Journal of Statistical and Econometric Methods, vol.4, no.3, 2015, 1-16
ISSN: 2241-0384 (print), 2241-0376 (online)
Scienpress Ltd, 2015

Imposing Frequency-Domain Restrictions on Time-Domain Forecasts

Erhard Reschenhofer[1] and Marek Chudy[2]

Abstract

This paper proposes a new model selection criterion for choosing the number of discontinuity points in piecewise constant frequency-domain models for stationary time series. In order to facilitate the use of this criterion in practice, penalties are calculated for various levels of complexity and sample sizes using an efficient algorithm which is based on the principle of dynamic programming. Moreover, it is shown how the selected frequency-domain model can be used to estimate in a first step the autocovariances via their spectral representation and then, in a second step, also the parameters of autoregressive models via the Durbin-Levinson algorithm. In an empirical study with macroeconomic data, the forecasts based on these restricted autoregressive models strikingly outperform conventional ARMA forecasts.

Mathematics Subject Classification: 62M10; 62M15; 62M20

Keywords: Forecasting; Step function; Spectral density; Model selection

---

[1] Department of Statistics and Operations Research, University of Vienna, Austria.
E-mail: erhard.reschenhofer@univie.ac.at

[2] Department of Statistics and Operations Research, University of Vienna, Austria.
E-mail: marek.chudy@univie.ac.at

Article Info: *Received* : April 27, 2015. *Revised* : May 26, 2015.
Published online : September 1, 2015.

1 Introduction

Like in the time domain, where abnormal time periods are usually excluded from the analysis, it makes also sense in the frequency domain to focus on certain frequency bands and disregard others. In the case where the relationship between variables depends on the frequency, the method of band-spectrum regression (see [1]-[2]) can be used (e.g. [3]-[6]). Of particular interest in this context are spectra which exhibit breaks. Taniguchi [7] developed the asymptotic estimation theory for piecewise continuous spectra. However, even in the simplest case of piecewise constant spectra (see [8]), the determination of the number of discontinuity points is still an unresolved problem. Just as in the case of a simple autoregressive model, there is one integer-valued parameter *K* determining the complexity of the model and *K* real-valued parameters that can be used to optimize the fit to the data. But there is one important difference. In the case of the step function, there are not only the *K* real-valued parameters determining the heights of the steps but additionally also $K-1$ integer-valued parameters determining the subsets of Fourier frequencies where the step function is constant. Like in the time domain, where we assume that structural breaks occur only at time points where observations are made, we assume that the jump discontinuities of the step function occur only at those frequencies where our frequency-domain data are observed. We do not really care what happens between two successive time-domain observations y_{t-1} and y_t, $t = 2,...,n$, or between two successive frequency-domain observations $I(\omega_{k-1})$ and $I(\omega_k)$, $k = 2,...,m = [(n-1)/2]$, where

$$I(\omega_k) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} y_t e^{i\omega_k t}\right|^2 \qquad (1)$$

is the value of the periodogram at the *k*th Fourier frequency $\omega_k = 2\pi k / n$.

It is a priori not clear how the location parameters should be penalized. Ninomiya's [9] suggestion that the penalty of an integer-valued location parameter

should be three times as large as that of a regular real-valued parameter is based on asymptotic arguments and the critical assumption that the number of breaks/steps is fixed and does not increase as the sample size increases. Similarly restrictive assumptions have been used for the derivation of consistent estimators for the number of jump discontinuities of a step function (e.g. [10]-[12]). Unfortunately, assumptions of this type are implausible in most applications. For example, economic time series typically exhibit structural breaks which occur every few years or decades.

A probably more promising approach is to eliminate all integer-valued location parameters by reducing the problem of determining the number of steps to the problem of selecting a suitable submodel of the linear regression model

$$\begin{pmatrix} I(\omega_1) \\ I(\omega_2) \\ \vdots \\ I(\omega_m) \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}}_{X} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} + \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix}. \tag{2}$$

For the latter task, several subset selection criteria are available (e.g. [13]-[16]). Unfortunately, these criteria have been derived under quite restrictive assumptions, including that of normality, and can therefore not be applied to the frequency-domain observations $I(\omega_1)$,..., $I(\omega_m)$ which approximately have independent exponential distributions with means $f(\omega_1),..., f(\omega_m)$, where f is the spectral density of the stationary process *y*. Any subset selection criterion that does not take into account the fact that extreme observations are much more likely in exponential samples than in normal samples would inevitably overestimate the number of steps and possibly even waste separate steps for individual outliers.

The next section therefore designs a new subset selection criterion for exponential samples and provides a table of penalties which have been calculated with the help of an efficient algorithm based on the principle of dynamic programming (see [17]). In Section 3, this criterion is used for choosing

parsimonious frequency-domain models. Associated time-domain models are obtained from estimates of the autocovariances implied by the selected frequency-domain models. Section 4 compares the forecasting performance of these restricted time-domain models with that of conventional ARMA models. Section 5 concludes.

2 A Subset-Selection Criterion for Exponential Samples

In the simplest case, the time-domain observations $y_1,...,y_n$ are i.i.d. N$(0,\sigma^2)$ and the frequency-domain observations $I(\omega_1)$,..., $I(\omega_m)$ are therefore i.i.d. Exp$(\sigma^2/(2\pi))$. Hence,

$$\underbrace{\begin{pmatrix} I(\omega_1) \\ I(\omega_2) \\ \vdots \\ I(\omega_m) \end{pmatrix}}_{\Im} = \frac{\sigma^2}{2\pi}\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}}_{X}\begin{pmatrix} \frac{\sigma^2}{2\pi} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \frac{\sigma^2}{2\pi}\begin{pmatrix} e_1 - 1 \\ e_2 - 1 \\ \vdots \\ e_m - 1 \end{pmatrix}, \tag{3}$$

where $e_1,...,e_m$ are i.i.d. Exp(1). In this case, the spectral density *f* is constant and therefore only the first column of the design matrix *X* is needed. However, if *f* can be adequately described by a piecewise constant function and both the number and the location of the discontinuity points is unknown, a suitable submatrix of *X* must be selected.

Let $X(S)$ denote that submatrix of *X*, the columns of which are determined by the proper subsequence $S=(S_1,...,S_K)$ of $(1,...,m)$. For any fixed *S*,

$$FPE(S) = \frac{m+K}{m-K}\left\| \Im - X(S)\underbrace{\left(X(S)'X(S)\right)^{-1}X(S)'\Im}_{\hat{\lambda}(S)}\right\|^2 \tag{4}$$

is an unbiased estimator of the mean squared prediction error

$$E\left\| \mathfrak{I}^* - \underbrace{X(S)\hat{\lambda}(s)}_{\hat{\mathfrak{I}}(S)} \right\|^2, \tag{5}$$

where $\mathfrak{I}^*$ is an independent sample which has the same distribution as $\mathfrak{I}$. According to Rothman [18] and Akaike [19] that *S* should be selected which minimizes $FPE(S)$. In our case, it is required that $S_1 = 1$ because Kolmogorov's formula

$$\sigma^2 = 2\pi \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log(f(\omega))\, d\omega \right) \tag{6}$$

would imply a vanishing innovation variance if the spectral density were zero on an interval.

The data-snooping bias of the naive estimator

$$E\left\| \mathfrak{I} - \hat{\mathfrak{I}}(S) \right\|^2 \tag{7}$$

of the mean squared prediction error (5) will clearly be much larger if *S* is not fixed but is rather found by minimization over all subsequences of length *K*. In this case, the criterion

$$FPE^*_{sub}(\hat{S}(K)) = \frac{m+1+\zeta_1(K-1,m-1)}{m-1-\zeta_1(K-1,m-1)} \left\| \mathfrak{I} - \hat{\mathfrak{I}}(\hat{S}(K)) \right\|^2, \tag{8}$$

where $\zeta_1(K,m)$ is the expected value of the sum of the *K* largest of *m* independent $\chi^2(1)$ -variables, would be more appropriate if the data were normally distributed and the regressors were orthogonal (see [20]). But since neither of these two assumptions is satisfied in (2), appropriate penalty factors for each *K* are obtained as

$$\frac{\sum_{j=1}^{r} \left\| \mathfrak{I}_j^* - \hat{\mathfrak{I}}_j(\hat{S}(K)) \right\|^2}{\sum_{j=1}^{r} \left\| \mathfrak{I}_j - \hat{\mathfrak{I}}_j(\hat{S}(K)) \right\|^2} \tag{9}$$

or, computationally more efficiently, as

$$P(K) = \frac{\sum_{j=1}^{r}(m + \left\|\hat{\Im}_j(\hat{S}(K))\right\|^2)}{\sum_{j=1}^{r}\left\|\Im_j - \hat{\Im}_j(\hat{S}(K))\right\|^2}, \tag{10}$$

where $\Im(j)$, $\Im^*(j)$, j=1,…,r, are independent samples of size m from a standard exponential distribution and $\hat{\Im}_K(j)$ is the best fit for $\Im(j)$ among all $\{1\}\subseteq S$ of size K. The efficient algorithm in [17] is used for the calculation of the penalty factors. Table 1 gives the increments $P_a(K)$–$P_a(K\text{-}1)$, K=2,…,10, of the additive penalties

$$P_a(K) = n\log(P(K)) \tag{11}$$

for m=20,30,40,…,250. Each table entry is based on r=100,000 random samples generated with the software R (see [21]).

At first sight, the non-monotonicity of the increments in the penalties is surprising because conventional criteria penalize new regressors to be included in a model either in the same way (e.g., AIC and BIC) or milder (e.g., MRIC and FPE-sub) than already included variables. However, Reschenhofer et al. [16], who investigated structural breaks in time-domain models, argued that clusters of unusual observations will not always occur just at the begin or at the end of the observation period but rather somewhere in the middle. In the latter case, two breaks are required for the description of each cluster. Consequently, the penalties for the second, fourth, and sixth break should be higher than those for the first, third and fifth break, respectively.

Table 1: Increments $P_a(K)-P_a(K\text{-}1)$, $K=2,\ldots,10$, of the additive penalties P_a for different sample sizes m. Each value is based on 100,000 random samples of size m from a standard exponential distribution.

K	2	3	4	5	6	7	8	9	10
m									
20	6.3	12.6	5.8	7.8	5.9	6.7	6.2	6.6	6.7
30	6.6	15.4	6.2	9.1	6.1	7.2	6.0	6.6	6.1
40	6.9	17.6	6.6	10.3	6.3	8.0	6.2	7.0	6.2
50	7.2	19.5	6.8	11.3	6.6	8.7	6.4	7.4	6.3
60	7.3	21.2	7.1	12.3	6.9	9.4	6.6	7.9	6.5
70	7.5	22.7	7.3	13.2	7.1	10.0	6.9	8.4	6.7
80	7.6	24.0	7.5	14.0	7.3	10.6	7.1	8.8	6.9
90	7.7	25.3	7.6	14.9	7.4	11.2	7.2	9.3	7.0
100	7.9	26.5	7.8	15.6	7.6	11.7	7.4	9.7	7.2
110	7.9	27.6	7.9	16.3	7.8	12.2	7.6	10.1	7.4
120	8.0	28.6	8.1	16.9	7.9	12.7	7.7	10.5	7.5
130	8.0	29.5	8.2	17.6	8.1	13.3	7.9	10.9	7.7
140	8.1	30.4	8.3	18.2	8.2	13.7	8.0	11.3	7.8
150	8.2	31.2	8.4	18.8	8.3	14.1	8.1	11.6	7.9
160	8.2	31.9	8.5	19.3	8.4	14.6	8.2	12.0	8.0
170	8.2	32.8	8.5	19.8	8.5	15.0	8.3	12.3	8.2
180	8.3	33.5	8.7	20.3	8.6	15.3	8.5	12.6	8.3
190	8.4	34.2	8.8	20.8	8.7	15.8	8.6	13.0	8.4
200	8.4	35.0	8.8	21.3	8.8	16.1	8.7	13.3	8.5
210	8.4	35.5	8.8	21.7	8.9	16.5	8.8	13.6	8.6
220	8.4	36.1	8.9	22.2	9.0	16.8	8.8	13.9	8.7
230	8.5	36.7	9.0	22.6	9.0	17.2	8.9	14.2	8.8
240	8.5	37.4	9.0	23.0	9.1	17.5	9.0	14.5	8.9
250	8.5	37.8	9.1	23.4	9.2	17.8	9.1	14.7	9.0

3 Obtaining Restricted Time-Domain Models from Frequency-Domain Models

The spectral representation of the autocovariance function of a stationary process (y_t) with piecewise constant spectral density

$$f(\omega) = \sum_{k=1}^{K} \lambda_k I_{(\omega_k - \omega_{k-1}]} , \tag{12}$$

where $0 = \omega_0 < \omega_1 < ... < \omega_K = \pi$ and $\lambda_1 \neq \lambda_2$, ..., $\lambda_{K-1} \neq \lambda_K$ yields

$$\begin{aligned} \gamma(j) &= \int_{-\pi}^{\pi} e^{i\omega j} f(\omega) d\omega \\ &= 2 \sum_{k=1}^{K} \lambda_k \int_{\omega_{k-1}}^{\omega_k} \cos(\omega j) d\omega \\ &= \frac{2}{j} \sum_{k=1}^{K} \lambda_k \left(\sin(\omega_k j) - \sin(\omega_{k-1} j) \right) \end{aligned} \tag{13}$$

if $j \neq 0$ and

$$\gamma(0) = 2 \sum_{k=1}^{K} \lambda_k (\omega_k - \omega_{k-1}) . \tag{14}$$

The parameters $\phi_1, ..., \phi_p$ of the minimum-mean-square-error predictor

$$\hat{y}_t = \phi_1 y_{t-1} + ... + \phi_p y_{t-p} \tag{15}$$

and the variance σ^2 of the prediction error $y_t - \hat{y}_t$ can be computed recursively from the autocovariances $\gamma(0), \gamma(1), ..., \gamma(p)$ with the Durbin-Levinson algorithm (see [22]-[23]). However, the autocovariances depend on the parameter vector λ which is unknown in practice and must therefore be estimated. Each component $\hat{\lambda}_k$ of the least squares estimator $\hat{\lambda}$ is just the sample mean of the periodogram ordinates in the respective frequency band. The sample mean is of course also the maximum-likelihood estimator of an i.i.d.

sample from an exponential distribution. For fixed K, the frequencies $\omega_1,...,\omega_{K-1}$ are estimated by global minimization of the sum of squared residuals.

As a simple example, consider the case where K=2 and p=1. Here the restricted estimator of the autoregressive parameter ϕ_1 is given by

$$\tilde{\phi}_1 = \frac{\tilde{\gamma}(1)}{\tilde{\gamma}(0)} = \frac{(\hat{\lambda}_1 - \hat{\lambda}_2)\sin(\hat{\omega}_1)}{\hat{\lambda}_1\hat{\omega}_1 + \hat{\lambda}_2(\pi - \hat{\omega}_1)} \tag{16}$$

and the conventional unrestricted estimator by

$$\hat{\phi}_1 = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)}, \tag{17}$$

where

$$\hat{\gamma}(j) = \frac{1}{n}\sum_{t=1}^{n-|j|}(x_t - \bar{x})(x_{t+|j|} - \bar{x}). \tag{18}$$

4 Empirical Results

The methods proposed in the previous sections are now used to forecast the quarterly real U.S. GDP from 1947Q1 to 2014Q3 (downloaded from FRED, Federal Reserve Economic Data, Fed St. Louis). Step functions of the form (12) with $K \leq 10$ steps are fitted to the periodogram of the first differences of the logarithms of this time series and the number of steps is chosen with AIC and BIC as well as with the new criterion based on the penalty factor (10). AIC and BIC select the maximum number of 10 steps, which is clearly a bad choice since macroeconomic time series of this type do not differ very much from white noise. Usually, the only apparent feature in their periodograms is a clustering of larger values in the low-frequency range. Figure 1.b shows that two steps already provide an adequate description of this typical spectral shape. Accordingly, the new criterion selects only two steps.

A similar spectral shape can be obtained by calculating the

autocovariances from the step function and using them for the calculation of the parameters of a conventional AR model as described in Section 3. However, this would require an absurdly large number of parameters (about 40; see Figure 1.b). On the other hand, unrestricted low-order ARMA(p,q) models (with $p,q \leq 3$) selected by AIC and BIC models imply spectral densities which are possibly too rich in detail given that the periodogram is rather featureless with the exception of the clustering mentioned above. Moreover, the spectral details of the different ARMA models are partly inconsistent with each other (see Figure 1.a).

Figure 1.c compares restricted low-order AR spectral densities with their unrestricted counterparts. Although the discrepancies appear to be relatively small, they have a large impact on the predictive power. Figure 2 shows the relative cumulative absolute forecast errors of various restricted and unrestricted ARMA(p,q) models (with $p,q \leq 2$). The unrestricted AR(1) model, which is typically selected by BIC, serves as benchmark. This benchmark model is consistently outperformed by the restricted AR(1) model throughout the whole forecasting period. However, the restricted AR(2) model is even better. It clearly outperforms all competing models. Increasing the AR order further up to 40 just increases the variance and has no positive effect on the forecasting performance. Similarly, the largest ARMA model, which is typically selected by AIC, performs worse than most other ARMA models.

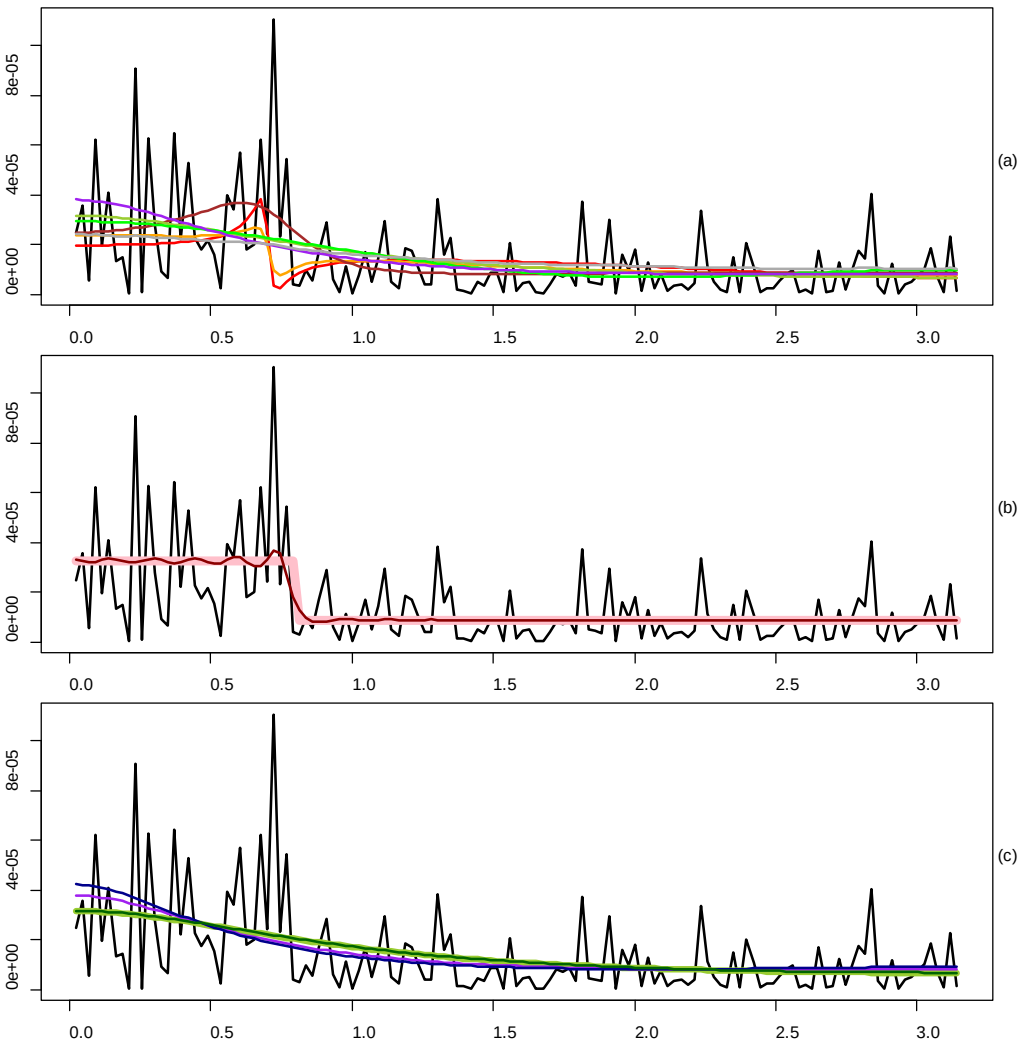

Figure 1: Smoothing the periodogram of U.S. GDP growth rates

(a) Best three ARMA spectra according to AIC (ARMA(3,3): red, ARMA(3,2): orange, ARMA(2,2): brown) and BIC (ARMA(1,0): yellowgreen, ARMA(0,2): green, ARMA(2,0): purple), respectively.

(b) AR(40) approximation (darkred) of step function (pink).

(c) Unrestricted AR(1) (yellowgreen) and AR(2) (purple) spectra vs. restricted AR(1) (darkgreen) and AR(2) (darkblue) spectra.

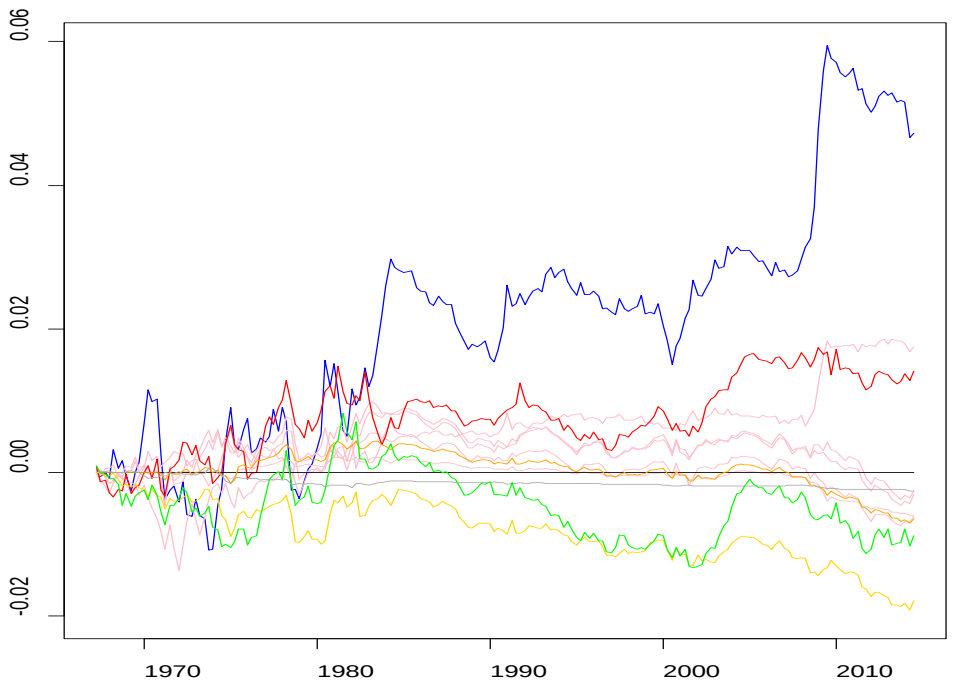

Figure 2: Relative cumulative absolute forecast errors of U.S. GDP growth rates

ARMA(0,0): blue, (1,0): black (benchmark, typically selected by BIC)

ARMA(2,0): orange, (2,2): red (typically selected by AIC)

ARMA(0,1),(0,2),(1,1),(1,2),(2,1): pink

Restricted AR(1): gray, (2): gold, (40): green

5 Conclusion

Leaving aside the fact that the GDP growth rates are not even stationary because of breaks in the first (e.g., the growth slowdown after the 1973 oil price shock) and second (e.g. the reduction in volatility starting in the 1980s which is called the Great Moderation) moments, the spectral densities implied by ARMA models still do not provide an adequate description of the clustering of large periodogram values in the low-frequency range (see Figure 1.a). While a simple step function appears to be more appropriate for this purpose, there exists no parsimonious time-domain model with a spectral density of this type. This paper therefore takes the pragmatic approach of first estimating the step function in the frequency domain with the help of a new model selection criterion for exponentially distributed samples and then using this estimate for imposing frequency-domain restrictions on conventional time-domain models.

When applied to the task of forecasting the GDP growth rates, this approach turns out to be extremely successful. The restricted forecasts clearly outperform their unrestricted counterparts. Since it is virtually impossible for any parsimonious model to take care of all the peculiarities of macroeconomic time series, it is very likely that the restricted forecasts also benefit from the fact that the use of step functions in this context typically produces a shrinkage effect (see Appendix A).

It is left to future research to extend the forecasting procedure introduced in this paper to the multivariate case and to investigate whether this extension will be a competitive alternative to the band-regression approach.

ACKNOWLEDGEMENTS. We would like to thank David Preinerstorfer and Lukas Steinberger for providing their implementation of the breakpoint-selection algorithm as well as for their valuable comments and suggestions.

Appendix A

Successive observations of a macroeconomic time series are typically positively autocorrelated, i.e., $\gamma(1) > 0$, which implies that the spectral density $f(\omega)$ of an AR(1) model decreases as the frequency increases. Using this property as well as the fact that $\cos(\omega)$ is also a decreasing function in the interval $[0, \pi]$, we obtain

$$\gamma(1) = \int_{-\pi}^{\pi} e^{i\omega} f(\omega) d\omega = 2 \sum_{k=1}^{K} \int_{\omega_{k-1}}^{\omega_k} \cos(\omega) f(\omega) d\omega$$

$$= 2 \sum_{k=1}^{K} \left(\int_{\omega_{k-1}}^{\omega_k} \cos(\omega) \lambda_k d\omega + \int_{\omega_{k-1}}^{\omega_k} \cos(\omega)(f(\omega) - \lambda_k) d\omega \right)$$

$$= 2 \sum_{k=1}^{K} \left(\int_{\omega_{k-1}}^{\omega_k} \cos(\omega) \lambda_k d\omega + \int_{\omega_{k-1}}^{\overline{\omega}_k} \cos(\omega) \underbrace{(f(\omega) - \lambda_k)}_{\geq 0} d\omega + \int_{\overline{\omega}_k}^{\omega_k} \cos(\omega) \underbrace{(f(\omega) - \lambda_k)}_{\leq 0} d\omega \right)$$

$$\geq 2 \sum_{k=1}^{K} \left(\int_{\omega_{k-1}}^{\omega_k} \cos(\omega) \lambda_k d\omega + \cos(\overline{\omega}_k) \int_{\omega_{k-1}}^{\overline{\omega}_k} (f(\omega) - \lambda_k) d\omega + \cos(\overline{\omega}_k) \int_{\overline{\omega}_k}^{\omega_k} (f(\omega) - \lambda_k) d\omega \right)$$

$$= 2 \sum_{k=1}^{K} \left(\int_{\omega_{k-1}}^{\omega_k} \cos(\omega) \lambda_k d\omega + \cos(\overline{\omega}_k) \underbrace{\int_{\omega_{k-1}}^{\omega_k} (f(\omega) - \lambda_k) d\omega}_{=0} \right)$$

$$= \int_{-\pi}^{\pi} e^{i\omega} \sum_{k=1}^{K} \lambda_k I_{(\omega_k - \omega_{k-1}]} d\omega,$$

where

$$\lambda_k = \int_{\omega_{k-1}}^{\omega_k} f(\omega) d\omega$$

and

$$0 = \omega_0 < \omega_1 < \ldots < \omega_K = \pi.$$

References

[1] E. J. Hannan, *Regression for Time Series*, in Time Series Analysis, Wiley, 1963a.

[2] E. J. Hannan, Regression for time series with errors of measurement, *Biometrika*, **50**, (1963b), 293-302.

[3] R. F. Engle, Band spectrum regression, *International Economic Review*, **15**, (1974), 1-11.

[4] P. C. B. Phillips, Spectral regression for co-integrated time series, *Nonparametric and semiparametric methods in economics and statistics*, Cambridge University Press, 1991.

[5] D. Corbae, S. Ouliaris and P. C. B. Phillips, Band spectral regression with trending data, *Econometrica*, **70**, (2002), 1067-1109.

[6] E. Reschenhofer and M. Chudy, Adjusting band-regression estimators for prediction: shrinkage and downweighting, *Working Paper*, (2015).

[7] M. Taniguchi, Non-regular estimation theory for piecewise continuous spectral densities, *Stochastic Processes and their Applications*, **118**, (2008), 153-170.

[8] E. Reschenhofer, Frequency domain modeling with piecewise constant spectra, *Journal of Modern Applied Statistical Methods*, **7**, (2008), 467-470.

[9] Y. Ninomiya, Information criterion for Gaussian change-point model, *Statistics and Probability Letters*, **72**, (2005), 237-247.

[10] Y. C. Yao, Estimating the number of change-points via Schwarz' criterion, *Statistics and Probability Letters*, **6**, (1988), 181-189.

[11] Y. C. Yao and S. T. Au, Least-squares estimation of a step function, *Sankhyā: The Indian Journal of Statistics*, Series A, **51**, (1989), 370-381.

[12] J. Liu, S. Wu and J. V. Zidek, On segmented multivariate regression, *Statistica Sinica*, **7**, (1997), 497-525.

[13] D. P. Foster and E. I. George, The risk inflation criterion for multiple regression, *Annals of Statistics*, **22**, (1994), 1947–1975.

[14] R. Tibshirani and K. Knight, The covariance inflation criterion for adaptive model selection, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **61**, (1999), 529-546.

[15] E. I. George and D. P. Foster, Calibration and empirical Bayes variable selection, *Biometrika*, **87**, (2000), 731–747.

[16] E. Reschenhofer, D. Preinerstorfer and L. Steinberger, Non-monotonic penalizing for the number of structural breaks, *Computational Statistics*, **28**, (2013), 2585–2598.

[17] J. Bai and P. Perron, "Computation and analysis of multiple structural change models, *Journal of Applied Econometrics*, **18**, (2003), 1-22.

[18] D. Rothman, Letter to the editor, *Technometrics*, **10**, (1968), 432.

[19] H. Akaike, "Fitting autoregressive models for prediction, *Annals of the Institute of Statistical Mathematics*, **21**, (1969), 243-247.

[20] E. Reschenhofer, M. Schilde, E. Oberecker, E. Payr, H. T. Tandogan and L. M. Wakolbinger, Identifying the determinants of foreign direct investment: a data-specific model selection approach, *Statistical Papers*, **53**, (2012), 739-752.

[21] R Core Team, R: A language and environment for statistical computing, *R Foundation for Statistical Computing*, Vienna, Austria, (2013). http://www.R-project.org/.

[22] N. Levinson, The Wiener RMS (root-mean-square) error criterion in filter design and prediction, *Journal of Mathematical Physics*, **25**, (1947), 261-278.

[23] J. Durbin, The fitting of time series models, *Review of the International Statistical Institute*, **28**, (1960), 233-244.