INVESTIGACIÓN & DESARROLLO, No. **12**, Vol. **1**: 33 – 38 (2012)
ISSN 2518-4431

BAYESIAN SEASONAL ANALYSIS WITH ROBUST PRIORS
ANALISIS ESTACIONAL BAYESIANO CON PRIORS ROBUSTOS

Rolando Gonzales Martínez
Universidad de Alcalá - España
gonzalesmartinez@gmail.com
(Recibido el 01 de octubre 2012, aceptado para publicación el 29 de enero 2012)

ABSTRACT

An analytical Bayesian approach to seasonal analysis is proposed, using robust priors to control for extreme observations. Seasonal fan charts were estimated with Bayesian predictive densities. Empirical applications to U.S. residential electricity consumption, Spain's tourism and Bolivian's inflation are presented. The results show that the Bayesian approach allows to investigate probabilistically the seasonal component of a time series, thus accounting for the uncertainty of the seasonal pattern.

RESUMEN

Se propone un enfoque Bayesiano para el análisis estacional, utilizando priors robustos para controlar el efecto de observaciones extremas. *Fan charts* estacionales fueron estimados con las densidades predictivas Bayesianas. Se presentan aplicaciones al consumo eléctrico residencial en EE.UU, el turismo en España, y la inflación en Bolivia. Los resultados indican que el enfoque Bayesiano permite investigar probabilísticamente el componente estacional de una serie de tiempo, considerando así la incertidumbre del patrón estacional.

Keywords: Seasonal Analysis, Bayesian Inference, Time Series.
Palabras clave: Análisis Estacional, Inferencia Bayesiana, Series Temporales.

1. INTRODUCTION

Seasonal analysis seeks to understand the periodic fluctuations of a variable that recur every year approximately with the same timing and with the same intensity[1]. Even if a usual procedure is to seasonally adjust a time series, in some cases the seasonal pattern can itself be of direct interest to the modeler e.g. in tourism or consumption applications. From the perspective of a policy-maker, the seasonal pattern of a time series can also be a useful tool to investigate the appropriate timing of issuing policy actions [4].

A Bayesian-least-squares regression approach to seasonal adjustment was suggested by Akaike [1]. Numerical results of Akaike and Ishiguro [2] showed that the approach of Akaike is superior to the X-11 seasonal adjustment, in the sense that no spurious fluctuations are introduced in the trend component during the seasonal adjustment. As stated by Young [14], Akaike's approach is based on solid inferential (namely, Bayesian) foundations, but its not resistant to outliers. To overcome this weakness, Young proposed a robust double-exponential error model where the predictive intervals are calculated with Gibbs sampling. Recent research analyses stochastic trends at the seasonal frequencies: Franses *et al*. [5] in the context of seasonal unit roots and Tommaso and Stefano [13] using Markov Chain Monte Carlo methods in a Bayesian stochastic model specification search for the selection of the unobserved seasonal component.

This paper proposes a simple, analytical, Bayesian approach to seasonal analysis. The seasonal analysis is not model-based, and the robustness to outlying observations is achieved using robust priors. Thus, analytic results for the predictive seasonal density can be calculated without using Markov Chain Monte Carlo methods. This seasonal predictive density —that accounts for the uncertainty in the estimation of the seasonal component— can be used to anticipate the seasonal dynamics of a time series, thus improving the decision-making process. The robust prior controls for the adverse effect of extreme observations, a traditional concern in seasonal analysis based on mean estimators[2].

Section 2 describes the Bayesian seasonal inference. Section 3 contains empirical applications to U.S. residential electricity consumption, Spain's tourism and Bolivian's inflation. Section 4 discusses the results.

[1] See Kallek [8]. A formal (mathematical) definition of seasonality in the frequency domain can be found in Granger [6]. According to Granger, the sources of these seasonal patterns are the weather, calendar effects, timing decisions or agent expectations.

[2] Due to the sensibility of the mean to outlying observations, a seasonal index obtained with data contaminated with outliers may be governed more by an exceptional deviation than by the systematic seasonal movement.

DOI: 10.23881/idupbo.012.1-3e

2. BAYESIAN SEASONAL ANALYSIS

Let $\mu_1, \mu_2, \dots, \mu_s$ be seasonal means of a stochastic process $\{y_t\}_{t=1}^{T}$ for n years, with $T = s \times n$, being s a seasonal set $s \in \{2, 3, 4, 6, 12\}$. Define $\xi_s = \{y_{s,1}, \dots, y_{s,n}\}$.

ASSUMPTION 1. Let $\mathrm{E}(\xi) = \mu_s$ with known variance $\mathrm{E}(\xi - \mu_s) = \sigma_S^2 < \infty$ and $\mathrm{P}(\bigcap_{i=1}^{n} y_{s,i}) = \prod_{i=1}^{n} \mathrm{P}(\bigcap_{i=1}^{n} y_{s,i})$, then $\xi = iid(\mu_s, \sigma_S^2)$

ASSUMPTION 2. Let $\sqrt{\mathrm{n}}\left(\left(\frac{1}{n}\sum_{i=1}^{n} y_{s,i}\right) - \mu_s\right) \xrightarrow{d} \mathscr{N}(0, \sigma_S^2)$ (Lindeberg - Lévy CLT, given assumption 1), then $\xi_s \sim \mathscr{N}(\mu_s, \sigma_S^2)$.

Even if ξ_s is not identically distributed, ASSUMPTION 2 can still hold due to the Lyapunov CLT *i.e.* if ξ_s satisfies the Lyapunov condition $\lim_{n\to\infty} \frac{1}{\vartheta_n^{2+\delta}} \sum_{i=1}^{n} \mathbb{E}(\,|y_{s,1} - \mu_s|^{2+\delta})$, for any $\delta > 0$.

For simplicity of notation, drop the seasonal subscript of $\mu \coloneqq \mu_s$ and $\sigma^2 \coloneqq \sigma_s^2$, then μ and σ^2 would be seasonal parameters for any $s \in \{2, 3, 4, 6, 12\}$, given N.

- **Bayesian Inference for a Single Seasonal Mean**

The posterior density $\pi(\mu | y_{s,1}, \dots, y_{s,n})$ of μ is proportional to a prior $\pi(\mu)$ times the likelihood $\mathscr{L}(y_{s,1}, \dots, y_{s,n} | \mu)$,

$$\pi(\mu \mid y_{s,1}, \dots, y_{s,n}) \propto \pi(\mu) \mathscr{L}(y_{s,1}, \dots, y_{s,n} | \mu)$$

If $y_{s,1}, \dots, y_{s,n}$ is a gaussian process with μ and σ^2 estimated with $\hat{\mu}$ and $\hat{\sigma}^2$ (ASSUMPTION 2),

$$\pi(\mu) \propto \exp\left(-\frac{1}{2\underline{\sigma}^2}(\mu - \underline{\mu})^2\right)$$

and from ASSUMPTION 1[3],

$$\mathscr{L}(y_{s,1}, \dots, y_{s,n} | \mu) \propto \prod_{i=1}^{n} \exp\left(-\frac{1}{2\hat{\sigma}^2}(y_i - \hat{\mu})^2\right) = \exp\left(-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(y_i - \hat{\mu})^2\right)$$

then,

$$\pi(\mu) \times \mathscr{L}(y_{s,1}, \dots, y_{s,n} | \mu) \propto \exp\left(-\frac{1}{2}\left(\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(y_i - \hat{\mu})^2 + \frac{1}{\underline{\sigma}^2}(\hat{\mu} - \underline{\mu})^2\right)\right)$$

expanding the squares,

$$\pi(\mu) \times \mathscr{L}(y_{s,1}, \dots, y_{s,n} | \mu) \propto \exp\left(-\frac{1}{2}\frac{1}{\hat{\sigma}^2 \underline{\sigma}^2}\left(\hat{\mu}^2(\hat{\sigma}^2 + n\underline{\sigma}^2) - 2\hat{\mu}(\underline{\mu}\hat{\sigma}^2 + n\underline{\sigma}^2\hat{y}) + (\underline{\mu}^2\hat{\sigma}^2 + \underline{\sigma}^2\sum_{i=1}^{n} y_i^{\,2})\right)\right)$$

since the last term can be treated as part of the normalizing constant k,

$$\propto \exp\left(-\frac{1}{2}\left(\mu^2\left(\frac{1}{\underline{\sigma}^2} + \frac{n}{\hat{\sigma}^2}\right) - 2\mu\left(\frac{\underline{\mu}}{\underline{\sigma}^2} + \frac{n\bar{y}}{\hat{\sigma}^2}\right) + k\right)\right)$$

[3] Assumption 1 can be relaxed in two directions: a full Bayesian treatment of σ_S^2 as an unknown parameter can be included in the estimation of μ_s, and the possible dependence in ξ_s implies modifying the likelihood. Note that even if $\{y_t\}_{t=1}^{T}$ is not iid, it is posible that $\xi_s = iid(\mu_s, \sigma_s^2)$, being ξ_s blocks of non-consecutive sequences of $\{y_t\}_{t=1}^{T}$.

$$\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\underline{\sigma}^2}+\frac{n}{\hat{\sigma}^2}\right)\left(\mu-\left(\frac{\frac{\underline{\mu}}{\underline{\sigma}^2}+\frac{n\bar{y}}{\hat{\sigma}^2}}{\frac{1}{\underline{\sigma}^2}+\frac{n}{\hat{\sigma}^2}}\right)\right)^2\right)$$

Thus, the Bayesian posterior precision estimator is,

$$\frac{1}{\bar{\sigma}^2}=\frac{1}{\underline{\sigma}^2}+\frac{n}{\hat{\sigma}^2}=\frac{\hat{\sigma}^2+n\underline{\sigma}^2}{\underline{\sigma}^2\hat{\sigma}^2}$$

with a posterior mean,

$$\bar{\mu}=\frac{\frac{1}{\underline{\sigma}^2}}{\frac{1}{\underline{\sigma}^2}+\frac{n}{\hat{\sigma}^2}}\underline{\mu}+\frac{\frac{n}{\hat{\sigma}^2}}{\frac{1}{\underline{\sigma}^2}+\frac{n}{\hat{\sigma}^2}}\bar{y}$$

- **Robust Priors**

A robust prior for μ is the value m such that $\mathbb{P}(y \leq m) \geq \frac{1}{2}$ and $\mathbb{P}(y \geq m) \geq \frac{1}{2}$, i.e. the median of $y_{s,1}, \dots, y_{s,n}$. It is know that the large sample variance of m is approximately $1/(4nf^2(m))$, with $f(m)$ a density function evaluated in m, see *inter alia* Rider [11], Siddiqui [12] and Maritz and Jarret [9]. In this case, $f(m) = (2\pi)^{-1/2} - (1/2)\exp(m^2)$, then a prior robust to extreme observations would be,

$$\pi(\mu)\propto \exp\left(-\frac{1}{2(4nf^2(m))^{-1}}(\hat{\mu}-m)^2\right)$$

- **Bayesian Seasonal Predictive Density**

The predictive density of the next seasonal observation $y_{s,n+1}$ given $y_{s,1}, \dots, y_{s,n}$ is the conditional density $f(y_{s,n+1}|y_{s,1}, \dots, y_{s,n})$. This can be found with the marginal posterior distribution,

$$\begin{aligned} f(y_{s,n+1}\mid y_{s,1},\dots,y_{s,n}) &= \int f(y_{s,n+1},\mu\mid y_{s,1},\dots,y_{s,n})d\mu \\ &= \int f(y_{s,n+1}\mid\mu)\times\pi(\mu\mid y_{s,1},\dots,y_{s,n})d\mu \end{aligned}$$

being both gaussian,

$$\begin{aligned} f(y_{s,n+1}\mid y_{s,1},\dots,y_{s,n}) &\propto \int \exp\left(-\frac{1}{2\hat{\sigma}^2}(y_{s,n+1}-\mu)^2\right)\times\exp\left(-\frac{1}{2\bar{\sigma}_n^2}(\mu-\bar{\mu})^2\right)d\mu \\ &\propto \int \exp\left\{-\frac{1}{2}\left[\left(\frac{1}{\hat{\sigma}^2}+\frac{1}{\bar{\sigma}_n^2}\right)\mu^2-2\left(\frac{y_{s,n+1}}{\hat{\sigma}^2}+\frac{\bar{\mu}}{\bar{\sigma}_n^2}\right)\mu+\frac{y_{s,n+1}^2}{\hat{\sigma}^2}+\frac{\bar{\mu}^2}{\bar{\sigma}_n^2}\right]\right\}d\mu \end{aligned}$$

which simplifies to[4],

$$f(y_{s,n+1}\mid y_{s,1},\dots,y_{s,n})\propto \exp\left(-\frac{1}{2(\hat{\sigma}^2+\bar{\sigma}_n^2)}(y_{s,n+1}-\bar{\mu})^2\right)$$

The posterior seasonal predictive density has a mean $\bar{\mu}$ and a variance equal to the observational variance plus the posterior variance, $\hat{\sigma}^2 + \bar{\sigma}_n^2$. Thus, the variance of the seasonal predictive density takes into consideration both the data uncertainty and the uncertainty in the estimation of the seasonal component μ.

[4] See Bolstad [4] for details.

3. RESULTS

Figure 1 shows the monthly series of tourism in Spain from January 1999 to December 2010[5] and the results of estimating the seasonal predictive density for this series using the Bayesian techniques described previously. A clear seasonal pattern can be appreciated in the time series of Spain's tourism, and the Bayesian seasonal analysis shows that tourism is higher in the middle of the year, during the European summer vacations, and it is lower in the coldest months (November, December, January and February). Also, a systematic increase in the number of tourists can be expect across the first part of the year. A peak of tourism is observed once a year, in August.

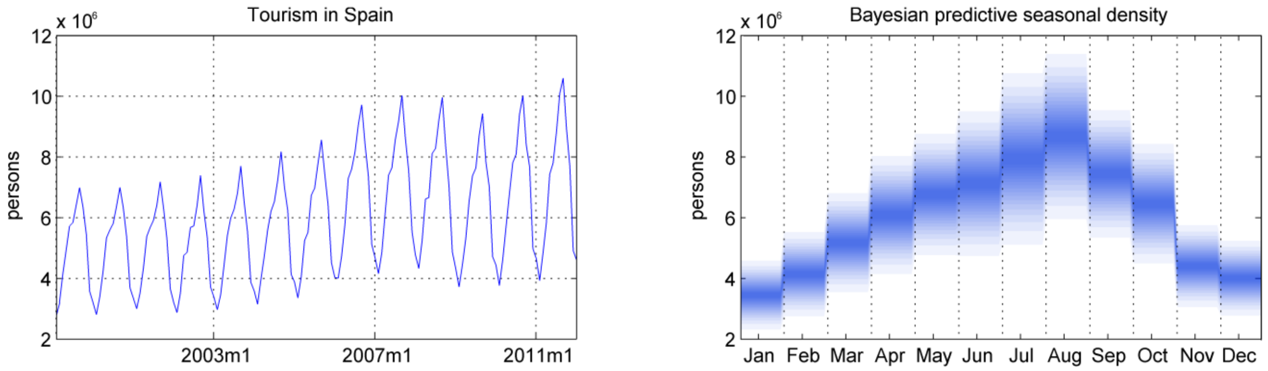

Figure 1 - Bayesian seasonal predictive density of the tourism in Spain

Figure 2 displays the US residential electricity consumption from January 1973 to December 2010, in trillion Btu[6]. The seasonal component of the consumption is evident in the time series, despite their volatility, and two spikes of electricity consumption are visible in the Bayesian seasonal predictive density (Figure 2, right): consumption peaks in summer due to air conditioning, and it is also higher at the beginning and the end of the year, due to the heating required during the winter months. See Jorgensen [7].

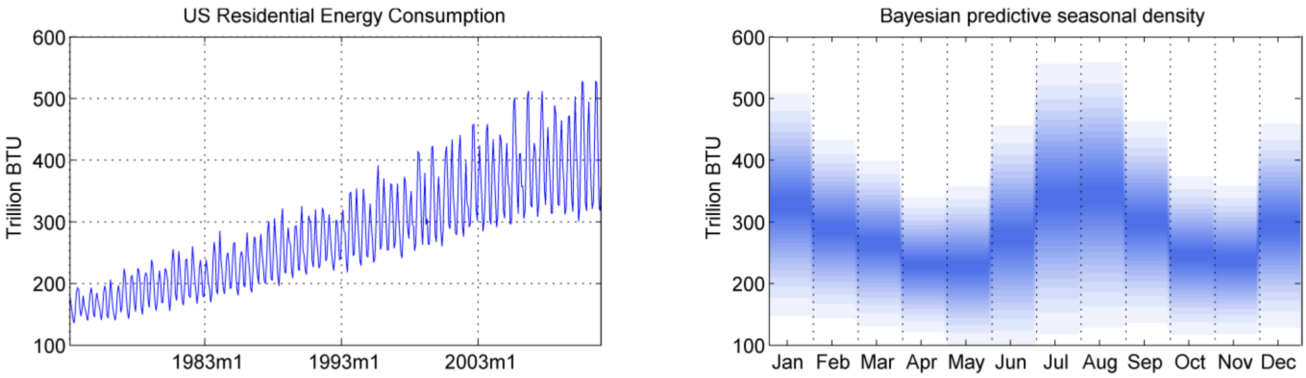

Figure 2 - Bayesian seasonal predictive density of U.S. residential electricity consumption

Finally, figure 3 (left) displays monthly data of the Bolivian inflation, from Janurary 1990 to December 2010. This series is different from the previous ones because in this case the seasonal component does not dominate the dynamics of the variable and it is only visible after estimating the Bayesian seasonal predictive density (figure 2, right). Also, the inflation data shows high volatility and some additive outliers in the months of January 1991 (where inflation was equal to 5.8 percent) and in November 2011 (where a deflation of 2.5 percent was observed). The Bayesian seasonal density suggests the presence of a seasonal pattern in inflation, possible due to the higher weight given to food products in the basket of the Bolivian price index, as these products are subject to the seasonal effects of weather and seasonal planting and harvesting.

[5] The data is measured in number of persons. The information was gathered from the Survey of Tourist Occupation (*Encuesta de Ocupación en Alojamientos Turísticos*) of the National Statistics Institute of Spain.
[6] Btu stands for British Thermal Units. The source of the data is the U.S. Energy Information Administration.

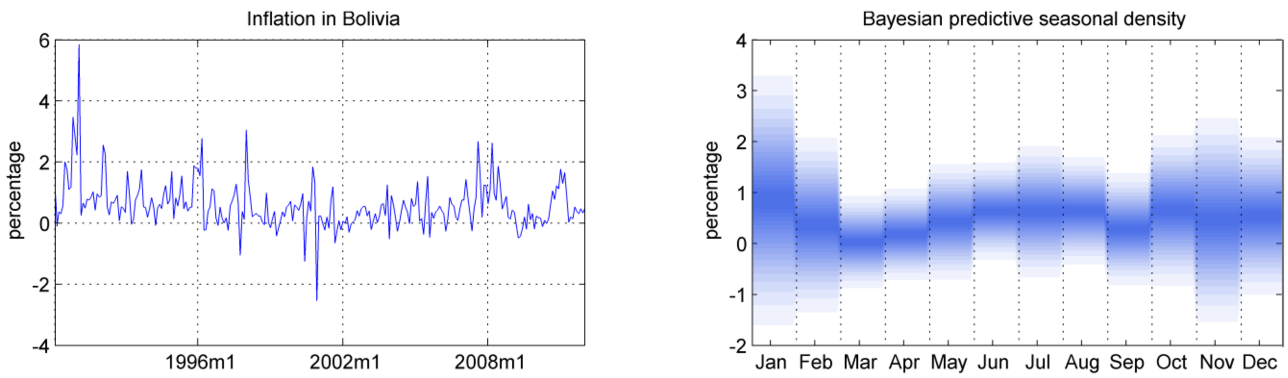

Figure 3 - Bayesian seasonal predictive density of the Bolivian inflation

The Bayesian seasonal analysis with robust priors shows that, on average, inflation in Bolivia tends to be lower in the months of March, April and September, and seems to be a little higher in May, June, July and August. In January, February, October, November and December, the concentration of the densities is above other months, but the high amplitude of the seasonal densities implies that it is more difficult to state conclusively that inflation is lower or higher in these periods than in other months. This is an advantage of the intervalic Bayesian seasonal analysis over a traditional seasonal analysis with point estimators: the Bayesian approach allows to take into consideration both the data and the estimation uncertainty to form seasonal fan charts, and thus properly account for the variance of the seasonal pattern. In March, for example, a 90 percent seasonal credible interval suggests that the seasonal component of inflation could be between -0.73 percent and 0.79 percent, and on the contrary, in December, the uncertainty about the pattern of seasonal inflation is higher, since the 90 percent credible interval is between -0.76 and 1.83. See Table 1 for a complete detail of the numerical values of the seasonal fan chart across a year.

TABLE 1 - BAYESIAN SEASONAL PREDICTIVE DENSITY: BOLIVIAN INFLATION

Bayesian Seasonal Interval (in percentage)			January	February	March	April	May	June	July	August	September	October	November	December
			2.90	1.80	0.79	0.93	1.37	1.43	1.71	1.52	1.20	1.88	2.13	1.83
			1.69	0.95	0.34	0.49	0.81	0.96	1.07	1.00	0.66	1.15	1.15	1.07
90	50	10	1.00	0.47	0.09	0.24	0.50	0.69	0.71	0.70	0.35	0.74	0.59	0.64
			0.69	0.25	-0.03	0.12	0.35	0.57	0.54	0.57	0.21	0.55	0.33	0.44
			0.00	-0.23	-0.28	-0.13	0.03	0.30	0.18	0.28	-0.10	0.14	-0.23	0.01
			-1.21	-1.07	-0.73	-0.58	-0.53	-0.17	-0.45	-0.25	-0.64	-0.59	-1.21	-0.76

4. CONCLUSIONS

The Bayesian approach to seasonal analysis was found to be useful both to investigate the seasonal component of a time series and to estimate seasonal predictive densities that adopt the form of seasonal fan charts. These Bayesian seasonal densities allow to properly judge the behavior of the seasonal pattern of a time series and the uncertainty about this seasonal component, an advantage over traditional (frequentist) seasonal point estimators. The issues of time aggregation and seasonal adjustment with the Bayesian methods described in this paper are left for future research[7].

[7] A Bayesian seasonal adjustment can easily be performed with the Bayesian seasonal index estimated in this paper, using in principle simple techniques as the ones describe in standard econometric textbooks as *inter alia* Pindyck and Rubinfeld [10]; nevertheless, it seems wise to explote the flexibility of Bayesian methods to explore the possibility of adapting the seasonal adjustment to the source of seasonal fluctuations, rather than produce a standard seasonal adjustment algorithm. On the other hand, in the case of composite series as e.g. price indexes, the effects of time aggregation can be explored comparing the results of estimating individual seasonal indexes with a Bayesian seasonal estimator of a (joint) multivariate gaussian distribution of the series with both mean and variance unknown.

5. REFERENCES

[1] H. Akaike. “Seasonal adjustment by a bayesian modeling.” *Journal of Time Series Analysis* vol. 1, no. 1, pp. 1-13, 1980.

[2] H. Akaike and M. Ishiguro. “Comparative study of the x- 11 and baysea procedures of seasonal adjustment,” in *Applied Time Series Analysis of Economic Data*, A. Zellner,1983.

[3] W. Bolstad. *Introduction to Bayesian Statistics*. John Wiley & Sons, 2004.

[4] P. H. Franses. *Time Series models for business and economic forecasting*. Cambridge University Press, 1998.

[5] P. H. Franses et al. “Bayesian analysis of seasonal unit roots and seasonal mean shifts.” *Journal of Econometrics*, vol. 78, no. 2, pp. 359-380, June, 1997.

[6] C. W. J. Granger. “Seasonality: Causation, interpretation, and implications,” in *Seasonal Analysis of Economic Time Series*, National Bureau of Economic Research, 1979, pp. 33-56.

[7] J. B. Jorgensen and F. Joutz. “Modelling and forecasting residential electricity consumption in the u.s. mountain region.” Working Papers 2012-003, The George Washington University, Department of Economics, Research Program on Forecasting, 2012.

[8] S. Kallek. “An overview of the objectives and framework of seasonal adjustment,” in *Seasonal Analysis of Economic Time Series*, National Bureau of Economic Research, 1978, pp. 3-32.

[9] J. S. Maritz and R. G. Jarrett. “A note on estimating the variance of the sample median.” *Journal of the American Statistical Association*, vol. 75, no. 361, pp. 194-196, 1978.

[10] R. S. Pindyck and D. L. Rubinfeld. *Econometria: Modelos y Pronosticos*. McGraw-Hill, 1998.

[11] P. R. Rider. “Variance of the median of small samples from several special populations.” *Journal of the American Statistical Association*, vol. 55, no. 289, pp. 148-150, 1960.

[12] M. M. Siddiqui. “Approximations to the moments of the sample median.” *The Annals of Mathematical Statistics*, vol. 33, no. 1, pp. 157-168, 1962.

[13] P. Tommaso and G. Stefano. “Bayesian stochastic model specification search for seasonal and calendar effects.” MPRA Paper 27305, University Library of Munich, Germany, 2010

[14] M. R. Young. “Robust seasonal adjustment by Bayesian modelling.” *Journal of Forecasting*, vol. 15, no. 5, pp. 355-367, 1996.