**Control and Cybernetics**

vol. **22** (1993) No. 1/2

# Time series forecasting on the basis of Markovian piecewise linear trend models

by

**Leszek Klukowski**

Systems Research Institute
Polish Academy of Sciences
Warsaw
Poland

In the paper two Markovian piecewise time series models are presented. They aim at forecasting of the turning points of the trend and of the slope of the future segments of the trend.

The first model, a univariate one can be represented as the sum of two components: the first component, called Markovian piecewise linear trend, is a discrete semi–Markov process, while the second one is the normal white noise. The realisation of the first component assumes values equal to those of some continuous piecewise linear function (for integer arguments); the sequence of slopes of the function is a realisation of the homogenous Markov chain with finite number of states, while the holding time of each state is a realisation of some random variable associated with this state. Model predictors are constructed for mean square criterion under assumption that incomplete information on current state of the process and its parameters is available only.

The second model, a bivariate one — is a combination of two piecewise trends: the first one is the mentioned above Markovian piecewise trend. The second one is assumed to be some piecewise linear trend (not necessarily continuous or with finite number of possible slopes) that provides information about turning points of the first process. Predictors of the bivariate model are determined under assumptions, similar to those adopted in the univariate model.

Examples of applications of these models for short term forecasting of commodity prices are presented. In the bivariate model the volume of the turnover is included, as the second process.

**Keywords:** Markovian piecewise models, turning points forecasting.

# 1. Introduction

Models presented in this paper are aimed at forecasting of time series with rapid changes of the trend slope. Such time series are typical for many phenomena, e.g. economies witnessing structural changes (especially the East European economies) or speculative markets (commodity markets, stock exchange markets, etc). For such phenomena the use of forecasting methods based on traditional linear time series or econometric models may be not suitable and more advanced models, taking into account changes of parameters, ought to be used. Recently some new approaches have been worked out for this purpose, especially methods based on: Kalman's filter (see e.g. Harvey (1989), Harrison and Stevens (1976)), Markovian switching models — regression and autoregression (see e.g. Goldfeld and Quandt (1973), Tyssedal and Tjostheim (1988)), threshold models (see Tong (1983)), cointegrated ARMA models (see Engle and Granger (1987)), Markovian trend models (see Hamilton (1989)).

Models considered in this paper belong to the last of the groups mentioned i.e. to Markovian trend models. The main idea of these models is that the trend of time series under consideration can be expressed as a sum of two components: the first one, representing the trend of time series, assumes values equal to those of some piecewise continuous linear function (for integer arguments), with slopes being the states of some homogeneous finite state Markov chain, and the second, representing disturbances, is normal white noise. Moreover, it is assumed that the holding time of each state (i.e. the time period the Markov chain spends in some state before making a transition to the next state) is a realisation of some random variable associated with this state. As a result the model of time series is a discrete semi-Markov process.

In the paper two types of Markovian piecewise models are considered : univariate and bivariate. For a bivariate model it is assumed that there exists a time series having turning points in the same instants of time as those of a predicted trend. Moreover, it is assumed that the second time series can be also represented as a sum of piecewise linear trend (not necessarily continuous or with finite number of possible slopes) and normal white noise. The second time series introduces information on turning points occurrence into the model.

Predictors of both models are determined for the mean-square error criterion under two basic assumptions: a) incomplete information about current state of the semi-Markov process is available only (i.e. the probability of being in each state can only be known); b) some parameters of both models are estimated (i.e. they are known with estimation error). These assumptions make the forecasting problem more complex but also more realistic.

It should be also emphasized that the assumption of randomness of holding time of Markov chain states is not applied in models considered by Goldfeld and Quandt (1973), Tyssedal and Tjostheim (1988) and Hamilton (1989), while the assumption on incomplete information (on current state) is taken into account by Hamilton (1989) only.

## 2. The univariate model

The univariate Markovian piecewise trend model (UMT) can be written in the form:

$$X_t = Y_t + \varepsilon_t, \quad t = 1, \dots \tag{1}$$

where: $Y_t$ — semi-Markov discrete process with realisations $y_1, \dots, y_T$ ($T$ — current time instant) of the form:

$$y_t = \begin{cases} a_0 + \beta_i t; & t = t^{(0)} + 1, \dots, t^{(1)} \\ a_1 + \beta_j(t - t^{(1)}); & t = t^{(1)} + 1, \dots, t^{(2)}; \quad \beta_j \neq \beta_i \\ \vdots & \\ a_{m-1} + \beta_k(t - t^{(m-1)}); & t = t^{(m-1)} + 1, \dots, t^{(m)} \\ a_m + \beta_l(t - t^{(m)}); & t = t^{(m)} + 1, \dots, T; \quad \beta_l \neq \beta_k, \end{cases} \tag{2}$$

where: $t^{(\nu)}$ — turning points of the trend, $\beta_i$, $\beta_j$, $\beta_k$, $\beta_l$ — slopes from the set $B = \{\beta_1, \dots, \beta_n\}$, $n < \infty$.

It is assumed that the following assumptions hold:

ASSUMPTION 1 *The sequence of slopes of the trend (i.e. the sequence of "beta's") is a realisation of homogenous Markov chain with the state set $B$ and transition matrix $\boldsymbol{P} = [p_{ij}]$, $(i, j = 1, \dots, n)$. The probabilities $p_{ij}$ have to satisfy the conditions: $0 \leq p_{ij} \leq 1$, $\sum_{j=1}^{n} p_{ij} = 1$ and $p_{ii} = 0$.*

ASSUMPTION 2 *The holding time of every state $r$, $(1 \leq r \leq n)$, is a realisation of the random variable $K_j$, $(j = 1, \dots, n)$, assuming values from the set $\{C_r, C_r + 1, \dots\}$, where $C_r$ — integer number satisfying condition $C_r \geq 2$; the random variables $K_j$ are independent.*

ASSUMPTION 3 *The constants $a_1, \dots, a_m$ of the trend satisfy the conditions: $a_\nu = a_{\nu-1} + \beta_j(t^{(\nu)} - t^{(\nu-1)})$, $(\nu = 1, 2, \dots)$, where $a_0$ is nonrandom constant (this assumption means that the piecewise function generating values $y_1, \dots, y_T$ is continuous).*

ASSUMPTION 4 *The disturbance term $\varepsilon_t$ is the normal white noise $N(0, \sigma_\varepsilon^2)$, uncorrelated with the variables of the process $\{Y_t, t = 1, \dots\}$.*

Distributions of the "future" variables $X_{T+h}$, $(h \geq 1)$, of the process $\{X_t\}$, can be easily determined under assumption that the values of "structural" parameters of the model (i.e. the slopes $\beta_r$, the probabilities $p_{ij}$, distributions of the random variables $K_j$ and variance $\sigma_\varepsilon^2$) as well as the "current state" parameters (the index $l$ of current slope $\beta_l$ and time period beetwen the last turning point $t^{(m)}$ and current time instant $T$) are known. In such a case, predictors of the process $\{X_t\}$ can be obtained directly on the basis of general formulae for the predictor of discrete semi-Markov process (see e.g. Gheorghe (1990)). For example, the mean square predictor $\hat{X}_{T+h}$ has the form of the expected value of the variable $Y_{T+h}$ (because of the fact that $E(\varepsilon_t = 0)$ and that variables of the processes $\{\varepsilon_t\}$ and $\{Y_t\}$ are independent), i.e.

$$\hat{X}_{T+h} = E(Y_{T+h}) = \sum_{\chi_{T+h}} p_{T+h}^{(\cdot)} y_{T+h}^{(\cdot)}, \tag{3}$$

where: $\chi_{T+h}$ — the set of values of the variable $Y_{T+h}$, $y_{T+h}^{(\cdot)}$ — an element of the set $\chi_{T+h}$, $p_{T+h}^{(\cdot)}$ — the probability of an event $\{Y_{T+h} = y_{T+h}^{(\cdot)}\}$, (the forms of $\chi_{T+h}$, $y_{T+h}$, and $p_{T+h}$ are presented in detail in Klukowski (1986)).

However, the assumption that values of structural as well as current state parameters of the process $\{X_t\}$ are known is not always satisfied in practice. This follows from the fact that the disturbance term $\varepsilon_t$ makes it difficult to determine the values of current state parameters, while values of structural parameters are usually to be estimated. Therefore in general one cannot make use of the predictor $\hat{X}_{T+h}$ and it is necessary to determine another predictor under more realistic assumptions. In this paper the following assumptions are used:

(i) the information on current state parameters is received with a certain delay $L$, $(L \geq 1)$, (with regard to current instant $T$);

(ii) the slopes from the set $B$ and the constants $a_\nu$, $(\nu = 0, \ldots, m)$, are estimated, i.e. the estimates $\hat{a}_0, \ldots, \hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n$ and their variances $D^2(\hat{a}_\nu)$, $(\nu = 1, \ldots, m)$, $D^2(\hat{\beta}_j)$, $(j = 1, \ldots, n)$, are known only.

Under assumptions — (i) and (ii) the mean square predictor can be obtained using the Bayesian approach. This predictor is (see Klukowski (1986)) an estimate of the expected value of the random variable $Y_{T+h}$ in the a posteriori distribution with regard to observations $x_{T-L+1}$, (i.e. the last $L$ elements of the time series) and the estimates $\hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n$. The predictor under consideration can be written in the form:

$$\tilde{X}_{T+h} \doteq \sum_{\tilde{\chi}_{T+h}} q_{T+h}^{(\cdot)} \hat{y}_{T+h,L}^{(\cdot)}, \tag{4}$$

where: $\tilde{\chi}_{T+h}$ — the set consisting of the estimates $\hat{y}_{T+h,L}^{(\cdot)}$ of the values $y_{T+h,L}^{(\cdot)}$ of the random variable $Y_{T+h}$; the values $y_{T+h,L}^{(\cdot)}$ are determined under the assumption (ii) — the estimates $\hat{y}_{T+h,L}^{(\cdot)}$ are obtained on the basis of the values $\hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n$; $q_{T+h}$ — probabilities of the events $\{Y_{T+h} = y_{T+h,L}^{(\cdot)}\}$ in the a posteriori distribution.
(The forms of the set $\tilde{\chi}_{T+h}$ and of the a posteriori probabilities $q_{T+h}^{(\cdot)}$ are described in detail in Klukowski (1986) and briefly in Appendix 1).

The variance of the error:

$$\tilde{\varepsilon}_{T+h} = X_{T+h} - \tilde{X}_{T+h},$$

of the predictor $\tilde{X}_{T+h}$, has the form (see Klukowski (1986)):

$$D^2(\tilde{\varepsilon}) = D^2(Y_{T+h}) + D^2(\tilde{X}_{T+h}) + \sigma_\varepsilon^2 \quad (5)$$

where: $D^2(Y_{T+h})$ — the variance of the random variable $Y_{T+h}$, $D^2(\tilde{X}_{T+h})$ — the variance of the predictor $\tilde{X}_{T+h}$.

The following properties of the predictor $\tilde{X}_{T+h}$ have been derived in Klukowski (1986):

a) for the mean square criterion the predictor $\tilde{X}_{T+h}$ is unbiased and optimal if estimates $\hat{a}_m$, $\hat{\beta}_1, \ldots, \hat{\beta}_n$, are unbiased and optimal (for the same criterion);

b) the predictor $\tilde{X}_{T+h}$ is not linear with respect to elements of the time series $x_1, \ldots, x_T$;

c) the probabilities of the a priori distribution result from the current state parameters of the process $\{Y_t\}$ (for the time instant $T - L$) and therefore are not subjective probabilities;

d) in general the distribution of the error $\tilde{\varepsilon}_{T+h}$ — is not symmetric and not the same for different moments $T'$ and $T''$ (in other words the accuracy of forecasts is not the same for forecasts obtained for different time moments $T'$ and $T''$);

e) the form of optimal predictors for the objective functions different from the mean square error (e.g. absolute error, minimax error) is not the same as the form of the predictor $\tilde{X}_{T+h}$ (some examples of such predictors have been proposed in Klukowski (1986)).

Let us note that the properties of predictor $\tilde{X}_{T+h}$, (points c) to e)) do not hold for other time series models, such as regression trend models, ARMA models, exponential smoothing, etc.

The predictor $\tilde{X}_{T+h}$ can be applied in the case when estimates of slopes $\beta_k$ and values of the structural parameters of the model (2) are known. In practice, these parameters are to be estimated on the basis of time series $x_1, \ldots, x_T$. Estimators of the parameters have been described in detail in Klukowski (1986); in the paper a brief discussion of estimation problems is presented only.

To estimate the structural parameters of the process $\{\tilde{X}_t\}$ one has to determine: the number of the states n in Markov chain, the slopes $\beta_j$ from the set $B$, the parameters of the random variables $K_j$, the transition probabilities $p_{ij}$ and the variance $\sigma_\varepsilon^2$.

It seems extremely difficult to obtain estimators of all the structural parameters of the model as a solution to one estimation problem. Therefore it is suggested to estimate individual parameters in four consecutive steps.

**Step 1:** estimation of the turning points $t^{(1)}, \ldots, t^{(m)}$.

Many procedures can be used for this purpose, especially those based on Chow's test, see e.g. Chow (1960), or Coopersmith (1979). Results of this step can also be used to estimate constants and slopes of sequential segments of the trend; estimates of slopes are denoted by $b_1, \ldots, b_m$.

### Step 2: estimation of the number of states $n$.

Assuming that all the possible states have occurred in the time interval $[1, T]$ (i.e. all the slopes from the set $B$), the problem of estimating the number n can be formulated as the classification problem of the form:
Accomplish partition of the set $B = \{b_1, ..., b_m\}$ into nonoverlaping family of subsets $B_1, \ldots, B_{\hat{n}}$ in such a way that each subset $B_k$, $(1 \leq k \leq \hat{n})$ includes parallel slopes only, while slopes from different subsets $B_i$ and $B_j$, $(j \neq i)$, are not parallel. The number $\hat{n}$ of subsets obtained can be regarded as an estimator of the number of states $n$.

The above classification problem can be solved using pairwise comparison algorithms. In the case under consideration the comparisons can be performed using parallelity tests for slopes, i.e. by verifying for each pair $b_i, b_j \in B$ the hypothesis $H_0$: $E(b_i) = E(b_j)$ under the alternative $H_1$: $E(b_i) \neq E(b_j)$, while the partition of the set $B$ can be made using the nearest adjoining order algorithm. Some probabilistic properties of such approach are presented in Klukowski (1990).

### Step 3: estimation of slopes $\beta_j \in B$.

Each slope $\beta_j$ $(1 \leq j \leq \hat{n})$ can be estimated as the average of slopes from the subset $B_j$, obtained in Step 2. However, more efficient approach is to estimate each value $\beta_j$ from one regression equation including appriopriate dummy variables.

The estimates $\hat{a}_m$ and $\hat{\beta}_l$ (i.e. the estimates of the constant and of the slope of the last — for $t \leq T$ — segment of the trend) can be obtained using the least squares method (on the basis of the values $x_{\hat{t}^{(m)}+1}, \ldots, x_T$, where $\hat{t}^{(m)}$ is the estimate of the last turning point of the trend, detected in Step 1).

### Step 4: estimation of the probabilities $p_{ij}$, parameters of the random variables $K_j$, and the variance $\sigma_\varepsilon^2$.

The natural estimator of each probability $p_{ij}$, $(i, j = 1, \ldots, \hat{n})$ is the ratio: the number of cases when an element of the subset $B_i$ is followed by an element of the subset $B_j$ divided by the number of elements of the subset $B_i$.

Parameters of each random variable $K_j$ are to be estimated on the basis of holding times of all slopes from the subset $B_j$, using appriopriate estimators (parametric or nonparametric).

The variance $\sigma_\varepsilon^2$ can be estimated on the basis of residuals obtained during the estimation of slopes $\beta_j$, $(j = 1, \ldots, \hat{n})$ in Step 3. Another approach, however

| Mean sq. error of the method | Leading period | | |
|---|---|---|---|
| | $h = 1$ | $h = 4$ | $h = 7$ |
| UMT — *ex ante* | 120 | 193 | 272 |
| UMT — *ex post* | 108 | 195 | 237 |
| H–S — *ex post* | 94 | 201 | 315 |

Table 1. The mean square errors for the Markovian trend model (UMT) and the Harrison–Stevens method.

less efficient one, is to use residuals related to the estimation of turning points of the trend in Step 1.

The model of the form (2) was applied for short term forecasting of weekly cocoa prices at New York Commodity Exchange. Forecasts were computed for every week of 1984 with seven steps ahead. Parameters of the model were obtained using weekly data from the period 1978–1983.

The *ex ante* and *ex post* mean square errors, obtained for leading periods 1, 4 and 7 are presented in Table 1. For the purpose of comparison, the *ex post* mean square errors obtained on the basis of Harrison–Stevens (H–S) method (see Harrison and Stevens (1976)), are also presented in this table.

The following conclusions can be drawn from the analysis of the forecasts generated (for details see Klukowski (1986)):

- Markovian trend model provides more accurate forecasts than the Harrison–Stevens method for the leading period greater than 1; for $h = 7$ the MSE of the Harrison–Stevens method is by about 1/3 greater than for the Markovian trend model;
- the difference beetween the MSE of Markovian trend model and Harrison–Stevens method is greater when leading period is increased (note that as the Harrison–Stevens method does not predict changes of the trend slope, the resulting forecasts errors are greater for longer leading periods);
- for the Markovian trend model the values of the mean square errors: ex ante and ex post are similar.

## 3. The bivariate model

In some periods forecasts obtained on the basis of the univariate Markovian trend model have relatively large errors. In particular, such errors occur when the values of disturbance term $\varepsilon_t$ in the interval $\langle t^{(m)} + 1, T\rangle$ are extremely high. In such cases information content in univariate time series may be too poor to generate forecasts with appropriate precision. Therefore, some additional data should be incorporated into the model. This can be done by inclusion of an additional variable into the model, connected with some stochastic relations with the forecasted time series.

In this paper it is assumed that such a variable can also be represented by a piecewise linear trend with turning points occurring at the same time instants as those of the time series to be forecasted. Thus the second variable introduces information on turning points into the model. For instance, in the case of the short term commodity prices forecasting, the volume of turnover is applied as the second variable.

The bivariate model can be expressed as the bivariate discrete stochastic process $\{X_{1t}, X_{2t}\}$, where $\{X_{1t}\}$ – the univariate Markovian trend model defined in Section 2, while $\{X_{2t}\}$ – the stochastic process of the form:

$$X_{2t} = Y_{2t} + \xi_t \tag{6}$$

where: $Y_{2t}$ – discrete stochastic process, with realisations $y_{21}, \ldots, y_{2T}$ of the form:

$$y_{2t} = \begin{cases} \delta_0 + \gamma_k t; & & t = t^{(0)} + 1, \ldots, t^{(1)}, \\ \delta_1 + \gamma_l (t - t^{(1)}); & \gamma_l \neq \gamma_k; & t = t^{(1)} + 1, \ldots, t^{(2)}, \\ \vdots & & \\ \delta_{m-1} + \gamma_p (t - t^{(m-1)}); & & t = t^{(m-1)} + 1, \ldots, t^{(m)}, \\ \delta_m + \gamma_q (t - t^{(m)}); & \gamma_q \neq \gamma_p; & t = t^{(m)} + 1, \ldots, T, \end{cases} \tag{7}$$

where: $t^{(\nu)}$ — turning points — the same as in the case of $\{X_{1t}\}$; $\gamma_k$, $\gamma_l$, $\gamma_p$, $\gamma_q$ — the slopes from some set $\Gamma$ (with finite or infinite number of elements); $\delta_\nu$ — constant of the $\nu$-th segment of the trend, $(\nu = 0, \ldots, m)$; $\xi_t$ — normal white noise $N(0, \sigma_\xi^2)$, uncorrelated with the processes $\{Y_{1t}\}$, $\{Y_{2t}\}$ and $\{\varepsilon_t\}$.

The definition of the $\{Y_{2t}\}$ process is more general than that of $\{Y_{1t}\}$, because the piecewise function, generating the values of $y_{2t}$ is not assumed to be continuous and to have the finite set of possible slopes. No assumption is also made on the existence of any relation between slopes of the processes $\{Y_{1t}\}$ and $\{Y_{2t}\}$; the common feature of both piecewise functions are the same turning points only.

The mean square predictor of the bivariate model has been determined under assumptions similar to those made for univariate one, i.e.: (i) the values of structural parameters of the process $\{X_{1t}\}$ are known with exception of the slopes from the set $B$, which are represented by their estimates, (ii) the values of current state parameters of the process $\{X_{1t}\}$ are received with delay $L$, $(L \geq 1)$, (iii) the slopes and constants of the process $\{X_{2t}\}$ are estimated and are known for time instants $t \leq T - L$.

It can be shown (see Klukowski (1988)) that the mean square predictor of the process $\{X_{1t}\}$ has the form:

$$\tilde{X}_{1,T+h} = \hat{E}(Y_{1,T+h} | \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}; \hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n; \hat{\delta}_m, \hat{\gamma}_q), \tag{8}$$

where:

$$\begin{aligned} \boldsymbol{x}^{(1)} &= [x_{1,T-L+1}, \ldots, x_{1,T}]', \\ \boldsymbol{x}^{(2)} &= [x_{2,T-L+1}, \ldots, x_{2,T}]', \end{aligned}$$

| Mean square error | Leading period | | |
|---|---|---|---|
| | $h = 1$ | $h = 4$ | $h = 7$ |
| *Ex ante* | 127 | 164 | 198 |
| *Ex post* | 122 | 147 | 169 |

Table 2. The mean square errors of forecasts resulting from the bivariate model.

$\hat{a}_m$, $\hat{\beta}_1, \ldots, \hat{\beta}_n$, $\hat{\delta}_m$, $\hat{\gamma}_q$ — the estimates of the parameters $a_m$, $\beta_1, \ldots, \beta_n$, $\delta_m$, $\gamma_q$.

The right-hand side of equality (8) is an estimate of the expected value of the random variable $Y_{1,T+h}$ in the a posteriori distribution with respect to vectors $\boldsymbol{x}^{(1)}$ and $\boldsymbol{x}^{(2)}$ and estimates $\hat{a}_m$, $\hat{\beta}_1, \ldots, \hat{\beta}_n$, $\hat{\delta}_m$, $\hat{\gamma}_q$. It can be shown (see Klukowski (1988)) that the predictor $\tilde{X}_{1,T+h}$ has the form:

$$\tilde{X}_{1,T+h} = \sum_{\tilde{\chi}_{T+h}} \tilde{q}^{(\cdot)}_{T+h} \tilde{y}^{(\cdot)}_{1,T+h} \tag{9}$$

where: $\tilde{\chi}_{T+h}$ — the set of estimates $\hat{y}^{(\cdot)}_{1,T+h}$ of the values $y^{(\cdot)}_{1,T+h}$ constituting the set of values of the variable $Y_{1,T+h}$; $\tilde{q}^{(\cdot)}_{T+h}$ — the probabilities of the a posteriori distribution of the variable $Y_{1,T+h}$.

The form of the set $\tilde{\chi}_{T+h}$ is the same as in the case of univariate model, while the form of a posteriori distribution is not, because the conditional distribution consists now of two subvectors $\boldsymbol{X}^{(1)}$ and $\boldsymbol{X}^{(2)}$ (in the univariate case — of $\boldsymbol{X}^{(1)}$ only). The form of this distribution, being quite complex, is described in detail in Klukowski (1988) and briefly in Appendix 2. The form of variance of the forecast error $X_{1,T+h} - \tilde{X}_{1,T+h}$ is also similar to that corresponding to the univariate model.

Estimation of the bivariate model parameters can be performed in similar way as in the case of the univariate one (see Klukowski (1988)).

The bivariate model has been examined using the same data (the cocoa commodity prices) as those for the univariate one. The mean square errors – *ex ante* and *ex post* – of the obtained forecasts, for leading periods 1, 4 and 7 are presented in Table 2.

The comparison of errors resulting from the univariate model, the bivariate model and the Harrison– Stevens method leads to the following conclusions:

The *ex post* mean square errors of forecasts obtained on the basis of the bivariate model are lower than those of the univariate model and the Harrison–Stevens method for leading periods $h > 1$. However, for one step ahead the bivariate model has the greatest error among all the models considered.

The difference among the mean square error of forecasts resulting from the bivariate model and the remaining two models increases with the increase of the leading period, e.g. for $h = 7$ the error corresponding to the bivariate model is about 50% of the error of the Harrison–Stevens method and about 66% of the error corresponding to the univariate model.

The *ex post* errors resulting from the bivariate model are smaller (for all leading periods) than *ex ante* errors. These differences may result from the fact that the *ex ante* error of the bivariate model is a function of the variances $\sigma_\varepsilon^2$ and $\sigma_\xi^2$; the latter variance is of extremely high value.

It should be emphasized, however, that application of the bivariate model significantly increases the computational cost and this approach is possible when a second time series with the same turning points as the predicted one is available.

## 4. Conclusions

The Markovian piecewise models presented in the paper are aimed at forecasting of time series with turning points and changes in the slope of the trend. Main features of the models discussed are:

1.° The optimal predictors of both models can be determined under assumption that information on "structural" parameters and "current state" parameters is incomplete.

2.° The forms of predictors corresponding to different objective functions are in general different.

3.° The empirical study concerning short term forecasting of the cocoa commodity prices for consecutive weeks of the year 1984 has shown that the accuracy of forecasts (measured by the mean square error) of both Markovian models was better than that of the Harrison–Stevens method for leading periods $h \geq 2$; these differences increased with the leading period. The best accuracy has been obtained for the bivariate model for $h \geq 2$; for $h = 7$ the mean square error of this model was about 50% of the error of the Harrison–Stevens method.

## References

CHOW G.,C., (1960) Test of Equality between sets of coefficients in two linear regressions. *Econometrica* **28**, 591–605.

COOPERSMITH L.,W., (1979) Automatic Forecasting Using FLEXICAST *Studies in the Management Sciences* **12** Forecasting, 265–278. North–Holland Publ. Comp. — Amsterdam, N. York, Oxford.

ENGLE R.,F., GRANGER C.,W.,J., (1987) Co–Integration and Error Correction: Representation, Estimation and Testing. *Econometrica* **55**, 251–276.

GHEORGHE A.,V., (1990) Decision Processes in Dynamic Probabilistic Systems. Kluwer Academic Publishers, Dordrecht.

GOLDFELD S.,M., QUANDT R., (1973) A Markov Model for Switching Regressions. JOURNAL OF ECONOMETRICS, **1**, 3–16.

GRANGER C.,W.,J., NEWBOLD P., (1977) Forecasting Economic Time Series. Academic Press, New York.

HAMILTON J.,D., (1989) A new Approach to the Economic Analysis of Nonstationary Time Series and the Bussiness Cycle. *Econometrica*, **57**, 357–384.

HARRISON P.,J., STEVENS C.,F., (1976) Bayesian Forecasting. *Journal of the Royal Statistical Society B*, **38**, 205–247.

HARVEY A.,C., (1989) Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge: Cambridge University Press.

KLUKOWSKI L.,(1986) The Method of Time Series Forecasting with Piecewise Linear Trend (in Polish), unpublished dissertation, Systems Research Institute Polish Academy of Sciences, Warsaw.

KLUKOWSKI L., (1988) Analysis of Application of Univariate and Bivariate Piecewise Linear Model to Short Term Forecasting (in Polish). Technical Report no.30/88/A1102 Systems Research Institute Polish Academy of Sciences, Warsaw.

KLUKOWSKI L., (1990) Algorithm for Classification of Samples in the Case of Unknown Number of Random Variables Generating Them (in Polish). *Przegląd Statystyczny*, **37**, no 3, 167–177.

KLUKOWSKI L., (1992) Time Series Forecasting on the Basis of Markovian Piecewise Linear Trend (in Polish). *Przegląd Statystyczny* **39**, no 3–4, 309–326.

TYSSEDAL J.,S., TJOSTHEIM D., (1988) An Autoregressive Model with Suddenly Changing Parameters and an Application to Stock Market Prices. *Applied Statistics*, **37**, 353–369.

TONG H., (1983) Threshold Models in Nonlinear Time Series Analysis. Lecture Notes in Statistics, **21**. New York Springer.

## Appendix 1. The analytic form of the predictor $\tilde{X}_{T+h}$.

For the sake of simplicity it is assumed that:

$$L, h \leq C_{\min} \tag{A1}$$

where: $C_{\min} = \min\{C_1, \ldots, C_n\}$, ($C_k$ is the minimal holding time of the $k$–th state in Markov chain).

As it was mentioned in Section 2 the predictor $\tilde{X}_{T+h}$ is an estimate of the expected value of the random variable $Y_{T+h}$ in the a posteriori distribution with regard to the last $L$ elements of time series $x_{T-L+1}, \ldots, x_T$ and estimates $\hat{a}_m$, $\hat{\beta}_1, \ldots, \hat{\beta}_n$. To determine the analytic form of expected value of the variable $Y_{T+h}$ one has to know:

- the set $\tilde{\chi}_{T+h}$ comprising the estimates $\hat{y}^{(\cdot)}_{T+h,L}$ of the values $y^{(\cdot)}_{T+h,L}$ of the random variable $Y_{T+h}$,
- the probabilities $q^{(\cdot)}_{T+h}$ of the events $\{Y_{T+h} = y^{(\cdot)}_{T+h,L}\}$ in the *a posteriori* distribution.

Assuming that in the time moment $T-L$ the process $\{Y_t\}$ was in the state $l$, $(1 \le l \le n)$, the set $\chi^{\star}_{T+h}$ of values of the variable $Y_{T+h}$ is of the form:

$$\begin{aligned}\chi^{\star}_{T+h} = {} & \{y^{(0)}_{T+h,L}\} \cup \{y^{(\tau_j)}_{T+h,L};\tau = 1,\dots,L+h; j \neq 1\} \cup \{y^{(\tau_{jdr})}_{T+h,L};\\ & \tau=C_j,\dots,L+h-1; j\neq 1; d=1,\dots,L+h-\tau; r\neq j\},\end{aligned} \tag{A2}$$

where:

$$\begin{aligned}y^{(0)}_{T+h,L} &= a_m + (N+L+h)\beta_l;\\ y^{(\tau_j)}_{T+h,L} &= a_m + (N+L+h-\tau)\beta_l + \tau\beta_j;\\ y^{(\tau_{jdr})}_{T+h,L} &= a_m + (N+L+h-\tau-d)\beta_l + \tau\beta_j + d\beta_r.\end{aligned}$$

The estimates $\hat{y}^{(\cdot)}_{T+h,L}$ constituting the set $\tilde{\chi}_{T+h}$ have the similar form as the values $y^{(\cdot)}_{T+h,L}$, but the parameters $a_m$, $\beta_l$. $\beta_j$, $\beta_r$ have to be replaced by their estimates: $\hat{a}_m$, $\hat{\beta}_l$, $\hat{\beta}_j$, $\hat{\beta}_r$.

The probabilities $q^{(\cdot)}_{T+h}$ are determined on the basis of both the a priori distribution and the conditional distribution.

The probabilities $p^{(\cdot)}_{T+h}$ of the a priori distribution have the form:

$$\left.\begin{aligned}p^{(0)} &= P(Y_{T+h} + y^{(0)}_{T+h,L}) = P(K_l \ge N+L+h | K_l \ge N),\\ p^{(\tau_j)} &= P(Y_{T+h} + y^{(\tau_j)}_{T+h,L}) = p_{lj}\, P(K_l = N+L+h|K_l \ge N)\cdot\\ &\quad \cdot P(K_j \ge \tau),\\ p^{(\tau_{jdr})} &= P(Y_{T+h} + y^{(\tau_{jdr})}_{T+h,L}) = p_{lj}p_{jr}P(K_l = N+h-\tau-d|\\ &\quad |K_l \ge N)P(K_j = \tau)P(K_r \ge d),\end{aligned}\right\} \tag{A3}$$

(probabilities $p_{lj}$, $p_{jr}$ are elements of the transition matrix $\boldsymbol{P}$).

The conditional distribution is constructed as the distribution of the random vector $\boldsymbol{U}^{(\cdot)}$ of the form:

$$\boldsymbol{U}^{(\cdot)} = \boldsymbol{X}_w - \hat{\boldsymbol{\mu}}^{(\cdot)}, \tag{A4}$$

where: $\boldsymbol{X}_w = [X_{T-L+1},\dots,X_T]'$, $\hat{\boldsymbol{\mu}}^{(\cdot)}$ – an estimate of the expected value of the vector $\boldsymbol{X}_w$, obtained on the basis of the estimates $\hat{a}_m$, $\hat{\beta}_l$, $\hat{\beta}_j$, $\hat{\beta}_r$, determined by the value $y^{(\cdot)}_{T+h,L}$, of the variable $Y_{T+h}$, i.e.:

- for $Y_{T+h} = y^{(0)}_{T+h,L}$ the vector $\boldsymbol{\mu}^{(0)}$ consists of elements $\mu^{(0)}_s$ of the form:
  $\mu^{(\cdot)}_s = \hat{a}_m + (N+s)\hat{\beta}_l, \quad s = 1,\dots,L;$

– for $Y_{T+h} = y^{(\tau_j)}_{T+h,L}$ the vector $\boldsymbol{\mu}^{(\tau_j)}$ consists of elements $\mu_s^{(\tau_j)}$ of the form:

$$\mu_s^{(\tau_j)} = \begin{cases} a_m + (N+s)\beta_l, & s \le L+h-\tau, \\ a_m + (N+L+h-\tau)\beta_l + (s-L-h-\tau)\beta_j; \\ \qquad L+h-\tau < s \le L, \end{cases}$$

– for $Y_{T+h} = y^{(\tau_{jd\tau})}_{T+h,L}$ the vector $\boldsymbol{\mu}^{(\tau_{jd\tau})}$ consists of elements $\mu_s^{(\tau_{jd\tau})}$ of the form:

$$\mu_s^{(\tau_{jd\tau})} = \begin{cases} a_m + (N+s)\beta_j, & s \le L+h-\tau-d, \\ a_m + (N+L+h-\tau-d)\beta_l + (s-L-h+\tau+d)\beta_j; \\ \qquad L+h-\tau-d < s \le L. \end{cases}$$

Assuming that estimates $\hat{a}_m$ and $\hat{\beta}_l$ have been obtained on the basis of least squares method (what implies that $\hat{a}_m \sim N(a_m, \mathrm{var}(\hat{a}_m))$, $\hat{\beta}_l \sim N(\beta_l, \mathrm{var}(\hat{\beta}_l))$, where $\mathrm{var}(\hat{a}_m) = \frac{2(2N+1)}{(N(N-1))}\sigma_\varepsilon^2$, $\mathrm{var}(\hat{\beta}_l) = \frac{6}{(N(N^2-1))}\sigma_\varepsilon^2$ and $\mathrm{cov}(\hat{a}_m, \hat{\beta}_l) = -\frac{6}{n(n-1)}\sigma_\varepsilon^2)$) the random vector $\boldsymbol{U}^{(0)}$ has the ($L$-dimensional) normal distribution $N(0, \Omega^{(0)})$, $\Omega^{(0)} = [\omega_{ij}^{(0)}]$, $(i,j = 1, \ldots, L)$, where:

$$\left.\begin{aligned} \omega_{ii}^{(0)} &= \mathrm{var}(\hat{a}_m) + (N+s)^2\,\mathrm{var}(\hat{\beta}_l) + 2(N+i)\,\mathrm{cov}(\hat{a}_m, \hat{\beta}_l) + \sigma_\varepsilon^2 \\ \omega_{ij}^{(0)} &= \mathrm{var}(\hat{a}_m) + (N+i)(N+j)\,\mathrm{var}(\hat{\beta}_l) + 2(N+i+j)\,\mathrm{cov}(\hat{a}_m, \hat{\beta}_l), \\ & \quad j \ne i, \end{aligned}\right\} \tag{A5}$$

each of the vectors $\boldsymbol{U}^{(\tau_j)}$ has the normal distribution $N(0, \Omega^{(\tau_j)})$, $\Omega^{(\tau_j)} = [\omega_{ij}^{(\tau_j)}]$, where:

$$\omega_{ii}^{(\tau_j)} = \begin{cases} \mathrm{var}(\hat{a}_m) + (N+i)^2\,\mathrm{var}(\hat{\beta}_l) + 2(N+s)\,\mathrm{cov}(\hat{a}_m, \hat{\beta}_l) + \sigma_\varepsilon^2; \\ \qquad i \le L+h-\tau; \\ \mathrm{var}(\hat{a}_m) + (N+L+h-\tau)^2\,\mathrm{var}(\hat{\beta}_l) \\ +2(N+L+h-\tau)\,\mathrm{cov}(\hat{a}_m, \hat{\beta}_l) + \\ +(i-L-h+\tau)^2\,\mathrm{var}(\hat{\beta}_l) + \sigma_\varepsilon^2; \quad L+h-\tau < i \le L,; \end{cases} \tag{A6}$$

$$\omega_{ij}^{(\tau_j)} = \begin{cases} \mathrm{var}(\hat{a}_m) + (N+i)(N+j)\,\mathrm{var}(\hat{\beta}_l) \\ +(2N+i+j)\,\mathrm{cov}(\hat{a}_m, \hat{\beta}_l); \\ \qquad i,j \le L+h-\tau; \\ \mathrm{var}(\hat{a}_m) + (N+i)(N+L-\tau)\,\mathrm{var}(\hat{\beta}_l) \\ +2(N+i+j+L+h-\tau)\,\mathrm{cov}(\hat{a}_m, \hat{\beta}_l); \\ \qquad i \le L+h-\tau < j \text{ or } j \le L+h-\tau < i; \\ \mathrm{var}(\hat{a}_m) + (N+L)\,\mathrm{var}(\hat{\beta}_l) \\ +2(N+L+h-\tau)\,\mathrm{cov}(\hat{a}_m, \hat{\beta}_l); \\ +(i-L-h+\tau)\,\mathrm{var}(\hat{\beta}_l); \quad i,j > L+h-\tau,\ i \ne j \end{cases} \tag{A7}$$

each of the vectors $\boldsymbol{U}^{(\tau_{jdr})}$ has the normal distribution $N(0, \Omega^{(\tau_{jdr})})$, $\Omega^{(\tau_{jdr})} = [\omega_{ij}^{(\tau_{jdr})}]$, where:

$$\omega_{ii}^{(\tau_{jdr})} = \begin{cases} \operatorname{var}(\hat{a}_m) + (N+i)^2 \operatorname{var}(\hat{\beta}_l) + 2(N+i)\operatorname{cov}(\hat{a}_m, \hat{\beta}_l) + \sigma_\varepsilon^2; \\ \qquad i \le L + h - \tau - d; \\ \operatorname{var}(\hat{a}_m) + (N + L + h - \tau - d)^2 \operatorname{var}(\hat{\beta}_l) \\ +2(N + L + h - \tau - d)\operatorname{cov}(\hat{a}_m, \hat{\beta}_l) + \\ +(i - L - h + \tau + d)^2 \operatorname{var}(\hat{\beta}_l) + \sigma_\varepsilon^2; \\ \qquad L + h - \tau - d < i \le L; \end{cases} \tag{A8}$$

$$\omega_{ij}^{(\tau_{jdr})} = \begin{cases} \operatorname{var}(\hat{a}_m) + (N+i)(N+j)\operatorname{var}(\hat{\beta}_l) \\ +(2N + i + j)\operatorname{cov}(\hat{a}_m, \hat{\beta}_l); \\ \qquad i \ne j;\ i, j \le L + h - \tau - d; \\ \operatorname{var}(\hat{a}_m) + (N+i)(N + L - \tau)\operatorname{var}(\hat{\beta}_l) \\ +2(N + i + j + L + h - \tau)\operatorname{cov}(\hat{a}_m, \hat{\beta}_l); \\ \qquad i \le L + h - \tau < j \text{ or } j \le L + h - \tau < i; \\ \operatorname{var}(\hat{a}_m) + (N + L + h - \tau)^2 \operatorname{var}(\hat{\beta}_l) \\ +2(N + L + h - \tau)\operatorname{cov}(\hat{a}_m, \hat{\beta}_l); \\ +(i - L - h + \tau)(j - L - h + \tau)\operatorname{var}(\hat{\beta}_l); \\ \qquad i \ne j; i, j > L + h - \tau. \end{cases} \tag{A9}$$

Denoting the probability distribution function of the vectors $\boldsymbol{U}^{(\cdot)}$ by $g^{(\cdot)}(\boldsymbol{u}^{(\cdot)} | Y_{T+h} = y_{T+h,L}^{(\cdot)}; \hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n)$, one can express the probabilities $q_{T+h}^{(\cdot)}$ of a posteriori distribution in the form:

$$q_{T+h}^{(0)} = \frac{p_{T+h}^{(0)} g^{(0)}(\boldsymbol{u}^{(0)} | Y_{T+h} = y_{T+h,L}^{(o)}; \hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n)}{Q}, \tag{A10}$$

$$q_{T+h}^{(\tau_j)} = \frac{p_{T+h}^{(\tau_j)} g^{(\tau_j)}(\boldsymbol{u}^{(\tau_j)} | Y_{T+h} = y_{T+h,L}^{(\tau_j)}; \hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n)}{Q}, \tag{A11}$$

$$q_{T+h}^{(\tau_{jdr})} = \frac{p_{T+h}^{(\tau_{jdr})} g^{(\tau_{jdr})}(\boldsymbol{u}^{(\tau_{jdr})} | Y_{T+h} = y_{T+h,L}^{(\tau_{jdr})}; \hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n)}{Q} \tag{A12}$$

where:

$$Q = \sum p_{T+h}^{(\cdot)} g^{(\cdot)}(\boldsymbol{u}^{(\cdot)} | Y_{T+h} = y_{T+h,L}^{(\cdot)}; \hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n).$$

The formulae (A2), (A10)–(A12) make it possible to evaluate the value of the predictor $\tilde{X}_{T+h}$ discussed in Section 2.

## Appendix 2. The analytic form of the predictor $\tilde{X}_{1,T+h}$

For the bivariate model the form of predictor $\tilde{X}_{1,T+h}$ has been also determined under assumption that the inequality (A1) from Appendix 1 is satisfied.

The bivariate predictor is an estimate of the expected value of the random variable $Y_{1,T+h}$ in the a posteriori distribution with regard to the vectors $\boldsymbol{x}^{(1)}$, $\boldsymbol{x}^{(2)}$ and the estimates $\hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n, \hat{\delta}_m, \hat{\gamma}_q$. To determine the value of this predictor it is neccesary to know: the set $\tilde{\chi}_{1,T+h}$ of estimates $\hat{y}^{(\cdot)}_{1,T+h}$ of the values $y^{(\cdot)}_{1,T+h}$ of the random variable $Y_{1,T+h}$, the a priori distribution on this set and the relevant conditional distribution. The form of the set of estimates $\tilde{\chi}_{1,T+h}$ and a priori distribution is the same as in the case of the univariate model (see formula (A2) and (A3) in Appendix 1)). As the basis for conditional distribution the random vector:

$$\boldsymbol{Z}^{(\cdot)} = \begin{bmatrix} \boldsymbol{U}^{(\cdot)} \\ \boldsymbol{V}^{(\cdot)} \end{bmatrix} \tag{A13}$$

is applied, where: $\boldsymbol{U}^{(\cdot)}$ — the same vector as in the univariate model (see (A4)); $\boldsymbol{V}^{(\cdot)}$ — the vector with elements $v_k$ of the form:

$$v_k = X_{2,t(m)+k} - \mu^{(v)}_k, \quad k = N+1, \ldots N+I \tag{A14}$$

where:

$$\mu^{(v)}_k = \hat{\delta}_m + \hat{\gamma}_q k$$

$$i = \begin{cases} L; \quad \text{for } Y_{1,T+h} = y^{(0)}_{1,T+h} \\ L - \tau,\ (\tau = 1, \ldots, L+h); \quad \text{for } Y_{1,T+h} = y^{(\tau_j)}_{1,T+h} \\ L - \tau - d,\ (\tau = C, \ldots, L+h-1;\ d = 1, \ldots, L+h-\tau); \\ \qquad\qquad\qquad \text{for } Y_{1,T+h} = y^{(\tau_{jdr})}_{1,T+h} \end{cases}$$

(the meaning of symbols: $\tau$, $j$, $d$ and $r$ is the same as in the univariate model; $y^{(\cdot)}_{1,T+h}$ is equal to $y^{(\cdot)}_{T+h,L}$ in the univariate model — see formulae (A2)).

Under the assumptions adopted in Section 3, the vector $\boldsymbol{Z}^{(\cdot)}$ has $L+I$ — dimensional normal distribution with expected value equal to zero and covariance matrix $\Omega^{(\cdot)}_Z$ of the form:

$$\Omega^{(\cdot)}_Z = \begin{bmatrix} \Omega^{(\cdot)}_U & 0 \\ 0 & \Omega^{(I)}_V \end{bmatrix}, \tag{A15}$$

where: $\Omega^{(\cdot)}_U$ — the same matrix as in the univariate model (see (A5)-(A9)), $\Omega^{(I)}_V$ — the covariance matrix of the subvector $\boldsymbol{V}^{(I)}$ with elements $\omega^{(I)}_{ij}$, $(i,j = 1, \ldots, I)$, of the form:

$$\omega^{(I)}_{ii} = \operatorname{var}(\hat{\delta}_m) + (N+i)\operatorname{var}(\hat{\gamma}_q) + 2(N+i)\operatorname{cov}(\hat{\delta}_m, \hat{\gamma}_q); \quad i = 1, \ldots, I$$
$$\omega^{(I)}_{ij} = \operatorname{var}(\hat{\delta}_m) + (N+i)(N+j)\operatorname{var}(\hat{\gamma}_q) + (2N+i+j)\operatorname{cov}(\hat{\delta}_m, \hat{\gamma}_q); \quad i \neq i$$

(the values of $\hat{\delta}_m$, $\hat{\gamma}_q$, $\mathrm{var}(\hat{\delta}_m)$, $var(\hat{\gamma}_q)$ and $\mathrm{cov}(\hat{\delta}_m, \hat{\gamma}_q)$ can be obtained in similar way as values of $\hat{a}_m$, $\hat{\beta}_l$, $\mathrm{var}(\hat{a}_m)$, $var(\hat{\beta}_l)$ and $\mathrm{cov}(\hat{a}_m, \hat{\beta}_l)$).

Due to the fact of block-diagonal form of the matrix $\Omega_Z^{(\cdot)}$ (resulting from the assumption on independency of disturbances $\varepsilon_t$ and $\xi_t$) the probability distribution function (pdf) $g^{(\cdot)}(\boldsymbol{z}|\cdot)$ of the vector $\boldsymbol{Z}$ is the product of pdf's of the vectors $\boldsymbol{U}^{(\cdot)}$ and $\boldsymbol{V}^{(I)}$, i.e., has the form:

$$
\begin{aligned}
&g_{UV}^{(\cdot)}(\boldsymbol{z}^{(\cdot)}|Y_{1,T+h} = y_{1,T+h}^{(\cdot)}; \hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n; \hat{\delta}_m, \hat{\gamma}_q) = \\
&= g^{(\cdot)}(\boldsymbol{u}^{(\cdot)}|Y_{1,T+h} = y_{1,T+h}^{(\cdot)}; \hat{a}_m, \hat{\beta}_1, \ldots, \hat{\beta}_n) \cdot \\
&\quad \cdot g_1^{(I)}(\boldsymbol{v}^{(I)}|Y_{1,T+h} = y_{1,T+h}^{(\cdot)}; \hat{\delta}_m, \hat{\gamma}_q) = \cdot \\
&= (2\pi)^{-(L+I)/2} |\Omega_U^{(\cdot)}|^{-1/2} \exp\{-\frac{1}{2}\boldsymbol{u}^{(\cdot)\prime}(\Omega_U^{(\cdot)})^{-1}\boldsymbol{u}^{(\cdot)}\} \cdot \\
&\quad \cdot |\Omega_V^{(I)}|^{-1/2} \exp\{-\frac{1}{2}\boldsymbol{v}^{(I)\prime}(\Omega_V^{(I)})^{-1}\boldsymbol{v}^{(I)}\}.
\end{aligned} \tag{A16}
$$

The number of elements of the vector $\boldsymbol{Z}^{(\cdot)}$ is equal $L+I$ and therefore is not the same for different values of the variable $Y_{1,T+h}$. Thus the realisation of vector $\boldsymbol{Z}^{(\cdot)}$ is a conjunction of $L+I$ random events. As a result, the values of function $g_{UV}^{(\cdot)}(\boldsymbol{z}|\cdot)$, corresponding to different values of $L+I$ (i.e. different number of components of the vector $\boldsymbol{Z}^{(\cdot)}$) are incomparable. Hence the distribution of the vector $\boldsymbol{Z}^{(\cdot)}$ cannot be (directly) applied as the conditional one. The natural way to assure comparability of functions $g_{UV}^{(\cdot)}(\boldsymbol{z}|\cdot)$, for different $L+I$ is to average the values of these functions with respect to $L+I$. An appriopriate way of such averaging is to apply the geometrical average (because the realisation of the vector $\boldsymbol{Z}^{(\cdot)}$ is a conjunction of $L+I$ random events). In such a case the pdf of conditional distribution has the form:

$$
f^{(\cdot)}(\boldsymbol{z}|\cdot) = [g_{UV}^{(\cdot)}(\boldsymbol{z}|\cdot)]^{1/(L+1)}. \tag{A17}
$$

The formula (A17) makes it possible to determine probabilities $\hat{q}_{T+h}^{(\cdot)}$ of the a posteriori distribution in similar way as in the case of the univariate model i.e. replacing the values $g^{(\cdot)}(\boldsymbol{u}|\cdot)$ with these $f^{(\cdot)}(\boldsymbol{z}^{(\cdot)}|\cdot)$ in formulas (A10)–(A12).