Institute of Economic Studies, Faculty of Social Sciences
Charles University in Prague

Correcting Predictive Models of Chaotic Reality

Petr Kadeřábek

IES Working Paper: 31/2006

Institute of Economic Studies,
Faculty of Social Sciences,
Charles University in Prague

[UK FSV – IES]

Opletalova 26
CZ-110 00, Prague
E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Institut ekonomických studií
Fakulta sociálních věd
Univerzita Karlova v Praze

Opletalova 26
110 00 Praha 1

E-mail : ies@fsv.cuni.cz
http://ies.fsv.cuni.cz

Bibliographic information:
Kadeřábek, P. (2006). " Correcting Predictive Models of Chaotic Reality " IES Working Paper 31/2006, IES FSV Charles University

This paper can be downloaded at: http://ies.fsv.cuni.cz

Correcting Predictive Models of Chaotic Reality

Petr Kadeřábek[#]

[#] IES FSV UK, Opletalova 26, Prague
E-mail: petrkaderabek@seznam.cz

December 2006

Abstract:

We will assume a chaotic (mixing) reality, can observe a substantially aggregated state vector only and want to predict one or more of its elements using a stochastic model. However, chaotic dynamics can be predicted in a short term only, while in the long term an ergodic distribution is the best predictor. Our stochastic model will thus be considered a local approximation with no predictive ability for the far future. Using an estimate of an ergodic distribution of the predicted scalar (or eventually vector), we get, under additional reasonable assumptions, the uniquely specified resulting model, containing information from both the local model and the ergodic distribution. For a small prediction horizon, if the local model converges in probability to a constant and additional technical assumption is fulfilled, the resulting model converges in L^1 norm to the local model. In long term, the resulting model converges in L^1 to the ergodic distribution. We propose also a formula for computing the resulting model from the nonparametric specification of the ergodic distribution (using past observations directly). Two examples follow.

Keywords: Chaotic system, Prediction, Bayesian Analysis, Local Approximation, Ergodic Distribution

JEL: C11, C53, C62

Acknowledgements:

This paper is part of the research, supported by the GACR under 402/03/H057.

1 Introduction

Throughout the article we will assume a chaotic (mixing) reality, see section 2.1. We have a partial information about a state vector only (can observe a function of the state vector) and want to predict one or more of its elements using a stochastic model. Dynamics of the chaotic reality can be predicted in the short term only, while in the long term an ergodic distribution is the best predictor.

We will thus consider our stochastic model to be a local approximation, in the neighborhood of the present state, with no predictive ability for the far future. Such a model can then be focused on the short term predictive power, while allowing for almost arbitrary long term dynamics (e.g. can be nonstationary). Next, we assume to have an estimate of an ergodic distribution of the predicted scalar (or eventually vector). Assumptions about the local model and ergodic distribution will be stated in section 2.2. The resulting model is defined in section 2.3.

We use a Bayesian approach, but do not have to assume explicitly any prior distribution, all the information necessary is derived from the local model and ergodic distribution under reasonable assumptions stated in section 2.4.

The resulting model is derived in section 3.1. It is uniquely specified and contains information from both the local model and the ergodic distribution. In section 3.2 we show that the resulting model converges to the ergodic distribution for the far future predictions. For a small prediction horizon, if the local model converges in probability to a constant[1], the resulting model converges in L^1 norm[2] to the local model.

We propose also a formula for computing the resulting model from a nonparametric specification of the ergodic distribution (using past observations directly) in section 3.3. Next, possibility of predicting a joint distribution of a vector, instead of a scalar variable distribution only, is discussed in section 3.4.

Two examples follow: One uses a possibly nonstationary AR(1) model as a local model (section 4.1) and the normal ergodic distribution. The other predicts the Lorentz attractor using a corrected AR(2) model and nonparametric specification of the ergodic distribution (section 4.2).

2 Assumptions and Definitions

2.1 Reality

First, we state an assumption about reality.

Assumption 1 (Reality) *Consider a deterministic reality described by a dynamical system* $\dot{\mathbf{z}}_t = \phi(\mathbf{z}_t)$, $\mathbf{z}_t \in \mathbb{R}^n$ *with* ϕ *being unknown, where the system is*

(i) mixing[3],

(ii) on its attractor.

We are not able to observe continuously the whole state vector $\mathbf{z}_t$ *but only in unit time steps the vector* $\mathbf{y}_t = \psi(\mathbf{z}_t)$, $\mathbf{y}_t \in \mathbb{S}' \subset \mathbb{R}^m$, ψ *is an unknown continuous function. We will predict the scalar* $x_t \equiv \mathbf{y}_{t,1} = \psi_1(\mathbf{z}_t) \in \mathbb{S} \subset \mathbb{R}$ *i.e.*

[1] The random variables $\boldsymbol{X}_i$ converge in probability to the constant x iff $\mathrm{P}\left[|\boldsymbol{X}_i - x| > \varepsilon\right] \to 0$ for all $\varepsilon > 0$. We will denote it $\operatorname{plim} \boldsymbol{X}_i = x$.

[2] The L^1 norm of a measurable function f on $\mathbb{S}$ is defined $||f||_1 := \int_{\mathbb{S}} |f(\xi)| \mathrm{d}\xi$. The sequence of measurable functions f_i is said to converge in L^1 to f iff $||f_i - f||_1 \to 0$.

[3] A dynamical system f is called *mixing* if for every pair of sets A and B we have $\lim_{n\to\infty} \mu(f^{-n}(A) \cap B) = \mu(A)\mu(B)$ (see Mañé (1987), p. 142). Note that mixing is a kind of a chaotic behavior and every mixing system is ergodic.

the first coordinate of the observed vector. We have not enough information to predict x_t deterministically. The present time value is $t = 0$.

This assumption about observing the function of the large continuous system's state vector in discrete time intervals is standard (see e. g. the reconstruction theorem in Diks (1999), p. 15). Due to the lack of information we are not able to precisely determine a state and dynamics of the system and so will use a stochastic model. Mixing reperesents a chaotic behavior. Being on an attractor is a standard assumption in systems, that exhibit complex (chaotic) dynamics (see (Eubank and Farmer , 1997)). In most cases we will assume $\mathbb{S}' = \mathbb{R}^m$ and $\mathbb{S} = \mathbb{R}$, however sometimes x_t may be restricted to the interval $\mathbb{S} = [0, 1]$, $\mathbb{S} = [0, \infty)$ or other.

We will suppose assumption 1 to hold throughout the article.

2.2 Available Information

The only information available for the purpose of building the resulting corrected model of x are the stochastic dynamical model and the ergodic distribution estimate. We will state assumptions about them in this section.

It may be possible to estimate the fluctuations of z_t in the near future. Due to continuity of ψ a trajectory of the observation of interest x_t will be continuous and thus, for a sufficiently small $t > 0$, x_t will be arbitrarily close to its known present value x_0. So it is possible to construct a model that can predict x_t in the near future:

Assumption 2 (Local Model) *Assume we have a local approximation of reality around the state $\mathbf{z}_t$ described by a stochastic model, containing information I_l and predicting x_t in each time $t \in (0, \infty)$ by a continuous cumulative distribution function (cdf) $L_t(x) \equiv F(x|I_l, t)$. We will denote the corresponding probability density function (pdf) $l_t(x) \equiv f(x|I_l, t)$, the random variable $\boldsymbol{L}_t$ and assume*

(i) there exist function $l_\infty(x)$ and constant K_t, dependent on time, such that

$$K_t l_t(x) \to l_\infty(x) \text{ for } t \to \infty and a.e. x \in \mathbb{S}, \tag{1}$$

(ii) fraction of pdf l_t and limit l_∞ is bounded almost everywhere (belongs to

the L^∞ space[4] *):*

$$||l_t(x)/l_\infty(x)||_\infty < \infty \textit{ for all } t > 0, \tag{2}$$

(iii) $l_t(x) > 0$ *for a.e.* $x \in \mathbb{S}$ *and all* $t > 0$.

We will call this model a local model.

Condition (i) ensures that l_t converges to some, possibly improper, distribution for $t \to \infty$. Restriction (ii) is merely technical. It is sufficient for product of $l_t(x)/l_\infty(x)$ and any integrable function on $\mathbb{S}$ to be integrable. For normally distributed pdf's l_t and l_∞ it states that the local model mut not have a lower variance than the limiting distribution – see example in the next paragraph. Reasonability of (iii) follows from the assumption that the local model is a local approximation only. Hence, it cannot reject a possibility of any state x being outside the neighborhood of x_0 (but inside the state space $\mathbb{S}$). Assumption of positive probability of x being in the neighborhood of x_0 is natural.

Consider, for example, that our local model is normally distributed for all t with mean μ_t and variance σ_t^2. Next, consider $\mu_t \to \mu_\infty$ and $\sigma_t \to \sigma_\infty$, where $|\mu_\infty| < \infty$ and $|\sigma_\infty| < \infty$. We will verify assumption 2 for this case:

ad (i) We can choose $K_t = 1$ for all $t > 0$ and l_∞ will thus be a pdf of a normal distribution with mean μ_∞ and variance σ_∞^2.

ad (ii) It can be shown that (2) holds iff for each $t > 0$ one of the following conditions holds:

(a) $\sigma_t < \sigma_\infty$,

(b) $\sigma_t = \sigma_\infty$ and $\mu_t = \mu_\infty$.

ad (iii) Pdf of a normal distribution is positive for all $x \in \mathbb{R}$, regardless of the parameters.

The local model can be seen as a local approximation of the system and we will adopt this point of view throughout the article. As the system is assumed to be mixing, it is predictable in a short time period only, and

[4] The L^∞ norm of a measurable function f on $\mathbb{S}$ is defined $||f||_\infty := \inf\{M \in [0,\infty] : |f(x)| \le M \text{ for a.e. } x\}$. Function f is said to belong to the L^∞ space iff $||f||_\infty < \infty$.

so knowledge of the complete (global) dynamics of the system may not be important. Hence, sufficiently good local approximation will be satisfactory for prediction, together with a sufficiently good approximation of the ergodic measure (see later the ergodic distribution).

There exists a vast number of models that we can use as a local model. We can start with classical linear models and end with various nonlinear ones, such as nonlinear theory-based models, radial basis functions (Kantz and Schreiber (1997) p. 212), neural networks (Kantz and Schreiber (1997) p. 213, Oliveira et al. (2000)), local linear models (Fan and Yao (2003) p. 20, Kantz and Schreiber (1997) p. 209), wavelet networks (Cao et al. (1995)) and many others. Judd and Mees (1995) discuss a method of building nonlinear models of possibly chaotic reality from data, while maintaining good robustness against noise.

Since we assume a mixing reality (by assumption 1), even if we were able to measure $\mathbf{z}_0$ (and $\mathbf{z}_{-1}, \mathbf{z}_{-2}, \ldots, \mathbf{z}_{-T+1}$) with an arbitrary accuracy $\varepsilon > 0$ and knew an exact form of ϕ, the measurement would yield no information about the location of future $\mathbf{z}_t$ for large t. Hence, the local model can not tell us anything about fluctuations of x_t for large t.

As reality is *on* the attractor, we can define the probability measure

$$\mu(A) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} I_A(x_t) \tag{3}$$

for all Borel sets A, where the indicator function $I_A(x) = 1$ for $x \in A$ and zero elsewhere.

As we are on the attractor and by invertibility of the system, we have

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} I_A(x_t) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=-T+1}^{0} I_A(x_t) \tag{4}$$

and we can define μ using the historical data,

$$\mu(A) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=-T+1}^{0} I_A(x_t). \tag{5}$$

Hence, probability measure $\mu(A)$ expresses the portion of time the system spent and will spend in a set A. For t large, due to the impossibility to predict the fluctuations, this measure is the best way of predicting possible values of x_t.

Assumption 3 (Ergodic Distribution) *Assume we have an* estimate *of the ergodic probability measure* μ *of* x_t*, containing information* I_m*, described by a continuous cdf* $M(x) \equiv F(x|I_m)$ *with the corresponding pdf* $m(x) \equiv f(x|I_m)$ *and random variable* $\boldsymbol{M}$*. This distribution will be called an* ergodic distribution*.*

In the contrary to the local model, the ergodic distribution describes a limiting distribution but no dynamics.

2.3 Resulting model

Our goal is to make a "synthesis" of the local model and ergodic distribution and obtain a resulting model, containing information from both models:

Definition 1 (Resulting Model) *The* resulting model *is defined by a continuous cdf* $R_t(x) \equiv F(x|I_m, I_l, t)$ *with the corresponding pdf* $r_t(x) \equiv f(x|I_m, I_l, t)$ *and random variable* $\boldsymbol{R}_t$*.*

We want to build the resulting distribution from knowledge of I_l and I_m only, assuming we have no other information. Model specifications I_l and I_m depend on present available information, thus the resulting distribution depends on present available information only indirectly through I_l and I_m – if we have some useful additional information, we should enclose it in I_l and I_m, i.e. modify the local model or ergodic distribution.

In contrast to many local models described in the literature, only a little attention has been given to a transition between the local model and ergodic distribution. Judd and Small (2000) take also a generally nonspecified local model and use the errors of its in-sample predictions to correct the out-of-sample predictions. However they do not use the ergodic distribution. The other way that can be used to treat this problem (but was not developed for this particular purpose) is the switching regime models (treshold autoregression). The problem concerning this approach is necessity of choosing the condition for switching and the resulting model is first *equal* to the local model and then switches immediately to the ergodic distribution. In case we use the regime switching determined by unobservable variables, we have a problem with the choice of the underlying stochastic process (see Franses and Dijk (2002)). However the most common approach is simply to use the (local) model for predicting the whole future. If it does not behave satisfactorily, e. g. explodes, we change it somehow or simply say it is valid only in a limited period of time only.

2.4 Model Construction

When building the resulting model, we will employ the Bayesian approach. All the information needed (such as prior distribution(s)) we derive from the local model and ergodic distribution under assumptions that we will state now.

In case we had no information about a state of the system, the best prediction of its future state (with arbitrary t) would be the ergodic distribution. Hence, it is suitable to use an estimate of the ergidic distribution as a prior for introducing some dynamics:

Assumption 4 (Prior Distribution) *The prior distribution of x_t, where $t > 0$, reflects our estimate of the ergodic distribution:*

(i) $f(x|I_m, t) = f(x|I_m)$ *under knowledge of* I_m*,*

(ii) $f(x|t) = f(x)$ *is considered to be a noninformative prior.*

In econometric practice it is common to omit distant past observations from the sample. The reason is parameter instability across a longer time horizon. To tell it in the language of our assumptions, the infinitely distant observation posseses no additional information for an estimate of the local model because the local model is a local approximation only:

Assumption 5 (Far Future Observation) *Knowledge of an infinitely distant future observation does not affect an estimate of the local model:*

$$f(I_l|x, t = \infty) = f(I_l). \tag{6}$$

As we have already mentioned, the local model can be seen as a local approximation of the system, while the ergodic distribution describes the best prediction in the long run. It is natural to consider that information about the ergodic distribution does not affect our estimate of the local model:

Assumption 6 (Model Independence)

$$f(I_l|I_m, x, t) = f(I_l|x, t). \tag{7}$$

3 Results

3.1 Assumptions Inference

According to the Bayes theorem and assumption 4,

$$r_t(x) \equiv f(x|I_l, I_m, t) = \frac{f(I_l|I_m, x, t)f(x|I_m)}{\int_{\mathbb{S}} f(I_l|I_m, \xi, t)f(\xi|I_m)\mathrm{d}\xi}, \quad (8)$$

$$l_t(x) \equiv f(x|I_l, t) = \frac{f(I_l|x, t)f(x)}{\int_{\mathbb{S}} f(I_l|\xi, t)f(\xi)\mathrm{d}\xi}. \quad (9)$$

Using assumption 5, equation (9) turns out to be for $t \to \infty$

$$l_\infty(x) \equiv f(x|I_l, t = \infty) = \frac{f(I_l)f(x)}{\int_{\mathbb{S}} f(I_l)f(\xi)\mathrm{d}\xi} = f(x). \quad (10)$$

Hence, if we behave consistently with assumptions 4 (ii) and 5, our *noninformative prior is equal to the limiting distribution of the local model.* This is fully compliant with our assumptions because the local model is a local approximation only and its predictive ability decreases with time, when the state of the system moves away from $\mathbf{z}_0$.

No distribution can be generally considered nininformative, it depends on the particular application. For a list of noninformative priors see Yang and Berger (1998). On the other hand, we cannot say that for any particular distribution there does not exist application, in which someone could reasonably consider it noninformative.

When choosing variability of the local model, we must take into account model uncertainty. If, e.g., the limiting distribution has a relatively low variability, compared to the ergodic distribution, and is moreover located far away from the ergodic distribution, the resulting model will reflect these our opinions. If we are not fully consistent with our choice, the resulting model may thus look surprising for us.

Employing assumption 6, we get

$$\frac{r_t(x)}{l_t(x)} = \frac{f(x|I_m)}{f(x)} \cdot \frac{\int_{\mathbb{S}} f(I_l|\xi, t)f(\xi)\mathrm{d}\xi}{\int_{\mathbb{S}} f(I_l|\xi, t)f(\xi|I_m)\mathrm{d}\xi}, \quad (11)$$

hence by substituting for $f(x|I_m)$ and $f(x)$ and multiplying by $l_t(x)$, the resulting density becomes

$$r_t(x) = \frac{l_t(x)m(x)/l_\infty(x)}{\int_{\mathbb{S}} l_t(\xi)m(\xi)/l_\infty(\xi)\mathrm{d}\xi}. \quad (12)$$

Due to assumption 2 (ii), an intergral in the denominator always exists and is finite. Assumption 2 (iii) is sufficient for the integral to be positive.

Note that conditions for existence of a positive and finite integral in the denominator can be specified in a different way. However, they should be specified separately for the local model and ergodic distribution, so that they would not force violating assumption 6. For example, we should not write directly the condition $\int_{\mathbb{S}} l_t(\xi)m(x)/l_\infty(\xi)\mathrm{d}\xi > 0$ to ensure that the integral will be positive.

3.2 Result Properties

The resulting model has two important properties. First, it converges almost everywhere to the ergodic distribution for $t \to \infty$, i.e.

$$\begin{aligned}
\lim_{t\to\infty} r_t(x) &= \lim_{t\to\infty} \frac{m(x)l_t(x)/l_\infty(x)}{\int_{\mathbb{S}} m(\xi)l_t(\xi)/l_\infty(\xi)\mathrm{d}\xi} = \\
&= \lim_{t\to\infty} \frac{m(x)\frac{1}{K_t}l_\infty(x)/l_\infty(x)}{\int_{\mathbb{S}} m(\xi)\frac{1}{K_t}l_\infty(\xi)/l_\infty(\xi)\mathrm{d}\xi} = \\
&= m(x) \text{ for a.e. } x \in \mathbb{S}. \qquad (13)
\end{aligned}$$

Note that almost everywhere convergence together with $||m||_1 = 1$ implies L^1 convergence.

Next, if $\operatorname{plim}_{t\to 0^+} \boldsymbol{L}_t = x_0$ and $m(x)/l_\infty(x)$ is continuous and positive in x_0, then $\lim_{t\to 0^+} ||r_t - l_t||_1 = 0$ (r_t converges in L^1 to l_t), because

$$\lim_{t\to 0^+} \int_{\mathbb{S}} |r_t(x) - l_t(x)|\,\mathrm{d}x = \lim_{t\to 0^+} \int_{x_0-\varepsilon}^{x_0+\varepsilon} \left| \frac{m(x_0)l_t(x)/l_\infty(x_0)}{\int_{x_0-\varepsilon}^{x_0+\varepsilon} m(x_0)l_t(\xi)/l_\infty(x_0)\mathrm{d}\xi} - l_t(x) \right| \mathrm{d}x = 0 \qquad (14)$$

for any $\varepsilon > 0$.

3.3 Nonparametric Ergodic Distribution

In some cases, the ergodic distribution can be nonstandard and thus difficult to express parametrically. In this section we describe a possibility to use historical values of x_t directly, without a need of specifying the ergodic distribution in a parametric form.

On the contrary to the ergodic distribution m, local model l_t is usually described in a parametric form (in most cases normally distributed).

By the Birkhoff ergodic theorem, the time average is equal to the space average, the space average is integrated according to the ergodic measure. In our model, the ergodic measure is represented by the ergodic distribution, particularly by pdf m. Hence, the theorem implies

$$\int_{\mathbb{S}} h(x)m(x)\mathrm{d}x = \lim_{T\to\infty} \frac{1}{T} \sum_{i=-T+1}^{0} h(x_i) \text{ a.s.}, \tag{15}$$

where h is a function integrable with respect to the ergodic measure.

Choosing $h(x) = l_t(x)/l_\infty(x)$,

$$\int_{\mathbb{S}} l_t(x)/l_\infty(x)m(x)\mathrm{d}x = \lim_{T\to\infty} \frac{1}{T} \sum_{i=-T+1}^{0} l_t(x_i)/l_\infty(x_i) \text{ a.s.}, \tag{16}$$

and analogically, when we choose $h(x) = g(x)l_t(x)/l_\infty(x)$, where g is a function integrable with respect to measure specified by r_t,

$$\int_{\mathbb{S}} g(x)l_t(x)/l_\infty(x)m(x)\mathrm{d}x = \lim_{T\to\infty} \frac{1}{T} \sum_{i=-T+1}^{0} g(x_i)l_t(x_i)/l_\infty(x_i) \text{ a.s.} \tag{17}$$

Substituting (16) and (17) into (12), we have

$$\int_{\mathbb{S}} g(x)r_t(x)\mathrm{d}x = \frac{\lim_{T\to\infty} \frac{1}{T} \sum_{i=-T+1}^{0} g(x_i)l_t(x_i)/l_\infty(x_i)}{\lim_{T\to\infty} \frac{1}{T} \sum_{i=-T+1}^{0} l_t(x_i)/l_\infty(x_i)} \text{ a.s.} \tag{18}$$

Hence, if we have a sufficiently long time series, we can use (18) for computing $\int_{\mathbb{S}} g(x)r_t(x)\mathrm{d}x$ without a need of specifying the ergodic distribution parametrically.

In case of knowing x_t in addition to the historical time series, we would probably leave our belief about I_m unchanged or would include value of x_t to the set of past observations. In both cases I_m would be a subset of information contained in the historical time series and x_t and so assumtion 6 would hold.

Note that according to the assumptions, for expressing the ergodic distribution nonparametrically we should use the whole historical time series available. On the other hand, for the local model estimation it is possible to drop old observations.

3.4 Predicting the Joint Distribution

Sometimes, unconditional prediction of only one variable is not satisfactory and we need to predict joint distribution of more variables instead. For example, in place of monthly price change we need the total price change for 12 successive months, or we may need the joint distribution of inflation and production growth.

One possibility, applicable to the first case, is changing the local model and ergodic distribution, so that they would predict directly the price change in successive 12 months instead of monthly price change.

However, in some cases we may not be able to modify the models, so that we got to predict only one variable (this may be the second example). In this case we can employ our approach in a slightly modified form: The resulting formula (12), respectively (18), remains the same, if we assume that x_t is a (short) vector in the very beginning. We will probably be able to construct marginal distributions of the (joint) ergodic distribution easily. The problem will be in specifying dependencies (using the known marginals, dependencies are specified by a copula, see Nelsen (1995)). The probably preferred way in this case will be the nonparametric approach, where dependencies need not be specified explicitly.

4 Examples

4.1 AR(1) Process

Consider the ergodic distribution with $\boldsymbol{M} \sim N(\mu_m, \sigma_m^2)$ and thus

$$m(x) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left[-\frac{1}{2} \cdot \frac{(x-\mu_m)}{\sigma_m^2}\right]. \tag{19}$$

Let the local model be AR(1), $x_{t+1} = ax_t + \varepsilon_{t+1}$ with $a \in (0, \infty)$, $t = 0, 1, 2, \ldots$ and $\varepsilon_{t+1} \sim N(0, \sigma_\varepsilon^2)$, i.i.d. Note that we did not allow for a negative a because we will be able to extend t to $(0, \infty)$ now. The unconditional prediction of the local model is

$$\boldsymbol{L}_t \sim N\left(a^t x_0, \sigma_\varepsilon^2 \frac{1-a^{2t}}{1-a^2}\right), \tag{20}$$

hence

$$l_t(x) = \frac{\sqrt{1-a^2}}{\sqrt{2\pi\sigma_\varepsilon^2(1-a^{2t})}} \exp\left[-\frac{1}{2} \cdot \frac{(x-a^t x_0)^2(1-a^2)}{\sigma_\varepsilon^2(1-a^{2t})}\right]. \tag{21}$$

We choose

$$K_t = \sqrt{2\pi\sigma_\varepsilon^2 \frac{1-a^{2t}}{1-a^2}}, \tag{22}$$

but it could be specified in many different ways. We have

$$l_\infty(x) = \lim_{t\to\infty} K_t l_t(x) = \begin{cases} \exp\left[-\frac{1}{2}\cdot\frac{(1-a^2)x^2}{\sigma_\varepsilon^2}\right] & \text{for } a \in (0,1), \\ \exp\left[-\frac{1}{2}\cdot\frac{(a^2-1)x_0^2}{\sigma_\varepsilon^2}\right] & \text{for } a \in [1,\infty) \end{cases} \tag{23}$$

and so assumption 2 (i) holds for all $a \in (0,\infty)$. We can express $l_t(x)/l_\infty(x)$ and verify that assumption 2 (ii) holds. Assumption 2 (iii) is fulfilled for any normally distributed local model.

Let us substitute (19), (21) and (23) into (12). After some algebra we realize that $\boldsymbol{R}_t \sim N(\mu_r(t,a), \sigma_r^2(t,a))$, where

$$\mu_r(t,a) = \begin{cases} \frac{(1-a^{2t})\mu_m\sigma_\varepsilon^2 + a^t(1-a^2)x_0\sigma_m^2}{(1-a^{2t})\sigma_\varepsilon^2 + a^{2t}(1-a^2)\sigma_m^2} & \text{for } a \in (0,1), \\ \frac{t\mu_m\sigma_\varepsilon^2 + x_0\sigma_m^2}{t\sigma_\varepsilon^2+\sigma_m^2} & \text{for } a = 1, \\ \frac{(a^{2t}-1)\mu_m\sigma_\varepsilon^2 + a^t(a^2-1)x_0\sigma_m^2}{(a^{2t}-1)\sigma_\varepsilon^2 + (a^2-1)\sigma_m^2} & \text{for } a \in (1,\infty), \end{cases} \tag{24}$$

$$\sigma_r^2(t,a) = \begin{cases} \frac{(1-a^{2t})\sigma_\varepsilon^2\sigma_m^2}{(1-a^{2t})\sigma_\varepsilon^2 + a^{2t}(1-a^2)\sigma_m^2} & \text{for } a \in (0,1), \\ \frac{t\sigma_\varepsilon^2\sigma_m^2}{t\sigma_\varepsilon^2+\sigma_m^2} & \text{for } a = 1, \\ \frac{(a^{2t}-1)\sigma_\varepsilon^2\sigma_m^2}{(a^{2t}-1)\sigma_\varepsilon^2 + (a^2-1)\sigma_m^2} & \text{for } a \in (1,\infty). \end{cases} \tag{25}$$

It is worth noting that

$$\lim_{a\to 1^-} \mu_r(t,a) = \mu_r(t,1) = \lim_{a\to 1^+} \mu_r(t,a), \tag{26}$$

$$\lim_{a\to 1^-} \sigma_r^2(t,a) = \sigma_r^2(t,1) = \lim_{a\to 1^+} \sigma_r^2(t,a), \tag{27}$$

hence $\mu_r(t,a)$ and $\sigma_r^2(t,a)$ are continuous in $a=1$ for any $t \in (0,\infty)$, no matter that the local model is stationary for $a \in (0,1)$ and nonstationary for $a \in [1,\infty)$. In the short term, stationarity is not an important question, however in case of using the local model even for the long term, we must keep it stationary and force $a \in (0,1)$ and thus change its short term behaviour. Our approach leaves the short term properties unrestricted while guaranteeing stationarity.

We can also verify that

$$\lim_{t\to 0^+} \frac{\mu_r(t,a)}{\mu_l(t,a)} = \lim_{t\to\infty} \frac{\mu_r(t,a)}{\mu_m} = \lim_{t\to 0^+} \frac{\sigma_r^2(t,a)}{\sigma_l^2(t,a)} = \lim_{t\to\infty} \frac{\sigma_r^2(t,a)}{\sigma_m^2} = 1. \tag{28}$$

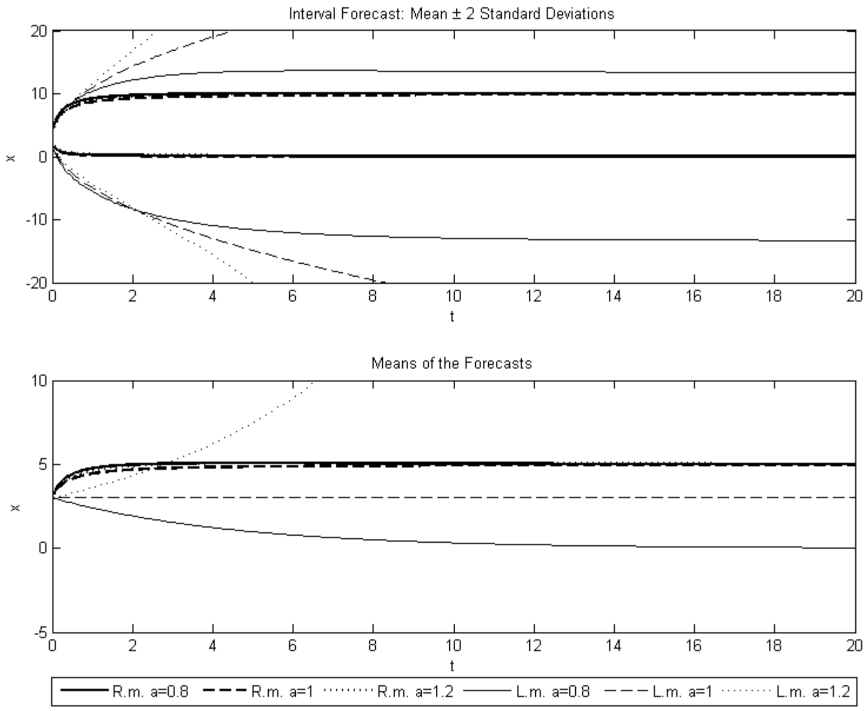

Figure 1: Means and two sigma confidence intervals of the AR(1) local model and the resulting model from example 4.1.

In figure 1 we can see the local models for $a \in \{0.8, 1, 1.2\}$. The corresponding resulting models are depicted too. The remaining parameters are same for both alternatives: $x_0 = 3$, $\sigma_\varepsilon = 4$, $\mu_m = 5$, $\sigma_m = 2.5$. While the local model is stationary for $a = 0.8$ and nonstationary for $a \in \{1, 1.2\}$, the resulting models are very similar. Thus there is no need to test stationarity of the local model.

As we have discussed in section 3.1, when choosing variance of the local model, we must take into account model uncertainty. To be consistent with the assumptions, the limiting distribution of the local model is considered to be noninformative, hence it should be, in most applications, flat or at least have a relatively high variance (compared to the ergodic distribution) and mean not far away from the mean of the ergodic distribution. We can experiment with values, not satisfying these suggestions, and see the consequences of considering such distribution to be noninformative.

4.2 Predicting the Lorenz System

We will assume that reality works according to the Lorenz system

$$
\begin{aligned}
\dot{z}_{1,t} &= 0.1(-10z_{1,t} + 10z_{2,t}) & (29)\\
\dot{z}_{2,t} &= 0.1(28z_{1,t} - z_{2,t} - z_{1,t}z_{3,t}) & (30)\\
\dot{z}_{3,t} &= 0.1(-\frac{8}{3}z_{3,t} + z_{1,t}z_{2,t}). & (31)
\end{aligned}
$$

These equations denote the Lorenz system with classical parameters, however the right hand side of each equation is multiplied by 0.1. This does not modify a shape of the attractor but rather changes the time scale.

To use symbols from assumption 1, the set of Lorenz equations describes the function ϕ, the state vector $\mathbf{z}_t \equiv [z_{1,t}, z_{2,t}, z_{3,t}]^T$. We can observe the scalar $\psi(\mathbf{z}_t) = y_t = x_t$ and this is also the variable we want to predict.

All the constants in this example will be rounded to 4 decimal places. The observable history of our Lorenz system starts in the state $\mathbf{z}_0 = [4.3606, 5.4437, 19.8638]^T$ that is very close to the attractor. The time series is 100 observations long.

We have built the local model as AR(2) model

$$
\begin{aligned}
x_t &= 1.4x_{t-1} - 1.2x_{t-2} + 5.5 + \varepsilon_t, \quad \varepsilon_t \sim N(0,1), i.i.d., & (32)\\
x_{-1} &= 4.3085, & (33)\\
x_0 &= 4.3606. & (34)
\end{aligned}
$$

This implies

$$\boldsymbol{L}_t \sim N(\mu_l(t), \sigma_l^2(t)). \quad (35)$$

where

$$
\begin{aligned}
\mu_l(t) &= \exp(0.0912t)\,[-2.5144\cos(0.8776t) + 1.5966\sin(0.8776t)] + \\
&\quad +6.8750, & (36)\\
\sigma_l^2(t) &= \exp(0.1823t)\,[-0.4059\cos(1.7552t) - .2885\sin(1.7552t) + \\
&\quad 4.2254] - 3.8194. & (37)
\end{aligned}
$$

When choosing

$$K_t = \sigma_l(t), \quad (38)$$

we get $l_\infty(x)$ being constant for all x and we can easily verify assumption 2.

We use the nonparametric approach to get the resulting model. It will be compared with the best possible prediction that is obtainable under a slightly different assumptions: knowledge of the Lorenz equations and ability to measure the whole present state vector $\mathbf{z}_0$ but with the normally distributed, independent measurement errors with zero mean and standard deviation of 0.05. The prediction is computed as the mean value of trajectories with the initial condition $\mathbf{z}_0$ plus the random measurement error (we used 1000 realisations). The results are depicted in figure 2.

5 Conclusion

In section 2 we have stated several assumptions. The rather technical ones were assumptions 1, 2 and 3, the key ones on the other hand 4, 5 and 6. If we are consistent with assumptions 4 (ii) and 5, we consider the limiting distribution of the local model for $t \to \infty$ to be a noninformative prior. Using assumption 4 (i) and 6 we derive the unique formula for the resulting model, that contains information from both the local model and the ergodic distribution. The noninformative prior distribution was replaced by an ergodic distribution.

We showed that the resulting model converges to the ergodic distribution for the far future predictions. For a small prediction horizon, if the local model converges in probability to a constant x_0 and both the ergodic pdf and limiting pdf of the local model are continuous and positive in x_0, the resulting model converges in L^1 to the local model.

As the ergodic distribution is nonstandard in many cases, we presented a possibility to express it nonparametrically. We can thus use the resulting model, computed directly from the local model and past observations. This is particularly useful when predicting joint distribution of a random vector, because specifying dependencies is often a difficult question.

In the end we presented two examples. In the first one we use a possibly nonstationary AR(1) model as a local model and a normal ergodic distribution. Consequences of considering the limiting distribution of the local model to be a noninformative prior are discussed. In the second example we predicted the Lorentz attractor using a corrected AR(2) model.

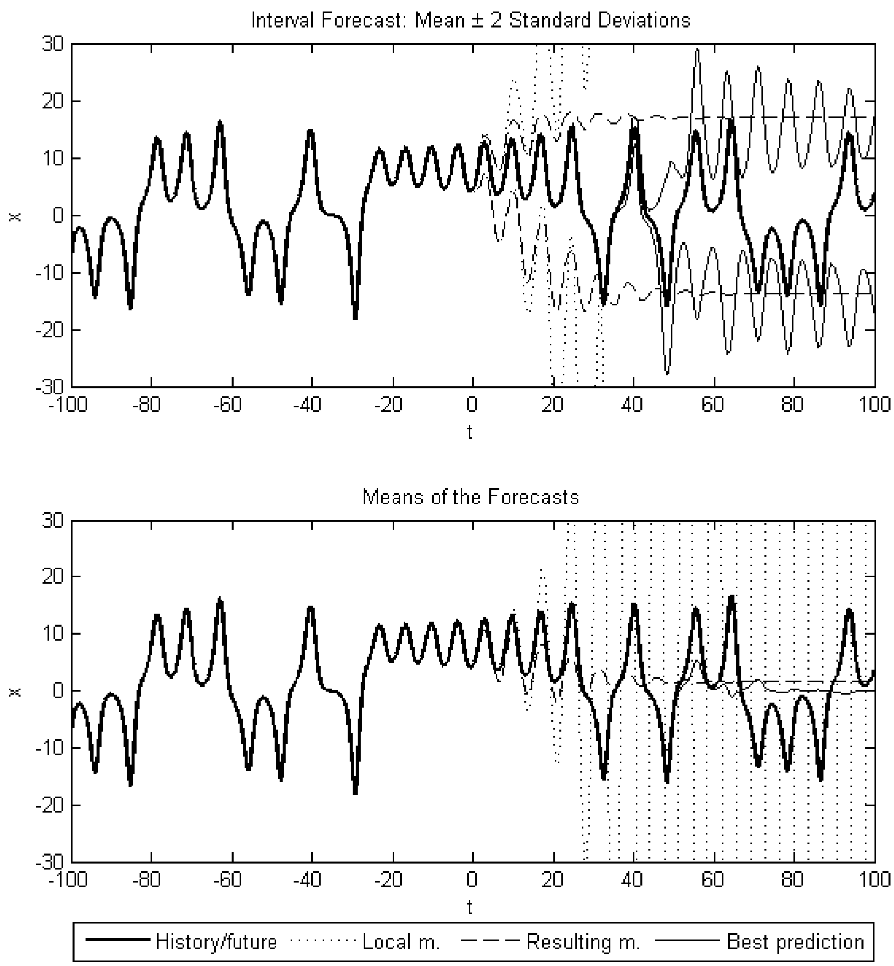

Figure 2: True historical/future values, means and two sigma confidence intervals of the local model, resulting model and best model from example 4.2.

References

Cao, L., Hong, Y., Fang, H., He, G., *Predicting Chaotic Time Series with Wavelet Networks*, Physica D 85 (1995), p. 225-238.

Diks, C., *Nonlinear Time Series Analysis, Methods and Application*, World Scientific Publishing, Singapore 1999. Springer-Verlag, New York 1994.

Fan, J., Yao, Q., *Nonlinear Time Series*, Springer-Verlag, New York 2003.

Franses, P., Dijk, D., *Non-linear Time Series Models in Empirical Finance*, Cambridge University Press, Cambridge 2002.

Guckenheimer, J., Holmes, P., *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer-Verlag, New York 1983.

Judd, K., Mees, A., *On Selecting Models for Nonlinear Time Series*, Physica D 82 (1995), p. 426-444.

Judd, K., Small, M., *Towards Long-Term Prediction*, Physica D 136 (2000), p. 31-44.

Kantz, H., Schreiber, T., *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge 1997.

Eubank, S., Farmer, J., *Introduction to Dynamical Systems*, in Lam, L. (Ed.), *Introduction to Nonlinear Physics*, p. 55-105, Springer, 1997.

Mañé, R., *Ergodic Theory and Differentiable Dynamics*, Springer-Verlag, Berlin Heidelberg 1987.

Nelsen, R., *Copulas, Characterization, Correlation, and Counterexamples*, Mathematics Magazine, Vol. 68, No. 3 (1995), p. 193-198.

Oliveira, K., Vannucci, Á., Silva, E., *Using Artificial Neural Networks to Forecast Chaotic Time Series*, Physica A 284 (2000), p. 393-404.

Yang, R., Berger, J., *A Catalog of Noninformative Priors*, http://www.stats.org.uk/priors/noninformative/YangBerger1998.pdf.

IES Working Paper Series

2005

15. Oldřich Dědek: *Rizika a výzvy měnové strategie k převzetí eura*
16. Karel Janda, Martin Čajka: *Srovnání vývoje českých a slovenských institucí v oblasti zemědělských finance*
17. Alexis Derviz: *Cross-border Risk Transmission by a Multinational Bank*
18. Karel Janda: *The Quantitative and Qualitative Analysis of the Budget Cost of the Czech Supporting and Guarantee Agricultural and Forestry Fund*
19. Tomáš Cahlík, Hana Pessrová: *Hodnocení pracovišť výzkumu a vývoje*
20. Martin Gregor: *Committed to Deficit: The Reverse Side of Fiscal Governance*
21. Tomáš Richter: *Slovenská rekodifikace insolvenčního práva: několik lekcí pro Českou republiku*
22. Jiří Hlaváček: *Nabídková funkce ve vysokoškolském vzdělávání*
23. Lukáš Vácha, Miloslav Vošvrda: *Heterogeneous Agents Model with the Worst Out Algorithm*
24. Kateřina Tsolov: *Potential of GDR/ADR in Central Europe*
25. Jan Kodera, Miroslav Vošvrda: *Production, Capital Stock and Price Dynamics in a Simple Model of Closed Economy*
26. Lubomír Mlčoch: *Ekonomie a štěstí – proč méně může být vice*
27. Tomáš Cahlík, Jana Marková: *Systém vysokých škol s procedurální racionalitou agentů*
28. Roman Horváth: *Financial Accelerator Effects in the Balance Sheets of Czech Firms*
29. Natálie Reichlová: *Can the Theory of Motivation Explain Migration Decisions?*
30. Adam Geršl: *Political Economy of Public Deficit: Perspectives for Constitutional Reform*
31. Tomáš Cahlík, Tomáš Honzák, Jana Honzáková, Marcel Jiřina, Natálie Reichlová: *Convergence of Consumption Structure*
32. Luděk Urban: *Koordinace hospodářské politiky zemí EU a její meze*

2006

1. Martin Gregor: *Globální, americké, panevropské a národní rankingy ekonomických pracovišť*
2. Ondřej Schneider: *Pension Reform in the Czech Republic: Not a Lost Case?*
3. Ondřej Knot and Ondřej Vychodil: *Czech Bankruptcy Procedures: Ex-Post Efficiency View*
4. Adam Geršl: *Development of formal and informal institutions in the Czech Republic and other new EU Member States before the EU entry: did the EU pressure have impact?*
5. Jan Zápal: *Relation between Cyclically Adjusted Budget Balance and Growth Accounting Method of Deriving 'Net fiscal Effort'*
6. Roman Horváth: *Mezinárodní migrace obyvatelstva v České republice: Role likviditních omezení*
7. Michal Skořepa: *Zpochybnění deskriptivnosti teorie očekávaného užitku*
8. Adam Geršl: *Political Pressure on Central Banks: The Case of the Czech National Bank*
9. Luděk Rychetník: *Čtyři mechanismy příjmové diferenciace*
10. Jan Kodera, Karel Sladký, Miloslav Vošvrda: *Neo-Keynesian and Neo-Classical Macroeconomic Models: Stability and Lyapunov Exponents*

11. Petr Jakubík: *Does Credit Risk Vary with Economic Cycles? The Case of Finland*
12. Julie Chytilová, Natálie Reichlová: *Systémy s mnoha rozhodujícími se jedinci v teoriích F. A. Hayeka a H. A. Simona*
13. Jan Zápal, Ondřej Schneider: *What Are Their Words Worth? Political Plans And Economic Pains Of Fiscal Consolidations In New Eu Member States*
14. Jiří Hlaváček, Michal Hlaváček*: Poptávková funkce na trhu s pojištěním: porovnání maximalizace paretovské pravděpodobnosti přežití s teorií EUT von-Neumanna a Morgensterna a s prospektovou teorií Kahnemana a Tverského*
15. Karel Janda, Martin Čajka: *Státní podpora českého zemědělského úvěru v období před vstupem do Evropské unie*
16. Nauro F. Campos, Roman Horváth: *Reform Redux: Measurement, Determinants and Reversals*
17. Michal Skořepa: *Three heuristics of search for a low price when initial information about the market is obsolete*
18. Michal Bauer, Julie Chytilová: *Opomíjená heterogenita lidí aneb Proč afrika dlouhodobě neroste*
19. Vít Bubák, Filip Žikeš: *The Price of Stock Trades: Evidence from the Prague Stock Exchange*
20. Vladimír Benáček, Jiří Podpiera a Ladislav Prokop: *Command Economy after the Shocks of Opening up: The Factors of Adjustment and Specialisation in the Czech Trade*
21. Lukáš Vácha, Miloslav Vošvrda: *Wavelet Applications to Heterogeneous Agents Model*
22. Lukáš Vácha, Miloslav Vošvrda: *"Morální hazard" a „nepříznivý výběr" při maximalizaci pravděpodobnosti ekonomického přežití*
23. Michal Bauer, Julie Chytilová, Pavel Streblov: *Effects of Education on Determinants of High Desired Fertility Evidence from Ugandan Villages*
24. Karel Janda: *Lender and Borrower as Principal and Agent*
25. Karel Janda: *Optimal Deterministic Debt Contracts*
26. Jiří Hlaváček: *Pojištění vkladů: současný stav, srovnání a perspektiva v kontextu EU*
27. Pavel Körner: *The determinants of corporate debt maturity structure: evidence from Czech firms*
28. Jarko Fidrmuc, Roman Horváth: *Credibility of Exchange Rate Policies in Selected EU New Members: Evidence from High Frequency Data*
29. Natálie Reichlová, Petr Švarc*: Strategic Referring in Labor Market Social Networks*
30. František Turnovec: *Publication Portfolio of the Czech Economists and Problems of Rankings*

All papers can be downloaded at: http://ies.fsv.cuni.cz

Univerzita Karlova v Praze, Fakulta sociálních věd
Institut ekonomických studií [UK FSV – IES] Praha 1, Opletalova 26
E-mail : ies@fsv.cuni.cz http://ies.fsv.cuni.cz