

Forecasting and Analysis Tools for Regional Industries' Dynamics [†]

Valeriy Semenychev ¹ and Anastasiya Korobetskaya ^{2,*} 

¹ Department of Mathematical Methods in Economics, Samara University, 34, Moskovskoye Shosse, 443086 Samara, Russia; 505tot@mail.ru

² System Integrator "Webzavod", 156, Galaktionovskaya Str., 443001 Samara, Russia

* Correspondence: kornast@yandex.ru

[†] Presented at the 7th International Conference on Time Series and Forecasting, Gran Canaria, Spain, 19–21 July 2021.

Abstract: The article is devoted to the author's approach and tools for regional industries' modeling, analysis and forecasting, following the general idea of splitting time series into four components: trend, cycles, seasonal component, and residuals. However, the authors introduce new approaches, models, metrics, and identification algorithms, and the components' interaction structures, having included the analysis of 12 industries in 82 regions of Russia. The models and forecast accuracy were tested on 3–12 month forecasts, thus proving their high accuracy. Therefore, the article proposes not only new systematic econometric tools but a methodology for decision making, developed to provide stable and adequate characteristics of complex non-linear evolutionary dynamics of Russian regions.

Keywords: time series; regional economy; forecasting; modeling; medians; wavelets; Russia



Citation: Semenychev, V.; Korobetskaya, A. Forecasting and Analysis Tools for Regional Industries' Dynamics. *Eng. Proc.* **2021**, *5*, 1. <https://doi.org/10.3390/engproc2021005001>

Academic Editors: Ignacio Rojas, Fernando Rojas, Luis Javier Herrera and Hector Pomare

Published: 24 June 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Regional industries, having both spatial and temporal dimensions, represent a complex meso-economic object of analysis [1]. On the one hand, they represent inter-related socio-economic systems, and their dynamics should be coherent with each other and the country. On the other hand, the regions' development level and their connectivity may vary a lot.

The common equilibrium approach is based on a mechanistic view of economic systems and is focused on a return to the equilibrium state. However, regional economies tend to shift to an evolutionary approach based on long-term forecasting [2] and the regions' abilities to change their economic structures [3]. Researchers also point out that regional differences in repeatability tend to become a significant factor in effective economic policy decisions [4].

The research aims at designing tools for modeling and forecasting, in the context of time series concerning regional industries' dynamics. The acquired results should be adequate and accurate to provide a facility for decision making and to support sustainable evolutionary development.

2. Data

As a statistical database, we are using an official data source provided through the Unified Interdepartmental Information and Statistical System (EMISS) by the Russian Federal State Statistic Service. The EMISS database possesses operational monthly data on the production level of each economic industry (real (volume) growth rate, percent) for each subject (region) of the Russian Federation. The industries are classified hierarchically by the All-Russian Classifier of Types of Economic Activity (OKVED2). We have chosen the twelve most important industries in terms of presence in different regions:

1. Extraction of minerals.
2. Crude oil and natural gas extraction.

3. Mining of metal ores.
4. Manufacturing.
5. Food production.
6. Petroleum production.
7. Chemical production.
8. Pharmaceuticals and materials used for medical purposes.
9. Rubber and plastic production.
10. Metallurgy.
11. Computers, electronics and optical production.
12. Automotive industry. Production of motor vehicles, trailers and semi-trailers.

The data reflect the situation in 82 regions of Russia (except Crimea, Sevastopol, and the Republic of Chechnya due to lack of statistics), as well as for Russia as a whole. However, some industries may be not presented in particular regions, so, in total, there are about 750 time series.

The analysis embraces the period from January 2005 to August 2020. This time interval seems interesting as it covers important periods and milestones within the Russian economy's evolution. It reveals the economic growth in the noughties of the XXI century, the crisis of 2008–2009, subsequent recession and recovery, the growth of political turbulence, the adoption of economic sanctions against Russia, and the beginning and extension of the COVID-19 pandemic.

3. Research Methods

3.1. Approach

To reach the research goal, we defined some approaches to be used in our algorithms and models. Basically, we are following the idea of splitting time series into trend, cycles, seasonal fluctuations, and residuals (stochastic component). However, we are trying to review the details and characteristics of economic objects at the meso-level.

Firstly, the meso-economic systems are non-linear and show evolutionary development. So, the simple models such as linear or exponential trends are applicable only for short periods of time. For longer periods, perspective changes inside the region, as well as its interrelations with other regions, affect the dynamics and should be appropriately reflected in the models. We provide a complex of different trends, possessing extremums, inflection points, asymptotes, asymmetry, and thus, the ability to adapt to such volatility.

Another approach is using points of structural change [5] to reveal the moments of time where dynamics change drastically and cannot anymore be described by the same model.

The other important factor is the model residuals. Traditionally, the distribution law of the residuals is supposed to be normal (Gaussian). However, real economic systems are rarely normally or even log-normally distributed. The practice shows very different asymmetric and heavy-tailed distributions. We examined all the above-mentioned industries and regions and discovered a wide variety of distributions. So, choosing some particular distribution law or even a fixed set of laws seems inappropriate. It is better to use tools that are more robust and do not depend upon the distribution law.

One of the basic and robust metrics is the median. Instead of choosing the one “best” model, we identify many different models, and at each moment of time use the median of all their fitted values. Thus, we eliminate one of the hardest problems—structural identification of the model. Using the median effectively filters inadequate models and provides sustainable fits.

To find the median, it is better to take the maximum possible fits for each point. Using the bootstrap procedure [6], it is possible to identify a few models with the same structure (formula) but different parameters. The bootstrap procedure is a common approach to increase small sample sizes.

The other important point is the criterion used to identify models' parameters. The most common approach is using the least squares. However, it is very sensitive to outliers

presented in heavy-tailed distributions. The least absolute deviations method is seen as more reliable [7]:

$$\sum_{t=1}^n |Y_t - \hat{Y}_t| \rightarrow \min, \quad (1)$$

where Y_t is the original time series, t is time (ordering indices from 1 to n), $\hat{Y}_t$ is the model's fitted values.

On the other hand, for multiplicative residuals, the least absolute percentage deviations seem more correct:

$$\sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \rightarrow \min. \quad (2)$$

Unfortunately, the choice between the additive and multiplicative residuals' structure is not obvious. The mixed additive–multiplicative structures can also be present. So, we defined combined measures to minimize both:

$$\frac{1}{\bar{Y}} \sum_{t=1}^n |Y_t - \hat{Y}_t| + \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \rightarrow \min. \quad (3)$$

The criterion (1) includes two parts to minimize both additive and multiplicative residuals, but the first part is divided by the time series average $\bar{Y}$ (which is constant and does not change extremum position), to underline the parts' comparability. The same effect may be achieved by multiplying the second part by $\bar{Y}$, but we prefer to use relative values.

It should be also mentioned that all the models and algorithms described below were implemented in the R language using both the authors' program code and open-source libraries.

3.2. Models

The most common models for time series structures appear as additive (2) and multiplicative (3):

$$Y_t = T_t + C_t + S_t + \varepsilon_t, \quad (4)$$

$$Y_t = T_t(1 + C_t)(1 + S_t)(1 + \varepsilon_t), \quad (5)$$

where T_t —trend values; C_t —cyclical component values; S_t —seasonal component levels; ε_t —stochastic component.

It is also reasonable to consider mixed additive–multiplicative structures:

$$Y_t = (T_t + C_t)S_t + \varepsilon_t, \quad (6)$$

$$Y_t = T_t(1 + C_t) + S_t + \varepsilon_t. \quad (7)$$

The authors' complex of trend models currently includes linear, generalized exponential, power trends, four cumulative logistic (S-shaped) and four impulse logistic (bell-shaped) trends with different asymmetry settings:

$$T_t = C_0 + A_0 t, \quad (8)$$

$$T_t = C_0 + A_0 t^\alpha, \quad (9)$$

$$T_t = C_0 + A_0 e^{\alpha t}, \quad (10)$$

$$T_t = C_0 + \frac{A_0}{1 + e^{-\alpha(t-t_0)}}, \quad (11)$$

$$T_t = C_0 + A_0 \arctg(\alpha(t - t_0)), \quad (12)$$

$$T_t = C_0 + A_0 \exp(-\exp(-\alpha(t - t_0))), \quad (13)$$

$$T_t = C_0 + A_0 (1 + \exp(-\alpha(t - t_0)))^\sigma, \quad (14)$$

$$T_t = C_0 + A_0 \exp(-\alpha(t - t_0)^2), \quad (15)$$

$$T_t = C_0 + \frac{A_0}{1 + \alpha(t - t_0)^2}, \quad (16)$$

$$T_t = C_0 + \frac{A_0}{1 + \alpha(t - t_0)^2} \cdot \frac{1}{1 + \exp(-\sigma(t - t_0))}, \quad (17)$$

$$T_t = C_0 + \frac{A_0}{1 + (\sigma(t)(t - t_0))^2}, \sigma(t) = \frac{1}{1 + \exp(-\sigma(t - t_0))}, \quad (18)$$

All the trend models (7)–(17) use the unified naming of parameters where: C_0 is the vertical shift constant and asymptotic level (if any), A_0 is the trend amplitude (vertical scale), α is the growth/decline velocity (horizontal scale), t_0 is the horizontal shift (inflection point for S-shaped trends, extremum point for bell-shaped trends), σ is the asymmetry coefficient. The models differ by their shape, growth velocity, and skewness (symmetric, fixed asymmetry or free asymmetry).

For each dynamic series, all the trend models are identified through the total sample length and can be grouped by means of structural changes (the points may be different for each model). Thus, we have up to 22 fitted values for each point in the time series.

As for cycles, we used two general approaches to modeling. The first approach is based on the E. Slutsky [8] hypothesis that any fluctuations could be presented as a sum of a few sinus functions with non-proportional frequencies:

$$C_t = \sum_{i=1}^{\infty} A_i \sin(\omega_i t + \varphi_i), \quad (19)$$

where A_i is the i th sinus amplitude, ω_i is the sinus frequency and φ_i is the sinus phase.

This approach is effective for modeling as it gives a well-smoothed model of the cycles. However, there is no guarantee that the amplitudes, phases, and frequencies that optimally described dynamics in the past will remain the same in the future. So, the extrapolation of such a model is simple but unproven. Thus, we turned our attention to wavelet transformation [9–11]. The wavelet transformation is used widely in signal processing to eliminate signal noise, but is now adopted in economics and other sciences for time series smoothing and forecasting.

Wavelets are seen as functions used to identify local non-periodical fluctuations and monitor their changes through time periods. The time series are decomposed on a few levels of so-called wavelet and scaling coefficients. These components may vary from high-frequency ones (representing the “noise”) to lower-frequency components representing local cycles. The low-frequency components of wavelet decomposition can be easily modeled and forecasted with ARMA models and reversely transformed back to provide a smoothed model and forecast.

The variety of wavelet functions’ families is wide. In this study, we used the most generalized discrete transformation from the wavelet families: Haar, Daubechies, etc. (in total, 42 wavelet functions).

3.3. Identification Algorithm

Based on the below-mentioned principles, an algorithm to identify time series models is designed with the following sequence of steps:

1. Preprocessing of the initial time series, removing random outliers and using R ’s standard library, then replacing them with median smoothed values.
2. Determining the structure of seasonal fluctuations (additive or multiplicative).
3. Detecting seasonal fluctuations using the STL function, which returns the smoothed trend, seasonal fluctuations, and random residuals based on LOESS smoothing. For multiplicative structures, the logarithms are used.
4. Deseasonalization (removing seasonal fluctuations from the initial series).

5. Determining the structure of cyclic fluctuations.
6. Building the median trend without structural shifts and without bootstrapping. To do this, all available types of trends are selected using criterion (1), and the median value from all trend estimates is taken at each point in the time series.
7. Detrending (removing a trend from a series).
8. Fitting cyclic fluctuations to the detrended and deseasonalized data.
9. Removal of cyclical fluctuations from the deseasonalized data.
10. Repeating step 6 but with structural shifts.
11. Repeating steps 7–9 for the newly fitted trend values.
12. Plotting the median trend with both the structural changes and bootstrapped values.
13. Repeating steps 7–9 for the newly fitted trend values.

The resulting estimates of dynamics and their components are used for modeling and forecasting. When studying the Russian regions' dynamics, regional models are built independently of each other, and general trends are revealed upon the modeling results.

The following methods are applied in the algorithm:

- LOESS smoothing to define seasonal coefficients as provided in the *stl* function in the *stats* package [12];
- The Breusch–Pagan test on heteroscedasticity to separate additive and multiplicative structures (using the *bptest* function in the *lmtest* package) [13];
- The probabilistic simulated annealing algorithm [14] for finding the global minimum area and initial estimates of model parameters;
- The *RPROP* algorithm [15,16], which is used to minimize errors in training neural networks;
- The minimization algorithm implemented in the standard *nlm* function [17];
- Wavelet transformation using the *wavelets* package;
- The ARIMA-models identification algorithm using the *forecast* package.

4. Results

The identification algorithm was applied for all analyzed time series. The results are shown in Figure 1.

The top part of the chart shows the original data (black points), fitted values (grey solid line), median trend fits (black dashed line) and the fits of all of the trends (dotted grey lines). The middle part of the chart demonstrates the median cycles model. The bottom part of the chart shows seasonal fluctuations. The titles of the middle and bottom of the chart appear as structures ('mult.' is abbreviation for multiplicative).

The example demonstrates a median-declining S-shaped trend, and the "cloud" of all the trends, depicting possible distributions of fits for estimates at each point. The cycles clearly show a decline in 2008–2009 (global crisis), 2014–2015 (economic sanctions against Russia) and 2020 (pandemic). Seasonality achieves its peak in December, and shows slow growth through the given period. This is one "typical" example of the dynamics, but for other regions and industries it varies drastically.

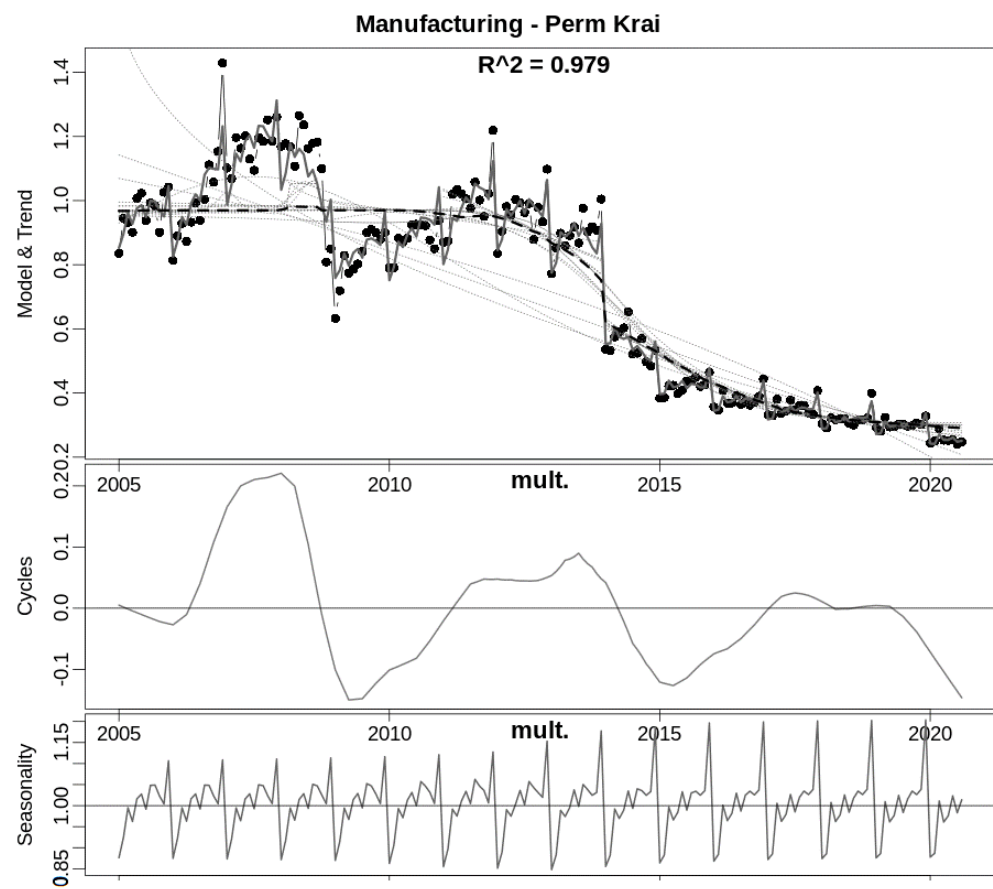


Figure 1. Modeling results for the manufacturing in Perm Krai.

Our research goal was, however, not only to obtain the forecasts and models but also to measure their accuracy. To achieve this, we split the time series into two parts: one to identify the model (working sample) and the other to measure forecast accuracy (test sample). At the regional level, short-term and middle term forecasts appear as the most useful. So, we tested the models on 3, 6, 9 and 12-months forecasts. We also varied the forecasting year from 2018 to 2020 to generalize the conclusions, and thus, verify the models' overall accuracy.

This study uses two common measures of accuracy. The determination coefficient is used to measure the modeling accuracy:

$$R^2 = 1 - \frac{\sum (Y_t - \hat{Y}_t)^2}{\sum (Y_t - \bar{Y}_t)^2} \quad (20)$$

The Theil's coefficient is used to measure the forecast error:

$$U_2 = \sqrt{\frac{\sum (Y_t - \hat{Y}_t)^2}{\sum Y_t + \sum \hat{Y}_t}} \times 100\% \quad (21)$$

For high accuracy, R^2 is supposed to be above 0.7 and U_2 below 30%.

Table 1 shows the median estimates of R^2 and U_2 among all regions, separated by industry. The industries are enumerated as mentioned in Section 2.

Table 1. Median estimates for models and forecasts on test samples.

Year	2018					2019					2020				2021
Forecast Depth, Months	-	3	6	9	12	-	3	6	9	12	-	3	6	8	-
Industry No	R ²	U ₂ , %				R ²	U ₂ , %				R ²	U ₂ , %			
1	0.888	4.7	6.4	8.4	10.5	0.908	3.9	6.6	9.0	10.3	0.910	5.0	9.1	11.5	0.894
2	0.966	2.3	4.1	4.9	6.1	0.973	1.3	2.4	3.2	3.5	0.976	2.2	6.7	7.8	0.956
3	0.897	4.4	5.1	6.6	7.4	0.911	3.6	5.2	8.3	9.3	0.915	6.2	7.6	7.5	0.917
4	0.881	4.3	5.2	7.0	7.7	0.883	4.0	5.8	6.9	9.8	0.891	4.8	9.2	11.6	0.892
5	0.891	3.9	4.8	6.0	7.1	0.893	3.5	4.8	5.9	6.8	0.896	4.5	5.9	7.5	0.905
6	0.772	3.1	5.9	7.9	8.6	0.797	3.9	8.6	9.3	9.4	0.798	4.2	10.1	10.2	0.791
7	0.851	4.6	6.9	9.5	10.6	0.856	5.0	7.0	10.4	12.1	0.862	7.2	11.3	11.7	0.853
8	0.712	12.3	17.9	18.5	18.0	0.723	8.8	19.0	21.1	25.9	0.696	17.3	28.8	29.8	0.725
9	0.916	6.8	9.1	9.7	11.7	0.916	5.4	8.2	11.1	11.9	0.916	6.5	11.3	13.0	0.914
10	0.846	8.5	10.7	12.0	13.4	0.851	6.5	9.9	12.0	14.8	0.858	8.8	12.6	14.2	0.868
11	0.650	14.0	21.1	26.6	33.7	0.581	12.2	20.2	26.8	34.7	0.582	23.2	36.0	38.6	0.626
12	0.864	10.4	16.0	22.8	26.3	0.859	9.4	15.0	17.8	23.6	0.877	13.4	26.0	28.5	0.873

Judging by the table, the forecast accuracy is generally high. R^2 estimates are above 0.7, and U_2 are below 20%, for most industries except for the pharmaceutical, electronic and computer production, and automotive industries. These industries are highly volatile at the meso-economic level in Russia, especially the electronic and computer production industry, which is highly subsidized by the state and depends on government support. More stable industries such as mineral extraction, manufacturing, the chemical industry and metallurgy demonstrate low forecast errors. The predictability of the industries' futures could be assessed as their stability indicator.

5. Conclusions

The key findings of the research are as follows:

1. The approaches used to analyze and forecast regional industries' dynamics are justified. They include time series decomposition, the median approach, increases in the models' variety, using weighted additive–multiplicative criterion, and applying wavelet transformation for cycles.
2. The complex of models and algorithms is designed, upgraded and applied in the form of a program code in the R language.
3. The designed tools are applied to 12 industries in 82 Russian regions. Decompositions and forecasts are obtained for each time series. The median trend model shows general tendencies (growth, decline and bell) and structural change points. Cycle models define cycle stages and reversion points (peaks, troughs and zero-points). Seasonal models describe calendar effects and their changes through years.
4. The results' accuracy is proven by short-term and mid-term forecasts (3–12 months), even including the pandemic period.

This paper mostly demonstrates individual series analysis. However, more significant results may be achieved by comparing different regions, both between each other and with Russia as a whole. In our previous research, we showed that cycles and trends in the regions are not synchronous [18] but they may be clustered in terms of model type and the values of parameters.

We plan to continue to develop tools by increasing the trends' variety, using bootstrapping at all algorithm steps, improving calculation methods, adding interval forecasts, and analyzing regions' interactions and neighborhoods.

Author Contributions: Conceptualization, V.S. and A.K.; methodology, V.S. and A.K.; software, A.K.; validation, A.K.; formal analysis, A.K.; investigation, V.S. and A.K.; resources, V.S. and A.K.; data curation, A.K.; writing—original draft preparation, V.S. and A.K.; writing—review and editing, A.K.; visualization, A.K.; supervision, V.S.; project administration, V.S.; funding acquisition, V.S. All authors have read and agreed to the published version of the manuscript.

Funding: The reported study was funded by RFBR, project number 20-010-00549.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Semenychev, V.K.; Korobetskaya, A.A. Tools for Estimation of “Deterministic Chaos” of Economic Sectoral Mesodynamic. In *Economic Systems in the New Era: Stable Systems in an Unstable World. IES 2020*; Ashmarina, S.I., Horák, J., Vrbka, J., Šuleř, P., Eds.; Lecture Notes in Networks and Systems; Springer: Cham, Switzerland, 2021; Volume 160. [\[CrossRef\]](#)
2. Boschma, R. Towards an Evolutionary Perspective on Regional Resilience. *Reg. Stud.* **2015**, *49*, 733–751. [\[CrossRef\]](#)
3. Simmie, J.; Martin, R. The economic resilience of regions: Towards an evolutionary approach. *Camb. J. Reg. Econ. Soc.* **2010**, *3*, 27–43. [\[CrossRef\]](#)
4. Beraja, M.; Hurst, E.; Ospina, J. The Aggregate Implications of Regional Business Cycles. *Econometrica* **2019**, *87*, 1789–1833. [\[CrossRef\]](#)
5. Perron, P. Structural change, econometrics of. In *Macroeconometrics and Time Series Analysis. The New Palgrave Economics Collection*; Durlauf, S.N., Blume, L.E., Eds.; Palgrave Macmillan: London, UK, 2010. [\[CrossRef\]](#)
6. Cavaliere, G.; Taylor, A. Bootstrap Unit Root Tests for Time Series with Nonstationary Volatility. *Econom. Theory* **2008**, *24*, 43–71. [\[CrossRef\]](#)
7. Tyrsin, A.N. Robust construction of regression models based on the generalized least absolute deviations method. *J. Math. Sci.* **2006**, *139*, 6634–6642. [\[CrossRef\]](#)
8. Slutsky, E.E. Slozhenie sluchajnyh prichin kak istochnik ciklicheskih processov. *Voprosy kon’yunktury* [Addition of random causes as a source of cyclic processes. *Market issues.*] **1927**, *3*, 34–64. (In Russian)
9. Morettin, P.A. Wavelets in Statistics. *Rev. Inst. Math. Stat. Univ. Sao Paulo* **1997**, *3*, 211–272.
10. Percival, D.B.; Walden, A.T. *Wavelet Methods for Time Series Analysis*; Cambridge University Press: London, UK, 2000.
11. Raihan, S.M.; Wen, Y.; Zeng, B. Joint Time-Frequency Distributions for Business Cycle Analysis. In *Wavelet Analysis and Its Applications. WAA 2001*; Tang, Y.Y., Yuen, P.C., Li, C., Wickerhauser, V., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Germany, 2001; Volume 2251. [\[CrossRef\]](#)
12. Cleveland, R.B.; Cleveland, W.S.; McRae, J.E.; Terpenning, I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *J. Off. Stat.* **1990**, *6*, 3–73.
13. Breusch, T.S.; Pagan, A.R. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* **1979**, *47*, 1287–1294. [\[CrossRef\]](#)
14. Xiang, Y.; Gubian, S.; Suomela, B.; Hoeng, J. Generalized Simulated Annealing for Efficient Global Optimization: The GenSA Package for R. *R J.* **2013**, *5*, 13–28. [\[CrossRef\]](#)
15. Igel, C.; Huesken, M. Empirical evaluation of the improved Rprop learning algorithms. *Neurocomputing* **2003**, *50*, 105–123. [\[CrossRef\]](#)
16. Riedmiller, M. Advanced supervised learning in multilayer perceptrons—From backpropagation to adaptive learning techniques. *Comput. Stand. Interfaces* **1994**, *16*, 265–278. [\[CrossRef\]](#)
17. Dennis, J.E.; Schnabel, R.B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*; Prentice-Hall: Englewood Cliffs, NJ, USA, 1983.
18. Khmeleva, G.A.; Semenychev, V.K.; Korobetskaya, A.A.; Kozhukhova, V.N.; Agaeva, L.K.; Burets, Y.S.; Egorova, K.S.; Zemtsov, S.P.; Koroleva, E.N.; Chertopyatov, D.A. *Rossiyskie Regiony v Usloviyakh Sanktsiy: Vozmozhnosti Opereshayushchego Razvitiya Ekonomiki na Osnove Innovatsiy* [Russian Regions in the Conditions of Sanctions: The Possibility of Priority Development of the Economy Based on Innovation]; Khmeleva, G.A., Ed.; Pub. Sam. State Univ.: Samara, Russia, 2019. (In Russian)