and its stability over time, see Pivetta and Reis (2007), Cogley, Primicieri and Sargent (2008), Cecchetti et al. (2007), among others. Recently Bos, Koopman and Ooms (2008) analyzed a U.S. core inflation series (excluding food and energy) as a long memory process subject to heteroscedastic shocks, and documented remarkable changes, taking place about at the time of the Great Moderation (1984), in the volatility of the series and the fractional integration parameter (which is the measure of persistence adopted in that paper).

In this paper we reconsider the unobserved components model of U.S. inflation estimated in Stock and Watson (2007), referred to as the local level model with stochastic volatility (UC-SV). The model provides a simple yet sufficiently rich framework for discussing the main stylized facts concerning inflation, such as the changes in persistence and predictability. The model postulates the decomposition of observed inflation into two components: the permanent component (or underlying inflation) which captures the trend in inflation, and the transitory component, which captures the deviations of inflation from its trend value. We will start from a specification such that both components are driven by disturbances whose variance evolves over time according to a stationary stochastic volatility process, and will attempt to assess the significance of the changing volatility in each of the components.

The contributions of this paper are the following: we provide a full Bayesian analysis, so that, unlike the current literature, we do not assume that some of the parameters, namely the variances of the stochastic volatility components, are known. Secondly, we carry out systematic model selection by comparing the marginal likelihood implied by the different models of inflation volatility. The marginal likelihood is estimated according to the Chib and Jeliazkov (2001) algorithm.

The interesting final result is that we find strong support for the specification with stochastic volatility in the permanent component, but not in both. We document that persistence is higher than in previous studies and is subject to a significant increase starting from the second half of the 90's, whereas predictability has decreased somewhat at about the same time.

This chapter is organized as follows. In Section 2.2 we present the local level model with stochastic volatility. Section 2.3 illustrates the Monte Carlo Markov Chain (MCMC) sampling scheme used to perform Bayesian inference for this model. In Section 2.4 we present and discuss the estimation results. In Section 2.5 and in Section 2.6 we provide a detailed description of the Chib and Jeliazkov (2001) technique an for the auxiliary particle filter that is needed to evaluate the conditional likelihood. In Section 2.7 we apply the Chib and Jeliazkov (2001) approach to the evaluation of the marginal likelihood. The results are used to select the final model among four competitors. Section 2.8 concludes the chapter.

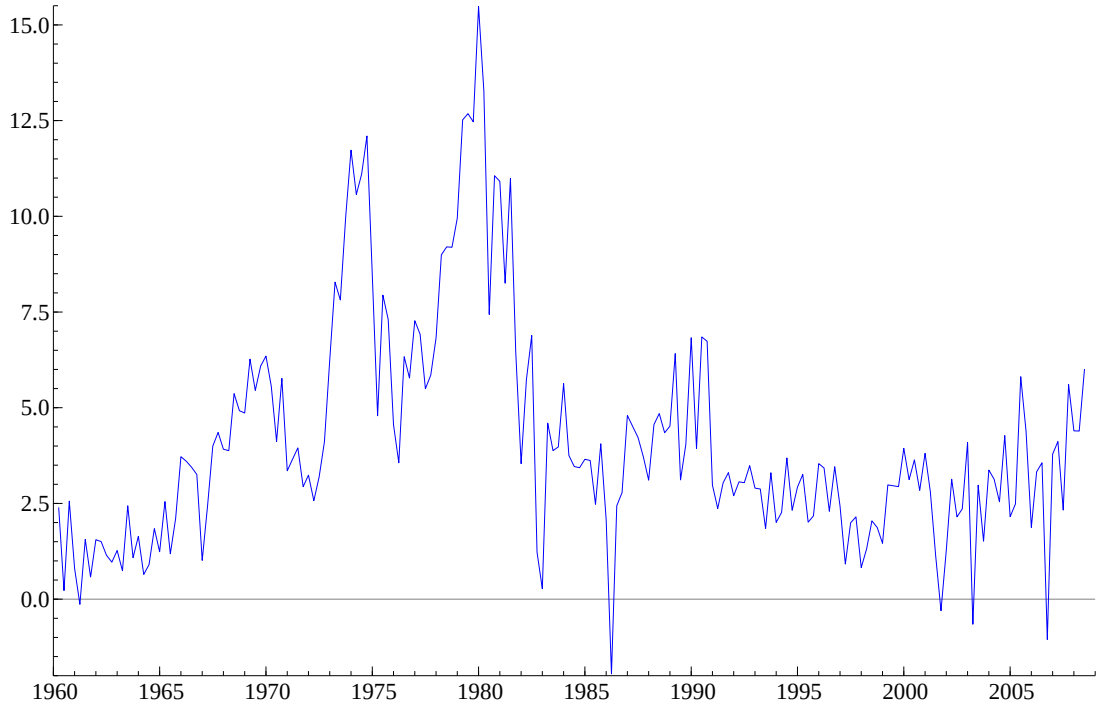


Figure 2.1: *Quarterly U.S. Inflation,* $y_t = 400\Delta \ln CPI_t$

2.2 The UC-SV model

The paper focuses on the quarterly inflation rate constructed from the Consumer Price Index (All Urban Consumers, seasonally adjusted), made available by the U.S. Bureau of Labor Statistics. The quarterly index is obtained from the monthly index by computing the average of the three months that make up each quarter; if we denote the quarterly series by CPI_t, the annualized quarterly inflation rate, denoted y_t, $t = 1, ..., T$, is computed as $y_t = 400\Delta \ln \text{CPI}_t$. The series is plotted in figure 2.1 and is available for the sample period 1960:q1 –2008:q3.

The most general specification of the UC-SV model with stochastic volatility represents inflation as the sum of an underlying level, denoted here by α_t, which evolves as a random walk, and a transitory component:

$$\begin{array}{rcll} y_t & = & \alpha_t + \sigma_{\varepsilon t}\varepsilon_t, & \varepsilon_t \sim N(0,1), \\ \alpha_t & = & \alpha_{t-1} + \sigma_{\eta t}\eta_t, & \eta_t \sim N(0,1), \end{array} \tag{2.1}$$

where ε_t and η_t are independent standard normal Gaussian disturbances and their size, $\sigma_{\eta t}$ and $\sigma_{\varepsilon t}$, respectively evolve over time according to a SV process. Denoting $h_{1t} = \ln \sigma^2_{\varepsilon t}$ and $h_{2t} = \ln \sigma^2_{\eta t}$,

$$\begin{array}{rcll} h_{1,t} & = & \mu_1 + \phi_1 h_{1,t-1} + \kappa_{1,t}, & h_{1,0} \sim N\left(0, \dfrac{\sigma^2_{\kappa_1}}{1-\phi_1^2}\right), \quad \kappa_1 \sim N(0, \sigma^2_{\kappa_1}), \\ h_{2,t} & = & \mu_2 + \phi_2 h_{2,t-1} + \kappa_{2,t}, & h_{2,0} \sim N\left(0, \dfrac{\sigma^2_{\kappa_2}}{1-\phi_2^2}\right), \quad \kappa_2 \sim N(0, \sigma^2_{\kappa_2}). \end{array} \tag{2.2}$$

The model encompasses the traditional stochastic volatility model that is widely used in finance (see for instance Shephard, 2006), which arises when the process α_t degenerates to a constant.

The specification of the stochastic volatility processes differ only slightly from Stock and Watson (2007) and Cecchetti et al. (2007), who assume a random walk process for the log-variances $h_{it}, i = 1, 2$. In fact, their specification is encompassed by (2.2), which is a more canonical specification of a volatility model (see for instance Jacquier et al., 1994, and Kim et al., 1998), since it arises as the discrete-time approximation to the Ornstein-Uhlenbeck continuous time process used in finance, and ensures the stationarity of η_t and ε_t, provided that $|\phi_i| < 1, i = 1, 2$. As a matter of fact, when the autoregressive coefficients ϕ_i are close to unity and the constants $\mu_i, i = 1, 2$, are close to zero, specification (2.2) is virtually indistinguishable from a random walk.

When both variances $\sigma^2_{\varepsilon t}$ and $\sigma^2_{\eta t}$ do not vary with time, the model reduces to the the traditional local level model. The latter has a IMA(1,1) reduced form, $\Delta y_t = \xi_t + \vartheta \xi_{t-1}$, with $\xi_t \sim \text{NID}(0, \sigma^2)$. The structural parameters are related to the reduced form parameters by the two equations $\sigma^2(1 + \vartheta^2) = \sigma^2_\eta + 2\sigma^2_\varepsilon$, $\sigma^2\vartheta = -\sigma^2_\varepsilon$, which are obtained by equating the autocovariances at lags 0 and 1, respectively; from these we obtain the moving average parameter $\vartheta = \left[(q^2 + 4q)^{\frac{1}{2}} - 2 - q\right]/2$, where $q = \sigma^2_\eta/\sigma^2_\varepsilon$ is the signal to noise ratio, and the prediction error variance (p.e.v.), $\sigma^2 = -\sigma^2_\varepsilon/\vartheta$. Notice that ϑ is restricted within the range [-1,0]. The local level model has a long tradition and a well-established role in the analysis of economic time series, since it provides the model-based interpretation for the popular forecasting technique known as *exponential smoothing*, which is widely used in applied economic forecasting and fares remarkably well in forecast competitions; see Muth (1960) and the comprehensive reviews by Gardner (1985, 2006). In the sequel we shall also consider the cases when either $\sigma^2_{\varepsilon t}$ or $\sigma^2_{\eta t}$, or both, are constant.

The UC-SV model can be considered as an IMA(1,1) model with time-varying p.e.v. and moving average parameter. This suggests taking, as a local measure of persistence, $P_t = 1 + \vartheta_t$, where ϑ_t varies with time according to the values of the time-varying signal to noise ratio $q_t = \sigma^2_{\eta t}/\sigma^2_{\varepsilon t}$. The quantity P_t decreases linearly from 0 to 1 as ϑ increases from -1 to 0. Cecchetti et al. (2007) use the implied time varying first order autocorrelation of Δy_t, as a measure of persistence; the local autocorrelation (i.e. conditional on $\sigma^2_{\eta t}$ and $\sigma^2_{\varepsilon t}$) is $\rho_t(1) = -1/(q_t + 2) = \vartheta_t/(1 + \vartheta^2_t)$. Alternatively, we can use the (conditional) normalized spectral generating function at the zero frequency, which is

$$P^*_t = \frac{\sigma^2_{\eta t}}{\sigma^2_{\eta t} + 2\sigma^2_{\varepsilon t}} = 1 + \frac{2\vartheta_t}{1 + \vartheta^2_t} = 1 + 2\rho_t(1).$$

This measure decreases monotonically from 0 to 1 as ϑ increases from -1 to 0.

As a measure of local predictability we can take the prediction error variance, conditional on $\{h_{it}, i = 1, 2, t = 1, \ldots\}$, which is defined as

$$\sigma^2_t = -\frac{\sigma^2_{\varepsilon t}}{\vartheta_t}.$$

A relative measure of predictability can be defined in terms of the Granger and Newbold (1986, p. 310) forecastability index:

$$\text{Pred}_t = 1 - \frac{\text{Var}(\xi_t|h_{it})}{\text{Var}(\Delta y_t|h_{it})} = \frac{\vartheta^2_t}{1 + \vartheta^2_t}. \tag{2.3}$$

In terms of the parameters of the UC-SV, the prediction error variance equals $\text{Var}(\xi_t|h_{it}) = \frac{\sigma^2_{\eta t}}{(1+\vartheta)^2}$, whereas the variance $\text{Var}(\Delta y_t|h_{it}) = \sigma^2_{\eta t} + 2\sigma^2_{\varepsilon t}$. The above measure ranges between 0 ($\vartheta_t = 0$) and

0.5 ($\vartheta_t = -1$), and it is negatively related to the persistence of the process. As a matter of fact, as ϑ_t ranges from -1 to 0, predictability decreases from its maximum, 0.5, to zero.

2.3 Bayesian Estimation

This section provides an overview of the MCMC methodology adopted for the estimation of the UC-SV model. All inferences are based on a Gibbs sampling scheme, according to which samples are drawn componentwise from the full conditionals; for the components which cannot be sampled directly a Metropolis-Hasting sub-chain is used within the Gibbs sampling cycle. In particular, the posterior of the AR parameters, ϕ_1 and ϕ_2, is not available in closed form; see Kim et al. (1998) and Bos and Shepard (2006). More details on the specification of the prior distributions, the full conditionals and the Metropolis-within Gibbs steps are provided in Appendix A.

Let $\theta = (\mu_1, \mu_2, \phi_1, \phi_2, \sigma^2_{\kappa_1}, \sigma^2_{\kappa_2})$ denote the vector of hyperparameters, $h_i, i = 1, 2$, be the collection of the values of the latent stochastic volatility processes, and α and y denote the stack of the values of permanent inflation and the series, respectively. The Gibbs sampling scheme can be sketched as follows:

1. Initialize h_i, θ
2. Draw a sample from $\theta, \alpha | y, h_i$
   a) Sample θ from $\theta | y, \alpha, h_i$ (see Appendix A).
   b) Sample α from $\alpha | y, \theta, h_i$, using the simulation smoother proposed by Durbin and Koopman (2002).
3. Sample $h_i, i = 1, 2$, from $h_i | \alpha, y, \theta$, using a Random Walk Metropolis-Hastings algorithm;
4. Go to 2.

The most complex part of the algorithm deals with the simulation of the stochastic volatility processes. We adopt a single move sampler based on the density:

$$h_{it} | h_{i,t+1}, h_{i,t-1}, y_t, \alpha_{t-1}, \alpha_t. \tag{2.4}$$

For this purpose, we implement a Random Walk Metropolis-Hastings algorithm, described in detail in Appendix A; see also Cappé et. al. (2007). In order to sample from the full conditional we use the following results:

$$f(h_{i,t} | h_{i,t-1}, h_{i,t+1}, y_t, \alpha_t, \alpha_{t-1}) \propto f(h_{i,t} | h_{i,t-1}) f(y_t | \alpha_t, h_{1,t}) f(\alpha_t | \alpha_{t-1}, h_{2,t}). \tag{2.5}$$

2.4 Estimation Results

This section reports on the main estimation results for the model presented in section 2.2. The MCMC sampler was initialized by setting all $h_{i,t} = 0$ and $\phi_i = 0.86$, $\sigma^2_i = 0.07$ and $\mu_i = 0.6$. We iterated the sampler for a burn-in period consisting of 12,500 iterations, before recording the draws from a subsequent 25,000 iterations. The programm is written in Ox v. 5.10 console (Doornik, 2007) using our own source code. The time needed for all calculations (including the additional

simulations required to evaluate the marginal likelihood with the Chib and Jeliazkov method) is about 35 minutes.

Figure 2.2 displays the inflation series with the posterior mean of the permanent component, and the interval estimates of two stochastic volatility components for the irregular and the permanent disturbances, $\sigma_{\varepsilon t}$ and $\sigma_{\eta t}$. The third panel shows that the volatility of the permanent component has been increasing from the 60ies up to 1982, and then is slowly decreasing. The volatility of the transitory component (central panel) is much more stable, instead. In the sequel of the paper we will address the question as to whether it can be considered as time invariant.

The estimates of the latent volatility processes are comparable to the corrected estimates obtained by Stock and Watson (2007) and displayed in their Figure 2.2, referring to CPI (all items), on page 8, panels (a) and (b), of the document available at http://www.princeton.edu/ mwatson. In particular, the estimated standard deviation of the permanent component shows two distinctive peaks in 1975 and 1981, and changes substantially over time; on the contrary, the volatility of the irregular component is much less evolutive. The difference that arise are due to the different sample considered and to the fact that Stock and Watson estimate a restricted version of the model (in particular, $\phi_i = 1, \mu_i = 0, \sigma^2_{\kappa i} = 0.2, i = 1, 2$; notice that the variance of the volatility shocks is not estimated).

Figure 2.3 displays the evolution of the Monte Carlo estimates of the posterior mean of the signal to noise ratio, q_t, of the persistence parameter, P_t, the prediction error variance and the predictability measure, Pred_t. The graph reveals that the size of the random walk component increases during the 70s, when the trend dynamics become more sustained, and it is lower in the 80s. Persistence is time varying at values well below 1 and there is evidence for a strong decreasing tendency during the 80s. The robustness of these results will be discussed later. As far as predictability is concerned, the prediction error variance undergoes a decline after 1982 (this is consistent with the results of Bos, Koopman and Ooms, 2008). In relative terms, the forecastability index shows an increase in the 80s.

Table 2.1 reports some summary statistics concerning the posterior distribution of the parameters and some convergence diagnostics. As for the assessment of convergence, we report the Geweke statistics: let $\theta^{(j)}$ denote the j-th sample of the sampling scheme for the generic parameter θ after the initial burn-in period. Let also $\bar{\theta}_a$ denote the average of the first n_a draws, $\bar{\theta}_b$ the average of the last n_b draws at the end of the convergence period, which are taken sufficiently remote to prevent any overlap, the Geweke's convergence statistic (Geweke, 1992, 2005) is defined as

$$C_G = \frac{\bar{\theta}_a - \bar{\theta}_b}{\sqrt{V_{L,a}/n_a + V_{L,b}/n_b}},$$

where

$$V_{L,k} = c_{0,k} + 2 \sum_{j=1}^{n_k - 1} w_j c_{j,k}, \quad k = a, b,$$

is the long run variance of the parameter sample path for the n_k draws, based on a weighted combination of the autocovariances of the draws at lag j, $c_{j,k}$, with weights w_j that are decreasing in j and ensure that $V_{L,k} \geq 0$. A customary choice is the set of linearly declining weights $w_j = \frac{l-j}{l+1}$, where l is the truncation parameter.

The inefficiency factor is $\text{INEF} = 1 + 2 \sum_{j=1}^{n-1} w_j \hat{\rho}_j$, where $\hat{\rho}_j$ is the sample autocorrelation of the draws at lag j. This can be interpreted as a normalized measure of persistence of the draws.

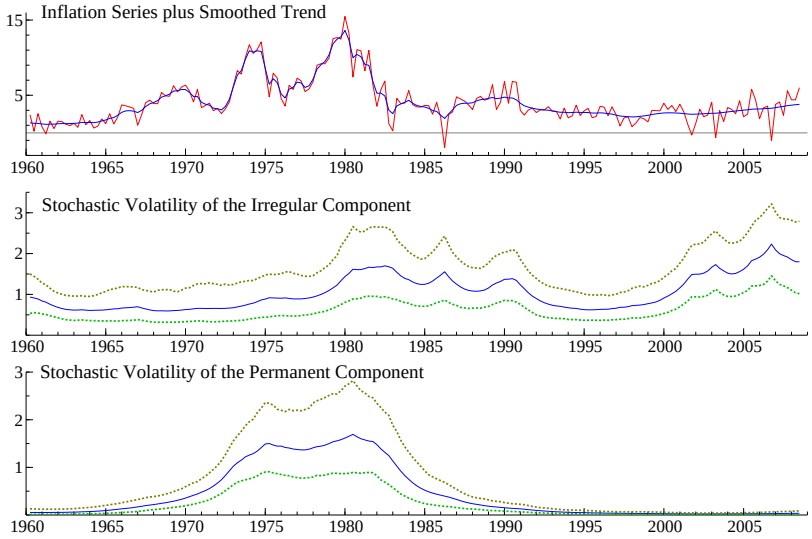


Figure 2.2: *Upper: Inflation and posterior mean of permanent component; Middle: Irregular Volatility component with 95 percent credible interval. Bottom: Permanent Volatility component with 95 percent credible interval.*

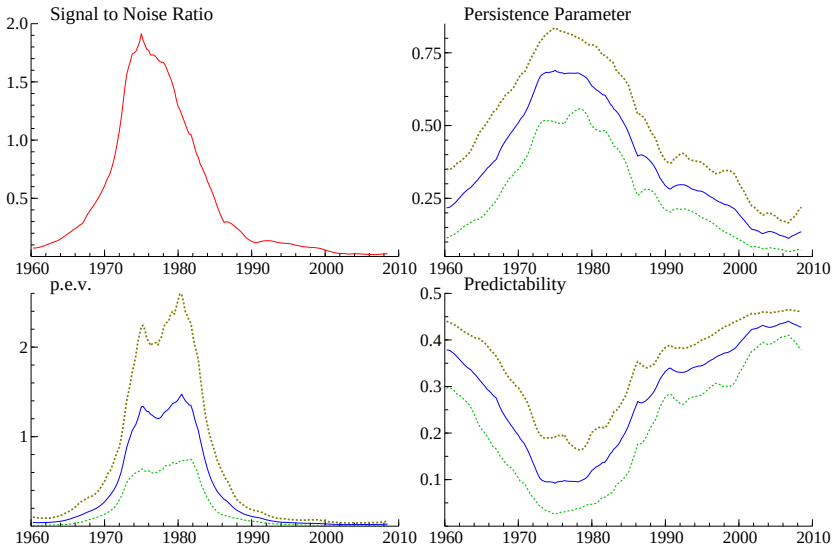


Figure 2.3: *Upper left: Signal to noise ratio. Upper Right: Persistence, with 95 percent credible interval. Bottom left: Prediction error variance, with 95 percent credible interval. Bottom right: Predictability, with 95 percent credible interval.*

Large values imply that the draws are strongly and positively autocorrelated (the spectral power is concentrated at the origin), so that the chain explores the parameter space very slowly and the additional information content of a draw is small.

The values reported in table 2.1 highlight that the convergence assessment of the chain are not fully satisfactory, since the Geweke statistic for some parameters, like μ_1 and ϕ_2, are strongly significant.

Table 2.1: *Posterior mean, Median, Geweke statistic and Inefficiency factor for UC-SV model*

Parameters	Mean	Median	Geweke' G_C	INEF
μ_1	-0.0017	-0.0015	-11.20	137.1
μ_2	-0.0253	-0.0252	-2.30	30.92
ϕ_1	0.9356	0.9372	1.18	13.00
ϕ_2	0.9885	0.9905	15.06	307.7
$\sigma^2_{\kappa_1}$	0.0491	0.0482	-2.38	24.43
$\sigma^2_{\kappa_2}$	0.0487	0.0479	-0.18	63.41

2.5 Chib and Jeliazkov Method

Chib (1995) and Chib and Jeliazkov (2001) proposed a technique for estimating the marginal likelihood useful to carry out Bayesian model selection. Define $M_1, \ldots, M_k$ as the k competitive models, $f_k(Y_T|\theta)$ as the sampling density (likelihood function) for model k, $Y_T = \{y_1, \ldots, y_t\}$ the available observations and $\pi_k(\theta)$ the associated prior density. Suppose that the Gibbs sampler has been applied to model k and we have a collection of samples:

$$\{\theta^{(g)}, z^{(g)}\}_{i=1}^{G}, \tag{2.6}$$

where $\theta^{(g)}$ represent the draws of the parameters and $z^{(g)}$ is the latent data, for example a latent process. The objective is to compute the marginal density $m(Y_T|M_k)$ for the output obtained from (2.6).

The posterior relation for each model is:

$$\pi_k(\theta|Y_T) = \frac{f_k(Y_T|\theta)\pi_k(\theta)}{m(Y_T|M_k)},$$

by virtue of $m(Y_T|M_k)$ to be the normalizing constant of the posterior density we can rewrite the relation in the following way:

$$m(Y_T|M_k) = \frac{f_k(Y_T|\theta)\pi_k(\theta)}{\pi_k(\theta|Y_T)}, \tag{2.7}$$

taking the logarithm and evaluating the resulting expression at the highest density point, θ^*, we get:

$$\log m(Y_T|M_k) = \log f_k(Y_T|\theta^*) + \log \pi_k(\theta^*) - \log \pi_k(\theta^*|Y_T). \tag{2.8}$$

This expression requires the evaluation at θ_* of the log-likelihood function, $\log f_k(Y_T|\theta^*)$, the log prior density, $\log \pi_k(\theta^*)$ and an estimate of the posterior ordinate, $\log \pi_k(\theta^*|Y_T)$. The original

paper of Chib (1995) proposes a numerical algorithm to evaluate the $\log \pi_k(\theta^*|Y_T)$ in the case of closed-form solution for the posteriors, this is easily done using the draws of the Gibbs sampler. Although the technique proposed by Chib (1995) is very flexible it can deals just with closed-form posterior of each parameter. Later Chib and Jeliazkov (2001) extended this technique to the cases of no closed-form expression for the posteriors. The algorithm is valid virtually for every MCMC algorithm, the main drawback rely on the high computational requirements. The main steps can be summarized as follows:

(1) After a Gibbs sampler has been performed for all possible models;

(2) Evaluate the $\log \pi_k(\theta^*)$ at a given point θ^*;

(3) Evaluate the log-likelihood $\log f_k(Y_T|\theta^*)$ at a given point θ^*. The log-likelihood of a linear state space model can be easily computed using the Kalman Filter, in case of nonlinear and non-Gaussian state space model the log-likelihood has to be evaluated with the particle filter;

(4) Evaluate the quantity $\log \pi_t(\theta^*|Y_T)$ using the relations given in Chib (1995) and Chib and Jeliazkov (2001);

(5) Evaluate formula (2.8) and repeat the steps 2, 3 and 4 for all the K possible models.

The model with the highest $\log m(Y_T|M_k)$ will be the chosen model.

2.6 Auxiliary Particle Filter

The auxiliary particle filter (APF) has been proposed by Pitt and Shepard (1999) to overcome one particular problem of the sequential importance sampling methods, it attempts to reduce the variability of the importance weights. The resampling step takes place at the conclusion of one iteration before moving to the next observation. It would be nicer and more efficient if one could pre-weight the particles prior to the resampling step in order to reflect how they are compatible with the next observation, this is essential the idea behind the auxiliary particle filter. The notation used in this section is the same as section 1.4.

Assuming that α_{t+1} is unknown the relationship, in the filtering and smoothing context, between α_t and y_{t+1} is:

$$p(y_{t+1}|\alpha_{1:t}, y_{1:t}) = p(\alpha_t|y_{1:t}) \int f_{t+1}(\alpha_{t+1}|\alpha_t) g_{t+1}(y_{t+1}|\alpha_{t+1}) d\alpha_{t+1} \tag{2.9}$$

which is the integral of the joint distribution of $\alpha_{t:t+1}$ and y_{t+1} given $y_{1:t}$. Defining $g_{t+1}(y_{t+1}|\alpha_t) = \int f_{t+1}(\alpha_{t+1}|\alpha_t) g_{t+1}(y_{t+1}|\alpha_{t+1}) d\alpha_{t+1}$ it would be desirable to use a term of this sort to determine how well a set of particles matches the next observation before the resampling step is performed. The auxiliary particle filter employs some approximation $\hat{g}_{t+1}(y_{t+1}|\alpha_t)$ of the *predictive likelihood* $g_{t+1}(y_{t+1}|\alpha_t)$ in an additional pre-weighting step and uses an auxiliary variable $\lambda^{(i)}$ to make use of these weights.

The general version of the auxiliary particle can be sketched as follows:

- At time $t = 1$:
  - Sample N particles from $\tilde{\boldsymbol{\alpha}}_1^{(i)} \sim q_1(\alpha_1)$;

  - Compute the weights $w_1(\boldsymbol{\alpha}_1^{(i)})$ and normalize them $W_1^{(i)} = \dfrac{w_1(\boldsymbol{\alpha}_1^{(i)})}{\sum_{j=1}^{N} w_1(\boldsymbol{\alpha}_1^{(j)})}$;
  - Resample $\{W_1^{(i)}, \boldsymbol{\alpha}_1^{(i)}\}$ to obtain N equally-weighted particles $\left\{\dfrac{1}{N}, \bar{\boldsymbol{\alpha}}_1^{(i)}\right\}$;

- For t>2, until $t = T$:
  - Calculate the auxiliary weights $\lambda_t^{(i)} \propto \hat{g}_t(y_t|\boldsymbol{\alpha}_{t-1}^{(i)})$ and normalize such that $\sum_{i=1}^{N} \lambda_t^{(i)} = 1$;
  - Sample $\varphi_t^{(i)}$ such that $P(\varphi_t^{(i)} = j) = \lambda_t^{(j)}$ (i.e. sample from the discrete distribution with parameter $\lambda_t^{(i)}$);
  - Sample N particles from $\tilde{\boldsymbol{\alpha}}_t^{(i)} \sim q_t(\alpha_t|\boldsymbol{\alpha}_{1:t-1}^{\varphi_t^{(i)}})$ and set $\boldsymbol{\alpha}_{1:t}^{(i)} \leftarrow \left(\boldsymbol{\alpha}_{1:t-1}^{\varphi_t^{(i)}}, \tilde{\boldsymbol{\alpha}}_t^{(i)}\right)$;
  - Compute the weights $\tilde{w}_t(\tilde{\boldsymbol{\alpha}}_t^{(i)})$ and normalize them $\tilde{W}_t^{(i)} = \dfrac{\tilde{w}_t(\tilde{\boldsymbol{\alpha}}_t^{(i)})}{\sum_{j=1}^{N} \tilde{w}_t(\tilde{\boldsymbol{\alpha}}_t^{(j)})}$;
  - Resample $\{\tilde{W}_t^{(i)}, \tilde{\boldsymbol{\alpha}}_t^{(i)}\}$ to obtain N equally-weighted particles $\left\{\dfrac{1}{N}, \bar{\boldsymbol{\alpha}}_t^{(i)}\right\}$.

The APF nests other algorithms in the literature as special cases, if we do not use a predictive likelihood then the APF reduces to the bootstrap filter of Gordon et al. (1993). Johansen and Doucet (2008) showed that it can interpreted as a standard sequential Monte Carlo algorithm applied to a sequence of target distributions that take into account the approximate predictive likelihood. Finally the auxiliary particle filter gives as output the log-likelihood as a weighted sum of the first and second stage weights, see Appendix B.

2.7 Model Selection

Thus far the literature has focused on fitting the UC-SV model (sometimes with arbitrary restrictions on the parameters $\sigma_{\kappa i}^2$) and describing the estimation result. There is a potential danger that the UC-SV model could be overfitting the data, but little or no attention has been devoted to this issue.

We thus turn our attention to Bayesian model selection. The models under comparison are the following four variants of the local level model:

- M_1: the Local Level Model with homoscedastic disturbances (UC);
- M_2: the Local Level Model with a SV disturbance only on the transitory component (UC-SVt);
- M_3: the Local Level Model with a SV disturbance only on the permanent component (UC-SVc);
- M_4: the Local Level Model with two SV disturbances (UC-SV).

Bayesian model comparison entails the computation of posterior model probabilities, see Geweke (2005) for more details. If the models have the same prior probability, the ratio of the posterior mode probabilities is the Bayes factor, which is the ratio of the marginal likelihoods of two rival specifications. The main difficulty lies with the evaluation of the marginal likelihood. For this purpose we adopt the method proposed by Chib and Jeliazkov (2001), which is based on the MCMC output, and additional draws from given partial full conditionals.

Denoting by $f(y|\theta_k, M_k)$ the conditional density of the data, given M_k and the parameter vector θ_k, and by $\pi(\theta_k|M_k)$, $\pi(\theta_k|y, M_k)$, the prior and posterior densities, respectively, of θ_k, the Chib and Jeliazkov (2001) approach is based on the following basic marginal likelihood identity:

$$m(y|M_k) = \frac{f(y|M_k, \theta_k)\pi(\theta_k|M_k)}{\pi(\theta_k|y, M_k)}, \qquad k = 1, 2, 3, 4, \tag{2.10}$$

where $m(y|M_k)$ is the marginal likelihood of model M_k.

The formal Bayesian approach for comparing any two rival specifications, M_k and M_r, is through the pairwise Bayes factor, defined as the ratio of marginal likelihoods:

$$B_{k,r} = \frac{m(y|M_k)}{m(y|M_r)},$$

which can also be interpreted as the posterior odds ratio the two models, when they are assumed to be equally likely a priori.

Taking logarithms of (2.10) and evaluating this function at some hight density point θ_k^*, such as the mean of the posterior density $\pi(\theta_k|y, M_k)$, we have:

$$\log m(y|M_k) = \log f(y|M_k, \theta_k^*) + \log \pi(\theta_k^*|M_k) - \log \pi(\theta_k^*|y, M_k) \tag{2.11}$$

The conditional likelihood appearing as the first term on the right hand side is evaluated with the support of the Kalman filter for the linear Gaussian homoscedastic local level model (M_1); for the other specifications, featuring stochastic volatility in at least one of the components, it is evaluated by sequential Monte Carlo methods (particle filters). Full details are provided in Appendix B. The second component is simply the product of the prior distribution for the parameters of each model. The last component, i.e. the normalized posterior density of the parameters, requires a specialized treatment. In Appendix C we provide the relevant details for its estimation, with particular reference to UC-SV specification.

Table 2.2: *Marginal likelihood for UC models of U.S. inflation.*

Models	$\log f(y\|M_k, \theta_k^*)$	$\log \pi(\theta_k^*\|M_k)$	$\log \pi(\theta_k^*\|y, M_k)$	Total
UC	-369.56	-11.48	-0.12	-380.93
UC-SVt	-367.80	-8.83	7.24	-383.87
UC-SVc	-366.71	-2.51	-13.5	-355.72
UC-SV	-356.10	-3.06	20.81	-379.96

The results, reproduced in Table 2.2, clearly point out that the model that performs best is the local level model with stochastic volatility in the permanent component. The variation in the

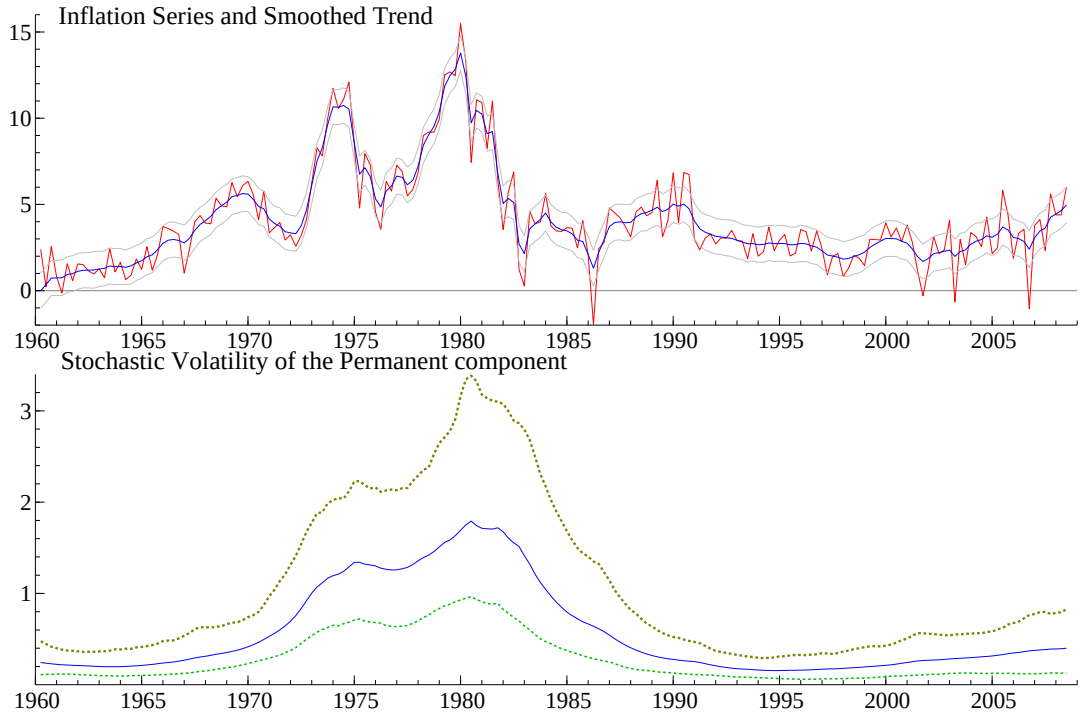


Figure 2.4: *Upper: Quarterly inflation and its posterior mean level; Bottom: Volatility of the permanent component with 95% percent credible interval.*

transitory one is by and large insignificant. The UC-SV has the highest conditional likelihood, but receives a high "penalty" from the term $\log \pi(\theta_k^*|y, M_k)$. As a result the posterior odds of model UC-SV against UC-SVc are close to zero. Hence, we conclude that the model with two stochastic volatility components is likely to over-fit the data.

Hence, our preferred model is the UC-SVc specification; table 2.3 and figures 2.4-2.5 report the main estimation results for this model. In particular, figure 2.4 displays the posterior mean of the permanent component, along with the 95% credible intervals. The bottom panel displays the posterior mean and the interval estimates of the process $\sigma_{\eta t}$. The plot illustrates that the volatility of the permanent component is subject to a steep decline in the years 1982-1995, whereas the trend is reversed after 1995. The first panel of figure 2.5 displays the evolution of the posterior mean of the signal to noise ratio, $\sigma_{\eta t}^2/\sigma_\varepsilon^2$. The persistence parameter, plotted in the top right panel of figure 2.5, declined during the great moderation, but has been increasing since 1995. The trend in predictability (see the bottom panels of figure 2.5) is the mirror image of persistence: predictability increases during the great moderation, but declines at the end of the sample.

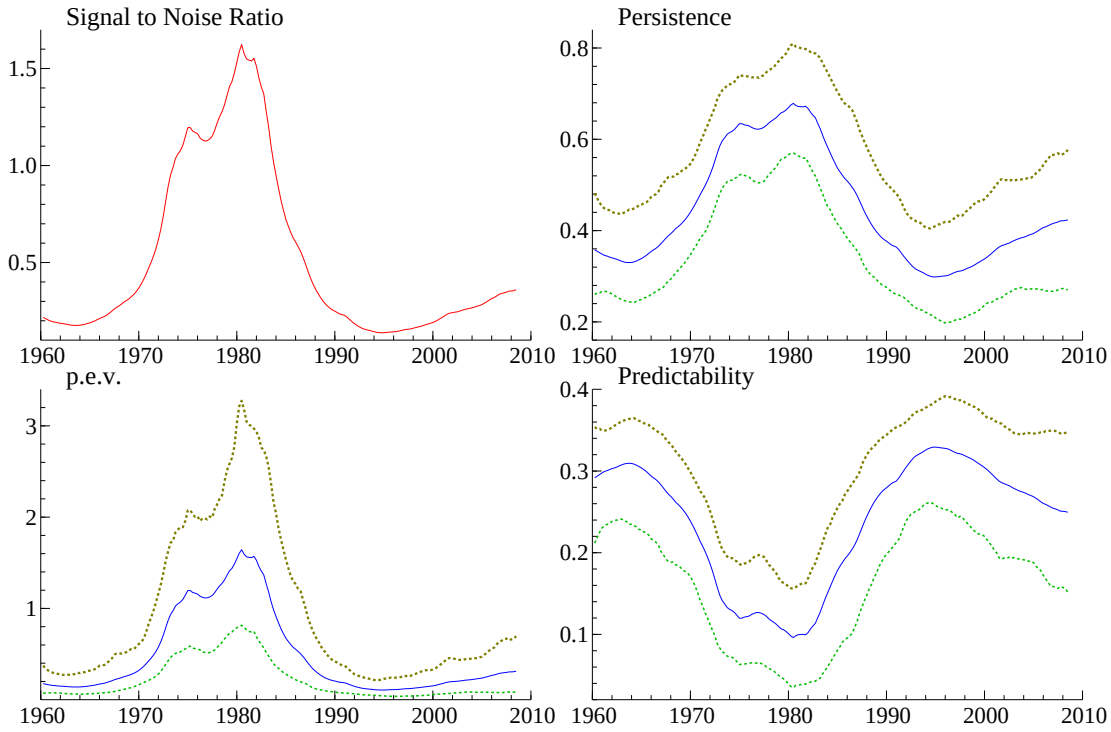


Figure 2.5: *Upper left: Signal to noise ratio; Upper Right: Persistence Parameter with 95 percent credible interval. Bottom left: Prediction error variance with 95 percent credible interval; Bottom right: Predictability with 95 percent credible interval.*

Finally, figure 2.6 displays the nonparametric estimates of the posterior density of the parameters of the permanent volatility process and the irregular variance, and table 2.3 presents summary statistics concerning the distribution of the parameters and the convergence of the MCMC sampling scheme. We notice in particular that the Geweke's convergence diagnostics are fully satisfactory.

Table 2.3: *Posterior mean, median, Geweke's statistic and Inefficiency factor for UC-SVc model.*

Parameters	Mean	Median	Geweke's G_C	INEF
μ_2	-0.0233	-0.0229	-0.93	26.61
ϕ_2	0.9791	0.9801	-0.13	69.74
$\sigma^2_{\kappa_2}$	0.0472	0.0463	-0.82	32.16
σ^2_{ε}	1.2509	1.2403	1.80	135.00

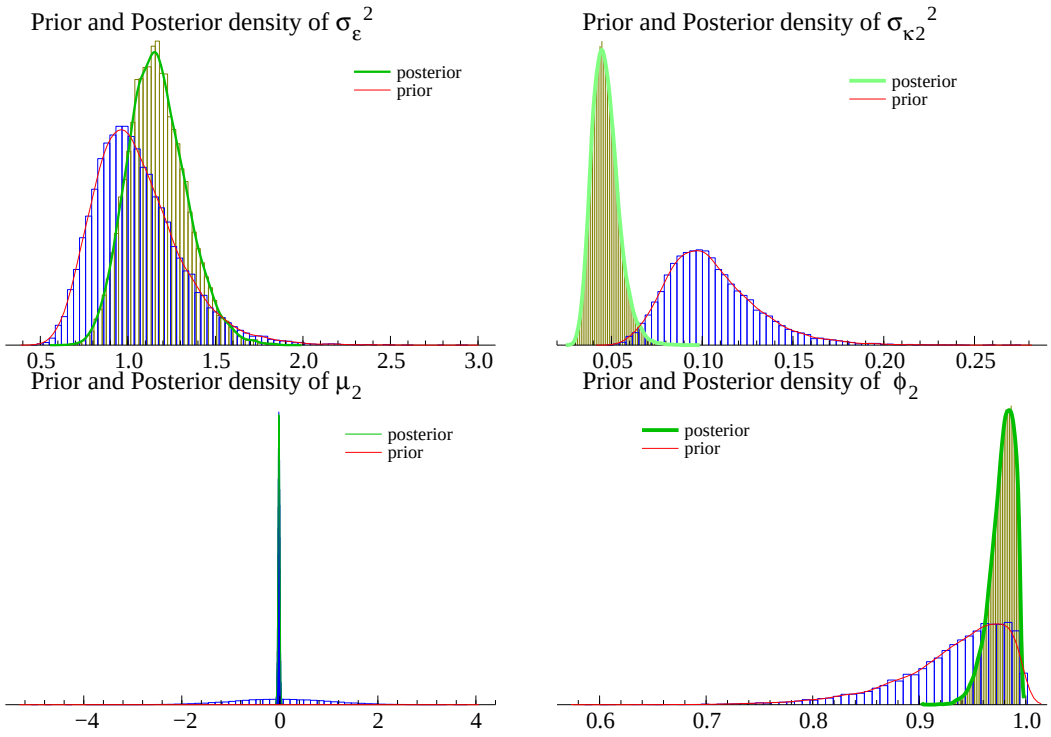


Figure 2.6: *Upper graph: Posterior density of the autoregressive parameter of the stochastic volatility process. Middle graph: Simulation against iteration for the variance of the stochastic volatility process. Bottom graph: Simulation against iterations for the constant in the stochastic volatility process.*

2.8 Conclusion

The paper has provided a full Bayesian analysis of the local level model with stochastic volatility proposed by Stock and Watson (2007) for the U.S. quarterly CPI inflation rate. The model provides a simple yet effective decomposition of U.S. inflation into a permanent component and a transitory one, with stochastic volatility in the disturbances driving the two components. Bayesian model selection enabled us to conclude that inflation's volatility is subject to significant changes over time, but the volatility affects only the permanent disturbances, not the transitory component.

The volatility of the permanent has been decreasing substantially after 1982, reaching a minimum around 1995, but has been increasing ever since, albeit at a small rate. The estimated volatility pattern support the view that a turning point took place around the mid-90ies and the great moderation is likely to be over. As correctly pointed out by a Referee, this result deserves further investigation as for its economic interpretation and implications. There are to possible explanation as to why it went undetected in previous analyzes: first and foremost, previous studies were conducted on a much shorter sample; for instance, the sample period consider by Stock and Watson (2007) ended in the fourth quarter of 2004, whereas our series ends in the 3rd quarter of 2008. The series, displayed in figure 2.1, does indeed display higher volatility at the end of the sample. Secondly, there are two substantial differences in model specification and estimation, that may play a role: on the one hand, our final specification, suggested by Bayesian model selection,

is such that the volatility of the transitory component is constant. Also, the parameters of the permanent disturbance volatility process are estimated, rather than fixed.

The persistence implied by the model has been decreasing during the years of the great moderation and it stayed at historical lows up to the mid-90ies. Recently, persistence has been rising again. Correspondingly, the predictability of inflation increased during the great moderation up to maximum occurring around 1995 and it has been going down ever since.

2.9 APPENDIX A: Metropolis - within - Gibbs sampling scheme

This Appendix illustrates the prior and posterior distributions used in our analysis. For the prior distribution we assume an independent structure between each block of variables and within each block so that $\pi(\theta, \alpha, h_1, h_2) = \pi(\theta)\pi(\alpha)\pi(h_1)\pi(h_2)$, and, for instance,

$$\pi(\theta) = \pi(\mu_1|c_1, d_1)\pi(\mu_2|c_2, d_2)\pi(\phi_1|a_1, b_1)\pi(\phi_2|a_2, b_2)\pi(\sigma^2_{\kappa_1}|\gamma_1, \beta_1)\pi(\sigma^2_{\kappa_2}|\gamma_2, \beta_2).$$

The prior distributions and their hyperparameters are reported in table 2.4.

Table 2.4: *Specification of the prior distributions*

θ	Prior	Hyperparameters	
μ_i	$N(c_i, d_i^2)$	$c_i = 0.00$	$d_i = 10.00$
ϕ_i	$Beta(a_i, b_i)$	$a_i = 20.50$	$b_i = 1.50$
$\sigma^2_{\kappa_1}$	$IG(\gamma_1, \beta_1)$	$\gamma_1 = 20.00$	$\beta_1 = 0.20$
$\sigma^2_{\kappa_2}$	$IG(\gamma_2, \beta_2)$	$\gamma_2 = 20.00$	$\beta_2 = 0.20$

The posterior densities are available in closed form for the core level of inflation (for which samples are drawn by a multimove sampler known as the simulation smoother, here implemented according to the algorithm presented in Durbin and Koopman (2002)), and for some elements of the vector θ for which we can exploit conditional conjugacy.

1. Given the choice of the prior distribution, the full conditional density of the parameter ϕ_1 (and similarly ϕ_2) is not available in closed form; therefore, to sample from the full conditional we employ a Metropolis-Hasting sampling algorithm, similar to the one described in Kim et al.(1998), which enforces the stationarity of the stochastic volatility process. Another possibility is to use a random walk Metropolis-Hasting that can be sketched as follows: if $\phi_i^{(j-1)}$ denotes the current value of the chain at the j-th iteration, we sample a new proposal $\phi_i^{(j)} = \phi_i^{(j-1)} + w_j$, where w_j is drawn a normal distribution with mean 0 and variance 0.1. If the proposal is within the stationary region then it is accepted with probability $\min\{1, g(\phi_i^{(j)})/g(\phi_i^{(j-1)})$, where

$$g(\phi_i) = \pi(\phi_i)f(h_i|\mu_i, \phi_i, \sigma^2_{\kappa_i})$$

and, apart from a constant term,

$$\log f(h_i|\mu_i, \phi_i, \sigma^2_{\kappa_i}) = -\frac{h^2_{i,0}}{2\sigma^2_{\kappa_i}} + \frac{1}{2}\log(1-\phi_i^2) - \frac{\sum_{t=1}^{n-1}(h_{i,t+1} - \phi_i h_{i,t} - \mu_i)^2}{2\sigma^2_{\kappa_i}}. \tag{2.12}$$

2. Using a Normal prior, the full conditional distribution of the parameters μ_i is $N(\hat{C}_i, \hat{D}_i)$ where:

$$\hat{C}_i = \hat{D}_i\left(\frac{C_i}{D_i^2} + \frac{1}{\sigma^2_{\kappa_i}}\sum_{t=1}^{T}(h_{i,t} - \phi_i h_{i,t-1})\right), \qquad \hat{D}_i = \left(\frac{1}{d_i^2} + \frac{T}{\sigma^2_{\kappa_i}}\right)^{-1}. \tag{2.13}$$

3. Using a conjugate Inverse Gamma prior, the full conditional of the variances of volatility processes are:

$$\sigma^2_{\kappa_i}|y,\alpha,h_i,\phi_i,\mu_i \sim IG\left\{\frac{n}{2}+\alpha_i,\beta_i+\frac{h^2_{i,0}+\sum_{t=1}^{n-1}(h_{i,t+1}-\mu_i-\phi_i h_{i,t})^2}{2}\right\}$$

4. To sample from $h_{1t}|h_{1,t-1},h_{1,t+1},y_t,\alpha_t,\theta$, we adopt the single move Metropolis-Hastings simulation step, based on the factorization:

$$f(h_{1t}|h_{1,t-1},h_{1,t+1},y_t,\alpha_t,\theta) \propto f(h_{1t}|h_{1,t-1},h_{1,t+1},\theta)f(y_t|\alpha_t,h_{1t}). \tag{2.14}$$

It can be shown that

$$f(h_{1t}|h_{1,t-1},h_{1,t+1},\theta) = f(h_{1t}|h_{1,t-1},\theta)f(h_{1,t+1}|h_{1t},\theta) \tag{2.15}$$

is a Gaussian density with mean

$$h^*_{1t} = \frac{\mu(1-\phi)+\phi(h_{1,t-1}+h_{1,t+1})}{(1+\phi_i^2)}$$

and variance

$$v_i^2 = \frac{\sigma^2_{\kappa_i}}{1+\phi_i^2}$$

(see Jacquier, Polson and Rossi, 1994). Random Walk proposals $h^{(j)}_{1t}$ can be made from this Gaussian density; their acceptance probability is $\min\{1, g(h^{(j)}_{1t})/g(h^{(j-1)}_{1t})\}$, where

$$\begin{aligned} g(h_{1t}) = &\exp\left[-\left\{\frac{(h_{1,t+1}-\mu_1-\phi_1 h_{1,t})^2}{2\sigma^2_{\kappa_1}}+\frac{(h_{1,t}-\mu_1-\phi_1 h_{1,t-1})^2}{2\sigma^2_{\kappa_1}}\right\}\right]\times \\ &\times\frac{1}{\exp(h_1/2)}\exp\left[-\frac{(y_t-\alpha_t)^2}{2\exp(h_1)}\right] \end{aligned} \tag{2.16}$$

for $t=1,\ldots,n$, whereas

$$g(h_{1,0}) = \exp\left\{-\frac{(h_{1,1}-\mu_1-\phi_1 h_{1,0})^2}{2\sigma^2_{\kappa_1}}-\frac{(1-\phi_1^2)h^2_{1,0}}{2\sigma^2_{\kappa_1}}\right\},$$

and, for $t=n$,

$$g(h_{1,n}) = \exp\left\{-\frac{(h_{1,n}-\mu_1-\phi_1 h_{1,n-1})^2}{2\sigma^2_{\kappa_1}}\right\}.$$

A similar sampling scheme is adopted for h_{2t}.

2.10 APPENDIX B: Auxiliary Particle Filter

For evaluating the conditional likelihood, $f(y|\theta_k, M_k)$, for the SV specifications, we implemented an auxiliary particle filter (see Pitt and Shephard, 1999). The latter estimates the one-step-ahead predictive densities which enter the factorization: $f(y|\theta_k, M_k) = \prod_t f(y_{t+1}|Y_t, \theta_k, M_k)$, where $Y_t = \{y_1, \ldots, y_t\}$, and the predictive density is evaluated by sequential Monte Carlo methods as follows:

$$f(y_{t+1}|Y_t, \theta_k, M_k) = \frac{1}{M}\sum_{i=1}^{M} w_{1,t}^{(i)} \times \frac{1}{R}\sum_{j=1}^{R} w_{2,t}^{(j)}. \tag{2.17}$$

Here M denotes the number of particles, $w_{1,t}$ are the so-called first stage weights, R is the number of daughter particles (see below) and $w_{2,t}$ are the so-called second stage weights.

All the inferences will be conditional on (θ_k, M_k); henceforth, for notational simplicity we will omit these conditioning elements. After initializing the weights $w_{1,0} = \frac{1}{M}$ and drawing samples $z_0^{(i)}, i = 1, \ldots, M$, from the initial distribution of the random vector $z_t = (\alpha_t, h_{1,t}, h_{2,t})$, at time $t = 0$, with

$$\alpha_0 \sim N(0, 1000) \qquad h_{1,0} \sim N\left(0, \frac{\sigma_{\kappa_1}^2}{1-\phi_1^2}\right) \qquad h_{2,0} \sim N\left(0, \frac{\sigma_{\kappa_2}^2}{1-\phi_2^2}\right), \tag{2.18}$$

we iterate, for $t = 1...T$, the following steps:

1. Set the first stage weights, $w_{1,t} \equiv \frac{1}{M}$

2. Predict the unobserved states one-step-ahead, and update the weights, by

$$\begin{aligned} \bar{z}_{t+1}^{(i)} &= E(z_{t+1}|z_t^{(i)}) \\ w_{1,t}^{(i)} &= w_{1,t}^{(i)} \times f(y_{t+1}|\bar{z}_{t+1}^{(i)}) \end{aligned} \tag{2.19}$$

   where $f(y_{t+1}|\bar{z}_{t+1}^{(i)})$ is a Gaussian density with mean $\bar{\alpha}_{t+1}^{(i)}$ and variance $\exp(h_{1,t+1})$. The $w_{1,t}^{(i)}$ are the first stage weights described in Pitt and Shephard(1999).

3. Resample the particles $z_t^{(i)}$ with replacement R times (by multinomial resampling). Let $\tilde{z}_t^{(i)}$ denote the resampled particles.

4. Sample $z_{t+1}^{(i)}, i = 1, \ldots, R$, from $z_{t+1}|\tilde{z}_t^{(i)}, y_{t+1}$, using the approach by Godsill and Clapp(2001), which is based on the factorization:

$$f(z_{t+1}|z_t, y_{t+1}) = f(\alpha_{t+1}|h_{2,t+1}, \alpha_t, y_{t+1}) f(h_{1,t+1}|h_{1,t}) f(h_{2,t+1}|h_{2,t}) \tag{2.20}$$

   where $f(h_{j,t+1}|h_{j,t}), j = 1, 2$, are Gaussian densities with mean $\mu_j + \phi_j h_{j,t}$ and variance $\sigma_{\kappa_j}^2$, and

$$\alpha_{t+1}|h_{2,t+1}, \alpha_t, y_{t+1} \sim \mathrm{N}(m, S^2)$$

   with

$$S^2 = \left(\frac{1}{\exp(h_{2,t+1})} + \frac{1}{\exp(h_{1,t+1})}\right)^{-1} \qquad m = S^2\left(\frac{y_{t+1}}{\exp(h_{1,t+1})} + \frac{\alpha_t}{\exp(h_{2,t+1})}\right) \tag{2.21}$$

5. Compute the second stage weights:

$$w_{2,t}^{(i)} = \frac{f(y_{t+1}|z_{t+1}^{(i)})f(z_{t+1}^{(i)}|\tilde{z}_t^{(i)})}{f(y_{t+1}|\tilde{z}_{t+1}^{(i)})f(\tilde{z}_{t+1}^{(i)}|\tilde{z}_t^{(i)})} \tag{2.22}$$

6. Resample M particles by multinomial resampling, with probabilities proportional to $w_{2,t}^{(i)}$.

7. Go to step 1.

2.11 APPENDIX C: Chib and Jeliazkov algorithm

This Appendix illustrates the steps of the Chib and Jeliazkov (2001) algorithm that are necessary to estimate the posterior density $\pi(\theta|y)$ for the UC-SV model at a high density point θ^*. The latter is the component of the basic marginal likelihood identity that is not automatically available from the MCMC output.

The estimate is constructed as follows: denoting $\theta = \{\theta_j, j = 1, \ldots, J\}$ the vector containing the hyperparameters, where the elements of the vector θ are $\{\mu_1, \phi_1, \sigma_{\kappa_1}^2, \mu_2, \phi_2, \sigma_{\kappa_1}^2\}$, consider the factorization of the joint conditional density:

$$\hat{\pi}(\theta^*|y) = \prod_{j=1}^{J} \hat{\pi}(\theta_j^*|y, \theta_1^*, \ldots, \theta_{j-1}^*)$$

Further, let $z = (h_1, h_2, \alpha)$.

The Chib and Jeliazkov (2001) algorithm takes the following steps:

- From the MCMC sample evaluate the posterior mean of μ_1 and set μ_1^* equal to this value. A Monte Carlo estimate of the first multiplicative factor, $\pi(\theta_1^*|y) = \pi(\mu_1^*|y)$, is obtained from the output of the MCMC sampling scheme by the technique known as Rao-Blackwellization.

- For estimating $\pi(\theta_2^*|y, \theta_1^*) = \pi(\phi_1^*|y, \mu_1^*)$ run a reduced Metropolis-Hastings within Gibbs chain for the following subset of parameters $\{\phi_1, \sigma_{\kappa_1}^2, \mu_2, \phi_2, \sigma_{\kappa_2}^2, z\}$, where the value of μ_1 is kept fixed at μ_1^*.

- Estimate the value of the density $\pi(\theta_2^*|y, \theta_1^*) = \pi(\phi_1^*|y, \mu_1^*)$, using the following steps:

  1. Simulate G draws from the posterior of $\{\phi_1^{(g)}, \sigma_{\kappa_1}^{2,(g)}, \mu_2^{(g)}, \phi_2^{(g)}, \sigma_{\kappa_2}^{2,(g)}, z^{(g)}\}, g = 1, \ldots, G$, by the same MCMC methods presented in appendix A, conditional on μ_1^*.
  2. Compute the posterior mean of ϕ_1 by averaging across the draws $\phi_1^{(g)}$ and denote it ϕ_1^*.
  3. Include ϕ_1^* in the conditioning set and sample J draws from the conditional distibutions:

$$\begin{aligned}
&\pi(\sigma_{\kappa_1}^2|y, z, \phi_1^*, \mu_1^*, \mu_2, \sigma_{\kappa_2}^2, \phi_2), \quad \pi(z|y, \sigma_{\kappa_1}^2, \mu_1^*, \phi_1^*, \mu_2, \phi_2, \sigma_{\kappa_2}^2),\\
&\pi(\mu_2|y, z, \mu_1^*, \phi_1^*, \sigma_{\kappa_1}^2, \phi_2, \sigma_{\kappa_2}^2), \quad \pi(\phi_2|y, z, \mu_1^*, \phi_1^*, \sigma_{\kappa_1}^2, \mu_2, \sigma_{\kappa_2}^2),\\
&\pi(\sigma_{\kappa_2}^2|y, z, \mu_1^*, \phi_1^*, \sigma_{\kappa_1}^2, \mu_2, \phi_2).
\end{aligned}$$

These iterations provide the sample $\{\sigma_{\kappa_1}^{2(j)}, \mu_2^{(j)}, \phi_2^{(j)}, \sigma_{\kappa_2}^{2(j)}, z^{(j)}\}_{j=1}^{J}$. Furthermore, at each iteration we generate

$$\phi_1^{(j)} \sim q(\phi_1^*, \phi_1 | y, z^{(j)}, \mu_1^*, \sigma_{\kappa_1}^{2,(j)}, \mu_2^{(j)}, \phi_2^{(j)}, \sigma_{\kappa_2}^{2,(j)})$$

where $q(\theta_j, \theta_j' | u)$ is the proposal density for the transition from θ_j to θ_j' conditional on u. As a result, the collection $\{\phi_1^{(j)}, \sigma_{\kappa_1}^{2(j)}, \mu_2^{(j)}, \phi_2^{(j)}, \sigma_{\kappa_2}^{2(j)}, z^{(j)}\}_{j=1}^{J}$ is are multiple (correlated) draws from the distribution:

$$\pi(\sigma_{\kappa_1}^2, \mu_2, \phi_2, \sigma_{\kappa_2}^2, z | y, \mu_1^*, \phi_1^*) \times q(\phi_1^*, \phi_1 | y, z, \mu_1, \sigma_{\kappa_1}^2, \mu_2, \phi_2, \sigma_{\kappa_2}^2).$$

4. Denoting the probability of a move by

$$\psi(\phi_1, \phi_1' | u) = \min \left\{ 1, \frac{f(y|\phi_1^*, \varsigma, z)\pi(\phi_1^*, \varsigma)}{f(y|\phi_1^{(g)}, \varsigma, z)\pi(\phi_1^{(g)}, \varsigma)} \frac{q(\phi_1^*, \phi_1^{(g)} | y, \varsigma, z)}{q(\phi_1^{(g)}, \phi_1^* | y, \varsigma, z)} \right\},$$

where ς is the collection of parameters $(\mu_1^*, \sigma_{\kappa_1}^2, \mu_2, \phi_2, \sigma_{\kappa_2}^2)$. The required marginal density at ϕ_1^* can now be estimated as

$$\hat{\pi}(\phi_1^* | y) = \frac{G^{-1} \sum_g \psi(\phi_1^{(g)}, \phi_1^* | y, z^{(g)}, \mu_1^*, \sigma_{\kappa_1}^{2(g)}, \mu_2^{(g)}, \phi_2^{(g)}, \sigma_{\kappa_2}^{2(g)}) \cdot q(\phi_1^{(g)}, \phi_1^* | y, z^{(g)}, \mu_1^*, \sigma_{\kappa_1}^{2(g)}, \mu_2^{(g)}, \phi_2^{(g)}, \sigma_{\kappa_2}^{2(g)})}{J^{-1} \sum_j \alpha(\phi_1^*, \phi_1^{(j)} | y, z^{(j)}, \mu_1^*, \sigma_{\kappa_1}^{2(j)}, \mu_2^{(j)}, \phi_2^{(j)}, \sigma_{\kappa_2}^{2(j)})}$$

- Run a reduced Gibbs sampling scheme on the following parameters $\{\sigma_{\kappa_1}^2, \mu_2, \phi_2, \sigma_{\kappa_2}^2, z\}$ and calculate $\sigma_{\kappa_1}^{2,(*)}$
- Run a reduced Gibbs sampling scheme and calculate the ϕ_2^* with the same procedure describe before noticing that the $\phi_1^*, \mu_1^*, \sigma_{\kappa_1}^{2,(*)}$ are fixed.
- Run a reduced Gibbs sampling scheme on the following parameters $\{\mu_2, \sigma_{\kappa_2}^2, z\}$ and calculate μ_2^*;
- Run a reduced sampling scheme Gibbs on the following parameters $\{\sigma_{\kappa_2}^2, z\}$ and calculate $\sigma_{\kappa_2}^{2,(*)}$

Chapter 3

Missing values in dynamic factor models

3.1 Introduction

It has long been observed that many countries experienced similar fluctuations in macroeconomic aggregates and that these fluctuations exhibit substantial synchronization across countries, see Moore and Zarnowitz (1986) for an early survey. Similarities in aggregate fluctuations in different countries has been interpreted as a challenge to economic theory, suggesting the development of business cycle theories which focus on the functioning of the market economies in general, rather than on individual countries. The similarity of cycles across countries and across different Regional areas is an important issue. If, for example, the business cycle is a worldwide phenomenon then it is not responsive to domestic shocks like country monetary policy. Although there is a convincing empirical evidence in favor of international business cycle linkages, see Gregory et al. (1997) and Kose et al. (2003), economists and econometricians still dispute on the causes, the consequences, as well as even on the measure of these comovements. For instance, the question concerning the predominance role of common shocks or common propagation mechanisms is far from being resolved. Stock and Watson (2005) use a factor-augmented vector autoregression (FAVAR) estimated over a pre and post-1983 subsample to investigate whether international or domestic shocks are the source of the decline in volatility. They find the presence of an emerging European factor and an emerging English-speaking group. Moreover for Japan, the sensitivity of international business cycle decline sharply but the variance of the domestic shock increased during time. Other studies like Heathcote and Perri (2004) found that the cross-country correlation among G-7 countries has declined over time, similar results has been reported by Del Negro and Otrok (2008). These studies are somehow limited due to *data availability* (missing values), and *econometric intractability* (inefficient estimation due to the presence of a lot of parameters), for this reasons they are focused on the analysis of a small group of countries, or to world aggregates. To overcome the second problem Kose et al. (2004) and Kose et al. (2008) using Bayesian dynamic factor model studied the dynamic comovement of macroeconomic aggregates in a broad cross section of countries. They provide an analysis of world, regional, even country factors, showing that the world factor is an important source of volatility for aggregates in most countries (world business cycle) and the region-specific factors play only a minor role in explaining fluctuations in economic activity. With this approach

they can handle many more series, but, they did not provide and efficient treatment of missing data, that it is a standard characteristic in economic time series.

We use a new maximum likelihood estimation method for high-dimensional dynamic multi-factor models in presence of large amount of missingness, to study the business cycle in a big unbalanced panel composed by 150 countries. Recently there has been an increasing interest in likelihood-based approach for the estimation of large scale data set. Since factors are explicitly modelled and the estimation method takes into account the model specification, the factor can represent aspects of economic theory. Doz et al. (2007) show, under mild conditions, that the estimates of the unobserved factors obtained from a likelihood-based analysis are consistent estimators for the true factors when $T \to \infty$ and $N \to \infty$, where N is the number of series and T is the number of observations, even if the dynamic factor model is misspecified. Furthermore, they present evidence that in some cases a likelihood-based analysis produces more precise estimates of the factors than a principal component method. In a recent paper Jungbacker and Koopman (2008) proposed a new method for likelihood based analysis for dynamic factor models; they demonstrated that when $N > q$, where q is the number of factors, the computational efficiency of Kalman filter and smoother can significantly be improved by a simple computational device based on the projection of the data on a reduced dimensional factor space. This device has been later extended to deal with missing data by a suitable state space formulation, see Jungbacker et al. (2009).

The model is estimated by maximum likelihood and therefore has a well-articulated statistical foundation for the estimates and inference under the maintained assumptions, nevertheless the estimation of a high-dimensional unbalanced panel with $N > T$, is a very challenging task. In contrast, a method like principal components is much simpler computationally but is unable to isolate jointly world, regional and country-specific factors in a large scale unbalanced panel.

We mainly extend the research program on global business cycle in three dimension. First our study is much more comprehensive then other studies as we use a large data set (150 countries) with a longer time span 1950-2007 allowing for missing values. Second, unlike most existing studies, we specifically consider the role played by the global factor and distinguish them from common cycle to each specific region. Finally we employ a new econometric tool that allow the estimation of all these quantities in a fast and efficient way even in presence of a large amount of missingness.

Our results can be summarized as follows. We analyze the same data set of Kose et al. (2004) and we get virtually the same conclusion. Then we extend the study to 150 countries and we provide a study of global, regional and country specific component. By computing the shares of the variances of the aggregate variables accounted for by fluctuations in the various dynamic factors, we quantify the aggregate fluctuations in each of the 150 countries. Although this kind of study has been proposed before as far as we know this is the first time that this exercise is carry out in a big data set, this allow us to provide new results about European accessing countries (Poland, Hungary, Romania and Bulgaria) and emerging economies like China and India. According to our calculation China and India are not related to the global and regional factor, these *emerging countries*, have *decoupled* from industrial economies, in the sense that their business cycles are not linked to them. This result is confirmed by the impressive growth performance of China and India, in fact, they seems not to have been affected by the growth slowdown in a number of industrialized countries.

We confirm the results of Kose et al. (2008) that there has been a decline over time in the relative importance of the global factor. Finally we provide new results about regional factors and we show that there is evidence of business cycle convergence within each industrialized region,

providing support for the standard neoclassical growth theory (Solow, 1956), and a divergence between rich and poor regions. The remainder of the paper is organized as follow.

The empirical questions and the data set are review in Section 3.2. The model specification and the estimation strategy are presented in Section 3.3. In Section 3.4 we apply this new technique to a different dimension data set. We use R^2 and Box-Ljung statistic to asses the accuracy of our model. Furthermore we provide the variance decomposition to measure the relative contributions of the world, region, and country factors. Finally we report some new results about world and regional convergence. Section 4.7 concludes the paper. The data set is presented in the Appendix A. The state space form and the derivatives useful for analytical maximum likelihood evaluation are in Appendix B.

3.2 A world economy: empirical questions and data set

3.2.1 Open questions

The phenomenon of globalization and the synchronized slowdown of major world economies alongside the contagious nature of financial crisis in emerging markets have recently incited the interest in understanding the propagation and synchronization of business cycle fluctuations across national borders as well as their evolution. Identify the degree of synchronization and analyzing the global, regional as well as country-specific determinants of international fluctuations are relevant from a number of perspectives including shot-run domestic and international policy coordination as well as assessment of the long-run feasibility of monetary unions. There are still a number of unanswered questions regards the technique and the data set used in those studies and our objective is to provide a comprehensive empirical characterization of global, regional and country specific factors. We focus on the following questions. What about the global convergence with a big number of countries? Do we have different results with respect to other studies if we use a differ data set with longer time span?

Stock and Watson (2005) using a Factor-Augmented vector autoregression (FAVAR) and Del Negro and Otrok (2008) using a Bayesian approach provided evidence of an emerging European business cycle and an English-speaking group. Can we confirm, using an exact maximum likelihood framework, these results?

In the literature it is still not deeply analyzed the effect that European union have on accession countries like Poland. Some studies like Artis et al. (2005) find that between the accession countries and the euro area the synchronization indicators are generally rather low, with the exception of Poland and Hungary. Can we confirm these results? What is the main source of the variability of those countries is it the global, the regional or the country specific component?

Shin and Wang (2003) pointed out that as trade integration deepens in East Asia, it is expected that there will be closer links in business cycles among East Asia countries. What can we say about the East Asia countries ? Are they more synchronized?

Finally what can we say about the poorest countries? Are they more related or detached to the world factor? Is there evidence of an emerging regional areas or is the country specific component the main source of variability?

The answers to these questions have important implications about the debate of global and regional convergence.

3.2.2 Data description

The data set used in our study is taken from Penn World Table (PWT 2009) freely available from the web side *http://pwt.econ.upenn.edu/*. The observations are Annual Real GDP per capita Constant Price: Laspeyres (base year 2005) and span the period between 1950 and 2007. The time series extracted are 150 with 57 observations available at the most for each series, this leads to an arbitrary pattern of missing data. After the logarithm transformation the series are assumed to be I(1), so we differentiate and standardize all of them. We do this in order to ensure that all series receive equal *weight* in the search of common factors. In addition it has been shown in Head (1995) that the volatilities of aggregates after the log first-difference vary systematically with country size, the smaller countries' series tend to be more volatile, these countries series would receive higher weigh as well. By standardizing the variance, we ensure that all the series are treated symmetrically and the econometric procedure that extracts common components does not distinguish between a 2-percent growth rate in the United States and a 2-percent growth rate in China, see Gregory et al. (1997) and Kose et al. (2004). We further analyze this issue for the G7 countries in Section 4. All the countries with the available sample are reported in Appendix A and they are divided in 6 different areas accordingly to Table 3.1. This division is somehow different from that proposed by Kose et al. (2004) because we include the Oceania countries, divided accordingly to developed and developing, in the Asia and Oceania Developed and Asia and Oceania Developing and Poor regions. In the following we assume the presence of one global factor (f^{global}) and six regional factors ($f^{regional}$ e.g. one each for North America, Latin America, Europe, Asia and Oceania Developed, Asia and Oceania Developing and Poor and Africa). Thus for the series i we have:

$$y_{i,t} = \lambda_i^{world} f_t^{world} + \lambda_i^{region} f_{r,t}^{region} + u_{i,t}, \quad t = 1, \ldots, T, \tag{3.1}$$

where $u_{i,t}$ are the *country specific* component, r is the region , $f_t^{(\cdot)}$ indicates the global and regional factors and $\lambda_i^{(\cdot)}$ are the global and regional factor loadings.

3.3 Dynamic Factor Model: specification and estimation

The estimation of the global and regional factors in a big unbalanced panel is a very challenging task. We address this and the related issues by implementing the maximum likelihood estimation for the state space form recently proposed by Jungbacker et al. (2009). Subsection 3.3.1 describes briefly the dynamic factor model and the strategy to disentangle the global and the regional factors. Subsection 3.3.2 presents the likelihood for unbalanced panel. Subsection 3.3.3 present the numerical performance of our algorithm. The state space form and the analytical derivatives are reported in Appendix B.

3.3.1 Dynamic Factor Model and Block Structure in the Loading Matrix

Consider a panel of N time series where we denote $y_{i,t}$ as the observation at time t in the i series, then the dynamic factor model is given by;

$$y_{i,t} = \Lambda_i^{'} f_t + u_{i,t}, \qquad t = 1, \ldots, T, \qquad i = 1, \ldots, N, \tag{3.2}$$

where Λ_i is a $q \times 1$ vector of factor loadings, f_t is a $q \times 1$ vector of unobserved common factors, $u_{i,t}$ is the *country-specific noise*, T is the number of observations and N is the number of series.

Table 3.1: *All the 146 countries divided accordingly to six different areas.*

North America	1-3				
	United States				
	Mexico				
	Canada				
Latin America	1-7	8-14	15-21	22-28	29-32
	Brazil	Jamaica	Barbados	Guatemala	Puerto Rico
	Argentina	Peru	Costa Rica	Honduras	Trinidad Tobago
	Bolivia	Paraguay	Dominica	Nicaragua	Suriname
	Cambodia	Uruguay	Dominica Republic	Panama	St. Lucia
	Chile	Venezuela	Ecuador	Bahamas	
	Colombia	Antigua	El Salvador	Bermuda	
	Cuba	Belize	Grenada	Haiti	
Europe	1-7	8-14	15-22	23-25	
	Germany	Portugal	Finland	Romania	
	France	Norway	Denmark	Poland	
	United Kingdom	Netherlands	Belgium	Bulgaria	
	Italy	Luxembourg	Austria		
	Sweden	Ireland	Cyprus		
	Switzerland	Iceland	Malta		
	Spain	Greece	Hungary		
Asia/Oceania Developed	1-4	5-8	9-12	13-15	
	Japan	Australia	Singapore	Saudi Arabia	
	China	New Zealand	Thailand	Kuwait	
	Taiwan	Malaysia	Republic of Korea	Qatar	
	Hong Kong	Turkey	United Arab Emirates	Israel	
Asia/Oceania Dev. and Poor	1-6	7-12	13-18	19-24	25-27
	Philippines	Kiribati	India	Bangladesh	Maldives
	Indonesia	Samoa	Pakistan	Iraq	Mongolia
	Brunei	Solomon Islands	Sri Lanka	Nepal	Syria
	Laos	Tonga	Mauritius	Oman	Vietnam
	Macao	Vanuatu	Iran	Bahrain	
	Papua New Guinea	Micronesia, Fed. Sts	Jordan	Bhutan	
Africa	1-10	11-20	21-30	31-40	41-47
	South Africa	Senegal	Comoros	Liberia	Sierra Leone
	Egypt	Somalia	Dem. Rep. Congo	Lesotho	Swaziland
	Morocco	Tunisia	Republic of Congo	Malawi	Tanzania
	Nigeria	Uganda	Equatorial Guinea	Mali	Togo
	Algeria	Cameroon	Gabon	Mauritania	Zambia
	Central African Republic	Botswana	Gambia	Namibia	Zimbabwe
	Cote d'Ivoire	Benin	Ghana	Niger	Sudan
	Ethiopia	Burundi	Guinea	Cape Verde	
	Madagascar	Burkina Faso	Guinea - Bissau	Mozambique	
	Rwanda	Chad	Kenya	Sao Tome and Principe	

The dynamic factor model given in (3.2) can be represented in matrix form as follows:

$$y_t = \Lambda f_t + u_t, \quad t = 1, \ldots, T, \tag{3.3}$$

where $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ is a matrix of factor loadings, f_t is a $q \times 1$ vector of unknown factors. The common factor f_t are modelled as a stationary first order vector autoregressive ($VAR(1)$) process, and the error components u_t are modelled as vector autoregressive of order one process ($VAR(1)$):

$$u_{t+1} = \phi u_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon), \tag{3.4}$$

where ϕ is an $N \times N$ diagonal matrix and the disturbance variance matrix Σ_ε is $N \times N$ diagonal matrix of unknown parameters. In Appendix B it is shown how to represent the model (3.3) and (3.4) in state space form.

The approximated factor models of Stock and Watson (1989) and Forni et al. (1998) are quite efficient in extracting factors, however they cannot be applied in situation where is necessary to impose zero restrictions on some factor loadings to identify a factor that belongs to all the series and some factors belonging to a particular group of countries. One of the main advantages of our estimation methodology is relatively easy way to impose constraints in the loading matrix that is useful to disentangle the global and regional factors. The block structure of the Λ matrix considered in this paper is:

$$\begin{pmatrix} Usa \\ Mexico \\ Canada \\ Brazil \\ \cdots \\ Germany \\ \cdots \\ \cdots \\ Sounth\ Africa \\ \cdots \end{pmatrix} = \begin{pmatrix} \lambda_1^{glob} & \lambda_1^{reg1} & 0 & 0 & 0 & 0 & 0 \\ \lambda_2^{glob} & \lambda_2^{reg1} & 0 & 0 & 0 & 0 & 0 \\ \lambda_3^{glob} & \lambda_3^{reg1} & 0 & 0 & 0 & 0 & 0 \\ \lambda_4^{glob} & 0 & \lambda_4^{reg2} & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \lambda_{36}^{glob} & 0 & 0 & \lambda_{36}^{reg3} & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \lambda_{103}^{glob} & 0 & 0 & 0 & 0 & 0 & \lambda_{103}^{reg6} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \begin{pmatrix} f_t^{glob} \\ f_t^{reg1} \\ f_t^{reg2} \\ f_t^{reg3} \\ f_t^{reg4} \\ f_t^{reg5} \\ f_t^{reg6} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \\ u_{4,t} \\ \cdots \\ u_{36,t} \\ \cdots \\ \cdots \\ u_{103,t} \\ \cdots \end{pmatrix}, \tag{3.5}$$

where λ_i^{glob} and λ_i^{reg} corresponds to the global and regional factor loadings.

The global factor loads to all the series and is associated with a pattern of co-movement among all the countries. The non zero elements, λ_i^{reg}, in the remaining columns represent the regional factor loadings, they represent the co-movement between the countries in a specific region. In case of unrestricted Λ, see section 3.4, to ensure that all parameters are identified, we set $\Lambda = (\lambda_1', \lambda_2')'$ where λ_1 is a $q \times q$ lower triangular matrix and λ_2 is a $(N - q) \times (N - q)$ full matrix.

3.3.2 Maximum Likelihood for Unbalanced Panel

The log-likelihood function for the considered model is:

$$l(y) = \log p(y_{1,t}, \ldots, y_{i,t}; \theta), \tag{3.6}$$

where $p(\cdot)$ is the Gaussian density function, $y_{i,t}$ is the observed data for the country i at time t and θ is the parameters vector. Following Jungbacker et al. (2009) and Appendix B, the likelihood of

our model can be expressed in this form:

$$l(y) = constant + l(y^L, y^{o,m}) + l(y^H), \tag{3.7}$$

where

$$y^L = \{A_t^L y_t(o_t, o_{t-1})\}, \quad y^{o,m} = \{y_t(o_t, m_{t-1})\}, \quad y^H = \{A_t^H y_t(o_t, o_{t-1})\}, \qquad t = 1, \ldots, T, \tag{3.8}$$

$y^L = \{A_t^L y_t(o_t, o_{t-1})\}$ and $y^{o,m} = \{y_t(o_t, m_{t-1})\}$ corresponds, respectively, to the reduced part of observed values at time t and $t-1$ and the observed values at time t but missing at time $t-1$. The likelihood related to this two components is evaluated with the Kalman Filter. Kalman Filter is not applied to $y^H = \{A_t^H y_t(o_t, o_{t-1})\}$, and this *partial* likelihood is calculated accordingly to:

$$l(y^H) = constant - \frac{1}{2}\sum_{t=1}^{T} \log\left(|\Sigma_{\varepsilon,t}|\right) - \frac{1}{2}\sum_{t=1}^{T} e_t^{'}\Sigma_{\varepsilon,t}^{-1} e_t, \tag{3.9}$$

where

$$e_t = (I - \Sigma_{\varepsilon,t} A_t^{L'} A_t^L [y_t(o_t, o_{t-1}) - \phi_t^{(o)} y_{t-1}(o_t, o_{t-1})] \quad t = 1, \ldots, T, \tag{3.10}$$

and A_t^L is a transformation matrix, see for a detailed discussion Jungbacker and Koopman (2008, Lemma 2).

The maximization of this likelihood with respect to the parameter vector θ, involves a high dimensional maximization problem. Large scale optimizations problems are solved, in general, by quasi-Newton type algorithm as described in Nocedal and Wright (1999). These algorithms require the evaluation of the log-likelihood function, $l(y)$, and the score vector at each iteration. Due to the high dimension of the parameters space numerical derivatives are not feasible, fortunately, analytical expression for the score is available and the derivation is provided in Appendix B. Following Koopman and Shephard (1992), we can write the likelihood for our model in the following way:

$$\begin{aligned}
\log f(y,\theta) = &-\frac{1}{2}\sum_{t=1}^{T} \log \mid H_t(\theta) \mid + \log \mid Q_t(\theta) \mid \\
&-\frac{1}{2}\sum_{t=1}^{T}\left[H_t(\theta)^{-1}\{(y_t - c_t - Z_t\alpha_{t|n})(y_t - c_t - Z_t\alpha_{t|n})^{'} + Z_t P_{t|n} Z_t^{'}\}\right] \\
&-\frac{1}{2}\sum_{t=1}^{T}\left[Q_t(\theta)^{-1}\{\eta_{t|n}\eta_{t|n}^{'} + P_{t|n} - T_t P_{t-1,t|n} - P_{t,t-1|n}T_t^{'} + T_t P_{t-1|n} T_t^{'}\}\right] \\
&-\frac{1}{2}\log \mid P_0 \mid -\frac{1}{2}(\alpha_0 - a_0)^{'} P_0^{-1}(\alpha_0 - a_0) - \log f(\alpha | Y_n; \theta),
\end{aligned} \tag{3.11}$$

where $\eta_{t|n} = (\alpha_{t+1|n} - d_t - T_t\alpha_{t|n})$, and $P_{t,t-1|n}$ is the smooth covariance between the states. To evaluate the likelihood and the derivative the smooth covariance $P_{t,t-1|n}$ as to be calculated according to the following formula:

$$P_{t,t-1|n} = \mathrm{Cov}(\alpha_t, \alpha_{t-1}|n) = T_{t-1}P_{t-1|n} - \Sigma_{\eta,t} N_{t-1} L_{t-1} P_{t-1|t-2}, \tag{3.12}$$

where all the quantities are given as output by the Kalman filter and smoother. Although other solutions could be applied those are, in presence of many missing data, highly inefficient or difficult to implement.

3.3.3 Numerical performance

Recall the division outlined in Table 3.1 and the formulation proposed for each series:

$$y_{i,t} = \lambda_i^{world} f_t^{world} + \lambda_i^{region} f_{r,t}^{region} + u_{i,t}, \quad t = 1, \ldots, T, \quad i = 1, \ldots, N, \tag{3.13}$$

where i denotes the countries and r denotes the region and the $u_{i,t}$ follows AR(1) processes given by

$$\begin{aligned} &u_{i,t} = \phi_i u_{i,t-1} + \varepsilon_{i,t}, && \varepsilon_{i,t} \sim N(0, \sigma^2_{\varepsilon,i}), \\ &E(\varepsilon_{i,t}\varepsilon_{j,t-s}) = 0, && for \ \ i \neq j \ , s > 0. \end{aligned}$$

The corresponding state space form with missing values is provided in Appendix B, see (3.26) and (3.27). The parameter vector θ consists of elements of Z and T, the diagonal elements of $\phi^{(0)}$ and Σ_ε, and the parameters corresponding to the missingness, indeed $\phi^{(1)}$ and $\phi^{(*)}$. In the case of 150 countries the number of coefficients to estimate in Λ are $2N \times 1 = 300$. Where $N \times 1$ coefficients corresponds to the global factor loadings and $N \times 1$ corresponds to the regional factor loadings, recall formula (3.5). Assuming seven factors, one global and six regional, in the matrix T we have $q^2 = 49$ and in the diagonal matrices $\phi^{(0)}$ and Σ_ε we have a total of $2N = 300$ parameters. Finally the number of $\phi^{(1)}$ and $\phi^{(*)}$ are dependent on the missingness, and they are equal to 96 and 1 respectively. The dimension of θ is therefore 746. The huge number of parameters and the random starting points gives multiple solutions for our likelihood, this kind of behavior is quite common with our parameters dimension. We repeat the maximization 2000 times, every time with random starting values, moreover, we keep track of the parameters stability. As a measure of parameters variation we use a normalized Euclidean distance:

$$D(\hat{\Theta}^{(i)}, \hat{\Theta}^*) = \frac{\parallel \hat{\Theta}^{(i)} - \hat{\Theta}^* \parallel}{\parallel \hat{\Theta}^* \parallel}, \tag{3.14}$$

where $\hat{\Theta}^*$ is the parameters value at the chosen highest maximum and $\hat{\Theta}^{(i)}$ is the parameters value at the i maximum. We report the likelihood values against iterations, the distance measure and the distribution of the distance measure in figure 3.1. The picture shows that the likelihood is moving around a range and the most significant maximum is reached different times. When we are closed to the highest maximum the distance measure goes to 0, indeed the parameters values are very near. The results presented in the following sections are taken from the most likely value reached by the likelihood, in other worlds we take as our likelihood an iteration with a value that falls in the mean value of upper graph of figure 3.1. One of the main advantage of this procedure is the reduce computational time, in Table 3.2 we report the evaluation time for the model considered in the paper. The program is written in Ox v. 5.10 console (Doornik, 2007) using our source code and is running on a standard desktop computer.

3.4 Empirical application

This paragraph examines the evolution of different factors and analyzes their ability to track important business cycle episode. We present the results for the Kose et al. (2004) data set (60 countries with time span 1960-1990) and the full data set (150 countries with time span 1950-2007)

Figure 3.1: *Upper graph: Likelihood variation against iteration. Middle graph: Distance measure for parameter's variation. Bottom graph: Distribution of the distance parameter.*

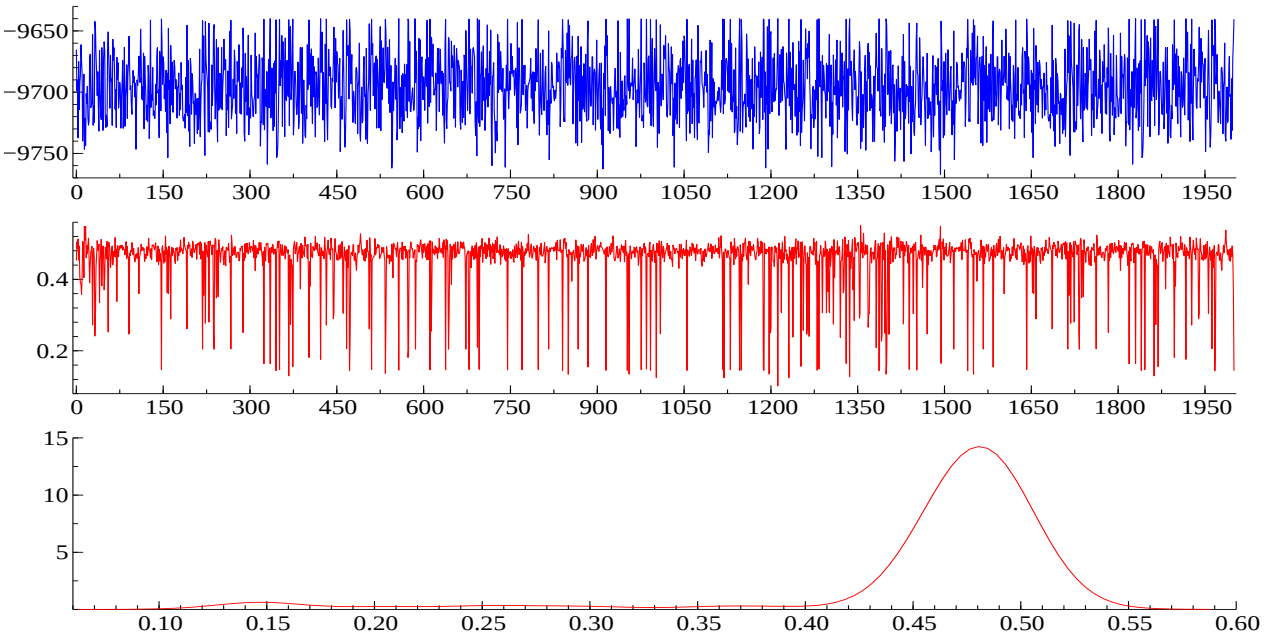
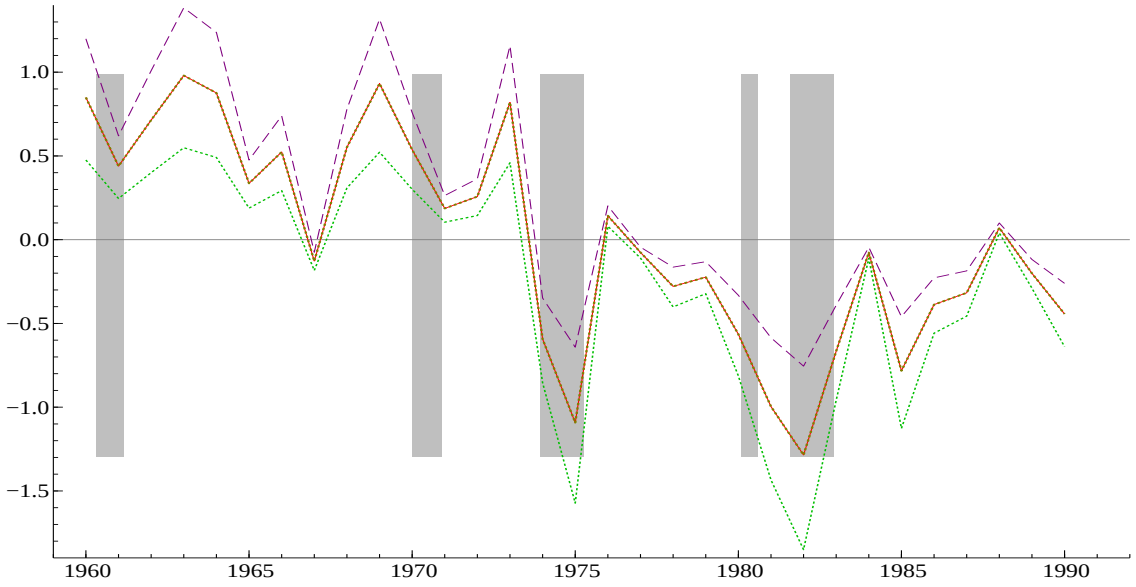


Table 3.2: *Evaluation Time*

Model	BFGS Iter	Time for 100 iter.	Evaluation Time	Data Points	Missing values	Parameters
Seven factors 60 countries period 1960-1990	102	4 minutes and 20 seconds	5 minutes	1891	30	361
Seven factors 150 countries period 1950-2007	220	9 minutes	21 minutes	8550	1245	746
Seven factors without restrictions 150 countries period 1950-2007	380	12 minutes	45 minutes	8550	1245	1699

Figure 3.2: *World Factor plus 33 and 66 quantile percent bands, estimated using the Kose et al. (2004) data set. We report the NBER recessions with the vertical lines.*



in Subsection 3.4.1 and 3.4.2 respectively. To check if our global factor rely mainly on biggest countries, we carry out a sensitivity experiment in Subsection 3.4.3. The diagnostic checking and model fit are presented in Subsection 3.4.4, the variance decomposition in Subsection 3.4.5, and finally in Subsection 3.4.6 we present new results about global and regional convergence.

3.4.1 Global and Regional factors using reduced data set

We now examine the global and regional factors using a data set similar to Kose et al. (2004). The 60 countries used in this study are reported in Appendix A and they are divided accordingly to Table 3.1. Figure 3.1 reports the estimated global factor with the 33 and 66 percent quantile bands. The fluctuations of the factor is very similar to Kose et al. (2004) and reflect the major economic events of those 30 years: the expansion in the period of the 1960's, the recession of the mid-1970's (same period of the first oil price shock), the strong recession in the early 1980's, caused by the debt crisis and the tight monetary policies started in Usa around 1979.

As in previous studies the estimated global factor confirms that the recession in the early 1980's was stronger than the recession of mid 1970's. The inclusion of the Latin America countries that suffered a lot from the debt crisis of the early 1980's, strongly influenced the global factors. Finally it is clearly shown the downturn of the early 1990's.

Thanks to an efficient treatment of the missing values we can estimate the global and regional factor using all the 150 countries, in this experiment the time period is still 1960-1990. In the bottom graph of figure 3.3 we report the estimated global factor with the 33 and 66 percent quantile bands. Looking at the graph we can notice that the movement is quite similar to that of figure 3.2. The inclusion of more countries does not change a lot the estimate of our global factor. The downturn of early 1970's seems to be less strong, this effect could be due to the inclusion of Asia and Oceania

Developed countries, that in the 70's experimented a long growth period, see figure 3.4.

3.4.2 Global and Regional factors using complete data set and all the time span

We begin by considering the general properties of our estimated world and regionals factors using all the 150 countries for whole period. Figure 3.3 upper graph reports the estimation of the global factor with the recessions as vertical lines. We use Usa recessions as a proxy of global recessions until 1985, before that year the global recession date are not available, see *http://www.imf.org/*. Figure 3.4 shows the estimated regional factors.

The global factor reflects the major economic events from the 1950 until 2007. The 1958 recession, the expansion in the period of the 1960's, the downturn of early 1970's, the recession of the mid-1970's and the strong recession in the early 1980's. Moreover the figure shows the global recessions that according to the IMF chronology, see *http://www.imf.org/*, corresponds to: 1990 - 1993, 1998 and 2001 - 2002. Finally we can notice the expansion of middle 2000. In some cases the thoughs of the world cycle corresponds quite closely with those of Usa output (NBER recession). In contrast the NBER thoughs in 1970 and 1980 appear to be principally Usa specific phenomena. The Usa seems not to have an overall tendency to lead the world business cycle. Figure 3.3, middle graph, reports the estimated factor using all the 150 countries for the time period 1950 - 2007, but we show only the results around the period 1960 - 1990. This graph is very similar to the global factor estimated using 150 countries for time period 1960 - 1990, bottom part of figure 3.3. This shows that using longer time span gives no different picture of our global factor, this is useful to asses the robustness of our estimation technique. To study the economic evolution of every single area we report in figure 3.4 the regional factors.

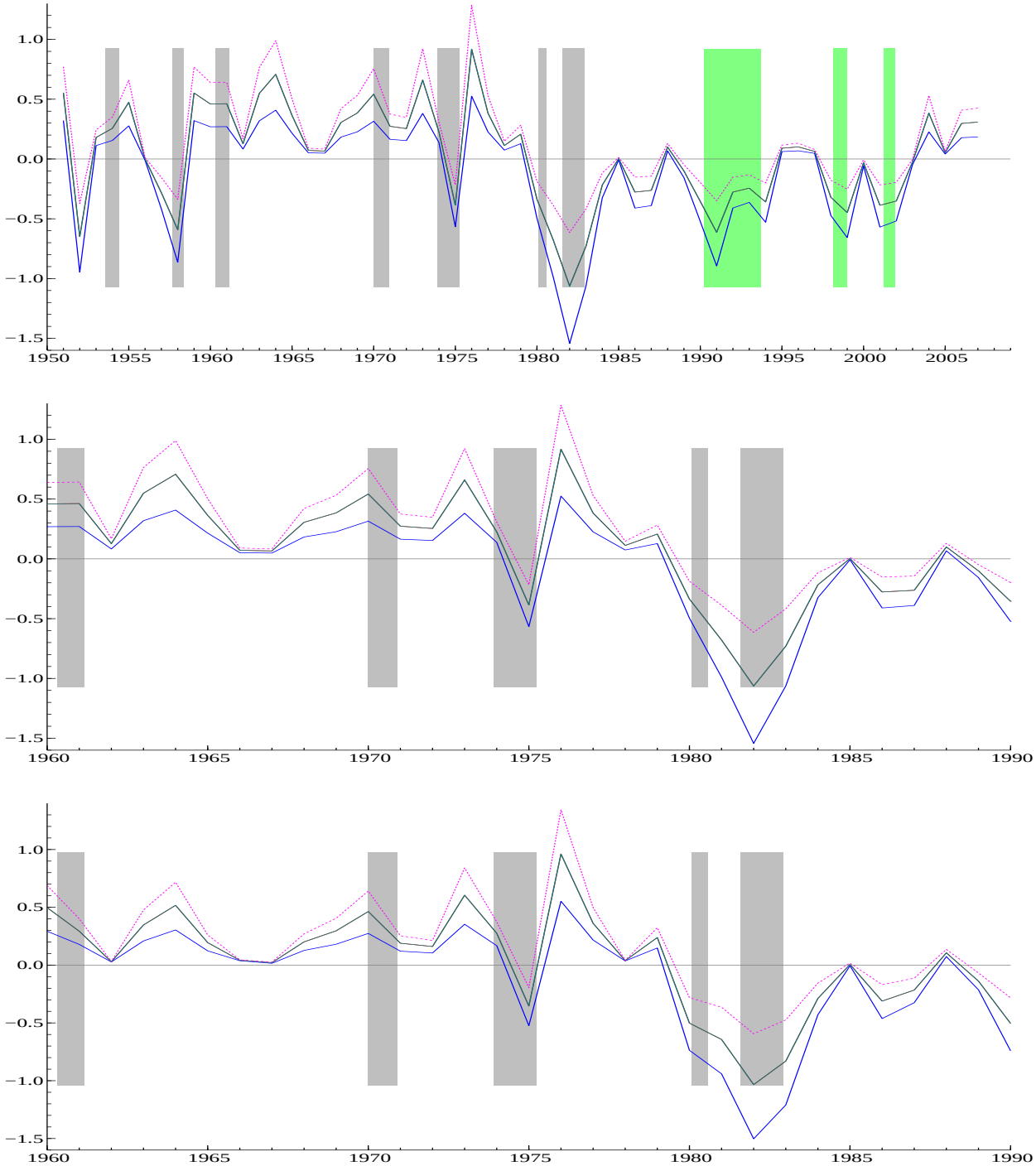
The upper left part of figure 3.4 reports the North America regional factor with NBER recessions as grey vertical line, as it is clear, this factor closely follows those recessions. In the last part of the graph we can notice a long period of growth until 2000 that coincides with the Clinton Era that strongly interested this area. After this period we have the 2001 recession caused by the collapsed of Dot-com bubble and September 2001 attacks, this recession seems to be different from the recent one, in fact, it has been neither so strong and nor so persistent. This result is in line with some findings in the literature, see for example Nordhaus (2002). Interestingly the downturn of 1994-1995 is not a Usa recession but a strong drop of the Mexico's GDP related to the Mexican peso crisis.

Factor 2, Latin America regional area, shows the debt crisis of the 1980's with decrement in the factor started around 1979 until 1983, see Weeks (2000). The trough clearly showed by the figure corresponds to a severe recession that took place in this area around 1989. The downturn of the factor in 1995 is the impact of the Mexican economic crisis of the Southern Cone (Argentina, Chile, Paraguay and Uruguay) and Brazil labeled as the *Tequila Effect*. Moreover it is evident the slowdown in the factor in the period 1999 - 2001 that corresponds to the Argentina's default and the global recession.

About European region, factor 3 shows that these countries suffer a lot from the recession of the mid - 70's but less from the recession of the early 1980's. Moreover, it displays the decline starting from 1990 and ending with a deep value of the factor in the 1992-1993, the same period of the EMS crisis, see Eichengreen (2001). As other areas in the world the European region was interested by the 2001 recession.

Factor 4, Asia and Oceania Developed region, shows the high growth period of the 1970's,

Figure 3.3: *Upper graph: World Factor estimated using all 150 countries for time period 1950-2007 plus 33 percent and 66 quantile bands. Gray vertical line, USA recessions, green vertical lines global recessions. Middle graph: World Factor estimated using all 150 countries for time period 1950 - 2007 but zoomed around 1960 - 1990 plus 33 percent and 66 quantile bands. Gray vertical line USA recessions. Bottom graph: World graph estimated using all 150 countries but using sub-sample 1960-1990 plus 33 and 66 percent quantile bands. Grey vertical line USA recessions.*



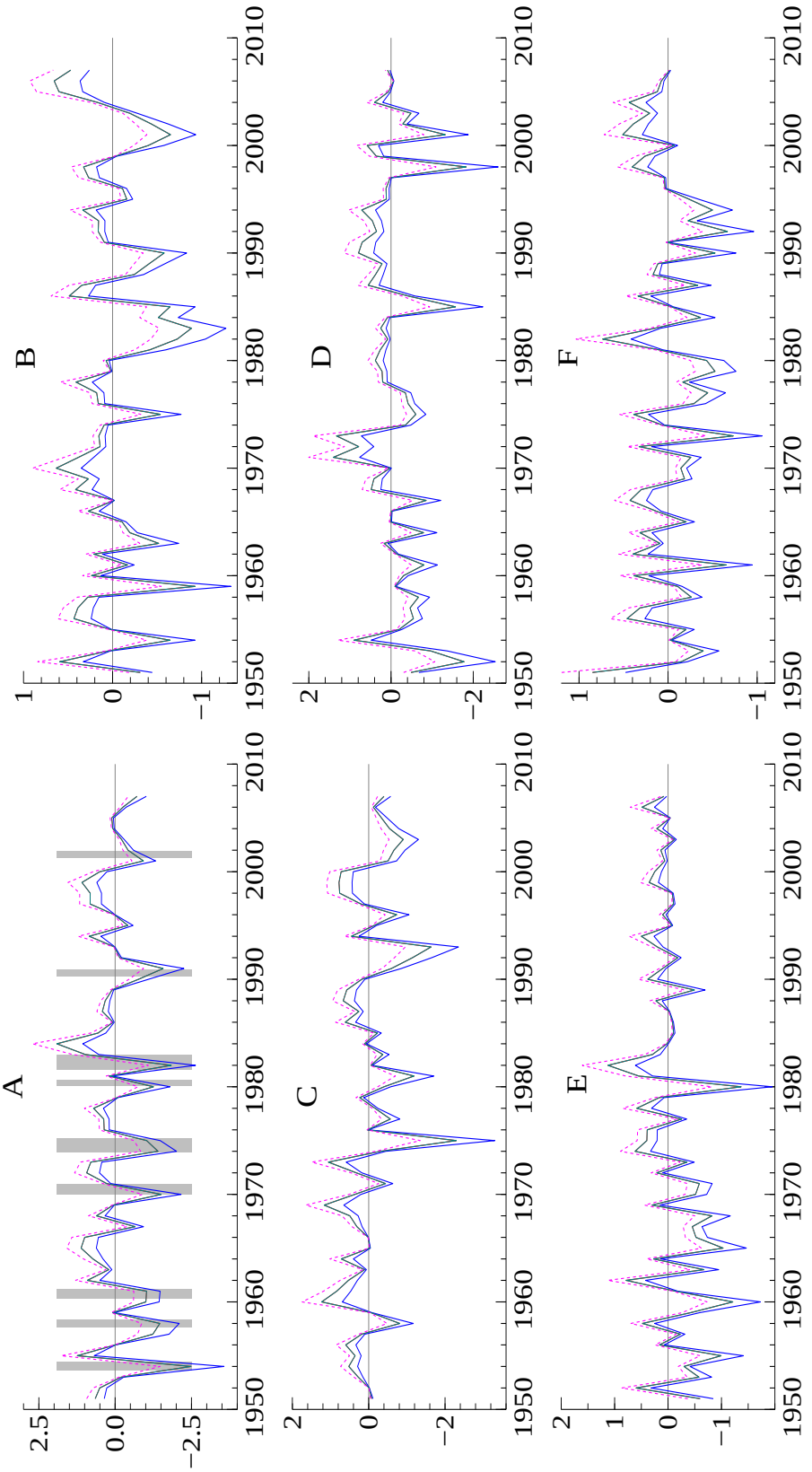


Figure 3.4: *Regional Factors with 33 percent and 66 percent quantile bands. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region. The vertical lines are the NBER USA recessions that we report just for the North America Region.*

Table 3.3: *Estimated VAR coefficients and Eigenvalues for 150 countries, period 1950 - 2007. Real is the real part of the eigenvalues, that ranges between a maximum of 0.36 and a minimum of 0.045. Img is the complex conjugate for the eigenvalues.*

Var Coefficients								Eigenvalues		
Factor	World	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6		*Real*	*Img*
World	0.171	0.005	0.030	0.030	0.030	0.011	0.004	Eig_1	0.36	0.000
Region 1	0.010	0.139	0.041	0.018	0.036	0.026	0.110	Eig_2	0.16	0.008
Region 2	0.136	0.026	0.182	0.003	0.023	0.004	0.070	Eig_3	0.16	-0.008
Region 3	0.111	0.112	0.061	0.156	0.023	0.010	0.105	Eig_4	0.095	0.043
Region 4	0.010	0.123	0.039	0.033	0.138	0.041	0.062	Eig_5	0.095	-0.043
Region 5	0.015	0.029	0.045	0.019	0.025	0.082	0.047	Eig_6	0.082	0.000
Region 6	0.039	0.035	0.010	0.003	0.029	0.033	0.110	Eig_7	0.045	0.000

moreover the debt crisis of the early 1980's seems to affect marginally this area. This factor shows clearly the financial crisis that took place in these countries around 1997. This region has been strongly affected by the 2001 global recession.

About Asia and Oceania Developing and Poor and Africa these regional factors seem to follow a different path, this finding is in line with Kose et al. (2004). We will justify this more rigorously in Subsections 4.4 and 4.5.

Table 3.3 reports the estimated VAR coefficients together with the eigenvalues organized in descending order. Looking at the table we can conclude that the factors are estimated as stationary and they seem to be quite persistent, the largest eigenvalue is around 0.36. We find the presence of two not persistent cyclical behavior in the factors since conjugate pair of complex eigenvalues are obtained whereas the real part is equal to 0.16 and 0.095.

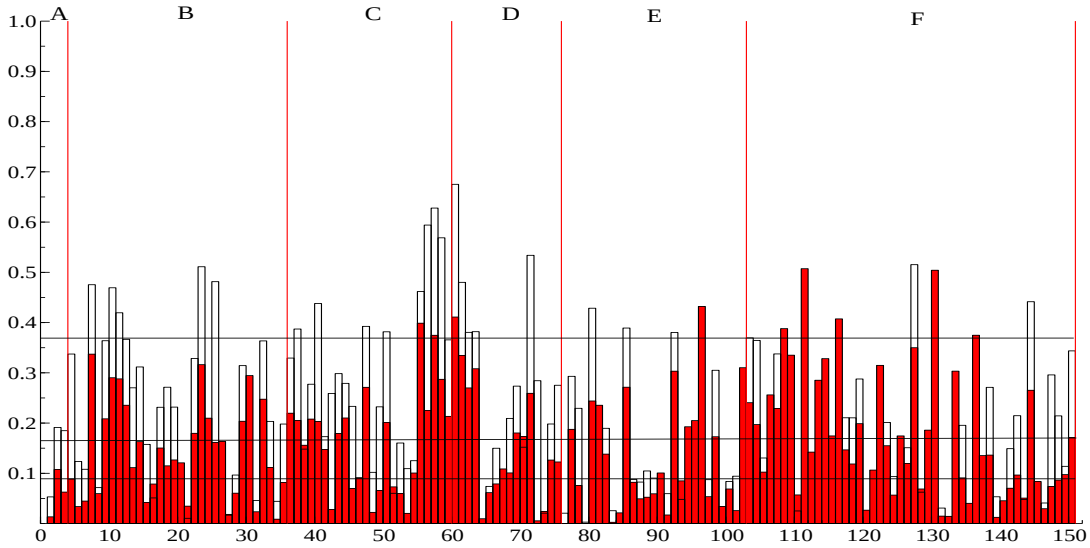
Figure 3.5 reports in a bar plot the estimate of the autoregressive parameters, $\phi^{(0)}$, that are used to calculate the percentage of country specific component. To facilitate the reading of this graph we divide it in the six regional areas using vertical lines. The autoregressive parameters range from a minimum around 0.01 to a maximum around 0.5. This figure reports in the horizontal line the 3 biggest eigenvalues of Table 3.3.

The figure also reports the first order autocorrelation, $\gamma(1)$, of the raw series. We can notice the reduction in the estimated autocorrelation given by our estimation procedure , this indicates that great part of the dynamics is picked up by our formulation. The highest values of the autoregressive parameter $\phi^{(0)}$ corresponds to few poorest countries, in this case they dynamics is not so well described.

3.4.3 Sensitivity experiment: G7 countries

To ensure that our results are due to the scope of the sample and not to our approach, we employ our procedure to estimate a dynamic factor model using the aggregate data of the G7 countries see Appendix A. This exercise is in line with Gregory et al. (1997) and we get virtually the same results. The estimated global and regional factors are presented in figure 3.6. Focusing just on G7 countries the global factor exhibit a more severe recession in 1974 that in 1982, the same observation has been done by Kose et al. (2004). About the regional factors just the North America regional factor is similar to the one presented in figure 3.4. This is can be explained by the fact that two

Figure 3.5: *Red bars: $\phi^{(0)}$ parameters for the autoregressive component, see formula (3.26). White bars: Sample autocorrelation of order one, $\gamma(1)$, of the raw series. The vertical lines divide the $\phi^{(0)}$ parameters between the areas. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region. The horizontal lines reports the largest eigenvalues estimate using the maximum likelihood, see Table 3.3.*



up to three countries in this region compose the G7 group, and Mexico seems not to be related to the North America region, see Subsection 4.5 and 4.6 for a detailed discussion.

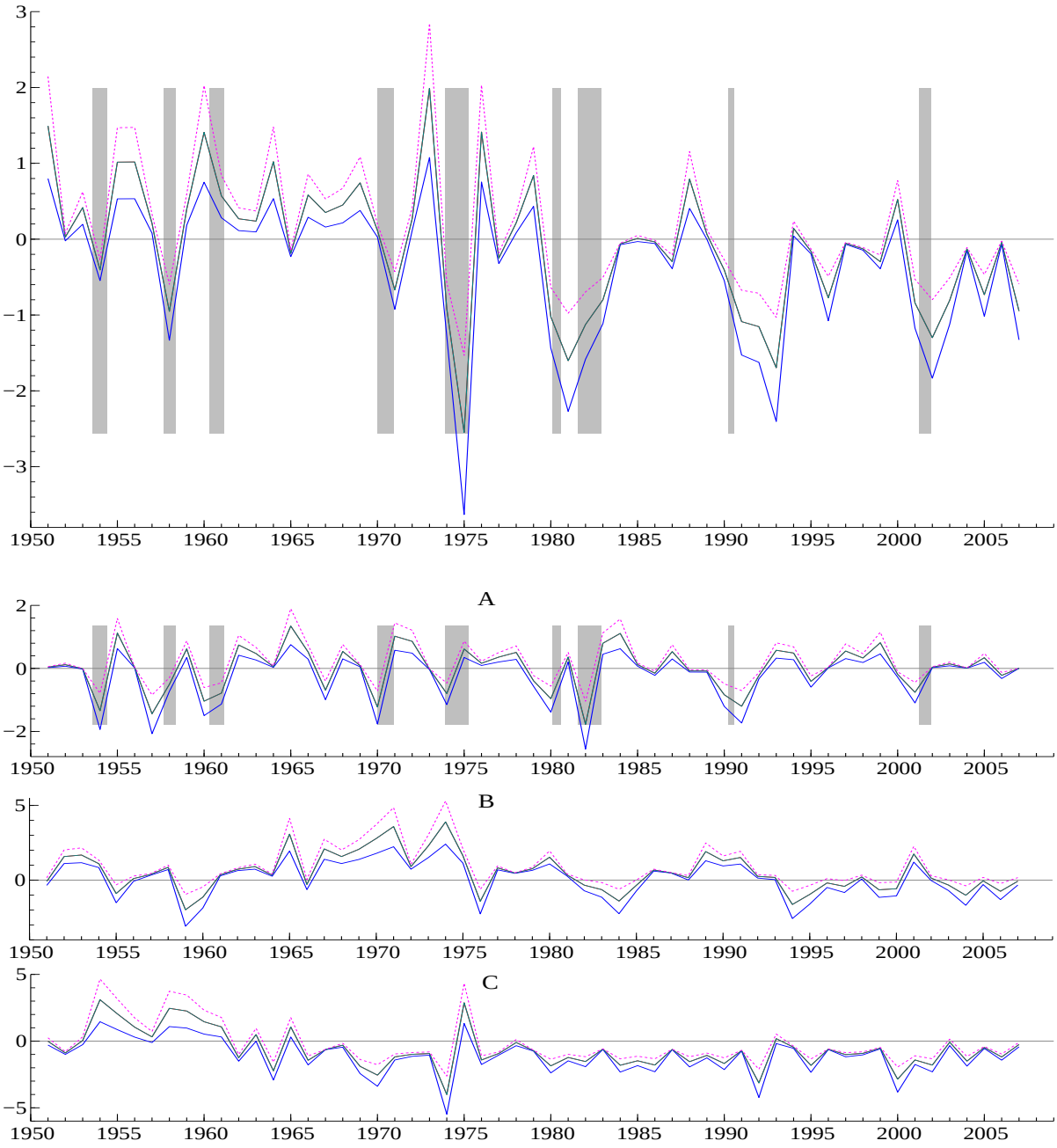
3.4.4 Diagnostic checking and model fit

In this subsection we discuss the model fit and the model diagnostics using the R^2 and Ljung-Box statistic. The actual estimate of Λ is not easy to interpret and therefore Stock and Watson (2002) proposed to focus on the R^2 goodness-of-fit statistics which is obtained by regressing the univariate time series $y_{i,t}$, for each $i = 1, \ldots, N$, on a constant and a particular principal component estimate. These R^2 statistics are then regarded as proxies for the correlations (in absolute values) between the series and each principal component. In our framework, we can evaluate the correlations between the series and each factor directly. The N regressions can be repeated for each principal component and the resulting N dimensional series of R^2 statistics can be displayed as an index plot for each principal component. We present the N series of R^2 statistics for the seven factors, in case of unrestricted and restricted Λ in figure 3.7 and in figure 3.8. To make it more readable the R^2 is split based on the areas.

Figure 3.8 shows the R^2, in the case of restricted Λ, for the global (left hand side) and the regional (right hand side) factors, divided accordingly to the different areas. In terms of the shares of variance accounted for by the common factor the quantitative importance of the world business cycle varies widely across countries.

The global factor is quite correlated with USA, with value around 0.22. Canada has almost the same value around 0.20. Mexico shows a good link with the global factor with a value for the

Figure 3.6: *Upper graph: World Factor estimated using the G7 countries for time period 1950-2007 plus 33 percent and 66 quantile bands. Gray vertical line, USA recessions. Buttom graph: Regional Factors for G7 countries with 33 percent and 66 percent quantile bands. (A) North America Region; (B) European Region; (C) Asia and Oceania Developed Region; The vertical lines are the NBER USA recessions that we report just for the North America Region.*



global R^2 around 0.18. More interestingly the regional R^2, is high correlated with USA and Canada with values around 0.85 for USA and 0.91 for Canada. This is quite natural if we consider that Canadian economy is strongly linked to USA. Mexico seems not to share a lot with the regional factor.

The global factor is not very correlated with the Latin America countries, this seems reasonable if we consider that almost all of those countries are very poor, moreover the regional factor seems not to be very important for this area as well.

The global factor has an important effect in Europe with highest value around 0.60. Germany presents a good link with the global factor similar to France and less than the UK, the global factor is very important for UK and Italy. This finding is coherent with the main features of the Italian economy mostly oriented to exportations. The regional factor is very important for the European countries in fact Germany, France, Italy, and Belgium are very correlated with the regional factor. United Kingdom is not influenced a lot by the European regional factor. We will analyze deeply those findings in Subsection 4.5.

The correlation between the Asia and Oceania Developed countries with respect to the global factor is quite interesting. The country with more correlation is the Saudi Arabia, this seems reasonable if we consider that Saudi Arabia's economy is petroleum-based and almost the 90 percent of export earnings come from the oil industry. Then, among other, Taiwan, Hong Kong and Japan are correlated with the global factor in this area. Two special cases are: Japan that seems to be quite correlated with the global factor but is not influenced by the regional one, and China that is not so correlated with both factors. It seems that all its variability is explained by the country specific component, see Subsection 4.5.

The correlation of Asia and Oceania Developing and Poor region with the global factor is small, with null value for some countries like Pakistan and Nepal. Finally the Africa macro area has not a strong correlation with the global factor, and the regional factor seems not to be important. For most of them the countries specific effect is the prominent element that explains the economic fluctuations, see Subsection 4.5. Finally looking at figure 3.7 and 3.8 we can notice the difference in the R^2 for the unrestricted and restricted case. The regional R^2 are very different between the two figures, showing the importance of the factor loading restrictions to disentangle global and regional factors.

Another advantage of this framework is to easily account for model misspecification tests and diagnostics concerning normality, heteroskedasticity and serial correlation and it can be seen as an effective tool for model selection. The Kalman filter allows us to calculate in few seconds the prediction errors for our data set even in presence of missing values. Thanks to this tool we can carry out easily the Ljung-Box test (1978). The Ljung-Box $Q(q)$ statistic is based on the first q sample autocorrelations r_k^*, $k = 1, \ldots, q$ of the residual series and is computed by $Q(q) = \sum_{k=1}^{q} r_k^2$.

The Ljung-Box statistics for the 150 series is presented as index plot in figure 3.9 for $q = 6$. Almost all the series are in the confidence interval, and we can conclude that the our specification well describe the collective dynamics in our data set. On the other hand some countries, mainly poor countries, have very high value of this statistic, an important exception is Japan that shows a very high value of the $Q(q)$ statistic. In fact Japanese economy seems to be more detached from other industrialized countries with the domestic shocks that explain big portion of its volatility, see Subsection 4.5. The figure also reports with the white bar the Ljung-Box statistics for the raw series, it is clearly showed the autocorrelation reduction after the estimation step.

Figure 3.7: *Global and Regional Factors in the unrestricted case. Left graph: R^2 for the Global Factor divided by areas accordingly to Table 3.1. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region. Right graph: R^2 for the Regional Factor divided by areas accordingly to Table 3.1. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region.*

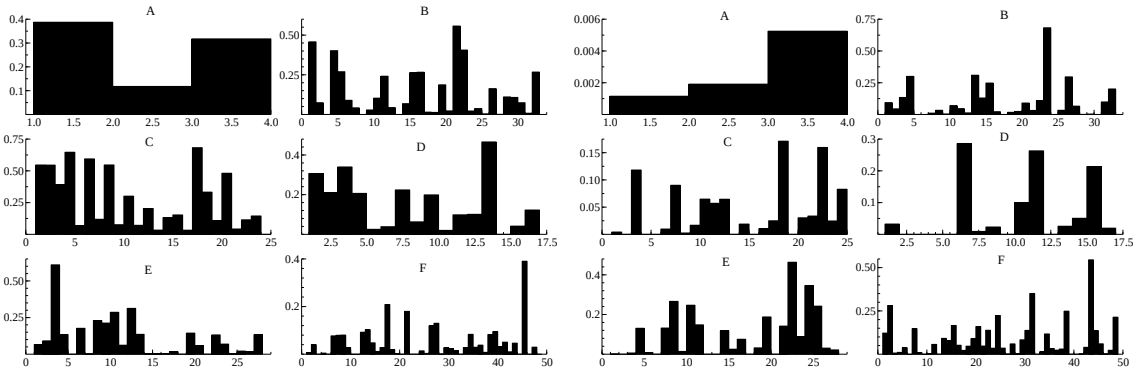


Figure 3.8: *Global and Regional Factors in the restricted case. Left graph: R^2 for the Global Factor divided by areas accordingly to Table 3.1. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region. Right graph: R^2 for the Regional Factor divided by areas accordingly to Table 3.1. (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region.*

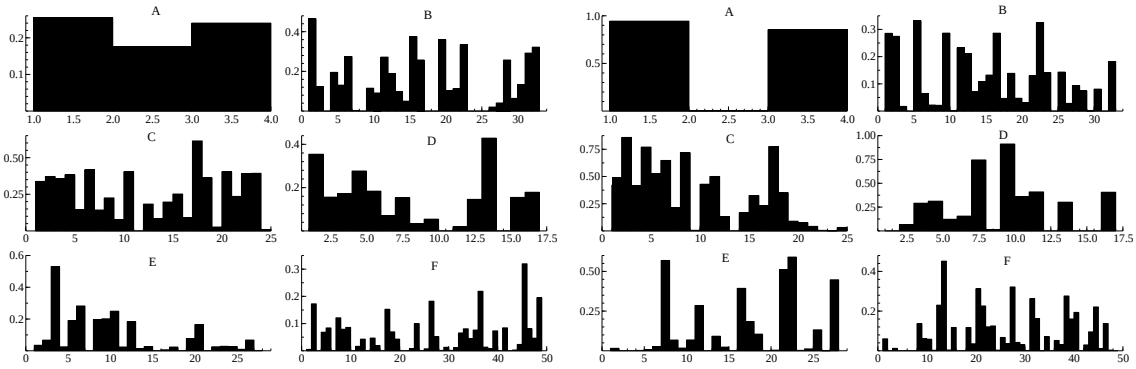
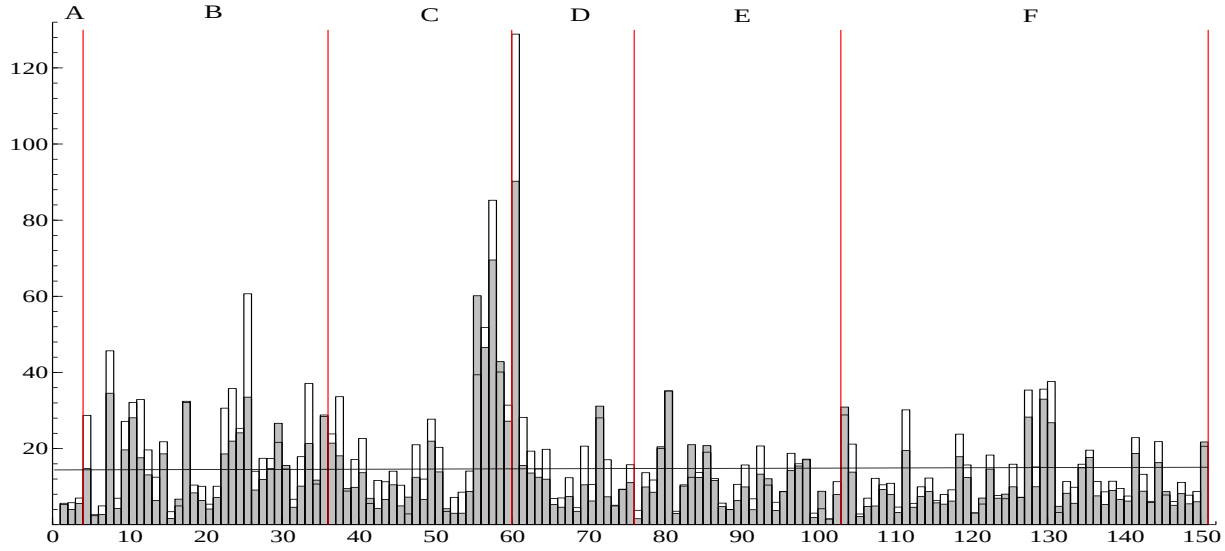


Figure 3.9: *Grey lines: Ljung-Box Q(6) statistics for the generalized least squares residuals of dynamic factor model. White lines: Ljung-Box Q(6) statistics for the raw series. The vertical lines divide the regional areas accordingly to: (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region. The horizontal line is the critical value.*



3.4.5 Variance Decomposition

To measure the relative contributions of the world, regional and country factors to variations in aggregate variables in each country, we estimate the share of the variance of each macroeconomic aggregate due to each factor. We decompose the variance of each observable into the fraction that is due to each global, regional and the country specific factors. With orthogonal factors the variance of observable i can be written in the following way:

$$\text{Var}(y_{i,t}) = (\lambda_i^{world})^2 \text{Var}(f_t^{world}) + (\lambda_i^{region})^2 \text{Var}(f_{r,t}) + \text{Var}(country_{r,t}), \tag{3.15}$$

where r is the region and the variance of the country component is given by the unconditional variance of AR(1) process. The fraction of the volatility explained by the global factor is given by:

$$\frac{(\lambda_i)^2 \text{Var}(f_t^{world})}{\text{Var}(y_{i,t})}, \tag{3.16}$$

this measure is calculated using the parameters and the factors estimated using maximum likelihood technique.

Table 3.4 contains the share of variance accounted for by each factors. This table displays the variance decomposition for the same countries used in Kose et al. (2004) study plus a selection of countries. The considered time span are two: 1960 - 1990 and 1950 - 2007. As measure of the importance of the factors we report the 33 percent and the 66 percent quantiles that are calculated based on a Gaussian approximation, therefore it is not guaranteed that they are in the bounds. The full tables are available from the authors upon request.

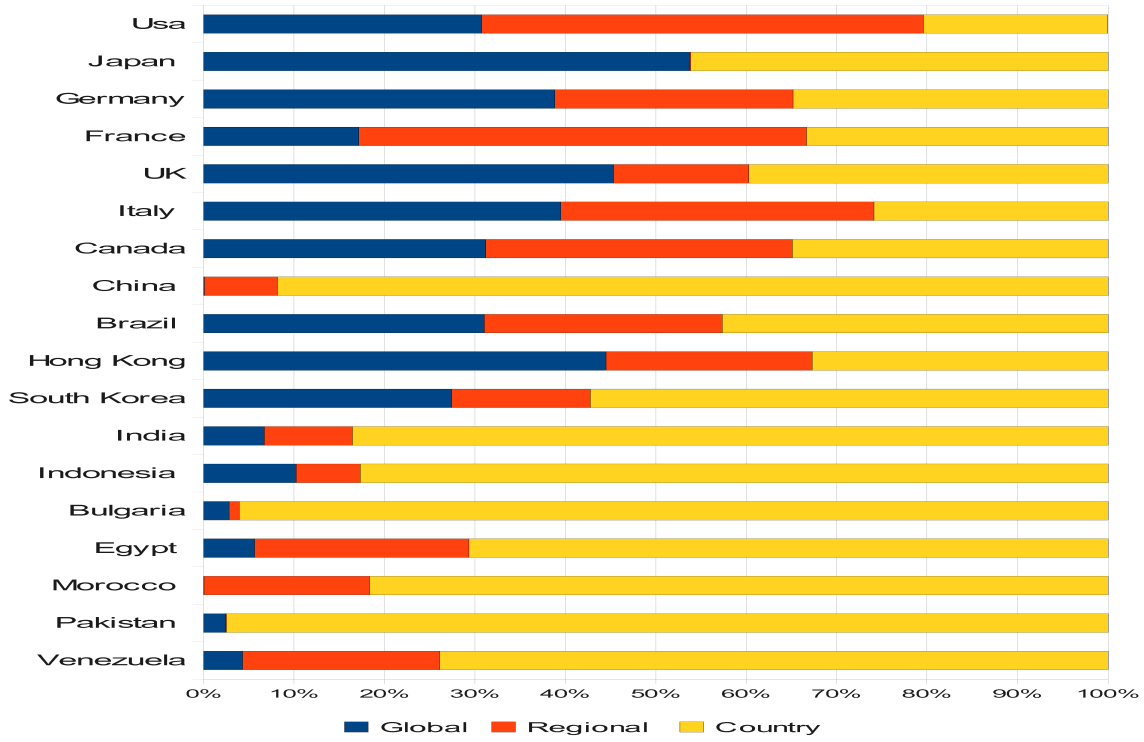
As Table 3.4 shows, the world factor explains a significant fraction of the fluctuations in all the North America countries. The fraction of variability explained by the global factor seems to be less important for USA and Canada and quite stable for Mexico in the period 1950-2007. The regional factor is very important for USA and Canada, moreover it shows a substantial increase for USA. As far as Mexico is concerned the share of variability explained by the regional factor is increased, this effect can be related to the NAFTA agreement.

As far as EU is concerned the table shows that the volatility explained by the global factor is very important between the two periods for almost all the countries. The variability explained by the regional factor is increased among other for Germany, Italy and France, the same conclusion has been reached by Bataa et al. (2009). According to Bataa et al. (2009) the contemporaneous correlations between on these Euro-area countries increased in 1984 and 1998. One important exception is United Kingdom where the variability explained by the global factor is substantially increased between the two periods and the variability explained by the regional factor has decreased. These values confirm the findings in Stock and Watson (2005), in particular UK seems not to be related to European factor any more. Table 4 also provides the variance decomposition of four accession countries in the European union namely Poland, Hungary, Romania and Bulgaria. Poland and Hungary have a very good link with the global factor, but just Hungary has a significant link with the regional factor. Almost all the variability of the other two countries, Romania and Bulgaria, is explained by the country specific component. Our estimations suggest that integration of these countries, except Hungary, with the European Union will require more time.

Though less important than in North America and Europe, the world and regional factors explain a noticeable fraction of aggregate volatility in few countries of the Latin America region. Among other it is interesting to analyze the variance decomposition evolution for Brazil and Venezuela. In Brazil the percentage of variability explained by the global factor has increased between the two periods and the regional factor has slightly decreased. Venezuela experimented a substantial decrease in the global factor, a substantially stable value for the regional factor and an increment in the country specific component. This effect could be due to the Venezuela crisis in the 80's and the corresponding slowdown of the economic activities. Moreover other countries, like Chile, seems to be more detached from the global factor with an increasing share in the regional one.

As far as Asia and Oceania Developed is concerned, Japan is less influenced by the global factor during the period 1950-2007. The regional factor seems not to be important in explaining the variability of this country as well great proportion of Japan's variability is explained by country specific component. If we compare these values with those corresponding to the period 1960-1990 we can notice that during the 1990's, cyclical fluctuations in Japanese GDP became more detached from the global factor, with domestic shocks explaining big portion of these movements. Hong Kong suffered from a substantial decrease in the percentage of the global factor between the two periods and a substantial increase in the regional factor. This suggests a stronger link of this economy with the other countries of this area. One important case is China that seems not to be influenced by the global and regional factor, all its variability is explained by the country specific component. South Korea experimented a substantial increase of the global factor and an almost stable percentage of regional factor between the two periods, moreover the country specific component has decreased substantially. These results are in line with the Korean economy as described in Pecotich and Shultz (2006). Korea is the seventh largest trading partner of the United States and the eighth largest trading partner of the European Union, moreover is the Asia's biggest exporter of refined

Figure 3.10: *Bar plot for the variance decomposition of the global, regional and country specific factor. We report the most industrialized economies and a selection of developing and poor countries.*



oil products.

Unlike North America and Europe for Asia and Oceania Developing and Poor and Africa regions, the country specific component explains almost all the output volatility, therefore these countries seem to be detached from the world economy. One important exception is the Philippines. In this country the global factor decreased between the two periods and the regional factor increased a lot, consequently the idiosyncratic component decreased substantially. This effect could be explained by the fact that during the 1960s, the economy was regarded as the second largest in Asia, second to Japan. However, the leadership of Ferdinand Marcos proved disastrous, by transforming the market economy into a centrally planned economy. The country suffered severe economic recession, and only recovered in the 1990s with a program of economic liberalization, see Gargan (1997). Table 3.4 shows an important regularity, the world factor plays a more important role in explaining the economic activities in advanced industrialized economies than it does in developing economies. Figure 3.10 further illustrate this point showing the variance decomposition for a selection of countries. The first 10 countries are selected between the most industrialized economies, the global and regional factors plays a major role in explaining their fluctuations. The main exception is the China where all its variability is explained by country specific component. This can be justify by the continuous growth of this country that not seems to follow the other industrialized economies. The country specific effect has the prominent role for the other countries.

Table 3.4: *Variance Decomposition for different data sets. The arrows indicates the variation between periods of the variance decomposition. The $\simeq$ indicates almost no variation between periods. The country reported with the (*) are the same countries used in Kose et al. (2004). We report the complete countries for the North America and Europe Region and a selection of countries for the other regions.*

Variance Decompositions for North America Region

		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
United States(*)	1960-1990	21.43	38.26	54.04		19.52	34.86	49.23		15.05	26.87	37.95	
	1950-2007	17.11	30.62	43.30	$\Downarrow$	27.46	48.95	69.10	$\Uparrow$	11.43	20.42	28.84	$\Downarrow$
Canada(*)	1960-1990	27.36	48.87	69.03		18.54	33.10	46.76		10.09	18.01	25.44	
	1950-2007	17.56	31.27	44.14	$\Downarrow$	19.00	33.93	47.92	$\simeq$	19.64	34.98	49.37	$\Uparrow$
Mexico(*)	1960-1990	10.15	18.12	25.59		1.13	2.02	2.85		44.72	79.85	112.80	
	1950-2007	9.12	16.50	23.52	$\Downarrow$	2.62	4.42	6.11	$\Uparrow$	44.26	79.02	111.62	$\simeq$

Variance Decompositions for European Region

		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
Germany(*)	1960-1990	19.62	35.04	49.49		5.33	9.52	13.45		31.04	55.42	78.28	
	1950-2007	21.81	38.89	54.92	$\Uparrow$	14.72	26.29	37.13	$\Uparrow$	19.48	34.81	49.18	$\Downarrow$
France(*)	1960-1990	38.91	69.48	98.14		12.12	21.64	30.56		4.97	8.87	12.53	
	1950-2007	9.67	17.22	24.30	$\Downarrow$	27.71	49.49	69.90	$\Uparrow$	18.61	33.38	47.03	$\Uparrow$
Italy(*)	1960-1990	30.84	55.07	77.79		9.79	17.48	24.69 0		15.37	27.44	38.76	
	1950-2007	22.10	39.48	55.77	$\Downarrow$	19.40	34.62	48.89	$\Uparrow$	14.50	25.89	36.57	$\Downarrow$
UK(*)	1960-1990	6.89	12.32	17.40		12.51	22.34	31.56		36.59	65.33	92.28	
	1950-2007	26.89	46.84	65.54	$\Uparrow$	6.86	13.43	19.59	$\Downarrow$	22.24	39.72	56.10	$\Downarrow$
Sweden(*)	1960-1990	20.20	36.07	50.95		2.04	3.65	5.16		33.75	60.26	85.12	
	1950-2007	20.39	36.40	51.42	$\simeq$	12.37	22.08	31.19	$\Uparrow$	23.25	41.51	58.63	$\Uparrow$
Switzerland(*)	1960-1990	15.16	27.08	38.25		20.97	37.45	52.90		19.86	35.46	50.091	
	1950-2007	32.86	58.68	82.884	$\Uparrow$	3.95	7.06	9.98	$\Downarrow$	19.18	34.25	48.38	$\Downarrow$
Spain(*)	1960-1990	27.15	48.49	68.49		2.41	4.32	6.09		26.43	47.19	66.65	
	1950-2007	3.15	6.03	8.72	$\Downarrow$	24.49	43.73	61.77	$\Uparrow$	28.35	50.23	70.74	$\Uparrow$
Portugal(*)	1960-1990	8.20	14.65	20.70		23.41	41.81	59.05		24.38	43.53	61.49	
	1950-2007	20.06	35.83	50.61	$\Uparrow$	10.68	19.07	26.94	$\Downarrow$	25.25	45.09	63.68	$\Uparrow$
Norway(*)	1960-1990	3.10	5.55	7.84		0.72	1.29	1.82		52.17	93.15	131.58	
	1950-2007	19.04	34.00	48.03	$\Uparrow$	3.35	5.98	8.45	$\Uparrow$	33.61	60.00	84.75	$\Downarrow$
Netherlands(*)	1960-1990	29.68	53.00	74.86		5.19	9.27	13.09		21.12	37.72	53.28	
	1950-2007	9.24	16.50	23.31	$\Downarrow$	22.30	39.42	55.47	$\Uparrow$	24.46	44.07	62.46	$\Uparrow$
Luxembourg(*)	1960-1990	5.69	10.16	14.35		25.60	45.72	64.57		24.70	44.11	62.31	
	1950-2007	1.12	1.29	1.32	$\Downarrow$	27.00	48.20	68.08	$\Uparrow$	27.88	50.56	71.83	$\Uparrow$
Ireland(*)	1960-1990	22.02	39.32	55.54		3.68	6.57	9.29		30.29	54.09	76.40	
	1950-2007	12.70	22.68	32.04	$\Downarrow$	14.87	26.54	37.49	$\Uparrow$	28.43	50.77	71.71	$\Downarrow$
Iceland(*)	1960-1990	13.05	23.30	32.91		0.72	1.28	1.81		42.23	75.41	106.52	
	1950-2007	13.86	24.75	34.96	$\Uparrow$	1.25	2.23	3.15	$\Downarrow$	40.89	73.01	103.13	$\Downarrow$
Greece(*)	1960-1990	18.25	32.58	46.02		3.85	6.87	9.70		33.90	60.54	85.51	
	1950-2007	12.81	23.50	33.52	$\Downarrow$	2.94	4.94	6.82	$\Downarrow$	40.24	71.54	100.90	$\Uparrow$
Finland(*)	1960-1990	16.27	29.06	41.05		6.73	12.02	16.98		32.99	58.91	83.21	
	1950-2007	14.43	25.77	36.40	$\Downarrow$	7.93	14.16	20.00	$\Uparrow$	33.64	60.06	84.84	$\Downarrow$
Denmark(*)	1960-1990	8.94	15.96	22.55		12.16	21.71	30.67		34.90	62.31	88.02	
	1950-2007	18.75	34.00	48.36	$\Uparrow$	10.68	18.55	25.93	$\Downarrow$	26.57	47.44	67.00	$\Downarrow$
Belgium(*)	1960-1990	33.98	60.67	85.70		14.24	25.42	35.91		7.78	13.89	19.62	
	1950-2007	35.53	63.42	89.53	$\Uparrow$	6.44	11.51	16.26	$\Downarrow$	14.01	25.06	35.42	$\Uparrow$
Austria(*)	1960-1990	25.00	44.63	63.04		7.09	12.67	17.89		23.91	42.69	60.30	
	1950-2007	30.53	54.51	76.99	$\Uparrow$	2.23	3.98	5.62	$\Downarrow$	23.24	41.50	58.63	$\Downarrow$
Cyprus	1960-1990	n.a.	n.a.	n.a.		n.a.	n.a.	n.a.		n.a.	n.a.	n.a.	
	1950-2007	5.20	9.30	13.13		0.91	1.62	2.30		49.88	89.07	125.81	
Malta	1960-1990	n.a.	n.a.	n.a.		n.a.	n.a.	n.a.		n.a.	n.a.	n.a.	
	1950-2007	21.79	38.91	54.95		12.46	22.26	31.44		21.74	38.83	54.84	
Hungary	1960-1990	n.a.	n.a.	n.a.		n.a.	n.a.	n.a.		n.a.	n.a.	n.a.	
	1950-2007	16.04	28.63	40.45		11.21	20.03	28.29		28.75	51.33	72.50	
Romania	1960-1990	n.a.	n.a.	n.a.		n.a.	n.a.	n.a.		n.a.	n.a.	n.a.	
	1950-2007	2.78	4.99	7.06		1.43	2.55	3.59		51.78	92.45	130.59	
Poland	1960-1990	n.a.	n.a.	n.a.		n.a.	n.a.	n.a.		n.a.	n.a.	n.a.	
	1950-2007	21.76	38.86	54.88		0.37	0.67	0.94		33.86	60.47	85.41	
Bulgaria	1960-1990	n.a.	n.a.	n.a.		n.a.	n.a.	n.a.		n.a.	n.a.	n.a.	
	1950-2007	1.70	2.94	4.10		0.65	1.17	1.65		54.97	94.80	136.81	

Variance Decompositions for Latin America Region

		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
Brazil(*)	1960-1990	17.02	30.39	42.93		20.31	36.27	51.23		18.66	33.32	47.07	
	1950-2007	17.40	31.06	43.88	⇑	14.70	26.28	37.13	⇓	23.91	42.65	60.23	⇑
Argentina(*)	1960-1990	22.88	40.85	57.70		6.05	10.80	15.26		27.07	48.34	68.28	
	1950-2007	18.86	33.63	47.49	⇓	17.23	30.82	43.54	⇑	19.90	35.54	50.20	⇓
Bolivia(*)	1960-1990	6.83	12.20	17.23		1.16	2.08	2.94		48.00	85.71	121.06	
	1950-2007	8.65	15.44	21.82	⇑	0.09	0.16	0.23	⇓	47.26	84.38	119.19	≃
Chile(*)	1960-1990	20.13	35.95	50.78		3.75	6.70	9.47		32.11	57.33	80.98	
	1950-2007	1.96	3.50	4.95	⇓	11.07	19.76	27.91	⇑	42.97	76.73	108.38	⇓
Venezuela(*)	1960-1990	23.68	42.28	59.72		5.85	10.46	14.77		26.46	47.25	66.74	
	1950-2007	2.37	4.25	6.02	⇓	12.19	21.78	30.81	⇑	41.45	73.97	104.46	⇑

Variance Decompositions for Asia and Oceania Developed Region

		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
Japan(*)	1960-1990	37.39	66.76	94.30		1.56	2.78	3.93		17.05	30.44	43.00	
	1950-2007	30.02	53.70	75.90	⇓	0.18	0.25	0.32	⇓	25.82	46.07	65.05	⇑
China	1960-1990	n.a.	n.a.	n.a.		n.a.	n.a.	n.a.		n.a.	n.a.	n.a.	
	1950-2007	0.04	0.07	0.14		4.56	8.11	11.39		51.40	91.81	129.70	
Taiwan	1960-1990	n.a.	n.a.	n.a.		n.a.	n.a.	n.a.		n.a.	n.a.	n.a.	
	1950-2007	2.07	3.70	5.23		5.93	10.50	14.79		47.99	85.78	121.22	
Hong Kong(*)	1960-1990	33.33	59.51	84.05		0.77	1.39	1.96		21.89	39.10	55.22	
	1950-2007	24.93	44.52	62.88	⇓	12.78	22.82	32.23	⇑	18.29	32.65	46.12	⇓
South Korea (*)	1960-1990	0.87	1.55	2.20		8.82	15.75	22.25		46.31	82.68	116.79	
	1950-2007	15.05	26.90	38.01	⇑	11.26	15.19	20.81	≃	29.68	56.90	82.42	⇓

Variance Decomposition for Asia and Oceania Developing and Poor Region

		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
India(*)	1960-1990	10.07	17.98	25.40		16.03	28.62	40.43		29.90	53.38	75.41	
	1950-2007	3.77	6.75	9.54	⇓	5.44	9.74	13.76	⇓	46.79	83.51	117.93	⇑
Indonesia(*)	1960-1990	9.43	16.84	23.78		0.04	0.08	0.12		46.52	83.07	117.33	
	1950-2007	5.71	10.19	14.40	⇓	4.00	7.15	10.11	⇑	46.28	82.64	116.73	≃
Philip.(*)	1960-1990	15.84	28.29	39.95		0.82	1.47	2.08		39.33	70.23	99.20	
	1950-2007	14.83	26.49	37.42	⇓	12.98	23.17	32.73	⇑	28.19	50.33	71.09	⇓
Pakistan(*)	1960-1990	12.14	21.68	30.63		1.42	2.55	3.60		42.43	75.76	107.01	
	1950-2007	1.42	2.50	3.52	⇓	0.07	0.10	0.13	⇓	54.59	97.47	137.67	⇑

Variance Decompositions for Africa Region

		World				*Regional*				*Country*			
		1/3	Med	2/3		1/3	Med	2/3		1/3	Med	2/3	
S.Africa(*)	1960-1990	12.36	22.07	31.18		3.64	6.50	9.18		40.00	71.42	100.88	
	1950-2007	6.18	11.04	15.59	⇓	0.07	0.13	0.19	⇓	49.74	88.82	125.46	⇑
Egypt	1960-1990	n.a.	n.a.	n.a.		n.a.	n.a.	n.a.		n.a.	n.a.	n.a.	
	1950-2007	3.15	5.63	7.95		13.28	23.72	33.51		39.56	70.64	99.77	
Morocco(*)	1960-1990	20.15	35.98	50.82		9.97	17.81	25.16		25.87	46.19	65.25	
	1950-2007	0.03	0.02	0.01	⇓	10.25	18.31	25.87	≃	45.74	81.68	115.37	⇑
Nigeria	1960-1990	n.a.	n.a.	n.a.		n.a.	n.a.	n.a.		n.a.	n.a.	n.a.	
	1950-2007	16.68	29.78	42.06		2.57	4.59	6.49		36.75	65.61	92.68	
Algeria	1960-1990	n.a.	n.a.	n.a.		n.a.	n.a.	n.a.		n.a.	n.a.	n.a.	
	1950-2007	23.62	42.17	59.57		0.33	0.60	0.85		32.04	57.22	80.82	

3.4.6 Global Convergence or Decoupling?

The phenomenon of globalization and regional convergence, which refers to the rising trade and financial integration of the world and regional economy, has been studied extensively in recent decades. Kose et al. (2003) showed that globalization leads to an increase in the degree of synchronization of business cycle. Helbling and Bayoumi (2003) examined the synchronization among the Group of Seven (G - 7) countries using simple descriptive statistics, they shown that synchronized slowdown, and in particular the slowdown of 2000 - 2001, has been the norm rather then the exception. Economic theory does not provide definitive guidance concerning the impact of increasing trade and financial linkages on the degree of business cycle synchronization between regional areas. Some studies, like Shin and Wang (2003), analyzed the impact of trade integration in East Asia countries and found an increasing synchronization in this area. Other studies like Stock and Watson (2005) and Del Negro and Otrok (2008) provide some results about Europe. As far as we know there is not a shared framework that gives insight to global and regional synchronization.

Here we want to analyze if the patterns of international business cycle synchronicity have changed over time in response to the globalization. We repeat the exercise presented in Kose et al. (2008) providing a variance decomposition for each region in two different subperiod: the pre-globalization (1950-1984) and the globalization (1985-2007) periods. Figure 3.11 reports the regional mean of the variance decomposition attributable to each factors for the two time periods, plus the mean value of the variance share attributable to each factors for all the countries. Contrary to the convergence hypothesis the average contribution of the global factor to output fluctuation falls around 10,7 percent between the two time periods confirming the results of Kose et al. (2008). More deeply the global factor explains around 21,6 percent of output growth variation among all countries in the period 1950-1984 and it also account for 19,3 percent variability in the period 1985-2007. While this numbers may seem small at first glance, note that they are calculated across a very large and diverse set of countries. In contrast to the global factor the regional factors have a different behavior, their importance has increased markedly for the developed regions, with the highest increment for the European union. The measure reported in figure 3.11 is a static representation of the variance evolution, for this reason we propose to compare over time the variance of the raw data set with the variance of the estimated residuals. The estimated residuals are given by:

$$v_t = y_t - \hat{\Lambda}\hat{f}_t, \quad t = 1, \ldots, T, \tag{3.17}$$

where y_t is the vector of GDPs at time t (recall that we have missing values), $\hat{\Lambda}$ are the estimated factor loadings and $\hat{f}_t$ are the estimated smoothed factors. The proposed *convergence* measure is calculated accordingly to the fraction of the series explained by the estimated factors, indeed:

$$\zeta_t^i = v_t^{i'} v_t^i, \quad t = 1, \ldots, T, \quad i = Global, Regional, \tag{3.18}$$

moreover the raw series variance is calculated as follows:

$$\varphi_t^i = y_t^{'} y_t, \quad t = 1, \ldots, T, \quad i = Global, Regional. \tag{3.19}$$

The difference between the two quantities is a measure of variance reduction introduced by the estimated factors, in other world if a factor, say the global factor, explains big portion of the

Figure 3.11: *Bar plot for the variance decomposition for the pre-globalization period (1950-1984) and the globalization period (1985-2007). First two quantities means of the global, regional and country specific composes. The other quantities are the percentage of the global, regional and country specific component for different regions: (A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region.*

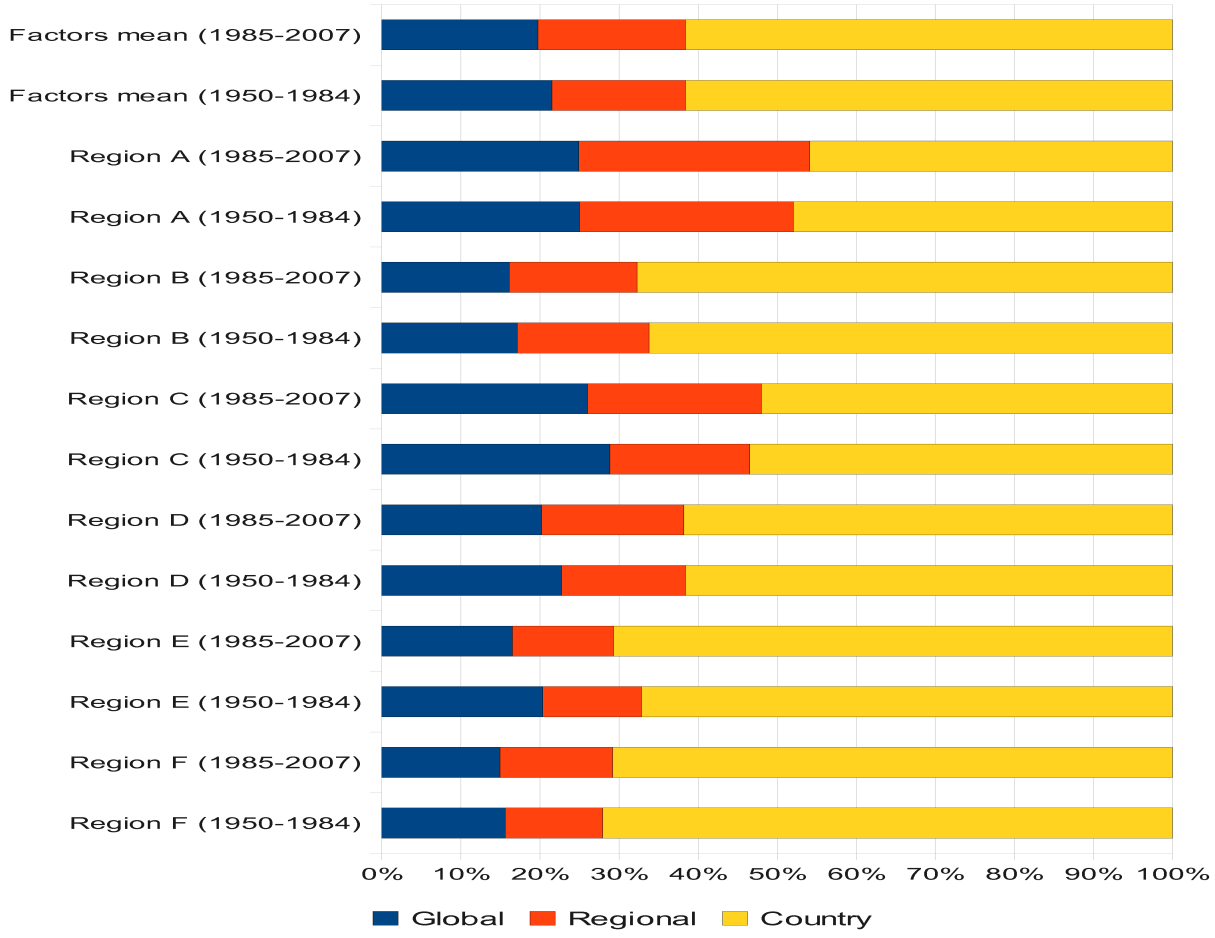
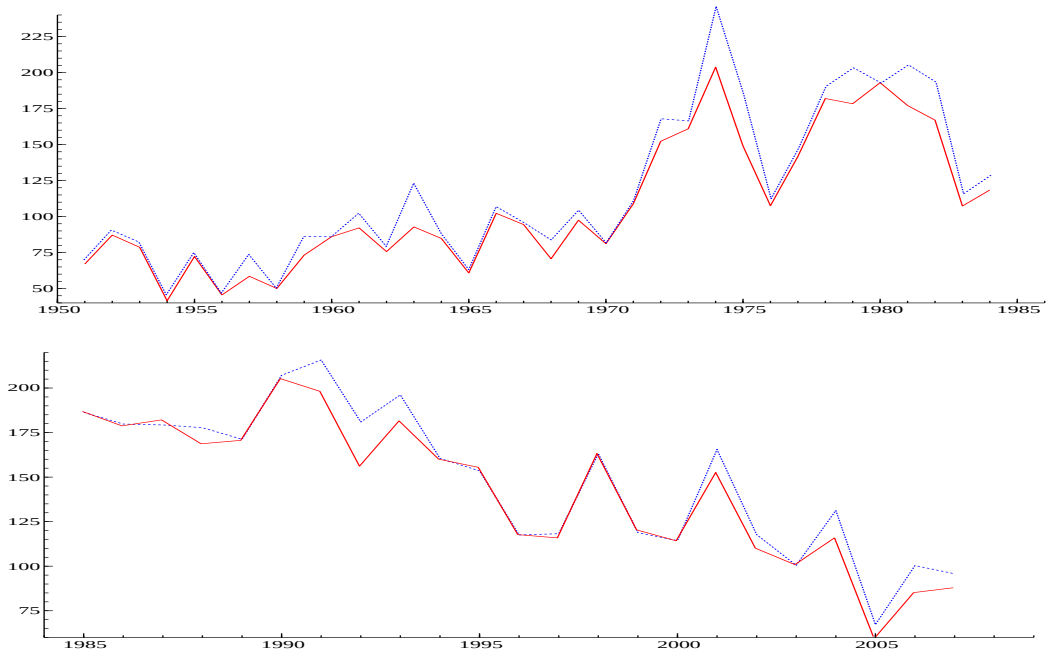


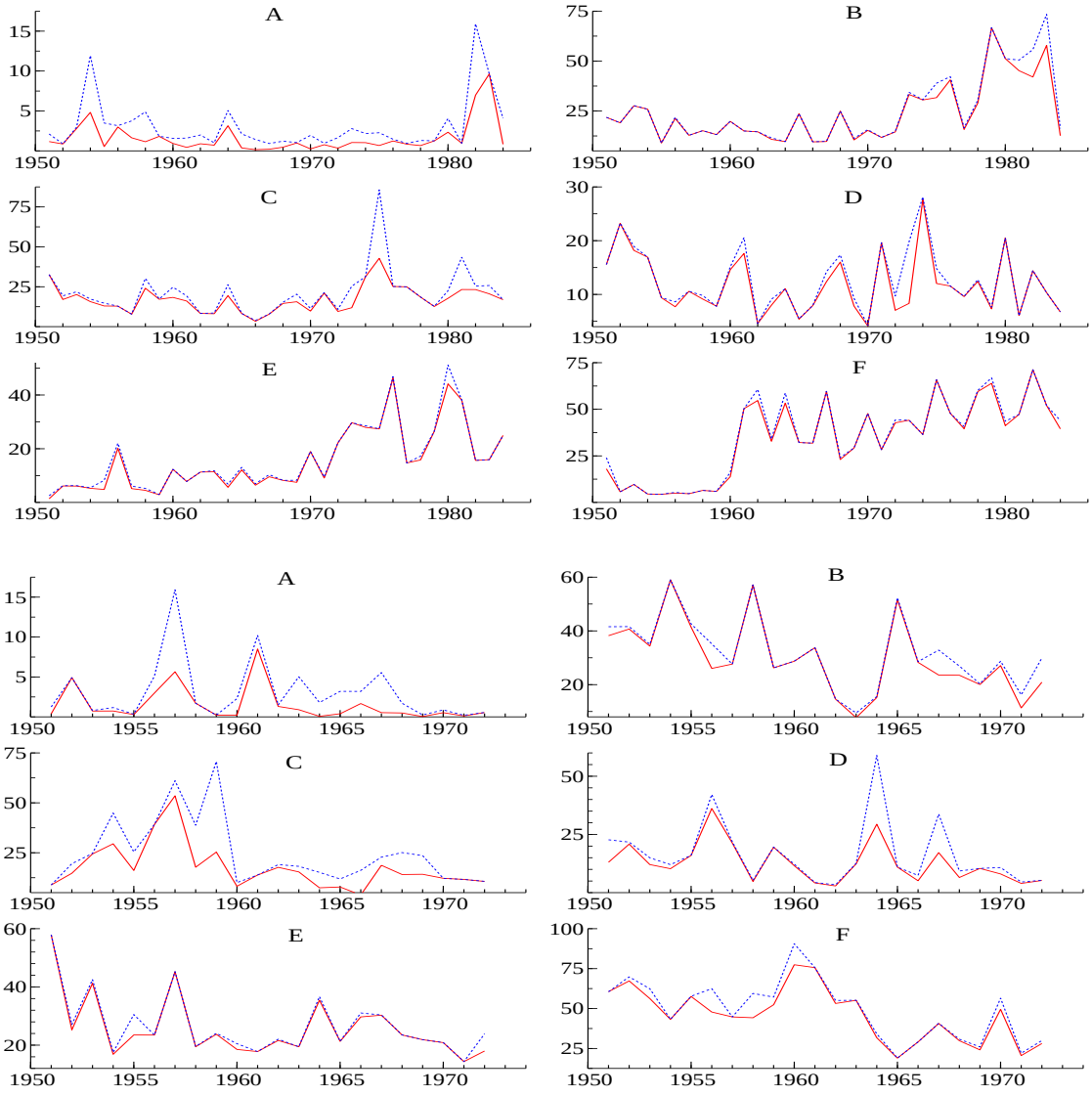
Figure 3.12: *Upper graph: Global φ_t (dotted line) against global ζ_t (solid line) for the pre-globalization period (1950-1984); Bottom graph: Global φ_t (dotted line) against global ζ_t (solid line) for the globalization period (1985-2007).*



variability the ζ_t^{Global} should be less then φ_t^{Global}. Figure 3.12 reports, divided accordingly to the pre-globalization and globalization period, the variance of the raw series φ_t and the convergence measure ζ_t for the global factor. Figure 3.13 reports the same indicators for each region.

Looking at figure 3.12 we can notice that the variance explained by the global factor is bigger in the first period respect to the second, confirming the fact that the global factor is much more important during the pre-globalization period. We get a completely opposed conclusion looking to figure 3.13. While there is no support for the hypothesis of global convergence, there is a higher degree of synchronization in business cycle within the regions composed by industrialized countries, indeed North America, Europe and Asia and Oceania developed region. At the same time looking at the mean of global and at the mean of regional factors, see figure 3.11, emerges the presence of stronger regional-specific factor provides partial support to the decoupling hypothesis. Among other European countries seems to be more influenced by region-specific factor, in globalization period then they were in pre-globalization period. How can we explain a global decoupling? According to Kose et al. (2008) there were large common disturbances during the pre-globalization period, for example the two oil shocks and the inflation reduction with the associated increment in the interest rates that occurred in the early 1980's, see Goodfriend and King (2005). From the mid-1980s onward common global disturbances have become less important with an associated decline of the global factor. How can we explain regional convergence? The standard neoclassical growth model (Solow, 1956 and Swan, 1956) asserts that per-capita output across countries converges when they have similar preferences, technology levels and institutional and legal systems. Thus gaps in

Figure 3.13: *Upper graph: Regional φ_t (dotted line) against regional ζ_t (solid line) for the pre-globalization period (1950-1984); Bottom graph: Regional φ_t (dotted line) against regional ζ_t (solid line) for the globalization period (1985-2007).(A) North America Region; (B) Latin America Region; (C) European Region; (D) Asia and Oceania Developed Region; (E) Asia and Oceania Developing and Poor Region; (F) Africa Region.*



national or regional outputs must disappear over time. On the other hand, the endogenous growth model (Romer, 1990 and Grossman and Helpman, 1991) asserts that per-capita income is mainly influenced by country-specific factors with endogenously influence output dynamics. If this is the case, countries will not converge over time given that per-capita income only responds to country specific factors. The elimination of trade barriers and adoption of common trade, industrial, fiscal and monetary policies in the European union has spurred regional convergence. At the same time similar preferences and legal system drove United Kingdom far away form European Union and more related to USA (English speaking group). Finally, common level of technology (for example South Korea and Taiwan for the information technology), elimination of the trade barriers create a new economic area for Asia and Oceania developed countries. These facts provide support to the neoclassical growth model for those regions. Finally, the endogenous growth model seems to explain well the dynamics of China, that is an important exception in its area, and the dynamics of the poor countries with an absence of country or regional convergence.

3.5 Conclusions

In this paper we have employed a new Maximum Likelihood approach to estimate the latent factors and to study the dynamic comovement of macroeconomic aggregates in a broad cross section of countries. We provide an analysis of comovement across the world and across regions, for different periods and different cross section dimensions. Our paper also makes a methodological contribution as it provides a useful framework to study factors in a large scale data set with different pattern of missing values. We find that the global factor has become less important in macroeconomic fluctuation both for developed and poor countries. The analysis was done using a pre-globalization (1950-1984) and globalization (1985-2007) period in the same line of Kose et al. (2008). Moreover we find that the regional factors are more important in explaining the macroeconomic fluctuation for the industrialized countries during the globalization period. These results indicate that rising trade and financial integration does not bring to global convergence but to a *different level of convergence*, indeed a regional convergence of the rich countries and a divergence of the poor countries with respect to them.

To check the robustness of our methodology we carry out two experiment. First we estimate the same data set of Kose et al. (2004) and we get virtually their conclusions, indeed there is a significant common global factor that explains a substantial fraction of the variability of developed economies. Moreover there is no evidence of regional business cycle except for the North America region. Second we estimate the world and regional factors for the G7 countries in the same spirit of Gregory et al. (1997) and we show that our estimated global factor is different form the one estimated using the whole data set reaching the same conclusion of Kose et al. (2004). Finally we provide the regional factors for the G7 countries.

Using variance decomposition we provide the following results. First the global factor explains great part of the variability for the most industrialized countries, but is not very important for the developing and poor countries. Second there is evidence of an emerging European cycle, indeed a big portion of the volatility of the European aggregates can be attributed to a common European factor. Third, Japan seems to be detached from the global factor and great portion of its variability, during the period 1950-2007, is explained by the country specific component. Finally United Kingdom seems to be more related to the global factor and less with the European one.

State space methods allows us to compute model misspecification test and diagnostics from one step head prediction error even in presence of missing data. We provide the Ljung-Box statistic for our data set and we find that this model specification is a good enough to represent our series.

There are many other applications of our methodology, for example it can be used in the analysis of mixed frequency data set, see Proietti (2008), in order to provide an exact solution to the problem, or it can be used in every framework that require the analysis of large scale unbalanced panel in fast and efficient way. The extension of this technique in order to incorporate a notion of regime switching to capture the business cycle asymmetries, is far beyond the scope of the present paper and we leave it as topic of further research.

3.6 Appendix A: Dataset

Europe	
Country	Sample
Germany*,1	1970 - 2007
France*,1	1950 - 2007
Italy*,1	1950 - 2007
United Kingdom*,1	1950 - 2007
Sweden[1]	1950 - 2007
Switzerland[1]	1950 - 2007
Spain[1]	1950 - 2007
Portugal[1]	1950 - 2007
Norway[1]	1950 - 2007
Netherlands[1]	1950 - 2007
Luxembourg[1]	1950 - 2007
Ireland[1]	1950 - 2007
Iceland[1]	1950 - 2007
Greece[1]	1951 - 2007
Finland[1]	1950 - 2007
Denmark[1]	1950 - 2007
Belgium [1]	1950 - 2007
Austria [1]	1950 - 2007
Cyprus	1950 - 2007
Malta	1970 - 2007
Hungary	1970 - 2007
Romania	1960 - 2007
Poland	1970 - 2007
Bulgaria	1970 - 2007

West and Central Asia	
Country	Sample
Israel	1950 - 2007
Turkey	1950 - 2007
India[1]	1950 - 2007
Pakistan[1]	1950 - 2007
Sri Lanka[1]	1950 - 2007
Mauritius	1950 - 2007
Iran	1955 - 2007
Jordan	1954 - 2007
Bangladesh[1]	1959 - 2007
Iraq	1970 - 2007
Nepal	1960 - 2007
Oman	1970 - 2007
U.A.E.	1970 - 2007
Bahrain	1970 - 2006
Bhutan	1970 - 2007
Maldives	1970 - 2007
Mongolia	1970 - 2007
Saudi Arabia	1970 - 2007
Kuwait	1970 - 2007
Qatar	1970 - 2007
Syria	1960 - 2007

America	
Country	Sample
United States*,1	1950 - 2007
Mexico[1]	1950 - 2007
Brazil[1]	1950 - 2007
Argentina[1]	1950 - 2007
Bolivia[1]	1950 - 2007
Cambodia	1970 - 2007
Canada*,1	1950 - 2007
Chile[1]	1950 - 2007
Colombia[1]	1950 - 2007
Cuba	1970 - 2007
Jamaica[1]	1953 - 2007
Peru[1]	1950 - 2007
Paraguay[1]	1951 - 2007
Uruguay[1]	1950 - 2007
Venezuela[1]	1950 - 2007
Antigua	1970 - 2007
Belize	1970 - 2007
Barbados	1960 - 2007
Costa Rica[1]	1950 - 2007
Dominica	1970 - 2007
Dominica Republic[1]	1951 - 2007
Ecuador[1]	1951 - 2007
El Salvador[1]	1950 - 2007
Grenada	1970 - 2007
Guatemala[1]	1950 - 2007
Honduras[1]	1950 - 2007
Nicaragua	1950 - 2007
Panama[1]	1950 - 2007
Bahamas	1970 - 2007
Bermuda	1970 - 2007
Haiti	1960 - 2007
Puerto Rico	1950 - 2007
Trinidad Tobago[1]	1950 - 2007
Suriname	1970 - 2007
St. Lucia	1970 - 2007

Table 3.5: [(1)] *Corresponds to countries that are used in Kose et al. (2004), (*) Corresponds to G7 countries*

Africa	
Country	Sample
South Africa[1]	1950 - 2007
Egypt	1950 - 2007
Morocco[1]	1950 - 2007
Nigeria	1950 - 2007
Algeria	1960 - 2007
Central African Republic	1960 - 2007
Cote d'Ivoire[1]	1960 - 2007
Ethiopia	1950 - 2007
Madagascar	1960 - 2007
Rwanda	1960 - 2007
Senegal[1]	1960 - 2007
Somalia	1970 - 2007
Tunisia	1960 - 2007
Uganda	1950 - 2007
Cameroon[1]	1960 - 2007
Botswana	1960 - 2007
Benin	1959 - 2007
Burundi	1960 - 2007
Burkina Faso	1959 - 2007
Chad	1960 - 2007
Comoros	1960 - 2007
Dem. Rep. Congo	1950 - 2007
Republic of Congo	1960 - 2007
Equatorial Guinea	1960 - 2007
Gabon	1960 - 2007
Gambia	1960 - 2007
Ghana	1955 - 2007
Guinea	1959 - 2007
Guinea - Bissau	1960 - 2007
Kenya[1]	1950 - 2007
Liberia	1970 - 2007
Lesotho	1960 - 2007
Malawi	1954 - 2007
Mali	1960 - 2007
Mauritania	1960 - 2007
Namibia	1960 - 2007
Niger	1960 - 2007
Cape Verde	1960 - 2007
Mozambique	1960 - 2007
Sao Tome and Principe	1970 - 2007
Sierra Leone	1961 - 2007
Swaziland	1970 - 2007
Tanzania	1960 - 2007
Togo	1960 - 2007
Zambia	1955 - 2007
Zimbabwe[1]	1954 - 2007
Sudan	1970 - 2007

East Asia and Oceania	
Country	Sample
Japan[*,1]	1950 - 2007
Australia[1]	1950 - 2007
New Zealand[1]	1950 - 2007
Philippines[1]	1950 - 2007
Thailand[1]	1950 - 2007
Taiwan	1951 - 2007
China	1952 - 2007
Indonesia[1]	1960 - 2007
Hong Kong[1]	1960 - 2007
Malaysia[1]	1955 - 2007
Singapore[1]	1960 - 2007
Republic of Korea[1]	1953 - 2007
Brunei	1970 - 2007
Laos	1970 - 2007
Macao	1970 - 2007
Papua New Guinea	1960 - 2007
Kiribati	1970 - 2007
Samoa	1970 - 2007
Solomon Islands	1970 - 2007
Tonga	1970 - 2007
Vanuatu	1970 - 2007
Micronesia, Fed. Sts.	1970 - 2007

Table 3.6: *[(1)] Corresponds to countries that are used in Kose et al. (2004), (*) Corresponds to G7 countries*

3.7 Appendix B: Analytical Score

We provide a review the state space formulation for dynamic factor model in presence of missing values, this section draws heavily for Jungbacker et al. (2009).

The dynamic factor model given in (3.2) can be represented in matrix form as follows:

$$y_t = \Lambda f_t + u_t, \quad t = 1, \ldots, T, \tag{3.20}$$

where $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ is a matrix of factor loadings, f_t is a $q \times 1$ vector of unknown factors, T is the number of observations and N is the number of series. The common factor f_t are modelled as a stationary first order vector autoregressive ($VAR(1)$) process, and the error components u_t are modelled as vector autoregressive of order one process ($VAR(1)$):

$$u_{t+1} = \phi u_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_\varepsilon), \tag{3.21}$$

where ϕ is a $N \times N$ diagonal matrix and the disturbance variance matrix Σ_ε is $N \times N$ diagonal matrix of unknown parameters.

The model (3.20) and (3.21) can be represented in state space form:

$$\begin{aligned} y_t &= Z\alpha_t + u_t, \\ \alpha_{t+1} &= T\alpha_t + \eta_t \quad \eta \sim N(0, \Sigma_\eta), \end{aligned} \tag{3.22}$$

in order to get the representation (3.22) consider that f_t can be rewritten as a linear combination of the unobserved state α_t using a suitable full rank selection matrix G, in particular:

$$f_t = G\alpha_t, \tag{3.23}$$

where α_t has the following state space representation

$$\alpha_{t+1} = T\alpha_t + \eta_t, \quad \eta_t \sim N(0, \Sigma_\eta), \tag{3.24}$$

and Σ_η is the variance covariance matrix, of the states. In our particular case the factors and the latent states follow a $VAR(1)$ process and the selection matrix G is just the unitary matrix. Finally the Z matrix is given by the combination of the Λ and the selection matrix G, indeed $Z = \Lambda G$.

The likelihood function for this model can therefore be calculated by means of the Kalman filter while the unobserved factors can be estimated using the associated smoothing algorithm. The optimal properties of Kalman filter only apply when the observation equation (3.22) has disturbances u_t that are not serially correlated. To get read of the errors we can premultiply the state space formulation (3.22) by polynomial lag operator $(1 - \phi L)$:

$$\begin{aligned} y_t &= \phi y_{t-1} + Z\alpha_t - \phi Z\alpha_{t-1} + \varepsilon_t && \varepsilon_t \sim N(0, \Sigma_\varepsilon), \\ &= c_t + \begin{pmatrix} Z & -\phi Z \end{pmatrix} \begin{pmatrix} \alpha_t \\ \alpha_{t-1} \end{pmatrix} + \varepsilon_t \\ \begin{pmatrix} \alpha_{t+1} \\ \alpha_t \end{pmatrix} &= \begin{pmatrix} T & 0 \\ I & 0 \end{pmatrix} \begin{pmatrix} \alpha_t \\ \alpha_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_t \\ 0 \end{pmatrix} && \eta_t \sim N(0, \Sigma_\eta), \end{aligned} \tag{3.25}$$

this formulation keeps the dimension of the states under control. The equations (3.25) defines the state space model for the observed values, this state space formulation will be enlarged temporarily with the u's that accounts for the missing values.

Now take a vector of time series y_t for $t = 1, \ldots, T$, the following expression $y_t(o_t, o_{t-1})$ indicates the observations present at time t and time $t-1$. In the same way we have $y_t(m_t, m_{t-1})$ for missing values at time t and $t-1$. Moreover we have all the other possible combinations: $y_t(m_t, o_{t-1})$, $y_t(o_t, m_{t-1})$, $y_t(m_{t+1}, o_t)$.

We accomplish this notation with the following state space formulation:

$$\begin{pmatrix} y_t(o_t, o_{t-1}) \\ y_t(o_t, m_{t-1}) \end{pmatrix} = \begin{pmatrix} \phi_t^{(o)} y_t(o_t, o_{t-1}) \\ 0 \end{pmatrix} + \begin{pmatrix} Z(o_t, o_{t-1}) & -\phi_t^{(o)} Z(o_t, o_{t-1}) & 0 & 0 \\ Z(o_t, m_{t-1}) & 0 & I & 0 \end{pmatrix} \dot{\alpha}_t + \begin{pmatrix} \varepsilon(o_t, o_{t-1}) \\ 0 \end{pmatrix}, \tag{3.26}$$

with state equation $\dot{\alpha}_t$ given by:

$$\dot{\alpha}_{t+1} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \phi_t^{(*)} y_t(m_{t+1}, o_t) \end{pmatrix} + \begin{pmatrix} T & 0 & 0 & 0 \\ I & 0 & 0 & 0 \\ 0 & 0 & 0 & J_t \phi_t^{(1)}(m_t, m_t) \\ \phi_t^{(*)} Z_t^* & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_t \\ \alpha_{t-1} \\ u_t(o_t, m_{t-1}) \\ u(m_t) \end{pmatrix} + \begin{pmatrix} \eta_t \\ 0 \\ J_t \varepsilon_t(m_t) \\ \varepsilon_t(m_{t+1}, o_t) \end{pmatrix} \tag{3.27}$$

where

$$u_t(m_t) = \begin{pmatrix} u_t(m_t, m_{t-1}) \\ u_t(m_t, o_{t-1}) \end{pmatrix}, \tag{3.28}$$

J_t is a selection matrix of 0's and 1's and $\phi_t^{(o)}$ is a diagonal matrix. We include those entries of u_t in the state vector that correspond to missing entries in y_t and/or y_{t-1}. When both element of y_t and y_{t-1} are present the state space form collapse to the one given in (3.25). The transition from $u_t(m_t)$ to $u_{t+1}(m_t)$ is the autoregressive update (3.28), see Jungbacker et al. (2009) for a more formal treatment of these equations.

We can still apply the computational device of Jungbacker and Koopman (2008) to the missing value state space formulation to get a significant computational gain. Define

$$A_t^L = C_t^{-1} Z_t^{+\prime} \Sigma_{\varepsilon,t}^{-1}, \qquad \tilde{Z}_t = \left[Z(o_t, o_{t-1}; .), -\phi_t^{(o)} Z(o_t, o_{t-1}; .) \right], \qquad \Sigma_{\varepsilon,t} = \Sigma_{\varepsilon,t}(o_t, o_{t-1}; o_t, o_{t-1}), \tag{3.29}$$

and C_t is chosen such that

$$C_t C_t' = \tilde{Z}_t' \Sigma_{\varepsilon,t}^{-1} \tilde{Z}_t', \quad t = 1, \ldots, T. \tag{3.30}$$

The transformation A^L is applied only to the $y_t(o_t, o_{t-1})$ and does not require to consider the element of the state vector associated with the u_t since they do not affect $y_t(o_t, o_{t-1})$.

Define the matrix

$$A_t = \begin{pmatrix} A_t^L \\ A_t^H \end{pmatrix}, \tag{3.31}$$

where A_t^H is chosen as $A_t^H \Sigma_{\varepsilon,t} A_t^{H\prime} = 0$ and A_t is a full rank matrix. Applying the transformation matrix given in (3.31) to (3.26) we get:

$$\begin{pmatrix} A_t^L y_t(o_t, o_{t-1}) \\ A_t^H y_t(o_t, o_{t-1}) \end{pmatrix} = \begin{pmatrix} A_t^L c_t^o \\ A_t^H c_t^o \end{pmatrix} + \begin{pmatrix} C_t' G & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \dot{\alpha}_t + A_t \varepsilon_t(o_t, o_{t-1}), \tag{3.32}$$

where $\mathrm{Var}[A_t \varepsilon_t(o_t, o_{t-1})]$ is a block diagonal variance matrix with upper block given by $\mathrm{Var}[A_t \varepsilon_t(o_t, o_{t-1})] = I$, and $\dot{\alpha}_t$ is given in formula (3.27). It follows that the second part $A^H y_t$ does not depend on the

state vector and is not considered in the KFS step. From equation (3.32) we can split the likelihood as in formula (3.7).

We now derive the exact score for the state space model of formulation of (3.26) and (3.27). The model can be rewritten more compactly as:

$$\begin{aligned} y_t &= c_t + Z_t\alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, H_t), \\ \alpha_{t+1} &= d_t + T_t\alpha_t + R_t\eta_t, \qquad \eta_t \sim N(0, Q_t), \end{aligned} \tag{3.33}$$

where $\alpha_1 \sim N(a_{1|0}, P_{1|0})$ as initial conditions, and $t = 1, \ldots, T$.

Taking the first derivatives of the likelihood respect to the system vectors and matrices c_t, d_t, Z_t, T_t, H_t and Q_t we obtain the following derivatives:

$$\begin{aligned} &\frac{\partial l(y)}{\partial d_t} = \tilde{R}_t Q_t^{-1}\tilde{R}_t(a_{t+1|n} - T_t a_{t|n} - d_t), && \frac{\partial l(y)}{\partial T_t} = \tilde{R}_t Q_t^{-1}\tilde{R}_t(M_{T_t} + d_t a_{t|n} - T_t M_{Z_t}), \\ &\frac{\partial l(y)}{\partial c_t} = \tilde{H}_t^{-1}(y_t - Z_t a_{t|n} - c_t), && \frac{\partial l(y)}{\partial Z_t} = \tilde{H}_t^{-1}((y_t - c_t)a_{t|n}^{'} - Z_t M_{Z_t}), \\ &\frac{\partial l(y)}{\partial Q_t} = Q_t^{-1}M_{Q_t}Q_t^{-1} - \frac{1}{2}diag\{Q_t^{-1}M_{Q_t}Q_t^{-1}\}, \\ &\frac{\partial l(y)}{\partial H_t} = \tilde{H}_t^{-1}M_{H_t}\tilde{H}_t^{-1} - \frac{1}{2}diag\{\tilde{H}_t^{-1}M_{H_t}\tilde{H}_t^{-1}\}, \end{aligned} \tag{3.34}$$

with:

$$\begin{aligned} &M_{Q_t} = E(\eta_t\eta_t^{'}|y_1, ..., y_n) - Q_t \qquad && M_{T_t} = a_{t+1|n}a_{t|n}^{'} + P_{t+1,t|n} \\ &M_{H_t} = (y_t - c_t - Z_t a_{t|n})(y_t - c_t - Z_t a_{t|n})^{'} + Z_t P_{t|n} Z_t^{'} - \tilde{H}_t \qquad && M_{Z_t} = a_{t|n}a_{t|n}^{'} + P_{t|n}. \end{aligned} \tag{3.35}$$

$\tilde{R}$ is a selection matrix and the matrix $\tilde{H}_t$ is constructed using the $y_t(o_t, m_{t-1})$ as index due to singularity of H_t. Finally the matrix M_{Q_t} can be evaluated by $\eta_t = \tilde{R}(\alpha_{t+1} - T_t\alpha_t - d_t)$ and by $P_{t+1,t|n}$ using formula (3.12) and $P_{t|n}$.
The system vectors and matrices further depends on our parameters of interest so applying the chain rule, see Magnus and Neudecker (2007), we have:

$$\begin{aligned} &\frac{\partial vec(l(y))}{\partial vec(T)} = \left(\frac{\partial vec(l(y))}{\partial vec(T_t)}\right)^{'}\left(\frac{\partial vec(T_t)}{\partial vec(T)}\right)^{'} && \frac{\partial vec(l(y))}{\partial vec\left(\phi_t^{(o)}\right)} = \left(\frac{\partial vec(l(y))}{\partial vec(c_t)}\right)^{'}\left(\frac{\partial vec(c_t)}{\partial vec\left(\phi_t^{(o)}\right)}\right)^{'} + \left(\frac{\partial vec(l(y))}{\partial vec(Z_t)}\right)^{'}\left(\frac{\partial vec(Z_t)}{\partial vec\left(\phi_t^{(o)}\right)}\right)^{'} \\ &\frac{\partial vec(l(y))}{\partial vec(Z_t)} = \left(\frac{\partial vec(l(y))}{\partial vec(Z_t)}\right)^{'}\left(\frac{\partial vec(Z_t)}{\partial vec(Z)}\right)^{'} && \frac{\partial vec(l(y))}{\partial vec\left(\phi_t^{(*)}\right)} = \left(\frac{\partial vec(l(y))}{\partial vec(d_t)}\right)^{'}\left(\frac{\partial vec(d_t)}{\partial vec\left(\phi_t^{(*)}\right)}\right)^{'} + \left(\frac{\partial vec(l(y))}{\partial vec(T_t)}\right)^{'}\left(\frac{\partial vec(T_t)}{\partial vec\left(\phi_t^{(*)}\right)}\right)^{'} \\ &\frac{\partial vec(l(y))}{\partial vec\left(\phi_t^{(1)}\right)} = \left(\frac{\partial vec(l(y))}{\partial vec(T_t)}\right)^{'}\left(\frac{\partial vec(T_t)}{\partial vec\left(\phi_t^{(1)}\right)}\right)^{'}. \end{aligned} \tag{3.36}$$

Chapter 4

Bayesian model selection

4.1 Introduction

Nelson and Plosser (1982) analyzed fourteen annual US macroeconomic time series for the presence of unit root using the Dickey-Fuller (1979) method. They pointed out that American macroeconomic time series present, quasi systematically, stochastic tendencies and unit root properties. Since this seminal work researchers have paid considerable attention to the presence of stochastic trends in macroeconomic variables, influencing the way how macroeconomists think about trends and short-run fluctuations.

If a time series contains a unit root, it should be analyzed after transforming it by taking first difference; this kind of time series is called a differences-stationary (DS) process. If a time series does not have a unit root and can be described using a time series model which includes a purely deterministic trend, it is called a trend-stationary (TS) process. The main reason to investigate unit roots in univariate time series is that the DS processes assume that shocks have a permanent effect, while such shocks only have a transitory effect for TS processes. Nelson and Plosser (1982) found that only one of the fourteen time series they investigate, the Unemployment series is TS and hence most of time series analyzed can be described by DS processes. Then, why testing, almost thirty years later, the presence of unit roots in the same macroeconomic series using Bayesian model selection? There are two main reasons for that.

The first concerns the distinction between trend stationary and difference stationary that is critical in some contexts. For example Real GNP is a particularly important series to investigate, the evidence for or against the existence of a unit root in this series provides support for the validity of competitive macroeconomic theories, see Mocan (1994).

The second reason relates to some researches that advocate forcefully for Bayesian alternatives over the classical approaches such as the ADF test, see Koop (1992), Sims (1988) and Sims and Uhlig (1991). These economists cited several advantages of the Bayesian approach over the classical approach. For example, it is well known that ADF tests have low power against plausible alternatives, especially against trend-stationary alternatives. The Bayesian approach, on the other hand, would reveals that both, the unit root and the trend-stationary hypotheses, would receive similar posterior probabilities, providing a more reasonable summary of sample informations. Another problem with the classical unit root tests is the discontinuity of the classical asymptotic theory in the presence of a unit root. The Bayesian approach is based on the likelihood function that does

not have the same discontinuity problem, moreover since it is conditional on the observed sample provides exact small sample results. This is not possible in classical tests because they suffer from a small sample bias, see Koop (1994) for a detailed discussion.

Despite the apparent advantages of the Bayesian approach over the classical one in unit-root testing only a relative small number of papers have appeared using this approach. This is mainly due to the critic of Phillips (1990). He argues that the selection between competitive models is very sensible to the chosen priors. Schotman and van Dijk (1991) gives some examples how the prior can be misleading in model selection.

Recently there has been an increasing interest in Bayesian model selection; Frühwirth-Schnatter and Wagner (2009) provided a new testing methodology for state space models. Model specification, is a challenging task for these models as one has to specify which components to include and to decide whether they are fixed or time-varying. Unfortunately in this context some regularity conditions are violated and standard criteria, like AIC and BIC, cannot be applied. The Bayesian approach is, in principle, able to deal with such non-regular testing problems. Given K models $M_1, \ldots, M_K$ as the potential data generating processes of a given time series $\mathbf{y} = \{y_1, \ldots, y_t\}$, we assign a prior probability $p(M_k)$ to each model that is necessary to derive the posterior probability $p(M_k|\mathbf{y})$. The model with the highest posterior probability is most likely to describe the data.

The main objective of this paper is to show the potentiality of the Bayesian approach to investigate whether stochastic trends are present in economic time series. We run two experiments and in each we test the Nelson and Plosser (1982) data set and a similar data set with updated observations. The first experiment provides a modify version of the test allowing for autoregressive parameters, that adds some dynamics to the available models. In this case the test can be seen as the Bayesian counterpart of the Leybourne et al. (1999) test and an extension of the test proposed by Koop and Van Dijk (2000). The second experiment uses seven competitive models but does not allow for autoregressive components.

The remainder of the paper is organized as follows. In section 4.2 we briefly review the methodology proposed by Frühwirth-Schnatter and Wagner (2009), the modified version of the test is also provided. In section 4.3 we discuss the prior specification and the Bayesian estimation. In section 4.4 we provide the empirical results of our testing procedure. In section 4.5 we provide a robustness analysis against different prior specifications. In section 4.6 we provide a Monte Carlo experiment. Section 4.7 concludes the chapter. The dataset is presented in Appendix A.

4.2 Methodology

Leybourne et al. (1999) suggested an optimal method for selecting the order of the autoregressive components, p, in the autoregressive integrated moving average model, $(ARIMA(p,d,q))$, on which the stationarity test is based. They considered a local level model with autoregressive parameters:

$$\begin{aligned} &\Phi(L)y_t = \mu + \alpha_t + \varepsilon_t, && \varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2), \\ &\alpha_t = \alpha_{t-1} + \eta_t, && \eta_t \sim i.i.d.(0, \sigma_\eta^2), \\ &\alpha_0 = 0, && t = 1, \ldots, T, \end{aligned} \tag{4.1}$$

where ε_t and η_t being mutually uncorrelated. Frühwirth-Schnatter and Wagner (2009) showed that the selection procedure proposed by Shively et al. (1999) can be extended to state space models.

Starting from a local linear trend model:

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \\ \alpha_t &= \alpha_t + \beta_t + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2), \\ \beta_t &= \beta_{t-1} + \zeta_t, & \zeta_t &\sim N(0, \sigma_\zeta^2), \end{aligned} \tag{4.2}$$

and writing the model in non-centered form, see section 4.2, they introduce indicator variables that are useful to add or delete the latent processes. This specification encompasses different models, for example, without the latent processes the model collapses to a $N(0, \sigma_\varepsilon^2)$. Usually economic time series have more dynamics then it is allows by the specification given in (4.2), this suggests to modify the test to get a more appropriate tool for empirical researches.

Extending the state space formulation (4.2) with autoregressive parameters provides the Bayesian counterpart of Leybourne et al. (1999) test and an extension of the Koop and Van Dijk (2000) test. The main advantage with respect to the Leybourne et al. (1999) test is that we do not apply a two-step procedure but we can, jointly, estimate all the model parameters and test for the presence of unit roots, moreover our test allows for the $I(2)$ specification.

4.2.1 General model

Consider modeling a time series $\mathbf{y} = (y_1, \ldots, y_T)$ where T is the number of observations, through the local linear trend model:

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \\ \mu_t &= \mu_{t-1} + a_{t-1} + \eta_t, & \eta_t &\sim N(0, \theta_1), \\ a_t &= a_{t-1} + \kappa_t, & \kappa_t &\sim N(0, \theta_2), \end{aligned} \tag{4.3}$$

where μ_t follows a random walk with a random drift starting from unknown initial values μ_0 and a_0. This model can be shown to be second-order equivalent in moments to the $ARIMA(0, 2, 2)$ process

$$(1 - L)^2 y_t = (1 - \theta L)^2 \zeta_t, \quad \zeta_t \sim N(0, \sigma_\zeta^2), \tag{4.4}$$

with complicated restrictions on the moving average roots, see Harvey (1989) for a detailed discussion.

The foregoing model can be extended to more general local linear trend model with autoregressive components:

$$\begin{aligned} \Phi(L) y_t &= \mu_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \\ \mu_t &= \mu_{t-1} + a_{t-1} + \eta_t, & \eta_t &\sim N(0, \theta_1), \\ a_t &= a_{t-1} + \kappa_t, & \kappa_t &\sim N(0, \theta_2), \end{aligned} \tag{4.5}$$

where $\Phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ is a pth-order autoregressive polynomial in the lag operator L with the roots outside the unit circle, clearly this formulation is equivalent to an $ARIMA(p, 2, 2)$ process. This model can be rewritten in the following form:

$$\begin{aligned} y_t &= \mu_t + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2), \\ \mu_t &= \mu_{t-1} + a_{t-1} + \eta_t, & \eta_t &\sim N(0, \theta_1), \\ a_t &= a_{t-1} + \kappa_t, & \kappa_t &\sim N(0, \theta_2), \end{aligned} \tag{4.6}$$

and finally in state space form as follows:

$$\begin{aligned} y_t &= X\beta + Z\alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\ \alpha_t &= T\alpha_{t-1} + W_t, \qquad W_t \sim N(0, Q), \end{aligned} \tag{4.7}$$

where

$$\begin{aligned} &X = \begin{pmatrix} y_{t-1} & \dots & y_{t-p} \end{pmatrix}, \quad \beta = \begin{pmatrix} \phi_1 & \dots & \phi_p \end{pmatrix}', \quad \alpha_t = \begin{pmatrix} \mu_t & a_t \end{pmatrix}, \\ &Z = \begin{pmatrix} 1 & 0 \end{pmatrix}, \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad Q = \begin{pmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{pmatrix}. \end{aligned} \tag{4.8}$$

The quantities given by $\begin{pmatrix} y_{t-1} & \dots & y_{t-p} \end{pmatrix}$ are constant because they are known at time t.

Following Frühwirth-Schnatter and Wagner (2009) we can rewrite the state space form (4.6) in non-centered form. The latter transforms the latent states into scale free and location free processes and transfers both location and scale to the measurement equation. The benefits are twofold. The non-centered representation enables the treatment of the stochastic levels and trends as regression effects, secondly it improves the convergence properties of the MCMC.

Define two independent random walk processes $\tilde{\mu}_t$ and $\tilde{a}_t$ with standard normal independent increments as well as an integrated process $\tilde{A}_t$:

$$\begin{aligned} \tilde{\mu}_t &= \tilde{\mu}_{t-1} + \tilde{\omega}_{1t}, \quad \tilde{\omega}_{1t} \sim N(0,1), \\ \tilde{a}_t &= \tilde{a}_{t-1} + \tilde{\omega}_{2t}, \quad \tilde{\omega}_{2t} \sim N(0,1), \\ \tilde{A}_t &= \tilde{A}_{t-1} + \tilde{a}_{t-1}. \end{aligned} \tag{4.9}$$

All the processes are assumed to start at zero: $\tilde{\mu}_0 = \tilde{a}_t = \tilde{A}_t = 0$. Combine the state equation (4.9) with the following observed equation:

$$y_t = \mu_0 + a_0 t + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \sqrt{\theta_1}\tilde{\mu}_t + \sqrt{\theta_2}\tilde{A}_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \tag{4.10}$$

where the constant μ_0 and the slope a_0 are equal to the initial values of the level and the drift component and θ_1 and θ_2 are equal to the variances of the latent processes defined in equation (4.6), in the following we assume an $AR(2)$ specification for the autoregressive part. Define the indicators vector $\Upsilon = \begin{pmatrix} \delta_0 & \delta_1 & \delta_2 & \gamma_1 & \gamma_2 \end{pmatrix}$ that are used to include or delete, the time trend, the autoregressive parameters, the stochastic trend and the stochastic slope, more concisely to select between competitive models. It turns out that the available models is equal to $2^5 = 32$. We restrict our analysis to 21 models avoiding the combinations with $\delta_1 = 0$ that corresponds to $\phi_1 = 0$, it is clear that these specifications can be easily handled in this framework.

The general model becomes:

$$\begin{aligned} y_t &= \mu_0 + \delta_0 a_0 t + \delta_1\phi_1 y_{t-1} + \delta_2\phi_2 y_{t-2} + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\ \tilde{\mu}_t &= \tilde{\mu}_{t-1} + \tilde{\omega}_{1t}, \quad \tilde{\omega}_{1t} \sim N(0,1), \\ \tilde{a}_t &= \tilde{a}_{t-1} + \tilde{\omega}_{2t}, \quad \tilde{\omega}_{2t} \sim N(0,1), \\ \tilde{A}_t &= \tilde{A}_{t-1} + \tilde{a}_{t-1}, \end{aligned} \tag{4.11}$$

that brings to the following 21 competitive models:

$$
\begin{aligned}
&(Model\ 1)\ y_t = \mu_0 + \varepsilon_t && \Upsilon = (0,0,0,0,0)\\
&(Model\ 2)\ y_t = \mu_0 + \delta_1\phi_1 y_{t-1} + \varepsilon_t && \Upsilon = (0,1,0,0,0)\\
&(Model\ 3)\ y_t = \mu_0 + \delta_1\phi_1 y_{t-1} + \delta_2\phi_2 y_{t-2} + \varepsilon_t && \Upsilon = (0,1,1,0,0)\\
&(Model\ 4)\ y_t = \mu_0 + a_0 t + \varepsilon_t && \Upsilon = (1,0,0,0,0)\\
&(Model\ 5)\ y_t = \mu_0 + \delta_0 a_0 t + \delta_1\phi_1 y_{t-1} + \varepsilon_t && \Upsilon = (1,1,0,0,0)\\
&(Model\ 6)\ y_t = \mu_0 + \delta_0 a_0 t + \delta_1\phi_1 y_{t-1} + \delta_2\phi_2 y_{t-2} + \varepsilon_t && \Upsilon = (1,1,1,0,0)\\
&(Model\ 7)\ y_t = \mu_0 + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \varepsilon_t && \Upsilon = (0,0,0,1,0)\\
&(Model\ 8)\ y_t = \mu_0 + \delta_1\phi_1 y_{t-1} + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \varepsilon_t && \Upsilon = (0,1,0,1,0)\\
&(Model\ 9)\ y_t = \mu_0 + \delta_1\phi_1 y_{t-1} + \delta_2\phi_2 y_{t-2} + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \varepsilon_t && \Upsilon = (0,1,1,1,0)\\
&(Model\ 10)\ y_t = \mu_0 + \delta_0 a_0 t + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \varepsilon_t && \Upsilon = (1,0,0,1,0)\\
&(Model\ 11)\ y_t = \mu_0 + \delta_0 a_0 t + \delta_1\phi_1 y_{t-1} + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \varepsilon_t && \Upsilon = (1,1,0,1,0)\\
&(Model\ 12)\ y_t = \mu_0 + \delta_0 a_0 t + \delta_1\phi_1 y_{t-1} + \delta_2\phi_2 y_{t-2} + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \varepsilon_t && \Upsilon = (1,1,1,1,0)\\
&(Model\ 13)\ y_t = \mu_0 + \delta_0 a_0 t + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (1,0,0,1,1)\\
&(Model\ 14)\ y_t = \mu_0 + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (0,0,0,1,1)\\
&(Model\ 15)\ y_t = \mu_0 + \delta_1\phi_1 y_{t-1} + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (0,1,0,1,1)\\
&(Model\ 16)\ y_t = \mu_0 + \delta_1\phi_1 y_{t-1} + \delta_2\phi_2 y_{t-2} + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (0,1,1,1,1)\\
&(Model\ 17)\ y_t = \mu_0 + \delta_0 a_0 t + \delta_1\phi_1 y_{t-1} + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (1,1,0,1,1)\\
&(Model\ 18)\ y_t = \mu_0 + \delta_0 a_0 t + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (1,0,0,0,1)\\
&(Model\ 19)\ y_t = \mu_0 + \delta_0 a_0 t + \delta_1\phi_1 y_{t-1} + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (1,1,0,0,1)\\
&(Model\ 20)\ y_t = \mu_0 + \delta_0 a_0 t + \delta_1\phi_1 y_{t-1} + \delta_2\phi_2 y_{t-2} + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (1,1,1,0,1)\\
&(Model\ 21)\ y_t = \mu_0 + \delta_0 a_0 t + \delta_1\phi_1 y_{t-1} + \delta_2\phi_2 y_{t-2} + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (1,1,1,1,1).
\end{aligned}
\tag{4.12}
$$

We can rewrite formulation (4.11) and (4.12) as follows:

$$
\begin{aligned}
&y_t = \tilde{X}\beta(\delta) + Z(\gamma)\boldsymbol{\alpha}_t + \varepsilon_t, && \varepsilon_t \sim N(0,\sigma_\varepsilon^2),\\
&\boldsymbol{\alpha}_t = T\boldsymbol{\alpha}_{t-1} + \xi_t, && \xi_t \sim N(\mathbf{0}, Q),
\end{aligned}
\tag{4.13}
$$

where $\boldsymbol{\alpha}_0 = \mathbf{0}_{3\times 1}$ and:

$$
\begin{aligned}
&Z(\gamma) = \begin{pmatrix}\gamma_1\sqrt{\theta_1} & 0 & \gamma_2\sqrt{\theta_2}\end{pmatrix}, && \tilde{X} = \begin{pmatrix}1 & t & y_{t-1} & y_{t-2}\end{pmatrix},\\
&\beta(\delta) = \begin{pmatrix}1 & \delta_0 t & \delta_1\phi_1 & \delta_2\phi_2\end{pmatrix}', && \boldsymbol{\alpha}_t = \begin{pmatrix}\tilde{\mu}_t & \tilde{a}_t & \tilde{A}_t\end{pmatrix}',\\
&T = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 1 & 1\end{pmatrix}, && Q = \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{pmatrix},
\end{aligned}
\tag{4.14}
$$

clearly this state space form changes depending on the chosen model, see Frühwirth-Schnatter and Wagner (2009).

4.3 Prior and Bayesian estimation

To perform the Bayesian estimation we have to choose a prior distribution for the indicators $p(\delta, \gamma)$, for the autoregressive parameters ϕ_1 and ϕ_2, for the $\sqrt{\theta_1}$ and $\sqrt{\theta_2}$ parameters and for the irregular variance σ_ε^2. Choosing the priors to perform a unit root test is a quite challenging problem as explained by Phillips (1990) where he showed that flat priors are not uninformative but unwittingly introduce a tendency toward stationary models. Schotman and van Dijk (1991) pointed out that improper priors, like the uniform and the Jeffrey's prior, are less suited for Bayesian inference on a sharp null hypothesis as the unit root.

Our framework is quite different, but the problem of choosing the right priors still remains. In fact choosing a sensible prior is very important in model selection. In this work we follow the rule of thumb mentioned in Koop (2003): *when comparing models it is acceptable to use uninformative priors over parameters which are common to all models. However, informative, proper priors should be used over all other parameters.*

The prior used in our study can be summarized as follows:

$$\begin{pmatrix}\theta_1\\ \theta_2\\ a_0\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\right) \qquad \begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} \sim \mathrm{TN}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right) \tag{4.15}$$

the remaining prior is $p(\mu_0) \propto 1$, a few comments are in order. First, we use a partially proper prior which combines the improper prior for the constant parameter μ_0 with a proper prior on the remaining unrestricted parameters, this implies some modifications to the standard marginal likelihood that are described below. Second we choose for the autoregressive parameters, ϕ_1 and ϕ_2, the truncate normal prior. $Beta(\cdot,\cdot)$ prior does not bring to closed-form expressions for the posteriors and for the marginal likelihood $p(\mathbf{y}|\delta, \gamma, \boldsymbol{\alpha})$. Drawing from the truncated Normal distribution, rather then simply a Normal adds a slight complication. However drawing from the truncated Normal distribution can be done by drawing from the untruncated variant and simply discarding the draws which fall outside the stationary region. Provided that the autoregressive parameters lie within (or not too far outside) the stationary region, this strategy works well. In our examples we have small amount of rejections and this algorithm is not very inefficient. To be more conservative about observation variance σ_ε^2, we follow the advise of Ley and Steel (2009) and Frühwirth-Schnatter and Wagner (2009). We choose the hierarchical prior because it increases the flexibility and decreases the dependence on prior assumptions. The prior for the irregular variance is given by $\sigma_\varepsilon^2 \sim IG(c_0/2, C_0/2)$ where $C_0 \sim G(g_0, C_0)$ with $c_0 = 2.5$, $g_0 = 5$ and $G_0 = (g_0/0.75 \cdot \mathrm{Var}(\mathbf{y})(c_0 - 1))$. The hierarchical prior requires an additional sampling step where C_0 is sampled conditional on σ_ε^2 from the conditional Gamma posterior $C_0|\sigma_\varepsilon^2 \sim G(g_0+c_0, G_0+1/\sigma_\varepsilon^2)$ at each sweep of the sample.

Define the following quantities, $\beta = \begin{pmatrix}\mu_0 & a_0 & \phi_1 & \phi_2 & \sqrt{\theta_1} & \sqrt{\theta_2}\end{pmatrix}'$ as the regression coefficients, $(\delta, \gamma) = \begin{pmatrix}1 & \delta_0 & \delta_1 & \delta_2 & \gamma_1 & \gamma_2\end{pmatrix}$ as the related indicators and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)$ as the non-centered latent states, see (4.13) and (4.14). The measurement equation defined in (4.13) can be written more compactly:

$$y_t = z_t^{\delta,\gamma}\beta^{\delta,\gamma} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \tag{4.16}$$

where $\beta^{\delta,\gamma} = \begin{pmatrix}\mu_0 & \delta_0 a_0 & \delta_1\phi_1 & \delta_2\phi_2 & \gamma_1\sqrt{\theta_1} & \gamma_2\sqrt{\theta_2}\end{pmatrix}'$ and $z_t^{\delta,\gamma} = \begin{pmatrix}1 & t & y_{t-1} & y_{t-2} & \tilde{\mu} & \tilde{A}_t\end{pmatrix}$, when all the indicators take value one. Otherwise the restricted parameters $\beta^{\delta,\gamma}$ and the corresponding predictors $z^{\delta,\gamma}$ contain the elements corresponding to the indicators equal to 1, e.g. $\beta^{\delta,\gamma} =$

$\begin{pmatrix}\mu_0 & \delta_0 a_0 & \delta_1\phi_1 & 0 & \gamma_1\sqrt{\theta_1} & 0\end{pmatrix}'$ brings to $z_t^{\delta,\gamma} = \begin{pmatrix}1 & t & y_{t-1} & 0 & \tilde{\mu} & 0\end{pmatrix}$. Equation (4.16) define a regression model that under the the conjugate priors:

$$\beta^{\delta,\gamma} \sim N(a_0^{\delta,\gamma}, A_0^{\delta,\gamma}), \quad \sigma_\varepsilon^2 \sim IG(c_0/2, C_0/2), \tag{4.17}$$

has the following marginal likelihood, see Geweke (2005):

$$p(\mathbf{y}_{T^*}|\delta,\gamma,\boldsymbol{\alpha}) = \frac{1}{(\pi)^{T^*/2}} \times \frac{|A_{T^*}^{\delta,\gamma}|^2}{|B_0^{\delta,\gamma}|^2} \frac{\Gamma(c_{T^*})(C_0^{c_0})}{\Gamma(c_0)(C_{T^*}^{c_{T^*}})}. \tag{4.18}$$

The quantities $A_{T^*}^{\delta,\gamma}$, c_{T^*} and $C_{T^*}^{\delta,\gamma}$ are the posterior moments of $\beta^{\delta,\gamma}$ and σ_ε^2 given below in (4.19). $\Gamma(\cdot)$ is the gamma function, $B_0^{\delta,\gamma}$ is a matrix containing the prior variances for the unrestricted parameters and T^* is given by the number of observations minus two, where two is the maximum number of autoregressive parameters. Finally the subscript T^* underlines the fact that we lose two observations, e.g. $\mathbf{y}_{T^*} = \{y_3, \ldots, y_t\}$. Under the conjugate prior the posterior moments are given by the following expressions:

$$\begin{aligned}
A_{T^*}^{\delta,\gamma} &= \left((Z_{T^*}^{\delta,\gamma})'(\sigma_\varepsilon^2)^{-1} Z_{T^*}^{\delta,\gamma} + (A_0^{\delta,\gamma})^{-1}\right)^{-1}, \\
a_{T^*}^{\delta,\gamma} &= A_{T^*}^{\delta,\gamma}\left((Z_{T^*}^{\delta,\gamma})'(\sigma_\varepsilon^2)^{-1}\mathbf{y}_{T^*} + (A_0^{\delta,\gamma})^{-1} a_0^{\delta,\gamma}\right), \\
c_{T^*} &= \frac{1}{2}(c_0 + T^*), \\
C_{T^*}^{\delta,\gamma} &= \frac{1}{2}(C_0 + (\mathbf{y}_{T^*} - Z_{T^*}^{\delta,\gamma} a_{T^*}^{\delta,\gamma})'(\mathbf{y}_{T^*} - Z_{T^*}^{\delta,\gamma} a_{T^*}^{\delta,\gamma})).
\end{aligned} \tag{4.19}$$

The Gibbs sampler can be sketched as follows:

(a) Draw the indicator $\begin{pmatrix}\delta_1 & \delta_2 & \delta_3 & \delta_4 & \gamma_1 & \gamma_2\end{pmatrix}$ the constant parameters $\begin{pmatrix}\mu_0 & a_0 & \phi_1 & \phi_2\end{pmatrix}$, the variance parameters $\sqrt{\theta_1}$ and $\sqrt{\theta_2}$ and the observation variance σ_ε^2 jointly in one block:

- Sample the indicator $p(\delta,\gamma|\mathbf{y}_{T^*},\boldsymbol{\alpha})$ obtained from the Bayes' theorem $p(\delta,\gamma|\mathbf{y}_{T^*},\boldsymbol{\alpha}) \propto p(\mathbf{y}_{T^*}|\delta,\gamma,\boldsymbol{\alpha})p(\delta,\gamma)$ where $p(\mathbf{y}_{T^*}|\delta,\gamma,\boldsymbol{\alpha})$ is the marginal likelihood defined in (4.18) and $p(\delta,\gamma)$ is the prior, uniform in our case.
- Sample σ_ε^2 from $IG(c_{T^*}, C_{T^*}^{\delta,\gamma})$ where c_{T^*} and C_{T^*} are given in equation (4.19);
- Conditional on σ_ε^2 sample μ_0 and a_0 (if unrestricted) and the unrestricted variance parameters $\sqrt{\theta_1}$ and $\sqrt{\theta_2}$ jointly from the normal posterior $N(a_{T^*}^{\delta,\gamma}, A_{T^*}^{\delta,\gamma})$ moreover sample ϕ_1 and ϕ_2 jointly from a $\text{TN}(a_{T^*}^{\delta,\gamma}, A_{T^*}^{\delta,\gamma})$ where a_{T^*} and A_{T^*} are given in equation (4.19). Notice that under our prior specification $a_0^{\delta,\gamma} = 0$ it follows that the quantity $(A_0^{\delta,\gamma})^{-1}a_0^{\delta,\gamma}$ of equation (4.19) can be safely removed, moreover $A_0 = \begin{pmatrix} 0 & \\ & B_0 \end{pmatrix}$;
- Set to 0 all the parameters that do not enter to the selected model.

(b) Sample $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)$ from the state space form (4.14) using the simulation smoothers of Carter and Kohn (1994), De Jong and Shepard (1995) or Durbin and Koopman (2002);

(c) Perform a random switch with probability 0.5 for $\sqrt{\theta_1}$ and $\tilde{\mu}_t, t = 1, \ldots, T$. Thus with probability 0.5 the draws of these parameters remain unchanged, while they are substituted by $-\sqrt{\theta_1}$ and $-\tilde{\mu}_t$ with the same probability. Perform another random sign switch for $\sqrt{\theta_2}$, $\tilde{a}$ and $\tilde{A}$ for $t = 1, \ldots, T$;

(d) Go to step (a).

Step (c) is necessary due to non-identifiability of the non-centered reparametrization. The sign of the variance parameters $\sqrt{\theta_i}$ for $i = 1, \ldots, k$, and the corresponding latent states can change by multiplying all the elements by -1 without changing the integrated likelihood, see Frühwirth-Schnatter and Wagner (2009) for a detailed discussion. In our study we use a uniform prior over all the possible models although other priors are possible, Ley and Steel (2009) propose to use a Beta prior that should outperform the uniform prior over large regression models.

4.4 Empirical Results

In this section we apply the testing procedure to different models and to different datasets. We consider first all the possible models given in (4.12), and then a subset of seven competitive models encompassed in specification (4.12). For each study we test two different datasets that are reviewed in Appendix A. To speed up the computation we do not sample from the components that are not selected, the program is written in Ox v. 5.10 console (Doornik, 2007) using our source code. We run the unrestricted model without variable selection for the first 1000 draws, this allows us to generate sensible starting values. Then the results are obtained by running the MCMC sampling for 40000 iterations after a burn-in period of 20000. The time needed with $T = 500$ observations is less then 25 minutes in a standard desktop computer. Subsection 4.1 presents the results for the first experiment and subsection 4.2 reports the results for the second experiment.

4.4.1 Full models

Table 4.1 reports the results for the 21 competitive models showed in section 2. The upper part of table 4.1 reports the results for Nelson and Plosser (1982) data set and it shows that the stationary hypothesis is rejected for almost all the series. Among them just three are stationary, the unemployment, the industrial production and the velocity of the money. The first one is stationary and is well represented by an $AR(2)$ process. The second one turns out to be better represented by a trend stationary process (TS) with one autoregressive parameter that takes value around 0.82. The third one has not clear evidence of a unit root, in this case the selected model is a first order autoregressive process ($AR(1)$) with root equal to 0.96 . Note that models with ϕ_1 very close to 1 still produce $I(0)$ series, where the memory decays exponentially, to deal with these borderline cases fractionally-integrated models have been introduced, see Palma (2007) for a review. Using fractional integrated processes, Caporale et al. (2005) showed that the velocity of the money is likely to be a stationary process with a very low value of the estimated fractional integration parameter. Our testing procedure cannot handle fractional integration but it seems to pick up the stationary representation for this series. As far as the other series are concerned we cannot reject the hypothesis of non-stationarity, confirming the results of Nelson and Plosser (1982).

The bottom part of table 4.1 reports the same experiment for the updated dataset. Among all the series the just the monthly hours worked turns out to be a stationary series that is better fitted by an $AR(2)$ process with autoregressive roots equal to $\phi_1 = 0.68$ and $\phi_2 = 0.26$. All the other series are non-stationary of order one, $I(1)$, with just the monthly CPI that is an $I(2)$ process, this is consistent with the monthly inflation rate selected as an $I(1)$ process.

The inflation rate, one of the most analyzed macroeconomic variables, has been an important subject of research, see among other Culver and Papell (1997). Whether inflation is best described as a stationary or unit root process has a number of economic implications. Conventional sticky-price models, such as Dornbusch (1976) imply stationarity of price level. While the *higher-order* Phillips curves of Calvo (1983) requires a stationary representation for the inflation series. Table 4.1 shows that the monthly and the quarterly inflation series are clearly non-stationary, with a different short-run behaviour. In fact the monthly inflation series is estimates as a local level model with one autoregressive parameter equal to 0.30, that corresponds to an $ARIMA(1,1,1)$ process.

The Phillips curve, based on the Non Accelerating Inflation Rate of Unemployment (NAIRU), is a fundamental instrument to regulate the inflation pressures, see Stiglitz (1997). The NAIRU is unobserved and has to be estimated using the unemployment series. For this reason it is very important to study the short and long run behaviour of this series. A new approach to evaluate the marginal likelihood that avoids MCMC has been proposed by Fiorentini et al. (2008). They showed, using this technique, that the unemployment series is better described by a random walk without drift. With our framework we can carry out the same analysis allowing for autoregressive effects, accordingly to our results the model that better fit the unemployment series is a local level model with two autoregressive roots equal to $\phi_1 = 0.60$ and $\phi_2 = 0.26$. This model corresponds to an $ARIMA(2,1,1)$ showing that the autoregressive roots together with stochastic trend are very important to describe the dynamics of this series.

4.4.2 Restricted models

Recall the models presented in section 4.2 and apply the right restriction on the vector $\Upsilon = \begin{pmatrix}\delta_0 & \delta_1 & \delta_2 & \gamma_1 & \gamma_2\end{pmatrix}$ we get the following subset of models:

$$\begin{aligned}
&(Model\ 1)\ y_t = \mu_0 + \varepsilon_t && \Upsilon = (0,0,0,0,0)\\
&(Model\ 2)\ y_t = \mu_0 + a_0 t + \varepsilon_t && \Upsilon = (1,0,0,0,0)\\
&(Model\ 3)\ y_t = \mu_0 + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \varepsilon_t && \Upsilon = (0,0,0,1,0)\\
&(Model\ 4)\ y_t = \mu_0 + \delta_0 a_0 t + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \varepsilon_t && \Upsilon = (1,0,0,1,0)\\
&(Model\ 5)\ y_t = \mu_0 + \delta_0 a_0 t + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (1,0,0,0,1)\\
&(Model\ 6)\ y_t = \mu_0 + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (1,0,0,1,1)\\
&(Model\ 7)\ y_t = \mu_0 + \delta_0 a_0 t + \gamma_1\sqrt{\theta_1}\tilde{\mu}_t + \gamma_2\sqrt{\theta_2}\tilde{A}_t + \varepsilon_t && \Upsilon = (1,0,0,1,1)
\end{aligned} \tag{4.20}$$

Table 4.2 reports the results for Nelson and Plosser data set (1982) and the results for the updated data set.

Looking at the upper part of table 4.2 just two series are stationary, although the evidence is quite different. On the one hand the unemployment is clearly selected to be stationary, on the

Table 4.1: *Upper table: percent results for Nelson and Plosser data set. Bottom table: percent results for the updated data set.*

Series	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	M_{13}	M_{14}	M_{15}	M_{16}	M_{17}	M_{18}	M_{19}	M_{20}	M_{21}
Rgnp	0	0	0	0	0	29.5	0	0.5	1.8	0	0.6	51.3	0	0	0.1	8.8	0.1	0	0	3.4	3.8
Ngnp	0	0	0	0	0	13.8	0	0	5.5	0	0	40.7	0	0	0	14.3	0	0	0	12.9	12.7
Rpcgnp	0	0	0	0	0	27.8	0	1.5	19.1	0	0.6	35.5	0	0	0.2	9.5	0.2	0	0	2.4	3.1
Iprod	0	0	0	0	48.2	0	0	0	0	0	49	0.5	0	0	0.3	0	0.8	0	0	0.8	0.1
Emply	0	0	0	0	0	26.4	0	1	2.5	0	3.1	57	0	0	0.27	2.6	0.2	0	0	2.8	4
Unempl	0	4.3	61.5	0	0	0	0	2.5	30.6	0	0	0	0	0	0	0.8	0	0	0	0.2	0.1
Gnpdefl	0	0	0	0	0	9.4	0	11.8	17	0	13.1	26.8	0	0	2.5	10.1	1.6	0	0	3.6	4.2
Pcons	0	0	0	0	0	0	0	0	30.2	0	0	13.7	0	0	0	20	0	0	0	18.3	17.8
Nwage	0	0	0	0	0	15.1	0	0.2	4.8	0	0.3	34.3	0	0	0.1	21.6	0	0	0	12.1	11.4
Rwage	0	0	0	0	6.6	17.6	0	8.3	0	0	30.7	27.5	0	0	2.8	2.6	1.4	0	0	1.1	1.1
Money	0	0	0	0	0	32.3	0	0	1.7	0	0.1	50.8	0	0	0	4.8	0.1	0	0	4.9	5.1
Veloc	0	39.4	0	0	0	0	0	31.6	2.2	0	8.1	2.8	0	0	1.9	5.4	0.7	0	0.2	4.1	3.4
Interest	0	0	0	0	0	0	0	0.31	0	0	0.1	0	0	0	62.7	0	17.4	0	19.5	0	0
Pstock	0	0	0	0	0	1.6	0	0.1	1.6	0	0	13.8	0	0	0	49.1	0	0	0	17.3	16.4

Series	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	M_{13}	M_{14}	M_{15}	M_{16}	M_{17}	M_{18}	M_{19}	M_{20}	M_{21}
MInfl	0	0	0	0	0	0	0	99.8	0.2	0	0	0	0	0	0	0	0	0	0	0	0
QInfl	0	0	0	0	0	0	99.7	0.3	0	0	0	0	0	0	0	0	0	0	0	0	0
Awhman	0	0	99.9	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0
Unempl	0	0.1	0	0	0	0	0	0	99.9	0	0	0	0	0	0	0	0	0	0	0	0
GNP	0	0	20.1	0	0	0	0	0	0	0	60.2	18.8	0	0	0	0	0.8	0	0	0	0.1
GDP	0	0	35.4	0	0	0	0	0	0	0	44.4	20.1	0	0	0	0	0.1	0	0	0	0
CPI	0	0	0	0	0	0	0	0	16.6	0	0	0	0	0	5.8	77	0	0	0	0.3	0.3
Mort	0	0	0	0	0	0	0	0	99.6	0	0	0.4	0	0	0	0	0	0	0	0	0
IP	0	0	0	0	0	0	0	95.7	0.5	0	3.8	0	0	0	0	0	0	0	0	0	0
ExcR	0	0	1	0	0	0.3	0	15	71.8	0	1.6	6.4	0	0	0.6	2.6	0.1	0	0	0	0.6
M2	0	0	0	0	0	0	0	99.8	0	0	0.2	0	0	0	0	0	0	0	0	0	0

Table 4.2: *Upper table: percent results for Nelson and Plosser data set. Bottom table: percent results for the updated data set.*

Series	M_1	M_2	M_3	M_4	M_5	M_6	M_7
$Rgnp$	0	1	0	37	42.9	4	15
$Ngnp$	0	16	0	33	27.7	2.1	21
$Rpcgnp$	0	0	0	42.3	36.2	1.3	20
$Iprod$	0	0	88	0.7	0	10	0.6
$Emply$	0	1.6	0	37	32.3	2.7	26.3
$Unempl$	49	0	37.5	3.6	2.6	3.6	3.5
$Gnpdefl$	0	31.2	0	34	16.7	2.7	15.4
$Pcons$	0	0.5	0.1	17.6	26.2	0.6	55.4
$Nwage$	0	32.7	0	30.3	18.5	2	16.4
$Rwage$	0	0.6	0	25.9	44.2	8.4	20.8
$Money$	0	15.8	0	38.1	16	3.9	26.1
$Veloc$	0	10.2	0.3	19	15.5	4	50.8
$Interest$	14.8	0	22.7	4.2	10.5	3.6	44.1
$Pstock$	0	0.8	0.2	8.1	13.1	25.5	52.2

Series	M_1	M_2	M_3	M_4	M_5	M_6	M_7
$MInfl$	0	0	100	0	0	0	0
$QInfl$	0	0	99.99	0.01	0	0	0
$Unempl$	0	0	1.9	6.7	12.3	7.6	71.5
GNP	0	16.5	0	20.1	7.1	5.5	51
CPI	0	0.3	0	2.2	4.2	1.9	91.3
GDP	0	0	0	16.6	40.4	0.37	42.5
$ExcRat$	0	0	0	0	12	79	9
IP	0	0	0	0.3	0.4	0	99.3
$Awhman$	0	48	7.4	15.5	7.9	8.1	12.4
$M2$	0	0		0	0	0	100
$Mort$	0	0	0	14.6	11.4	9.8	64

other hand the percentages of the nominal wages are very spread out between model 2 and model 8. The testing procedure underlines that the other series are non-stationary, with the CPI selected to be a $I(2)$ process.

Table 4.2, bottom part, reports the results for the updated data set. The table shows that the unemployment rate is selected to be $I(2)$ in sharp contrast with table 4.1 and Fiorentini et al. (2008). The testing procedure carried out on a restrict set of models seems to overfit the non-stationary order. This can be explained by the sum of the autoregressive parameters equal to $\phi_1 + \phi_2 = 0.86$ that guides the selection procedure to overfit the integration order. The same behaviour can be found in the industrial production series, where model (8) of equation (4.12) better fits the data, with $\phi_1 = 0.86$. As far as the monthly and quarterly inflation series are concerned, these series are always selected as non-stationary. Looking at table 4.1 and table 4.2 we can notice that the quarterly inflation series is estimated as a local level model in both cases. On the contrary the monthly inflation series is estimated to be a local level model with an autoregressive root in the extended version of the test. Finally the monthly hours worked is always selected as a stationary process.

Table 4.3: *The most frequently visited models for diffent prior specifications.*

Prior	Selected Model	Frequency in percentage
$p(\mu_0) \propto 1$, $B_0 = 0.1$	8	90
$p(\mu_0) \propto 1$, $B_0 = 1$	8	99.8
$p(\mu_0) \propto 1$, $B_0 = 10$	8	100
$p(\mu_0) \propto 1$, $B_0 = 100$	8	100
$p(\mu_0) \propto 1$, $B_0 = 1000$	8	100

4.5 Robustness analysis

This section provides a robustness analysis of the testing procedure under different priors specification. The robustness analysis is carried out using the monthly inflation series. Table 4.3 provides clear evidence in favor of the unit root hypothesis, despide the fact that it has been considered a wide range of priors. It can been seen that most evidence for a unit root is found when the prior allocates less weight to the region near zero, in this case the percentage of the choice increases, this results is consistent with Frühwirth-Schnatter and Wagner (2009). For a discussion about robustness and prior specification for the Nelson and Plosser data set as well has for seasonal time series see Koop and Van Dijk (2000).

4.6 Monte Carlo Experiment

This section illustrates the results of Monte Carlo experiment for the testing procedure presented in this chapter.

4.6.1 Experiment Design

The series to be estimated and selected are generated by an $ARIMA(1,1,1)$ process:

$$y_t = \mu_0 + a_0 t + \phi_1 y_{t-1} + \sqrt{\theta_1}\tilde{\mu}_t + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \quad t = 1, \ldots, T. \tag{4.21}$$

for different value of a_0, $\sqrt{\theta_1}$, ϕ_1 and different sample sizes. The design of the experiment is the following. The log of the normalized parameter is considered for the deterministic trend and for the stochastic trend variance, indeed $\log\left(a_0/\sigma_\varepsilon^2\right)$ and $\log\left(\sqrt{\theta_1}/\sigma_\varepsilon^2\right)$. Their values are given by 10^i where $i = 2, \ldots, -6$ and therefore they range between 100 to 10^{-6}. The other parameters are, the autoregressive components that range from 0 to 0.90; the constant parameter μ_0 and the irregular variance σ_ε^2 are fixed to 1 in all the simulations. Finally the considered sample sizes are $T = 100$, $T = 250$ and $T = 500$. To ensure that the testing procedure do not select the best model just by chance, we simulate, estimate and a select a model, at each parameters combination, for 20 times.

Due to the computational burdensome of this technique, 20 times is the maximum number of repetitions that we can handle before the Monte Carlo experiment become unfeasible.

The simulated exercise proceeds as follows:

1 For $T = 100$, $T = 250$ and $T = 500$:

(a) For $\phi_1 = 0$, $\phi_1 = 0.25$, $\phi_1 = 0.50$, $\phi_1 = 0.75$ and $\phi_1 = 0.90$, repeat 20 times the following:

For each value of $\log\left(\sqrt{\theta_1}/\sigma_\varepsilon^2\right) = 10^i$, $i = 100, \ldots, -6$, simulate, estimate and select the model for all the possible values of $\log\left(a_0/\sigma_\varepsilon^2\right) = 10^i$, $i = 100, \ldots, -6$. This gives a total of $9 \times 9 = 81$ combinations.

(b) Report the most frequently chosen model for each parameters combination and go to step (a);

2 Go to step (1).

4.6.2 Monte Carlo Results

Table 4.4 reports the results for $T = 100$, table 4.5 and table 4.6 report the results for $T = 250$ and $T = 500$ respectively. The tables also report the number of times (in percentage) the model has been selected in 20 iterations. This percentage varies across models and across parameter combinations.

Table 4.4 shows that with a small sample size the test is biased for high values of the deterministic trend and for the stochastic trend variance. In these cases the percentages are very low at least for the first two columns of each panel of table 4.4. This indicates the difficulty of the testing procedure to select correctly a model with these extreme parameter values. As the autoregressive parameter rises towards 0.90, in combination with high values for the stochastic trend variance, the procedure spuriously picks up two autoregressive roots with very high percentages, this behaviour is still present with $T = 250$ and $T = 500$. The selection procedure, despite this bias, always selects the non-stationary model when the stochastic trend variance is different from zero. As this parameter decreases the model with deterministic trend becomes more important.

As the values of both parameters (deterministic trend and stochastic trend variance) go to zero, the procedure selects the stationary model. In this case the autoregressive order is always rightly chosen and the number of times the model has been selected is very high.

Table 4.5 reports the same exercise with 250 observations, and it shows that the bias introduced by high values of the stochastic trend variance is less pronounced with respect to table 4.4. For example the first columns of the first panel of table 4.5 shows that in some cases the procedure selects the right model. Correspondingly to very high values for the deterministic trend and for the stochastic trend variance, the selection procedure still picks up the wrong models, with a tendency to overfit the non-stationary order. As the values of these parameters decrease the right model is selected with an increasing percentage.

Table 4.6 reports the results for 500 observations. In this case the problem with the extreme parameter values still remains, although less pronounced. Table 4.6 shows that the procedure in this case is more precise and very often picks up the right model.

To conclude, tables 4.4, 4.5 and 4.6 show that, as the sample size increases, the precision of our testing procedure increases. High values of the stochastic trend variance brings to overestimate the autoregressive parameters. Moreover the procedure spuriously selects an $I(2)$ process in the case of high values for the stochastic trend variance and for the autoregressive parameters. This bias is less severe with 500 observations.

In the case of small values for the deterministic trend and for the stochastic trend variance, e.g. 10^{-6}, the procedure always selects the right autoregressive order but is oriented to stationary models. This is consistent with our simulation scheme that assumes the irregular variance σ_ε^2 fixed to 1; in these cases it becomes the most important source of variability.

Table 4.4: *Results for simulated series with 100 observations. First panel $\phi_1 = 0$, second panel $\phi_1 = 0.25$, third panel $\phi_1 = 0.5$, forth panel $\phi_1 = 0.75$ and final panel $\phi_1 = 0.90$. The rows are the values for the deterministic trend. The columns are the values for the stochastic trend variance. Percentage of times the model is selected in brackets.*

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	5 (50)	11 (75)	10 (55)	4 (70)	4 (95)	4 (100)	4 (100)	4 (100)	4 (100)
10	15 (45)	5 (55)	10 (45)	10 (85)	4 (100)	4 (95)	4 (100)	4 (100)	4 (95)
1	15 (45)	11 (50)	10 (95)	10 (70)	10 (80)	4 (95)	4 (100)	4 (95)	4 (100)
0.1	8 (35)	8 (75)	10 (75)	10 (75)	10 (90)	10 (95)	4 (95)	4 (100)	4 (100)
0.01	8 (55)	8 (55)	7 (90)	7 (60)	7 (50)	4 (50)	4 (75)	7 (65)	7 (90)
0.001	8 (60)	8 (65)	7 (80)	7 (45)	1 (95)	1 (100)	1 (100)	1 (100)	1 (70)
0.0001	8 (65)	8 (65)	7 (85)	7 (50)	1 (95)	1 (100)	1 (95)	1 (100)	1 (100)
10^{-5}	8 (50)	8 (40)	7 (90)	1 (50)	1 (90)	1 (95)	1 (90)	1 (95)	1 (95)
10^{-6}	5 (55)	8 (65)	7 (80)	1 (60)	1 (95)	1 (100)	1 (95)	1 (100)	1 (100)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	5 (26)	5 (45)	5 (85)	10 (80)	10 (65)	5 (50)	4 (60)	10 (70)	4 (70)
10	11 (25)	12 (55)	5 (80)	10 (60)	10 (65)	10 (75)	10 (55)	4 (90)	4 (65)
1	11 (35)	11 (40)	11 (45)	11 (50)	11 (50)	5 (65)	5 (60)	10 (50)	5 (55)
0.1	8 (20)	15 (40)	11 (45)	11 (50)	10 (45)	5 (50)	5 (55)	5 (45)	5 (40)
0.01	15 (45)	11 (45)	8 (55)	8 (60)	5 (55)	7 (55)	8 (50)	7 (45)	8 (60)
0.001	15 (25)	8 (40)	8 (55)	8 (60)	5 (60)	2 (45)	2 (55)	2 (50)	2 (45)
0.0001	15 (30)	8 (44)	8 (55)	8 (45)	2 (40)	2 (50)	2 (65)	2 (50)	2 (60)
10^{-5}	11 (20)	8 (45)	8 (40)	8 (45)	2 (60)	2 (45)	2 (60)	2 (55)	2 (55)
10^{-6}	11 (35)	8 (50)	8 (45)	8 (75)	2 (45)	2 (65)	2 (45)	2 (60)	2 (50)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	5 (80)	5 (90)	5 (100)	5 (100)	5 (100)	5 (85)	5 (85)	5 (100)	5 (100)
10	15 (45)	11 (66)	5 (55)	11 (85)	11 (70)	5 (100)	5 (100)	11 (75)	11 (95)
1	15 (42)	9 (35)	11 (75)	11 (75)	5 (85)	5 (90)	5 (90)	5 (90)	5 (80)
0.1	8 (50)	9 (50)	11 (80)	11 (55)	5 (95)	5 (80)	5 (80)	11 (75)	5 (90)
0.01	8 (50)	8 (38)	8 (85)	11 (35)	11 (40)	11 (50)	5 (95)	8 (65)	5 (40)
0.001	11 (40)	8 (84)	8 (80)	8 (65)	11 (80)	2 (95)	2 (50)	2 (90)	2 (95)
0.0001	11 (40)	8 (70)	8 (95)	8 (45)	8 (90)	2 (90)	2 (90)	2 (100)	2 (90)
10^{-5}	15 (40)	8 (65)	8 (80)	8 (45)	2 (85)	2 (90)	2 (90)	2 (85)	2 (95)
10^{-6}	8 (70)	8 (60)	8 (90)	8 (45)	2 (90)	2 (90)	2 (80)	2 (95)	2 (95)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	6 (57)	6 (56)	5 (100)	5 (90)	5 (40)	5 (100)	5(90)	5 (90)	5 (100)
10	12 (50)	6 (50)	11 (75)	5 (100)	5 (50)	5 (75)	5 (85)	5 (90)	5 (95)
1	16 (45)	6 (55)	11 (75)	11 (85)	5 (100)	5 (100)	5 (100)	11 (100)	5 (95)
0.1	16 (55)	12 (42)	11 (80)	11 (75)	5 (100)	11 (100)	11 (95)	11 (95)	5 (55)
0.01	15 (50)	12 (45)	11 (85)	11 (40)	8 (90)	8 (95)	8 (90)	2 (95)	2 (40)
0.001	15 (60)	12 (55)	8 (90)	8 (55)	8 (65)	8 (85)	2 (65)	2 (70)	2 (80)
0.0001	12 (60)	9 (38)	8 (75)	8 (35)	2 (55)	2 (45)	2 (40)	2 (60)	2 (80)
10^{-5}	12 (45)	12 (40)	8 (95)	8 (35)	2 (80)	2 (75)	2 (80)	2 (75)	2 (70)
10^{-6}	6 (85)	12 (53)	8 (90)	8 (35)	2 (75)	2 (90)	2 (90)	2 (75)	2 (85)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	12 (53)	12 (85)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)
10	16 (60)	12 (53)	11 (65)	11 (85)	5 (100)	5 (100)	5 (95)	5 (100)	5 (95)
1	16 (75)	12 (73)	11 (80)	11 (80)	5 (95)	5 (100)	5 (95)	5 (90)	5 (95)
0.1	6 (70)	16 (75)	11 (80)	11 (60)	11 (45)	5 (45)	5 (40)	11(40)	5 (50)
0.01	16 (55)	16 (68)	11 (75)	11 (50)	11 (65)	11 (60)	5 (45)	8 (50)	5 (60)
0.001	16 (75)	12 (78)	8 (90)	8 (80)	8 (75)	8 (60)	2 (75)	2 (65)	2 (75)
0.0001	16 (70)	9 (70)	8 (85)	8 (60)	8 (85)	2 (75)	2 (75)	2 (80)	2 (85)
10^{-5}	16 (65)	9 (77)	8 (75)	8 (50)	2 (80)	2 (80)	2 (60)	2 (70)	2 (75)
10^{-6}	16 (80)	9 (63)	8 (83)	8 (60)	2 (75)	2 (60)	2 (75)	2 (65)	2 (75)

Table 4.5: *Results for simulated series with 250 observations. First panel $\phi_1 = 0$, second panel $\phi_1 = 0.25$, third panel $\phi_1 = 0.5$, forth panel $\phi_1 = 0.75$ and final panel $\phi_1 = 0.90$. The rows are the values for the deterministic trend. The columns are the values for the stochastic trend variance. Percentage of times the model is selected in brackets.*

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	5 (55)	5 (85)	5 (100)	10 (75)	4 (100)	4 (100)	4 (100)	4 (100)	4 (95)
10	15 (40)	11 (80)	10 (75)	10 (85)	4 (100)	4 (100)	4 (100)	4 (100)	4 (100)
1	8 (30)	11 (80)	10 (90)	10 (65)	4 (95)	4 (90)	4 (100)	4 (100)	4 (95)
0.1	10 (35)	8 (85)	10 (80)	10 (75)	7 (100)	4 (95)	4 (95)	4 (95)	4 (100)
0.01	10 (55)	8 (90)	10 (80)	10 (60)	7 (80)	7 (80)	4 (70)	4 (75)	4 (100)
0.001	8 (50)	8 (95)	7 (80)	7 (85)	1 (90)	1 (95)	7 (100)	1 (95)	1 (100)
0.0001	10 (45)	8 (85)	7 (90)	7 (100)	1 (90)	1 (100)	1 (100)	1 (90)	1 (95)
10^{-5}	15 (55)	8 (90)	7 (95)	7 (100)	1 (95)	1 (100)	1 (100)	1 (100)	1 (100)
10^{-6}	8 (30)	8 (85)	7 (70)	7 (90)	1 (90)	1 (100)	1 (100)	1 (100)	1 (95)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	5 (60)	5 (80)	5 (80)	10 (100)	10 (80)	5 (60)	4 (75)	4 (80)	4 (80)
10	12 (50)	12 (60)	11 (50)	10 (95)	10 (70)	5 (85)	4 (90)	4 (80)	4 (80)
1	11 (55)	9 (45)	11 (75)	11 (55)	11 (80)	5 (75)	5 (85)	5 (90)	5 (80)
0.1	16 (40)	9 (70)	11 (35)	11 (45)	11 (85)	5 (90)	5 (80)	5 (100)	5 (95)
0.01	11 (30)	9 (55)	8 (75)	8 (55)	8 (45)	8 (70)	5 (65)	5 (60)	5 (60)
0.001	11 (30)	9 (65)	8 (55)	8 (45)	8 (95)	2 (100)	2 (80)	2 (80)	2 (100)
0.0001	16 (40)	9 (75)	8 (65)	8 (60)	8 (80)	2 (90)	2 (90)	2 (90)	2 (85)
10^{-5}	8 (35)	9 (85)	8 (75)	8 (80)	8 (100)	2 (100)	2 (90)	2 (95)	2 (90)
10^{-6}	17 (50)	9 (80)	8 (55)	7 (45)	7 (100)	2 (90)	2 (85)	2 (90)	2 (90)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	12 (55)	6 (70)	5 (95)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)
10	9 (40)	12 (95)	11 (40)	11 (40)	5 (95)	5 (95)	5 (100)	5 (95)	5 (95)
1	9 (65)	9 (75)	11 (100)	11 (95)	11 (100)	5 (95)	5 (100)	5 (100)	5 (95)
0.1	9 (55)	9 (95)	11 (60)	11 (100)	8 (100)	5 (100)	5 (100)	5 (100)	5 (100)
0.01	9 (70)	9 (90)	8 (100)	11 (95)	8 (85)	8 (100)	5 (90)	5 (100)	5 (100)
0.001	9 (45)	9 (95)	8 (90)	8 (95)	2 (55)	2 (55)	2 (70)	8 (60)	2 (80)
0.0001	9 (60)	9 (90)	8 (95)	8 (95)	2 (90)	2 (100)	2 (100)	2 (100)	2 (100)
10^{-5}	9 (65)	9 (95)	8 (100)	8 (100)	2 (90)	2 (100)	2 (100)	2 (95)	2 (100)
10^{-6}	9 (50)	9 (90)	8 (95)	8 (100)	2 (90)	2 (100)	2 (100)	2 (100)	2 (100)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	12 (55)	6 (90)	6 (99)	6 (96)	6 (100)	6 (100)	6 (100)	6 (100)	6 (100)
10	15 (55)	12 (90)	11 (55)	11 (65)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)
1	16 (40)	9 (75)	11 (98)	11 (65)	5 (95)	5 (95)	5 (100)	5 (95)	5 (100)
0.1	16 (60)	12 (84)	11 (85)	11 (55)	5 (100)	5 (90)	8 (95)	5 (95)	5 (90)
0.01	15 (40)	9 (85)	8 (90)	8 (80)	8 (55)	8 (60)	8 (50)	5 (55)	5 (65)
0.001	15 (55)	9 (90)	8 (95)	8 (65)	8 (90)	2 (100)	2 (95)	2 (100)	2 (85)
0.0001	15 (40)	9 (84)	11 (90)	8 (85)	2 (95)	2 (95)	2 (90)	2 (95)	2 (100)
10^{-5}	15 (35)	9 (79)	8 (95)	8 (60)	2 (100)	2 (95)	2 (100)	2 (95)	2 (95)
10^{-6}	15 (55)	9 (65)	11 (95)	8 (70)	2 (100)	2 (90)	2 (100)	2 (95)	2 (95)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	12 (73)	6 (45)	6 (100)	5 (85)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)
10	16 (95)	12 (95)	11 (63)	11 (70)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)
1	16 (65)	9 (84)	11 (90)	11 (80)	5 (90)	5 (100)	5 (95)	5 (95)	5 (100)
0.1	12 (75)	9 (70)	11 (75)	11 (85)	5 (70)	5 (55)	5 (65)	5 (65)	5 (85)
0.01	16 (85)	9 (72)	11 (70)	8 (60)	8 (80)	5 (40)	5 (50)	5 (55)	5 (45)
0.001	12 (65)	9 (70)	11 (85)	8 (75)	8 (65)	2 (85)	2 (85)	2 (80)	2 (80)
0.0001	16 (70)	9 (68)	8 (90)	8 (85)	2 (90)	2 (90)	2 (95)	2 (90)	2 (95)
10^{-5}	16 (80)	16 (52)	8 (90)	8 (80)	2 (100)	2 (85)	2 (85)	2 (90)	2 (95)
10^{-6}	16 (45)	16 (52)	8 (100)	2 (70)	2 (90)	2 (85)	2 (90)	2 (95)	2 (95)

Table 4.6: *Results for simulated series with 500 observations. First panel $\phi_1 = 0$, second panel $\phi_1 = 0.25$, third panel $\phi_1 = 0.5$, forth panel $\phi_1 = 0.75$ and final panel $\phi_1 = 0.90$. The rows are the values for the deterministic trend. The columns are the values for the stochastic trend variance. Percentage of times the model is selected in brackets.*

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	10 (55)	11 (90)	10 (95)	10 (95)	4 (100)	4 (100)	4 (100)	4 (100)	4 (100)
10	10 (40)	11 (95)	10 (70)	10 (100)	4 (100)	4 (100)	4 (100)	4 (100)	4 (100)
1	10 (50)	11 (50)	10 (95)	10 (100)	10 (100)	4 (100)	4 (100)	4 (100)	4 (100)
0.1	7 (60)	8 (80)	10 (65)	10 (95)	10 (100)	10 (95)	4 (100)	4 (95)	4 (95)
0.01	8 (55)	8 (95)	10 (100)	10 (100)	10 (100)	10 (100)	10 (100)	4 (60)	4 (55)
0.001	8 (50)	8 (95)	7 (100)	7 (90)	7 (55)	7 (60)	7 (70)	1 (95)	4 (95)
0.0001	8 (55)	8 (100)	7 (80)	7 (95)	7 (90)	1 (95)	1 (100)	1 (100)	1 (95)
10^{-5}	8 (50)	8 (90)	7 (85)	7 (95)	1 (95)	1 (100)	1 (100)	1 (100)	1 (100)
10^{-6}	8 (70)	8 (75)	7 (95)	7 (90)	1 (85)	1 (65)	1 (100)	1 (100)	1 (98)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	11 (50)	5 (80)	5 (65)	11 (100)	5 (85)	5 (85)	4 (95)	4 (80)	4 (85)
10	11 (40)	11 (95)	11 (85)	11 (86)	11 (45)	5 (50)	5 (75)	5 (55)	5 (65)
1	11 (45)	11 (50)	11 (75)	11 (90)	11 (100)	5 (95)	5 (100)	5 (100)	5 (100)
0.1	11 (65)	9 (95)	11 (85)	11 (90)	11 (100)	11 (100)	5 (100)	5 (100)	5 (100)
0.01	8 (55)	9 (80)	11 (60)	11 (90)	8 (95)	11 (95)	5 (100)	5 (90)	5 (100)
0.001	9 (60)	9 (85)	8 (70)	8 (95)	8 (60)	8 (70)	8 (55)	8 (60)	2 (50)
0.0001	9 (70)	9 (70)	8 (60)	8 (100)	8 (85)	2 (100)	2 (95)	2 (95)	2 (100)
10^{-5}	9 (70)	9 (95)	8 (65)	8 (95)	2 (90)	2 (100)	2 (100)	2 (100)	2 (100)
10^{-6}	9 (60)	9 (90)	8 (65)	8 (95)	2 (100)	2 (100)	2 (100)	2 (100)	2 (95)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	11 (55)	6 (70)	5 (95)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)
10	9 (40)	12 (95)	11 (40)	11 (40)	5 (95)	5 (95)	5 (100)	5 (95)	5 (95)
1	11 (65)	9 (75)	11 (100)	11 (95)	5 (100)	5 (95)	5 (100)	5 (100)	5 (95)
0.1	11 (55)	9 (95)	11 (60)	11 (100)	11 (100)	11 (100)	5 (100)	5 (100)	5 (100)
0.01	9 (70)	9 (90)	11 (100)	11 (95)	11 (85)	11 (100)	5 (90)	5 (100)	5 (100)
0.001	9 (45)	9 (95)	8 (90)	11 (95)	11 (55)	11 (55)	8 (70)	8 (60)	2 (80)
0.0001	9 (60)	9 (90)	8 (95)	8 (95)	8 (90)	2 (100)	2 (100)	2 (100)	2 (100)
10^{-5}	9 (65)	9 (95)	8 (100)	8 (100)	2 (90)	2 (100)	2 (100)	2 (95)	2 (100)
10^{-6}	9 (50)	9 (90)	8 (95)	8 (100)	2 (90)	2 (100)	2 (100)	2 (100)	2 (100)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	12 (50)	6 (90)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)
10	9 (30)	12 (95)	11 (60)	11 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (95)
1	9 (60)	9 (60)	11 (100)	11 (75)	11 (100)	5 (95)	5 (90)	5 (100)	5 (100)
0.1	9 (50)	9 (95)	11 (60)	11 (70)	11 (95)	11 (100)	5 (100)	5 (100)	5 (100)
0.01	9 (60)	9 (95)	11 (95)	11 (85)	11 (85)	11 (100)	5 (90)	5 (100)	5 (100)
0.001	9 (70)	9 (100)	11 (95)	11 (95)	11 (55)	11 (70)	8 (60)	2 (100)	2 (95)
0.0001	9 (60)	9 (95)	8 (95)	8 (90)	2 (95)	2 (100)	2 (100)	2 (100)	2 (100)
10^{-5}	9 (40)	9 (85)	8 (95)	8 (80)	2 (65)	2 (100)	2 (100)	2 (100)	2 (95)
10^{-6}	9 (50)	9 (90)	8 (95)	8 (95)	2 (85)	2 (100)	2 (100)	2 (65)	2 (95)

	100	10	1	0.1	0.01	0.001	0.0001	10^{-5}	10^{-6}
100	12 (84)	6 (84)	5 (100)	11 (53)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)
10	12 (31)	12 (92)	11 (61)	11 (100)	5 (100)	5 (100)	5 (100)	5 (100)	5 (100)
1	9 (61)	9 (61)	11 (100)	11 (92)	11 (100)	5 (100)	5 (100)	5 (100)	5 (92)
0.1	9 (38)	9 (77)	11 (61)	11 (92)	11 (92)	5 (100)	5 (100)	5 (100)	5 (92)
0.01	16 (54)	9 (100)	11 (77)	11 (61)	11 (100)	11 (92)	5 (92)	5 (92)	5 (84)
0.001	9 (46)	9 (84)	11 (70)	11 (84)	11 (53)	8 (69)	5 (53)	2 (61)	2 (70)
0.0001	9 (61)	9 (84)	8 (84)	8 (92)	2 (92)	2 (100)	2 (100)	2 (100)	2 (100)
10^{-5}	9 (38)	9 (100)	8 (92)	8 (92)	2 (100)	2 (92)	2 (92)	2 (92)	2 (100)
10^{-6}	9 (46)	9 (92)	8 (92)	8 (69)	2 (84)	2 (100)	2 (100)	2 (92)	2 (100)

4.7 Conclusions

There is no doubt that unit root tests do suffer from low power in many situations of interests. Researchers generally agree that the Bayesian approach offers an alternative and useful way to the classical approach in empirical modeling.

In this paper, we provide a modify version of the Bayesian test proposed by Frühwirth-Schnatter and Wagner (2009). This test can be seen as the Bayesian counterpart of the Leybourne et al. (1999) test and an extended version of the test proposed by Koop and Van Dijk (2000). After a detailed discussion about the prior specification we applied this test to Nelson and Plosser (1982) data set and to an updated data set.

Empirical evidence, using real data, indicates that this approach is both simple to use and yields reasonable results. For the Nelson and Plosser data set, the stationary hypothesis for the unemployment rate, the velocity of the money as well as for the industrial production cannot be rejected. In the updated data set the only stationary series is the monthly hours worked that is well represented as an $AR(2)$ process. We then presented the results for a subset of models that do not allow for autoregressive parameters and we showed the high risk to overfit the integration order in this case. We provided a robustness analysis for the monthly inflation series and we showed the robustness of this procedure. Finally a Monte Carlo experiment confirmed the reliability of the selection procedure. The extension of this testing procedure to time varying parameters, seasonality models and to structural instability is far beyond the scope of this chapter and we leave it as a topic of further research.

Table 4.7: *Upper table:Nelson a Plosser data set. Bottom table: updated data set.*

Name	Time span	Curtailment
Real GNP	1909 - 1970	Rgnp
Nominal GNP	1909 - 1970	Ngnp
Real per capita GNP	1909 - 1970	Rpcgnp
Industrial product index	1860 - 1970	Iprod
Total employment	1890 - 1970	Emply
Total unemployment rate	1890 - 1970	Unempl
GNP deflator	1889 - 1970	Gnpdefl
Consumer price index	1860 - 1970	Pcons
Nominal wage	1900 - 1970	Nwage
Real wages	1900 - 1970	Rwage
Money stock M2	1889 - 1970	Money
Velocity of money	1869 - 1970	Veloc
Bond yield 30-year corporate	1900 - 1970	Interest
Stock prices	1871 - 1970	Pstock

Name	Time span	Curtailment
Monthly Cpi all items	1960(1) - 2008(12)	CPI
Monthly Inflation rate	1960(1) - 2008(12)	MInfl
Quarterly Inflation rate	1960(1) - 2008(12)	QInfl
Monthly Industrial product	1960(1) - 2009(10)	IP
Quarterly GNP Deflator	1947(1) - 2009(3)	GNPD
Quarterly GDP	1947(1) - 2009(4)	GDP
Monthly Conventional Mortgage Rate 30 years	1971(4) - 2009(10)	Mort
Monthly Unemployment rate	1960(1) - 2010(10)	Unempl
Monthly U.S./Euro Foreign Exchange Rate	1999(1) - 2009(11)	ExcRat
Weekly M2 Money Stock	1980(11/3) - 2010(25/1)	M2
Monthly Hours worked	1948(1) - 2009(11)	Awhman

4.8 Appendix A: Dataset

The Nelson and Plosser (1982) data set are taken from the database of the Gretl software freely available from the web side *http://gretl.sourceforge.net/*. The sample size range from 62 to 111 observations for each series. Except for the bold yield, the raw data are transformed into natural logarithms. The more recent data set are taken from Federal Reserve Bank of St.Louis freely available from the web side *http://research.stlouisfed.org/fred2/*. The raw data are transformed into natural logarithms.

Chapter 5

General Conclusions

Since the seminal paper of Kalman (1960) and Kalman and Bucy (1961), that introduced the Kalman filter, a lot of work has been done in state space methodology. The usefulness of linear and nonlinear state space model has been demonstrated in many applications in the last decades. In this dissertation we have showed new fields of applicability of this methodology. Linear and nonlinear state space models, classical and Bayesian estimation strategies as well as model testing in a Bayesian framework have been considered. All the examples showed both the practical usefulness and the economically relevant meaning of the state space approach, by extending the research agenda of state space models in different directions.

After a review of linear and nonlinear and non-Gaussian state space models with the related filtering and estimation strategies in chapter 1, in chapter 2 we studied the US inflation using a nonlinear state space model. A Bayesian estimation strategy has been provided, along with a Bayesian model testing procedure. This illustration showed the flexibility of the state space methodology to resolve such complex problems that cannot easily be handled using the ARIMA framework. In chapter 3 the state space methodology has been used to estimate the unknown factors for a large unbalanced panel. A suitable state space form with the corresponding analytical derivatives has been used to estimate in a fast and efficient way the unknown factors for a big unbalanced panel. The chapter further shows new results for global and regional convergence.

Finally, in chapter 4, we provide a new Bayesian test for linear and Gaussian state space models, showing how to include autoregressive components to the state space formulation. This new test has been applied to the Nelson and Plosser (1982) dataset and to an updated dataset. A Monte Carlo experiment confirmed the reliability of the selection methodology.

References

Akaike, H. (1974). "A new look at the statistical model identification". *IEEE Transactions on Automatic Control*, AC-19, 716 - 723.

Andrieu, C., Arnaud Doucet, A. and Tadic, V. (2005). "On-Line Parameter Estimation in General State-Space Models", in *Proceedings of the 44th IEEE Conference on Decision and Control*, Sevilla, Spain.

Ansley, C., and Kohn, R. (1986). "Prediction Mean Square Error for State Space Models with Estimated Paramaters", *Biometrika*, 73, 467 - 473.

Artis, M., Marcellino, M and Proietti, T. (2005). "Business Cycle in the New EU Member Countries and their Conformity with the Euro Area", *Journal of Business Cycle Measurement and Analysis*, Vol 2, No. 1.

Arulampalam, M.S., Maskell, S., Gordon, N. and Clapp, T. (2002). "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking", *IEEE Trans. Signal Processing*, 50, 174 - 188.

Backus, D.K. and Crucini, M.J. (2000). "Oil Prices and the Terms of Trade", *Journal of International Economics*, 50, 203 - 231.

Bai, J. and NG, S. (2008). "Large dimensional factor analysis". *Foundations and Trends in Econometrics*, 3, 89 - 163. Working Paper, Mimeo.

Backus, D.K., Kehoe, P. J. and Kydland, F.E. (1995). "International Business Cycles: Theory and Evidence" in Thomas F. Cooley, ed., *Frontiers of business cycle research* Princeton, NJ: Princeton University Press, 1995, 331 - 357.

Bataa E., Osborn, D.R., Sensier, M. and van Dijk D. (2009). "Changes in International Business Cycle Affiliations", *Centre for Growth and Business Cycle Research*, Discussion Paper No. 132.

Baxter, M. (1995). "International Trade and Business Cycles" in Gene M. Grossman and Kenneth Rogoff, eds., *Handbook of international economics*, volume 3. Amsterdam: North-Holland, 1801 - 1864.

Bos, C.S. and Shephard N. (2006). "Inference for Adaptive Time Series Models: Stochastic Volatility and Conditionally Guassian State Space Form", *Econometric Reviews*, 25 , 219 - 244.

Bos, C.S., Koopman S.J. and Ooms M. (2008). "Long Memory Modelling of Inflation with Stochastic Variance and Structural Breaks", Discussion paper TI 07-099/4, Tinbergen Institute.

Box, G. E. P. and Jenkins, G.M. (1979). *Time Series Analysis: Forecasting and Control*, Holden-Day.

Broto, C. and Ruiz, E. (2004). "Estimation Methods for Stochastic Volatility Models: A Survey", *Journal of Economic Survey*, Vol 18 , No. 5.

Burt, C. (1909). "Experimental tests of general intelligence". *British Journal of Psychology*, 4, 94 - 177.

Calvo, G. (1983). "Staggered prices in a utility-maximizing framework", *Journal of Monetary Economics*, 12, 383 - 398.

Caporale, M. G. and Gil-Alana, L. (2005). "Nelson and Plosser Revisited: Evidence from Fraction ARIMA Models", *American Journal of Applied Sciences*, 2(4), 860 - 872.

Cappé, O., Moulines, E. and Rydén, T. (2007). *Inference in Hidden Markow Models*, Springer.

Carter, C. K. and Kohn, R. (1994). "On Gibbs sampling for state space models", *Biometrika*, 81, 541 - 553.

Cecchetti, G., Hooper, P., Kasman, B.C., Shoenholtz, K. and Watson, M. (2007). "Understanding the Evolving Inflation Process", *U.S. Monetary Policy Forum 2007*.

Chen, R. and Liu, J.S. (2000). "Mixture Kalman filters", *Journal of Royal Statistical Society B*, 62, 493 - 508.

Chib, S. (1995) "Marginal Likelihood from the Gibbs Output", *Journal of the American Statistical Association*, 90, 1313 - 1321.

Chib, S. and Jeliazkov, I. (2001). "Marginal Likelihood from the Metropolis-Hastings Output", *Journal of the American Statistical Association*, 96, 270 - 281.

Cogley, T., Primicieri, G. and Sargent, T.J. (2008). "Inflation-Gap persistence in the U.S.", NBER Working Paper, No. 13749.

Culver, S.E. and Papell, D.H. (2008). "Is there a Unit Root in the Inflation Rate? Evidence from Sequential Break and Panel Data Models", *Journal of Applied Econometrics*, Vol. 12, No. 4, 435 - 444.

de Jong, P. (1991). "The diffuse Kalman filter", *Annals of Statistics*, 19, 1073 - 1083.

de Jong, P. and Shephard, N. (1996). "The simulation smoother", *Biometrika*, 2, 339 - 350.

Dempster, A.P., Laird N. and Rubin D.B. (1977). "Maximum likelihood from incomplete data via the EM algorithm", *Journal of Royal Statistical Society B*, 39, 1 - 38.

Del Negro, M. and Otrok, C. (2008). "Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles", Federal Reserve Bank of New York, Staff Report, No. 326.

Dickey, D.A. and Fuller, W.A. (1979). "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, 74, 427 - 431.

DiCiccio, T.J., Kass, R.E., Raftery, A. and Wasserman, L. (1997). "Computing Bayes factors by combining simulation and asymptotic approximations", *Journal of the American Statistical Association*, 92, 903 - 915.

Dornbusch, R. (1976). "Expectations and exchange rate dynamics", *Journal of Political Economy*, 84, 1161 - 1176.

Doornik, J.A. (2007). *Ox: An Object-Oriented Matrix Programming Language*, Timberlake Consultants Press, London.

Doucet, A. and Johansen, A.M. (2008). "A Tutorial on Particle Filter and Smoothing: Fifteen years later", *Technical report*, 197 - 208.

Doucet, A. and Tadic, V. B. (2003). "Parameter Estimation in General State-Space Models using Particle Methods", *Annals of the Institute of Statistical Mathematics*, Vol. 55, No. 2, 409 - 422.

Doucet, A., de Freitas, N. and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*, New York, Springer.

Doucet, A., Godsill, S.J. and Andrieu, C. (2000). "On sequential Monte Carlo sampling methods for Bayesian filtering", *Statistics and Computing*, 3, 197 - 208.

Doz, C., Giannone, D. and Reichlin, L. (2007). "A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models", CEPR Discussion Paper, No. 5724.

Durbin, J. and Koopman, S.J. (2001). *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford.

Durbin, J. and Koopman, S.J. (2002). "A simple and efficient simulation smoother for state space time series analysis", *Biometrika*, 89, 603 - 615.

Eichengreen, B. (2001). "The EMS Crisis in Retrospect", *CEPR Discussion Papers*, No. 2704.

Engle, R.F. and Watson, M.W. (1981). "A one-factor multivariate time series model of metropolitan wage rates", *Journal of the American Statistical Association*, 76, 774 - 781.

Fiorentini, G., Planas, C. and Rossi, A. (2008). "The Marginal Likelihood of Structural Time Series Models, with Application to the Euro Area and US Nairu", *The Rimini Centre for Economic Analysis*, Working Paper, No. 21-08.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). "The Generalized Dynamic-Factor Model: Identification and Estimation", *Review of Economics and Statistics*, 82, 540 - 545.

Frühwirth-Schnatter, S. (1994a). "Applied state space modelling of non-Gaussian time series using integration-based Kalman filtering ", *Statistics and Computing*, 259 - 269.

Frühwirth-Schnatter, S. (1994b). "Data augmentation and dynamic linear models ", *Journal of Time series Analysis*, 15, 183 - 202.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer Series in Statistics. Springer, New York.

Frühwirth-Schnatter, S. and Wagner, H. (2009). "Stochastic model specification search for Gaussian and partial non-Gaussian state space models", *Journal of Econometrics*(forthcoming).

Gardner, E.S. (1985). "Exponential smoothing: the state of the art", *Journal of Forecasting*, 4, 1 - 28.

Gardner, E.S. (2006). "Exponential smoothing: the state of the art. Part II", *International Journal of Forecasting*, 22, 637 - 666.

Gargan, E. A. (1997). "Last Laugh for the Philippines; Onetime Joke Economy Avoids Much of Asia's Turmoil", New York Times, December 11.

Geman, S. and Geman, D. (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6 (6), 721 - 741.

Gelman, A. and Smith, A. (1990). "Sampling based approaches to calculating marginal densities", *Journal of the American Statistical Association*, 85, 479 - 488.

Gentry, J.W., Jun, S., Chun, S., Kang, H. and Ko, G. (2006). "Korea: Two Countries, Sharp Contrasts, but a Common Heritage", in Pecotich, A., Shultz, C.J. *Handbook of Markets and Economies*, M.E. Sharpe.

George, E. I. and McCulloch, R. (1993). "Variable selection via Gibbs sampling", *Journal of the American Statistical Association*, 88, 881 - 889.

George, E. I. and McCulloch, R. (1997). "Approaches for Bayesian variable selection", *Statistica Sinica*, 7, 339 - 373.

Geweke, J.F. (2005). *Contemporary Bayesian econometrics and statistics*, Wiley, New York.

Geweke, J.F. (1992). "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments", in J.O. Berger, J.M. Bernardo, A.P. Dawid, and A.F.M. Smith (eds.), *Bayesian Statistics*, 4, 169 - 194. Oxford: Oxford University Press, 1992.

Godsill, S. J. and Clapp, T. C. (2001). "Improvement strategies for Monte Carlo particle filters". In A Doucet, J. F. G. De Freitas, and N. J. Gordon, editors, Sequential Monte Carlo Methods in Practice. New York, Springer-Verlag.

Goodfriend, M. and King, R. G. (2005). "The incredible Volcker disinflation", *Journal of Monetary Economics*, 52, 981 - 1015.

Gordon, N.J., Salmond, D.J. and Smith A.F.M. (1993). "Novel approach to nonlinear/ non-Gaussian Bayesian state estimation, *IEE Proc.-F*, 140, No. 2, 107 - 113.

Granger, C.W.J. and Newbold, P. (1986). *Forecasting in Business and Economics*, Academic Press, Orlando, Second Edition.

Green, P. J. (1995). "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination", *Biometrika*, 82, 711 - 732.

Gregory, A. W., Head, A. C. and Raynauld, J. (1997). "Measuring World Business Cycles", *International Economic Review*, 38(3), 677 - 702.

Grewal, M. S. and Andrews, A. P. (2001). *Kalman Filtering: Theory and Practice Using MATLAB*, Second Edition, John Wiley and Sons.

Grossman, G. and Helpman, E. (1991). *Innovation and growth in the global economy*, Cambridge, MIT Press.

Hammersley, J.M. and Morton, K.W. (1954). "Poor mans Monte Carlo", *Journal of the Royal Statistical Society, Series B*, 16, No. 1, 23 - 38.

Harvey, A. C. and Proietti, T. (2005). *Readings in Unobserved Components Models*, Advanced Texts in Econometrics. Oxford University Press, Oxford, UK.

Harvey, A. C., Ruiz, E. and Shephard, N. G. (1994). "Multivariate stochastic variance models", *Review of Economic Studies*, 61, 247 - 264.

Harvey, A.C. (1989). *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge University Press, Cambridge, UK.

Harvey, A. C. and Phillips, G. D. A. (1979). "The estimation of regression models with autoregressive-moving average disturbances", *Biometrika*, 66, 49 - 58.

Hasting, W. (1970). "Monte Carlo sampling methods using Markov chains and their application", *Biometrika*, 57, 97 - 109.

Head, A.C. (1995). "Country Size, Aggregate Fluctuations, and International Risk Sharing", *Canadian Journal of Economics*, 28, 1096 - 1119.

Heathcote, J. and Perri, F. (2004). "Financial Globalization and Real Regionalization", *Journal of Economic Theory*, 119 (1), 2007 - 2043.

Helbling, T. and Bayoumi, T. (2003). "Are They All in the Same Boat? The 2000-2001 Growth Slowdown and the G-7 Business Cycle Linkages", *IMF Working Paper*, No. 03/46.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994). "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business & Economic Statistics*, 12, 371 - 389.

Johansen, A.M. and Doucet, A. (2008). "A note on auxiliary particle filters", *Statistics and Probability Letters*, 78, 1498 - 1504.

Johansen A.M., Doucet, A. and Davy, M. (2006). "Maximum Likelihood Parameter Estimation for Latent Variable Models Using Sequential Monte Carlo", in 2006 IEEE International Conference on Acoustics, Speech and Signal Processing, Toulouse, France.

Julier, S.J. and Uhlmann, J.K. (1997). "A new extension of the Kalman filter to nonlinear systems", *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, 3.

Jungbacker, B. and Koopman, S.J. (2008). "Likelihood-based analysis for dynamic factor models", *VU University of Amsterdam*, Working Paper.

Jungbacker, B., Koopman, S.J., and van der Wel, M.(2009). "Dynamic Factor Analysis in the Presence of Missing Data", *Tinbergen Institute*, Discussion Paper, No. 010-4.

Kalman, R.E. (1960). "A new approach to linear filtering and prediction problems", *J. Basic Engineering, Transactions ASMA, Series D*, 82, 35 - 45.

Kalman, R.E. and Bucy, R.S. (1960). "New results in linear filtering and prediction theory ", *Transactions of the ASME. Series D, Journal of Basic Engineering*, 83, 95 - 107.

Kapetanios, G. and Marcellino, M. (2006). "A parametric estimation method for dynamic factor models of large dimensions", *Centre for Economic Policy Research*, Discussion Paper, No. 5620.

Kim, S., Shepard, N. and Chib, S. (1998)."Stochastic Volatility: Likelihood Inference and Comparison with ARCH model", *Review of Economic Studies*, 65, 361 - 393.

Kong, A., Liu, J.S. and Wong, W.H. (1994). "Sequential imputation and Bayesian missing data problems", *Journal of the American Statistical Association*, 89, 1032 - 1044

Koop, G. (1992). "Objective' Bayesian Unit root tests", *Journal of Applied Econometrics*, 7, 65 - 82.

Koop, G. (1994). "Recent progress in applied Bayesian Econometrics", *Journal of Economic Survey*, 8, 1 - 34.

Koop, G. (2003). *Bayesian Econometrics*, John Wiley and Sons Ltd.

Koop, G. and Van Dijk, H. (2000). "Testing for integration using evolving trend and seasonals models: A Bayesian approach", *Journal of Econometrics*, 97, 261 - 291.

Koopman, S.J. and Shephard, N. (1992). "Exact Score for Time Series Models in State Space Form", *Biometrika*, Vol. 79, No 4.

Kose, M. A. (2002). "Explaining Business Cycles in Small Open Economies: How Much Do World Prices Matter?", *Journal of International Economics*, 56(2), 299 - 327.

Kose, M. A., Prasad, E. S. and Terrones, M.E. (2003). "How does Globalization Affect the Synchronization of Business Cycles?", *The American Economic Review*, Vol. 93, No. 2, 57 - 62.

Kose, M.A, Otrok, C. and Whiteman, C.H. (2004). "International Business Cycle: World, Region, and Country-Specific Factors", *The American Economic Review*, Vol. 93, No. 4, 1216 - 1239.

Kose, M.A, Otrok, C. and Prasad, E. (2008). "Global Business Cycle: Convergence or Decoupling?", *IZA* Discussion Paper, No. 3442.

Lewis, S.M. and Raftery, A.E. (1997). "Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator", *Journal of the American Statistical Association*, 92, 648-655.

Ley, E. and Steel, M. F. J. (2009). "On the effect of prior assumptions in Bayesian model averaging with applications to growth regression", *Journal of Applied Econometrics*, 24, 651 - 674.

Leybourne, S.J., McCabe, B.P. and Tremayne, A.R. (1996). "Can Economic Time Series Be Differenced to Stationatity?", *Journal of Business and Economic Statistics*, Vol. 14, No. 4, 435 - 446.

Leybourne, S.J. and McCabe, B.P. (1999). "Modified Stationarity Tests with Data-Dependent Model-Selection Rules", *Journal of Business and Economic Statistics*, Vol 17, No. 2, 264 - 270.

Liu, J.S. (2001). "Monte Carlo Strategies in Scientific Computing", Springer.

Ljung, G.M. and Box, G.E.P. (1978). "On a measure of lack of fit in time series models", *Biometrika*, 66, 67 - 72.

Magnus, J.R. and Neudecker, H. (2007). "Matrix Differential Calculus with Applications in Statistics and Econometrics", Third Edition, John Wiley and Sons.

McGee, L.A. and Schmidt, S.F. (1985). "Discovery of the Kalman Filter as a Practical Tool for Aerospace and Industry", NASA Technical Memorandum 86847, Moffett Field, California.

Mendoza, E.G. (1995). "The Terms of Trade, the Real Exchange Rate, and Economic Fluctuations", *International Economic*, 36(1), 101 - 137.

Meng, X.L. and Wong, W.H. (1996). "Simulating ratios of normalizing constants via a simple identity: A theoretical exploration", *Statistica Sinica*, 6, 831 - 860.

Metropolis, N., Rosenbluth, A., Resenbluth, M., Teller, A. and Teller, E. (1953). "Equations of state calcultations by fast computing machines", *Journal of Chemical Physichs*, 21, 1087 - 1092.

Mocan H.N. (1994). "Is there a unit root in U.S. real GNP ? A re-assessment", *Economics Letters*, 45, 23 - 31.

Molenaar, P.C.M. (1985). "A dynamic factor model for the analysis of multivariate time series", *Psycometrika*, 50, 181 - 202.

Moore, G.H. and Zarnowitz, V. (1986). "The Development and Role of the National Bureau of Economic Research's Business Cycle Chronologies", in R.J. Gordon ed., *The American Business Cycle: Continuity and Change*, University of Chicago Press, 735 - 779.

Muth, J.F. (1960). "Optimal properties of exponentially weighted forecasts", *Journal of the American Statistical Association*, 55, 299 - 306.

Nelson, D. B. (1988). "The time series behavior of stock market volatility and returns", Unpublished PhD Dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Nelson, C. R. and Plosser, C.I. (1982). "Trend and Random Walk in Macroeconomic Time Series", *Journal of Monetary Economic*, 10, 139 - 162.

Nocedal, J. and Wright, S.J. (1999). *Numerical Optimization*, New York, Springer-Verlag.

Nordhaus, D. W. (2002). "The Mildest Recession: Output, Profits, and Stock Prices as the U.S. Emerges from the 2001 Recession", *NBER Working Paper*, No. 8938.

Orderud, F. (2005). "Comparison of Kalman Filter Estimation Approaches for State Space Models with Nonlinear Measurements" , in proceedings of *Scandinavian Conference on Simulation and Modeling*, October 13-14, Trondheim, Norway.

Palma, W. (2007). *Long-Memory Time Series*, Wiley-Interscience, a John Wiley and Sons, Inc.

Pecotich A. and Shultz, C.J. (2005). *Handbook of Markets and Economies: East Asia, Australia, New Zealand*, Armonk: New York. M.E. Sharpe, Inc.

Phillips, P.C.B. (1990). "To criticize the critics: An objective Bayesian analysis of stochastic trends", *Journal of Applied Econometrics*, Vol. 6(4), 333 - 364.

Pitt, M.K. and Shephard, N. (1999a). "Filtering via simulation: Auxiliary particle filters", *Journal of the American Statistical Association*, 94(446), 590 - 599.

Pivetta, F. and Reis, R. (2007). "The persistence of inflation in the United States", *Journal of Economic Dynamics and Control*, 31, 1326 - 1358.

Proietti T. (2008). "Estimation of Common Factors under Cross-Sectional and Temporal Aggregation Constraints: Nowcasting Monthly GDP and its Main Components", *University Library of Munich*, Germany, MPRA Paper No. 6860.

PWT 6.3 (August 2009) by Alan Heston, Robert Summers and Bettina Aten, *Penn World Table Version 6.3*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania.

Robert, P.C and Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer.

Romer, P. (1990). "Endogenous technical change", *Journal of Political Economy*, 98, 71 - 102.

Rosenbluth, M.N. and Rosenbluth A.W. (1956). "Monte Carlo calculation of the average extension of molecular chains", *Journal of Chemical Physics*, Vol. 23, No. 2, 356 - 359.

Ruiz, E. (1994). "Quasi-maximum likelihood estimation of stochastic volatility models", *Journal of Econometrics*, 63, 289 - 306.

Schmidt, S. F. (1966). "Applications of state-space methods to navigation problems" in *Advances in Control Systems*, (C. T. Leondes, Ed.), Vol. 3, Academic, New York.

Schweppe, F. (1965). "Evaluation of likelihood functions for Gaussian signals", *IEEE Transactions on Information Theory 11*, 61 - 70.

Schotman, P. C. and van Dijk, H.K. (1991). "On Bayesian Routes to Unit Roots", *Journal of Applied Econometrics*, Vol. 6(4), 387 - 401.

Shepard, N. (2006). *Stochastic Volatility. Selected readings*, Advanced Text in Econometrics, Oxford University Press.

Shephard, N. (1996). "Statistical aspects of ARCH and stochastic volatility", In O.E. Barndorff-Nielsen, D. R. Cox and D. V.Hinkley (eds.), *Statistical Models in Econo- metrisc, Finance and Other Fields*, 167. London: Chapman and Hall.

Shin, K. and Wang, Y. (2003). "Trade integration and Busines Cycle Synchronization in East Asia", *Asian Economic Papers*, MIT Press, Vol. 2, No. 3, 1 - 20.

Shively, T. S. and Kohn, R. (1997). "A Bayesian approach to model selection in stochastic coefficient regression models and structural time series models", *Journal of Econometrics*, 76, 39 - 52.

Shumway, R.H. and Stoffer, D.S. (1982). "An approach to time series smoothing and forecasting using the EM algorithm", *Journal Time Series Analysis*, 3, 253 - 264.

Sims, C.A. (1988). "Bayesian skepticism on unit root econometrics", *Journal of Economic Dynamic and Control*, 12, 463 - 474.

Sims, C.A. and Uhlig, H. (1991). "Understanding unit rooters: A helicopter tour", *Econometrica*, 59, 1591 - 1599.

Smith, A.F.M. and West, M. (1983). "Monitoring renal transplants: an application of the multiprocess Kalman filter", *Biometrics*, 39, 867 - 878.

Solow, R. (1956). "A Contribution to the theory of economic growth", *Quarterly Journal of Economics*, 70, 312 - 320.

Stiglitz, J. (1997). "Reflections on the natural rate hypothesis", *Journal of Economic Perspectives*, 11, 1, 3 - 10.

Stock, J.H. and Watson, M. W. (1989). "New Indexes of Coincident and Leading Economic Indicators", *NBER Macroeconomics Annual*, 351 - 394.

Stock, J.H. and Watson, M. W. (2002). "Macroeconomic Forecasting Using Diffusion Index", *Journal of Business and Economic Statistics*, Vol. 20, No. 2, 147 - 163.

Stock, J.H and Watson, M. W. (2005). "Understanding Changes in International Business Cycle Dynamics". *Journal of European Economic Association*, Vol. 3(5), 968 - 1006.

Stock, J.H. and Watson, M.W. (2006). "Forecasting with many predictors", in Elliott, G., Granger, C.W.J. and Timmermann, A. (eds.) *Handbook of Economic Forecasting*, Vol. 1, Elsevier.

Stock, J.H. and Watson, M.W. (2007). "Why Has U.S. Inflation Become Harder to Forecast?", *Journal of Money, Credit and Banking*, 39, 3 - 34.

Swan, T. (1956). "Economic growth and capital accumulation", *Economic Record*, 32, 344 - 361.

Taylor, S. (1994). "Modeling stochastic volatility: a review and comparative study", *Mathematical Finance*, 4, 183 - 204.

Watson, M.W. and Engle, R.F. (1983). "Alternative Algorithm for the estimation of dynamic factor, MIMIC and varying coefficient regression", *Journal of Econometrics*, 23, 385 - 400.

Weeks, J. (2000). "Latin America and the 'high performing Asian economies': growth and debt", *Journal of International Development*, 625 - 654.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, New York, Springer-Verlag.