DOI: 10.5604/01.3001.0014.1203

PRZEGLĄD STATYSTYCZNY
R. LXIII – ZESZYT 3 – 2016

ŁUKASZ LENART[1], BŁAŻEJ MAZUR[2]

ON BAYESIAN INFERENCE FOR ALMOST PERIODIC IN MEAN AUTOREGRESSIVE MODELS[3]

1. INTRODUCTION

In this paper we discuss Bayesian approach in case of autoregressive model with time-varying mean function. The focus is on providing an effective numerical method for posterior inference in a rather specific, highly non-linear case. Our discussion of general prior assumptions and model specification issues is therefore somewhat limited.

We make use of the idea of almost periodic time series (used in non-parametric statistics) and consider its parametric counterpart in which e.g. unconditional mean is represented by so-called Flexible Fourier Form of Gallant (1981). Models based on Fourier form with unknown set of frequency parameters are highly nonlinear and therefore difficult to estimate in case when the number of frequencies (characterizing the fluctuations) is greater than one, which is exactly the case of empirically interesting specifications.

Models of this kind are often referred to as deterministic cycle models (see for example Harvey, 2004). However, within a Bayesian approach and with non-trivial number of estimated frequencies the resulting pattern of fluctuations is quite complicated and the models can be considered competitive to stochastic cycle specifications, especially for relatively short series of data. The problems of Bayesian inference stem from the fact that the resulting posterior distribution can be multimodal and therefore difficult to explore by standard MCMC methods. One might also notice that the multimodal posterior (resulting from multimodal likelihood function) results in substantial differences between results obtained by Maximum Likelihood (ML in short) and Bayesian methods, as the multimodal posterior cannot be accurately approximated by multivariate Gaussian distribution.

Our suggestion on how to explore the posterior distribution with MCMC methods is actually two-fold. Firstly, following by the results presented in Bretthorst (1988) we

[1] Cracow University of Economics, Faculty of Finance, Department of Mathematics, 27 Rakowicka St., 31-510 Cracow, Poland, corresponding author – e-mail: lenartl@uek.krakow.pl.

[2] Cracow University of Economics, Faculty of Management, Department of Econometrics and Operations Research, 27 Rakowicka St., 31-510 Cracow, Poland.

[3] This research was supported by the Polish National Science Center based on decision number DEC-2013/09/B/HS4/01945.

make use of a non-parametrically motivated estimator to construct a proposal density for the frequency parameters. Secondly, we demonstrate how the standard conjugate results (with respect to other model parameters) can be used to reduce dimensionality of the problem. The latter step is quite interesting as it takes precisely the opposite direction compared to the usual augmentation strategy that expands the parameter space.

The remaining part of the paper has the following structure: we begin by introducing the idea of almost periodicity and recall basic results on non-parametric models with non-periodicity in mean. We also indicate some relationships between parametric Bayesian and non-parametric estimates in a very simple case. Subsequently we develop a parametric counterpart to a model representing almost-periodicity in mean which makes use of a Flexible Fourier Form. Eventually we consider two parametric models representing the process of interest. The first model, labelled "approximate" allows for taking full advantage of the standard conjugate results in Bayesian partially linear (or conditionally linear) models. In the model it is possible to obtain the kernel of marginal posterior density for frequency parameters using analytical integration only, with generates a closed-form solution (up to a normalizing constant).

However, the approximation model is not satisfactory being quite restrictive as to the way the prior information can be introduced. It does not allow for clear elaboration of prior knowledge as to the unconditional men of the process without interference with information on its autocovariance structure. Moreover, the stationarity restriction of the autoregressive part is somewhat more difficult to handle in the setup.

We therefore consider another Bayesian model based on modified parametrization, labeled "final", which is free of such inconveniences. The two Bayesian models (the "approximate" and the "final") are built upon sampling models (likelihoods) that are observationally equivalent, however only the latter has desirable overall properties. Our ultimate goal is to develop a practical MCMC algorithm for estimation of the "final" model.

We claim that the standard MCMC approaches applied to the final model are very likely to fail to explore the full posterior (and the failure is not easy to detect based just on the MCMC output). We make use of the approximate model to demonstrate the problematic structure of the posterior distribution (in particular its multimodality). The demonstration is not contaminated by possible numerical inaccuracies since it is based on analytical results.

After discussing the reasons that are likely to make the standard algorithms impractical we introduce two ideas that alleviate the problem. The first one amounts to indicating that certain non-parametric results can be used to create an efficient proposal for one group of parameters that display multimodality. The second one is based on the fact that for some other vector of parameters a standard full conditional distribution is available. The fact is often used to build a Gibbs sampler exploring the posterior, but in the case considered here such a strategy would lead to numerical inefficiency. Instead, we use the analytical results to integrate out a sub-vector of parameters from the posterior.

Our amended numerical method therefore targets a marginalized posterior kernel for a sub-vector of all the remaining model parameters. The marginalized posterior kernel is likely to be less irregular compared to the full kernel. The remaining parameters (that have been integrated out) can be sampled outside the MCMC by direct sampling, which has no negative effect on numerical efficiency. We show that using the amended algorithm for the final model we obtain the results that are in line with the analytical results from the approximate model (and the models differ only by the priors). The above problems are illustrated using both simulated and real data.

2. NON-PARAMETRIC APPROACH

The models with periodic mean or autocovariance function are broadly used in econometrics (see for example: Parzen, Pagano, 1979; Osborn, Smith, 1989; Franses, 1996; Franses, Dijk, 2005; Bollerslev, Ghysels, 1996; Burridge, Taylor, 2001; Mazur, Pipień, 2012; Lenart, Pipień, 2013a; 2013b). Formally we say that a second order real valued time series $\{Y_t\colon t \in \mathbb{Z}\}$ is periodically correlated (in short PC) if the mean function $\mu(t) = E(Y_t)$ and the autocovariance function $B(t, \tau) = \mathrm{cov}(Y_t, Y_{t+\tau})$ exists (for any $T \in \mathbb{Z}$) and are periodic functions at variable t with period T. In this parer we consider broader class, the class of almost periodically correlated time series (in short APC). In this class of time series the mean function and the autocovariance function are assumed to be almost periodic in time (see Corduneanu, 1989). This class of time series was applied in business fluctuations analysis in Lenart, Pipień (2013a) with subsampling application. Mazur, Pipień (2012) used this class of time series in modeling volatility of daily financial returns. In ACP case the mean function and the autocovariance function has Fourier representation (see for example Hurd, 1989; Hurd, 1991; Dehay, Hurd, 1994):

$$\mu(t) \sim \sum_{\varphi\in\Psi} m(\varphi)e^{i\varphi t}, \tag{1}$$

$$B(t,\tau) \sim \sum_{\lambda\in\Lambda_\tau} a(\lambda,\tau)e^{i\lambda t}, \tag{2}$$

where the Fourier coefficients $m(\varphi)$ and $a(\lambda,\tau)$ are given by the limits:

$$m(\varphi) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} \mu(t)e^{-i\varphi t},\ a(\lambda,\tau) = \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} B(j,\tau)e^{-i\lambda j}, \tag{3}$$

and the sets $\Psi = \{\varphi \in [0,2\pi)\colon |m(\varphi)| \neq 0\}$ and $\Lambda_\tau = \{\lambda \in [0,2\pi)\colon |a(\lambda,\tau)| \neq 0\}$ are countable.

In non-parametric approach the natural estimator based on sample $\{X_1, X_2,\ldots, X_n\}$ of Fourier coefficients $m(\varphi)$ has the following form

$$\widehat{m}_n(\varphi) = \frac{1}{n}\sum_{j=1}^{n} X_j e^{-ij\varphi}, \tag{4}$$

where $\varphi \in [0, 2\pi)$. As was shown in Lenart (2013) this estimator after appropriate normalizing is asymptotically normal distributed with zero mean and variance-covariance matrix that depends on spectral density function. Unfortunately in non-parametric approach the spectral density estimation is still an open problem in the case of unknown set of frequency Ψ. Therefore it is not passible to use plug in technique in statistical inference. Therefore authors use subsampling method to estimate asymptotic distribution, where knowledge about exact parameters is not necessary. An applications of the non-parametric methodology to business cycle analysis was presented by Lenart (2013) and Lenart, Pipień (2013a). Details concerning subsampling methodology in general problems are discussed e.g. by Politis et al. (1999).

In our future consideration we weaken the assumption concerning the set Ψ. For the set Ψ we assume that is finite. Therefore the equivalent representation for $\mu(t)$ takes the form:

$$\mu(t) = \delta_0 + \sum_{f=1}^{F} a_f \sin(t\varphi_f) + \sum_{f=1}^{F} b_f \cos(t\varphi_f), \tag{5}$$

where F is an unknown nonnegative integer, $\delta_0 \in \mathbb{R}$, $\mathbf{a} = (a_1, a_2, \ldots, a_F) \in \mathbb{R}^F$, $\mathbf{b} = (b_1, b_2, \ldots, b_F) \in \mathbb{R}^F$ and $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \ldots, \varphi_F) \in (0, \pi]^F$. Parameters **a** and **b** are below referred to as amplitudes, whereas elements of $\boldsymbol{\varphi}$ are labeled frequencies.

3. PARAMETRIC BAYESIAN APPROACH

In what follows we confine our attention to parametric models with time-varying unconditional mean given by (5), which (for known F) corresponds to a special case of Flexible Fourier Form discussed by Gallant (1981). One might notice that without further assumptions the parameters in (5) are not identified (due to so-called label switching). This is one source of multimodality of the joint posterior kernel and can be relatively easily eliminated by introducing a restriction of the form $0 < \varphi_1 < \varphi_2 < \ldots < \varphi_F \leq \pi$. However, here we do not impose it, though it can be easily be done in post-processing of MCMC output if desired. Our point is that there exists another source of multimodality driven by properties (5) and typical features of macroeconomic data, and it can be seen even in the case of $F = 1$, where no identification issues arise (as discussed below).

Moreover, here we do not discuss how one choses the value of F. However, within the Bayesian paradigm the models representing whole sequence with $0 < F < F_{\max}$ can be compared and the inference on regular fluctuations in mean or prediction can be based on the pooled results taking into account various values of F.

Here we assume that the deviations from the mean take the autoregressive form with J lags:

$$L(B)(y_t - \mu(t)) = \varepsilon_t, \tag{6}$$

where the function $\mu(t)$ is given by (5) and $L(B) = 1 - \eta_1 B - \eta_2 B^2 - \ldots - \eta_J B^J$ with backshift operator B: $B^k y_t = y_{t-k}$, $\{\varepsilon_t\}$ being a Gaussian (therefore strict) white noise process with precision $\tau > 0$. Notice that the observable series y_t is non-stationary in mean, though its covariance structure (under standard assumptions for coefficients of polynomial $L(B)$) corresponds to that of a covariance stationary process.

The sampling model (6) is observationally equivalent to:

$$L(B)y_t = \mu^*(t) + \varepsilon_t, \tag{7}$$

though of course in (7) $\mu^*(t) = L(B)\mu(t)$ is no longer an unconditional mean of y_t. We refer to (6) as to a final model, whereas (7) is labeled approximate.

Bretthorst (1988) has shown that in a simple case with $F = 1$ and $J = 0$ the posterior distribution of φ_1 (under uniform priors for amplitudes and Jeffreys prior for τ) can be approximated by:

$$p(\varphi_1|y) \propto \left[1 - \frac{2n|\hat{m}_n(\varphi_1)|^2}{\sum_{t=1}^{n} y_t^2}\right]^{\frac{2-n}{2}}. \tag{8}$$

It is easy to see that distribution with kernel (8) is generally a multimodal distribution. The number of modes is the same as the number of local maxima of the periodogram. The kernel (8) is a differentiable function on variable φ_1 and the derivative can be express as a product of derivative of the function $|\hat{m}_n(\varphi_1)|$ and some function with positive values on considered interval.

4. POSTERIOR DISTRIBUTION OF FREQUENCY PARAMETERS IN THE APPROXIMATE MODEL

In this section we obtain explicit formulation of marginal distribution for vector of frequency parameters in the approximate model (7) using the results based on the use of conjugate priors in conditionally linear models (see. e.g. Osiewalski, 1991). Note that the approximate model can be equivalently written as:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \tag{9}$$

where $\mathbf{y} = (y_1\ y_2\ \ldots\ y_T)'$,

$$\mathbf{X} = \begin{bmatrix} 1 & y_0 & y_{-1} & \ldots & y_{-J+1} & \sin(\varphi_1) & \cos(\varphi_1) & \sin(\varphi_2) & \cos(\varphi_2) & \ldots & \sin(\varphi_F) & \cos(\varphi_F) \\ 1 & y_1 & y_0 & \ldots & y_{-J+2} & \sin(2\varphi_1) & \cos(2\varphi_1) & \sin(2\varphi_2) & \cos(2\varphi_2) & \ldots & \sin(2\varphi_F) & \cos(2\varphi_F) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & y_{T-1} & y_{T-2} & \ldots & y_{T-J} & \sin(T\varphi_1) & \cos(T\varphi_1) & \sin(T\varphi_2) & \cos(T\varphi_2) & \ldots & \sin(T\varphi_F) & \cos(T\varphi_F) \end{bmatrix},$$

$$\beta = (\delta_0\ \eta_1\ \eta_2\ \ldots\ \eta_J\ a_1\ b_1\ a_2\ b_2\ \ldots\ a_F\ b_F)',$$

$$\varepsilon = (\varepsilon_1\ \varepsilon_2\ \ldots\ \varepsilon_T)', \varepsilon_t \sim N(0, \tau^{-1}), \text{ for } t = 1, 2, \ldots, T.$$

The $N(0,\tau^{-1})$ denotes Gaussian distribution with zero mean and variance τ^{-1}. In the above $\mathbf{X}$ depends on $\boldsymbol{\varphi}$'s being model parameters, but we suppress that to keep notation simple and to highlight relationships with standard conjugate results obtained in linear regression. Denote $\theta = (\beta, \tau, \varphi)$. Then the likelihood function has the following form:

$$p(\mathbf{y}|\theta) = \frac{1}{\sqrt{(2\pi)^n}} \tau^{\frac{n}{2}} \exp\left\{-\frac{\tau}{2}(\mathbf{y}-\mathbf{X}\beta)'(\mathbf{y}-\mathbf{X}\beta)\right\}. \quad (10)$$

Following the standard conjugate approach we assume the following prior structure:

$$p(\theta) = p(\beta,\tau)p(\boldsymbol{\varphi}) = p(\beta|\tau)p(\tau)p(\boldsymbol{\varphi}),$$

with $\beta|\,\tau \sim N(\mathbf{c},(\tau\mathbf{B})^{-1})$ and $\tau \sim G\left(\frac{n_0}{2},\frac{s_0}{2}\right)$, where $G\left(\frac{n_0}{2},\frac{s_0}{2}\right)$ denotes the Gamma distribution with expectation $\frac{n_0}{s_0}$ and variance $\frac{2n_0}{s_0^2}$ and $\mathbf{c}$, $\mathbf{B}$, n_0, s_0 are hyperparameters. This implies:

$$p(\beta|\tau) = (2\pi)^{-k/2}(\det(\mathbf{B}))^{1/2}\tau^{k/2}\exp\left\{-\frac{\tau}{2}(\beta-\mathbf{c})'\mathbf{B}(\beta-\mathbf{c})\right\},$$

$$p(\tau) = \frac{(s_0/2)^{\frac{n_0}{2}}}{\Gamma\left(\frac{n_0}{2}\right)}\tau^{\frac{n_0}{2}-1}\exp\left(-\frac{s_0\tau}{2}\right).$$

For the frequency parameter we assume uniform prior distribution:

$$p(\boldsymbol{\varphi}) = \prod_{i=1}^{F} p(\varphi_i) \text{ and } \varphi_i \sim U(0,\pi),$$

where $U(0,\pi)$ denotes uniform distribution on interval $(0,\pi)$. The above implies that:

$$p(\theta|\mathbf{y}) \propto \tau^{\frac{n+k+n_0}{2}-1}\exp\left\{-\frac{\tau}{2}[(\beta-\mathbf{d})'\mathbf{D}(\beta-\mathbf{d})]\right\}\exp\left\{-\frac{\tau}{2}[-\mathbf{d}'\mathbf{D}\mathbf{d}+\mathbf{c}'\mathbf{B}\mathbf{c}+\mathbf{y}'\mathbf{y}+s_0]\right\},$$

where $\mathbf{D} = \mathbf{X'X} + \mathbf{B}$ and $\mathbf{d} = \mathbf{D}^{-1}(\mathbf{X'y} + \mathbf{Bc})$. Integrating over β and over τ we get[4]

$$\begin{aligned} &p(\varphi|\mathbf{y}) \\ &\propto (\det(\mathbf{D}))^{-1/2}(\mathbf{y}'\mathbf{y}-\mathbf{d}'\mathbf{D}\mathbf{d}+\mathbf{c}'\mathbf{B}\mathbf{c}+s_0)^{-\frac{n+n_0}{2}} \\ &\propto (\det(\mathbf{X}'\mathbf{X}+\mathbf{B}))^{-1/2} \\ &\cdot(\mathbf{y}'\mathbf{y}-(\mathbf{X}'\mathbf{y}+\mathbf{B}\mathbf{c})'(\mathbf{X}'\mathbf{X}+\mathbf{B})^{-1}(\mathbf{X}'\mathbf{y}+\mathbf{B}\mathbf{b})+\mathbf{c}'\mathbf{B}\mathbf{c}+s_0)^{-\frac{n+n_0}{2}}. \end{aligned} \quad (11)$$

[4] Note that we assume here that the parameters η_1 η_2 … η_J are unrestricted so we do not consider stationarity restrictions due to complexity of the issue.

Assuming $\mathbf{c} = \mathbf{0}$ for simplicity we obtain the analytical solution:

$$p(\boldsymbol{\varphi}|\mathbf{y}) \propto (\det(\mathbf{X}'\mathbf{X}+\mathbf{B}))^{-1/2}(\mathbf{y}'[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X}+\mathbf{B})^{-1}\mathbf{X}']\mathbf{y}+s_0)^{-\frac{n+n_0}{2}}. \quad (12)$$

Note that distribution with kernel (12) is bounded on the set $(\varphi_1, \varphi_2, \ldots, \varphi_F) = [0,\pi)^F$, hence all posterior moments exist and it is symmetric, which follows directly from model equation (7). Unfortunately, the kernel (12) does not characterize any known distribution in the literature. In addition, contrary to the result (8) of Bretthorst (1988), the direct theoretical relation to periodogram in the case is not obvious (see distribution (12)). Hence, to illustrate the linkages between (14) and a periodogram we consider a short simulation study.

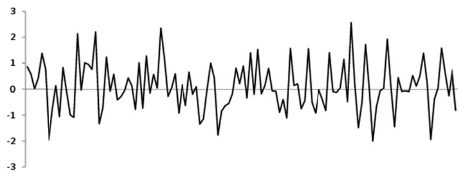
5. A SIMULATION STUDY

We restrict the attention only to the case $J = 0$ to examine the relation of (12) to usual periodogram function without additional relation to autoregressive part. We consider three cases with $F = 0, 1, 2, 3$. At each case we generate $n = 120$ realizations from considered model and we determine the distribution (12). In practice we try to choose the best F, therefore in simulation study at each case we compare the periodogram with the univariate distribution (12) (under model assumption $F = 1$) and bivariate distribution (12) (under model assumption $F = 2$). To make the results visible we use additionally the logarithmic scale. For the hyperparamiters we take $\mathbf{B}^{-1} = 100\,\mathbf{I}$, $n_0 = 2.1$ and $s_0 = 1.05$.

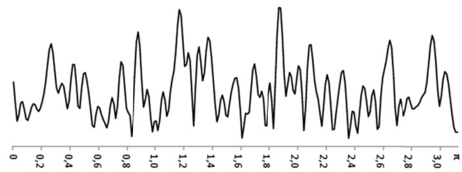
When sample is generated in the case $F = 0$ (see figure 1), the distribution (12) turns out to be multimodal under assumption of $F = 1$ and $F = 2$. Two peaks for posterior distribution (11) with $F = 1$ (see figure 1(c)) and four peaks with $F = 2$ (see figure 1(e)) correspond clearly to two dominant peaks on periodogram (see figure 1(b)).

Figure 2(a) shows a sample from model with one frequency ($\varphi_1 = 0.15$) with relatively large amplitude as compared to the variance of the white noise (see figure 2(b)). In this case the mass of probability in the posterior distribution (12) is strongly concentrated around the point where $\varphi = 0.15$ (under model assumption $F = 1$ and around sets: $0.15 \times (0, \pi)$ and $(0, \pi) \times 0.15$ (under model assumption $F = 2$).

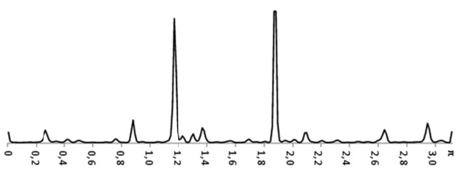
If we consider sample obtained from the model with two different frequencies ($\varphi_1 = 0.15$, $\varphi_2 = 0.5$) with different amplitudes (see figure 3(a–b)), the posterior distribution (12) with $F = 1$ (see figure 3(c)) has only one dominating peak around the frequency with larger amplitude (in this case: φ_1). The probability mass concentrated around the second frequency (φ_2) is much lower (see figure 3(d)). The posterior distribution (12) under assumption of $F = 2$ (see figure 3(e)) has two symmetric peaks that clearly correspond to points $(\varphi_1, \varphi_2) = (0.15, 0.5)$ and $(\varphi_1, \varphi_2) = (0.5, 0.15)$.



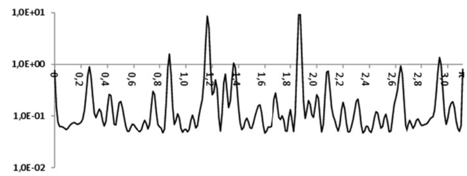
(a) Realizations from model (6)



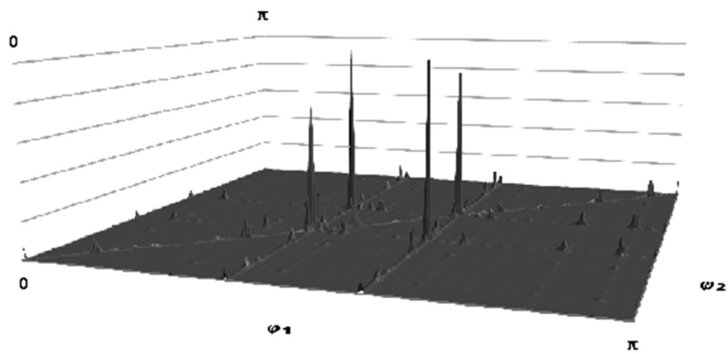
(b) The value of normalized periodogram: $|\hat{m}_n(\varphi)|$ for $\psi \in (0, \pi)$



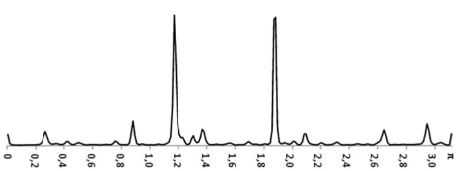
(c) Posterior distribution (12) for $\varphi \in (0, \pi)$ under model assumption $F = 1$



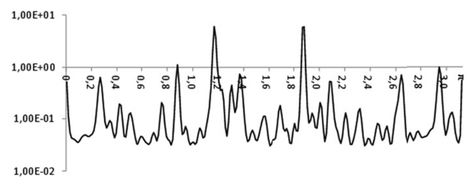
(d) Posterior distribution (12) in logarithmic scale for $\varphi \in (0, \pi)$ under model assumption $F = 1$



(e) Posterior distribution (12) for $(\varphi_1, \varphi_2) \in (0, \pi)^2$, under assumption $F = 2$



(f) Marginal posterior distribution (12) for $\varphi \in (0, \pi)$ under assumption $F = 2$

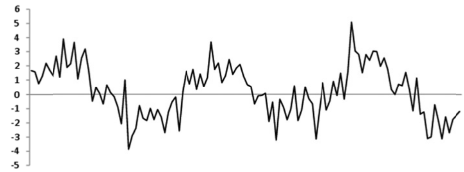


(g) Marginal posterior distribution (12) in logarithmic scale for $\varphi \in (0, \pi)$ under assumption $F = 2$

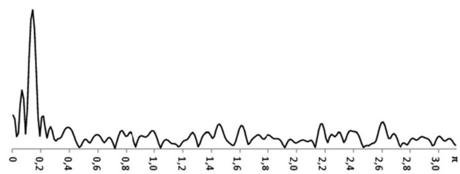
Figure 1. Posterior distributions in case of sample with length $n = 120$ generated from considered model (6) with $F = 0$ and $\tau = 1$

Source: own calculations.

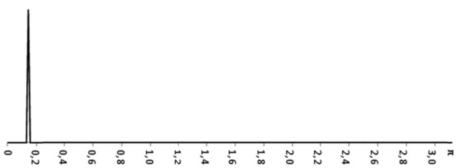
The last case, where sample is generated from model with three frequencies: $\varphi_1 = 0.15$, $\varphi_2 = 0.5$ and $\varphi_3 = 2.2$ is presented on figure 4. The amplitude for the first frequency is the biggest, while for second and third frequencies are equal (see figure 4(b)). Univariate distribution (under model assumption $F = 1$) has only one peak that clearly corresponds to frequency with the highest amplitude (φ_1). Two peaks that correspond to second and third frequency are visible only in the logarithmic scale (on marginal distribution). The bivariate distribution for frequency (under model assumption $F = 2$) has four peaks that clearly corresponds to the points (0.15, 0.5), (0.5, 0.15), (0.15, 2.2) and (2.2, 0.15).



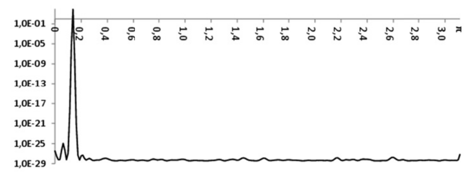
(a) Realizations from model (6)



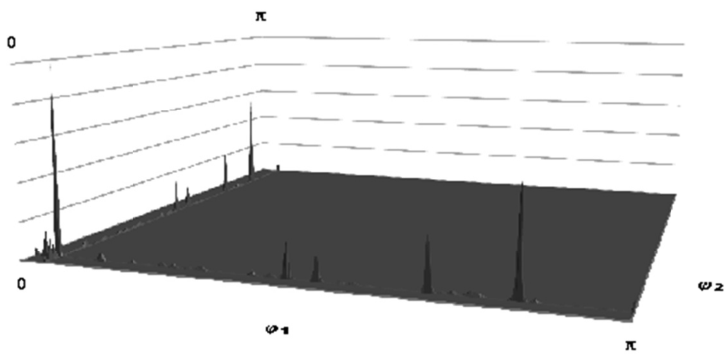
(b) The value of normalized periodogram: $|\hat{m}_n(\varphi)|$ for $\psi \in (0, \pi)$



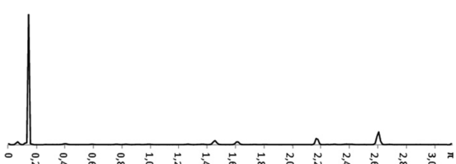
(c) Posterior distribution (12) for $\varphi \in (0, \pi)$ under model assumption $F = 1$



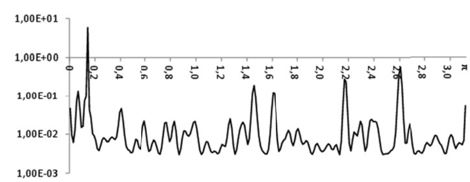
(d) Posterior distribution (12) in logarithmic scale for $\varphi \in (0, \pi)$ under model assumption $F = 1$



(e) Posterior distribution (12) for $(\varphi_1, \varphi_2) \in (0, \pi)^2$, under assumption $F = 2$



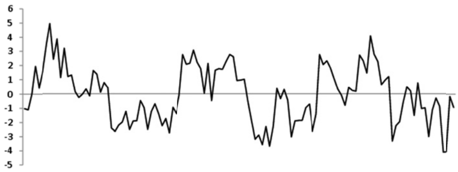
(f) Marginal posterior distribution (12) for $\varphi \in (0, \pi)$ under assumption $F = 2$



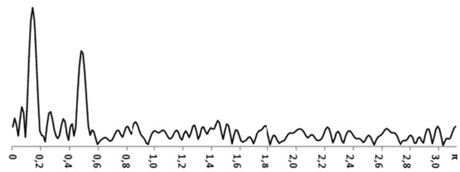
(g) Marginal posterior distribution (12) in logarithmic scale for $\varphi \in (0, \pi)$ under assumption $F = 2$

Figure 2. Posterior distributions in case of sample with length $n = 120$ generated from considered model (6) with $F = 1$, $\varphi_1 = 0.15$, $a_1 = 2$, $b_1 = 0$, $\tau = 1$

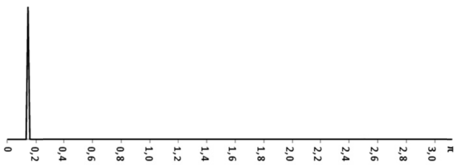
Source: own calculations.



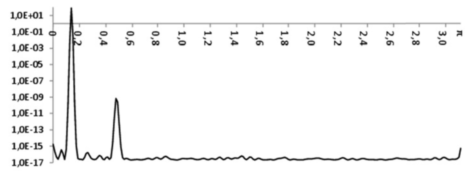
(a) Realizations from model (6)



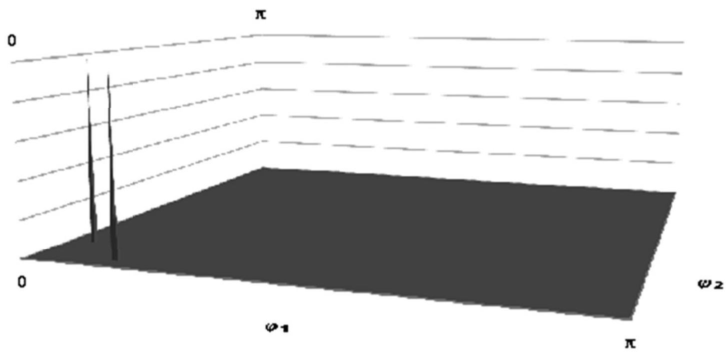
(b) The value of normalized periodogram: $|\hat{m}_n(\varphi)|$ for $\psi \in (0, \pi)$



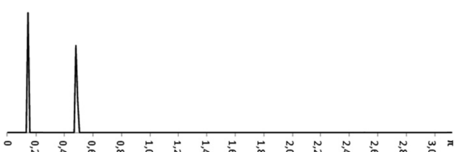
(c) Posterior distribution (12) for $\varphi \in (0, \pi)$ under model assumption $F = 1$



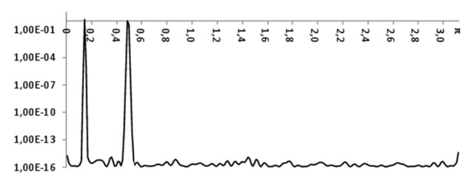
(d) Posterior distribution (12) in logarithmic scale for $\varphi \in (0, \pi)$ under model assumption $F = 1$



(e) Posterior distribution (12) for $(\varphi_1, \varphi_2) \in (0, \pi)^2$, under assumption $F = 2$



(f) Marginal posterior distribution (12) for $\varphi \in (0, \pi)$ under assumption $F = 2$

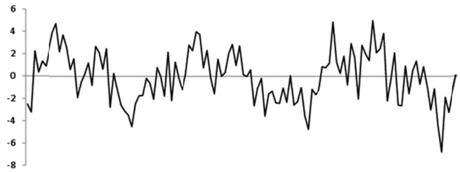


(g) Marginal posterior distribution (12) in logarithmic scale for $\varphi \in (0, \pi)$ under assumption $F = 2$

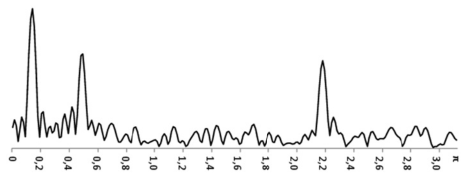
Figure 3. Posterior distributions in case of sample with length $n = 120$ generated from considered model (6) with $F = 2$, $\varphi_1 = 0.15$, $\varphi_2 = 0.5$, $a_1 = 2$, $b_1 = 0$, $a_2 = -1.5$, $b_2 = 0$, $\tau = 1$

Source: own calculations.

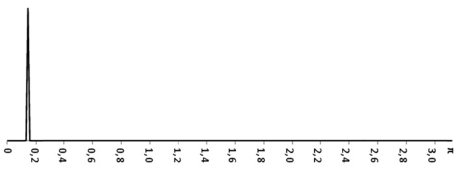
The above simulation study strongly exposes the relationship between shape of the periodogram and related posterior distributions for frequency parameters. Most importantly we demonstrate that in cases corresponding to the number of observation that characterizes typical macroeconomic applications, the resulting posterior might have highly irregular shape. Sources of the multimodality go well beyond the non-identification issue arising from the label switching. Moreover, the lack of global identification generates no theoretical problems within the Bayesian approach and can be easily resolved without any change of the MCMC algorithm discussed here.



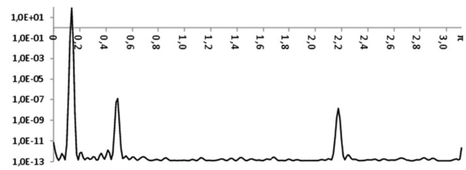
(a) Realizations from model (6)



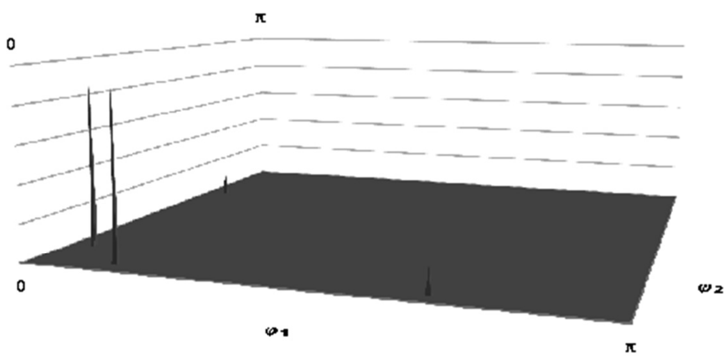
(b) The value of normalized periodogram: $|\hat{m}_n(\varphi)|$ for $\psi \in (0, \pi)$



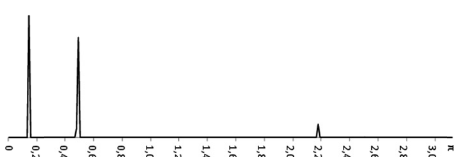
(c) Posterior distribution (12) for $\varphi \in (0, \pi)$ under model assumption $F = 1$



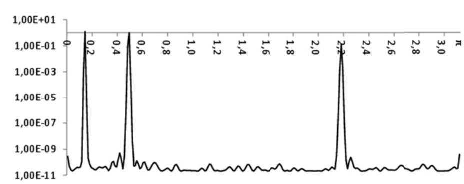
(d) Posterior distribution (12) in logarithmic scale for $\varphi \in (0, \pi)$ under model assumption $F = 1$



(e) Posterior distribution (12) for $(\varphi_1, \varphi_2) \in (0, \pi)^2$, under assumption $F = 2$



(f) Marginal posterior distribution (12) for $\varphi \in (0, \pi)$ under assumption $F = 2$



(g) Marginal posterior distribution (12) in logarithmic scale for $\varphi \in (0, \pi)$ under assumption $F = 2$

Figure 4. Posterior distributions in case of sample with length $n = 120$ generated from considered model (6) with $F = 2$, $\varphi_1 = 0.15$, $\varphi_2 = 0.5$, $\varphi_3 = 2.2$, $a_1 = 2$, $b_1 = 0$, $a_2 = -1.5$, $b_2 = 0$, $a_3 = 0$, $b_3 = 1.5$, $\tau = 1$

Source: own calculations.

6. MCMC SAMPLER FOR POSTERIOR INFERENCE

In the following discussion we assume basic knowledge of MCMC algorithm used in Bayesian inference – for an accessible review see e.g. Osiewalski (2001). An obvious approach to Bayesian estimation of the final model would be to sample from the full posterior $p(\theta|\mathbf{y})$ using a Gibbs sampler. Such a sampler would be based on factorization of $p(\theta|\mathbf{y})$ into full conditionals for sub-vectors of θ, of which at least some have a standard form (as conjugate-type priors are used). In particular a sampler consisting of four steps, for linear parameters of the mean $(\delta_0, a_1, a_2, \ldots, a_F, b_1, b_2, \ldots, b_F)$, τ, $\boldsymbol{\eta} = (\eta_1\ \eta_2\ \ldots\ \eta_J)$ and $\boldsymbol{\varphi}$ respectively, is an obvious solution. However, we point out that two difficulties would arise. Firstly, one would need a good proposal for the frequency parameters sampled within a Mertopolis-Hastings (M-H in short) step, as the full conditional posterior is definitely not a standard one in this case. Secondly, even after addressing that, such a sampler could fail to achieve convergence to true posterior within a finite and practical timespan. This is because in practical cases the joint posterior would be multimodal and a move from one mode to another would require a change in parameters belonging to two separate Gibbs blocks (namely frequencies and amplitudes). Under fairly weak conditions such a change has a very low chance and this arises just from the conditioning inherent in such a sampler, which therefore would fail to visit all the relevant modes.[5]

In order to solve the issue we first introduce the idea of posterior marginalization. Consider the following general factorization of a posterior distribution:

$$p(\theta|\mathbf{y}) = p(\theta^{(1)} \,|\, \mathbf{y},\theta^{(2)})p(\theta^{(2)} \,|\, \mathbf{y}) \propto k(\theta^{(1)} \,|\, \mathbf{y},\theta^{(2)})k(\theta^{(2)} \,|\, \mathbf{y}) = k(\theta|\mathbf{y}),$$

where $\theta' = [\theta^{(1)\prime}\ \theta^{(2)\prime}]$ We assume that $p(\theta^{(1)} \,|\, \mathbf{y},\theta^{(2)})$ represents full conditional posterior for $\theta^{(1)}$ that has a known form. Its kernel is $k(\theta^{(1)} \,|\, \mathbf{y},\theta^{(2)})$ and the normalizing constant is known (and depends on $\theta^{(2)}$). Consequently, $\theta^{(1)}$ can be integrated out from the posterior using analytical techniques, resulting in a closed form of marginal posterior kernel for $\theta^{(2)}$ only. The resulting marginal kernel usually retains all the terms from $k(\theta|\mathbf{y})$ not included in $k(\theta^{(1)} \,|\, \mathbf{y},\theta^{(2)})$ and inverse of the normalizing constant of $p(\theta^{(1)} \,|\, \mathbf{y},\theta^{(2)})$ being a function of $\theta^{(2)}$.

By the virtue of marginalization, $p(\theta^{(2)} \,|\, \mathbf{y})$ is likely to have more regular shape compared to $p(\theta|\mathbf{y})$. Essentially, we aim to improve properties of the MCMC algorithm by adjusting its target distribution, replacing $p(\theta|\mathbf{y})$ with $p(\theta^{(2)} \,|\, \mathbf{y})$ being potentially more regular. Of course finally we draw from $p(\theta|\mathbf{y})$, but this can be achieved by

[5] One might imagine the example of sampling from a bivariate target (with one variable in each Gibbs step) when the target distribution is a mixture of two bivariate normal densities with modes that are separated in both dimensions and variances that are small relative to the difference in modes. Conditionally on MCMC chain visiting one mode, a move to the other one would require occurrence of a very particular tail event in the first step.

additional direct sampling since $p(\theta^{(1)} \mid \mathbf{y}, \theta^{(2)})$ has known, standard form. In the final model (6), $\theta^{(1)}$ corresponds to parameters that appear in the unconditional mean (5) linearly, i.e. $\delta_0, a_1, a_2, \ldots, a_F, b_1, b_2, \ldots, b_F$. The resulting marginal posterior kernel is $p(\theta^{(2)} \mid \mathbf{y})$ where $\theta^{(2)}$ includes $\boldsymbol{\eta} = (\eta_1\ \eta_2\ \ldots\ \eta_J)$, τ and φ and it does not include amplitude parameters, as these are integrated out.

In order to sample from $p(\theta^{(2)} \mid \mathbf{y})$ one might construct another Gibbs sampler with M-H steps for $\boldsymbol{\eta} = (\eta_1\ \eta_2\ \ldots\ \eta_J)$, τ and $\boldsymbol{\varphi}$ (with stationary restrictions imposed on $\boldsymbol{\eta}$). Here again a crucial problem to be solved would be the one of sampling frequency parameters $\boldsymbol{\varphi}$. We suggest using a M-H step with a proposal density being a product of identical (normalized) magnitude of periodogram functions (4) restricted to interval of interest for the frequencies. The one-dimensional problem with finite support can be handled numerically in an effective way (using one-dimensional numerical representation of the univariate density generated with an arbitrary precision). The distribution would allow for a simple design of a M-H step with an independent proposal.

7. REAL DATA EXAMPLE

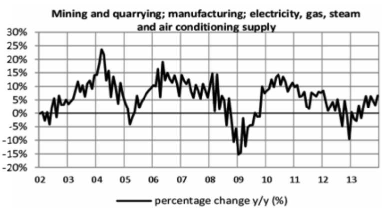
In this section we consider two data sets from the Polish economy concerning growth rates of monthly production in industry (percentage change compared to corresponding period of the previous year, y-o-y in short): *Mining and quarrying; manufacturing; electricity, gas, steam and air conditioning supply* and *MIG – Non-durable consumer goods*[6]. The samples start at January 2002 and end at December 2013. These two economic processes belong to main cyclical indicators of economy. In comparison analysis we take $J = 0$ and the same prior distributions as in section 4, since under the assumptions our “approximate” and “final” formulations coincide exactly, therefore the MCMC output[7] (see detailed algorithm in the previous section) can be compared to an exact, analytical benchmark (12).

The two real data examples show that in the bivariate case the shape of the distribution for the frequency parameters obtained by proposed MCMC sampler is comparable with theoretical distribution (12) (see results on figures 5–6). This demonstrates the efficiency of the sampler proposed here in a real data example. For extreme cases with really high number of unknown frequencies (that are unlikely to be encountered within macroeconomic applications) the approach could be refined by taking a proposal for frequency parameters based on the marginal posterior (12) obtained analytically from the approximate model (this would also require a numerical approximation of a marginalized univariate version of (12) instead of periodogram).

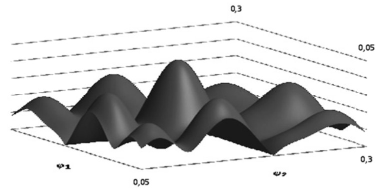
---

[6] Source: Eurostat.

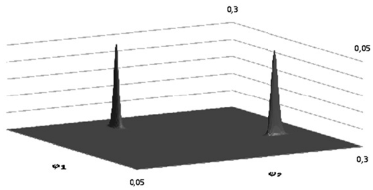
[7] 50 000 burn-in cycles and 1 000 000 final cycles.



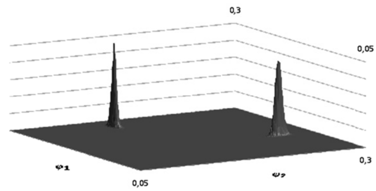
The data



The value of $|\hat{m}_n(\varphi_1)\ \hat{m}_n(\varphi_2)|$ – proposal kernel of distribution in Metropolis-Hastings step



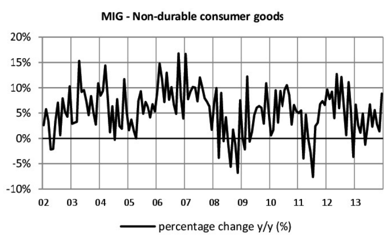
Analytical marginal posterior for frequency parameters (12) under assumption $F = 2$



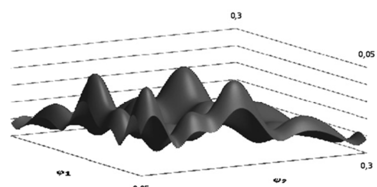
Histogram (marginal posterior) for frequency parameters based on MCMC sample from the algorithm proposed in the paper

Figure 5. The comparison of posterior distribution (12) with obtained MCMC sample

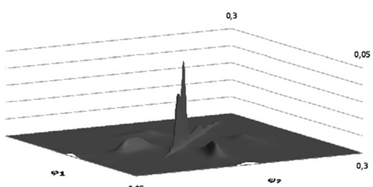
Source: own calculations.



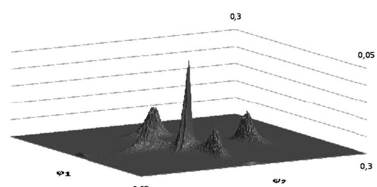
The data



The value of $|\hat{m}_n(\varphi_1)\ \hat{m}_n(\varphi_2)|$ – proposal kernel of distribution in Metropolis-Hastings step



Analytical marginal posterior for frequencies (12) under assumption $F = 2$



Two dimensional histogram (marginal posterior) for frequency based on MCMC sample from the algorithm proposed in the paper

Figure 6. The comparison of posterior distribution (12) with obtained MCMC sample

Source: own calculations.

8. CONCLUSIONS

In the paper we highlight some problems that arise in Bayesian estimation of parametric time-series model with fluctuations (corresponding to e.g. business cycle) are modelled using Flexible Fourier Form of Gallant (1981). The problems appear in empirically appealing cases with more than one unknown frequency parameter. We demonstrate that the resulting posterior is likely to be highly multimodal. This cast doubts on applicability of ML estimation, but can also result in problems within the Bayesian approach, as standard MCMC methods might fail to explore the whole posterior, especially when the modes are separated.

We demonstrate that the multimodality is actually an issue using the exact solution (i.e. an analytical marginal posterior) in an approximate model. The approximate model differs from our target (final) specification by the prior assumptions only. The posterior multimodality seems to be most severe within the joint space of amplitude and frequency parameters.

We address that problem using two essential steps. Firstly, we integrate the posterior with respect to amplitude parameters, which can be carried out analytically. Secondly, we propose a non-parametrically motivated proposal for the frequency parameters. This allows for construction of an improved MCMC sampler that effectively explores the space of all the model parameters, with the amplitudes sampled by the direct approach outside the MCMC chain.

Using the improved algorithm we are able to estimate our target specification which allows prior information to be introduced in a reasonable way: parameters characterizing unconditional mean are separated from those describing autocovariances. In particular one can express prior knowledge on possible amplitudes of regular fluctuations in mean (by setting prior precision of amplitude parameters) or cycle length (by specifying φ_L and φ_U). The approach can be therefore used to "filter" cyclical fluctuations characterized by cycle lengths within a given range. The "extracted" pattern of regular fluctuations would be described by posterior distribution of unconditional mean (as a function of model parameters given by (5)).

Moreover, the causality restriction can be imposed on autoregressive parameters (so that e.g. explosive paths are ruled out *a priori*). In our experience the approach is feasible even with quite high lag order of the autoregressive process.

REFERENCES

Bollerslev T., Ghysels E., (1996), Periodic Autoregressive Conditional Heteroscedasticity, *Journal of Business and Economic Statistics*, 14 (2), 139–152.

Bretthorst G. L., (1988), *Bayesian Spectrum Analysis and Parameter Estimation*, Springer-Verlag, Germany.

Burridge P., Taylor A. M., (2001), On Regression-based Tests for Seasonal Unit Roots in the Presence of Periodic Heteroscedasticity, *Journal of Econometrics*, 104, 91–117.

Corduneanu C., (1989), *Almost Periodic Functions*, Chelsea, New York.

Dehay D., Hurd H., (1994), Representation and Estimation for Periodically and Almost Periodically Correlated Random Processes, in: Gardner W. A., (ed.), *Cyclostationarity in Communications and Signal Processing, IEEE Press*, 295–329.

Franses P. H., Dijk D., (2005), The Forecasting Performance of Various Models for Seasonality and Nonlinearity for Quarterly Industrial Production, *International Journal of Forecasting*, 21, 87–102.

Franses P. H., (1996), *Stochastic Trends in Economic Time Series*, Oxford University Press, New York.

Gallant A. R., (1981), On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Flexible Fourier Form, *Journal of Econometrics*, 15, 211–245.

Harvey A., (2004). Tests for cycles, in: Harvey A. C., Koopman S. J., Shephard N., (eds.), *State space and unobserved component models, CUP*, pages 102–119.

Hurd H., (1989), Representation of Strongly Harmonizable Periodically Correlated Process and their Covariances, *Journal of Multivariate Analysis*, 29, 53–67.

Hurd H., (1991), Correlation Theory of Almost Periodically Correlated Processes, *Journal of Multivariate Analysis*, 37, 24–45.

Lenart Ł., (2013), Non-parametric Frequency Identification and Estimation in Mean Function for Almost Periodically Correlated Time Series, *Journal of Multivariate Analysis*, 115, 252–269.

Lenart Ł., Pipień M., (2013a), Almost Periodically Correlated Time Series in Business Fluctuations Analysis, *Acta Physica Polonica A*, 123 (3), 567–583.

Lenart Ł., Pipień M., (2013b), Seasonality Revisited - Statistical Testing for Almost Periodically Correlated Processes, *Central European Journal of Economic modelling and Econometrics*, 5, 85–102.

Mazur B., Pipień M., (2012), On the Empirical Importance of Periodicity in the Volatility of Financial Returns - Time Varying GARCH as a Second Order APC(2) Process, *Central European Journal of Economic modelling and Econometrics*, 4, 95–116.

Osborn D. R., Smith J. P., (1989), The Performance of Periodic Autoregressive Models in Forecasting Seasonal U.K. Consumption, *Journal of Business & Economic Statistics*, 9, 117–127.

Osiewalski J., (1991), Bayesowska estymacja i predykcja dla jednorównaniowych modeli ekonometrycznych, *Zeszyty Naukowe / Akademia Ekonomiczna w Krakowie. Seria Specjalna, Monografie nr 100.*

Osiewalski J., (2001), *Ekonometria bayesowska w zastosowaniach*, wyd. AE w Krakowie.

Parzen E., Pagano M., (1979), An Approach to Modeling Seasonally Stationary Time Series, *Journal of Econometrics*, 9, 137–153.

Politis D., Romano J., Wolf M., (1999), *Subsampling*, Springer-Verlag, New York.

WNIOSKOWANIE BAYESOWSKIE DLA ZMIENNEJ W CZASIE PRAWIE OKRESOWEJ FUNKCJI WARTOŚCI OCZEKIWANEJ W MODELU AUTOREGRESJI

Streszczenie

Artykuł ma na celu przedstawienie problematyki bayesowskiej estymacji klasy jednowymiarowych modeli dla danych charakteryzujących się występowaniem skomplikowanych wahań cyklicznych w średniej. Koncentrujemy się na zagadnieniach powstających w estymacji parametrycznych modeli dla szeregów czasowych wykorzystujących tzw. giętką formę Fouriera (Flexible Fourier Form, zob. Gallant, 1981), której parametry opisują amplitudę i częstotliwość wahań. Wskazujemy, iż w takich modelach łączny rozkład *a posteriori* charakteryzuje się silną wielomodalnością, przez co standardowe metody numeryczne typu MCMC mogą okazać się raczej zawodnym narzędziem wnioskowania. Ma to miejsce, gdy próbnik MCMC nie odwiedza (w praktyce) wszystkich modalnych badanego rozkładu. Wykorzystując dokładne rozwiązanie analityczne w bardzo zbliżonym modelu wykazujemy, iż wzmiankowana wielomodalność faktycznie ma miejsce. Proponujemy dwa rozwiązania szczegółowe. Po pierwsze wycałkowujemy analitycznie z rozkładu *a posteriori* parametry odpowiadające za amplitudę wahań. Po drugie przedstawiamy specjalnie dobrany rozkład proponujący dla parametrów częstotliwości wyspecyfikowany z wykorzystaniem wyników otrzymanych na gruncie podejścia nieparametrycznego. Tak otrzymany próbnik MCMC w ramach praktycznie użytecznej liczby losowań jest w stanie skutecznie przemieszczać się w (zredukowanej) przestrzeni parametrów. Wycałkowane parametry są dolosowywane poza algorytmem MCMC poprzez losowanie bezpośrednie ze standardowego rozkładu warunkowego. Ilustrujemy omawianą problematykę wykorzystując dane symulacyjne a także dwa przykłady danych rzeczywistych.

Słowa kluczowe: wnioskowanie bayesowskie, funkcja prawie okresowa wartości oczekiwanej, model autoregresji, próbnik MCMC

ON BAYESIAN INFERENCE FOR ALMOST PERIODIC
IN MEAN AUTOREGRESSIVE MODELS

Abstract

The goal of the paper is to discuss Bayesian estimation of a class of univariate time-series models being able to represent complicated patterns of “cyclical” fluctuations in mean function. We highlight problems that arise in Bayesian estimation of parametric time-series model using the Flexible Fourier Form of Gallant (1981). We demonstrate that the resulting posterior is likely to be highly multimodal, therefore standard Markov Chain Monte Carlo (MCMC in short) methods might fail to explore the whole posterior, especially when the modes are separated. We show that the multimodality is actually an issue using the exact solution (i.e. an analytical marginal posterior) in an approximate model. We address that problem using two essential steps. Firstly, we integrate the posterior with respect to amplitude parameters, which can be carried out analytically. Secondly, we propose a non-parametrically motivated proposal for the frequency parameters. This allows for construction of an improved MCMC sampler that effectively explores the space of all the model parameters, with the amplitudes sampled by the direct approach outside the MCMC chain. We illustrate the problem using simulations and demonstrate our solution using two real-data examples.

Keywords: Bayesian inference, almost periodic mean function, autoregressive model, MCMC sampler