



SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY



CREATES

Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2011-16

**Characterizing economic trends by Bayesian stochastic
model specification search**

Stefano Grassi and Tommaso Proietti

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Characterizing economic trends by Bayesian stochastic model specification search *

Stefano Grassi[†]

Tommaso Proietti[‡]

CREATES and Aarhus University

University of Rome “Tor Vergata”

May 5, 2011

Abstract

We extend a recently proposed Bayesian model selection technique, known as stochastic model specification search, for characterising the nature of the trend in macroeconomic time series. In particular, we focus on autoregressive models with possibly time-varying intercept and slope and decide on whether their parameters are fixed or evolutive.

Stochastic model specification is carried out to discriminate two alternative hypotheses concerning the generation of trends: the trend-stationary hypothesis, on the one hand, for which the trend is a deterministic function of time and the short run dynamics are represented by a stationary autoregressive process; the difference-stationary hypothesis, on the other, according to which the trend results from the cumulation of the effects of random disturbances.

We illustrate the methodology for a set of U.S. macroeconomic time series, which includes the traditional Nelson and Plosser dataset. The broad conclusion is that most series are better represented by autoregressive models with time-invariant intercept and slope and coefficients that are close to boundary of the stationarity region. The posterior distribution of the autoregressive parameters, estimated by a suitable Gibbs sampling scheme, provides useful insight on quasi-integrated nature of the specifications selected.

Keywords: Bayesian model selection; stationarity; unit roots; stochastic trends; variable selection.

JEL codes: C22, C52

*The first author acknowledges the research support of CREATES (funded by the Danish National Research Foundation).

[†]School of Economics and Management, Aarhus University, Bartholins Alle 10, 1322 University Park, DK-8000 Aarhus C, Denmark; phone: +45 89 421 547. E-mail address: sgrassi@creates.au.dk

[‡]S.E.F. e ME.Q., University of Rome “Tor Vergata”, Via Columbia 2, 000133 Rome, Italy. E-mail address: tommaso.proietti@uniroma2.it

1 Introduction

Characterizing the nature of the trends observed in economic time series is a widely debated topic in time series analysis. An issue that has attracted a lot of attention is whether the trend is best captured by deterministic or integrated stochastic processes.

The historically oldest approach is to view the trend as a deterministic, possibly unknown, function of time, and the deviations from trend as a stationary process (thus, the series is said to be trend-stationary). According to this interpretation, the trend is an entirely exogenous component, that can be estimated e.g. by global or local polynomial approximations.

An alternative view is that trends arise endogenously as a result of the persistent effects of economic shocks, that are cumulated in the level of the series. This behaviour is the characteristic property of the class of integrated, or unit root, processes. As the series can be made stationary after suitable differencing, it is also said to be difference-stationary. The distinction between what is permanent and what is transitory in economic dynamics has important implications for interpretation and policy.

The econometric literature has envisaged formal statistical tests for discriminating the two trend generation hypotheses. Unit root tests, see Dickey-Fuller (1979) and Phillips and Perron (1988), test the null of integration versus a stationary alternative; on the contrary, stationary tests, see Nyblom and Makelainen (1983) and Kwiatkowski et al. (1992), test trend stationarity against the alternative of integration. The implications for the interpretation of macroeconomic dynamics were considered in a seminal paper by Nelson and Plosser (1982), in which they applied the Dickey-Fuller test on a representative set of annual U.S. macroeconomic time series, and were unable to reject the null of integration for most of the series.

A rich literature has discussed the limitations of the testing approach, see among others DeJong et al. (1992), Schwert (1989) and Caner and Kilian (2001), and has proposed refinements and enhancements. Important references are Perron (1989), Elliott, Rothenberg and Stock (1996), Ng and Perron (2001) for unit roots tests, and Leybourne and McCabe (1994) for stationarity tests; see also Harvey (2001) for a review.

The Bayesian approach to unit root testing has been considered by DeJong and Whiteman (1991), Koop (1992), Sims (1988), Sims and Uhlig (1991), Phillips (1991), Schotman and van Dijk (1991), Phillips and Ploberger (1994), among others; the literature has focused on the selection of noninformative priors for the autoregressive coefficients and on assessing the sensitivity of model selection on the prior choice.

The problem of discriminating fixed trends from stochastically evolving ones has been addressed by Frühwirth-Schnatter (1995) and Koop and van Dijk (2000). The research question that we posit

in this paper is similar to that of the two aforementioned articles, in that our ultimate aim is establishing which trend model appears to provide the most plausible explanation for the behaviour of economic time series. However, our approach is different as we capitalize on the recent developments in Bayesian model selection. In particular, we apply the stochastic model specification search recently proposed by Frühwirth-Schnatter and Wagner (2010, FS-W henceforth). The different trend models are nested inside a more general hierarchical state space model and are obtained by imposing exclusion restrictions, so that discriminating the trend hypothesis amounts to performing variable selection within the regression framework considered by George and McCulloch (1993, 1997). We will argue that this approach can shed further light on the issue of characterising trends in macroeconomic time series.

The plan of the paper is the following. The next section introduces the approach in the simple case when we are interested in discriminating a fixed level versus a random walk level. Section 3 brings into the analysis a possibly stochastic drift. Model selection and estimation by Monte Carlo Markov Chain is discussed in section 4. Illustrations are provided in section 5 with respect to the traditional Nelson and Plosser (1982) dataset and other key macroeconomic time series. In section 6 we draw some conclusions.

2 Discriminating Level Stationarity and random walk trends.

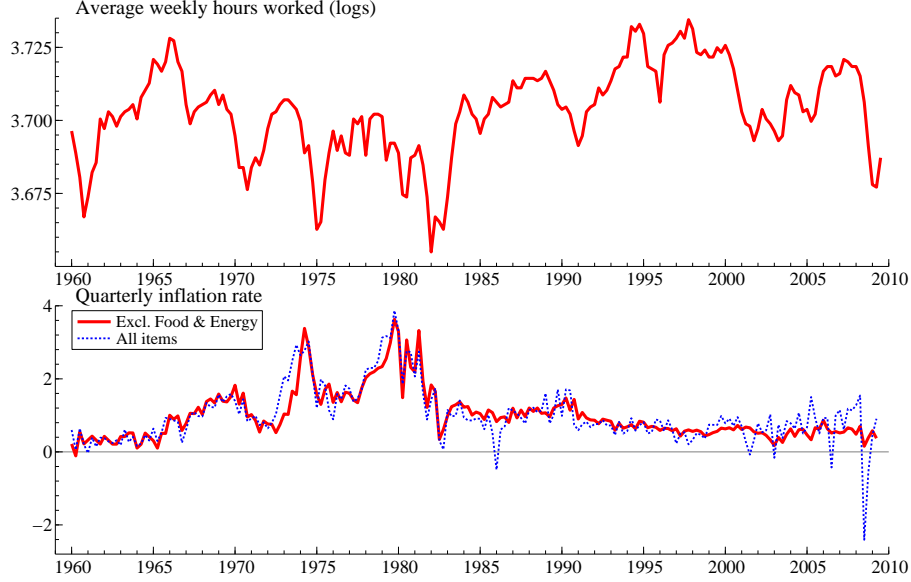
Figure 1 displays the quarterly series of U.S. average weekly hours worked (AWHMAN) for the manufacturing sector and the quarterly CPI and core (ex. food and energy) inflation rate for the period 1960:1-2009.4 (Source: U.S. Census Bureau). These series have been extensively investigated in macroeconomic applications. For instance, as far as AWHMAN is concerned, the order of integration of the series is a crucial issue, as the response of the labour market to technology shocks crucially depends on the stationarity of this series. Opposite conclusions are reached whether one uses differences or levels in a structural vector autoregressive model: in the former case (see Galí, 1999) technology shocks induce a short run reduction in hours worked; in the second, hours worked increase, see Christiano et al. (2003).

In this section we present an approach based on Bayesian model selection to investigate the issue as to whether the long run evolution of hours and inflation is better characterized by a fixed level or a slowly evolving component driven by permanent shocks.

Let us consider, as a starting point, the following AR(p) model with time-varying intercept:

$$\begin{aligned}\phi(L)y_t &= \mu_t + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \sigma_\epsilon^2), & t = 1, \dots, T, \\ \mu_t &= \mu_{t-1} + \eta_t, & \eta_t &\sim \text{NID}(0, \sigma_\eta^2),\end{aligned}\tag{1}$$

Figure 1: U.S. Average Weekly Hours Worked (Manufacturing) and Quarterly Inflation rate (Core and all items). Source: US Census Bureau, 1960:q1-2009.q4.



such that $\phi(L)$ is a stationary AR polynomial, and ϵ_t and η_t are mutually uncorrelated at all leads and lags.

The locally best invariant test of the null $H_0 : \sigma_\eta^2 = 0$ versus the alternative $H_1 : \sigma_\eta^2 > 0$ has been studied by Leybourne and McCabe (1994). Model (1) nests two trend generation hypotheses of interest:

- a) y_t is a level-stationary process. This occurs if $\sigma_\eta^2 = 0$, and $\phi(L)$ is a stationary polynomial, i.e. $\phi(z) = 0 \iff |z| > 1$.
- b) y_t is a difference stationary process. There are two ways by which difference stationary processes can arise as particular cases of (1).

1. y_t is a time-invariant $AR(p)$ process characterized by a unit root. This occurs when $\sigma_\eta^2 = 0$ and $\phi(z) = 0 \iff z = 1$. As $\phi(L) = \Delta\phi(L)^*$, where $\phi(L)^* = 1 - \phi_1^*L - \dots - \phi_{p-1}^*L^{p-1}$, $\phi_j^* = -\sum_{i=j+1}^p \phi_i$, Δy_t is a stationary $AR(p-1)$ process. Thereby, we say that y_t is a purely autoregressive difference stationary process. It admits a trend-cycle decomposition (see Beveridge and Nelson, 1981) with components driven by perfectly correlated disturbances: $y_t = m_t + c_t$, $m_t = m_{t-1} + \epsilon_t^*$ and $\phi^*(L)c_t = \theta^*(L)\epsilon_t^*$, where $\epsilon_t^* = [\phi^*(1)]^{-1}\epsilon_t$ and $\theta^*(L)$ is the order $p-2$ moving average polynomial $\Delta^{-1}[\phi^*(1) - \phi^*(L)]$.

Under restrictive conditions on the ϕ_j^* coefficients, discussed in Proietti and Harvey (2000), the model can be interpreted as the reduced form of an orthogonal trend-cycle decomposition of y_t into a random walk trend and an $\text{AR}(p-1)$ cycle; the Beveridge-Nelson components would provide the real time estimates of the trend and the cycle in this case. For instance, when $p = 2$, we need $0 < \phi_2 < 1$, along with the unit root condition $\phi_2 = 1 - \phi_1$.

2. y_t is a difference stationary process with multiple sources of disturbances, which occurs when $\sigma_\eta^2 > 0$, and $\phi(L)$ is a stationary polynomial, i.e. $\phi(z) = 0 \iff |z| > 1$. As a result, Δy_t has a moving average (MA) feature, namely $\Delta y_t \sim \text{ARIMA}(p, 1, 1)$, with a negative MA coefficient, as it can be readily seen from the fact that the autocovariance of the process $\phi(L)\Delta y_t = \eta_t + \Delta \epsilon_t$ at lag 1 is $-\sigma_\epsilon^2$. Again, y_t admits the Beveridge-Nelson decomposition into a random walk trend and a stationary $\text{ARMA}(p, p-1)$ cycle, and, under restrictive conditions, an orthogonal trend-cycle decomposition. The conditions relate to the possibility of interpreting the process $\eta_t/[\phi(L)\Delta]$ as the reduced form of an orthogonal decomposition of into a random walk trend and an $\text{AR}(p)$ cycle; see Proietti and Harvey (2000). The specification with $\phi_j = 0, j = 1, \dots, p$, is known as the local level model (Harvey, 1989).

Notice that we do not restrict the AR roots to be in the stationarity region, otherwise the first category of difference stationary models could not be represented among the competing models.

The stochastic model specification search methodology proposed by FS-W is based on a reparameterisation of (1), known as the non-centred representation, with respect to location and scale (see also Gelfand et al., 1995, Frühwirth-Schnatter, 2004, Strickland et al. 2007), which is obtained by writing

$$\begin{aligned} \mu_t &= \mu_0 + \sigma_\eta \tilde{\mu}_t, & t = 1, \dots, T, \\ \tilde{\mu}_t &= \tilde{\mu}_{t-1} + \tilde{\eta}_t, \quad \tilde{\eta}_t \sim \text{NID}(0, 1), \quad \tilde{\mu}_0 = 0, \end{aligned} \tag{2}$$

where μ_0 is the starting value of the random walk and $\tilde{\mu}_t \sim \text{N}(0, t)$.

The non-centred representation is useful not only for the efficiency of Bayesian estimation by Monte Carlo Markov Chain (MCMC) methods (in particular, when σ_η^2 is small in comparison to σ_ϵ^2), but also since it paves the way to performing model selection in a regression framework via the stochastic search variable selection (SSVS) approach proposed by George and McCulloch (1993, 1997).

FS-W's key idea is that the non-centred representation is not identified since model (2) with $(-\sigma_\eta)(-\tilde{\mu}_t)$ is observationally equivalent. As a consequence, the likelihood function is symmetric around zero along the σ_η dimension and bimodal, if the true σ_η is larger than zero. This fact can

be exploited to quantify how far the posterior of σ_η is removed from zero. Thus, letting B denote a Bernoulli random variable with $E(B) = 0.5$, independent of y_t , writing

$$\sigma_\eta \tilde{\mu}_t = \beta_\mu \mu_t^*, \quad \beta_\mu = (-1)^B \sigma_\eta, \quad \mu_t^* = (-1)^B \tilde{\mu}_t,$$

and replacing into (2) and subsequently into (1), yields:

$$y_t = \mu_0 + \beta_\mu \mu_t^* + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t.$$

By this clever expedient a standard deviation is transformed into a regression coefficient and SSVS can be applied. Hence the selection of the trend generating process is reconducted to the inclusion of a particular regressor. A seemingly ancillary issue is the presence and the selection of the AR component. This is only apparently a secondary issue, since misspecification of the short run dynamics, implied by lagged values of y_t , has important implications on the fundamental issue concerning the nature of the trend component trend. We will assume throughout that the AR polynomial has maximum order equal to 2 (we judge a second order model sufficiently general for our purposes).

Using the non-centred representation, for $p = 2$, there are three potential explanatory variables for y_t ; if we assume that every subset of the 3 explanatory variables is admissible, there are $K = 2^3$ possible models in competition. We now introduce the binary indicator variable γ_1 , taking value 1 if the random effect μ_t^* is present in the model and 0 if it is excluded, along with a pair of binary indicators for the two AR effects, δ_1, δ_2 , each taking values (0,1) according to as to whether the term $\phi_i y_{t-i}$, $i = 1, 2$, is included in the model. Hence, the 8 models in competition are nested within the following representation:

$$\begin{aligned} y_t &= \mu_0 + \gamma_1 \beta_\mu \mu_t^* + \delta_1 \phi_1 y_{t-1} + \delta_2 \phi_2 y_{t-2} + \epsilon_t, & \epsilon_t &\sim \text{NID}(0, \sigma_\epsilon^2), \\ \mu_t^* &= \mu_{t-1}^* + \eta_t^*, \quad \eta_t^* \sim \text{NID}(0, 1), \quad \mu_0^* = 0 \end{aligned} \quad (3)$$

Collecting the binary indicators in the vector $\Upsilon = (\gamma_1, \delta_1, \delta_2)$, the 8 possible models are listed below:

Label	Υ	Equation
M_1	(0,0,0)	$y_t = \mu_0 + \epsilon_t$
M_2	(0,0,1)	$y_t = \mu_0 + \phi_2 y_{t-2} + \epsilon_t$
M_3	(0,1,0)	$y_t = \mu_0 + \phi_1 y_{t-1} + \epsilon_t$
M_4	(0,1,1)	$y_t = \mu_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$
M_5	(1,0,0)	$y_t = \mu_0 + \beta_\mu \mu_t^* + \epsilon_t$
M_6	(1,0,1)	$y_t = \mu_0 + \beta_\mu \mu_t^* + \phi_2 y_{t-2} + \epsilon_t$
M_7	(1,1,0)	$y_t = \mu_0 + \beta_\mu \mu_t^* + \phi_1 y_{t-1} + \epsilon_t$
M_8	(1,1,1)	$y_t = \mu_0 + \beta_\mu \mu_t^* + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$

Models M_1 - M_4 are level stationary if $\phi(z) = 0 \iff |z| > 1$. If $\phi(z) = 1$ they are purely AR difference stationary models. Models M_5 - M_8 with $\phi(L) = 0 \iff |z| > 1$ are difference stationary with an MA(1) feature.

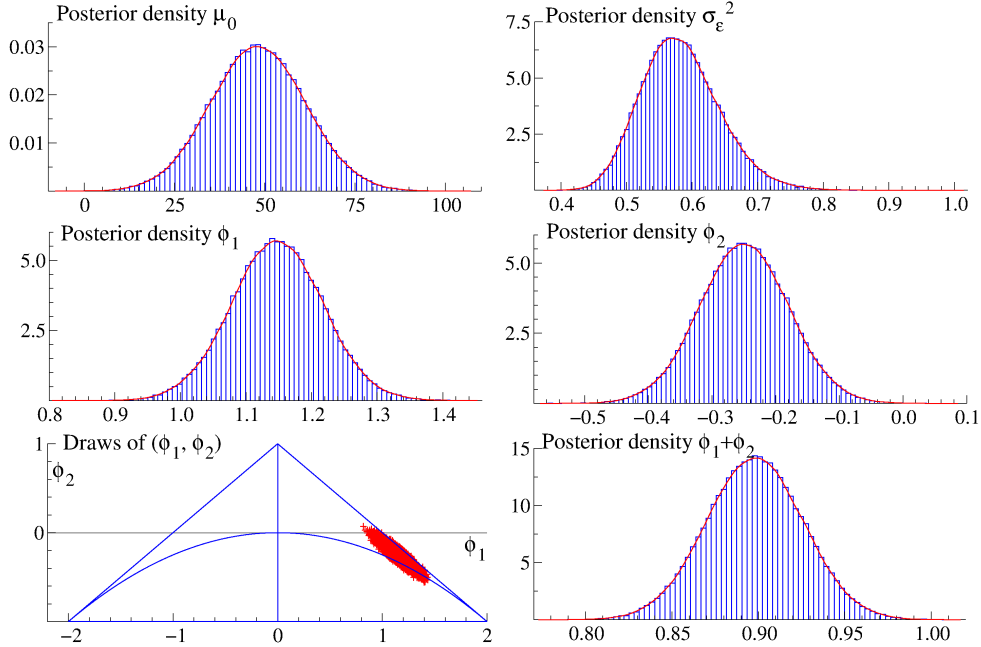
We will assume that the models $M_k, k = 1, \dots, K$, are equally likely a priori, that is $\pi(M_k) \propto 1$, or equivalently $\pi(\Upsilon) = 2^{-3}$, where $\pi(\cdot)$ denotes the density or the probability function of the argument. A distinctive trait of FS-W stochastic specification search is the adoption of Gaussian prior, centred at zero, for the parameter β_μ ; for instance, in model M_5 , $\pi(\mu_0, \beta_\mu, \sigma_\epsilon) = \pi(\mu_0)\pi(\beta_\mu, \sigma_\epsilon^2)$ and the prior for the joint distribution of $(\beta_\mu, \sigma_\epsilon^2)$ is normal-inverse gamma, that is $\pi(\sigma_\epsilon^2) \sim \text{IG}(c_0, C_0)$, see section 4, and $\pi(\beta_\mu | \sigma_\epsilon^2) \sim \text{N}(0, \kappa_\mu \sigma_\epsilon^2)$.

Not only this allows conjugate analysis, but FS-W show that inference will benefit substantially from the use of a normal prior for $\beta_\mu = \pm\sigma_\eta$, en lieu of the usual inverse gamma prior for the variance parameter σ_η^2 . In fact, a major problem that arises when the IG prior is used is the high sensitivity of the posterior distribution of σ_η^2 to the hyperparameters of the IG distribution, when the true σ_η^2 is close to zero; as a result the MCMC draws will mix very slowly or even lack convergence. On the contrary, the posterior distribution of β_μ is not too sensitive to the choice of κ and Monte Carlo inference is much more efficient.

Notice that $\beta_\mu | \sigma_\eta, \gamma_1 = 1$, is a random variable which takes the values $-\sigma_\eta$ and σ_η with probabilities both equal to 1/2 so that a Gaussian prior centred at zero is reasonable; furthermore, this choice amounts to specify a hierarchical mixture prior to the parameter β_μ , of the form $\pi(\beta_\mu) = (1 - \gamma_1)I_0 + \gamma_1\text{N}(0, \kappa\sigma_\epsilon^2)$ where I_0 is a degenerate density with point mass at zero, see Smith and Kohn (1996). As pointed out in George and McCulloch (1997), this prior entails that a stochastic trend will be included if β_μ can be distinguished from zero irrespective of its absolute size. An alternative, not explored yet for SSVS, is to $\pi(\beta_\mu) = (1 - \gamma_1)\text{N}(0, \kappa_0\sigma_\epsilon^2) + \gamma_1\text{N}(0, \kappa_1\sigma_\epsilon^2)$ such that κ_0 is small in comparison to κ_1 , in which case selection is based on practical significance of the stochastic variation in the level.

As far as model selection is concerned, given the limited number of specifications, one possibility is to compute the posterior model probabilities and select that specification which has the largest. However, this entails the evaluation of the marginal likelihood for each model. This evaluation is computationally intensive and the accuracy may be poor (see the discussion in FS-W and the references therein). Rather than computing the posterior probabilities of all the possible models, it is computationally more attractive to simulate samples from their posterior distribution by MCMC methods. In particular, exploiting the conditional independence structure of the model, and given the availability of the full conditional posterior distribution of Υ in closed form, the multinomial vector Υ is sampled along with the model parameters by using a Gibbs sampling scheme and a

Figure 2: U.S. Average Weekly Hours Worked (Manufacturing) 1960:q1-2009.q4.

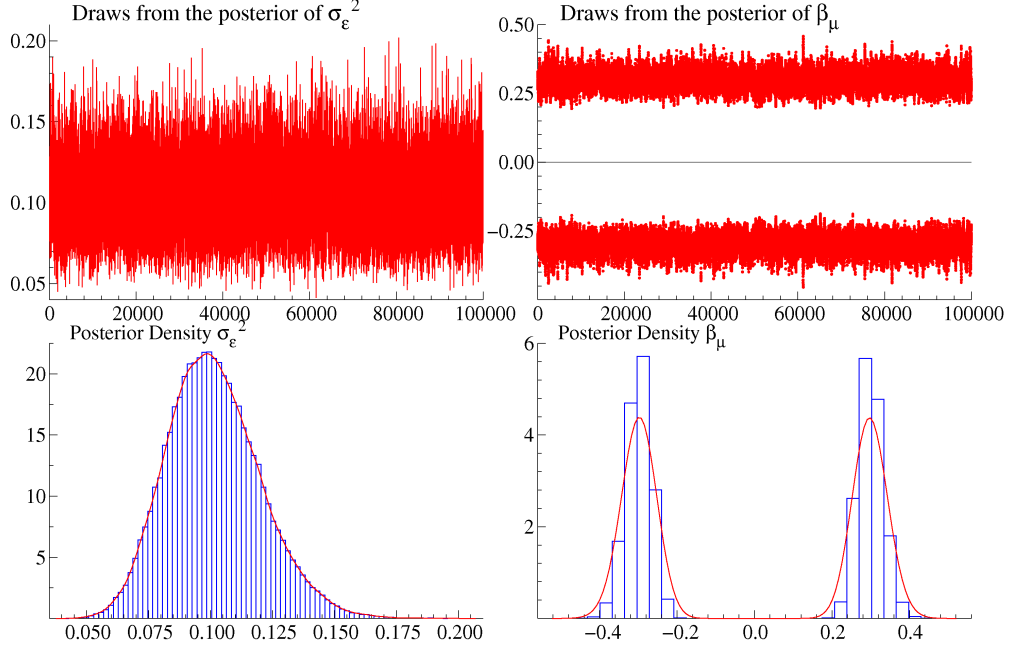


stochastic search of the most likely explanation of the observed time series is sought. After a large number of iterations of the GS scheme, model selection (and averaging, if one wishes) can be based on $\pi(\Upsilon|y)$, as estimated by the proportion of times a particular specification was drawn.

Full details on the statistical treatment will be postponed to section 4. We conclude this section by highlighting some estimation results for the series AWHMAN and core inflation. For hours worked, the model selected by the stochastic model specification search is M_4 , i.e. a stationary AR(2). The proportion of times model M_4 is selected varies slightly with the values of κ_μ ; if $\kappa_\mu = 1$, the proportion is 91.5%, whereas model M_8 is selected in 8.2% of the draws. If we let κ_μ increase, the proportion for M_4 quickly goes to 100%. Figure 2 displays the estimated posterior distribution obtained from 100,000 GS draws after a burn-in of 50,000 iterations. It should be noticed that the AR coefficients are close to the boundary of the stationary region, represented by the triangle of vertices $(\phi_1, \phi_2) = (-2, -1), (2, 1), (0, 1)$; furthermore they are close to the complex roots regions, but the sum of the AR coefficients has a posterior mean of about 0.9. When model M_8 is estimated, the posterior distribution of β_η has a large mass around zero, which is taken as evidence that the time variation in the intercept is not statistically detectable.

The U.S. inflation series provides also an interesting case study. We prefer to analyse the so

Figure 3: U.S. quarterly inflation (ex. food and energy) 1960:q1-2009.q4.



called core inflation series (excluding food and energy), because it is less affected by outliers. In this case the evidence is in favour of model M_5 , that is a local level model with no AR effects (see Harvey, 1989, Stock and Watson, 2007). Figure 3 displays 100,000 draws from the posterior of the parameters σ_ϵ^2 and β_μ , arising from a Gibbs sampling scheme with a burn-in of 50,000 iterations, along with the estimated posterior. It should be noticed that the posterior of β_μ is bimodal and symmetric around zero; the fact that the two modes are well separated is taken as evidence that a stochastic level, driven by disturbances with non zero variance, is present.

3 Stochastic and Deterministic Linear trends

The model for μ_t can be generalised to include a slope component. We are interested in investigating whether this further component is fixed or time varying; in the latter case the evolution over time is described by a random walk. This leads to the consideration of the local linear trend model for the component μ_t in (1):

$$\begin{aligned} \mu_t &= \mu_{t-1} + a_{t-1} + \eta_t, & \eta_t &\sim \text{NID}(0, \sigma_\eta^2) \\ a_t &= a_{t-1} + \zeta_t, & \zeta_t &\sim \text{NID}(0, \sigma_\zeta^2) \end{aligned} \quad (4)$$

where a_t is the slope component and we assume that η_t and ζ_t are mutually uncorrelated and independent of ϵ_t (see Harvey, 1989, and West and Harrison, 1997).

Denoting by μ_0 and a_0 the initial values of the level and slope components, the non-centred representation of (4) is the following:

$$\begin{aligned}\mu_t &= \mu_0 + a_0 t + \sigma_\eta \tilde{\mu}_t + \sigma_\zeta \tilde{A}_t, \\ \tilde{\mu}_t &= \tilde{\mu}_{t-1} + \tilde{\eta}_t, & \tilde{\eta}_t &\sim \text{NID}(0, 1), \\ \tilde{A}_t &= \tilde{A}_{t-1} + \tilde{a}_{t-1}, & \tilde{a}_t &= \tilde{a}_{t-1} + \tilde{\zeta}_t, \quad \tilde{\zeta}_t \sim \text{NID}(0, 1),\end{aligned}\tag{5}$$

so that $\tilde{\mu}_0 = \tilde{A}_0 = \tilde{a}_0 = 0$, and $\tilde{\zeta}_t = \zeta_{t-1}/\sigma_\zeta$. Thus, in the non-centred representation the mean function is explicitly written as a linear function of time and the stochastic part is the combination of a random walk and an integrated random walk, both starting off at the origin and driven by standardised independent disturbances.

As before, the non-centred representation is identified up to a sign switch, that $(-\sigma_\eta)(-\tilde{\mu}_t)$ has the same likelihood as $(\sigma_\eta)(\tilde{\mu}_t)$, and the same holds for the pair $(-\sigma_\zeta)(-\tilde{A}_t)$ and $(\sigma_\zeta)(\tilde{A}_t)$. Defining $\beta_\mu = (-1)^{B_1} \sigma_\eta \beta_A = (-1)^{B_2} \sigma_\zeta$, where B_1, B_2 are independent Bernoulli random variables with parameter 0.5, and correspondingly, $\mu_t^* = (-1)^{B_1} \tilde{\mu}_t$, $A_t^* = (-1)^{B_2} \tilde{A}_t$, we can reparameterise the model for y_t as

$$y_t = \mu_0 + \beta_\mu \mu_t^* + \beta_A A_t^* + a_0 t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t.$$

Further defining the multinomial vector $\Upsilon = (\gamma_1, \gamma_2, \delta_0, \delta_1, \delta_2)$, collecting the 0-1 binary indicator variables for the inclusion of the regression effects $\mu_t^*, A_t^*, t, y_{t-1}, y_{t-2}$, respectively, the general specification encompassing all the possible models is:

$$\begin{aligned}y_t &= \mu_0 + \gamma_1 \beta_\mu \mu_t^* + \gamma_2 \beta_A A_t^* + \delta_0 a_0 t + \delta_1 \phi_1 y_{t-1} + \delta_2 \phi_2 y_{t-2} + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\epsilon^2) \\ \tilde{\mu}_t &= \tilde{\mu}_{t-1} + \tilde{\eta}_t, & \tilde{\eta}_t &\sim \text{NID}(0, 1), \\ A_t^* &= A_{t-1}^* + a_{t-1}^*, a_t^* = a_{t-1}^* + \tilde{\zeta}_t, & \tilde{\zeta}_t &\sim \text{NID}(0, 1).\end{aligned}\tag{6}$$

The number of available models is $K = 2^5$. The different models will be labelled by

$$M_k, \quad k = 1 + 2^4 \gamma_1 + 2^3 \gamma_2 + 2^2 \delta_0 + 2 \delta_1 + \delta_2.$$

The model (6) nests several important special cases, that can be grouped as follows.

- a) Trend-stationary processes. They occur when $\sigma_\eta^2 = \sigma_\zeta^2 = 0$, and $\phi(L)$ is a stationary polynomial, i.e. $\phi(z) = 0 \iff |z| > 1$. Models $M_1 - M_8$ with stationary roots belong to this group.

b) y_t is a first or second order difference stationary process (i.e. $\Delta^d y_t$ has a stationary Wold representation, $d = 1, 2$). There are several ways by which difference stationary processes can arise as particular cases of (6).

1. y_t is a time-invariant AR(p) process characterized by at least one unit root. This occurs when $\sigma_\eta^2 = 0$ and $\phi(L)$ can be written as $\Delta\phi(L)^*$ (models M_3 and M_7) or $\phi(L) = \Delta^2$ (models M_4 and M_8).
2. y_t is a difference stationary process with multiple sources of disturbances, which occurs when either $\sigma_\eta^2 > 0$ or $\sigma_\zeta^2 > 0$, or both, and $\phi(L)$ is a stationary polynomial (Models $M_9 - M_{32}$). When $\sigma_\eta^2, \sigma_\zeta^2 > 0$ (Models $M_{25} - M_{32}$), the specification is ARIMA($p, 2, 2$). Models $M_9, M_{13}, M_{17}, M_{21}, M_{25}, M_{29}$, are variants of the local linear trend model (Harvey, 1989). In particular, in the first two cases the trend is an integrated random walk and the stationary component is white noise. The slope is fixed for models M_{17}, M_{21} : model M_{17} has $\gamma_1 = 1, \gamma_2 = \delta_0 = \delta_1 = \delta_2 = 0$, and it is coincident with the local level model, with centred representation $y_t = \mu_t + \epsilon_t, \mu_t = \mu_{t-1} + \eta_t$. Model M_{21} is the random walk with drift plus noise model $y_t = \mu_t + \epsilon_t, \mu_t = a + \mu_{t-1} + \eta_t$.

4 Statistical Treatment

The statistical treatment of model (6) is based on FS-W. In this section we discuss how perform model selection is carried out. In particular, we discuss our prior choices, and describe the algorithm used for computing the posterior distribution and the full conditional distributions.

Partition Υ as $\Upsilon = (\gamma, \delta)$, where $\gamma = (\gamma_1, \gamma_2)$ and $\delta = (\delta_0, \delta_1, \delta_2)$. According to the value of Υ , any particular model admits the the non-centered representation as linear mixed model of the following kind:

$$\begin{aligned} y_t &= x'_{\delta,t} \rho_\delta + z'_{\gamma,t} \alpha_{\gamma,t} + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2) \\ \alpha_{\gamma,t} &= T_\gamma \alpha_{\gamma,t-1} + R_\gamma u_{\gamma,t}, \quad u_{\gamma,t} \sim \text{NID}(0, I), \end{aligned} \quad (7)$$

where $\alpha_{\gamma,0} = 0$ and the intercept is always included in the vector $x_{\delta,t}$. For the full model, M_{32} , $\delta = (1, 1, 1)$ and $\gamma = (1, 1)$,

$$x_{\delta,t} = (1, t, y_{t-1}, y_{t-2})', \rho_\delta = (\mu_0, a_0, \phi_1, \phi_2)', z_{\gamma,t} = (\beta_\mu, \beta_A, 0)', \alpha_{\gamma,t} = (\mu_t^*, A_t^*, a_t^*).$$

$$T_\gamma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad R_\gamma = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

4.1 Prior specification

Let y denote the collection of time series values $\{y_t\}$ and α denote that of the latent states $\{\alpha_t\}$, ψ_Υ collects the parameters $(\mu_0, \beta_\mu, \beta_A, a_0, \phi_1, \phi_2)$ that enter the corresponding model.

The prior assumes an independent structure between each block of variables, such that:

$$\pi(\Upsilon, \psi, \sigma_\epsilon^2, \alpha) = \pi(\Upsilon)\pi(\sigma_\epsilon^2)\pi(\psi|\Upsilon, \sigma_\epsilon^2)\pi(\alpha|\Upsilon).$$

The prior distribution over the model space is uniform, that is $\pi(\Upsilon) = 2^{-5}$.

For the irregular variance a hierarchical inverse gamma prior is adopted, $\sigma_\epsilon^2 \sim \text{IG}(c_0, C_0)$, where $C_0 \sim \text{G}(g_0, G_0)$, $\text{G}(\cdot)$ denoting the Gamma distribution, $c_0 = 2.5$, $g_0 = 5$, and $G_0 = g_0/[0.75\text{Var}(y_t)(c_0 - 1)]$, as in FS-W. The hierarchical prior makes the posterior distributions less sensitive to the choice of the hyperparameters of the IG distribution; it obviously requires an additional sampling step where C_0 is sampled conditional on σ_ϵ^2 from the conditional Gamma posterior $C_0|\sigma_\epsilon^2 \sim \text{G}(g_0 + c_0, G_0 + 1/\sigma_\epsilon^2)$ at each sweep of the sample.

As far as the vector ψ_Υ is concerned, we adhere to the general prescription by Koop (2003), according to which, 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. For instance, for the full model,

$$\pi(\psi_\Upsilon|\Upsilon, \sigma_\epsilon^2) = \pi(\mu_0|\sigma_\epsilon^2)\pi(\beta_\mu|\sigma_\epsilon^2)\pi(\beta_A|\sigma_\epsilon^2)\pi(a_0|\sigma_\epsilon^2)\pi(\phi_1|\sigma_\epsilon^2)\pi(\phi_2|\sigma_\epsilon^2).$$

we take the conjugate priors $\beta_\mu|\sigma_\epsilon^2 \sim \text{N}(0, \kappa_\mu\sigma_\epsilon^2)$ and $\beta_A|\sigma_\epsilon^2 \sim \text{N}(0, \kappa_A\sigma_\epsilon^2)$, whereas for the autoregressive parameters we adopt a normal prior of the type: $\phi_i|\sigma_\epsilon^2 \sim \text{N}(0, d_i\sigma_\epsilon^2)$, $i = 1, 2$. Also, $a_0|\sigma_\epsilon^2 \sim \text{N}(0, d_0\sigma_\epsilon^2)$. For the constant term we adopt the uninformative prior $\pi(\mu_0|\sigma_\epsilon^2) \propto 1$ or the proper conjugate prior $\mu_0|\sigma_\epsilon^2 \sim \text{N}(0, q_0\sigma_\epsilon^2)$, where q_0 is a large number. The default values chosen for the scale factors are $\kappa_\mu = \kappa_A = d_0 = d_1 = d_2 = 10$.

Finally, the prior for α is provided by the Gaussian dynamic model (7), so that,

$$p(\alpha) = p(\alpha_{\gamma 0}) \prod_{t=1}^n p(\alpha_{\gamma t}|\alpha_{\gamma, t-1}),$$

with $\alpha_{\gamma t}|\alpha_{\gamma, t-1} \sim \text{N}(T_\gamma\alpha_{\gamma, t-1}, R_\gamma R'_\gamma)$ and $\alpha_{\gamma, 0} = 0$.

4.2 MCMC Estimation

Model selection requires the evaluation of the posterior probability function of the multinomial vector Υ , denoted $\pi(\Upsilon|y)$. Also, for the selected model we are interested in the marginal posterior distributions of the parameters $\pi(\psi|y)$ and the states $\pi(\alpha|y)$. The required posteriors are not

available in closed form, but we are capable of drawing samples from them by Monte Carlo Markov Chain methods and, in particular, by a Gibbs sampling (GS) scheme that we now are going to discuss in some detail. The GS scheme produces correlated random draws from the posteriors by repeatedly sampling an ergodic Markov chain whose invariant distribution is the target density; see e.g. Chib (2001), Robert and Casella (2004), Gamerman and Lopes (2007). In essence, it defines a homogeneous Markov Chain such that the transition kernel is formed by the full conditional distributions and the invariant distribution is the unavailable target density.

The GS scheme can be sketched as follows. Specify a set of initial values $\Upsilon^{(0)}, \sigma_\epsilon^{2(0)}, \alpha^{(0)}, \psi^{(0)}$. For $i = 1, 2, \dots, M$, iterate the following operations:

- a. Draw $\Upsilon^{(i)} \sim \pi(\Upsilon|\alpha^{(i-1)}, y)$
- b. Draw $\sigma_\epsilon^{2(i)} \sim \pi(\sigma_\epsilon^2|\Upsilon^{(i)}, \psi^{(i-1)}, \alpha^{(i-1)}, y)$
- c. Draw $\psi^{(i)} \sim \pi(\psi|\Upsilon^{(i)}, \sigma_\epsilon^{2(i)}, \alpha^{(i-1)}, y)$
- d. Draw $\alpha^{(i)} \sim \pi(\alpha|\Upsilon^{(i)}, \sigma_\epsilon^{2(i)}, \psi^{(i)}, y)$

The above complete conditional densities are available, up to a normalizing constant, from the form of the likelihood and the prior.

For the sake of notation, let us write the regression model as $y = Z_\Upsilon \psi_\Upsilon + \epsilon$, where y and ϵ are vectors stacking the values $\{y_t\}$ and $\{\epsilon_t\}$, respectively, and the generic row of matrix Z_Υ contains the relevant subset of the explanatory variables $(1, \mu_t^*, A_t^*, t, y_{t-1}, y_{t-2})$.

Step a. is carried out by sampling the indicators with probabilities proportional to the conditional likelihood of the regression model, as

$$\begin{aligned} \pi(\Upsilon|\alpha, y) &\propto \pi(\Upsilon)\pi(y|\Upsilon, \alpha) \\ &\propto \pi(y|\Upsilon, \alpha), \end{aligned}$$

which is available in closed form (see below).

Under the normal-inverse gamma conjugate prior for $(\psi_\Upsilon, \sigma_\epsilon^2)$

$$\sigma_\epsilon^2 \sim \text{IG}(c_0, C_0), \quad \psi_\Upsilon|\sigma_\epsilon^2 \sim \text{N}(0, \sigma_\epsilon^2 D_\Upsilon),$$

where, e.g. for the model $\Upsilon = (1, 1, 1, 1, 1)$, $D_\Upsilon = \text{diag}(q_0, \kappa_\mu, \kappa_A, d_0, d_1, d_2)$, steps b. and c. are carried out by sampling from the posteriors

$$\begin{aligned} \sigma_\epsilon^2|\Upsilon, \alpha, y &\sim \text{IG}(c_{T^*}, C_{T^*}) \\ \psi_\Upsilon|\Upsilon, \sigma_\epsilon^2, \alpha, y &\sim \text{N}(m, \sigma_\epsilon^2 S) \end{aligned}$$

where

$$\begin{aligned} S &= (Z'_{\Upsilon} Z_{\Upsilon} + D_{\Upsilon}^{-1})^{-1}, & m &= SZ'_{\Upsilon} y \\ c_{T^*} &= c_0 + T^*/2, & C_{T^*} &= C_0 + \frac{1}{2} (y'y - m'S^{-1}m). \end{aligned}$$

Finally,

$$\pi(y|\Upsilon, \alpha) \propto \frac{|S|^{0.5} \Gamma(c_{T^*}) C_0^{c_0}}{|D_{\Upsilon}|^{0.5} \Gamma(c_0) C_{T^*}^{c_{T^*}}}$$

see e.g. Geweke (2005), where $\Gamma(\cdot)$ denotes the Gamma function.

The sample from the posterior distribution of the latent states, conditional on the model and its parameters, in step d., is obtained by the conditional simulation smoother proposed by Durbin and Koopman (2002) for linear and Gaussian state space models.

Finally, the draw of the parameters β_{μ} and β_A are obtained by performing a final random sign permutation. This is achieved by drawing independently Bernoulli random variables B_r , $r = 1, 2$ (here again we refer to the model including both level and slope), with probability 0.5, and recording $(-1)^{B_1}(\sigma_{\eta}, \tilde{\mu}_t)$, and $(-1)^{B_2}(\sigma_{\zeta}, \tilde{A}_t, a_t)$.

5 Empirical Results

We apply variable selection to two data sets dealing with U.S. macroeconomic time series. The first is the original Nelson and Plosser (1982, NP) data set, consisting of 15 annual time series which are a testbed for unit root and stationary testing. The series are listed in table 1; the sample sizes range from 62 to 111 observations for each series. Except for the bold yield, all the series are transformed into natural logarithms. The conclusions reached by NP, matured on the evidence of unit root tests, have been revisited since then many times as new methodologies were proposed. We follow suit by performing stochastic model specification search according to the methodology presented in the previous sections.

We also consider an additional data set (ADS) made up of some relevant quarterly and monthly and quarterly time series (made available at the Federal Reserve Bank of St.Louis Economic Data website, <http://research.stlouisfed.org/fred2/>), listed in table 2.

Variable selection is implemented in Ox 6.0 (Doornik, 2007) using our source code. Tables 3 and 4 present the frequency by which model M_k , $k = 1 + 2^4\gamma_1 + 2^3\gamma_2 + 2^2\delta_0 + 2\delta_1 + \delta_2$, was selected in 100,000 iterations of the GS scheme outlined in the previous section, after a burn-in of 50,000 iterations (sensible starting values for the parameters are obtained by running the unrestricted model, model 32, without variable selection for the first 1000 draws). The time needed to perform MCMC model selection for a time series of around 500 observations is less than 30 minutes using a standard desktop computer.

Table 1: The Nelson and Plosser data set.

Series description	Sample period	Name
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	Empl
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 yields, 30-year corporate	1900 - 1970	Interest
Stock prices	1871 - 1970	Pstock

Table 2: Data set ADS.

Series description	Sample period	Name
Gross Domestic Product (chained 2005 volumes)	Quarterly	1947.1 - 2009.4 GDP
Gross National Product (chained 2005 volumes)	Quarterly	1947.1 - 2009.4 GNP
Consumer Price Index (all items)	Monthly	1960.1 - 2009.12 CPI
Inflation rate (all items)	Monthly	1960.1 - 2009.12 MInfl
Inflation rate	Quarterly	1960.1 - 2009.4 QInfl
GDP Deflator	Quarterly	1947.1 - 2009.4 GDPdefl
Industrial Production Index	Monthly	1960.1 - 2009.12 IP
Unemployment rate	Monthly	1960.1 - 2009.12 MUnempl
Unemployment rate	Quarterly	1960.1 - 2009.4 QUnempl
Average Weekly Hours Worked (Manuf.)	Monthly	1960.1 - 2009.12 MHWorked
Average Weekly Hours Worked (Manuf.)	Quarterly	1960.1 - 2009.4 QHWorked

Table 3: Stochastic Model Specification Search for the Nelson and Plosser data set. Percentages by which model M_k is selected in 100,000 iterations of the Gibbs sampler (burn in period 50,000 GS iterations).

Series	Model																
	M_3	M_7	M_8	M_9	M_{11}	M_{12}	M_{13}	M_{15}	M_{16}	M_{17}	M_{18}	M_{19}	M_{20}	M_{23}	M_{24}	M_{25}	M_{28}
<i>Rgnp</i>			94.5			0.2						2.2	0.3	1.7	1.1		
<i>Ngnp</i>			90.2			0.9			0.4			2.9	1.2	2.1	2.3		
<i>Rpcgnp</i>			89.6									4.8	1.6	2.3	1.7		
<i>Iprod</i>		99.8													0.2		
<i>Empl</i>			98.2									0.5	0.2	0.4	0.7		
<i>Unempl</i>	87.5									1.8	1.5	2.6	4.6		2.0		
<i>Gnpdefl</i>			76.3			0.3						12.9	1.3	7.2	2.0		
<i>Pcons</i>						20.6			10.5				44.0		22.5		2.4
<i>Nwage</i>			93.7			0.7				0.4	1.7		1.0	0.9	1.6		
<i>Rwage</i>		98.4										0.5		1.1			
<i>Money</i>			98.5			0.2			0.1				0.3		0.9		
<i>Veloc</i>	93.2											3.9		2.9			
<i>Interest</i>				50.3	8.5		21.8	6.0				5.5		3.1		4.8	
<i>Pstock</i>	21.7	31.3				2.9			1.4			9.4	12.1	6.3	14.9		

As far as the NP data set is concerned, stochastic model specification search leads to the selection of model M_8 for most of the series, not only for real gross national product, but also for nominal GNP, nominal wages and the GNP deflator. Notice that this model has $\Upsilon = (0, 0, 1, 1, 1)$, that is a trend stationary AR(2) model:

$$y_t = \mu_0 + a_0 t + \phi_1 y_{t-1} + \phi_2 y_{t-1} + \epsilon_t, \epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2).$$

Also, similar results are obtained for industrial production and real wages, for which the selected model is M_7 , i.e. $y_t = \mu_0 + a_0 t + \phi_1 y_{t-1} + \epsilon_t$. For unemployment model M_3 , which is a stationary AR(1) model with intercept term and no slope, is selected in 88% of the draws. Apparently, only this result is consistent with NP findings. This is a striking and new result for the unit root literature, although similar evidence is found in Schotman and van Dijk (1991) and Koop and van Dijk (2000), at least for GNP and IP. Also, the results for GNP, employment, the unemployment rate, the GNP deflator and money are in accord with Phillips (1991) and Kwiatowski et al. (1992).

On the contrary, for the consumer price index (*Pcons*) the modal specification is M_{20} , which features a stochastic level, whereas for interest rates and stock prices the estimated posterior model probabilities are rather sparse.

These results were obtained by assuming a Gaussian conjugate prior with scale factors equal to 10, i.e. $\kappa_\mu = \kappa_A = d_0 = d_1 = d_2 = 10$. The original time series were scaled by the variance of the second order differences of the series $\Delta^2 y_t$ so as to avoid that the draws of σ_ϵ^2 were too small.

We assessed prior sensitivity by letting the prior variance be smaller, $\kappa_\mu = \kappa_A = d_0 = d_1 = d_2 = 1$, and larger, $\kappa_\mu = \kappa_A = d_0 = d_1 = d_2 = 100$. The evidence is that the distribution of $\Upsilon|y$ is less concentrated when the prior variance is small, in that the proportion of times stochastic levels and slopes are selected is larger, and tending to be more concentrated on the modal model when the prior variance is large as compared to σ_ϵ^2 .

The evidence arising from the application of MCMC model specification search on the data set ADS, described in table 2, confirms the trend stationarity of the real gross national and domestic product. However, for monthly industrial production we select model 23, which is

$$y_t = \mu_0 + a_0 t + \sigma_\eta \tilde{\mu}_t + \phi_1 y_{t-1} + \epsilon_t, \epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2).$$

Against this background, the overwhelming evidence in favour of pure AR models requires a closer investigation. Further insight on the problem is obtained by estimating model M_8 by MCMC and by considering the estimated posteriors of the AR parameters (ϕ_1, ϕ_2) . Figure 4 displays the sample distribution of the 100,000 draws from the joint posterior of (ϕ_1, ϕ_2) along with the estimated posterior density of $\phi_1 + \phi_2$, both for the NP GNP annual series, and for the quarterly GDP series

Table 4: Stochastic Model Specification Search for selected U.S. macroeconomic time series (see table 2). Percentages by which model M_k is selected in 100,000 iterations of the Gibbs sampler (burn in period 50,000 GS iterations).

Series	Model														
	M_4	M_8	M_{12}	M_{16}	M_{17}	M_{18}	M_{19}	M_{20}	M_{21}	M_{22}	M_{23}	M_{24}	M_{27}	M_{28}	M_{31}
<i>GDP</i>	100.0														
<i>GNP</i>	100.0														
<i>CPI</i>								48.0				52.0			
<i>MInfl</i>						65.5	2.2		31.2		1.1				
<i>QInfl</i>					52.0	4.2	9.2	0.5	26.35	5.3					
<i>GDPDefl</i>			15.8	7.6									49.9	3.8	22.9
<i>IP</i>							41.8				58.0	0.2			
<i>MUnempl</i>								67.1				32.9			
<i>QUnempl</i>	90.8						0.4	6.2				2.6			
<i>MHWorked</i>					67.2		2.4		29.2		1.2				
<i>QHWorked</i>	91.5						3.3	2.5			1.5	1.2			

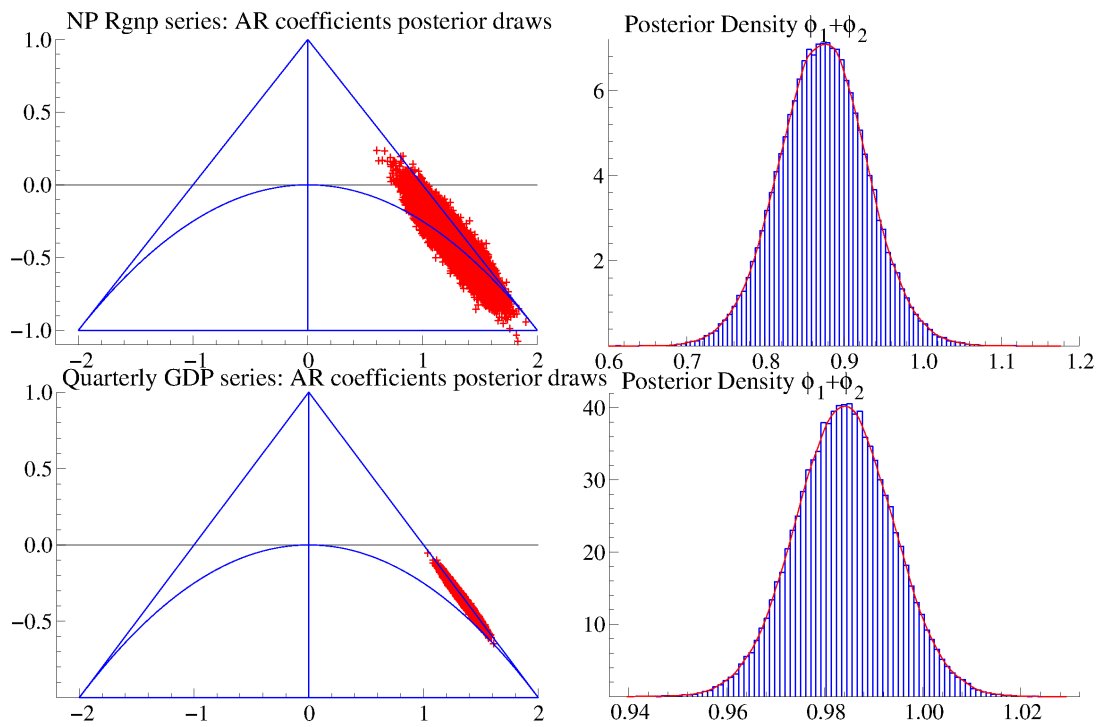
(similar results being obtained for quarterly GNP). We notice that the series cover very different sample periods.

The exploration of these posteriors highlights that the AR polynomial is close to the nonstationarity region. This is particularly true of quarterly GDP. This fact is rather general: most of the series for which pure AR models were selected can be described as quasi-integrated time series.

We should remark at this point that the model selection procedure that we outlined has a characteristic property, which can be illustrated with reference to the case when the true model is a pure random walk with drift: $\Delta y_t = c + \eta_t$. If a series is actually generated in this way, the only way the true model can be reflected in our set is as a special case of M_2 ($\Upsilon = (0, 0, 0, 1, 0)$), with the AR coefficient sufficiently close to unity. The specification $\Upsilon = (1, 0, 0, 0, 0)$, corresponding to model M_{17} , whose non-centred representation is $y_t = \mu_0 + \sigma_\eta \tilde{\mu}_t + \epsilon_t$, with σ_η strictly greater than zero, would be characterised by a smaller marginal likelihood, since we face a random walk plus noise model, with ARIMA(1,1,1) reduced form, unless $\sigma_\epsilon^2 = 0$.

Now, an intrinsic assumption of variable selection as applied to unobserved components models is that σ_ϵ^2 is strictly greater than zero. MCMC inference breaks down when σ_ϵ^2 is allowed to be zero. Assuming $\sigma_\epsilon^2 > 0$ is both a strong point and a limitation of the approach; a strong point, as

Figure 4: U.S. annual GNP (NP data set) and quarterly GDP (1947.1-2009.4). Draws from the posterior distribution of the AR coefficients (ϕ_1, ϕ_2) and estimated posterior density of $\phi_1 + \phi_2$.



is allows to carry out variable selection for unobserved random effects using a regression framework (see George and McCulloch, 1993). The random walk plus noise model, with strictly positive noise variance, implies some kind of mean reversion (or low persistence) that is not supported for series like those belonging to the NP dataset. Temporal aggregation (the series are annual) may play a role in determining a highly persistent behaviour.

In conclusion, the selection of models $M_k, k \leq 8$ does not necessarily entails that the series are trend stationary. A closer look at the posteriors reveals that the AR coefficients are close to the boundaries of the stationarity region. In the light of the above discussion, we interpret the outcome of variable selection as pointing out the presence of quasi unit root behaviour, with no significant mean reversion (attributable to a negative moving average root).

The analysis of the monthly CPI and inflation provides an interesting case study. The modal choice for monthly inflation ΔCPI_t ($MInfl$) is the AR(1) with stochastic level

$$(1 - \phi_1 L)\Delta CPI_t = \mu_0 + \sigma_\eta \tilde{\mu}_t + \epsilon_t$$

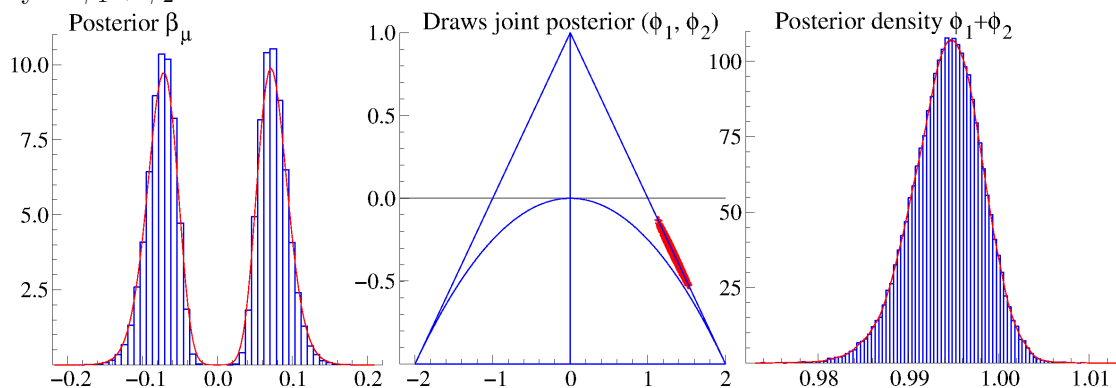
The posterior mean of the AR coefficient is 0.33. The fact that for the series $QInfl$ the selected model has no AR feature (M_{17}) can be attributed to temporal aggregation. A few comments are due for explaining the CPI result: Bayesian model selection leads to model M_{24} , whereas we would expect M_{15} which results from integrating the model selected for ΔCPI_t , which is M_{19} . Despite the fact that the model for CPI, $(1 - \phi_1 L - \phi_2 L^2)CPI_t = \mu_0 + \sigma_\eta \tilde{\mu}_t + \epsilon_t$, has no integrated random walk component, the distribution of the AR coefficients has a root close to 1, as it evident from the posterior draws and the distribution of the sum of the AR coefficients, displayed in figure 5. This suggests that the selected model for CPI can be actually reparameterised as:

$$(1 - \phi_1 L)\Delta CPI_t = \mu_0 + \sigma_\eta \tilde{\mu}_t + \epsilon_t.$$

Hence, the conflict with the model selected for monthly inflation is only apparent. Notice that, integrating both sides, $(1 - \phi_1 L)CPI_t$ can be rewritten as a linear trend plus an integrated random walk, with no irregular term. The only way in which such a model could arise in our model set is indeed via M_{17} , allowing for a unit root in the AR coefficients.

A final comment is due for unemployment and hours worked. While the quarterly series are stationary around a fixed level, the monthly series appears to be integrated. There is a possible explanation, related to temporal aggregation, especially in the case of unemployment. The fact is that the quarterly AR(2) coefficients are close to the boundary of the stationary region and imply a pseudo-cyclical behaviour with a very long period. The monthly series provides a better separation of the low frequency spectral peak due to the cycle from the long run trend (the zero frequency).

Figure 5: U.S. Consumer price index (all urban consumers) (1960.1-2009.12). Posterior density of β_μ , draws from the posterior distribution of the AR coefficients (ϕ_1, ϕ_2) and estimated posterior density of $\phi_1 + \phi_2$.



6 Conclusions

The paper has considered Bayesian model selection via MCMC methods for assessing the nature of trends in macroeconomic time series. The contribution of this paper to the literature, and in particular with respect to Frühwirth-Schnatter and Wagner (2010), is the inclusion of autoregressive terms into the model selection problem as well as the application to a large data set consisting of the original Nelson and Plosser series and a set of key economic indicators of the U.S. economy.

We conclude that the consideration of the autoregressive component is essential for the characterisation of the selected model. For instance, when the outcome of the selection is a pure autoregressive model with a deterministic intercept, the posterior distribution of the autoregressive coefficients provides a remarkably clear indication of the distance from a unit root process; this information is less easy to appraise from unit roots and stationarity tests.

The empirical application has shown that most annual time series in the Nelson and Plosser data set are better characterised as trend stationary; however, the posterior distribution of the sum of the autoregressive coefficients is in some cases highly concentrated on the boundary of the stationary region, leading to a quasi unit root process. With the notable exception of the monthly inflation rate, we found no support for the presence of an orthogonal decomposition into a stochastic trend and a cyclical component.

References

- Beveridge, S. and Nelson, C. R. (1981). “A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement

- of the Business Cycle”, *Journal of Monetary Economics*, Vol. 7, 151 - 174.
- Caner, M. and Kilian, L. (2001). “Size distortions of tests of the null hypothesis of stationarity: evidence and implications for the PPP debate”, *Journal of International Money and Finance*, Vol. 20, 639 - 657.
- Christiano, L., Eichenbaum, M., and Vigfusson, R. (2003). What happens after a technology shock? *Federal Reserve Board, International Finance Discussion Paper*, 2003 - 2768.
- Chib, S. (2001). “Markov Chain Monte Carlo Methods: Computation and Inference”, in *Handbook of Econometrics*, Vol. 5, (eds J.J. Heckman and E. Leamer), North Holland, Amsterdam, 3569 - 3649.
- De Jong, D. N. and Whiteman, C. H. (1991). “The Case for Trend-Stationarity Is Stronger Than We Thought”, *Journal of Applied Econometrics*, Vol. 6, 413 - 421.
- DeJong, D. N., Nankervis, John, C., Savin, N. E. and Whiteman, Charles H. (1992). “The power problems of unit root test in time series with autoregressive errors,” *Journal of Econometrics*, Vol. 53, 323 - 343.
- 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*, Vol. 74, 427 - 431.
- Doornik, J.A. (2007), *Ox: An Object-Oriented Matrix Programming Language*, Timberlake Consultants Press, London.
- Durbin, J. and Koopman, S.J. (2002). “A simple and efficient simulation smoother for state space time series analysis”, *Biometrika*, Vol. 89, 603 - 615.
- Elliott, G., Rothenberg, T.J. and Stock, J.H. (1996). “Efficient Tests for an Autoregressive Unit Root”, *Econometrica*, Vol. 64, 813 - 836.
- Frühwirth-Schnatter, S. (1995). “Bayesian model discrimination and Bayes factor for linear gaussian state space models”. *Journal of the Royal Statistical Society, series B*, Vol. 57, 237 - 246.
- Frühwirth-Schnatter, S. (2004). “Efficient Bayesian Parameter Estimation for State Space Models Based on Reparameterizations”. In Harvey, A.C., Koopman, S.J. and Shephard, N. (eds.), *State Space and Unobserved Component Models: Theory and Applications*, Proceedings of a Conference in Honour of James Durbin, Cambridge University Press, 123 - 151.

- Frühwirth-Schnatter, S. and Wagner, H. (2010). “Stochastic model specification search for Gaussian and partial non-Gaussian state space models”, *Journal of Econometrics*, Vol. 154, 85 - 100.
- Galí, G. (1999). “Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?”. *American Economic Review*, Vol. 89, 249 - 271.
- Gamerman, D. and Lopes, F. H. (2007). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Second Edition (Chapman and Hall/CRC Texts in Statistical Science).
- Gelfand, A.E., Sahu, S.K., and Carlin, B.P. (1995), “Efficient Parameterizations for Normal Linear Mixed Models”, *Biometrika*, Vol. 82, 479 - 488.
- George, E. I. and McCulloch, R. (1993). “Variable selection via Gibbs sampling”, *Journal of the American Statistical Association*, Vol. 88, 881 - 889.
- George, E. I. and McCulloch, R. (1997). “Approaches for Bayesian variable selection”, *Statistica Sinica*, Vol. 7, 339 - 373.
- John F. Geweke, J.F. (2005). *Contemporary Bayesian econometrics and statistics*, Wiley, New York.
- Harvey, A.C. (1989). *Forecasting, Structural Time Series and the Kalman Filter*, Cambridge University Press, Cambridge, UK.
- Harvey, A. (2001). “Testing in Unobserved Components Models”, *Journal of Forecasting*, Vol. 20, 1 - 19.
- Koop, G. (1992). “Objective’ Bayesian Unit root tests”, *Journal of Applied Econometrics*, Vol. 7, 65 - 82.
- Koop, G. (1994). “Recent progress in applied Bayesian Econometrics”, *Journal of Economic Survey*, Vol. 8, 1 - 34.
- Koop, G. (2003). *Bayesian Econometrics*, John Wiley and Sons Ltd, England.
- Koop, G. and Van Dijk, H. (2000). “Testing for integration using evolving trend and seasonals models: A Bayesian approach”, *Journal of Econometrics*, Vol. 97, 261 - 291.
- Kwiatkowski, D. , Phillips, P. C. B. , Schmidt, P. and Y. Shin (1992). “Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root”. *Journal of Econometrics*, Vol. 54, 159 - 178.

- Leybourne, S. J. and McCabe, B. P. M. (1994). "A Consistent Test for a Unit Root", *Journal of Business and Economic Statistics*, Vol. 12, 157 - 166.
- Nelson, C. R. and Plosser, C. I. (1982). "Trend and Random Walk in Macroeconomic Time Series", *Journal of Monetary Economics*, Vol. 10, 139 - 162.
- Ng, S. and Perron, P. (2001). "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power", *Econometrica*, Vol. 69, 1519 - 1554.
- Nyblom J. and Makelainen T. (1983). "Comparisons of Tests for the Presence of Random Walk Coefficients in a Simple Linear Model", *Journal of the American Statistical Association*, Vol. 78, 865 - 864.
- Perron, P. (1989). "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", *Econometrica*, Vol. 57, 1361 - 1401.
- Phillips, P.C.B. (1991). "To criticize the critics: An objective Bayesian analysis of stochastic trends", *Journal of Applied Econometrics*, Vol. 6, 333 - 364.
- Phillips, P.C.B and Perron, P. (1988), "Testing for a Unit Root in Time Series Regression", *Biometrika*, Vol. 75, 335 - 346.
- Phillips, P.C.B., and Ploberger, W. (1994). "Posterior odds testing for a unit root with data-based model selection", *Econometric Theory*, Vol. 10, 774 - 808.
- Proietti, T. and Harvey, A.C. (2000). "A Beveridge Nelson Smoother", *Economics Letters*, Vol. 67, 139 - 146.
- Robert, P.C and Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer, New York, USA.
- Schotman, P. C. and van Dijk, H.K. (1991). "On Bayesian Routes to Unit Roots", *Journal of Applied Econometrics*, Vol. 6, 387 - 401.
- Schwert, G. W. (1989). "Tests for Unit Roots: A Monte Carlo Investigation", *Journal of Business and Economic Statistics*, Vol. 7, 147 - 159.
- Sims, C.A. (1988). "Bayesian skepticism on unit root econometrics", *Journal of Economic Dynamic and Control*, Vol. 12, 463 - 474.
- Sims, C.A. and Uhlig, H. (1991). "Understanding unit rooters: A helicopter tour", *Econometrica*, Vol. 59, 1591 - 1599.

- Smith, M. and Kohn, R. (1996). “Nonparametric Regression using Bayesian Variable Selection”, *Journal of Econometrics*, Vol. 75, 317 - 343.
- Stock, J.H. and Watson, M.W. (2007). “Why Has U.S. Inflation Become Harder to Forecast?”, *Journal of Money, Credit and Banking*, Vol. 39, 3 - 34.
- Strickland, C. M., Martin, G.M. and Forbes, C.S. (2007). “Parameterisation and efficient MCMC estimation of non-Gaussian state space models”, *Computational Statistical and Data Analysis*, Vol. 52, 2911 - 2930.
- West, M., and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*, 2nd ed., New York, Springer-Verlag.

Research Papers 2011



- 2011-02: Timo Teräsvirta: Nonlinear models for autoregressive conditional heteroskedasticity
- 2011-03: Roxana Halbleib and Valeri Voev: Forecasting Covariance Matrices: A Mixed Frequency Approach
- 2011-04: Mark Podolskij and Mathieu Rosenbaum: Testing the local volatility assumption: a statistical approach
- 2011-05: Michael Sørensen: Prediction-based estimating functions: review and new developments
- 2011-06: Søren Johansen: An extension of cointegration to fractional autoregressive processes
- 2011-07: Tom Engsted and Stig V. Møller: Cross-sectional consumption-based asset pricing: The importance of consumption timing and the inclusion of severe crises
- 2011-08: Tommaso Proietti and Stefano Grassi: Bayesian stochastic model specification search for seasonal and calendar effects
- 2011-09: Matt P. Dziubinski: Option valuation with the simplified component GARCH model
- 2011-10: Tim A. Kroencke, Felix Schindler and Andreas Schrimpf: International Diversification Benefits with Foreign Exchange Investment Styles
- 2011-11: Eduardo Rossi and Paolo Santucci de Magistris: Estimation of long memory in integrated variance
- 2011-12: Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Generalized Jackknife Estimators of Weighted Average Derivatives
- 2011-13: Dennis Kristensen: Nonparametric Detection and Estimation of Structural Change
- 2011-14: Stefano Grassi and Paolo Santucci de Magistris: When Long Memory Meets the Kalman Filter: A Comparative Study
- 2011-15: Antonio E. Noriega and Daniel Ventosa-Santaularia: A Simple Test for Spurious Regressions
- 2011-16: Stefano Grassi and Tommaso Proietti: Characterizing economic trends by Bayesian stochastic model specification search