DYNAMIC ECONOMETRIC MODELS
Vol. 7 – Nicolaus Copernicus University – Toruń – 2006

Mariola Piłatowska
Nicolaus Copernicus University in Toruń

The Effects of the Incorrect Identification of Non-stationarity of Economic Processes for Prediction Mean Square Error

1. Introduction

Analysing the relationships among non-stationary economic processes the identification of the two type of non-stationarity, i.e. in mean and in variance, is of great importance considering the negative effects for estimation and statistical inference being appeared in the case of incorrect identification of non-stationarity. These negative effects[1] can be avoided if the information about the internal structure of economic processes (trend, periodicity, autoregression) will be used when building the econometric model. Such approach ensures that the residual process will have the white noise properties which are the most desired and the congruence postulate (in Granger sense (1981) or in Zieliński sense (1984)) will be satisfied. If the congruence postulate is realized the coexistence of models with different specification as well for levels as for differences is possible. Hence, the two competitive model specification, i.e. for levels (strategy 'always take levels') and for differences (strategy 'always difference') when economic processes are non-stationary, can be reconciled. However, it should be remembered that the estimates of parameters in models for levels and models for differences are in general different. These models will be statistically acceptable, but will differ with regard to the specification, economic interpretation of parameters and behavior in forecasting.

The purpose of the paper is to evaluate the effects of incorrect identification of processes non-stationary in mean (stationary fluctuations around determinis-

[1] There are the effects of *overdifferencing* and *underdifferencing*. See: Piłatowska (2003a), (2003b), (2004).

tic trend, trend stationary (TS)) and processes non-stationary in variance (integrated processes, difference stationary (DS)) for the behavior in forecasting of econometric models specified for levels (the TS model – strategy 'always take levels') and for differences (the DS and EC (error correction) model – strategy 'always difference'). The evaluation of these effects will be curried out with the use of the Monte Carlo experiments for processes generated with a given structure (under different assumptions)

2. The Description of Simulation Experiment

The scenario of experiments assuming the relationship between processes Y_t and X_t , which are non-stationary in mean or in variance is depicted the Table 1.

The experiments 1A and 2A are aimed at evaluating the effects of incorrect identification of processes non-stationary in mean and in variance respectively considered the same relationship in the whole frequency band. This means that the parameters measuring relationships among different frequency components of Y_t and X_t (for example low and high frequency components) are the same. Then, the parameters in models for differences measuring the relationship in high frequency band (because the difference filter eliminates the low frequency band), will be the same as the parameters in models for levels measuring the relationship in the whole frequency band.

Data generating model in experiment 1A assumes that the linear combination of two processes Y_t and X_t, which are non-stationary in mean (they are trend stationary), i.e. $Y_t - \alpha X_t = \mu + \varepsilon_t$, is stationary (does not have deterministic trend). This means that the vector $[1\ -\alpha]$ eliminates the non-stationarity in mean and at the same time reflects the relationship between Y_t and X_t on a stationary level (because the relationship in the whole frequency band is the same). Then, it is said that Y_t and X_t are co-trended.

While in experiment 1A the co-trending effect is observed, in experiment 1B the data generating model has the deterministic trend component. This means that the co-trending does not occur, i.e. the parameters measuring relationship between Y_t and X_t on a stationary level can not be used in eliminating the deterministic trend from Y_t and X_t. The lack of co-trending may be interpreted in terms of omitted important variable[2].

[2] The results of simulation experiments in Kufel (2002), pp. 180–183, indicate that in the congruent model reduced to significant variables the deterministic trend plays the role of balancing the structure of model when the important variable including in the data generating model of Y_t was omitted, and the case of non-stationarity in mean was occurred. See also Kufel, Piłatowska, Zieliński (1996).

Table 1. Scenario of experiments

Type of non-stationarity	Relationship of low and high frequency components of generated processes Y_t and X_t	
	the same for low and high frequencies	different for low and high frequencies
non-stationarity in mean	***Experiment* 1A** $Y_t = 3X_t + \mu_t + \varepsilon_t$, $X_t = \gamma_{0x} + \gamma_x t + \eta_{x,t}$, $\eta_{x,t} = \beta_x \eta_{x,t-1} + \varepsilon_{x,t}$	***Experiment* 1B** $Y_t = 3\eta_{x,t}^{low} + 2\eta_{x,t}^{high} + \gamma_{0y} + \gamma_y t + \varepsilon_t$, $X_t = \gamma_{0x} + \gamma_x t + \eta_{x,t}$, $\eta_{x,t} = \beta_x \eta_{x,t-1} + \varepsilon_{x,t}$,
	Parameter values taken in experiments: – $\beta_x = (0.6, 0.8, 0.9, 1)$, $\varepsilon_t \sim N(0,\sigma), \sigma = 1, 3$, n = 120, 60, 30, – significance level in the selection method: 0.01, 0.05, 0.1	
non-stationarity in variance	***Experiment* 2A** $Y_t = 3X_t + \varepsilon_t$, $X_t = \sum_{s=0}^{t-1} X_{AR,t-s}$, $X_{AR,t} = \beta_x X_{AR,t-1} + \varepsilon_{AR,t}$	***Experiment* 2B** $Y_t = 3X_t^{low} + 2X_t^{high} + \varepsilon_t$, $X_t = \sum_{s=0}^{t-1} X_{AR,t-s}$, $X_{AR,t} = \beta_x X_{AR,t-1} + \varepsilon_{AR,t}$,
	Parameter values taken in experiments: – $\beta_x = (0.6, 0.7, 0.8, 0.9)$, $\varepsilon_t \sim N(0,\sigma), \sigma = 1, 3$, n = 120, 60, 30, – significance level in the selection method: 0.01, 0.05, 0.1	

Low frequency components, i.e. $\eta_{x,t}^{low}$ in the 1B experiment and X_t^{low} in the 2B experiment are obtained through the filtration of process $\eta_{x,t}$ and process X_t respectively by the means of the Spencer moving average[3], and high frequency components will be calculated as: $\eta_{x,t}^{high} = \eta_{x,t} - \eta_{x,t}^{low}$, $X_t^{high} = X_t - X_t^{low}$. The parameter values γ_{1x}, γ_{1y} of deterministic trend are generated from the symmetric distribution with parameters (0.05; 0.02), (0.075; 0.025) respectively, and $\gamma_{0x} = \gamma_{0y}$ are equal 100.

The purpose of experiments 1B and 2B is to evaluate the effects of incorrect identification of processes non-stationary in mean and in variance considered the different relationship for different frequency components. Then, the parameters in models for differences are not the same as the parameters in models for levels, which as before measure the relationship in the whole frequency band, but this time they are averaged with weights equal to the proportion of variance of different frequency components in the total variance of X_t.

In experiment 2A the data generating model assumes that the processes Y_t and X_t are cointegrated, i.e. the combination $Y_t - \alpha X_t = \varepsilon_t$ is stationary although processes Y_t and X_t are first order integrated processes (have a trend in vari-

[3] The Spencer moving average in the form: $1/350[5]^2[7][-1, 0, 1, 1, 2, \ldots]$ eliminates high frequency components (see: Yule, Kendall (1966)).

ance). This means that the vector [1 –α] eliminates the non-stationarity in variance and at the same time measures the relationship between processes Y_t and X_t on a stationary level.

In experiment 2B the data generating model also assumes the existence of cointegration, but additionally the different relationship for low and high frequency components of processes Y_t and X_t ($\alpha_1 \neq \alpha_2$). As a result the parameters measuring the relationship between processes Y_t and X_t are averaged with appropriate weights.

In all experiments, 1A, 1B, 2A, 2B, the OLS method was used to estimate the following models: the TS model within the strategy 'always take levels', the DS and EC models within the strategy 'always difference'. The models are of the form:

– *strategy 'always take levels'*:

the TS model: $$Y_t = \sum_{s=1}^{q_y} \beta_{ys} Y_{t-s} + \sum_{s=0}^{q_x} \beta_{xs} X_{t-s} + \delta_1 t + \delta_0 + \varepsilon_t, \qquad (1)$$

– *strategy 'always difference'*:

the DS model: $$\Delta Y_t = \sum_{s=1}^{q_y} \beta_{ys} \Delta Y_{t-s} + \sum_{s=0}^{q_x} \beta_{xs} \Delta X_{t-s} + \mu + \eta_t, \qquad (2)$$

the EC model: $$\Delta Y_t = \sum_{s=1}^{q_y} \beta_{ys} \Delta Y_{t-s} + \sum_{s=0}^{q_x} \beta_{xs} \Delta X_{t-s} + \theta EC_{t-1} + \mu + \eta_t. \qquad (3)$$

When the case of the relationship among processes non-stationary in mean occurring (experiment 1A, 1B), model (1) is treated as a true model, and the models for differences (models (2) and (3)) are treated as alternative ones with regard to model (1) for levels. In model (3) EC_{t-1} denotes the error correction which in experiment 1A assuming the same relationship in the whole frequency band is equal $EC_{t-1} = Y_{t-1} - \hat{\alpha} X_{t-1}$, and in experiment 1B assuming different relationships for different frequency components – $EC_{t-1} = Y^*_{t-1} - \hat{\alpha} X^*_{t-1}$, where Y^*_{t-1} and X^*_{t-1} stand for processes after elimination of linear trend.

When the relationship among integrated processes (non-stationary in variance) is considered (comp. experiment 2A, 2B), models (2) and (3) are treated as true models, and model (1) for levels is treated as an alternative one.

The results of the Monte Carlo simulation from all experiments (1A, 1B, 2A, 2B) referred to the comparison of models TS, DS and EC with regard to: specification, residual process properties, estimates of parameters by X_t and ΔX_t in models for levels and differences respectively, distribution of the *t*-Student and Durbin-Watson statistics, distribution of determination coefficient R^2. The detailed results within the mentioned capacity are in Piłatowska (2003).

Whereas the behavior of forecasting models TS, DS and EC provided the different relationships for different frequency components (experiment 1B, 2B) for samples of n = 60, 30, is presented below[4]. The models reduced to significant variables are used to calculate dynamic forecasts with horizon $h = 1, 2, \ldots, 15$ for $n = 60$ and $h = 1, 2, \ldots, 10$ for $n = 30$, and also to calculate the prediction mean square errors, where realizations $y_{n+1}, y_{n+2}, \ldots, y_{n+h}$ and $x_{n+1}, x_{n+2}, \ldots, x_{n+h}$ were obtained from appropriate data generating models of Y_t and X_t. Forecasts of Y_t were calculated under assumption of known values of explanatory process X_t.

3. Performance of Forecasting Models TS, DS and EC

The estimates of parameters in reduced models (1), (2) and (3) distinctly differ (see Piłatowska (2003)) as a result of different relationship for different frequency components (experiment 1B). Therefore it is expected that the performance of model TS (reduced model (1)) and models DS and EC (reduced model (2) and (3)) in forecasting will also be different. Forecasts of Y_t are obtained from the model for levels, which describes relationships in the whole frequency band (although parameters are averaged with appropriate weights) and hence takes into account the relationships in long and short run. Whereas forecasts of Y_t are calculated from the model describing relationships for high frequencies (i.e. after eliminating the low frequencies by difference filter), i.e. from the model referring to relationships in short run. The comparison of forecasting properties of models TS, DS and EC will be curried out by the means of the ratios of prediction mean square errors (PMSE) calculated from each forecasting model, i.e. PMSE(DS)/PMSE(TS), PMSE(EC)/PMSE(TS), which allow to compare the strategy ‘always difference’ with the strategy ‘always take levels’.

Ratios of Prediction Mean Square Errors in Experiment 1B

The PMSE(DS)/PMSE(TS) ratios show (fig. 1, 2) that model TS outperforms model DS for all parameters values β_x, significance levels α, size of disturbances σ and sample sizes n at the whole forecast horizon, because ratios PMSE(DS)/PMSE(TS) are greater than one. As the forecast horizon and parameter values β_x increase, the domination of model TS is greater at the whole forecast horizon. Hence model DS cannot compete with model TS.

From the comparison of the performance of forecasting models TS and EC results that model TS can compete with model TS. Model EC outperforms

[4] The behavior of forecasting models TS, DS and EC provided the same relationship in the whole frequency band (experiment 1A, 2A) is presented in Piłatowska (2003).

model TS (ratios PMSE(EC)/PMSE(TS) are lower than one), see fig. 3 and 4, for small disturbance ($\sigma = 1$) in both sample sizes (n = 60, 30), for large disturbance in sample size of 60 and for conservative strategy in eliminating insignificant processes (especially at the 1%, but also at 5% significance level). In those cases, as a result of conservative strategy the specification of model TS is too parsimonious and does not include lagged processes such x_{t-1}, y_{t-1}, y_{t-2} which play the role of balancing the harmonic structure of both sides of model. Therefore such specification is not sufficient to describe changes of Y_t in the case of different relationships for low and high frequency components.

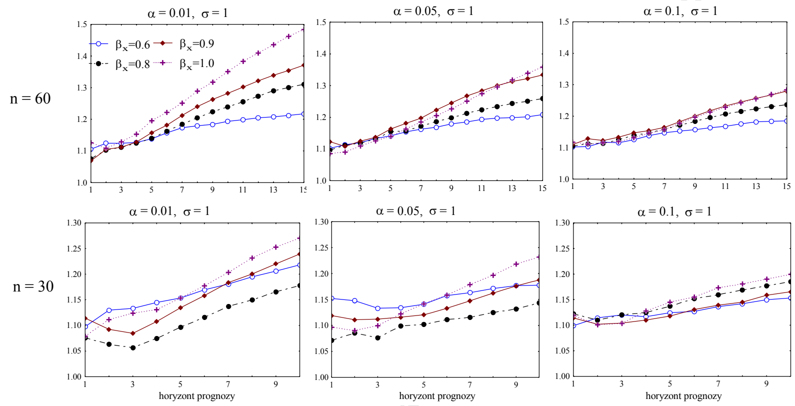

Fig 1. Ratios PMSE(DS)/PMSE(TS) in different sample size (n = 60, 30) and disturbance $\sigma = 1$ in experiment 1B

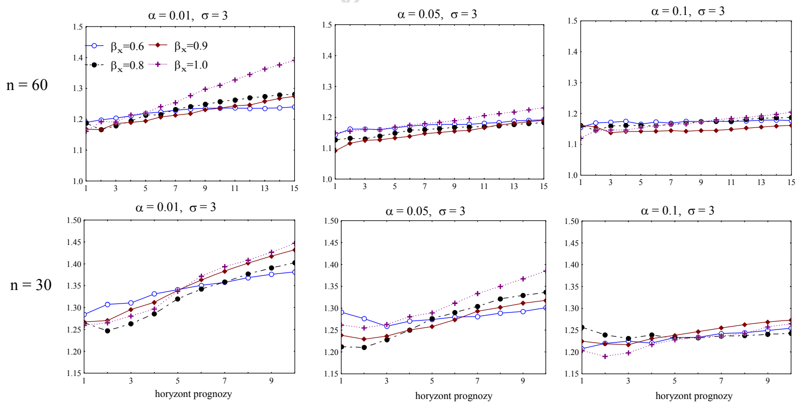

Fig. 2. Ratios PMSE(DS)/PMSE(TS) in different sample size (n = 60, 30) and disturbance $\sigma = 3$ in experiment 1B

Model TS slightly outperforms model EC (fig. 3 and 4) at almost whole forecast horizon (except short horizon) at the 5% and 1% significance level (liberal strategy in eliminating insignificant processes). As the significance level increases the specification of model TS includes more frequently additional elements, such y_{t-1}, y_{t-2}, x_{t-1}, t, balancing the structure of model, than at the 1% significance level. For large disturbance ($\sigma = 3$) model TS provides forecasts with lower PMSE at all significance levels and in all sample sizes. This domination is kept also in small sample (n = 30) at the 5% and 1% significance level in spite of parsimonious specification of model TS (most frequently 'x_t' and 'x_t, t'). This means that these specifications can well approximate the data generating model in the case of different relationships for low and high frequency components.

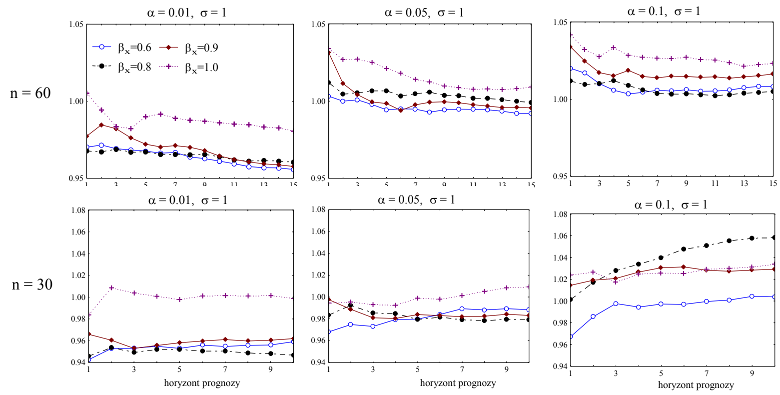

Fig. 3. Ratios PMSE(EC)/PMSE(TS) in different sample size (n = 60, 30) and disturbance $\sigma = 1$ in experiment 1B

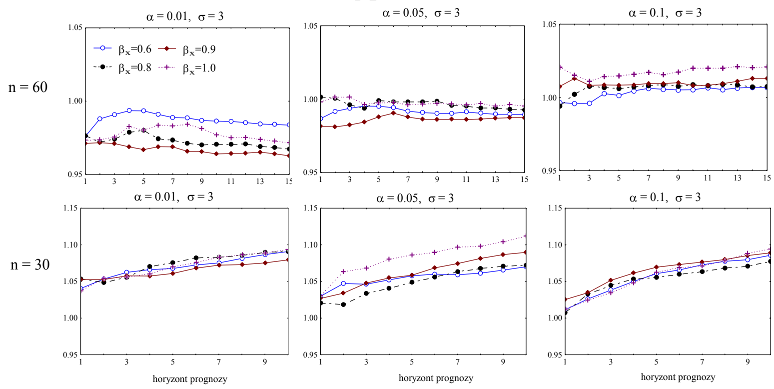

Fig. 4. Ratios PMSE(EC)/PMSE(TS) in different sample size (n = 60, 30) and disturbance $\sigma = 3$ in experiment 1B

Ratios of Prediction Mean Square Errors in Experiment 2B

In experiment 2B the ratios PMSE(DS)/PMSE(TS) indicates that model TS outperforms model DS for all parameter values β_x, significance levels, size of disturbance and all sample sizes at the whole forecast horizon, because the ratios PMSE(DS)/PMSE(TS) are greater than one (fig. 5 and 6). As the forecast horizon increases, the domination of model TS grows at the whole forecast horizon.

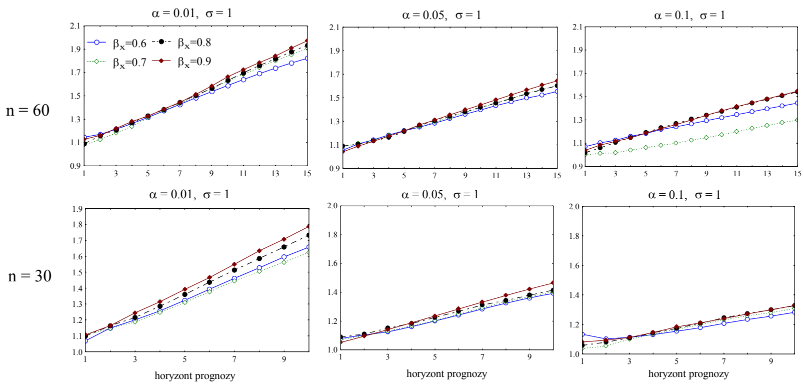

Fig. 5. Ratios PMSE(DS)/PMSE(TS) for different sample size ($n = 60, 30$) and disturbance $\sigma = 1$ in experiment 2B

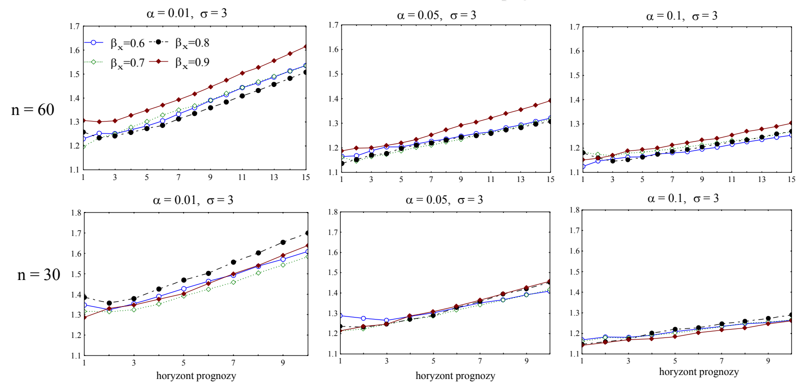

Fig. 6. Ratios PMSE(DS)/PMSE(TS) for different sample size ($n = 60, 30$) and disturbance $\sigma = 3$ in experiment 2B

Ratios PMSE(EC)/PMSE(TS) show the different performance of forecasting models TS and EC depending on the size of disturbance σ. For small disturbance ($\sigma = 1$) model EC is in general superior for all sample sizes n, significance levels and parameter values β_x at the whole forecast horizon (except $\beta_x = 0.6$, 0.8 at the 1%), because the ratios are lower than one (fig. 7 and 8). However, for large disturbance ($\sigma = 3$) model TS gives lower prediction mean square errors (ratios PMSE(EC)/PMSE(TS) are slightly greater than one).

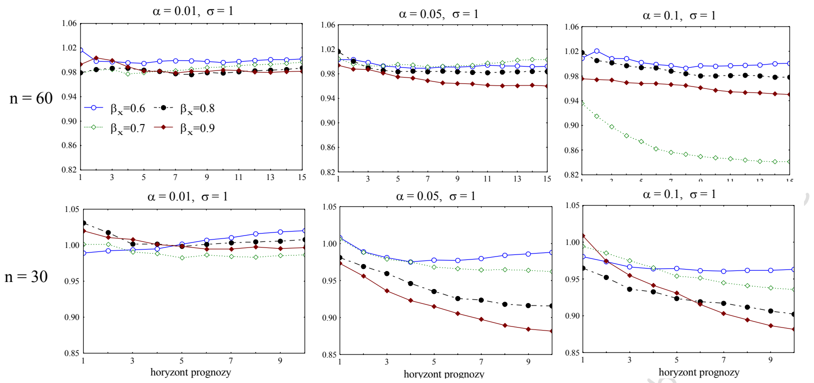

Fig. 7. Ratios PMSE(EC)/PMSE(TS) for different sample size ($n = 60$, 30) and disturbance $\sigma = 1$ in experiment 2B

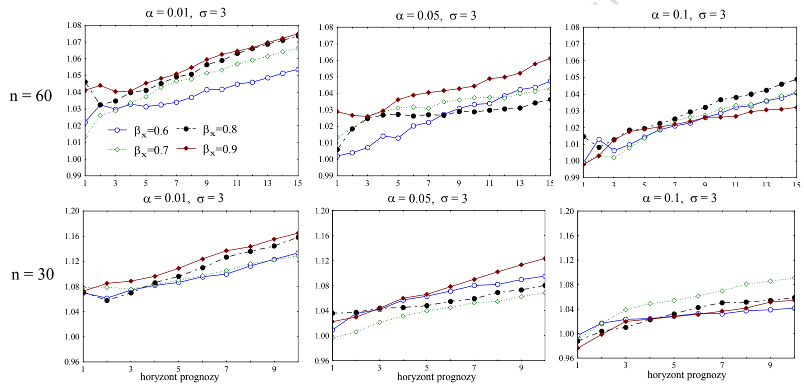

Fig. 8. Ratios PMSE(EC)/PMSE(TS) for different sample size ($n = 60$, 30) and disturbance $\sigma = 3$ in experiment 2B

4. Conclusions

The comparison of the strategy „always take levels" with the strategy „always difference" indicates that neither strategy has the clear advantage in forecasting. In other words, models for levels (the TS models) can compete with models for differences (the EC models) even in the case of correct identification of non-stationarity (in mean or in variance). This suggests the usefulness of models for levels as well as models for difference independently of the type of non-stationarity, however provided that models satisfy the congruence postulate consisting in specifying the model in such a way that the residual process has white noise properties. Moreover this may indicate the rule of thumb to build models at the same time for levels and for differences, and then to choose the model having better statistical properties as well as better acceptance from the economic point of view. On the other hand both models can be used in forecasting to calculate combined forecasts.

Literature

Granger, C. W. J. (1981), Some Properties of Time Series Data and their Use in Econometric Model Specification, *Journal of Econometrics*, 16, 121–130.

Kufel, T. (2002), *Postulat zgodności w dynamicznych modelach ekonometrycznych (Congruence Postulate in Dynamic Econometric Models)*, Wydawnictwo UMK, Toruń.

Kufel, T., Piłatowska, M., Zieliński, Z.(1996), Symulacyjna analiza poznawczych własności dynamicznych modeli zgodnych (Simulation Analysis of Cognitive Properties of Dynamic Congruent Models), in: A. Zeliaś (red.), Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych (Spatial Time Series Modelling and Forecasting), Wydawnictwo AE, Kraków, 47–59.

Piłatowska, M. (2003a), Skutki nadmiernego i niewystarczającego różnicowania procesów ekonomicznych. Analiza symulacyjna (Effects of Overdifferencing and Underdifferencing. Simulation Analysis), *Przegląd Statystyczny (Statistical Survey)*, z. 1, 59–74.

Piłatowska, M. (2003b), *Modelowanie niestacjonarnych procesów ekonomicznych. Studium metodologiczne (Modelling Non-stationary Economic Processes. Methodological Study)*, Wydawnictwo UMK, Toruń.

Piłatowska, M. (2004), Realizacja postulatu zgodności jako metoda uniknięcia skutków pozornej zależności (Realization of Congruence Postulate as a Metod of Avoiding the Effects of a Spurious Relationship), *Przegląd Statystyczny (Statistical Survey)*, z. 1, 13-22.

Wooldridge, J. M. (1999), Asymptotic Properties of Some Specification Tests In Linear Models with Integrated Processes, w: R. F. Engle, H. White, *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W. J. Granger*, Oxford University Press, 366–384.

Yule, G. U., Kendall, M. G. (1966), *Wstęp do statystyki* (*Introduction to Statistics*), PWN, Warszawa.

Zieliński, Z. (1984), Zmienność w czasie strukturalnych parametrów modelu ekonometrycznego (Time Variability of Structural Parameters in Econometric Model), *Przegląd Statystyczny (Statistical Survey)*, z. 1/2, 135–148.