# The marginal likelihood of Structural Time Series Models, with application to the US and the euro area NAIRU *

G.Fiorentini[(a)], C.Planas[(b)] and A.Rossi[(b)]

January 2008

**Abstract**

We propose a new result that simplifies the evaluation of the marginal likelihood in Gaussian Structural Time Series (STS) models. For this we exploit the statistical properties of STS models and a theorem in Dickey (1968) to obtain the likelihood marginally to all variance parameters. Our strategy applies under inverted gamma-2 prior distributions for the structural shocks variances. In general, we show that marginalizing with respect to variance parameters greatly improves the accuracy of the Laplace method. Moreover, in some empirically relevant cases such as the local level and the local linear trend, it yields the marginal likelihood by single or double integration over a finite support. We use our methodology to analyze models for the NAIRU in the US and in the euro area.

KEYWORDS: Bayesian model selection, bridge sampling, Laplace method, Markov Chain Monte Carlo, unobserved components.

*The ideas expressed here are those of the authors and do not necessarily reflect the positions of the European Commission. Thanks are due to Daniel Grenouilleau, Martin Hradisky, and Werner Roeger for useful discussions.

($a$) University of Florence, Department of Statistics, and Rimini Centre for Economic Analysis.

($b$) Joint Research Centre of the European Commission.

E-Mails: *Fiorentini@ds.unifi.it*, *Christophe.Planas@jrc.it*, *Alessandro.Rossi@jrc.it*.

# 1 Introduction

Since the first studies in the 1970's (see Pagan, 1975), Structural Time Series (STS) models have become quite widespread in empirical macroeconomics. Researchers typically resort to structural components for describing potential output (Clark, 1987), technological growth (Hansen, 1997), permanent income (Hall and Mishkin, 1982), reservation wage (Planas, Roeger and Rossi, 2007), and core inflation (Cogley and Sargent, 2005; Stock and Watson, 2007). Ideally macroeconomics should guide the time series model choice. Economic theories however rarely point to a unique specification, and discriminating between different possibilities can be a difficult task, especially when classical inference becomes non-standard. In the STS context, examples are testing for a deterministic versus a stochastic component (see Harvey, 2001) and inferring about a stationary against a non-stationary slope. Through the marginal likelihood, the Bayesian framework offers a conceptually simple answer to the model selection problem (see Kass and Raftery, 1995, or Koop, 2003), with the important advantage of involving exact finite sample distributions instead of asymptotic approximations. Estimating the marginal likelihood can however be non-trivial: the number of parameters to be integrated out is usually relevant and the likelihood function is typically highly concentrated with respect to the prior distribution (see Fruhwirth-Schnatter, 2004).

In this paper we simplify the evaluation of the marginal likelihood in Gaussian STS models. For this we exploit the statistical properties of STS models and a theorem by Dickey (1968) to obtain the likelihood function marginally to all variance parameters. Our strategy applies under inverted gamma-2 prior distributions for the structural shocks variances, an assumption that is prevailing in the time series literature (see Fruhwirth-Schnatter, 1994; Chib and Greenberg, 1994). In general, we show that marginalizing out the variance parameters greatly improves the accuracy of the Laplace method (see Tierney and Kadane, 1986). Moreover, in some empirically relevant cases such as the local level and the local linear trend, it yields the marginal likelihood by single or double integration over a finite support, without any Monte Carlo Markov Chain (MCMC) sampling.

In Section 2 we derive simple expressions for the likelihood function marginally to the variance parameters. For most trend plus noise models, this gives an easy tool for isolating the most likely trend equation. We also give an insight about the posterior distribution of the unobserved components in the local level model. More complex models are considered in Section 3. For these we suggest to integrate out the conditional mean parameters by applying the Laplace method after marginalizing with respect to the variance parameters. We show in a simulation study that this strategy greatly improves over the traditional Laplace marginal likelihood estimator, and that the results are comparable to bridge sampling (Meng and Wong, 1996) although no importance sampling is needed.

This methodology gives us the opportunity to weight models for the NAIRU in the US and in the euro area (EA). The NAIRU is of particular interest because it is related to the imperfect equilibrium of the labor market. The European Commission (EC) uses it for evaluating the potential growth of Member States and for the cyclical adjustment of budget balances, in application of the Stability and Growth Pact (see Denis, Grenouilleau, Mc Morrow and Roeger, 2006). Central Banks also scrutinize the NAIRU when monitoring inflation, following Phillips curve theory (see for instance Stiglitz, 1997). A debate is however still open about the dynamic properties of the NAIRU, especially for the US (see for instance Staiger, Stock and Watson, 1997; Ball and Mankiw, 2002). In Section 4 we rank twenty seven models for the US and for the EA NAIRU. We end up with some empirical evidence that, we believe, can be of interest to policy makers and analysts. Section 5 concludes.

# 2 The marginal likelihood of trend plus noise models

## 2.1 Background

The theoretical results in this paper use a theorem due to Dickey that we reproduce below:

**Theorem** *(Dickey, 1968, p.1623, Theorem 2) Let* $\tau_1,\ldots,\tau_K$ *have independent standard* $q_k$*-dimensional multivariate t distributions with* $\nu_1,\ldots,\nu_K > 0$ *degrees of freedom (centers 0, matrices* $\nu_k^{-1} I_{q_k}$*). Then the random r-vector* $\delta$*,*

$$\delta = \sum B_k \tau_k$$

*has the representation,*

$$\delta = (\sum u_k^{-1} \nu_k B_k B_k')^{1/2} z$$

*where the $u_k$ are independently chi-squared distributed with $\nu_k$ degrees of freedom, and $z$ is an independent $r$-dimensional standard normal vector. Consequently, $\delta$ has the further representation,*

$$\delta = (\sum v_k^{-1}(\nu_k/\nu)B_k B_k')^{1/2}\tau$$

*where, with $\nu = \sum \nu_k$, the $v_k = u_k/\sum u_k$, $v_1$,...,$v_K$ are jointly Dirichlet distributed: $v_k > 0$, $\sum v_k = 1$, with density $\Gamma(\nu/2)\prod_k v_k^{\nu_k/2-1}/\Gamma(\nu_k/2)$ in $v_1$,...,$v_{K-1}$, and $\tau$ has an $r$-dimensional standard multivariate $t$ distribution with $\nu$ degrees of freedom. If the matrix $\sum B_k B_k'$ is non singular, the distribution of $\delta$ is non degenerate with the density function*

$$f(\delta) = \Gamma(\frac{\nu+r}{2})\pi^{-r/2}/\prod_k \Gamma(\nu_k/2)\times$$
$$\int_\sigma \prod_k v_k^{\nu_k/2-1} \mid \sum v_k^{-1}\nu_k B_k B_k') \mid^{-1/2} [1+\delta'(\sum v_k^{-1}\nu_k B_k B_k')^{-1}\ \delta]^{-\frac{\nu+r}{2}} dv_1 \cdots dv_{K-1}$$

*the range $\sigma$ of the $v_k$ as above.*

PROOF: see Dickey (1968).

The above Theorem expresses the density of a linear combination of independently distributed multivariate $t$ vectors as an integral of dimension one less than the number of summands. Assuming standardized $t$-distributions for the $\tau_k$ vectors, the density of $\delta$ is obtained as a function of the degrees of freedom $\nu_k$ and of the products $\nu_k B_k B_k'$. As we turn to see, this result can be exploited to simplify the computation of the marginal likelihood of Gaussian STS models with inverted gamma-2 (IG) priors on the variance parameters. We start with simple trend plus noise processes.

## 2.2 First-order random walk trends

We first consider the case of a time series $y_t$ that is made up of a random walk $p_t$ plus a noise $c_t$ like in:

$$\begin{aligned} y_t &= p_t + c_t & \\ \Delta p_t &= a_{pt} & a_{pt}|V_p \sim N(0, V_p) \\ c_t &= a_{ct} & a_{ct}|V_c \sim N(0, V_c) \end{aligned} \tag{2.1}$$

where $\Delta \equiv 1 - L$ and $L$ is the lag operator. Model (2.1) is also known as the local level model (see Durbin and Koopman, 2001, Chap.2). Given their respective variances, the structural shocks $a_{ct}$ and $a_{pt}$ are independent and Normally distributed. The variance parameters $V_c$ and $V_p$ are assumed to be random variables with IG prior distribution:

$$V_\ell \sim IG(s_{\ell 0}, \nu_{\ell 0}) \qquad \ell = c, p. \tag{2.2}$$

We shall denote $x_k^T \equiv (x_k, \cdots, x_T)'$ and in short $x \equiv x_1^T$. IG-priors for variance parameters have been intensively used in time series analysis (see for instance Chib, 1993; Chib and Greenberg, 1994). This assumption implies that the corresponding shocks are marginally distributed according to the Student density:

$$f(a_\ell) = t(0, s_{\ell 0}, I_T, \nu_{\ell 0}) \qquad \ell = c, p$$

where $I_T$ is the $T \times T$ identity matrix. The structural shocks can be expressed as $a_\ell = (s_{\ell 0}/\nu_{\ell 0})^{1/2}\tau_\ell$, where $\tau_\ell$ is the random vector with standard t-distribution:

$$f(\tau_\ell) = t(0, 1, I_T/\nu_{\ell 0}, \nu_{\ell 0}) \qquad \ell = c, p \tag{2.3}$$

Let us define $D_1$ the $T-1 \times T$ first-order difference matrix, i.e. $D_1(i,i) = -1$, $D_1(i, i+1) = 1$, and 0 elsewhere. The stationary transformation of the observed process $y$ verifies:

$$\begin{aligned} D_1 y &= (s_{p0}/\nu_{p0})^{1/2}\tau_p + (s_{c0}/\nu_{c0})^{1/2} D_1 \tau_c \\ &= B_p \tau_p + B_c \tau_c \end{aligned}$$

Since the vectors $\tau_p$ and $\tau_c$ are mutually independent, Dickey's Theorem applies to the local level model with $B_p = (s_{p0}/\nu_{p0})^{1/2} I_{T-1}$ and $B_c = (s_{c0}/\nu_{c0})^{1/2} D_1$. This proves our first result:

**Lemma 1** *The marginal likelihood of the local level model (2.1) with IG-prior (2.2) on the variance parameters is such that:*

$$\begin{aligned} f_D(y) &= \pi^{-\frac{T-1}{2}} \Gamma(\frac{\nu}{2}) \Gamma(\frac{\nu_{p0}}{2})^{-1} \Gamma(\frac{\nu_{c0}}{2})^{-1} \\ &\times \int_0^1 u^{\nu_{p0}/2-1}(1-u)^{\nu_{c0}/2-1} \mid \frac{s_{p0}}{u} M_p + \frac{s_{c0}}{1-u} M_c \mid^{-1/2} \\ &\times [1 + (D_1 y - \mu_y)'(\frac{s_{p0}}{u} M_p + \frac{s_{c0}}{1-u} M_c)^{-1}(D_1 y - \mu_y)]^{-\frac{\nu}{2}} du \end{aligned} \tag{2.4}$$

*with* $\nu = \nu_{p0} + \nu_{c0} + T - 1$, $M_p = I_{T-1}$, $M_c = D_1 D_1'$, *and* $\mu_y = 0$.

Lemma 1 reduces the problem of evaluating the marginal likelihood of the local level model with T observations to a simple one-dimensional integration over the $(0, 1)$ open interval. Notice that for this model, the evaluation of the marginal likelihood does not require any MCMC simulation. To facilitate computations, let $M_c = P_c \Lambda_c P_c'$ be the eigenvalues and eigenvectors factorization of $M_c$. Since $M_p = I_{T-1}$ we can write:

$$(\frac{s_{p0}}{u} M_p + \frac{s_{c0}}{1-u} M_c)^{-1} = P_c \{\frac{s_{p0}}{u} I_{T-1} + \frac{s_{c0}}{1-u} \Lambda_c\}^{-1} P_c' \tag{2.5}$$

where the central term is a diagonal matrix and the eigenvectors do not depend on $u$. Since $|P_c| = 1$, expression (2.5) also simplifies the computation of the determinant in (2.4).

In empirical macroeconomics, a constant slope is often added to the trend (see for instance Stock and Watson, 1988):

$$\Delta p_t = \mu_p + a_{pt} \tag{2.6}$$

Considering the standard assumption that $\mu_p, V_p$ are jointly NIG-distributed like in

$$\mu_p | V_p \sim N(\mu_{p0}, V_p v_{\mu 0}), \tag{2.7}$$

the distribution of the trend growth marginally to the parameters $\mu_p$ and $V_p$ becomes:

$$f(D_1 p) = t(\mu_{p0}, s_{p0}, (I_{T-1} + 1_{T-1} v_{\mu 0})^{-1}, \nu_{p0})$$

where $1_k$ is the $k \times k$ matrix of ones (see Bauwens et al., 1999, p.300, 304). In terms of the standardized $t$-variables defined in (2.3), we have now $\tau_p = (s_{p0}/\nu_{p0})^{-1/2}(I_{T-1} + 1_{T-1} v_{\mu 0})^{-1/2} a_p$. The stationary transformation of the observed series verifies:

$$D_1 y = \mu_{p0} + (s_{p0}/\nu_{p0})^{1/2}(I_{T-1} + 1_{T-1} v_{\mu 0})^{1/2} \tau_p + (s_{c0}/\nu_{c0})^{1/2} D_1 \tau_c$$

Hence Dickey's Theorem applies with $B_p = (s_{p0}/\nu_{p0})^{1/2}(I_{T-1} + 1_{T-1} v_{\mu 0})^{1/2}$ and $B_c = (s_{c0}/\nu_{c0})^{1/2} D_1$. We get:

**Lemma 2** *The marginal likelihood of the random walk with drift plus noise model with prior distributions (2.2) and (2.7) is given by equation (2.4) with* $\nu = \nu_{p0} + \nu_{c0} + T - 1$, $M_p = I_{T-1} + 1_{T-1} v_{\mu 0}$, $M_c = D_1 D_1'$, *and* $\mu_y = \mu_{p0}$.

In this case the simple diagonalization (2.5) cannot be used anymore for speeding up the integration. One must instead resort to the simultaneous diagonalization $Q'M_pQ = I_{T-1}$ and $Q'M_cQ = \Lambda_c$, where $\Lambda_c$ is a diagonal matrix (see Magnus and Neudecker, 1988, p.22). This yields:

$$(\frac{s_{p0}}{u}M_p + \frac{s_{c0}}{1-u}M_c)^{-1} = Q\{\frac{s_{p0}}{u}I_{T-1} + \frac{s_{c0}}{1-u}\Lambda_c\}^{-1}Q'$$

The matrix $Q$ is obtained as $Q = P_p\Lambda_p^{-1/2}P^\star$, where $P_p$ and $\Lambda_p$ are the eigenvectors and eigenvalues matrices of $M_p$, and $P^\star$ is the eigenvector matrix of $(P_p\Lambda_p^{-1/2})'M_cP_p\Lambda_p^{-1/2}$.

## 2.3 Second-order random walk trends

For some macroeconomic variables like unemployment, the hypothesis of constant growth is unrealistic. A possibility is to assume an integrated slope. The trend equation becomes:

$$\begin{aligned} \Delta p_t &= \mu_{t-1} + a_{pt} \qquad && a_{pt}|V_p \sim N(0, V_p) \\ \Delta \mu_t &= a_{\mu t} && a_{\mu t}|V_\mu \sim N(0, V_\mu) \end{aligned} \tag{2.8}$$

Model (2.8) is known as the local linear trend. If $V_p = 0$, it reduces to the I(2) plus noise process that is implicitly considered in Hodrick-Prescott (HP) filtering (see Hodrick and Prescott, 1997, and Harvey and Jaeger, 1993). Equation (2.8) introduces one more latent variable, $\mu_t$, with associated variance parameter $V_\mu$ for which we assume the IG-prior distribution:

$$V_\mu \sim IG(s_{\mu 0}, \nu_{\mu 0}) \tag{2.9}$$

Like previously, the IG-prior hypothesis implies that marginally to the variance parameters, the shocks of the slope equation $a_\mu$ re-scaled as $\tau_\mu = (s_{\mu 0}/\nu_{\mu 0})^{-1/2}a_\mu$ has a standard t-density. Let $D_2$ denote the $T-2 \times T$ second-order difference matrix. In terms of the $\tau$-variables, the measurement equation can be written as:

$$D_2 y = (s_{\mu 0}/\nu_{\mu 0})^{1/2}\tau_\mu + (s_{p0}/\nu_{p0})^{1/2}D_1\tau_p + (s_{c0}/\nu_{c0})^{1/2}D_2\tau_c$$

Clearly, for such models $D_1$ is a $T-2 \times T-1$ matrix. We have now:

**Lemma 3** *The marginal likelihood of the local linear trend (2.1)-(2.8) with IG-priors (2.2) and (2.9) verifies:*

$$f_D(y) = \pi^{-\frac{T-2}{2}}\Gamma(\frac{\nu}{2})\Gamma(\frac{\nu_{\mu 0}}{2})^{-1}\Gamma(\frac{\nu_{p0}}{2})^{-1}\Gamma(\frac{\nu_{c0}}{2})^{-1}$$

$$
\begin{aligned}
&\times \int_0^1 \int_0^1 u_1^{\nu_{\mu 0}/2-1} u_2^{\nu_{p0}/2-1} (1-u_1-u_2)^{\nu_{c0}/2-1} \\
&\times \mid \frac{s_{\mu 0}}{u_1} M_\mu + \frac{s_{p0}}{u_2} M_p + \frac{s_{c0}}{1-u_1-u_2} M_c \mid^{-1/2} \\
&\times [1 + D_2 y' (\frac{s_{\mu 0}}{u_1} M_\mu + \frac{s_{p0}}{u_2} M_p + \frac{s_{c0}}{1-u_1-u_2} M_c)^{-1} D_2 y]^{-\frac{\nu}{2}} du_1 du_2
\end{aligned}
\tag{2.10}
$$

*with $\nu = \nu_{\mu 0} + \nu_{p0} + \nu_{c0} + T - 2$, $M_\mu = I_{T-2}$, $M_p = D_1 D_1'$, and $M_c = D_2 D_2'$. When $V_p = 0$, the marginal likelihood of the I(2)-trend plus noise model reduces to:*

$$
\begin{aligned}
f_D(y) &= \pi^{-\frac{T-2}{2}} \Gamma(\frac{\nu}{2}) \Gamma(\frac{\nu_{\mu 0}}{2})^{-1} \Gamma(\frac{\nu_{c0}}{2})^{-1} \\
&\times \int_0^1 u_1^{\nu_{\mu 0}/2-1} (1-u_1)^{\nu_{c0}/2-1} \mid \frac{s_{\mu 0}}{u_1} M_\mu + \frac{s_{c0}}{1-u_1-u_2} M_c \mid^{-1/2} \\
&\times \quad [1 + D_2 y' (\frac{s_{\mu 0}}{u_1} M_\mu + \frac{s_{c0}}{1-u_1} M_c)^{-1} D_2 y]^{-\frac{\nu}{2}} du_1
\end{aligned}
$$

*with $\nu = \nu_{\mu 0} + \nu_{c0} + T - 2$, $M_\mu = I_{T-2}$, and $M_c = D_2 D_2'$.*

Because simultaneous diagonalizations do not extend to the three-matrix case, the double-integration (2.10) is computationally more demanding than for first-order random walk models. Lemma 3 can be easily extended to the m-th order trend plus noise models discussed by Harvey and Trimbur (2003). We now examine models with conditional mean dynamics.

## 2.4 STS models with conditional mean dynamics

When the structural processes contain some dynamics, Lemmas 1-3 give the likelihood only conditionally on some parameters. For instance, the damped trend model assumes a stationary zero-mean autoregressive slope such as (see Harvey, 1989, p.46):

$$
\mu_t = \phi_\mu \mu_{t-1} + a_{\mu t}
$$

Let $\Sigma_\mu$ denote the variance-covariance matrix of $\mu/V_\mu^{1/2}$, i.e. $\Sigma_\mu = V(\mu_1^{T-1})/V_\mu$. Marginally to $V_\mu$, the distribution of $\mu$ given $\phi_\mu$ is a Student-t density with precision matrix $\Sigma_\mu^{-1}$. In this case:

**Lemma 4** *For the damped trend plus noise model with IG priors (2.2) and (2.9) on variance parameters, the likelihood of $y$ conditional on $\phi_\mu$ is given by equation (2.10) with $M_p = I_{T-1}$, $M_c = D_1 D_1'$, $M_\mu = \Sigma_\mu$, and replacing $D_2 y$ with $D_1 y$.*

More often however it is the short-term component that evolves dynamically. Indeed many macroeconomic time series display recurrent short-term movements, usually related to the business cycle, and for such series the STS model must complement the long-term trend with a cyclical component. The regularity of the cyclical fluctuations can be reproduced using an AR(2) process with complex roots parameterized in terms of amplitude $A$ and periodicity $\varphi$ as in:

$$(1 - 2A\cos\frac{2\pi}{\varphi}L + A^2L^2)c_t = a_{ct} \qquad a_{ct}|V_c \sim N(0, V_c) \tag{2.11}$$

This specification is closely related to the stochastic cycle discussed in Harvey (1989, p.46). The amplitude-periodicity parameterization is appealing as it suits well the prior information available about the business cycle (see Planas, Rossi and Fiorentini, 2008). Notice that our results do not depend on the particular choice of the prior distribution for the conditional mean parameters. For the prior distribution of $V_c$, we keep Assumption (2.2). Let $\Sigma_c$ denote the variance-covariance matrix of $c/V_c^{1/2}$, i.e. $\Sigma_c = V(c)/V_c$. Its inverse gives the precision matrix of the $t$ distribution of $c$ given $A$ and $\varphi$. If the AR(2)-cycle complements a driftless random walk trend, then setting $M_c = D_1\Sigma_c D_1'$ in equation (2.4), Lemma 1, yields the likelihood of $y$ given $A$ and $\varphi$. The same correction applies to Lemma 2 for random walks with drift. If instead the trend is described as a second-order random walk, then it is enough to set $M_c = D_2\Sigma_c D_2'$ in equation (2.10), Lemma 3. As can be seen, the extension of Lemmas 1-4 to models with short-term dynamics is straightforward.

For obtaining the marginal likelihood, it remains to integrate out the conditional mean parameters. Before turning to this problem, we briefly discuss another consequence of the IG-priors on variance parameters in STS models.

## 2.5 The posterior distribution of the latent components

By the Bayes theorem, the posterior distribution of the latent component $c$ verifies:

$$
\begin{aligned}
f(c|y) &= f(y|c)f(c)/f(y) \\
&= f(p|p = y - c)f(c)/f(y)
\end{aligned}
$$

Hence the posterior distribution of the unobservables is given by the product of the prior distributions of $c$ and $p$ evaluated along the direction $p + c = y$ with the marginal likelihood as re-scaling factor. Let us focus for instance on the local level model. Since marginally to the variance parameters, the unobserved variables have a $t$-prior distribution, we have:

**Corollary 1** *The posterior distribution of the increments $D_1 c$ in the local level model (2.1)-(2.2) is the poly-t 2-0 density:*

$$f(D_1 c|y) \propto t(D_1 y, s_{p0}, I_{T-1}, \nu_{p0}) \times t(0, s_{c0}, \{D_1 D_1'\}^{-1}, \nu_{c0})$$

In the expression above, the first $t$-kernel corresponds to the prior distribution of $D_1 p$ evaluated at $D_1 p = D_1 y - D_1 c$ while the second term is the prior distribution of $D_1 c$. Given Corollary 1, the complete posterior distribution of the latent variables $c$ and $p$ can be easily retrieved. Since $f(c|y) = f(c_1|D_1 c, y) \times f(D_1 c|y)$, it can be seen that given the increments $D_1 c$, the data do not bring further information about the starting point $c_1$, i.e. $f(c_1|D_1 c, y) = f(c_1|D_1 c)$. Given the $t$ marginal prior for $c$, the factor $f(c_1|D_1 c)$ is a Student density. Multiplying it by the distribution in Corollary 1 yields the kernel of the posterior distribution of the unobserved components. A constant drift in the trend can be handled similarly.

This makes possible the use of Richard and Tompa's (1980) results to draw posterior samples of the unobservables marginally to the variance parameters in two steps: first the increments, for instance following Appendix B.4.6 in Bauwens et al. (1999, p.321) for sampling from poly-$t$ 2-0 densities, and then the starting point given the increments. Program simplicity would be the main appeal: neither diffuse Kalman Filter initialization (see deJong, 1991) nor smoothing algorithm is needed. The cost however would be a substantial computing time delay due to the resorting to matrix computations. If the sampling of the state variable is inserted into a MCMC scheme, for instance when

Corollary 1 holds conditionally on any other random quantity, such a delay can become prohibitive. For these cases, a recursive scheme such as the Carter and Kohn (1994) state-sampler remains preferable. At least so long as a procedure for factorizing poly-$t$ 2-0 densities is not available.

Corollary 1 does not extend to second-order random walk trends. In this case the posterior distribution of the components increments is an unknown form; it is a poly-$t$ 2-0 density only when conditioning on one unobservable, for instance the slope. This is one reason why we prefer to use Dickey's results instead of Richard and Tompa (1980)'s work about poly-$t$ densities.

# 3 Dickey-Laplace approximation to the marginal likelihood

Let $\Lambda$ denote the set of conditional mean parameters, for instance $\Lambda = (A, \varphi)$ or $(A, \varphi, \phi_\mu)$. For integrating $\Lambda$ out we bring into play the Laplace method (see Tierney and Kadane, 1986). We describe below the details of the Dickey-Laplace combination.

## 3.1 Dickey-Laplace

The Laplace method has been used in the STS context for instance by Harvey et al. (2007). It solves the marginal likelihood integral in the neighborhood of the posterior mode using a normal estimates of the posterior density: the more precise the normal approximation, the better the marginal likelihood evaluation. The strategy we put forward here aims at improving the normal approximation by integrating out the variance parameters using Section 2 results. Besides the benefit arising from the reduced number of parameters, an improvement is expected because variances are typically the main responsible for the posteriors' departure from normality (see for instance Figure 2 in Harvey et al., 2007). Of course, given the asymptotics at work, the smaller the sample size the larger should be the gain in accuracy.

A requirement is that posterior samples of model parameters are available. They can be obtained following the MCMC scheme proposed for instance in Planas et al. (2008). Let $\tilde{\Lambda}$ denote the posterior mode of the parameters $\Lambda$ and $\Sigma(\tilde{\Lambda})$ represent minus the inverted Hessian matrix of the logarithm of the non-normalized marginal posterior

$f_D(y|\Lambda)f(\Lambda)$ evaluated at $\tilde{\Lambda}$:

$$\Sigma(\tilde{\Lambda}) = -[\frac{\partial^2 \ log\{f_D(y|\Lambda)f(\Lambda)\}}{\partial\Lambda'\partial\Lambda} \ |_{\Lambda=\tilde{\Lambda}}]^{-1}$$

The second-order expansion of the non-normalized log-posterior around its mode is such that:

$$log \ \{f_D(y|\Lambda)f(\Lambda)\} \simeq log \ \{f_D(y|\tilde{\Lambda})f(\tilde{\Lambda})\} - \frac{1}{2}(\Lambda - \tilde{\Lambda})\Sigma(\tilde{\Lambda})^{-1}(\Lambda - \tilde{\Lambda})'$$

The last term above takes the form of the kernel of a normal distribution with mean $\tilde{\Lambda}$ and variance-covariance matrix $\Sigma(\tilde{\Lambda})$. Exponentiating and integrating out $\Lambda$ yields:

$$f_{DL}(y) = (2\pi)^{d/2}|\Sigma(\tilde{\Lambda})|^{1/2} f_D(y|\tilde{\Lambda})f(\tilde{\Lambda}) \tag{3.1}$$

where $d$ is the dimension of $\Lambda$. We shall refer to equation (3.1) as the Dickey-Laplace marginal likelihood estimates. In (3.1), the term $f(\tilde{\Lambda})$ assigns a prior weight to the posterior mode while $f_D(y|\tilde{\Lambda})$ is the model likelihood marginally to the IG-variance parameters given in Lemmas 1-4.

## 3.2 Comparison with Laplace and bridge sampling

We evaluate the Laplace (LP) and Dickey-Laplace (D-LP) marginal likelihood estimators in a simulation exercise using the Meng-Wong (MW, 1996) bridge sampler as benchmark. During the last two decades, econometricians have most often resorted to importance sampling for computing marginal likelihoods (see Kloek and Van Dijck, 1978; Geweke, 1989). MW's technique is an extension that re-weights both the importance function and the posterior density through a bridge function. We adopt this method as benchmark as a consensus is emerging about its potential superiority over the other estimators available (see Meng and Schilling, 1996; diCiccio et al., 1997; and Fruhwirth-Schnatter, 2004).

Let $S$ and $S_q$ denote the support of the parameter posterior distribution and of an importance function, say $q(\theta)$, where $\theta$ represents the model parameters including variances. Let also $h(\theta)$ represent a function defined over $S \bigcap S_q$. The MW marginal likelihood estimate is obtained from (see also Gelman and Wong, 1998):

$$f(y) = \frac{\int_{S_q} \frac{h(\theta)}{q(\theta)} dq(\theta)}{\int_S \frac{h(\theta)}{f(y|\theta)f(\theta)} df(\theta|y)}$$

Equivalent formulations are sometimes given in terms of a function $\gamma(\theta)$ such that $\gamma(\theta)q(\theta)f(y|\theta)f(\theta) = h(\theta)$ (see Fruhwirth-Schnatter, 2004). The bridge function $h(\theta)$ reduces the estimation error when located at an intermediate position between the importance function and the parameter posterior distribution. MW propose as optimal choice a recursive procedure based on:

$$h(\theta) \propto \frac{q(\theta)f(\theta|y)}{n_q q(\theta) + n_y f(\theta|y)}$$

where the constants $n_q$ and $n_y$ refer to the number of draws from the importance function and from the posterior density, respectively. The recursions are introduced through the term $f(\theta|y)$ that involves a preliminary marginal likelihood estimate. We initialize the algorithm using the standard Laplace approximation and then iterate for ten rounds; no further sampling is needed for iterating. The MW estimator can also be built around likelihood functions marginal to the variance coefficients, i.e. using $f_D(y|\Lambda)$ in place of $f(y|\theta)$. Because Dickey's integral would need to be evaluated for every sample out of the importance function, we discard this possibility for its computational cost.

We simulate three series of respective length $T = 25, 100, 250$ from a random walk with drift plus AR(2) cycle like in (2.6) and (2.11). The coefficients are set to $\mu_p = .1$, $A = .8$, $\varphi = 10$, $V_p = .01$ and $V_c = .05$. The marginal likelihoods of the simulated series are estimated using the LP, D-LP and MW methods for eight models obtained as combinations of four trend models, i.e. I(2), integrated random walk (irw), random walk plus drift (rw), and damped trend (dt), with two models for the cycle, i.e. the white noise (wn) and the autoregressive model (ar2) in (2.11). The prior distributions are omitted for the sake of space. For each model, we record two thousand samples from the parameters posterior distribution out of two hundred thousand simulations using the Gibbs sampling scheme detailed in Planas et al. (2008), after a burn-in of ten thousand iterations. The sparse recording serves at lowering chain autocorrelations. This MCMC output is then used to compute the marginal likelihoods, and the whole computations are repeated twenty times in order to get numerical averages and standard deviations. Lemmas 1-4 integrals are calculated over grids of one thousand points in dimension one, and over squares of four hundred points in each side in dimension two; their standard deviations are neglected as of irrelevant size. Notice that when the STS model includes some autoregressive dynamics, the normalizing constants of the full conditional distributions are not entirely known so Chib's (1995) marginal likelihood estimator does not apply.

Table 1 displays the results. The average marginal likelihoods are reported with a minus sign and the numerical standard deviations lie between brackets. The models are ranked according to the number of parameters, from the I(2) plus noise model with 2 parameters to the damped trend plus AR(2) cycle with 6 parameters. For models with only variance coefficients, MW and Lemmas 1-3 based estimates are in close agreement: the differences are of the third digit order, whatever the sample size. When dynamic parameters are introduced, the deviations get to the first digit order. The error in the LP estimates can instead reach a unit, especially in short sample. As can be seen, marginalizing out the variance parameters always improves the approximation. For the models and sample sizes considered, the improvement is such as to make the D-LP estimate almost as accurate as the MW one. This result is interesting because no further sampling from an importance function is needed with the D-LP approach.

All marginal likelihood estimates point to the random walk with drift plus AR(2) cycle as the most adequate model. Mispecifying the short-term dynamics implies quite a large drop in the marginal likelihood. It could be argued that the Laplace estimator remains useful for model discrimination in spite of the approximation errors, but such a conclusion depends on the discrepancies between the alternatives considered. We shall see in the next Section that discriminating between models with similar properties can become difficult with the Laplace estimator. Moreover, when the model misses some important pattern such as a cycle, the mispecification can yield posterior distributions with bi-modal characteristics. In such cases the LP marginal likelihood is unreliable. For this experiment we mitigated this problem by carefully tuning the prior distributions and, in a few cases, by trimming the output.

**Table 1 Minus average log marginal likelihood**

| Trend | Cycle | MW | | LP | | D-LP | |
|---|---|---|---|---|---|---|---|
| | | **T = 25** | | | | | |
| **i2** | **wn** | 23.772 | [ .015 ] | 23.143 | [ .092 ] | 23.765 | [ ⋆⋆ ] |
| **irw** | **wn** | 23.200 | [ .022 ] | 20.925 | [ .101 ] | 23.202 | [ ⋆⋆ ] |
| **rw** | **wn** | 21.378 | [ .021 ] | 19.510 | [ .169 ] | 21.378 | [ ⋆⋆ ] |
| **dt** | **wn** | 21.368 | [ .023 ] | 19.484 | [ .123 ] | 21.504 | [ .088 ] |
| **i2** | **ar2** | 11.012 | [ .025 ] | 9.693 | [ .106 ] | 10.286 | [ .084 ] |
| **irw** | **ar2** | 11.230 | [ .027 ] | 9.113 | [ .138 ] | 10.532 | [ .084 ] |
| **rw** | **ar2** | 9.075 | [ .026 ] | 7.406 | [ .179 ] | 8.383 | [ .084 ] |
| **dt** | **ar2** | 9.700 | [ .028 ] | 7.915 | [ .115 ] | 9.170 | [ .090 ] |
| | | **T = 100** | | | | | |
| | | **MW** | | **LP** | | **D-LP** | |
| **i2** | **wn** | 77.607 | [ .011 ] | 77.512 | [ .079 ] | 77.606 | [ ⋆⋆ ] |
| **irw** | **wn** | 68.518 | [ .014 ] | 68.009 | [ .298 ] | 68.518 | [ ⋆⋆ ] |
| **rw** | **wn** | 64.946 | [ .011 ] | 64.840 | [ .058 ] | 64.946 | [ ⋆⋆ ] |
| **dt** | **wn** | 66.744 | [ .080 ] | 64.191 | [ .311 ] | 66.930 | [ .140 ] |
| **i2** | **ar2** | 21.893 | [ .016 ] | 21.428 | [ .116 ] | 21.707 | [ .080 ] |
| **irw** | **ar2** | 21.702 | [ .021 ] | 20.872 | [ .090 ] | 21.551 | [ .090 ] |
| **rw** | **ar2** | 15.627 | [ .021 ] | 14.928 | [ .148 ] | 15.492 | [ .076 ] |
| **dt** | **ar2** | 24.612 | [ .021 ] | 23.270 | [ .176 ] | 24.230 | [ .132 ] |
| | | **T = 250** | | | | | |
| | | **MW** | | **LP** | | **D-LP** | |
| **i2** | **wn** | 141.749 | [ .009 ] | 141.763 | [ .082 ] | 141.746 | [ ⋆⋆ ] |
| **irw** | **wn** | 124.670 | [ .013 ] | 124.444 | [ .077 ] | 124.697 | [ ⋆⋆ ] |
| **rw** | **wn** | 114.732 | [ .011 ] | 114.751 | [ .109 ] | 114.730 | [ ⋆⋆ ] |
| **dt** | **wn** | 114.727 | [ .018 ] | 114.348 | [ .119 ] | 114.717 | [ .034 ] |
| **i2** | **ar2** | 54.653 | [ .015 ] | 54.345 | [ .131 ] | 54.468 | [ .080 ] |
| **irw** | **ar2** | 54.817 | [ .018 ] | 54.192 | [ .158 ] | 54.720 | [ .125 ] |
| **rw** | **ar2** | 37.354 | [ .016 ] | 36.860 | [ .100 ] | 37.251 | [ .084 ] |
| **dt** | **ar2** | 57.289 | [ .019 ] | 56.439 | [ .174 ] | 57.065 | [ .091 ] |

Notes: **MW** Meng-Wong; **LP** Laplace; **D-LP** Dickey-Laplace. Models: **rw** random walk; **irw** second-order random walk; **i2** I(2); **dt** damped trend; **wn** white noise; **ar2** AR(2). Standard deviations between brackets, ⋆⋆ means that the Laplace approximation is not needed.

# 4 The US and the euro area NAIRU

We apply this methodology to the analysis of the NAIRU in the EA and in the US. The NAIRU is of particular interest as it is related to the imperfect equilibrium of the labor market. Characterizing the NAIRU is however difficult, mainly because of its unobserved and changing nature. Staiger, Stock and Watson (1997) underlined the lack of precision of estimates obtained with standard specifications. Also, although its time-varying behavior is now well-accepted, not much is known about its actual variability. For instance, the widely-used HP filter requires a prior hypothesis about the signal to noise ratio, but this hypothesis is rarely confronted to the data. Here we take advantage of the Bayesian framework to address the following questions: which STS models best describe the EA and the US NAIRU? How smooth are these NAIRU? And how precise can be their univariate STS estimates?

The EA unemployment series has been collected from AMECO, the national accounts database of the EC Directorate General Economic and Financial Affairs, available at $europa.eu.int/comm/economy\_finance$. The US data have been downloaded from the Bureau of Labor Statistics web-site $www.bls.gov$. Both series are annual averages over 1960-2007, the last figure being preliminary. Following standard practice in the NAIRU literature, we describe these two series as made up of a trend plus a cycle. The cyclical dynamics are represented with an AR(2) process parameterized as in (2.11), with amplitude and periodicity parameters assumed to be Beta-distributed. The prior distribution of the former is tuned so as to yield an average amplitude of 0.8 for the EA and of 0.7 for the US, in agreement with previous empirical business cycle studies (see for instance Kuttner, 1994, Gerlach and Smets, 1999). The standard deviations are set to one-tenth of the mean so as to not impose too much precision. Namely, we use a Beta$(19.2, 4.8)$ for the EA cycle amplitude and a Beta$(29.3, 12.6)$ for the US one. The periodicity parameter is also assumed to be Beta-distributed, with support translated to $[2, T]$ where $T = 48$ is the sample length. Still according to business cycle studies, we tune the periodicity prior distribution so as to get cycles of mean length 9 years for the EA and 8 years for the US, with standard deviations of 2 and 1.5 years, respectively. Formally, we assume $(\varphi - 2)/(48 - 2) \sim$ Beta$(10.2, 57)$ for the EA and $(\varphi - 2)/(48 - 2) \sim$ Beta$(13.8, 91.9)$ for the US. Finally, the IG-distribution for the short-term shocks variance has been set so as to add a mean deviation of 0.5 for euro and 0.7 for US, with the distributions $IG(1.9, 9.6)$ and $IG(2.6, 7.1)$ respectively.

For describing the NAIRU we consider the four specifications discussed in Section 2, namely the driftless random walk, the I(2), the second-order random walk and the damped trend. Three different prior distributions are used for the variance parameters $V_p$ and $V_\mu$: IG(.08,6), IG(.28,6) and IG(.80,6). The expected values of the variances implied by these priors are 0.02, 0.07, and 0.20, respectively; we shall refer to these priors as low (L), medium (M) and high (H). As can be seen in Figure 1, they cover quite a wide range of patterns. For its empirical relevance we also consider the damped trend model without shocks on the trend level, i.e. $V_p = 0$. Finally, the prior for the slope autoregressive parameter $\phi_\mu$ has been set to the Normal distribution $N(.85, 1/30)I_{(0,1)}$ truncated to the stationary positive region. Altogether, the combination of the trend model specifications with the different variance priors yields twenty seven models for the NAIRU.

Table 2 reports the posterior probabilities of each model and ranks the models accordingly. These posterior probabilities are defined as $p(M_i|y) = f(y|M_i)p(M_i)/\sum_k f(y|M_k)$ $\times p(M_k)$, where a discrete uniform prior on the models is assumed. The densities $f(y|M_k)$ have been evaluated using the Dickey-Laplace approximation to the marginal likelihood discussed in Section 3. For both EA and US series, the first five models receive a total posterior weight greater than 50%. Of these best fitting models, all but one are integrated of order 1: the data strongly support the I(1) hypothesis. With all I(2) models ranked last, the evidence is particularly striking for the US. This result can be related to the failure of I(2) models to produce reasonable long-term forecasts of unemployment rate series. The data also express an overwhelming preference for the damped trend model, i.e. the model that accounts for a time-varying slope with moderate persistence. Finally, the EA NAIRU seems to have received larger shocks on its slope than on its level, perhaps explaining why for the EA some I(2) models receive a relevant posterior weight. On the contrary, the shocks on US structural unemployment seem to have hit mostly its level, an observation that pleads against the use of the HP filter for estimating the US NAIRU.

**Table 2 Posterior model probabilities**

| | EA | | | US | | |
|---|---|---|---|---|---|---|
| Rank | Trend | $V_p$-$V_\mu$ | $p(M_i\|y)$ | Trend | $V_p$-$V_\mu$ | $p(M_i\|y)$ |
| 1 | **dt** | **0-M** | .126 | **rw** | **H-0** | .141 |
| 2 | **dt** | **0-L** | .114 | **dt** | **H-L** | .113 |
| 3 | **dt** | **L-L** | .108 | **dt** | **M-L** | .098 |
| 4 | **dt** | **L-M** | .102 | **dt** | **L-L** | .085 |
| 5 | **i2** | **0-L** | .089 | **dt** | **0-L** | .081 |
| 6 | **irw** | **L-L** | .081 | **rw** | **M-0** | .065 |
| 7 | **dt** | **M-L** | .058 | **dt** | **0-M** | .064 |
| 8 | **dt** | **M-M** | .052 | **dt** | **M-M** | .062 |
| 9 | **irw** | **M-L** | .047 | **dt** | **H-M** | .061 |
| 10 | **dt** | **0-H** | .043 | **dt** | **L-M** | .055 |
| 11 | **i2** | **0-M** | .034 | **dt** | **L-H** | .027 |
| 12 | **irw** | **L-M** | .033 | **dt** | **M-H** | .024 |
| 13 | **dt** | **L-H** | .029 | **dt** | **0-H** | .024 |
| 14 | **irw** | **M-M** | .017 | **rw** | **L-0** | .019 |
| 15 | **dt** | **M-H** | .014 | **dt** | **H-H** | .019 |
| 16 | **dt** | **H-L** | .014 | **irw** | **H-L** | .012 |
| 17 | **irw** | **H-L** | .010 | **irw** | **M-L** | .011 |
| 18 | **dt** | **H-M** | .010 | **irw** | **L-L** | .011 |
| 19 | **i2** | **0-H** | .006 | **i2** | **0-L** | .010 |
| 20 | **irw** | **L-H** | .005 | **irw** | **M-M** | .004 |
| 21 | **irw** | **H-M** | .003 | **irw** | **H-M** | .004 |
| 22 | **dt** | **H-H** | .002 | **irw** | **L-M** | .003 |
| 23 | **irw** | **M-H** | .002 | **i2** | **0-M** | .003 |
| 24 | **irw** | **H-H** | .000 | **irw** | **M-H** | .001 |
| 25 | **rw** | **H-0** | .000 | **irw** | **L-H** | .001 |
| 26 | **rw** | **M-0** | .000 | **irw** | **H-H** | .001 |
| 27 | **rw** | **L-0** | .000 | **i2** | **0-H** | .001 |

Notes: **rw** random walk; **irw** second-order random walk; **i2** I(2); **dt** damped trend. Priors for variance parameters: for $\ell = p, \mu$, **L** $V_\ell \sim IG(.08, 6)$; **M** $V_\ell \sim IG(.28, 6)$; **H** $V_\ell \sim IG(.80, 6)$; **0** $V_\ell = 0$.

Given Section 3 results, the robustness of the model classification to the marginal likelihood estimator is worth verifying. For the best five models, Table 3 shows the posterior weights and the ranking obtained with the MW and LP estimators. The D-LP results are reported for comparison. As can be seen, D-LP and MW are in very close agreement: the discrepancies between the posterior weights are always less than 10% of the estimates and the classification differs in only one occasion, two successive models being permuted. The LP outcome is instead quite different: the posterior weights show variations that can reach 100% of the estimates and the ranking is upset. Hence for these models we consider the LP marginal likelihood as unreliable. Probably because 48 observations are not enough to make the posterior distribution of the variance parameters close to the Normal density.

**Table 3 Classification robustness**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | **EA** | | | |
| Trend | $V_p$-$V_\mu$ | **D-LP** | | **MW** | | **LP** | |
| | | Rank | $P(M_i\|y)$ | Rank | $P(M_i\|y)$ | Rank | $P(M_i\|y)$ |
| **dt** | **0-M** | 1 | .126 | 1 | .132 | 5 | .088 |
| **dt** | **0-L** | 2 | .114 | 2 | .107 | 3 | .112 |
| **dt** | **L-L** | 3 | .108 | 4 | .097 | 1 | .167 |
| **dt** | **L-M** | 4 | .102 | 3 | .105 | 2 | .121 |
| **i2** | **0-L** | 5 | .089 | 5 | .088 | 6 | .063 |
| | | | | **US** | | | |
| Trend | $V_p$-$V_\mu$ | **D-LP** | | **MW** | | **LP** | |
| | | Rank | $P(M_i\|y)$ | Rank | $P(M_i\|y)$ | Rank | $P(M_i\|y)$ |
| **rw** | **H-0** | 1 | .141 | 1 | .138 | 6 | .062 |
| **dt** | **H-L** | 2 | .113 | 2 | .106 | 3 | .110 |
| **dt** | **M-L** | 3 | .098 | 3 | .101 | 4 | .115 |
| **dt** | **L-L** | 4 | .085 | 4 | .088 | 1 | .200 |
| **dt** | **0-L** | 5 | .081 | 5 | .076 | 7 | .061 |

Notes: **D-LP** Dickey-Laplace; **LP** Laplace; **MW** Meng-Wong. Models: **rw** random walk; **i2** I(2); **dt** damped trend. Priors for variance parameters: for $\ell = p, \mu$, **L** $V_\ell \sim IG(.08, 6)$; **M** $V_\ell \sim IG(.28, 6)$; **H** $V_\ell \sim IG(.80, 6)$; **0** $V_\ell = 0$.

The best fitting model is, for the EA, the damped trend with no level shocks and for the US, the random walk without drift. This last has been frequently used in empirical studies of the US NAIRU, for instance by Staiger, Stock and Watson (1997) and by Gordon (1998). Figure 2 shows the corresponding estimates. As can be seen, the EA NAIRU is continuously decreasing since the mid-1990's peak. It is tempting to see here the effect of new regulations for increasing the flexibility of EA labor markets, following the European Employment Strategy (1997) within the Lisbon agenda. The US NAIRU seems to be almost constant in the last ten years, after fifteen years of steady decrease between 1982 and 1997. As expected given the relative labor markets flexibility, it embodies more short-term dynamics than the EA one.

In order to analyze the NAIRU smoothness, we compare the ratios between the variance of the cycle and the variance of the trend second difference, i.e. $V(c_t)/V(\Delta^2 p_t)$. This quantity is a slight generalization of the inverse signal to noise ratio typically considered in HP filtering. More elaborate measures of smoothness have been proposed in the literature (see for instance Froeba and Koyak, 1994), but the acquaintance of economists with the HP filter gives such variance ratios the advantage of immediacy. Figure 3 shows the posterior distribution of the generalized inverse signal to noise ratio; the continuous line refers to the best model and the dashed one is obtained marginally to the model choice. As can be seen, the US NAIRU participates more to the unemployment fluctuations than the EA one, and this evidence is strong enough to hold marginally to the model choice: the variance ratio mode is about 2 for the US against 8.0 for the EA with the best model, 5.0 after model marginalizing. These results are in broad agreement with the findings by Ravn and Uhlig (2002) about annual data. Notice that the posterior distribution of the variance ratio is quite diffuse for the EA, perhaps reflecting a substantial time-varying behavior.

Finally, Figure 4 shows the posterior distribution of the 2007 NAIRU for the EA and for the US. Again, the continuous line refers to the best model and the dashed one corresponds to the model average. The current NAIRU is measured about 7.5 for the EA and about 5-5.1 for the US. This result is obtained with both the best model and marginally to the model specification. A 95% confidence band around the modes covers about 2.6 points, with the interval (6.2,8.9) for the EA and (3.7,6.3) for the US. This is comparable with the uncertainty that Staiger et al. (1997) reported for the US NAIRU in 1991's first quarter using also inflation data. There is a close matching between the posterior distribution obtained with the best model and the one obtained marginally to the model specification, mainly because the models that receive the highest posterior weights yield similar NAIRU estimates. Hence, as long as a reasonable model is used like for instance Table 2's first five, researchers should not worry too much about model uncertainty.

# 5 Conclusion

We obtain simple expressions for evaluating the marginal likelihood of Gaussian STS models by taking benefit of the model properties and of a theorem by Dickey (1968). For trend plus noise models, they only require an integration over a finite support. For trend plus cycle models, we show that combining this approach with the Laplace method yields a substantial gain in accuracy with respect to the traditional Laplace marginal likelihood estimator. Overall the precision is comparable to that of the MW estimator, without requiring any importance sampling.

We apply this methodology to the analysis of the EA and the US NAIRU. As best models, we found the damped trend for the EA and the driftless random walk for the US; these would be our recommendation to practitioners. The NAIRU smoothness seems compatible with the inverse signal to noise ratio suggested by Ravhn and Uhlig (2002) for HP-detrending annual data. Model uncertainty does not seem to add much variation to the NAIRU estimates, at least so long as a reasonable model is used. Although this specification has been somewhat overlooked in empirical macroeconomics, the damped trend dominates the other models for both EA and US unemployment series. We could see that conducting this analysis with the traditional version of the Laplace marginal likelihood gives misleading results, perhaps because of the limited sample size. The methodology we propose can be extended to STS models including a third unobserved variable such as the irregular component, and can also be adapted to bivariate system such as the Kuttner (1994) Phillips-curve augmented model for output gap.

# References


Ball L. and Mankiw G.N. (2002), "The NAIRU in theory and practice", *Journal of Economic Perspectives*, 16, 4, 115-136.

Bauwens L., Lubrano M. and Richard J. (1999), *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press.

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

Chib S. (1993), "Bayesian regression with autoregressive errors. A Gibbs sampling approach", *Journal of Econometrics*, 58, 275-94.

Chib S. (1995), "Marginal likelihood from the Gibbs output", *Journal of the American Statistical Association*, 90, 432, 1313-1321.

Chib S. and Greenberg E. (1994), "Bayesian inference in regression models with ARMA(p,q) errors", *Journal of Econometrics*, 64, 183-206.

Clark P. (1987), "The cyclical comovement of U.S. economic activity", *Quarterly Journal of Economics*, 797-814.

Cogley T. and Sargent T.J (2005), "Drifts and volatilities: monetary policies and outcomes in the post war US", *Review of Economic Dynamics*, 8, 262-302.

de Jong P. (1991), "The diffuse Kalman filter", *Annals of Statistics*, 2, 1073-83.

Denis C., Grenouilleau D., Mc Morrow K. and Roeger W. (2006), "Calculating potential growth rates and output gaps - A revised production function approach", European Economy, Economic Papers nr. 246, European Commission.

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*, 91, 435, 1254-1267.

Dickey J. (1968), "Three multidimensional-integral identities with Bayesian applications", *Annals of Mathematical Statistics*, 39, 5, 1615-1637.

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

Froeba L. and Koyak R. (1994), "Measuring and comparing smoothness in time series: the production smoothing hypothesis", *Journal of Econometrics*, 64, 1-2, 97-122.

Fruhwirth-Schnatter S. (1994), "Data augmentation and dynamic linear models", *Journal of Time Series Analysis*, 15, 2, 183-202.

Fruhwirth-Schnatter S. (2004), "Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques", *Econometrics Journal*, 7, 143-167.

Gelman A. and Wong X.L. (1998), "Simulating normalizing constants: from importance sampling to bridge sampling to path sampling", *Statistical Science*, 13, 2, 163-185.

Gerlach S. and Smets F. (1999), "Output gap and monetary policy in the EMU area", *European Economic Review*, 43, 801-812.

Geweke J. (1989), "Bayesian inference in econometric models using Monte-Carlo integration", *Econometrica*, 57, 1317-1339.

Hall R.E. and Mishkin F.S. (1982), "The sensitivity of consumption to transitory income: estimates from panel data on households", *Econometrica*, 50, 2, 461-481.

Hansen G. (1997), "Technical progress and aggregate fluctuations", *Journal of Economic Dynamics and Control*, 21, 6, 1005-1023.

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

Harvey, A.C. (2001), "Testing in unobserved components models", *Journal of Forecasting*, 20, 1-19.

Harvey, A.C. and Jaeger, A. (1993), "Detrending, stylised facts and the business cycles", *Journal of Applied Econometrics*, 8, 231-247.

Harvey, A.C. and Trimbur T.M. (2003), "General model-based filters for extracting cycles and trends in economic time series", *Review of Economics and Statistics*, 85, 2, 244-255.

Harvey A.C., Trimbur T.M. and van Dijk H.K. (2007), "Trends and cycles in economic time series: a Bayesian approach", *Journal of Econometrics*, 140, 2, 618-649.

Hodrick R.J. and E.C. Prescott (1997), "Post-war U.S. business cycle: an empirical investigation", *Journal of Money, Credit and Banking*, 29, 1, 1-16.

Kloek T. and Van Dijck H.K (1978), "Bayesian estimates of equation system parameters: an application of integration by Monte Carlo", *Econometrica*, 46, 1-19.

Koop G. (2003), *Bayesian Econometrics*. Chichester: Wiley.

Kuttner K. (1994), "Estimating potential output as a latent variable", *Journal of Business and Economic Statistics*, 12, 3, 361-368.

Magnus J.R. and Neudecker H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley.

Meng X.L. and Schilling S. (1996), "Fitting full-information item factor models and an empirical investigation of bridge sampling", *Journal of the American Statistical Association*, 92, 439, 903-915.

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.

Pagan A. (1975), "A note on the extraction of components from time series", *Econometrica*, 43, 1, 163-183.

Planas C., Roeger W. and Rossi A. (2007), "How much has labour taxation contributed to European structural unemployment?", *Journal of Economic Dynamics and Control*, 31, 1359-1375.

Planas C., Rossi A. and Fiorentini G. (2008), "Bayesian analysis of the output gap", *Journal of Business & Economic Statistics*, 26, 1, 18-32.

Richard J.F. and Tompa H. (1980), "On the evaluation of poly-t density functions", *Journal of Econometrics*, 12, 335-351.

Staiger D., Stock J.H. and Watson M.W. (1997), "How precise are estimates of the natural rate of unemployment?", in *Reducing Inflation: Motivation and Strategy*, by C.D. Romer and D.H. Romer, eds. Chicago: University of Chicago Press, 195-246.

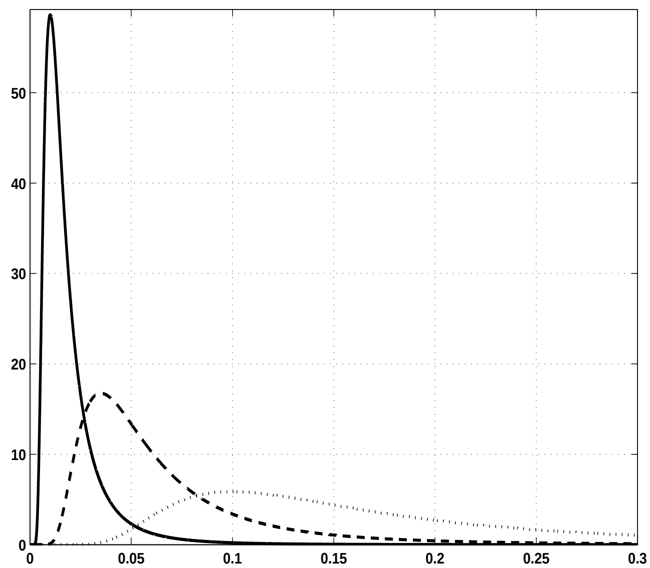
Stiglitz, J. (1997), "Reflections on the natural rate hypothesis", *Journal of Economic Perspectives*, 11, 1, 3-10.

Stock J.H. and Watson M.W. (1988), "Variables trends in economic time series", *Journal of Economic Perspectives*, 2, 147-174.

Stock J.H. and Watson M.W. (2007), "Why has US inflation become harder to forecast?", *Journal of Money, Credit and Banking*, 39, 1, 3-33.

Tierney L. and Kadane J.B. (1986), "Accurate approximations for posterior moments and marginal densities", *Journal of the American Statistical Association*, 81, 393, 82-86.

**Figure 1**
**Prior distributions for $V_p$ and $V_\mu$**




— L: IG(.08,6)  – – M: IG(.28,6)  ··· H: IG(.8,6)

**Figure 2**
**NAIRU (—) and 95% confidence bands (- -)**
**Best model**

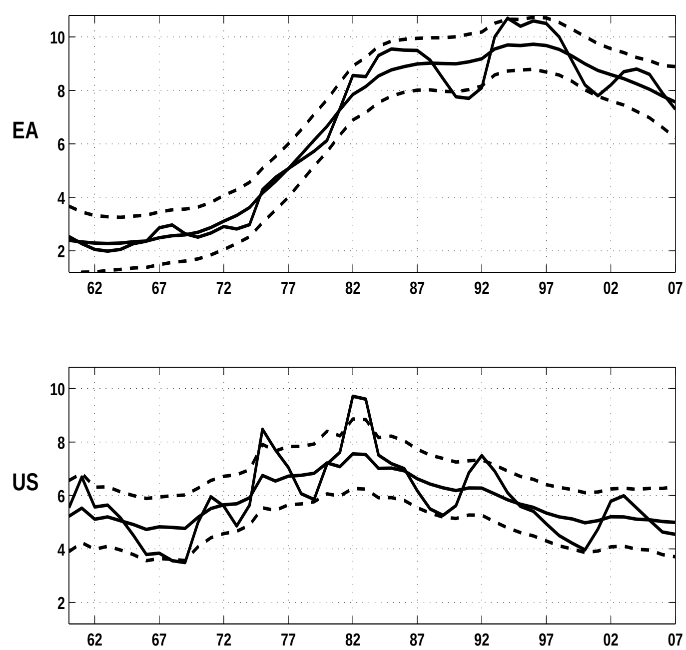
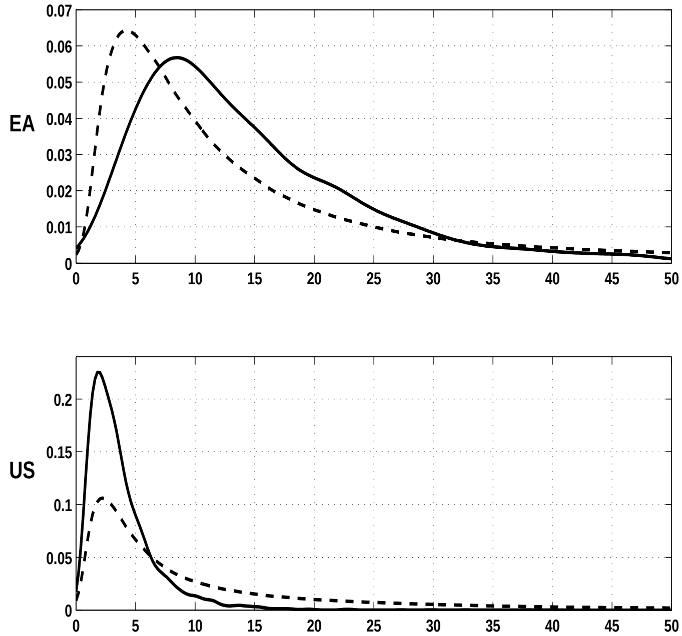


**Figure 3**

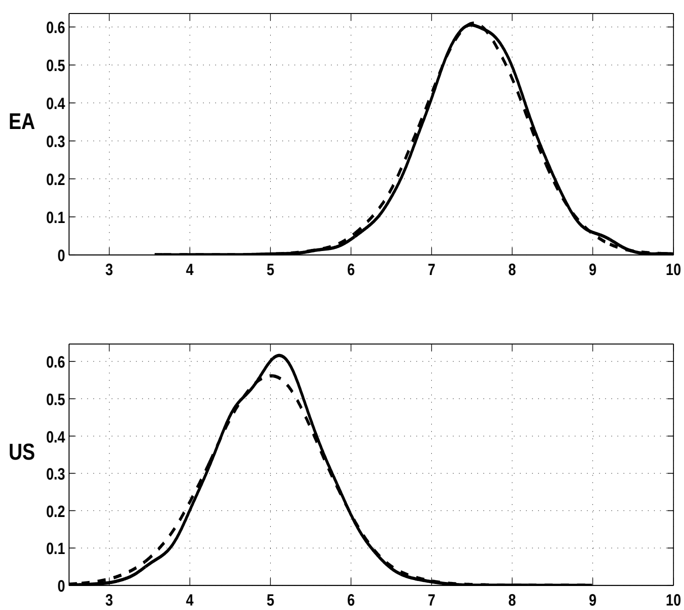
**Variance ratio $V(c)/V(\Delta^2 p)$ posterior distribution**




—— Best model – – Model average

**Figure 4**

**Posterior distribution of 2007 NAIRU**




—— Best model – – Model average