Extracting the Cyclical Component in Hours Worked: a Bayesian Approach

Bernardi, Mauro and Della Corte, Giuseppe and Proietti, Tommaso

University of Rome, Tor Vergata

May 2008

Extracting the Cyclical Component in Hours Worked: a Bayesian Approach

Mauro Bernardi **Giuseppe della Corte**
Tommaso Proietti
Dipartimento S.E.F. e ME.Q., Via Columbia 2, 00133 Roma, Italy.

Abstract

The series on average hours worked in the manufacturing sector is a key leading indicator of the U.S. business cycle. The paper deals with robust estimation of the cyclical component for the seasonally adjusted time series. This is achieved by an unobserved components model featuring an irregular component that is represented by a Gaussian mixture with two components. The mixture aims at capturing the kurtosis which characterizes the data. After presenting a Gibbs sampling scheme, we illustrate that the Gaussian mixture model provides a satisfactory representation of the data, allowing for the robust estimation of the cyclical component of per capita hours worked. Another important piece of evidence is that the outlying observations are not scattered randomly throughout the sample, but have a distinctive seasonal pattern. Therefore, seasonal adjustment plays a role. We finally show that, if a flexible seasonal model is adopted for the unadjusted series, the level of outlier contamination is drastically reduced.

Keywords: Gaussian Mixtures. Robust signal extraction. State Space Models. Bayesian model selection. Seasonality.
JEL codes: C11, C22, C52, E32.

1 Introduction

The series of average weekly hours in manufacturing (AWH, henceforth) is an important indicator of the state of the U.S. economy. It is considered, in particular, a leading economic indicator of output and employment in manufacturing (see for instance Glosser and Golden, 1998), since firms usually tend to respond to business cycle conditions by decreasing or increasing hours worked, before hiring or laying off workers. According to Cho and Cooley (1994), a sizeable share of the adjustment in total hours over the business cycle represents adjustment in average hours, while the remainder concerns changes in employment. Also, the procyclicality of per capita hours worked and their response to technology shocks is the subject of a vivid ongoing debate, see Galí and Rabanal (2005) and the references therein. Finally, AWH is listed among the 10 indicators that make up the Conference Board composite index of leading indicators (CBO, 2001).

The series measures an average of the number of hours worked per week by production workers in U.S. manufacturing industries. It is produced by the U. S. Bureau of Labor Statistics (http://www.bls.gov/ces/), as part of the Current Employment Statistics (CES) monthly survey, which obtains payroll hours, employment and earnings from business establishments.

The series is usually analyzed in seasonally adjusted (SA) form. Seasonal adjustment is carried out by the BLS according to the methodology reported at http://www-.bls.gov/ces/cesseasadj.htm. A particular problem is posed by the treatment of calendar related fluctuations, and in particular by the treatment of moving festivals (Easter and Labor day). As a matter of facts, the data are collected from the respondents' payroll records for the pay period that includes the 12th of each month. Thus, if Easter or Labor day falls during the week including the 12th, the number of hours worked will be reduced.

The seasonally adjusted series is plotted in the top panel of figure 1; the superimposed shaded ares locate the NBER recessions. The plot confirms the tendency to lead the business cycle peaks, and reveals the presence of a few occasional large drops in hours worked. The standardized fourth moment of the monthly growth rates is equal to 22.5, and the Jarque-Bera (JB, 1980) normality test statistics for this series is 11518. When computed on the yearly growth rates, the kurtosis drops to 3.79, and the JB normality statistics is a mere, though significant, 18.9. The unadjusted series is presented in the bottom plot. A noticeable feature is that the seasonal component does not look regular and evolves over time. It is perhaps interesting to single out the period 1975-1980, when the SA series is characterized by the same troughs as the unadjusted series.

AWH provides an interesting case study in Bayesian estimation of the business cycle in an environment heavily contaminated by outlying observations: the latter may have economic interpretation, as they can be due to strikes or overtime hours, but we document in this paper that a large share may be due to underadjustment of the seasonal component. The main objective of this paper is thus to carry out robust signal extraction with reference to the cyclical component. This is based on an unobserved components model featuring the decomposition into a long-run trend, a cycle and with

finite mixtures provide a flexible tool to model outliers and skewed distribution.

Outlying observations and structural breaks in the components can be handled by the inclusion of appropriate dummy variables on the right hand side of the measurement and transition equations. However, this strategy has several drawbacks: for instance, when a dummy is used to capture an additive outlier, this amounts to considering the corresponding observation as missing, so that a weight of zero is assigned to it in signal extraction and forecasting; on the contrary, the observation may still contain some information, which could be elicited by suitable downweighting. An alternative strategy consists in allowing the disturbances of a structural model to possess a heavy tailed density, such as Students' t-distribution, the general error distribution (Durbin and Koopman, 1997), or a mixture of Gaussian (Harrison and Stevens, 1976, Frühwirth-Schnatter, 2006, Giordani, Kohn and van Dijk, 2007). The second strategy is preferred in this paper since it allows to single out which observations are outlying.

The unobserved components model Gaussian mixture model provides a good fit to the data and is a significant improvement over a standard linear model, as it is revealed by model selection according to a proposal by Chib and Jeliazkov (2001).

The paper is structured as follows. Section 2 presents the unobserved components model for the seasonally adjusted series. Bayesian inference is discussed in section 3, whereas section 4 addresses the issue of model selection and deals with the estimation of the marginal likelihood of our model, using the approach suggested by Chib and Jeliazkov (2001). In section 5 we presents and discuss the estimation results, which finally lead us to the specification and the estimation of an unobserved components Gaussian mixture model for the unadjusted series (section 6). Section 7 concludes the paper.

2 The Gaussian mixture model

Let $y_t^{SA}, t = 1, \ldots, n$, denote the logarithm of seasonally adjusted AWH. The model for the series, plotted in the upper panel of figure 1, is an additive decomposition into a trend component, μ_t, representing the underlying long run evolution of the series, a stationary cycle, ψ_t, and an irregular component, ϵ_t, which is specified as follows:

$$
\begin{array}{lcll}
y_t^{SA} & = & \mu_t + \psi_t + \epsilon_t, & \\
 & & & \\
\mu_t & = & \mu_{t-1} + \eta_t, & \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \\
\psi_t & = & \phi_1 \psi_{t-1} + \phi_2 \psi_{t-2} + \kappa_t, & \kappa_t \sim \mathrm{N}(0, \sigma_\kappa^2) \\
\epsilon_t & = & (1 - S_t)\epsilon_{0t} + S_t \epsilon_{1t} & \epsilon_{it} \sim \mathrm{N}(0, \sigma_{\epsilon i}^2), \\
S_t & \sim & \text{IID Bernoulli}(\omega) &
\end{array}
\tag{1}
$$

The trend is a random walk, whereas the cycle is an autoregressive process of order 2, AR(2), with stationary complex roots, which is achieved via the reparameterization in terms of the modulus and the phase of the roots of the AR polynomial. In particular, we write $\phi_1 = 2\rho \cos \lambda_c$ and $\phi_2 = -\rho^2$, where ρ is defined in the interval (0,1), and λ_c is in $[0, \pi]$.

We assume that the irregular component has a finite mixture distribution, with two components. Denoting by $f(\epsilon_t)$ the probability density function of ϵ_t and by $g(\epsilon_t; \mu, \sigma_\epsilon^2)$ the univariate Gaussian density with mean μ and variance σ_ϵ^2,

$$f(\epsilon_t) = (1-\omega)g(\epsilon_t; 0, \sigma_{\epsilon 0}^2) + \omega g(\epsilon_t; 0, \sigma_{\epsilon 1}^2). \tag{2}$$

We identify the mixture parameters by imposing the ordering restriction $\sigma_{\epsilon 0}^2 < \sigma_{\epsilon 1}^2$. Letting S_t denote an IID Bernoulli indicator variable taking the values $0, 1$ with probability $\omega = P(S_t = 1)$ and $1 - \omega = P(S_t = 0)$, respectively, when the series is in a low volatility state 0, the irregular variance is $\sigma_{\epsilon 0}^2$; $\sigma_{\epsilon 1}^2$ is the irregular variance in the high volatility state ($S_t = 1$). This identification constraint avoids the label switching problem; see Geweke (2005) and Frühwirth-Schnatter (2006) for details.

We further assume that the random disturbances $\eta_t, \kappa_t, \epsilon_{0t}, \epsilon_{1t}$ are mutually independent. Conditionally on S_t, the model (1) is a linear Gaussian state space model with measurement equation given by the first equation and transition equation built from the Markovian representation for μ_t and ψ_t.

3 Bayesian Estimation

In this section we discuss how we make inference for the model presented in section 2. In particular, we discuss our prior choices, and we describe the algorithm used for computing the posterior distribution and the full conditional distributions.

Let $x = \{\mu_t, \psi_t, t = 0, \ldots, n\}$ denote the collection of the unobserved components, $\Psi = (\sigma_\eta^2, \sigma_\kappa^2, \sigma_{\epsilon 0}^2, \sigma_{\epsilon 1}^2, \lambda_c, \rho, \omega)$ the vector of parameters of the model as defined in section 2. We further denote by $S = \{S_1, \ldots, S_t, \ldots, S_n\}$ the $n-$dimensional vector of indicator variables S_t.

Our objective is to estimate the parameters of the posterior distribution of the vector of parameters Ψ, of the the unobserved states and the latent indicator, by generating random draws from the joint posterior of the unobserved components and vector of parameters itself, $p(\Psi, S, x|y)$, where $y = \left(y_1^{SA}, y_2^{SA}, \ldots, y_n^{SA}\right)$ denotes the vector of seasonally adjusted observations. A Metropolis-Hastings within Gibbs sampling algorithm is implemented, where we take the vector S, the matrix of unobserved states as a single block and partition the vector Ψ into three blocks, $\Psi = (\Sigma', \Lambda', \omega)'$, where $\Sigma = \left(\sigma_\eta^2, \sigma_\kappa^2, \sigma_{\epsilon 0}^2, \sigma_{\epsilon 1}^2\right)$, $\Lambda = (\lambda_c, \rho)$, and ω is the mixing probability.

3.1 Prior Distributions

In a mixture context being fully non-informative and obtaining proper posterior distributions is not feasible. Since there is always the possibility that no observations are allocated to one or more components, and so the data are uninformative about them, standard choices of independent improper non-informative prior distributions for the component parameters can not be used (Diebold and Robert, 1994). For this reason, the choice of the a-priori distributions is guided by a desire to ensure that posterior computations are relatively straightforward and that, as far as possible, the observed data is allowed to speak for itself without strong prior information being imposed.

Table 1: Prior hyper-parameters, initial values of the chain and lower bounds.

Parameter	Initial Value	α_J^0	β_J^0	Lower Bound
σ_η^2	2.83E-07	4.0	5.67E-07	1.00E-010
σ_κ^2	3.66E-04	4.0	7.32E-04	1.00E-06
$\sigma_{\epsilon_0}^2$	4.83E-06	4.0	9.67E-06	1.00E-07
$\sigma_{\epsilon_1}^2$	4.83E-05	4.0	9.67E-05	-

The usual Inverted Gamma prior distribution is chosen for the scale parameters in Σ, while a Beta distribution is chosen for the cycle parameters (λ_c, ρ), and for the mixing probability ω. This gives rise to the following structure of prior distributions:

$$p(\Sigma, \Lambda, \omega) \propto p(\Sigma)\, p(\Lambda)\, p(\omega) \tag{3}$$

where we assume an independent structure between each block of variables and within each block

$$p(\Sigma) \propto \prod_J \mathcal{IG}\left(\sigma_J^2, \frac{\alpha_J^0}{2}, \frac{\beta_J^0}{2}\right), \quad \forall J = \{\epsilon_0, \epsilon_1, \eta, \kappa\} \tag{4}$$

$$p(\Lambda) \propto \mathcal{B}e\left(\rho | p_1^0, p_2^0\right) \times \pi \mathcal{B}e\left(\lambda_c | c_1^0, c_2^0\right) \mathbb{I}_{[0,\pi]} \tag{5}$$

$$p(\omega) \propto \mathcal{B}e\left(\omega | s_1^0, s_2^0\right). \tag{6}$$

The choice of an inverted gamma structure for the variance parameters in Σ is motivated by the aforementioned need of avoiding improperness of the posterior distributions in the mixture framework. Under these priors the variance parameters have independent inverted gamma conditional posteriors, even in the case on no observations allocated to one of the two mixture components. However, as pointed out by Harvey, Trimbur and van Dijk (2007), the use of very slow shape and scale parameters for these distributions may lead to problems in estimation, in particular for parameters that tend to take on small values such as the variance of the trend component. In order to avoid distortions in the estimates of the variance components and degeneracy of the sampler, we introduced the lower bounds reproduced in the last column of table 1, even if in the post-processing of the simulations we realize that these bounds are not binding most of the time. The choice of a uniform prior for the cycle parameters (ρ, λ_c), over the stationarity region, which is a bounded region in $\mathbb{R} \times \mathbb{R}$, i.e. $(\rho, \lambda_c) \in (0,1) \times [0, \pi]$, does not pose a problem in terms of successfully generating competitive parameter values. As described in section 3.3 above, we use a Metropolis step for the simulation of variate from this full conditional distribution.

3.2 Likelihood and posterior

The complete-data likelihood is

$$p(y, x, S|\Psi) = \prod_{t=1}^{n} p(y_t|x_t, S_t, \Psi)\, p(x_t|x_{t-1}, \Psi)\, p(x_0), \tag{7}$$

where

$$p\left(y_t|x_t, S_t, \Psi\right) = \left\{\omega g\left(y_t|x_t, S_t = 0, \sigma_{\epsilon_0}^2\right)\right\}^{1-S_t}\left\{\left(1-\omega\right) g\left(y_t|x_t, S_t = 1, \sigma_{\epsilon_1}^2\right)\right\}^{S_t}, \quad (8)$$

represents the density of the mixture, $p\left(x_t|x_{t-1}, \Psi\right)$ is the transition density of the state space model which is markovian for the representation of the unobserved components proposed we use here, (see, e.g. Harvey, 1989), and $p\left(x_0\right)$ is the prior distribution on the initial vector of states x_0, which is diffuse for non-stationary components and centered around the long-term mean for the stationary component. The joint posterior distribution of the parameters and the unobservable components, $p\left(\Psi, x, S|y\right)$ will be proportional to the product of the likelihood and the prior distributions given in the previous subsection.

3.3 The Gibbs sampler

The Gibbs sampling approach to estimating the model parameters involves sampling from the complete conditional distribution of each parameter in a systematic manner, conditional on the previous sampled values of the other parameters. This approach is always possible, since the complete conditional densities are available, up to a normalizing constant, from the form of the likelihood and the prior (see, Geman and Geman (1994), and de Pooter, Segers and Van Dijk (2006) for an up to date overview of the state of the art in Bayesian computation using Gibbs sampler). When some of these conditional densities do not have standard form, as is often the case, the Metropolis-Hastings algorithm may be used to obtain realizations from a Markov chain having the required stationary distribution (see e.g., Casella and Robert (2004), and Gamerman and Lopes (2007)).

After choosing a set of initial values for the parameter vector $\Psi^{(0)}$, simulations $\left\{\Psi^{(i)}, x^{(i)}, S^{(i)}\right\}, i = 1, 2, \ldots,$ from the posterior distribution are obtained by iterating the following steps of the Gibbs sampler.

(i) *Update the indicator variable*

$$\begin{aligned} p(S_t = 1|\Psi^{(i)}, x^{(i)}, y) \quad \propto \quad & \frac{\omega}{\sigma_{\epsilon_0}} \exp\left\{-\frac{1}{2\sigma_{\epsilon_0}^2}\left(y_t - \mu_t - \psi_t\right)^2\right\} \\ & + \frac{(1-\omega)}{\sigma_{\epsilon_1}} \exp\left\{-\frac{1}{2\sigma_{\epsilon_1}^2}\left(y_t - \mu_t - \psi_t\right)^2\right\} \end{aligned} \quad (9)$$

with $t = 1, \ldots, n$.

(ii) *Simulate the matrix of unobserved components* $x^{(i+1)}$*, from the complete full conditional distribution* $p(x|\Psi^{(i)}, S^{(i+1)}, y)$*, where* $S^{(i+1)}$ *is the vector of indicator variables generate at the previous step of the Gibbs sampler.*

(iii) *Simulate the cycle parameters* $(\rho, \lambda_c)^{(i+1)}$ *from the full conditional distributions*

$$\begin{aligned} p\left(\rho|\lambda_c^{(i)}, \Sigma^{(i)}, \omega^{(i)}, x^{(i+1)}, S^{(i+1)}, y\right) & \propto p\left(\rho\right) \\ & \times \exp\left\{\frac{1}{2\sigma_\kappa^2}\sum_{t=1}^{n}\left(\psi_t - \rho\cos\lambda_c\,\psi_{t-1} + \rho^2\psi_{t-2}\right)^2\right\}, \end{aligned} \quad (10)$$

and

$$p\left(\lambda_c|\rho^{(i+1)},\Sigma^{(i)},\omega^{(i)},x^{(i+1)},S^{(i+1)},y\right)\propto p\left(\lambda_c\right)$$
$$\times\exp\left\{\frac{1}{2\sigma_\kappa^2}\sum_{t=1}^{n}\left(\psi_t-\rho\cos\lambda_c\,\psi_{t-1}+\rho^2\psi_{t-2}\right)^2\right\}, \quad (11)$$

where $\psi_t, t=0,1,\ldots,n$, *is defined as* $\psi_t=y_t-\mu_t$, *and the prior distributions are beta, as described in the previous section.*

(iv) *Simulate* $\Sigma^{(i+1)}$ *from the complete full conditional distributions*

$$p\left(\sigma_J^2|\Lambda^{(i+1)},x^{(i+1)},S^{(i+1)},y\right)\propto\mathcal{IG}\left(\sigma_J^2|\alpha_J,\beta_J\right), \quad \forall J=\{\epsilon_0,\epsilon_1,\eta,\kappa\}. \quad (12)$$

The parameters of these posterior distributions are

$$\alpha_J=\frac{\alpha_J^0}{2}+\frac{n_{\epsilon_J}}{2}, \qquad \beta_J=\frac{\beta_J^0}{2}+\frac{\bar{S}_J}{2},$$

where for $J=\{\epsilon_0,\epsilon_1\}$, $n_{\epsilon_0}=\sum_{t=1}^{n}S_t$ *and* $n_{\epsilon_1}=\sum_{t=1}^{n}(1-S_t)$ *are are the number of observations allocated to the two components of the mixture,*

$$\bar{S}_{\epsilon_0}=\sum_{i=1}^{n}S_t\left(y_t-\mu_t-\psi_t\right)^2, \quad \bar{S}_{\epsilon_1}=\sum_{i=1}^{n}\left(1-S_t\right)\left(y_t-\mu_t-\psi_t\right)^2,$$

and, for the trend and cycle posterior variance, $J=\{\eta,\kappa\}$, *we have* $\alpha_J=\frac{a_J^0+n}{2}$, *and*

$$\begin{aligned}\beta_\eta &= \frac{\alpha_\eta^0}{2}+\frac{1}{2}\sum_{t=1}^{n}\left(\mu_t-\mu_{t-1}\right)^2\\ \beta_\kappa &= \frac{\alpha_\kappa^0}{2}+\frac{1}{2}\sum_{t=2}^{n}\left(\psi_t-\rho\cos\lambda_c\psi_{t-1}-\rho^2\psi_{t-2}\right)^2.\end{aligned}$$

(v) *Simulate the mixing probability* $\omega^{(i+1)}$, *from the complete full conditional distribution*

$$p\left(\omega|\Lambda^{(i+1)},\beta^{(i+1)},\Sigma^{(i+1)},S^{(i+1)},S^{(i+1)},y\right)\propto\mathcal{B}e\left(\omega|g_0+n_{\epsilon_0},h_0+n_{\epsilon_1}\right). \quad (13)$$

We generate random draws from the full conditional distribution of the states $p(x|\Psi,S,y)$, at point (ii), using the simulation smoother for a linear state space model, developed by de Jong and Shephard, (1996) and Durbin and Koopman (2001)). All the computations are carried out in Ox 4.10 by Doornik (2006). For the simulation smoother we use the library Ssfpack, version 2.3, see Koopman et al. (1999).

The conditional posterior distributions of the cycle parameters (ρ,λ_c), at point (iii), have no explicit forms of well known distributions that can be easily sampled. Thus, we propose to update each parameter separately using Metropolis-Hastings (see

Metropolis et al. (1953) and Hastings (1970)) steps where the candidate distributions are chosen to have the same supports than the conditional posteriors and the variances can be calibrated using a small number of iterations of the algorithm in order to have an acceptance ratio of about 50%. More specifically, in order to sample from the full conditional posterior of $\rho^{(i)}$, for example, we generate candidate values from a Beta distribution, $\mathcal{B}e(a,b)$, where we pose the mean of the Beta distribution to be equal to the previously generated value of the parameter $\rho^{(i-1)}$. By inverting the relation linking the mean and the variance of the Beta distribution to its parameters (a,b), we obtain the following relation

$$\begin{cases} a = \frac{\rho_{i-1}^2 (1-\rho_{i-1}) - v\,\rho_{i-1}}{v} \\ b = a\frac{1-\rho_{i-1}}{\rho_{i-1}} \end{cases}$$

where ρ_{i-1} is the value generated by the Metropolis-Hastings sub-chain at step $(i-1)$ and v is the variance of the proposal distribution. This parameters choice allows us also to avoid numerical problems related to the evaluation of the Metropolis-Hastings acceptance ratio in the presence of fat tailed and quite spiked likelihood functions. We carry out the same operations for the parameter λ_c.

4 Model Selection

In this section, we give a brief account of Chib and Jeliazkov (2001) method to calculate an estimate of the marginal density using the output from the Metropolis-Hastings algorithm. This methodology allows us to discriminate which model, among g different proposals, provides a better representation of a vector a data $y = (y_1, y_2, \ldots, y_n)$, in an appropriate statistical sense. Euristically, we want to compare how likely are the data under the g different models, integrating over the parameter space.

Let us denote by $p(y|\Psi_k, \mathcal{M}_k)$ the density function of the data under model $\mathcal{M}_k$ and a vector of parameters, Ψ_k, and $p(\Psi_k|\mathcal{M}_k)$ as the prior density. In this paper, we compare two different models $\mathcal{M}_k$: $\mathcal{M}_1$ is a standard Gaussian state space model where the structural components are trend and cycle, while $\mathcal{M}_2$ is the Gaussian mixture model defined in equation 1. These models differ for the treatment of the error component, which is a standard Gaussian distribution in $\mathcal{M}_1$, and a mixture of two normal distribution in $\mathcal{M}_2$. Since the two components of the mixture distribution have the same mean but different variances, the implied unconditional distribution exhibits fatter tails than in the normal case. The marginal density of the data under model $\mathcal{M}_k$ is

$$m(y|\mathcal{M}_k) = \int p(y|\Psi_k, \mathcal{M}_k)\, p(\Psi_k, \mathcal{M}_k)\; d\Psi_k, \qquad k = 1, 2, \tag{14}$$

where $p(\Psi_k, \mathcal{M}_k)$ represents the prior distribution of the parameter vector Ψ_k under model $\mathcal{M}_k$. The formal bayesian approach for comparing model $\mathcal{M}_1$ and $\mathcal{M}_2$ is through the pairwise Bayes factor, defined as the ratio of marginal likelihoods

$$B_{1,2} = \frac{m(y|\mathcal{M}_1)}{m(y|\mathcal{M}_2)}, \tag{15}$$

which can also be interpreted as the posterior probability of model $\mathcal{M}_1$, when both models are, a priori, equally likely. For a exhaustive discussion on Bayesian model comparison using Bayes factors, we refer to the paper of Kass and Raftery (1995), Chipman et al. (2001), Pericchi (2005) and the bibliography quoted therein, in particular for the explanation of why it may be misleading to use arbitrary vague priors and improper priors in a model selection framework.

Calculating the marginal density, and hence the Bayes factor, is a daunting task when estimation of the unknown parameters has been done by Markov chain Monte Carlo methods. Recently, several Monte Carlo methods for computing the marginal likelihoods have been developed, which include the Chib estimator, Chib (1995), the ratio of importance sampling, Chen and Shao (1997), the path sampling, Gelman and Meng (1998), the bridge sampling, Meng and Shilling (2002), and the extension of the Chib (1995) estimator for the case of some Metropolis steps within the Gibbs sampler, Chib and Jeliazkov (2001). In what follows we adapt the Chib and Jeliazkov (2001) method to our general framework, in which some blocks of parameters of the posterior distribution are updated by means of Gibbs steps, while some other are updated by applying the more general Metropolis-Hastings algorithm.

The approach developed by Chib and Jeliazkov (2001) is based on the so called marginal likelihood identity:

$$\log m\left(y|\mathcal{M}_k\right) = \log p\left(y|\Psi_k, \mathcal{M}_k\right) + \log p\left(\Psi_k|\mathcal{M}_k\right) - \log p\left(\Psi_k|y, \mathcal{M}_k\right). \tag{16}$$

The first terms of the RHS of equation (16) have a closed form expression, and can be evaluated, for both models, by running the Kalman filter for the relevant state space model, while the second component is simply the product of the prior distribution for the parameters of each model. The last component, i.e. the normalized posterior density of the parameters, requires a careful treatment. In fact, the computation of $p\left(\Psi_k|y, \mathcal{M}_k\right)$ is possible only conditioning on a particular set of unobservable components. The advantage of the approach proposed by Chib and Jeliazkov (2001) is to provide a method to integrate this function over the space of the unobservable components directly using the output of the a general MCMC algorithm. This operation differs if the estimates of the parameters come from a pure Gibbs sampler (Σ, ω), or from a Metropolis-within Gibbs step (Λ). By using the law of total probabilities, we express the joint posterior, evaluated at Ψ^*, (Chib and Jeliazkov (2001) suggests to choose a high density point in the support of the posterior, to assure the stability of the estimate of the marginal likelihood), as the product of conditional distributions, which in case of model $\mathcal{M}_1$ becomes (for convenience, we now suppress model conditioning notation)

$$p\left(\Psi^*|y\right) = p\left(\Sigma^*|y\right) p\left(\omega^*|\Sigma^*, y\right) p\left(\Lambda^*|\Sigma^*, \omega^*, y\right) \tag{17}$$

where $p\left(\Sigma^*|y\right)$ is the marginal density ordinate of Σ, $p\left(\omega^*|\Sigma^*, y\right)$ is a reduced conditional ordinate, and $p\left(\Lambda^*|\Sigma^*, \omega^*, y\right)$ is the full conditional density ordinate. In Appendix, we give a detailed description of how we apply the Chib estimator to evaluate the posterior density ordinate, in our specific case.

5 Estimation Results

For the seasonally adjusted AWH series we estimated the standard linear Gaussian decomposition and the mixture model (1); the former is obtained by setting $S_t = 0, t = 1, \ldots, n$ in 1. The estimation results presented in this section are based on a sample of 80000 draws from the Gibbs sampling scheme outlined in the previous section, with a burn-in of 20000 iterations.

Our assessment of the convergence of the chain is based on the statistical properties of the observed chain. Table 2 summarizes some aspects of the posterior distribution of the parameters and presents the main convergence diagnostics. The acronim LinSA refers to the linear Gaussian model; MixSA to the mixture model for the SA series. If we let

$$V_L = c_0 + 2\sum_{j=1}^{l} w_j c_j, \quad w_j = \frac{l-j}{l+1}$$

denote the long run variance of a parameter sample path, where c_j is the autocovariance at lag j and l is the truncation parameter, *persistence* is defined as S_L divided by the variance (c_0) (i.e., an estimate of the normalized spectral density at the zero frequency). Moreover, if $\psi^{(j)}$ denotes the j-th sample of the GS scheme, after the burn-in period, and $\bar{\psi}_a$ denotes the average of the first n_a draws, $\bar{\psi}_b$ is the average of the last n_b draws at the end of the convergence period, which are sufficiently remote to prevent any overlap, the Geweke's convergence statistic (Geweke, 1992, 2005) is

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

Figure 2 summarises aspects of the posterior distribution of the level, the cycle and the irregular for the linear Gaussian model. The posterior distribution of variance of the trend disturbances, σ_η^2, and the weakly evolutive pattern of the estimated level confirm that unit root and stationarity tests may provide little guidance regarding the as to whether AWH are level stationary or difference stationary.

The plot reveals the presence of a number of observations which are highly influential for the estimates of the cycle and the irregular. On the contrary, the estimated level is quite robust to the influence of those observations.

Modeling the irregular as a mixture of two Gaussian distributions has several benefits. First and foremost, the cycle estimates, plotted in figure 3, are smoother and are not affected by the outliers. The comparison of figure 3 with the estimated components for the standard linear model (see 2) suggests that part of the variability affecting the cycle estimates has been reallocated to the irregular component. This enables a clearer description of the business cycle phases and improves the characterisation of the turning points. Secondly, the reliability of the cycle estimates increases, as the posterior variance of the cycle estimate reduces. That the mixture model outperforms the Gaussian linear model is confirmed by the comparison of the estimated marginal likelihoods under the two models. The following table reports the Chib and Jeliazkov (2001) estimator of the marginal likelihood for the mixture model $\mathcal{M}_1$ and the Gaussian model $\mathcal{M}_2$:

Model	$\log p(y\|\Psi,\mathcal{M}_k)$	$\log p(\Psi_k,\mathcal{M}_k)$	$\log \bar{p}(\Psi_k\|y,\mathcal{M}_k)$	$\log m(y\|\mathcal{M}_k)$
$\mathcal{M}_1$	1.77	9.45	19.42	-8.20
$\mathcal{M}_2$	137.50	8.19	21.36	124.44

Table 2: Posterior means, medians, variances and convergence diagnostics.

	Mean		Median		Variance		Persistence		Geweke's C_G	
	LinSA	MixSA	LinSA	MixSA	LinSA	MixSA	LinSA	MixSA	LinSA	MixSA
σ^2_η	6.79e-07	3.92e-07	4.60e-07	3.07e-07	1.03	0.74	8696	79.96	-0.89	1.57
λ_c	0.04	0.04	0.04	0.032	0.69	0.69	43.81	37.56	-0.46	-0.85
ρ	0.73	0.76	0.73	0.76	0.038	0.03	56.17	38.23	-0.10	-.99
σ^2_κ	1.08e-05	1.00e-05	1.06e-05	8.16e-06	0.13	0.11	24.52	18.95	1.63	1.59
$\sigma^2_{\epsilon 0}$	1.44e-05	2.40e-06	1.44e-05	2.39e-06	0.086	0.20	1696	30.96	0.26	-1.01
$\sigma^2_{\epsilon 1}$		1.61e-04		1.55e-04		0.26		8.68		-0.16
ω		0.08		0.08		0.02		15.38		-0.00

An interesting stylized fact emerges from the consideration of the estimated posterior probabilities of being in a high volatility state, $P(S_t = 1|y, \Psi)$, which are displayed in figure 6. The Monte Carlo estimate of the posterior mean of ω is 0.08, i.e. about one in ten observations is likely to be outlying. However, a closer inspection reveals that the the outlying observations are clustered in certain months of the year. This is clear from the bottom panel, which is the plot by month of the posterior probabilities. The solid line connects the monthly averages, and the dotted line is drawn at the average 0.08. In particular, the outliers are clustered in the initial months of the year, namely January (for which the average value of the posterior probability is 0.18), February, April and December. The average probability is a mere 0.03 in June.

If the estimated $\hat{p}_t = P(S_t = 1|y, \psi)$ are regressed on a constant and 11 seasonal dummies, the test for the joint significance of the seasonal dummies takes the value 37.7 with a p-value of 8.9E-005. If the covariance matrix of the ordinary least squares estimator is corrected for heteroscedasticity and autocorrelation (see Newey and West, 1987), the evidence is unchanged. This suggests that the outlying observations have a marked periodic pattern.

6 Modeling Seasonality

The analysis of the seasonally adjusted data allows us to conclude that the Gaussian mixture model provides a useful representation of the data, allowing for the robust estimation of the cyclical component. Another important piece of evidence is that the outlying observations are not allocated randomly throughout the sample, but have a distinctive seasonal pattern. This may be the consequence of seasonal under and/or overadjustment. In this section we further investigate whether this is indeed the case, by estimating the same mixture model on the unadjusted series, $y_t, t = 1, \ldots, n$. Naturally, we have to extend the model by considering a seasonal component and a calendar component.

The model for the seasonal time series is specified as follows:

$$y_t = y_t^{SA} + \gamma_t + \beta_1 E_t + \beta_2 L_t, \tag{18}$$

where y_t^{SA} was given above in 1. The seasonal component, γ_t, is modeled as follows:

$$\gamma_t = z_t' \chi_t, \quad \chi_t = \chi_{t-1} + \zeta_t \tag{19}$$

where $z_t' = [D_{1t}, \ldots, D_{st}]$, with $D_{jt} = 1$ in season j and 0 otherwise. The vector χ_t contains the effects associated to each season and changes over time according to a multivariate random walk; ζ_t is a zero-mean multivariate white noise with covariance matrix which enforces the constraint $i_s' \text{Var}(\zeta_t) = 0$. This formulation is known in the literature as the Harrison and Stevens (1976) specification. The distinguishing feature of this approach is that it is formulated directly in terms of the effect of a particular season, thereby enhancing flexibility needed to model seasonal heteroscedasticity (that is when there are seasons which are 'more variables' than others, see Proietti, 1998). The appropriate action for this model to deal with heteroscedasticity is to define the covariance matrix of the seasonal innovations as follows:

$$\text{Var}(\zeta_t) = \Omega = D - \frac{1}{i_s' D i_s} D i_s i_s' D \tag{20}$$

where D is a diagonal matrix, $D = \text{diag}\{d_j, j = 1, \ldots, 12\}$. Our preferred specification has the d_j constant across groups of seasons. In particular, it envisages two groups of seasons, made up respectively by January, February and April, and the remaining months, characterized by the two parameters, d_a and d_b, which are constant across the months belonging to the same group.

The regression component, $\beta_1 E_t + \beta_2 L_t$, captures calendar effects: E_t and L_t are deterministic dummy variables taking value 1 if respectively Easter and Labor day fall within the observations week (that containing the 12th of each month).

6.1 Bayesian Estimation

Estimation is carried out in the same fashion as for the nonseasonal model in section (3). The unobserved states collected in the vector x now include also the seasonal component, i.e., $x = \{\mu_t, \psi_t, \gamma_t, t = 0, \ldots, n\}$. The vector of unknown parameters $\Psi = (\sigma_\eta^2, \sigma_\kappa^2, \sigma_{\epsilon 0}^2, \sigma_{\epsilon 1}^2, \lambda_c, \rho, \omega, \beta_1, \beta_2, d_a, d_b)$, is also updated to keep into account the block of exogenous regressors β and the block of seasonal parameters d, introduced in the previous section.

To make Bayesian inference on this seasonal model we need to specify the prior distributions on these new parameters. To the extent of being noninformative, we specify the prior distribution for the β's as the Jeffrey's reference prior for location parameters $p(\beta_1, \beta_2) \propto 1$, and we specify a Uniform distribution over a reasonable support for both the seasonal parameters $p(d_a, d_b) \propto \text{U}_{(0,10]} \times \text{U}_{(0,10]}$.

Now one iteration of the Gibbs sampling is completed by adding the following two steps, just before updating the mixing probabilities in the previous version of the algorithm

(v) *Simulate* $\beta^{(i+1)} = \left(\beta_1^{(i+1)}, \beta_2^{(i+1)}\right)$ *from the complete full conditional distribution*

$$p\left(\beta|\Lambda^{(i+1)}, \Sigma^{(i+1)}, \omega^{(i)}, x, S^{(i+1)}, y\right) \propto N_2\left(\beta|\delta, \tau^2\right).$$

The posterior location and scale parameters are, respectively

$$\tau^{-2} = \left(F'F\right), \qquad \delta = \left(F'F\right)^{-1} F'\tilde{y}$$

where F *is a* $(n \times 2)-$*matrix of observations of the two regressors* E_1 *and* L_1*, and* $\tilde{y}_t = y_t - \psi_t - \mu_t - \gamma_t$ *represent the difference between the original series* y_t *and the unobserved components* μ_t*,* ψ_t *and* γ_t*.*

(vi) *Simulate* $(d_a, d_b)^{(i+1)}$ *from the complete full conditional distribution*

$$p\left(d_a, d_b|\beta^{(i+1)}, \Lambda^{(i+1)}, \Sigma^{(i+1)}, \omega^{(i)}, x, S^{(i+1)}, y\right) \propto \frac{1}{|\Omega|} \exp\left\{-\frac{1}{2}\sum_{t=1}^{n} \zeta_t' \, \Omega^{-1} \, \zeta_t\right\}$$

where $\zeta_t, t = 1, 2, \ldots, n$ *and* Ω *have been defined in equation (19) and (20) respectively.*

The form of the full conditional distribution of the seasonal parameters (d_a, d_b) is not known due to the way the two parameters enters the likelihood, through the relation specified in equation (20). To simulate from such a distribution we use a Gaussian multiplicative random walk Metropolis-Hastings algorithm where proposed values are generated using the relation

$$\begin{bmatrix} \log \tilde{d}_a \\ \log \tilde{d}_b \end{bmatrix} = \begin{bmatrix} \log d_a^{(i-1)} \\ \log d_b^{(i-1)} \end{bmatrix} + \begin{bmatrix} \xi_a \\ \xi_b \end{bmatrix}, \tag{21}$$

where $\left(d_a^{(i-1)}, d_b^{(i-1)}\right)$ are the values sampled at the previous iteration of the algorithm, while $\left(\tilde{d}_a, \tilde{d}_b\right)$ are the proposed values, and the two innovations terms (ζ_a, ζ_b) are distributed accordingly to

$$\begin{bmatrix} \xi_a \\ \xi_b \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \nu_a^2 & 0 \\ 0 & \nu_b^2 \end{bmatrix}\right).$$

The two variance parameters $\left(\nu_a^2, \nu_b^2\right)$ are calibrated in order to obtain a acceptance ratio equal to about $50 - 60\%$. The acceptance probability of this Metropolis step is equal to

$$\alpha\left(\tilde{d}, d^{(i-1)}\right) = \min\left\{1, \frac{|\widetilde{\Omega}|}{|\Omega|} \exp\left[-\frac{1}{2}\sum_{t=1}^{n} \zeta_t' \, \widetilde{\Omega}^{-1} \, \zeta_t + \frac{1}{2}\sum_{t=1}^{n} \zeta_t' \left(\Omega^{(i-1)}\right)^{-1} \zeta_t\right] \mathrm{J}\right\}$$

where $\widetilde{\Omega} = \Omega\left(\tilde{d}_a, \tilde{d}_b\right)$, represent the proposed value of the variance of ζ_t and $\Omega^{(i-1)} = \Omega\left(\tilde{d}_a^{(i-1)}, \tilde{d}_b^{(i-1)}\right)$ represents the previous value of the same quantity, while $\mathrm{J} = \left|\frac{\tilde{d}_a}{d_a^{(i-1)}} \frac{\tilde{d}_b}{d_b^{(i-1)}}\right|$, is the inverse of the jacobian of the transformation defined in equation (21).

6.2 Estimation Results

The estimation results presented in this section are based on a sample of 50000 draws from the Gibbs sampling scheme above, with a burn-in of 50000 iterations. The draws and the nonparametric estimates of the posterior densities are reproduced in figure 5. Table 3 reports the posterior means, medians, variances and convergence diagnostics of the parameters in the vector Ψ. It is remarkable that the estimated proportion of outlying observation is much reduced (the posterior mean of the mixture parameter ω is a mere 0.01). Actually, as it is shown in the first panel of figure 6, which displays the posterior probabilities of the high variance component, $P(S_t = 1|y, \Psi)$, there is one single observation belonging to the high volatility state with posterior probability greater than 0.5. Morever, the periodic feature has disappeared.

The second panel of the figure, which plots the estimated seasonal component, $\mathrm{E}(\gamma_t|y)$, reveals that the seasonal pattern is highly evolutive over time. Moreover, the seasonal component absorbs a relevant part of the volatility that the model fitted to the SA series ascribed to the irregular component.

The cycle estimated from the model (bottom panel of figure 6) does not differ from that of the estimated using the MixSA model.

Table 3: Posterior means, medians, variances and convergence diagnostics.

	Mean	Median	Variance	Persistence	Geweke's C_G
σ^2_η	4.67e-007	3.38e-007	0.84	83.07	2.86
λ_c	0.047	0.041	0.10	23.42	0.87
ρ	0.74	0.74	0.03	28.01	-2.04
σ^2_κ	1.03e-005	1.0e-05	0.10	38.05	-1.49
$\sigma^2_{\epsilon 0}$	5.1e-07	4.8e-07	0.30	43.12	1.41
$\sigma^2_{\epsilon 1}$	6637e-07	4886e-07	1.00	3.95	-0.23
ω	0.01	0.01	0.70	24.25	0.70
d_a	6.8e-06	6.8e-06	0.08	71.13	-1.89
d_b	1.1e-05	1.1e-05	0.08	82.49	-1.48
β_1	0.01	0.01	0.08	11.17	-2.53
β_2	0.01	0.01	0.13	1.30	-0.57

7 Conclusions

For the U.S. seasonal adjusted average weekly hours worked time series we have proposed an unobserved component which decomposes the series into a local level component, an AR(2) cycle and an irregular term represented by a Gaussian scale mixture model. The model was fitted using Monte Carlo Markov chain methods and proved to be a significant improvement over the linear model with no mixture component, in the light of the comparison of the marginal likelihood of the two models. The latter was computed using the Chib and Jeliazkov (2001) approach. The average level of outlier contamination was estimated to be equal to 8%.

The mixture model allows to robustify the estimates of the cyclical components, which can already be considered the end result of this paper. However, we documented that the pattern of the observations belonging to the high volatility state has a marked periodic pattern, which may be the consequence of under- or over-adjustment. As a matter of fact, if we model the unadjusted time series by a straightforward extension of the model, contemplating the presence of a time evolutive seasonal component, along with a regression component capturing the role of moving festivals, the level of outlier contamination undergoes a dramatic reduction (only one observation could be categorized as outlying).

A Model Selection

In this appendix we discuss in detail how we implement the Chib and Jeliazkov (2001) estimator of the marginal likelihood for the Gaussian mixture model defined in section 1. We evaluate each of the components of the posterior distribution of the parameters in equation 17 by means of the following steps.

(i) The draws from the full MCMC run described in section 3.3, are used to estimate the marginal density ordinate of the block parameters Σ, by

$$\begin{aligned} p\left(\Sigma^{*}|y\right) &= \int \prod_{J} \mathcal{IG}\left(\sigma_{J}^{2}|\alpha_{J},\beta_{J}\right) p\left(\omega,\Lambda,x,S|y\right)\, d\omega\, d\Lambda\, dx\, dS \\ &\simeq \frac{1}{M}\sum_{m=1}^{M}\prod_{J}\mathcal{IG}\left(\sigma_{J}^{2}|\alpha_{J}^{(m)},\beta_{J}^{(m)}\right), \quad J=\{\epsilon_{0},\epsilon_{1},\eta,\kappa\} \end{aligned}$$

where $\left\{\alpha_{J}^{(m)},\beta_{J}^{(m)}\right\}_{m=1}^{M},\forall J$ are the parameters of the complete full conditional distributions of the MCMC sampler described in section 3.3 point *(iv)*, computed at the generated values.

(ii) Next, we fix the parameter Σ to Σ^{*}, and we obtain new draws $\left\{\omega^{(g)},\Lambda^{(g)},x^{(g)},S^{(g)}\right\}_{g=1}^{G}$, from a reduced MCMC with densities

$$p\left(\omega|y,x,\Lambda,\Sigma^{*},S\right), \quad p\left(\Lambda|y,x,\Sigma^{*},\omega,S\right), \quad p\left(x|y,\omega,\Sigma^{*},\Lambda,S\right), \quad p\left(S|y,x,\omega,\Sigma^{*},\Lambda\right)$$

which are used to estimate the reduced posterior ordinate of the parameter ω

$$\begin{aligned} p\left(\omega^{*}|\Sigma^{*},y\right) &= \int \mathcal{B}e\left(\omega|g_{0}+n_{\epsilon_{0}},h_{0}+n_{\epsilon_{1}}\right) p\left(\Lambda,x,S|y,\Sigma^{*}\right)\, d\Lambda\, dx\, dx \\ &\simeq \frac{1}{M}\sum_{m=1}^{M}\mathcal{B}e\left(\omega|g_{0}+n_{0}^{(g)},h_{0}+n_{1}^{(g)}\right), \end{aligned}$$

where $n_{\epsilon_{0}}^{(g)}$ and $n_{\epsilon_{1}}^{(g)}$, are the number of observations allocated to the two components of the mixture by the new reduced MCMC.

(iii) Finally, the full conditional density ordinate

$$p\left(\Lambda^{*}|\Sigma^{*},\omega^{*},y\right) = \int p\left(\Lambda^{*}|y,x,\Sigma^{*},\omega^{*},S\right) p\left(x,S|y,\Sigma^{*},\omega^{*}\right)\, dx\, dS$$

can not be estimated by Rao-Blackwellization, as in the previous cases, because the normalizing constant of the full conditional distribution $p\left(\Lambda^{*}|y,x,\Sigma^{*},\omega^{*},S\right)$ is not known. In particular, as described in section 3.3, we use two Metropolis-Hastings within Gibbs steps to generate random variates from the full conditionals distributions in equation 10 and 11. Starting from the reversibility condition of the M-H sub-kernel, Chib and Jeliazkov (2001) prove that the reduced conditional ordinate of a parameter, for which the normalizing constant is not know,

can be expressed as the ratio of two expectations and estimated by MC averages of the MCMC output, in a similar way as we did in the previous steps. In what follows, we adapt the general formula of Chib and Jeliazkov (2001) to our problem of evaluating the normalizing constant of the reduced conditional ordinate $p(\rho^*, \lambda_c^*|\Sigma^*, \omega^*, y)$. Let $\alpha(\rho, \rho^*|y, \lambda_c, \omega^*, \Sigma^*, S)$ denotes the acceptance probability of the M-H algorithm implemented within the Gibbs sampling to sample from the full conditional distribution of ρ, (see equation 10)[1], and let $q(\rho, \rho^*)$ the proposal distribution for the transition form ρ to the new value ρ^*[2], then $p(\rho^*, \lambda_c^*|\Sigma^*, \omega^*, y)$ can be expressed as the ratio of two expectations, in the following way

$$p(\rho^*|\Sigma^*, \omega^*, y) = \frac{\mathbb{E}_1\left[\alpha(\rho, \rho^*|y, \lambda_c, \omega^*, \Sigma^*, S)\, q(\rho, \rho^*)\right]}{\mathbb{E}_2\left[\alpha(\rho^*, \rho|y, \lambda_c, \omega^*, \Sigma^*, S)\right]}, \tag{22}$$

where the expectation at the numerator $\mathbb{E}_1$ is taken with respect to the density $p(\rho, \lambda_c, x, S|y, \omega^*, \Sigma^*)$, while the expectation at the denominator $\mathbb{E}_2$ is taken with respect to the density $p(\lambda_c, x, S|y, \rho^*, \Sigma^*, \omega^*)\, q(\rho^*, \rho|y, \lambda_c, \Sigma^*, \omega^*, S)$. Each of the integrals in equation 22 can be estimated by the output of the MCMC algorithm.

1. To estimate the numerator, fix Σ and ω to the maximum a posteriori (Σ^*, ω^*), run a reduced MCMC with densities

   $$p(\Lambda|y, x, \Sigma^*, \omega^*, S), \quad p(x|y, \omega^*, \Sigma^*, \Lambda, S), \quad p(S|y, x, \omega^*, \Sigma^*, \Lambda),$$

   and take the draws $\left\{\Lambda^{(l)}, x^{(l)}, S^{(l)}\right\}_{l=1}^{L}$ to average the quantity $\alpha(\rho, \rho^*|y, \lambda_c, \omega^*, \Sigma^*, S)\, q(\rho, \rho^*)$, i.e.

   $$\widehat{\mathbb{E}}_1\left[\alpha(\rho, \rho^*|y, \lambda_c, \omega^*, \Sigma^*, S)\, q(\rho, \rho^*)\right] \simeq \frac{1}{L}\sum_{l=1}^{L} \alpha\left(\rho^*, \rho|y, \lambda_c^{(l)}, \omega^*, \Sigma^*, S^{(l)}\right) q\left(\rho^{(l)}, \rho^*\right).$$

2. For the denominator of equation 22, because the expectation is conditioned on ρ^*, we run an additional reduced MCMC algorithm with ρ fixed at ρ^*, and full conditionals

   $$p(\lambda_c|y, x, \Sigma^*, \omega^*, \rho^*, S), \quad p(x|y, \omega^*, \Sigma^*, \lambda_c, \rho^*, S), \quad p(S|y, x, \omega^*, \Sigma^*, \rho^*, \lambda_c).$$

   At each iteration of this reduced sampler, a value of ρ is drawn from the proposal distribution of the Metropolis step $q(\rho^*, \rho)$, conditional on the previous draws of $\left(\lambda_c^{(h)}, x^{(h)}, S^{(h)}\right)$ from the reduced sampler, i.e.

   $$\rho^{(h)} \sim q(\rho^*, \rho)$$

[1] The acceptance probability of the M-H step, $\alpha(\rho, \rho^*|y, \lambda_c, \omega^*, \Sigma^*, S)$, depends on the parameters $(\lambda_c, \omega, \Sigma)$ and on the latent vector S, through the likelihood ratio.

[2] The proposal distribution is a Beta distribution with mean equal to the previous value of the sub-chain, as discussed in section 3.3.

leading to the sample $\left\{\lambda_c^{(h)}, x^{(h)}, S^{(h)}, \rho^{(h)}\right\}_{h=1}^{H}$ from the distribution

$$p\left(\lambda_c, x, S|y, \rho^*, \Sigma^*, \omega^*\right) q\left(\rho^*, \rho\right).$$

The denominator of equation 22 is then estimated by

$$\widehat{\mathbb{E}}_2\left[\alpha\left(\rho^*, \rho|y, \lambda_c, \omega^*, \Sigma^*, S\right)\right] = \frac{1}{H}\sum_{h=1}^{H} \alpha\left(\rho^*, \rho^{(h)}|y, \lambda_c^{(h)}, \omega^*, \Sigma^*, S^{(h)}\right),$$

and this allows us to estimate the ordinate $p\left(\rho^*|y, \Sigma^*, \omega^*\right)$

$$\widehat{p}\left(\rho^*|y, \Sigma^*, \omega^*\right) = \frac{\widehat{\mathbb{E}}_1\left[\alpha\left(\rho, \rho^*|y, \lambda_c, \omega^*, \Sigma^*, S\right) q\left(\rho, \rho^*\right)\right]}{\widehat{\mathbb{E}}_2\left[\alpha\left(\rho^*, \rho|y, \lambda_c, \omega^*, \Sigma^*, S\right)\right]} \tag{23}$$

Exactly in the same way as we did for the parameter ρ, we can now estimate the ordinate for λ_c, using the following relation

$$\widehat{p}\left(\lambda_c^*|\Sigma^*, \omega^*, \rho^*, y\right) = \frac{\widehat{\mathbb{E}}_1\left[\alpha\left(\lambda_c, \lambda_c^*|y, \rho^*, \omega^*, \Sigma^*, S\right) q\left(\lambda_c, \lambda_c^*\right)\right]}{\widehat{\mathbb{E}}_2\left[\alpha\left(\lambda_c^*, \lambda_c|y, \rho^*, \omega^*, \Sigma^*, S\right)\right]}, \tag{24}$$

where the expectation in the numerator of equation 24 is taken with respect to $p\left(\lambda_c, x, S|y, \rho^*, \Sigma^*, \omega^*\right)$ and it is estimated by averaging $\alpha\left(\lambda_c, \lambda_c^*|y, \rho^*, \omega^*, \Sigma^*, S\right) q\left(\lambda_c, \lambda_c^*\right)$ using the same draws form the reduced MCMC described in the previous point 2. The expectation in the denominator of equation 24 is taken with respect to the distribution $p\left(x, S|y, \rho^*, \Sigma^*, \omega^*, \lambda_c^*\right) q\left(\lambda_c^*, \lambda_c\right)$, and it is estimated by averaging $\alpha\left(\lambda_c^*, \lambda_c|y, \rho^*, \omega^*, \Sigma^*, S\right)$ using the output of a reduced MCMC with densities

$$p\left(x|y, \omega^*, \Sigma^*, \lambda_c^*, \rho^*, S\right), \quad p\left(S|y, x, \omega^*, \Sigma^*, \lambda_c^*, \rho^*\right)$$

with the additional simulation, at each step, of a random variates from the proposal distribution $q\left(\lambda_c^*, \lambda_c\right)$ conditional to the generated value of S.

References

Bernardo, J. M. (2005). Reference Analysis. In *Handbook of Statistics, 25*, Elsevier, North-Holland, Amsterdam, 459-507.

Casella, G., and Robert, C. P. (2004). *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer, New York.

Chen M.-H and Q.-M. Shao (1997). On Monte Carlo methods for estimating ratios of normalizing constants. *The Annals of Statistics*, 25, 1563-1594.

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

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

Chipman, H., E. I. George, and E. McCulloch (2001). The practical implementation of Bayesian model selection. *IMS Lecture Notes - Monograph series (2001)*, 38.

Cho, J-O. and Cooley, T. F. (1994). Employment and hours over the business cycle. *Journal of Economic Dynamics and Control*, 18, 411-432.

Conference Board (2001). Business Cycle Indicators Handbook. Available at `http://www.conference-board.org/pdf_free/economics/bci/BCI-Handbook.pdf`

de Jong, P., and Shephard, N. (1996). The simulation smoother. *Biometrika*, 2, 339-50.

de Pooter, M. D., Segers, R., and van Dijk, H. K. (2006). On the Practice of Bayesian Inference in Basic Economic Time Series Models using Gibbs Sampling. *Tinbergen Institute Discussion Paper*, 076/4.

Diebold, J. and Robert, C. P., (1994). Estimation of finite mixture distributions through Bayesian sampling. *Biometrika*, 56, 363 - 375.

Doornik, J.A. (2006), *Ox: An Object-Oriented Matrix Programming Language*, Timberlake Consultants Press, London.

Durbin, J., and S.J. Koopman (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika*, 84, 669-84.

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

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

Galí, J., and Rabanal, P. (2005). Technology Shocks and Aggregate Fluctuations: How Well Does the RBC Model Fit Postwar U.S. Data? In *NBER Macroeconomics Annual* 2004, (Mark Gertler and Kenneth Rogoff eds.), 22588. Cambridge, MIT Press.

Gamerman, D., and Lopes H. F. (2007). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Second Edition. Text in Statistical Science Series. Wiley.

Gelman, A., and X.-L. Meng (1998). Simulating normalizing constants: from importance sampling to bridge sampling to path sampling *Statistical Sciences*, 18, 163-185.

Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

Geweke, J., (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In J. M. Bernardo, J. Berger, A. P. Dawid, and A. F. M. Smith, eds., *Bayesian Statistics 4*, Oxford University Press, pp. 169-193.

Geweke, J., (2005). *Contemporary Bayesian Econometrics and Statistics*. Wiley Series in Probability and Statistics. Wiley, Hoboken.

Giordani, P., Kohn, R., and van Dijk, D. (2007). A unified approach to nonlinearity, structural change, and outliers. *Journal of Econometrics*, 127, 112–133.

Glosser, S. M., and Golden, L. (1997). Average work hours as a leading economic variable in US manufacturing industries. *International Journal of Forecasting*, 13, 175–195.

Harrison, P. and C. Stevens (1976). Bayesian forecasting. *Journal of the Royal Statistical Society. Series B*, 38, 205-247.

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

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, 618 - 649.

Hastings W. K., (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97-109.

Kass, R. E., and A. E. Raftery. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773-795.

Koopman S.J., Shepard, N., and Doornik, J.A.: "Statistical algorithms for models in state space using SsfPack 2.2", *Econometrics Journal*, 2 (1999), 113-166.

Newey, W.K., and West, K.D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55, 703–708.

Meng, X.-L, and S. Schilling (2002). Warp bridge sampling. *Journal of Computational and Graphical Statistics*, 11, 552-586.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087-1091.

Pericchi, L. R. (2005). Model selection and hypothesis testing based on objective probabilities and Bayes factors. *Handbook of Statistics Vol. 25*.

Proietti, T. (1998). Seasonal heteroscedasticity and trends. *Journal of Forecasting*, 17, 1-17.

Figure 1: Average weekly hours in manufacturing: seasonally adjusted series (logarithms) and unadjusted series. The shaded areas flag recessionary periods, according to the NBER chronology.

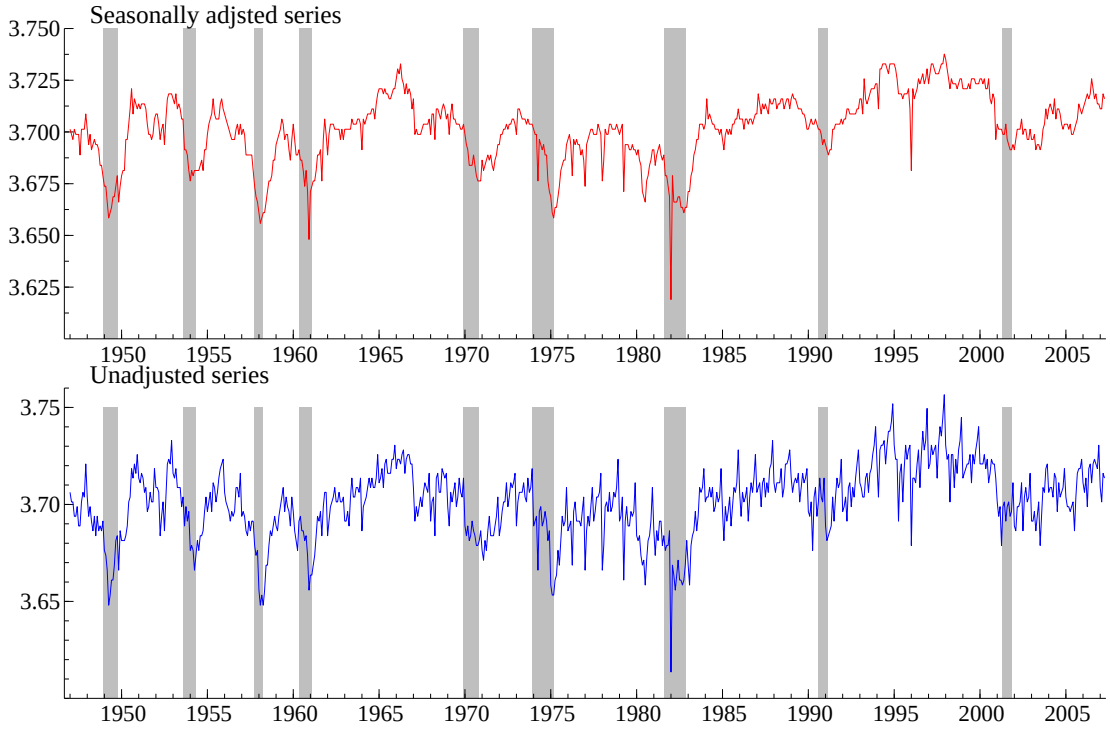


Figure 2: Seasonally adjusted AWH, linear Gaussian model: estimates of unobserved components.

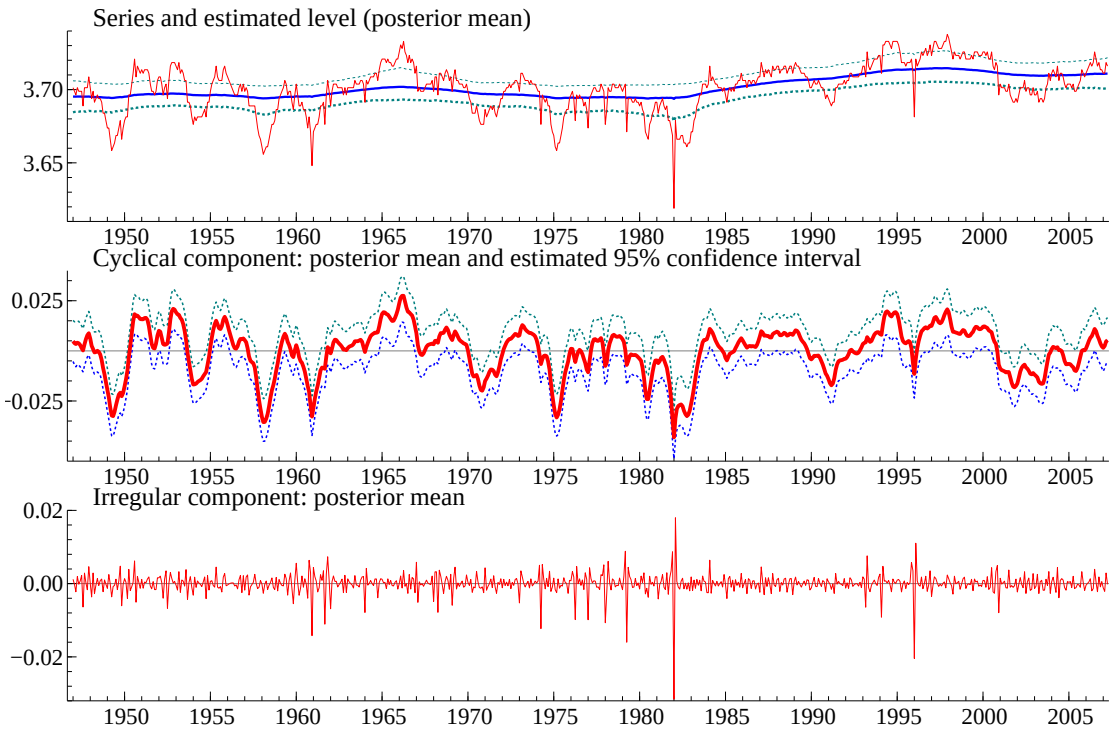


Figure 3: Seasonally adjusted AWH, model with normal mixture irregular: estimates of unobserved components.

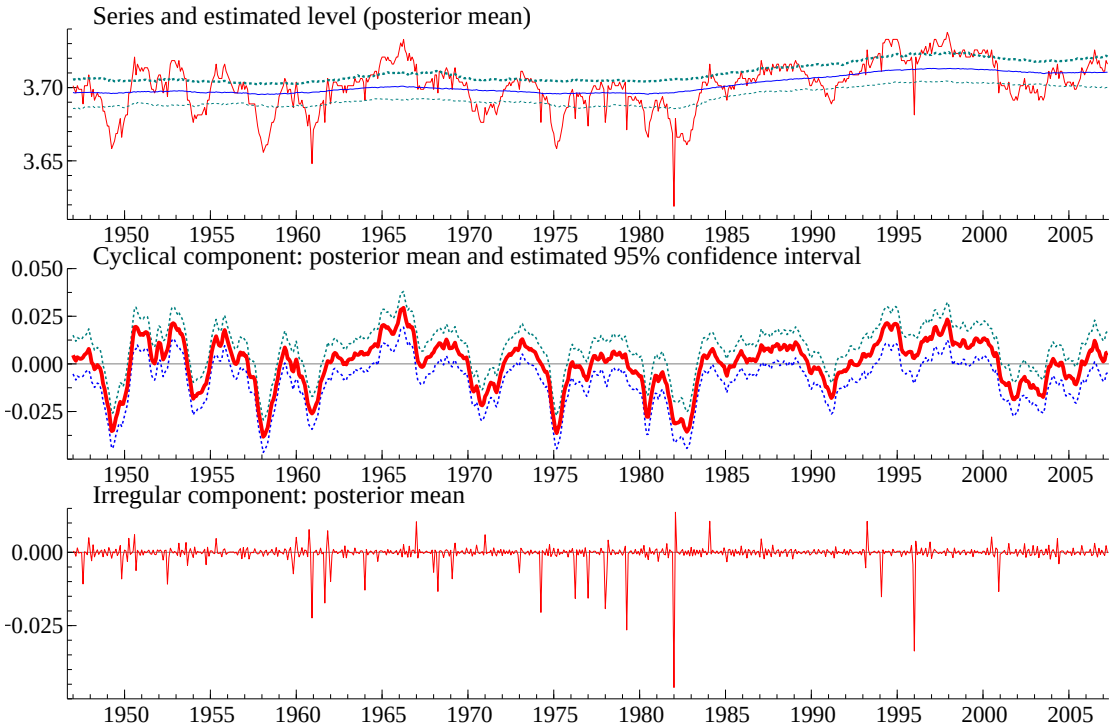


Figure 4: Seasonally adjusted AWH, model with normal mixture irregular: posterior probabilities of high variance component, $P(S_t = 1|y, \Psi)$ (top panel), and monthplot.

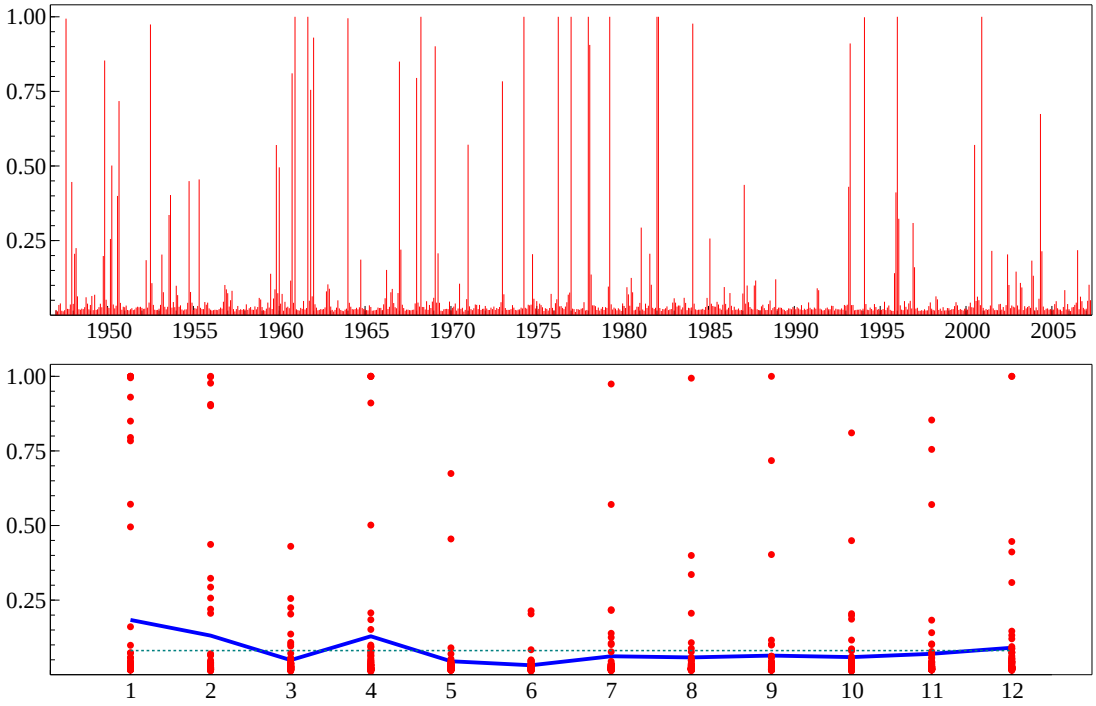
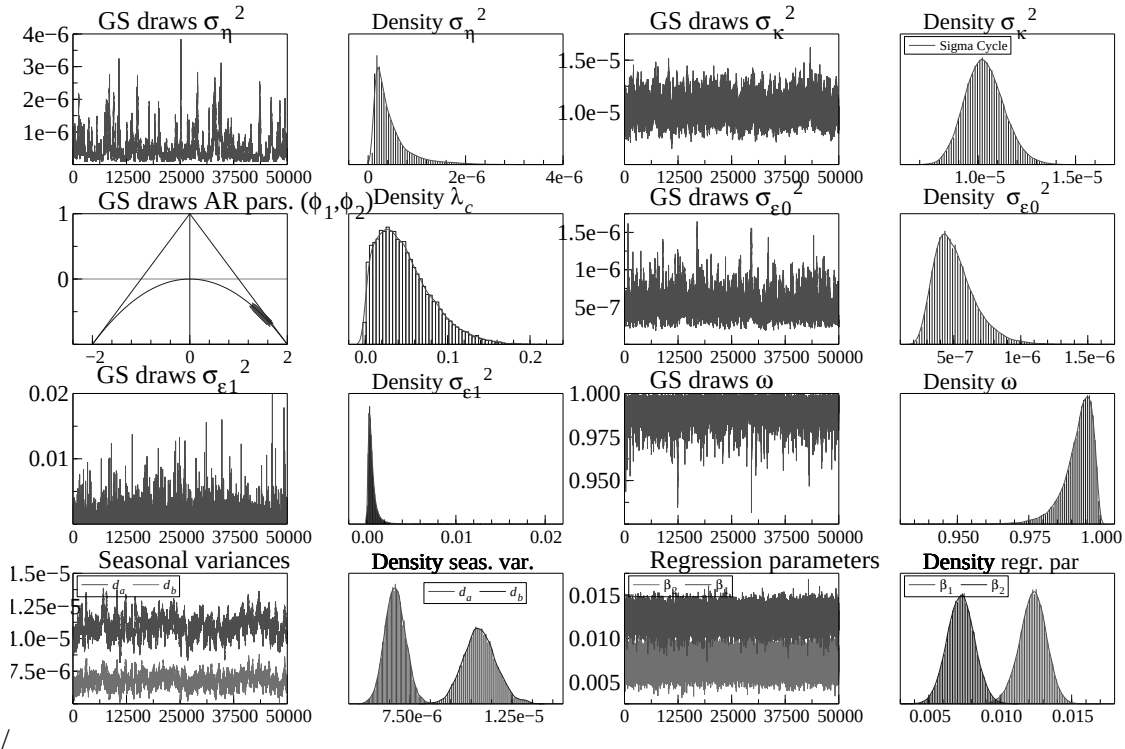


Figure 5: Unadjusted AWH series model: MCMC draws and posterior densities of the parameters.



//

Figure 6: Estimated cycle (posterior mean of ψ_t) and seasonal component (posterior mean of γ_t): posterior probabilities of high variance component, $P(S_t = 1|y, \Psi)$ (top panel), and monthplot.