ОПТИМАЛЬНОСТЬ И РОБАСТНОСТЬ В СТАТИСТИЧЕСКОМ ПРОГНОЗИРОВАНИИ

Ю. С. Харин

Белорусский государственный университет
Научно-исследовательский институт прикладных проблем математики и информатики
Минск, Беларусь
e-mail: kharin@bsu.by

Рассматривается проблема статистического прогнозирования (предсказания) временных рядов для типовой на практике ситуации, когда искажена гипотетическая модель данных. Дается обзор наших решений для следующих актуальных проблем оптимальности и робастности в статистическом прогнозировании: математическое описание искажений типовых гипотетических моделей временных рядов; количественная оценка робастности риска (анализ чувствительности) при наличии искажений для традиционных прогнозирующих статистик (которые минимизируют риск для гипотетических моделей); оценивание критических уровней искажений; построение новых робастных прогнозирующих статистик.

Ключевые слова: статистические прогнозирование; экстраполяция; временной ряд; искажение; робастность; прогнозирующая статистика.

OPTIMALITY AND ROBUSTNESS IN STATISTICAL FORECASTING

Yu. S. Kharin

Belarusian State University, Research Institute for Applied Problems of Mathematics and Informatics
Minsk, Belarus

The statistical forecasting (prediction) problem of time series is considered for the typical in practice situation where the underlying hypothetical data model is distorted. We present a review of our solutions for the following topical problems of optimality and robustness in statistical forecasting: mathematical description of distortions for typical hypothetical models of time series; quantitative evaluation of the risk-robustness (sensitivity analysis) under distortions for traditional forecasting statistics (that are optimal under hypothetical models); evaluation of critical distortion levels; construction of new robust forecasting statistics.

Keywords: statistical forecasting; prediction; time series; distortion; robustness; forecasting statistic.

1. INTRODUCTION

Many significant applied problems in economics, finance, medicine and other fields lead to the global problem in mathematical statistics: statistical forecasting of random processes with discrete time, that are usually called random sequences or time series [1–6].

Mathematical substance of the Forecasting (Prediction) Problem is very simple: to estimate (evaluate) the future value of the random process $x_{T+\tau} \in R^d$ in $\tau \geq 1$ steps ahead, based on the before observed data (fig. 1).

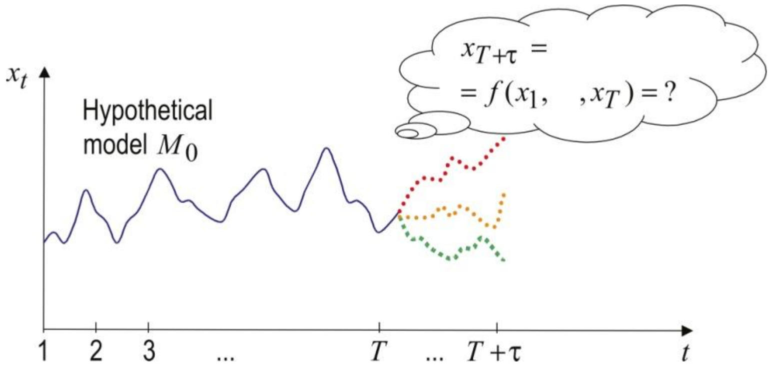

Fig 1. Illustration of the Forecasting Problem

We can clearly detect two stages in the history of attacking the Forecasting Problem [1–6].

First stage (up to the year 1974): construction of optimal forecasting statistics $\hat{x}_{T+\tau} = f_0(x_1, \ldots, x_T): R^{dT} \to R^d$ for various hypothetical models M_0 w.r.t. to minimization of some forecast risk functional (e. g., mean square error of forecasting). A. N. Kolmogorov was the first who considered the forecasting problem in strict mathematical form [1].

Second stage (from the year 1974). It was detected and announced by P. Huber at his lecture on the Vancouver International Congress of Mathematicians [2] that «statistical inferences (including forecasts) are based only in part upon the observations; an equally important base is formed by prior assumptions about the underlying situation (that is the hypothetical model M_0)». In practice, the hypothetical models are usually distorted, and the risk of optimal forecasting statistic (that was constructed under M_0) becomes much more than the hypothetical risk. It was proposed by P. Huber to construct *robust* statistical forecasts that «are weakly sensitive w.r.t. small distortions of the hypothetical model M_0».

A list of researchers influencing the field of robust statistical analysis of time series is given here: J. Tukey, P. Huber, F. Hampel, C. Croux, R. Dahlhaus, P. Filzmoser, R. Fried, R. Dutter, V. Gather, M. Genton, Yu. Kharin, R. Maronna, R. D. Martin, S. Morgethaler, C. Mueller, D. Pena, G. Tiao, H. Rieder, E. Ronchetti, P. J. Rousseeuw, R. Tsay, W. Wefelmeyer, V. J. Yohai. The majority of publications on robustness in statistical time series analysis are concentrated on estimation of parameters and hypotheses testing. Although these problems are fundamentals, they do not completely cover the problem of robustness in statistical forecasting of distortions that includes the following topical tasks considered in this talk: A) mathematical description for typical hypothetical models of time series; B) quantitative evaluation of the risk-robustness (sensitivity analysis) under distortions for traditional forecasting statistics (that are optimal under hypothetical models); C) evaluation of critical distortion levels; D) construction of new robust forecasting statistics.

2. DISTORTIONS OF HYPOTHETICAL MODELS

Introduce the notation: $x_{T+\tau} \in \mathrm{R}^d$ is an observed d-variate time series with discrete time $t \in \mathrm{Z}$, $X = \left(x'_1, \ldots, x'_T\right)' \in \mathrm{R}^{Td}$ is the composed vector of observations for T successive time points (prime means transposition), $x_{T+\tau} \in \mathrm{R}^d$ is the non-observable random vector to be predicted at the future time point $T+\tau$, $\tau \in \mathrm{N}$ (in economic applications the value T is called «the base of forecasting», τ – «the horizon of forecasting»). The probability model of the observed time series under distortions is determined by a family of probability measures

$$\left\{P^{\varepsilon}_{T,\,\theta^0}\left(A\right),\ A \in B^{Td}: T \in \mathrm{N},\ \theta^0 \in \Theta \subseteq \mathrm{R}^m,\ \varepsilon \in \left[0, \varepsilon_+\right]\right\}, \qquad (1)$$

where B^{Td} is the Borel σ-algebra in R^{Td}, θ^0 is an unknown true value of model parameters, ε is the distortion level, $\varepsilon_+ \geq 0$ is its maximal admissible value. If $\varepsilon_+ = 0$, then the distortions are absent, and we have the hypothetical model M_0.

Classification and mathematical description of typical for practice kinds of distortions (1) is given in [3]. In fig. 2 we present a short version of the classification scheme from [6].

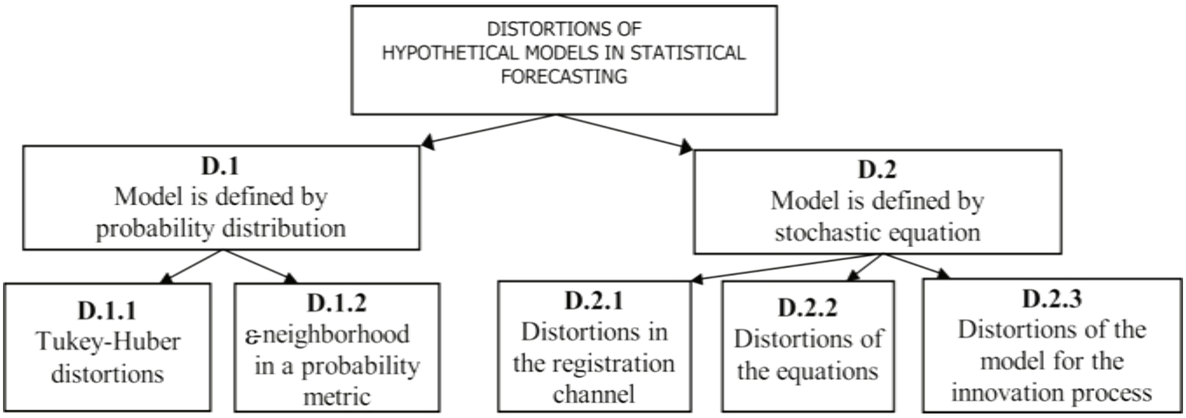

Fig. 2. Classification scheme for types of distortions

Let us give short mathematical descriptions for classes of distortions in fig. 2.

The class of distortions D.1 consists of two subclasses. *Tukey-Huber distortions D.1.1* for the observation vector X are described by the mixture: $p\left(X\right) = \left(1-\varepsilon\right)p^0\left(X\right) + \varepsilon \cdot h\left(X\right)$, $p^0\left(\cdot\right)$ is «non-distorted» (hypothetical) p.d.f., $h\left(\cdot\right)$ is so-called «contaminating» p.d.f., $\varepsilon \in \left[0,\ 1\right)$ is the distortion level. If $\varepsilon = 0$, then $p\left(\cdot\right) = p^0\left(\cdot\right)$, and distortions are absent.

Distortions of the type D.1.2 are described by ε-neighborhoods in some probability metric: $0 \leq \rho\left(p\left(\cdot\right), p^0\left(\cdot\right)\right) \leq \varepsilon$, $\rho(\cdot)$ is a probability metric, e. g. Kolmogorov and Hellinger probability metrics:

$$\rho\left(p, p^0\right) = \frac{1}{2}\int_{\mathrm{R}^{Td}} \left|p\left(X\right) - p^0\left(X\right)\right| dX \in \left[0,\ 1\right];$$

$$\rho\left(p,\, p^0\right) = \frac{1}{2}\int_{\mathrm{R}^{Td}} \left(\sqrt{p(X)} - \sqrt{p^0(X)}\right)^2 dX \in [0,\, 1].$$

The class of distortions D.2 describes the hypothetical model by some stochastic equation:

$$x_t = G\left(x_{t-1},\ \ldots,\ x_{t-s},\ u_t,\ u_{t-1},\ u_{t-L};\ \theta^0\right),\ t \in \mathrm{Z},$$

where $u_t \in \mathrm{R}^{\nu}$ is innovation process on the probability space $(\Omega,\ F,\ P)$, $s, L \in \mathrm{N}$ are some natural numbers indicating the memory depth, $\theta^0 \in \Theta \subseteq \mathrm{R}^m$ is vector of model parameters; $G(\cdot) : \mathrm{R}^{ds} \times \mathrm{R}^{\nu(L+1)} \times \Theta \to \mathrm{R}^d$ is some Borel function.

The class D.2 consists of 3 subclasses. *The subclass D.2.1* describes distortions in the observation channel: $X = H\left(X^0,\ U\right)$, where $X^0 \in \mathrm{R}^{Td}$ is «non-observable prehistory» of the process, $X \in \mathrm{R}^{Td}$ is observation results, that is the «observable prehistory», $U = \left(u'_1,\ \ldots, u'_T\right) \in \mathrm{R}^{Td}$ is non-observable random vector of distortions (errors in the observation channel), $H(\cdot)$ is a function that describes the registration algorithm.

Subclass D.2.1 includes following types of distortions: additive, multiplicative, ε-non-homogeneities, «outliers», «missing values», censoring, etc. *Subclass D.2.2* describes distortions of the generating stochastic equation («misspecification»), includes two types of distortions: *parametric distortions*, when instead of the true parameter value θ^0 we get (estimate by statistical data) a different value $\tilde{\theta}$, with $\left|\tilde{\theta} - \theta^0\right| \le \varepsilon$, where ε is the distortion level; *functional distortions*, when instead of the true function $G(\cdot)$ we get a different function $\tilde{G}(\cdot)$, and in some metric $\left\|\tilde{G}(\cdot) - G(\cdot)\right\| \le \varepsilon$. *Subclass D.2.3* describes distortions of the innovation process $u_t \in \mathrm{R}^{\nu}$, $t \in \mathrm{Z}$, in the generating stochastic equation and includes distortions of three types: ε-non-homogeneities, probabilistic dependence, «outliers».

Note, that in practice, the real data can be corrupted by distortions of two or more indicated classes simultaneously, e. g. outliers and missing values.

3. FUNCTIONALS OF ROBUSTNESS IN STATISTICAL FORECASTING

Let $x_{T+\tau} = f_{T,\,\tau}(X) : \mathrm{R}^{Td} \to \mathrm{R}^d$ be a forecasting statistic (Borel function). Introduce the notation: $\mathbf{E}_{\varepsilon}\{\cdot\}$ is expectation w.r.t. to the measure $P^{\varepsilon}_{T,\,\theta^0}(\cdot)$ from (1); $\pi(\theta)\colon\ \Theta \to \mathrm{R}^1$ is some p.d.f. on Θ; $\hat{\theta} \in \mathrm{R}^m$ is some consistent estimator of θ^0 by X. To analyze optimality and robustness of forecasting statistics we will use the following functionals.

Point risk of forecasting: $\rho_{\varepsilon} = \rho_{\varepsilon}\left(f_{T,\,\tau};\ \theta^0\right) = \mathbf{E}_{\varepsilon}\left\{\left\|\hat{x}_{T+\tau} - x_{T+\tau}\right\|^2\right\} \ge 0$.

Integral risk of forecasting: $r_{\varepsilon} = r_{\varepsilon}\left(f_{T,\,\tau}\right) = \int_{\Theta} \rho_{\varepsilon}\left(f_{T,\,\tau};\ \theta\right)\pi(\theta)\,d\theta \ge 0$.

Guaranteed (upper) risk: $r_{+} = r_{+}\left(f_{T,\,\tau}\right) = \sup\limits_{0 \le \varepsilon \le \varepsilon_{+}} r_{\varepsilon}\left(f_{T,\,\tau}\right)$.

Optimal forecasting statistic (for the hypothetical model M_0): $\hat{x}^0_{T,\tau} = f^0_{T,\tau}\left(X;\ \theta^0\right)$ with the hypothetical risk:

$$\rho_0\left(f^0_{T,\tau};\ \theta^0\right) = \inf_{f_{T,\tau}} \rho_0\left(f_{T,\tau};\ \theta^0\right);\ r_0 = \int_{\Theta} \rho_0\left(f^0_{T,\tau};\ \theta\right)\pi(\theta)d\theta. \qquad (2)$$

"Plug-in" forecasting statistic:

$$\hat{x}_{T+\tau} = f_{T,\tau}(X) = f^0_{T,\tau}\left(X;\ \hat{\theta}\right), \qquad (3)$$

Coefficient of risk instability ($r_0 > 0$) [7]:

$$\kappa = \kappa\left(f_{T,\tau}\right) = \left(r_+\left(f_{T,\tau}\right) - r_0\right)/r_0 \geq 0. \qquad (4)$$

δ*-admissible (critical) distortion level* ($\delta > 0$) [7]:

$$\varepsilon^* = \varepsilon^*(\delta) = \sup\left\{\varepsilon:\ \kappa\left(f_{T,\tau}\right) \leq \delta\right\}. \qquad (5)$$

Minimax risk-robust forecasting statistic $\hat{x}_{T+\tau} = f^*_{T,\tau}(X)$ [7]:

$$\kappa\left(f^*_{T,\tau}\right) = \inf_{f_{T,\tau}(\cdot)} \kappa\left(f_{T,\tau}\right). \qquad (6)$$

4. MAIN RESULTS

By the functionals of robustness (2)–(6) given in Section 3 we solve [3–10] (and present results of solution in this talk) the problems A-D indicated in Section 1 for the following typical in practice distortions of hypothetical models:

- regression time series under "additive outliers";
- $\mathrm{AR}(p)$ model under non-homogenety in mean;
- $\mathrm{AR}(p)$ model under bilinear distortion;
- $\mathrm{AR}(p)$ model under simultaneously influencing "outliers" and missing values.

5. CONCLUSION

The results given in this presentation provide a Statistician (a Forecaster) with quantitative estimates of the guaranteed upper risk, with the risk instability coefficient and with the δ-admissible (critical) distortion level for the «plug-in» statistical forecasting of time series under some typical distortions of the underlying hypothetical models, i.e. trend, regression and autoregressive models. These estimates reveal the influence of distortions on the risk, indicate the risk limits of the safe forecasting by traditional algorithms under distortions, and outline approaches to the robustification of forecasting procedures.

Some new minimax risk-robust forecasting statistics are constructed and compared to the traditional forecasting algorithms.

The theoretical results have been tested on simulated data, on real statistical data and applied in the development of the software package ROSTATFOR (RObust STATistical FORecasting) in the Belarusian State University.

The results are applied in statistical forecasting of macroeconomic time series, spatio-temporal processes and in information protection.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Kolmogorov A. Interpolation and extrapolation of stationary stochastic series // Izv. Akad. Nauk SSSR. Ser. Mat. 1941. Vol. 5. № 1. P. 3–14.
2. Huber P. Some mathematical problems arising in robust statistics // Proceedings of the International Congress of Mathematicians. Vancouver, 1974. P. 821–824.
3. Kharin Yu. Robustness in statistical forecasting. N. Y. ; Dordrecht ; London : Springer, 2013.
4. Kharin Yu. Robustness in statistical pattern recognition. Boston ; London : Kluwer Academic Publishers, 1996.
5. Kharin Yu. Optimality and robustness in statistical forecasting // International Encyclopedia of Statistical Sciences. N.Y. : Springer, 2011. P. 1034–1037.
6. Kharin Yu. Robustness in statistical forecasting (chapter 14) // Robustness and Complex Data Structures / ed.: C. Becker, R. Fried, S. Kuhnt. N.Y. : Springer, 2013. P. 225–242.
7. Kharin Yu. Robustness analysis in forecasting of time series // Developments in Robust Statistics. N.Y., 2000. P. 180–193.
8. Kharin Yu. Robust forecasting of parametric trends // Studies in Classification and Data Analysis. 2000. Vol. 17. P. 197–206.
9. Kharin Yu., Maevskiy V. Robust regressive forecasting under functional distortions // Automation and Remote Control. 2002. Vol. 11. P. 125–136.
10. Kharin Yu. Robustness of mean square risk in forecasting of regression time series / Yu. Kharin // Comm. in Statist. – Theory and Methods. 2011. Vol. 40. № 16. P. 2893–2906.