Study of a BVAR(p) process applied to U.S. commodity market data

Jan Šindelář

Abstract—The paper presents an applied study of a multivariate AR(p) process fitted to daily data from U.S. commodity futures markets with the use of Bayesian statistics. In the first part a detailed description of the methods used is given. In the second part two BVAR models are chosen one with assumption of log-normal, the second with normal distribution of prices conditioned on the parameters. For a comparison two simple benchmark models are chosen that are commonly used in todays Financial Mathematics. The article compares the quality of predictions of all the models, tries to find an adequate rate of forgetting of information and questions the validity of Efficient Market Hypothesis in the semi-strong form.

Keywords—Vector Auto-regression, Forecasting, Financial, Bayesian, Efficient Markets.

I. INTRODUCTION

In the sixties E.Fama together with P.Samuelson laid the ground for Efficient Market Hypothesis (EMH) in Fama's PhD. thesis and a series of famous articles [1], [2], [3], following earlier work of L.Bachelier [4]. Since then, validity of the hypothesis has been widely examined, but was mostly accepted by mathematicians and financial engineers, because of the economic arguments, indecisive experimental proofs against it and greater ease of computation [5] of related modeling problems, mainly on the side of optimization and decision making. The forecastability of financial market returns remains an open problem and is discussed for example in [6], [7]. The second article states that a model able to predict financial returns has to adapt to market changes quickly and catch local dependencies in price movements. As such a model, we propose a vector auto-regression used quite successfully for forecasting evolution of various time series of non-financial kind and even economic time series [8].

In this paper, we try to compare two kinds of BVAR(p) models with the most widely accepted price evolution models in modern finance (for a continuous-time case see for example [9]), which are in agreement with the semi-strong form of EMH. The abbreviation BVAR stands for Bayesian Vector Auto-Regression process, to be fully defined later. For the comparison we use daily data from U.S. commodity futures markets.

To be able to compare the competing models we need to be able to compute a multi-step ahead prediction of the related BVAR processes. Such predictions have been already proposed under the assumptions of classical statistics [10] or when mean

J. Šindelář is with the Department of Adaptive Systems, Institute of Information Theory and Automation CAS, Pod Vodárenskou věží 4, Prague 182 08, Czech Republic e-mail:jenik.sindelar@gmail.com

values of estimated parameters were taken as true values [11]. In this paper we compute an approximate solution of h-step prediction of evolution of the stochastic process related to the data in a full distribution form proposed in [12], although only the first moment of the distribution will be used in the experimental part. To obtain the distribution we draw Monte Carlo samples from the estimated parameter distribution, we let the vector auto-regression process evolve with the sampled parameters and reconstruct the predictive distribution by a simple numerical integration.

Since in two of the models we predict variables that are log-normally distributed, when the values of the parameters of the model are known, we can work with the related data channels logarithmized, but then we have to be careful when computing predictions.

In the paper we give a detailed description of the computation of predictions, when certain channels are modeled by a conditionally log-normal auto-regression process.

The paper is organized as follows: In section II the theoretical background to the problem presented is given. Subsection II-A presents basic terms for the reader to understand the issue. Subsection II-B presents the model chosen as one possibility for modeling time series evolution. Subsection II-C references the reader to a source, where systematic structure estimation of the model is presented. Subsection II-D guides the reader through the Bayesian estimation of parameters. Subsection II-E describes an approximation for time evolution of parameters through exponential forgetting. Subsection II-F describes the approximate computation of future price distribution through Monte Carlo sampling and gives a recipe for finding first two central moments of the distribution.

Finally subsection II-G presents inclusion of log-normal data channels into the model. Section III presents the results obtained from selected models, when applied to U.S. commodity market data, with description of the quality measures of point predictions of the individual models. Section IV concludes the paper.

II. MODELLING

A. Basic concepts and notation

We try to model the time series of a market price y_t of some commodity futures contract, where t is from a discrete finite set of times – an index set $\mathcal{T} = \{1, \ldots, T\}$. We suppose that the price evolution is also influenced by other observable data and we collect all these data channels into a vector of data $\mathbf{d}_t = (d_{1;t} = y_t, d_{2;t}, \ldots, d_{k;t})'$, where the apostrophe stands for transposition.

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:10, 2009

To such data, we try to assign a discrete-time stochastic process $\mathbf{D}_t$ – an *adapted stochastic process* meaning that now the $\mathbf{D}_t$ are *random vectors*, defined on a probability space $(\Omega, \mathcal{F}, \mu)$, where the space is equipped with a filtration $\mathcal{F}_t$[1] (collection of σ-algebras) and for each time $s \geq t$ the random vector $\mathbf{D}_t$ is measurable $\mathcal{F}_s$ and $\mathbf{D}_t = \mathbf{d}_t$ – the realization of the random vector $\mathbf{D}_t$ is known to the observer from time t on. Capital letters are used for random variables and matrices and small letters for realizations and values.

We suppose, the joint probability distribution of $\mathbf{D}_1, \ldots, \mathbf{D}_T$ is absolutely continuous with respect to the underlying Lebesgue measure λ^{Tk}, so that there exists a joint probability density $f(\mathbf{d}_1, \ldots, \mathbf{d}_T)$ specifying the distribution. For a more detailed discussion on this issue see Chapter 6 in [14].

Remark 1: In the following text densities of a different functional form can be denoted by f if they differ in arguments (either in number or type). This is a concept similar to that of function overloading often used in computer programming and leads to a less complicated notation.

Except for random vectors $\mathbf{D}_t$, there are other random variables $\boldsymbol{\theta}$ defined on $(\Omega, \mathcal{F}, \mu)$, called *parameters*, some of which can describe the relations between $\mathbf{D}_1, \ldots, \mathbf{D}_T$ by a parametric model – one of such will be introduced in the next section. These variables are not measurable $\mathcal{F}_t, \forall t \in \mathcal{T}$, but are measurable $\mathcal{F}$. We again suppose, there exists a joint density of data and parameters determining their joint distribution.

B. Model choice

There are various ways how to choose an appropriate *parametric* model to describe the behavior of stochastic process $\mathbf{D}_1, \ldots, \mathbf{D}_T$ (see for example [15],[16]). We will use a special type of ARMA processes - multivariate AR processes reduced to the first p time-lags (AR processes were first systematically studied by Box and Jenkins in [17] and are also presented in detail in [18]). The reason for choosing such processes is their relative richness and also computational ease, when it comes to their parameter estimation. We start by splitting the joint probability density mentioned earlier into factors similar to each other, but shifted in time with the use of basic theorems of probability theory [14], [19]

$$f(\mathbf{d}_1, \ldots, \mathbf{d}_T, \boldsymbol{\theta}) = \tag{1}$$
$$f(\mathbf{d}_T, \boldsymbol{\theta}_T | \boldsymbol{\theta}_{T-1}, \ldots, \boldsymbol{\theta}_p, \mathcal{F}_{T-1})$$
$$f(\mathbf{d}_{T-1}, \boldsymbol{\theta}_{T-1} | \boldsymbol{\theta}_{T-2}, \ldots, \boldsymbol{\theta}_p, \mathcal{F}_{T-2}) \cdots$$
$$f(\mathbf{d}_{p+1}, \boldsymbol{\theta}_{p+1} | \boldsymbol{\theta}_p, \mathcal{F}_p)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_T, \ldots, \boldsymbol{\theta}_p)$. We start modeling at time $p + 1$, when the first p data values are available.

In a general case, the number of parameters $\boldsymbol{\theta}$ can be very large and they can evolve over time similarly to the measured data. In such a case, we would also need to associate a stochastic process with the parameter evolution. In such a situation, we have to index the parameters by $t \in \mathcal{T}$ and model them similarly to the data. We would then have to describe the causal dependence of the parameters in terms

---

[1]See [13] page 51 for details

of probability density $f(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \ldots, \boldsymbol{\theta}_p, \mathcal{F}_{t-1}) \ \forall t \in \mathcal{T}$. Instead we choose a smaller set of parameters $\boldsymbol{\theta}$, which we believe evolve slowly over time, so that we can account for their evolution by applying an exponential forgetting in their estimation [15],[20]. In this approximation we have

$$f(\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \ldots, \boldsymbol{\theta}_p, \mathcal{F}_{t-1}) = f(\boldsymbol{\theta} | \mathcal{F}_{t-1}) \quad \forall t \in \mathcal{T} \tag{2}$$

Remark 2: Since the previously defined set of time-dependent parameters will not be needed anymore throughout the text, we used the same notation $\boldsymbol{\theta}$ for the restricted set of time-independent parameters.

We now come to the modeling of individual factors in (1). We split the factors again to get

$$f(\mathbf{d}_t, \boldsymbol{\theta} | \mathcal{F}_{t-1}) = f(\mathbf{d}_t | \boldsymbol{\theta}, \mathcal{F}_{t-1}) f(\boldsymbol{\theta} | \mathcal{F}_{t-1}) \tag{3}$$

where the first factor on the right-hand side is the actual parametric model we will choose now conditioned on the past and the parameters. The second factor in (3) is the posterior probability density we will estimate from past data.

We choose the probability density of data conditioned on the parameters to be multivariate Gaussian of the form

$$f(\mathbf{d}_t | \boldsymbol{\theta}, \mathcal{F}_{t-1}) = \tag{4}$$
$$\frac{1}{[2\pi |\mathbf{R}|]^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \mathrm{tr} \left[\mathbf{R}^{-1} \begin{bmatrix} \mathbf{I} \\ -\mathbf{A} \end{bmatrix}' \begin{bmatrix} \mathbf{d}_t \\ \phi_{t-1} \end{bmatrix} \begin{bmatrix} \mathbf{d}_t \\ \phi_{t-1} \end{bmatrix}' \begin{bmatrix} \mathbf{I} \\ -\mathbf{A} \end{bmatrix} \right] \right\}$$

where $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{R}\}$ are matrices of parameters, $\phi_t = [\mathbf{d}_{t-1} \ \mathbf{d}_{t-2} \cdots \mathbf{d}_{t-p} 1]'$. Parameter $\mathbf{R}$ stands for a covariance matrix of the model, $\mathbf{A}$ are parameters of the auto-regression relating past observations of data to present or future observations (more detailed description will follow). There are several reasons for choosing such a model:

- Bayesian estimation of parameters of such a model is feasible, since it is from the so called *exponential family* of models [15], [20] and reduces the possibly difficult assimilation of data to a simple algebraic operation.
- The model contains as a subset the kind of models used by modern Financial Mathematics in the case that Efficient Market Hypothesis holds. For a more detailed presentation of financial models of such kind see [13], [21], [22], [23] and others.

Remark 3: This is not completely true, since the models used in modern Financial Mathematics are often of the continuous-time type – they have to be discretized to form a subset of a discrete-time AR model. Such a discretization of a Wiener type processes is rather intuitive and is discussed in [24].

Since the probability density in (4) represents a conditional density of random vector $\mathbf{D}_t$, we can use the rules of probability theory [14] and decompose the random vector into conditional mean value and innovation

$$\mathbf{D}_t = \mathbb{E}[\mathbf{D}_t | \boldsymbol{\theta}, \mathcal{F}_{t-1}] + \boldsymbol{\varepsilon}_t = \mathbf{A} \boldsymbol{\Phi}_{t-1} + \boldsymbol{\Sigma} \mathbf{e}_t = \tag{5}$$
$$= \mathbf{A}_1 \mathbf{D}_{t-1} + \mathbf{A}_2 \mathbf{D}_{t-2} + \cdots + \mathbf{A}_p \mathbf{D}_{t-p} + \mathbf{c} + \boldsymbol{\Sigma} \mathbf{e}_t$$

where $\mathbf{e}_t$ is a noise vector having a normal distribution with mean value $\mathbf{0}$ and covariance matrix $\mathbf{I}$ and $\mathbb{E}$ stands for mathematical expectation. The matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$, vector

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:10, 2009

c and matrix $\boldsymbol{\Sigma}$, representing square root of the covariance **R** being now the parameters of the model. If the original noise is uncorrelated, the matrix $\boldsymbol{\Sigma}$ is diagonal.

Remark 4: Note we have changed the realizations of data variables on the right-hand side of (5) to their random variable representation. Since at time $t-1$ or greater the σ-algebra $\mathcal{F}_{t-1}$ is available and $\boldsymbol{\Phi}_{t-1}$ is measurable $\mathcal{F}_{t-1}$, the relationship (5) therefore holds unchanged. The new feature of (5) is that it holds also for future times, when the data in the condition are unknown. This feature will allow us to construct predictions of data evolution into the future.

For further computation, previous equation can be embedded into a wider scheme to be recursive and have the Markov property – see [21] page 49

$$
\underbrace{\begin{bmatrix} \mathbf{D}_t \\ \mathbf{D}_{t-1} \\ \mathbf{D}_{t-2} \\ \vdots \\ \mathbf{D}_{t-p} \\ 1 \end{bmatrix}}_{\boldsymbol{\Phi}_t} = \underbrace{\begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_{p-1} & \mathbf{A}_p & \mathbf{c} \\ \mathbf{I} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{I} & 0 & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \\ 0 & 0 & \mathbf{I} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{D}_{t-1} \\ \mathbf{D}_{t-2} \\ \mathbf{D}_{t-3} \\ \vdots \\ \mathbf{D}_{t-1-p} \\ 1 \end{bmatrix}}_{\boldsymbol{\Phi}_{t-1}} +
$$

$$
+ \underbrace{\begin{bmatrix} \boldsymbol{\Sigma} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 \dots 0 \\ \vdots & & \vdots \\ 0 & 0 & 0 \end{bmatrix}}_{\boldsymbol{\Sigma}} \underbrace{\begin{bmatrix} \mathbf{e}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{e}_t} \qquad (6)
$$

where we have used same notation for the embedded as for some of the original quantities, which should cause no confusion, as we will speak only of the extended quantities from now on.

C. Structure estimation

Although we could use a full dimensional model of a predefined maximal time-lag and predefined number of data channels k such a practice can lead to large models with unnecessary AR component or even unnecessary channel, bringing additional inaccuracy to the model. For that reason Kárný and Kulhavý [25] have proposed a systematic way of Bayesian hypothesis testing for finding the best model through maximum a posteriori likelihood estimation.

D. Parameter estimation

To estimate parameters from past data means to evaluate the second factor on the right-hand side in (3) at each time. We will perform the estimation in a Bayesian manner [26],[27] and use the Bayes theorem. First we have to note that the new σ-algebra $\mathcal{F}_t$ contains only information that has no impact on parameters or $\mathbf{D}_\tau$, $\forall \tau \in \mathcal{T}$ and the information making $\mathbf{D}_t$ measurable $\mathcal{F}_t$. Now we can use Bayes theorem to write

$$
f(\boldsymbol{\theta}|\mathcal{F}_t) = \frac{f(\mathbf{d}_t|\boldsymbol{\theta}, \mathcal{F}_{t-1}) f(\boldsymbol{\theta}|\mathcal{F}_{t-1})}{\int_{\mathcal{R}} f(\mathbf{d}_t|\boldsymbol{\theta}, \mathcal{F}_{t-1}) f(\boldsymbol{\theta}|\mathcal{F}_{t-1}) d\boldsymbol{\theta}} \qquad (7)
$$

where $\mathcal{R}$ is the range of the parameters. Usually $\mathcal{R} = \mathbb{R}^n$, where n is the number of parameters and $\mathbb{R}$ is the set of real numbers. We can use this procedure at every time step to update the posterior probability density function (pdf), since we have already chosen the first factor in numerator on right hand-side in (4). At the first estimation step we need the initial condition - the *prior* pdf. Since we do not know anything about the time series, before the first data are obtained, we should choose a *non-informative* or *Jeffrey's* prior pdf. Such a choice is a special case of a wider class of prior distributions defined by a *conjugate* prior density [26]. Since the model chosen is from the exponential family of models a *conjugate* prior pdf can be chosen in a closed form. In [25] it is shown, such a pdf is of Gauss-Inverse-Wishart (Inverse-Gamma in one-dimensional case) type

$$
GiW(\mathbf{V}, \nu) \propto \qquad (8)
$$
$$
|\mathbf{R}|^{-\frac{\nu+kp+1-k}{2}} \exp \left\{ -\frac{1}{2} \left[\mathbf{R}^{-1} \begin{bmatrix} \mathbf{I} \\ -\mathbf{A} \end{bmatrix}' \mathbf{V} \begin{bmatrix} \mathbf{I} \\ -\mathbf{A} \end{bmatrix} \right] \right\}
$$

where **V** is a positive definite *extended information matrix* and ν is a positive number of degrees of freedom. These parameters have to be chosen before the estimation starts. Such a choice can have a considerable impact on estimation of parameters $\mathbf{A}, \mathbf{R}$, but since in time series analysis $\mathcal{T}$ is usually large, the prior pdf choice can be treated with a little less care.

As described in [26], conjugate priors are chosen so that they are *self-reproducing* when estimation (7) is performed - the probability density function retains the form (8) with $\mathbf{V}$, ν replaced by $\mathbf{V}_t$, ν_t respectively. The estimation step comes down to simple algebraic operation on these parameters, written recursively

$$
\mathbf{V}_t = \mathbf{V}_{t-1} + \begin{bmatrix} \mathbf{d}_t \\ \phi_{t-1} \end{bmatrix} \begin{bmatrix} \mathbf{d}_t' & \phi_{t-1}' \end{bmatrix} \qquad (9)
$$
$$
\nu_t = \nu_{t-1} + 1
$$

where

$$
\mathbf{V}_0 = \mathbf{V} \qquad \nu_0 = \nu \qquad (10)
$$

Remark 5: For computational reasons, the model can be decomposed into individual one dimensional regression models as follows. Because **R** is a regular positive definite and symmetric covariance matrix, it can be LD-decomposed [28] and we obtain

$$
\mathbf{R} = \boldsymbol{\Sigma}\boldsymbol{\Sigma}' = \mathbf{L}\mathbf{D}\mathbf{L}' \qquad (11)
$$

L is a lower triangular matrix with units on the diagonal and **D** is diagonal with non-negative entries. The inverse $\mathbf{L}^{-1}$ is also a lower triangular matrix with units on the diagonal. By multiplying the model in (6) by $\mathbf{L}^{-1}$ and moving additional terms from left-hand to right-hand side we obtain

$$
\boldsymbol{\Phi}_t = \left[\mathbf{I} - \mathbf{L}^{-1} \right] \boldsymbol{\Phi}_t + \mathbf{L}^{-1}\mathbf{A}\boldsymbol{\Phi}_{t-1} + \mathbf{D}^{\frac{1}{2}}\mathbf{e}_t \qquad (12)
$$

where the square-root of **D** is well defined, since all the diagonal elements of **D** are non-negative. The channels of such a model are no longer correlated and the parameters of the model can be estimated for k univariate models instead. Then by a backward transformation, the model can be brought back to its original form. When such a transformation is carried through, the prior information on the parameters is also

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:10, 2009

transformed. Anyhow, in time series analysis, such a change should not be important for the reasons mentioned.

E. Exponential forgetting

In case of slow parameter evolution, we can use the exponential forgetting technique, described in [15], page 46. Before the parameter estimation step a forgetting step is added, accounting for parameter evolution. This step replaces the parameter time-update step [15]. A forgetting factor $\kappa \in (0, 1)$ is chosen and previous parameter probability density is flattened

$$f(\boldsymbol{\theta}_t = \boldsymbol{\theta}|\mathcal{F}_{t-1}) = f(\boldsymbol{\theta}_{t-1} = \boldsymbol{\theta}|\mathcal{F}_{t-1})^{\kappa} \tag{13}$$

For GiW model, inclusion of the forgetting causes a change of (9) to

$$\begin{aligned} \mathbf{V}_t &= \kappa \mathbf{V}_{t-1} + \begin{bmatrix} \mathbf{d}_t \\ \boldsymbol{\phi}_{t-1} \end{bmatrix} \begin{bmatrix} \mathbf{d}'_t & \boldsymbol{\phi}'_{t-1} \end{bmatrix} \\ \nu_t &= \kappa \nu_{t-1} + 1 \end{aligned} \tag{14}$$

The choice of optimal forgetting factor and also the structure of the forgetting are difficult tasks. These tasks can be left for consideration of an expert, but attempts were made to choose this factor systematically [29].

Remark 6: Exponential forgetting influences the model structure choice discussed in subsection (B). To the best knowledge of the author, no satisfactory feasible algorithm of structure estimation has been proposed yet for $\kappa < 1$. Therefore, the practice used is to estimate model structure using $\kappa = 1$.

F. Prediction using Monte Carlo sampling from parameter distribution

Let's now suppose we know the model parameters perfectly – e.g. their estimated joint probability density is a delta function $f(\boldsymbol{\theta}|\mathcal{F}_t) = \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0 = [\mathbf{A}, \boldsymbol{\Sigma}])$ where $\mathbf{A}, \boldsymbol{\Sigma}$ are now matrices of numbers, not random variables. We now want to construct the *prediction* of the stochastic process $\boldsymbol{\Phi}_t$ up to a horizon $t + h$, with the information contained in $\mathcal{F}_t$. We therefore need to evaluate probability density $f(\boldsymbol{\phi}_{t+h}|\mathcal{F}_t)$, characterizing the distribution. Since we know the stochastic process evolves as in (6), if the parameters are known we obtain a predicted random variable

$$\boldsymbol{\Phi}_{t+h} = \mathbf{A}^h \boldsymbol{\Phi}_t + \sum_{i=0}^{h-1} \mathbf{A}^i \boldsymbol{\Sigma} \mathbf{e}_{t+h-i} \tag{15}$$

and we can also compute the mean value (point prediction) and covariance of this random variable

$$\boldsymbol{\mu}_h = \mathbb{E}\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t, \boldsymbol{\theta}\right] = \tag{16}$$

$$\mathbb{E}\left[\mathbf{A}^h \boldsymbol{\Phi}_t|\mathcal{F}_t, \boldsymbol{\theta}\right] + \mathbb{E}\left[\sum_{i=0}^{h-1} \mathbf{A}^i \boldsymbol{\Sigma} \mathbf{e}_{t+h-i}|\mathcal{F}_t, \boldsymbol{\theta}\right] =$$

$$\mathbf{A}^h \boldsymbol{\Phi}_t + \sum_{i=0}^{h-1} \mathbf{A}^i \boldsymbol{\Sigma} \mathbb{E}\left[\mathbf{e}_{t+h-i}|\mathcal{F}_t, \boldsymbol{\theta}\right] = \mathbf{A}^h \boldsymbol{\Phi}_t$$

$$\mathbf{R}_h = \text{cov}\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t, \boldsymbol{\theta}\right] = \tag{17}$$

$$\text{cov}\left[\mathbf{A}^h \boldsymbol{\Phi}_t|\mathcal{F}_t, \boldsymbol{\theta}\right] + \text{cov}\left[\sum_{i=0}^{h-1} \mathbf{A}^i \boldsymbol{\Sigma} \mathbf{e}_{t+h-i}|\mathcal{F}_t, \boldsymbol{\theta}\right] =$$

$$\sum_{i=0}^{h-1} \left(\mathbf{A}^i \boldsymbol{\Sigma}\right) \underbrace{\text{cov}\left[\mathbf{e}_{t+h-i}|\mathcal{F}_t, \boldsymbol{\theta}\right]}_{\mathbf{I}} \left(\mathbf{A}^i \boldsymbol{\Sigma}\right)' =$$

$$\sum_{i=0}^{h-1} \left[\mathbf{A}^i \boldsymbol{\Sigma} \left(\mathbf{A}^i \boldsymbol{\Sigma}\right)'\right]$$

where $\mathbf{A}^0 \equiv \mathbf{I}$, since we consider a model with $\text{cov}\left[\mathbf{e}_t, \mathbf{e}_s|\mathcal{F}_t\right] = \mathbf{0}$ for $s > t$. These first two moments of distribution of $\boldsymbol{\Phi}_{t+h}$ will be important in the next paragraph. Since $\mathbf{e}_t$ are normally distributed $\forall t \in \mathcal{T}$, we obtain the pdf of random variable on the left-hand side in (15) as

$$f(\boldsymbol{\phi}_{t+h}|\mathcal{F}_t) = \varphi_{\boldsymbol{\mu}_h, \mathbf{R}_h}(\boldsymbol{\phi}_{t+h}) \tag{18}$$

where $\varphi_{\boldsymbol{\mu}, \mathbf{R}}$ is the normal multivariate probability density function with mean value $\boldsymbol{\mu}$ and covariance $\mathbf{R}$.

If instead of the delta function, the parameter values are uncertain, we can proceed with the multi-step ahead prediction by drawing N *Monte Carlo* samples $\boldsymbol{\theta}_i$, where $i \in \{1, \ldots, N\}$ from their probability distribution characterized by $f(\boldsymbol{\theta}|\mathcal{F}_t)$. After having evaluated the predictions, we reconstruct the predictive probability density by simple numerical integration.

In an exact computation, we should consider the predicted value for $t+1$ to estimate new probability density $f(\boldsymbol{\theta}|\mathcal{F}_{t+1})$ and we should perform the forgetting (13). Instead, for computational feasibility, we use a so called *receding horizon* or *moving window* approximation. In this approximation

$$f(\boldsymbol{\theta}|\mathcal{F}_s) = f(\boldsymbol{\theta}|\mathcal{F}_t) \qquad t < s \leq t + h \tag{19}$$

for purposes of prediction up to time $t+h$ – we do not update the parameter distribution with the use of *estimated* data. Once we obtain new real data at time $t + 1$, we proceed with parameter estimation (7), forget (13), again fix the distribution, draw new N samples and predict up to horizon $t+1+h$ and so on. This approximation allows us to use the previously obtained result (18) for multistep-ahead predictive probability density function, except now we obtain N such result, each conditioned on the parameter value $\boldsymbol{\theta}_i$ drawn from $f(\boldsymbol{\theta}|\mathcal{F}_t)$.

We now reconstruct the final predictive probability density function $f(\boldsymbol{\phi}_{t+h}|\mathcal{F}_t)$ by integrating out the parameters, which in the Monte Carlo approximation results in an averaging

$$f_N(\boldsymbol{\phi}_{t+h}|\mathcal{F}_t) = \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{\phi}_{t+h}|\boldsymbol{\theta}_i, \mathcal{F}_t) \tag{20}$$

In the moving window approximation such a probability density function should converge point wise to the distribution obtained by a general integration for $N \to \infty$.

From the distribution obtained we can generally compute its central moments, which characterize the distribution of the predicted values of the stochastic process $\boldsymbol{\Phi}_t$. For illustration we now compute the first two central moments of the distribution. With the first two moments known, we could fit a

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:10, 2009

normal distribution to the prediction, although it is certain, that the uncertainty in parameters $\boldsymbol{\theta}$ causes the real predictive distribution to be heavy–tailed. For the mean value we get

$$\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t\right] = \int_{\mathbb{R}^{kp+1}} \boldsymbol{\phi}_{t+h} f_N(\boldsymbol{\phi}_{t+h}|\mathcal{F}_t)\, d\boldsymbol{\phi}_{t+h} =$$

$$\int_{\mathbb{R}^{kp+1}} \boldsymbol{\phi}_{t+h} \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{\phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i)\, d\boldsymbol{\phi}_{t+h} \quad (21)$$

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i\right]$$

if the sum and integral can be transposed, which we assume. From knowledge of the covariance matrices $\mathbf{R}_{h;i} = \operatorname{cov}(\boldsymbol{\phi}_{t+h}|\boldsymbol{\theta}_i,\mathcal{F}_t)$ we can compute

$$\operatorname{cov}\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t\right] = \quad (22)$$

$$\mathbb{E}_N\left[\left(\boldsymbol{\phi}_{t+h}-\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t\right]\right)\left(\boldsymbol{\phi}_{t+h}-\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t\right]\right)'|\mathcal{F}_t\right]$$

$$= \int_{\Omega}\left(\boldsymbol{\phi}_{t+h}-\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_j\right]\right)\cdot$$

$$\left(\boldsymbol{\phi}_{t+h}-\frac{1}{N}\sum_{l=1}^{N}\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_l\right]\right)'\cdot$$

$$\frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{\phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i)\, d\boldsymbol{\phi}_{t+h} =$$

$$\frac{1}{N^3}\sum_{i=1}^{N}\int_{\Omega}\sum_{j=1}^{N}\left(\boldsymbol{\phi}_{t+h}-\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_j\right]\right)\cdot$$

$$\sum_{l=1}^{N}\left(\boldsymbol{\phi}_{t+h}-\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_l\right]\right)' f(\boldsymbol{\phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i)\, d\boldsymbol{\phi}_{t+h} =$$

$$\frac{1}{N^3}\sum_{i,j,l}\int_{\Omega}(\boldsymbol{\phi}_{t+h}\boldsymbol{\phi}'_{t+h}-\boldsymbol{\phi}_{t+h}\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_l\right]'-$$

$$\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_j\right]\boldsymbol{\phi}'_{t+h}+$$

$$\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_j\right]\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_l\right]')f(\boldsymbol{\phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i)\, d\boldsymbol{\phi}_{t+h} =$$

$$\frac{1}{N^3}\sum_{i,j,l}(\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}\boldsymbol{\phi}'_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i\right]-$$

$$\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i\right]\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_l\right]'-$$

$$\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_j\right]\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i\right]'+$$

$$\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_j\right]\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_l\right]') =$$

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{R}_{h;i}+\frac{1}{N}\sum_{i=1}^{N}(\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i\right])(\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i\right])'-$$

$$\frac{1}{N^2}\left(\sum_{i=1}^{N}\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_i\right]\right)\left(\sum_{j=1}^{N}\mathbb{E}_N\left[\boldsymbol{\Phi}_{t+h}|\mathcal{F}_t,\boldsymbol{\theta}_j\right]\right)'$$

where $\Omega = \mathbb{R}^{kp+1}$. We again suppose, we can transpose sums and integrals. Higher moments could be computed to obtain a more accurate approximation of the final distribution.

G. Incorporating log-normally distributed data channels into the model

In the Bayesian framework models from exponential family have an exclusive position [15],[26],[27] for feasible parameter inference. The family is quite restricted, but except for the normal model described above, it also contains a log-normal model, which is sometimes a more adequate choice for the problem at hand, but brings a few difficulties, when used for multi-step ahead prediction. We will present and resolve these difficulties in this paragraph. We will also assume a mixed model containing normal and log-normal components.

Now the data vector $\mathbf{d}_t$ contains channels $\mathbf{n}_t$ and to these channels we associate a normal autoregressive process $\mathbf{N}_t$ as before and channels $\mathbf{l}_t$, data that need to be logarithmized in order to follow a normal stochastic process $\ln\mathbf{L}_t$. The new model equation follows to clarify the situation

$$\mathbf{D}_t = \begin{bmatrix}\ln\mathbf{L}_t \\ \mathbf{N}_t\end{bmatrix} = \quad (23)$$
$$\mathbf{A}_1\mathbf{D}_{t-1}+\mathbf{A}_2\mathbf{D}_{t-2}+\cdots+\mathbf{A}_p\mathbf{D}_{t-p}+\mathbf{c}+\boldsymbol{\Sigma}\,\mathbf{e}_t =$$
$$\mathbf{A}_1\begin{bmatrix}\ln\mathbf{L}_{t-1} \\ \mathbf{N}_{t-1}\end{bmatrix}+\cdots+\mathbf{A}_p\begin{bmatrix}\ln\mathbf{L}_{t-p} \\ \mathbf{N}_{t-p}\end{bmatrix}+\mathbf{c}+\boldsymbol{\Sigma}\,\mathbf{e}_t$$

where $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0},\mathbf{I})$ for all $t\in\mathcal{T}$. The first problem is that the random walk is biased in the space of backward transformed data.

Lemma 1: Let $\mathbf{L}$ be a random vector that when logarithmized is jointly normally distributed – $\ln\mathbf{L}\sim\mathcal{N}(\boldsymbol{\mu},\mathbf{R})$ having k entries. Then $\mathbf{L}$ is characterized by multivariate log-normal probability density function

$$f(l_1,\ldots,l_k) = \quad (24)$$
$$[2\pi]^{-\frac{k}{2}}|\mathbf{R}|^{-\frac{1}{2}}\prod_{i=1}^{k}\frac{1}{l_i}\exp\left\{-\left[\ln\mathbf{l}-\boldsymbol{\mu}\right]^T\mathbf{R}^{-1}\left[\ln\mathbf{l}-\boldsymbol{\mu}\right]\right\}$$

where $\ln\mathbf{l} = [\ln l_1\cdots\ln l_k]'$ and the first two moments of its distribution for are given by

$$\mathbb{E}\left[L_i\right] = e^{\mu_i+R_{ii}/2} \quad (25)$$

and

$$\operatorname{cov}\left[L_i,L_j\right] = \left[e^{R_{ij}}-1\right]e^{\mu_i+\mu_j+\frac{R_{ii}+R_{jj}}{2}} \quad (26)$$

where $i,j\in\{1,\ldots,k\}$.

Proof: For proof using moment generating functions see [30]. ∎

Application of (25) and (26) to the mixed vector in (23) leads to a one-step ahead mean-value prediction conditioned on the parameters

$$\mathbb{E}\left(\begin{bmatrix}\mathbf{l}_{t+1} \\ \mathbf{n}_{t+1}\end{bmatrix}|\mathcal{F}_t,\boldsymbol{\theta}\right) = \begin{bmatrix}\mathbb{E}\left(\mathbf{l}_{t+1}|\mathcal{F}_t,\boldsymbol{\theta}\right) \\ \mathbb{E}\left(\mathbf{n}_{t+1}|\mathcal{F}_t,\boldsymbol{\theta}\right)\end{bmatrix} = \quad (27)$$
$$\begin{bmatrix}\exp\left(\boldsymbol{\mu}_l+\frac{\sigma_l^2}{2}\right) \\ \boldsymbol{\mu}_n\end{bmatrix} \neq \begin{bmatrix}\exp\mathbb{E}\left(\ln\mathbf{l}_{t+1}|\mathcal{F}_t,\boldsymbol{\theta}\right) \\ \mathbb{E}\left(\mathbf{n}_{t+1}|\mathcal{F}_t,\boldsymbol{\theta}\right)\end{bmatrix}$$

where $\boldsymbol{\sigma}^2 = \operatorname{diag}\mathbf{R}$ is a vector of diagonal components of matrix $\mathbf{R}$, $\boldsymbol{\mu}$ is the mean-values vector and the indices l and n specify the parts of the vectors belonging to the normally and log-normally distributed components. In the above equation we can see that the prediction of a process associated with a log-normally distributed data channel i is now positively biased when backward-transformed into the original data space by a multiplicative factor of $e^{\sigma_i^2/2}$.

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:10, 2009

We now want to predict in such a model up to horizon h. We apply equation (15) again, but we now have to be careful when taking expectations, because we need the backward-transformed predictions

$$\mathbb{E}\left[\begin{bmatrix} \mathbf{L}_{t+h} & \mathbf{N}_{t+h} & \cdots & \mathbf{L}_{t+h-p} & \mathbf{N}_{t+h-p} \end{bmatrix}'|\mathcal{F}_t, \boldsymbol{\theta}\right] = \mathbb{E}\left[\boldsymbol{\phi}_{t+h}|\mathcal{F}_t, \boldsymbol{\theta}\right] \quad (28)$$

To obtain such predictions, we have to combine (15), (25) and (26) and a little extra

Lemma 2: Let $\boldsymbol{\rho} = \begin{bmatrix} \ln \mathbf{L} \\ \mathbf{N} \end{bmatrix}$ of $l + n$ elements belonging respectively to the individual parts and $\boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{R})$. Then for a vector $\boldsymbol{\phi} = \begin{bmatrix} \mathbf{L} \\ \mathbf{N} \end{bmatrix}$ the mixed covariance

$$\text{cov}\left[\phi_i, \phi_j\right] = r_{ij} \exp\left\{\mu_i + \frac{r_{ii}}{2}\right\} \quad (29)$$

for $i \in \{1, \ldots, l\}, j \in \{l+1, \ldots, l+n\}$.

Proof: We start from the moment generating function

$$\mathbb{E}\left[\exp\left(\mathbf{t}'\boldsymbol{\rho}\right)\right] = \exp\left(\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\mathbf{R}\,\mathbf{t}\right) = \mathbb{E}\left\{\left[\prod_{i=1}^{l} \phi_i^{t_i}\right]\left[\exp\left(\sum_{j=l+1}^{l+n} \phi_j t_j\right)\right]\right\} \quad (30)$$

where the third term on the right-hand side is the multivariate moment generating function of normal distribution [30]. From here wee see that

$$\mathbb{E}\left[\phi_i\phi_j\right] = \lim_{\substack{t_k \to 0 \\ t_i \to 1}} \frac{\partial}{\partial t_j}\left[\exp\left(\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\mathbf{R}\,\mathbf{t}\right)\right] \quad (31)$$

for $i \in \{1, \ldots, l\}, j \in \{l+1, \ldots, l+n\}$. Since

$$\text{cov}\left[\phi_i, \phi_j\right] = \mathbb{E}\left[\phi_i\phi_j\right] - \mathbb{E}\left[\phi_i\right]\mathbb{E}\left[\phi_j\right] \quad (32)$$

we use (25) and we plug in the limit values in the equation above to obtain the desired result. ∎

Corollary 1: The above proposition can be generalized to a permuted random vector $\boldsymbol{\rho}_\mathbf{P} = \mathbf{P}\boldsymbol{\rho}$, where $\mathbf{P}$ is an arbitrary permutation matrix ($|\mathbf{P}| = 1, \forall i, j \ P_{ij} \in \{0, 1\}, \sum_i P_{ij} = 1, \sum_j P_{ij} = 1$).

III. EXPERIMENTAL RESULTS

The model introduced in previous section was tested on daily data from 11 U.S. commodity futures markets in period from 2.1.1990 to 9.8.2005. The markets considered have been

1) Australian Dollar [AD] (Currency, CME)
2) British Pound [BP] (Currency, CME)
3) Cocoa [CC] (Soft, CSCE)
4) Canadian Dollar [CD] (Currency, CME)
5) Light Crude Oil [CL] (Energy, NYMEX)
6) Cotton [CT] (Grain, NYCE)
7) Feeder Cattle [FC] (Livestock, CME)
8) Gold [GC] (Metal, COMEX)
9) Heating Oil [HO] (Energy, NYMEX)
10) Gasoline [HU] (Energy, NYMEX)

11) Wheat [W] (Grain, CBOT)

From these markets we have incorporated 33 information channels into the model. Among these channels we have assumed: opening, highest, lowest and closing prices, trading volume, open interest, spot price, commitment of traders information and a few others. From this information we tried to predict the closing price of the futures contracts up to a horizon of $h = 14$ days. The data at hand contained 3928 trading days from which 1700 trading days were used to train the model and estimate model structure. After the warm-up period multi-step ahead out-of-sample forecasts of closing price distribution were made.

Remark 7: Futures contract usually are not traded for a period of 15 years. Therefore the time series had to be merged from data of more different contracts. The time series used were synthesized as follows: The prices at the end of the trading period are real market prices and as we go back in time, when the active contract (the contract with highest trade activity) changes, we switch to the previous active contract with price adjusted by an additive constant, so that there is no gap at the time of change. This way we create an artificial time series, which will differ from the time series of real prices. In the future this artificial transformation of data will be removed and the models will be used on a single contract time series.

For prediction, we have used four models of the presented type. Two of them were obtained from a structure estimation procedure described above and two were benchmark models used extensively in todays financial mathematics. Maximum order $p = 2$ of the AR process has been chosen for computational feasibility of the structure estimation.

So far, all the criteria of success measured, were based only on the conditional value $\mathbb{E}\left[Y_{t+h}|\mathcal{F}_t\right]$ – on the *point* prediction of future price evolution. Therefore from now on prediction always means such a point prediction.

In the first BVAR model with estimated structure and the first benchmark model the price channels have been transformed using logarithmic transformation. After the prediction has been made for the future evolution of closing price, the predictions were backward transformed into the space of real-world prices – we have obtained point predictions of the closing price by the methods described in the theoretical part of the paper.

In the second BVAR model and a second benchmark model, no transformation was performed and a direct prediction of future evolution of closing price has been made.

The two univariate benchmark models were the ones commonly used in todays financial mathematics. These models are in agreement with the semi-strong form of Efficient Market Hypothesis. In the first model the closing price

$$Y_{t+1} = Y_t \exp\{\alpha + \sigma e_{t+1}\} \quad (33)$$

or equivalently

$$\ln Y_{t+1} - \ln Y_t = \alpha + \sigma e_{t+1} \quad (34)$$

where in the second benchmark model

$$Y_{t+1} = Y_t + \alpha + \sigma e_{t+1} \quad (35)$$

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:10, 2009

TABLE I
OUT-OF-SAMPLE MEDIAN RELATIVE ERROR COMPUTED FOR THE HIGHEST LIKELIHOOD LOG-NORMAL AND NORMAL BVAR(2) MODEL AND LOG-NORMAL (33) AND NORMAL (35) BENCHMARK MODEL FOR HORIZONS $1, \ldots, 7$. THE ERROR IS SHOWN IN PARTS PER MILLE ($\cdot 10^3$).

mkt.	model	horizon 1 frg	MERE	2 frg	MERE	3 frg	MERE	4 frg	MERE	5 frg	MERE	6 frg	MERE	7 frg	MERE
AD	BVAR(log)	**3**	**4.68**	**17**	**6.87**	2	8.70	14	10.40	**3**	**11.46**	**3**	**12.47**	**3**	**13.67**
	benchmark(log)	4	4.70	6	6.91	**3**	**8.57**	**1**	**10.02**	2	11.48	3	12.56	4	13.84
	BVAR(normal)	**3**	**4.68**	5	6.89	4	8.73	3	10.27	3	11.57	3	12.65	2	13.78
	benchmark(normal)	6	6.91	2	8.57	1	10.04	2	11.47	1	12.57	1	13.65	2	14.41
BP	BVAR(log)	7	3.34	6	4.90	5	6.08	5	7.13	3	8.09	5	9.16	4	10.12
	benchmark(log)	**3**	**3.31**	**5**	**4.79**	**1**	**5.99**	**18**	**7.04**	**2**	**8.01**	**7**	**8.93**	**3**	**9.88**
	BVAR(normal)	7	3.36	6	4.89	5	6.08	4	7.08	5	8.09	5	9.15	5	10.12
	benchmark(normal)	4	4.77	1	5.89	2	6.97	2	7.92	17	8.93	1	9.77	2	10.29
CC	BVAR(log)	21	9.79	**11**	**13.39**	**4**	**16.89**	3	19.11	6	22.20	3	24.61	5	26.73
	benchmark(log)	**21**	**9.77**	12	13.44	**4**	**16.89**	**2**	**19.07**	4	22.28	3	24.92	5	26.81
	BVAR(normal)	21	9.78	17	13.41	3	17.05	2	19.18	**4**	**22.13**	**3**	**24.49**	**4**	**26.64**
	benchmark(normal)	2	13.43	1	16.80	1	19.05	1	22.18	2	24.71	4	26.91	11	29.14
CD	BVAR(log)	**6**	**2.40**	**5**	**3.56**	6	4.47	**10**	**5.23**	2	6.09	10	6.59	**21**	**6.94**
	benchmark(log)	3	2.45	5	3.60	2	4.53	2	5.33	1	6.11	1	6.72	3	7.10
	BVAR(normal)	7	2.42	5	3.58	**10**	**4.46**	7	5.25	**21**	**6.01**	**20**	**6.55**	14	6.98
	benchmark(normal)	4	3.57	1	4.46	1	5.33	1	6.06	1	6.70	1	7.09	1	7.56
CL	BVAR(log)	10	7.06	2	10.47	9	12.66	6	14.80	**13**	**16.62**	9	18.82	**6**	**19.81**
	benchmark(log)	**11**	**6.99**	**2**	**10.34**	**3**	**12.53**	**4**	**14.78**	6	16.66	**3**	**18.64**	9	19.93
	BVAR(normal)	4	7.12	3	10.59	5	12.88	4	15.44	6	18.34	7	NaN	7	NaN
	benchmark(normal)	1	10.34	1	12.59	3	14.96	7	16.73	1	18.64	2	19.68	3	20.79
CT	BVAR(log)	**17**	**5.36**	4	7.65	2	9.65	5	11.17	**1**	**12.30**	**3**	**13.59**	4	15.09
	benchmark(log)	7	5.42	17	7.67	**21**	**9.62**	3	11.20	1	12.44	3	13.74	**1**	**14.94**
	BVAR(normal)	10	5.37	**6**	**7.62**	6	9.64	**2**	**11.15**	14	12.58	2	13.68	4	15.11
	benchmark(normal)	11	7.66	6	9.54	1	11.18	1	12.33	1	13.33	1	15.02	1	15.91
FC	BVAR(log)	6	4.82	4	7.20	1	8.75	9	9.91	13	10.88	**21**	**12.15**	**7**	**13.30**
	benchmark(log)	6	4.79	**8**	**7.17**	12	8.77	7	9.91	**21**	**10.86**	19	12.23	4	13.40
	BVAR(normal)	**6**	**4.78**	6	7.18	**5**	**8.74**	**9**	**9.90**	19	10.91	21	12.27	4	13.39
	benchmark(normal)	5	7.17	1	8.72	1	9.93	18	10.98	11	12.23	5	13.33	3	14.36
GC	BVAR(log)	21	3.95	5	6.00	8	7.64	7	8.65	6	9.94	4	11.31	4	12.65
	benchmark(log)	14	3.93	2	5.94	5	7.52	**9**	**8.53**	**5**	**9.76**	2	11.02	5	12.20
	BVAR(normal)	**12**	**3.92**	**4**	**5.93**	**4**	**7.51**	7	8.53	3	9.77	**3**	**10.93**	**4**	**12.15**
	benchmark(normal)	3	5.96	5	7.58	3	8.49	3	9.71	3	10.96	4	12.16	5	12.95
HO	BVAR(log)	7	10.14	4	14.98	2	18.46	3	22.10	2	24.50	5	27.87	3	29.85
	benchmark(log)	17	10.16	**4**	**14.79**	**3**	**18.33**	**5**	**21.71**	**3**	**24.22**	**2**	**27.08**	11	**29.42**
	BVAR(normal)	**5**	**10.13**	5	14.93	2	18.42	5	21.82	3	24.40	2	27.49	5	29.56
	benchmark(normal)	1	14.66	1	17.93	2	21.83	1	24.07	2	27.09	2	29.10	7	30.82
HU	BVAR(log)	4	7.17	8	10.78	4	13.58	6	15.59	5	17.76	**16**	**18.88**	**10**	**19.92**
	benchmark(log)	3	7.18	**4**	**10.66**	**3**	**13.51**	**2**	**15.41**	**7**	**17.66**	7	19.02	6	19.97
	BVAR(normal)	**4**	**7.16**	6	10.74	3	13.55	4	15.68	8	17.70	6	19.09	5	20.13
	benchmark(normal)	1	10.62	1	13.53	2	15.46	4	17.75	9	19.09	8	19.91	3	21.16
W	BVAR(log)	13	5.86	5	8.27	4	9.96	**5**	**11.58**	5	13.30	4	14.14	12	15.53
	benchmark(log)	9	5.86	**2**	**8.19**	5	**9.86**	7	11.61	3	13.17	12	14.18	**4**	**15.40**
	BVAR(normal)	**6**	**5.85**	2	8.20	4	9.91	6	11.59	**4**	**13.16**	5	**14.07**	5	15.51
	benchmark(normal)	2	8.24	4	9.88	2	11.61	1	13.09	3	14.09	3	15.49	4	16.90

Both these models are in the set of models considered in the structure estimation step and are ruled out because of a lower likelihood, but we still measure their out-of-sample performance, against the out-of-sample performance of their highest likelihood BVAR counterparts.

Twenty forgetting factors κ_i have been chosen as values of

$$\kappa_i = 1 - \exp\{-x_i\} \qquad (36)$$

where $x_1 = 4, x_2 = 4.5, x_3 = 5, \ldots, x_{20} = 14$ so that the forgetting factor ranges from about 0.98 to almost 1. The forgetting factor has been kept constant for all channels and parameters.

As the measure of success of the prediction we first take the median relative error of the forecast

$$MERE = \text{median} \left| \underbrace{\mathbb{E}\left[Y_{t+k}|\mathcal{F}_t\right]}_{\hat{y}_{t+k}} / y_{t+k} - 1 \right| \qquad k \in \{1, \ldots, h\}$$
(37)

where $\hat{y}_{t+k}$ is the already mentioned point estimate of price at time $t + k$ and y_{t+k} is real price at the same time. Median error is chosen for its robustness, since some of the models that are unstable may produce heavily outlying predictions.

Remark 8: The predicted change in price is small compared to the price in vast majority of cases and especially in the cases with small prediction error. Therefore negative prices are not predicted and the ratio in MERE computation stays positive,

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:10, 2009

TABLE II

OUT-OF-SAMPLE MEDIAN RELATIVE ERROR COMPUTED FOR THE HIGHEST LIKELIHOOD LOG-NORMAL AND NORMAL BVAR(2) MODEL AND LOG-NORMAL (33) AND NORMAL (35) BENCHMARK MODEL FOR HORIZONS $8, \ldots, 14$. THE ERROR IS SHOWN IN PARTS PER MILLE ($\cdot 10^3$).

horizon		8		9		10		11		12		13		14	
mkt.	model	frg	MERE	frg	MERE	frg	MERE	frg	MERE	frg	MERE	frg	MERE	frg	MERE
AD	BVAR(log)	3	14.48	2	15.17	3	16.41	3	17.18	2	17.73	2	18.23	2	19.59
	benchmark(log)	**4**	**14.41**	**1**	**15.08**	**3**	**16.02**	**1**	**16.88**	**2**	**17.65**	**2**	**18.20**	**2**	**19.54**
	BVAR(normal)	3	14.54	3	15.10	4	18.56	4	224.97	4	NaN	4	NaN	4	NaN
	benchmark(normal)	1	14.99	1	15.95	1	16.82	2	17.44	2	17.93	1	19.17	1	20.35
BP	BVAR(log)	6	10.57	4	11.43	4	12.26	6	12.78	5	13.58	5	14.12	7	14.62
	benchmark(log)	**4**	**10.37**	**14**	**11.10**	**2**	**11.93**	**5**	**12.58**	**6**	**13.31**	**19**	**13.86**	**17**	**14.43**
	BVAR(normal)	6	10.61	5	11.36	5	12.31	6	14.26	7	232.98	7	NaN	7	NaN
	benchmark(normal)	2	11.05	2	11.92	4	12.46	9	13.24	9	13.86	2	14.38	2	14.99
CC	BVAR(log)	9	29.13	11	31.54	7	33.20	18	34.99	14	36.30	21	37.57	21	37.90
	benchmark(log)	6	29.02	4	31.34	5	33.17	**4**	**34.73**	**4**	**36.08**	**2**	**36.73**	19	37.84
	BVAR(normal)	**4**	**29.00**	**3**	**30.70**	**4**	**32.46**	2	34.77	4	36.83	3	37.60	4	38.45
	benchmark(normal)	2	30.83	2	32.93	3	34.83	4	35.70	2	36.60	21	37.77	3	38.55
CD	BVAR(log)	**7**	**7.50**	18	7.87	7	8.52	7	8.87	5	9.40	8	9.69	15	10.06
	benchmark(log)	8	7.65	8	8.04	5	8.74	2	9.05	2	9.60	2	10.03	3	10.35
	BVAR(normal)	18	7.51	**14**	**7.84**	**14**	**8.48**	**12**	**8.86**	**11**	**9.34**	**10**	**9.65**	**7**	**10.04**
	benchmark(normal)	21	8.05	1	8.64	1	8.99	4	9.58	4	9.99	2	10.28	3	10.60
CL	BVAR(log)	**4**	**21.01**	5	22.24	6	23.37	**6**	**24.44**	**4**	**25.69**	6	27.75	5	28.92
	benchmark(log)	**2**	**21.01**	**2**	**21.64**	**3**	**23.01**	2	24.46	3	26.13	**2**	**27.60**	**2**	**28.63**
	BVAR(normal)	7	NaN	7	NaN	7	NaN	7	NaN	7	NaN	7	NaN	7	NaN
	benchmark(normal)	2	21.65	2	22.74	2	23.88	2	25.71	2	27.47	1	28.54	2	28.65
CT	BVAR(log)	4	15.97	3	17.13	3	18.31	**3**	**19.41**	**5**	**20.40**	**3**	**21.27**	3	22.33
	benchmark(log)	2	15.88	**3**	**17.11**	**4**	**18.13**	5	19.46	5	20.48	4	21.28	**3**	**22.13**
	BVAR(normal)	**2**	**15.84**	3	17.15	4	18.17	2	19.53	4	20.56	4	21.61	3	22.49
	benchmark(normal)	3	16.87	3	18.21	4	19.26	4	20.44	4	21.03	3	22.03	4	23.10
FC	BVAR(log)	**6**	**14.28**	**3**	**15.02**	**8**	**15.67**	6	16.58	6	17.47	**6**	**18.14**	**7**	**18.59**
	benchmark(log)	4	14.41	14	15.26	13	15.76	5	16.58	4	17.51	4	18.19	21	18.89
	BVAR(normal)	3	14.38	9	15.35	21	15.89	**11**	**16.55**	**3**	**17.45**	4	18.55	7	18.83
	benchmark(normal)	2	15.24	18	15.90	7	16.57	2	17.39	2	18.36	10	18.93	6	19.32
GC	BVAR(log)	5	13.25	5	14.45	5	15.44	5	16.18	6	17.23	7	17.81	8	18.24
	benchmark(log)	5	13.08	**3**	**13.90**	**4**	**14.94**	4	15.78	4	16.59	21	17.25	4	17.32
	BVAR(normal)	**6**	**12.96**	4	13.98	5	15.00	**5**	**15.66**	**5**	**16.54**	**6**	**17.24**	**5**	**17.28**
	benchmark(normal)	3	13.87	3	15.09	4	15.84	4	16.47	19	17.29	6	17.31	5	18.23
HO	BVAR(log)	4	31.72	4	33.13	3	35.09	5	36.90	6	38.60	4	40.30	6	41.35
	benchmark(log)	**12**	**30.73**	**11**	**32.36**	**19**	**34.16**	**10**	**35.42**	**3**	**37.96**	**3**	**39.13**	**14**	**40.17**
	BVAR(normal)	5	31.07	7	33.02	4	34.38	7	NaN	7	NaN	7	NaN	7	NaN
	benchmark(normal)	2	32.34	3	33.97	3	34.66	2	36.38	2	38.19	20	39.81	8	41.91
HU	BVAR(log)	4	21.46	3	22.84	3	24.32	5	25.27	4	27.06	3	27.89	11	29.09
	benchmark(log)	**3**	**21.24**	**3**	**22.70**	**3**	**23.87**	4	25.08	**2**	**26.47**	**2**	**27.42**	**2**	**28.49**
	BVAR(normal)	4	21.39	5	22.81	4	24.28	**5**	**24.99**	4	26.72	2	27.62	3	28.61
	benchmark(normal)	3	22.84	2	23.96	3	24.87	2	26.07	3	27.20	2	28.12	2	29.84
W	BVAR(log)	8	16.86	**7**	**18.12**	6	19.27	3	20.51	**5**	**21.42**	**6**	**22.64**	3	23.17
	benchmark(log)	4	16.94	6	18.27	6	19.35	3	20.68	5	21.63	6	22.67	**3**	**23.16**
	BVAR(normal)	**3**	**16.82**	4	18.15	**5**	**19.18**	**3**	**20.46**	6	21.45	5	22.72	3	23.24
	benchmark(normal)	5	18.21	5	19.23	3	20.55	2	21.46	6	22.49	5	23.27	3	23.65

so that it is a good measure of a relative error. Asymmetry between positive and negative relative error is neglected.

The results for $h = \{1, \ldots, 7\}$ and $h = \{8, \ldots, 15\}$ are summarized in Tables I and II respectively. The index of forgetting factor κ_i with minimal MERE chosen for each model and each horizon is written in the first column, the error per mille of the actual price in the second one. At every given horizon, the results of the model with lowest MERE among the four competitors is highlighted by the use of a bold font.

For the purposes of a real agent trading in the markets the magnitude of prediction error is not the most important indicator of trading success. For an honest measure of such a success a Dynamic Programming optimization should be performed for the predictions, see [31], [32], transaction fees should be considered and constraints should be set for number

of held futures contracts or trading capital. All these properties of trading have to be taken seriously, but are out of scope of the presented paper. We instead measure something, possibly called "trading potential".

If the direction of the trade misses the direction of market evolution, the agent looses money even if MERE is low, where if the directions agree, she makes money even if the error is high. The situation is shown in Figure 1, where $\hat{y}_{t+1}$ and $\bar{y}_{t+1}$ are the two possible predictions made at time t for time $t+1$ and y_t, y_{t+1} are the real prices at times t, $t+1$ respectively. We therefore use a second simple measure of trading success based on a an open-loop buy and hold strategy. The agent at time t buys a single contract if she expects the closing price to rise $\hat{y}_{t+k} > y_t$ and sells if she expects a price fall $- \hat{y}_{t+k} < y_t$ for a given k. We compute a cumulative gain

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:10, 2009

TABLE III
TRADING POTENTIAL COMPUTED FOR THE HIGHEST LIKELIHOOD LOG-NORMAL AND NORMAL BVAR(2) MODEL AND LOG-NORMAL (33) AND NORMAL (35) BENCHMARK MODEL FOR HORIZONS $1, \ldots, 7$. THE POTENTIAL IS SHOWN IN US\$, TRANSACTION COSTS AND LIMITED TRADING CAPITAL ARE NOT TAKEN INTO ACCOUNT.

	horizon	1		2		3		4		5		6		7	
mkt.	model	frg	GAINS	frg	GAINS	frg	GAINS	frg	GAINS	frg	GAINS	frg	GAINS	frg	GAINS
AD	BVAR(log)	**2**	**58530**	3	66020	3	109600	3	165450	3	202520	3	246260	3	272440
	benchmark(log)	4	44290	**4**	**86720**	**4**	**132440**	**4**	**174750**	**4**	**207700**	**4**	**251200**	**4**	**291300**
	BVAR(normal)	1	50230	2	74600	3	128460	2	144950	2	189300	2	221980	2	248100
	benchmark(normal)	3	47530	1	14600	2	45880	1	63930	1	88780	1	130520	2	166000
BP	BVAR(log)	15	78019	6	32037	6	42437	1	-7375	2	33637	2	-22706	7	-24363
	benchmark(log)	6	40481	6	**82350**	6	**112475**	6	**129925**	6	**173300**	6	**202419**	6	**230325**
	BVAR(normal)	**7**	**79006**	9	44900	6	101825	6	80825	6	74025	6	76569	4	50112
	benchmark(normal)	15	14406	11	42875	11	74250	8	32988	3	29688	3	62156	3	100275
CC	BVAR(log)	20	7970	14	15170	**15**	25230	**15**	33020	18	43370	12	41320	12	45200
	benchmark(log)	**20**	**8930**	15	12490	16	17590	17	23020	17	29030	17	34620	17	40380
	BVAR(normal)	9	5750	9	11450	9	17170	9	23020	9	29030	9	34620	9	40380
	benchmark(normal)	2	1810	**20**	**17790**	4	7010	3	16000	**2**	**43630**	**3**	**49740**	**3**	**81440**
CD	BVAR(log)	**9**	**75040**	18	82420	6	102470	**10**	**123990**	**12**	**175630**	**12**	**211200**	**12**	**244690**
	benchmark(log)	5	23720	6	40860	5	63330	5	82270	5	103330	5	125220	5	142590
	BVAR(normal)	7	73460	8	80320	**7**	**102970**	7	118950	7	159190	7	177760	10	201390
	benchmark(normal)	4	17200	7	12100	5	12710	8	41810	1	59350	1	83160	2	103470
CL	BVAR(log)	**5**	**57450**	5	56610	5	98100	6	165720	6	212610	**6**	**271300**	6	**336040**
	benchmark(log)	3	55730	**3**	**91710**	**3**	**131320**	**3**	**170840**	**3**	**216690**	3	249800	3	300700
	BVAR(normal)	6	48230	1	-10250	1	-9180	1	-15020	3	64650	12	70840	12	91420
	benchmark(normal)	7	-30050	5	11910	1	29040	3	54000	3	65170	4	78620	1	120460
CT	BVAR(log)	16	58725	7	55050	4	72775	**4**	**123240**	**4**	**147110**	**4**	**172970**	**4**	**201075**
	benchmark(log)	5	21335	4	32770	4	51135	5	70160	5	98080	5	125150	5	139445
	BVAR(normal)	**7**	**61515**	7	59470	4	73065	3	93760	4	131210	4	171070	4	199745
	benchmark(normal)	16	5225	**4**	**72090**	**3**	**88135**	1	73400	4	66290	2	76290	4	106125
FC	BVAR(log)	**8**	**29775**	**11**	46950	20	103675	9	132825	8	**149813**	16	**163250**	18	139813
	benchmark(log)	4	21500	11	41500	11	61825	11	82175	11	102062	11	121275	**11**	**140538**
	BVAR(normal)	11	20525	11	41500	11	61825	11	82175	11	102062	11	121275	11	140538
	benchmark(normal)	2	-45325	4	-15000	1	20150	1	7250	2	10813	2	26500	2	60288
GC	BVAR(log)	**5**	**36010**	7	38270	1	37700	2	46050	2	44230	2	66420	2	40930
	benchmark(log)	5	18430	**5**	**39130**	**5**	**55040**	**5**	**76070**	**5**	**93750**	**5**	**111820**	**5**	**131470**
	BVAR(normal)	4	15570	4	29910	4	41600	4	57470	5	73110	4	90560	4	101790
	benchmark(normal)	12	4710	2	-14450	2	-4960	4	-2450	4	2250	5	26800	5	31270
HO	BVAR(log)	3	42067	**3**	**107785**	3	133426	3	190180	3	232877	3	263668	3	250631
	benchmark(log)	**4**	**51778**	4	105827	**6**	**156500**	**6**	**204620**	**6**	**255146**	**4**	**310514**	**4**	**370726**
	BVAR(normal)	14	22109	3	74277	3	91812	3	148214	3	189164	3	246162	4	285449
	benchmark(normal)	6	-22554	7	-323	10	8560	1	15628	1	63995	2	88847	2	158634
HU	BVAR(log)	3	62513	6	68813	6	119910	6	169088	6	195002	6	241878	6	266692
	benchmark(log)	16	63731	**16**	**128268**	**16**	**193603**	**16**	**258010**	**12**	**321766**	**12**	**384350**	**12**	**448266**
	BVAR(normal)	**4**	**81942**	7	50946	7	90989	5	184720	5	260791	5	287759	5	358176
	benchmark(normal)	5	-35465	3	21193	12	-21420	3	40627	3	70892	3	154837	5	155123
W	BVAR(log)	15	5000	**11**	**38275**	15	47463	**17**	**68750**	17	70500	16	82750	14	93538
	benchmark(log)	**2**	**23525**	7	33450	7	47688	7	60950	7	76100	7	86800	7	102638
	BVAR(normal)	7	21375	7	34850	**7**	**47713**	7	61750	**7**	**76950**	**7**	**91025**	**7**	**104788**
	benchmark(normal)	7	10375	13	6650	2	18413	15	23175	4	50500	4	70750	3	63063

from this strategy by simply adding up all the profits (losses) at times $t \in \{1700, \ldots, T\}$ from such a strategy, where we compute the sums for all horizons $k \in \{1, \ldots, h = 15\}$.

The results are presented in Tables III and IV. The tables are organized similarly to Tables I, II. In the first column for each horizon is the index of forgetting factor κ_i with maximal gain potential and in the second column the trading potential in US\$.

IV. CONCLUSIONS AND FUTURE WORK

A. Conclusions

We have proposed a method for computing prediction of a normal and log-normal stochastic BVAR(p) process and applied it to U.S. commodity market data. We have compared the prediction power of such models to two benchmark models commonly used in todays financial practice. Due to the curse of dimensionality we have used only a BVAR(2) model to obtain the experimental results now to be presented.

As we can see from Tables I,II,III,IV the results of the comparison is highly market-dependent.

In case of median relative error (MERE), there are markets, such as the British Pound and the energy markets (CL,HO,HU), where the benchmark model performs better and we were not able to employ the information at hand to reach a better result with the BVAR(2) model of either normal or log-normal kind. On the other hand, there are markets, although fewer in number, where a linear structure of the BVAR kind seems to be present. In our selection, these markets are mainly Canadian Dollar, Feeder Cattle and Gold. In the rest of the

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:10, 2009

TABLE IV
TRADING POTENTIAL COMPUTED FOR THE HIGHEST LIKELIHOOD LOG-NORMAL AND NORMAL BVAR(2) MODEL AND LOG-NORMAL (33) AND NORMAL (35) BENCHMARK MODEL FOR HORIZONS 8, . . . , 14. THE POTENTIAL IS SHOWN IN US$, TRANSACTION COSTS AND LIMITED TRADING CAPITAL ARE NOT TAKEN INTO ACCOUNT.

| horizon | | 8 | | 9 | | 10 | | 11 | | 12 | | 13 | | 14 | |
mkt.	model	frg	GAINS	frg	GAINS	frg	GAINS	frg	GAINS	frg	GAINS	frg	GAINS	frg	GAINS
AD	BVAR(log)	3	308910	3	338740	3	362290	3	400340	3	403990	3	433220	3	479660
	benchmark(log)	**4**	**324530**	**4**	**361840**	**4**	**396130**	**4**	**429480**	**4**	**465610**	**4**	**511880**	**4**	**546160**
	BVAR(normal)	3	273450	2	320440	4	258570	18	112800	18	102750	18	107280	18	110620
	benchmark(normal)	1	159890	3	169500	3	230950	2	219740	2	270630	2	285260	3	378500
BP	BVAR(log)	6	11331	3	63194	1	158200	1	204087	1	241256	1	274994	1	327856
	benchmark(log)	**6**	**261094**	**3**	**289794**	**5**	**316500**	**5**	**348150**	**5**	**370031**	**5**	**396619**	**5**	**432356**
	BVAR(normal)	5	59894	5	66194	5	36363	1	-16800	4	124994	9	173844	4	105856
	benchmark(normal)	3	48506	3	79506	3	129700	3	117500	4	162681	4	186544	4	216994
CC	BVAR(log)	16	49860	18	54010	3	67280	**3**	**90550**	**3**	**100070**	**3**	**102760**	5	105580
	benchmark(log)	17	46220	17	52190	17	57980	17	63650	17	69470	21	75160	17	80760
	BVAR(normal)	9	46220	9	52190	9	57980	9	63650	9	69470	9	75120	9	80760
	benchmark(normal)	**5**	**54620**	**5**	**79590**	**5**	**90140**	4	41850	4	57770	4	67900	4	96480
CD	BVAR(log)	**12**	**278990**	**7**	**300180**	**15**	**339290**	**15**	**367700**	**15**	**404580**	**15**	**434040**	**18**	**466440**
	benchmark(log)	5	160850	5	177360	6	192830	5	215480	5	238300	5	256660	5	275200
	BVAR(normal)	8	224350	7	252360	7	284930	7	306360	7	342700	7	371740	7	406200
	benchmark(normal)	1	114770	1	123000	1	129130	1	132520	1	127520	1	131820	1	155860
CL	BVAR(log)	**5**	**393480**	**5**	**473520**	**5**	**535720**	**5**	**592150**	**5**	**611360**	**7**	**655530**	**6**	**700250**
	benchmark(log)	3	349860	3	375780	3	426520	3	461990	3	510580	3	562910	3	600650
	BVAR(normal)	19	115740	1	123820	12	157120	12	189390	12	178480	12	154890	12	187070
	benchmark(normal)	1	191140	1	214280	3	238200	3	298310	3	301900	3	327670	3	321610
CT	BVAR(log)	4	222130	4	238225	4	259605	4	293320	4	297530	4	309045	4	344085
	benchmark(log)	4	148410	4	179785	4	197195	4	215060	4	231390	4	261075	4	276165
	BVAR(normal)	**4**	**223640**	**4**	**246795**	**4**	**279925**	**4**	**298600**	**4**	**319340**	**4**	**347265**	**4**	**369055**
	benchmark(normal)	4	93550	4	98475	4	92515	4	97260	4	117380	4	99095	4	106705
FC	BVAR(log)	**17**	**162850**	21	175938	6	180475	6	207275	7	221363	7	247600	7	272075
	benchmark(log)	11	159700	**11**	**178963**	**11**	**198850**	**11**	**219450**	**11**	**239487**	**11**	**259250**	**11**	**279475**
	BVAR(normal)	11	159700	**11**	**178963**	**11**	**198850**	**11**	**219450**	**11**	**239487**	**11**	**259250**	**11**	**279475**
	benchmark(normal)	4	82950	4	74438	4	83300	4	114500	4	126538	2	142025	2	154125
GC	BVAR(log)	2	16610	3	-22570	1	-63550	1	-57690	7	-49970	1	-64860	2	-75770
	benchmark(log)	**5**	**147930**	**5**	**163830**	**5**	**186690**	**5**	**203630**	**5**	**224170**	**5**	**245980**	**5**	**266790**
	BVAR(normal)	5	111990	4	132710	5	145190	5	161770	5	178470	5	196660	5	215330
	benchmark(normal)	5	33990	5	63690	5	54610	5	90510	5	88510	4	101560	5	132170
HO	BVAR(log)	3	280720	3	302072	2	377198	2	497981	2	557075	2	657842	**2**	**739255**
	benchmark(log)	**4**	**412264**	**5**	**462084**	**5**	**519343**	**5**	**563930**	**5**	**621134**	**5**	**674533**	5	738440
	BVAR(normal)	3	346693	3	387374	3	292089	12	233138	20	191743	12	237707	20	290611
	benchmark(normal)	2	214368	3	258989	3	245024	3	326185	3	356727	3	376039	2	414233
HU	BVAR(log)	6	323463	6	346794	5	421294	5	464646	5	493374	5	534341	5	581385
	benchmark(log)	**12**	**513286**	**12**	**576752**	**12**	**640441**	**12**	**704155**	**12**	**769835**	12	834767	16	900182
	BVAR(normal)	5	442625	5	512736	5	584018	5	666112	5	743627	**5**	**835220**	**5**	**934168**
	benchmark(normal)	3	166492	5	255259	2	254671	2	316168	2	289019	2	369760	2	303328
W	BVAR(log)	14	106513	11	118513	21	127500	10	140800	10	159863	10	165900	10	174725
	benchmark(log)	7	116838	7	130563	7	140725	7	156450	7	172638	7	184950	7	199550
	BVAR(normal)	**7**	**124263**	**7**	**137463**	**7**	**154225**	**7**	**169725**	**7**	**187913**	**7**	**201400**	**7**	**218400**
	benchmark(normal)	4	65363	4	66913	3	88250	3	119875	3	125863	3	161150	3	163000

markets the presence of a BVAR structure is uncertain.

From the point of view of market potential the situation should be somewhat similar, although here we only measure the success of sign prediction, where in case of MERE we measure the success of prediction of the joint pair sign-amplitude. We see from Tables III,IV that the markets, where BVAR models are more suitable for sign prediction than their benchmark counterparts are again Canadian Dollar and Feeder Cattle, but now also Wheat and Cotton and at longer horizons Light Crude Oil.

We see that in certain markets, the BVAR(2) models were outperformed by the simple benchmark models, especially the one, where the distribution of prices conditional on the parameters is log-normal. Such an outcome is in agreement with the Efficient Market Hypothesis and can be a result of an efficient market – low information value of the data channels used, but there are other possibilities that should be a part of future research. The structure estimation procedure could not be performed for a model with forgetting $\kappa < 1$. That can lead to an overparameterized model.

At further horizons the difference between the point predictions of the competing models is lower. This effect is caused by the dominating effect of the constant term in the VAR process on prediction, when h is higher.

An interesting result in agreement with [7] is, that the ideal forgetting factor is often around 0.9975 (indices $i = 3, 4, 5, 6$) and is in the lower part of the range chosen for our experiments. In the graphs of error dependence on forgetting factor observed while performing the experiments, a quite robust minimum could be observed around this value. A slow shift

World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:3, No:10, 2009

Fig. 1. The picture shows that lower error does not necessarily mean higher gain. The absolute error for one-step ahead prediction $\bar{y}_{t+1}$ is greater than the one for $\hat{y}_{t+1}$, but the trader gains only if he buys. The figure clearly illustrates such situation for an absolute error, for MERE the ratios have to be considered.
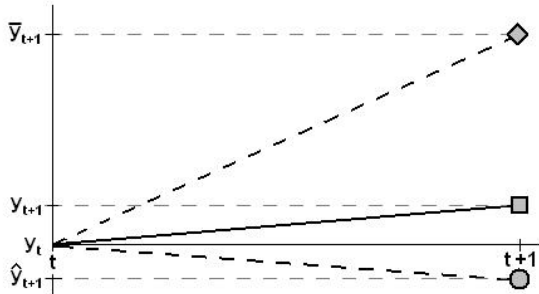
in optimal forgetting toward one was observed as h grew. In a few markets, an ideal forgetting was rather close to 1.

The gains presented in Tables III and IV are much higher than a real market gain. Even if we forget the transaction fees, the unlimited trading capital is not a realistic assumption. The result should serve as an illustration of success of the competing models.

B. Future Work

In the future, further research of the modeling of price evolution is needed. The structure estimation of a BVAR model should be questioned. One possibility for a correct estimation of structure is parallel computation of few most promising candidates for best BVAR models, where the maximum likelihood model would be chosen as the locally correct one. The original structure estimation procedure would be used only to select promising candidates (or possibly only to select the most important data channels), where the maximum likelihood model selected in the original procedure would be used as an upper bound for parameter matrix **A** regarding non-zero parameters of the model.

Once we have a reliable model to build upon a task at least as difficult as the modeling is the optimization of trading under limited capital and transaction costs. A solution to this problem can be very difficult, since the predictive random variables are not independent (In case of independent variables, the problem has been quite successfully solved, see for example [33],[34]) and analysis of the associated Bellman function (optimal cost to go) is difficult. Approximate solution has been proposed by a member of our project in the case of a flat utility function [35], but since we obtain the full predictive distribution of future price, we could perform a mean-risk, mean-variance or full utility function optimization for an adequate utility function of the agent.

The models should be tested for different frequencies of data from high frequency data gathered during the day to monthly data. Raw futures data for single expiration futures contracts should also be used for future experiments.

Finally, different models of price prediction should be tested in the future, where wavelet transform or neural network models seem as promising candidates.

Jan Šindelář received a MSc. degree in Mathematical Physics at the Faculty of Mathematics and Physics of Charles University in Prague in 2006. He currently studies Econometrics and Operational Research at the Department of Probability and Mathematical Statistics at the same faculty and works on his PhD. thesis at the Department of Adaptive Systems of Institute of Automation and Information Theory of the Czech Academy of Sciences under supervision of Dr. Miroslav Kárný.

REFERENCES

[1] E. Fama, "The behavior of stock market prices," *Journal of Business*, vol. 38, p. 34105, 1965.

[2] P. Samuelson, "Proof that properly anticipated prices fluctuate randomly," *Industrial Management Review*, vol. 6, pp. 44–49, 1965.

[3] E. Fama, "Efficient capital markets: A review of theory and empirical work," *Journal of Finance*, vol. 25, pp. 383–417, 1970.

[4] L. Bachelier, *Théorie de la speculation*. Paris: Gauthier-Villars, 1900, reprinted in P.H.Cootner: The Random Character of Stock Market Prices(1964).

[5] E. Fama, "Efficient capital markets: II," *International Journal of Finance*, vol. 46, pp. 1575–1617, 1991.

[6] C. Granger, "Forecasting stock market prices: Lessons for forecasters," *International Journal of Forecasting*, vol. 8, pp. 3–13, 1992.

[7] A. Timmermann and C. Granger, "Efficient market hypothesis and forecasting," *International Journal of Forecasting*, vol. 20, pp. 15–27, 2004.

[8] R. Litterman, "Forecasting with bayesian vector autoregressions – five years of experience," *Journal of Business & Economic Statistics*, vol. 4, pp. 25–38, 1986.

[9] S. Shreve and H. Soner, "Optimal investment and consumption with transaction costs," *The Annals of Applied Probability*, vol. 4, pp. 609–692, 1994.

[10] T. Yamamoto, "Asymptotic mean square prediction error for an autoregressive model with estimated coefficients," *Applied Statistics*, vol. 25, pp. 123–127, 1976.

[11] C. Ing, "Multistep prediction in autoregressive processes," *Econometric theory*, vol. 19, pp. 254–279, 2003.

[12] J. Šindelář, "Construction of multi-step ahead predictions in a normal bvar(p) model using monte carlo sampling," in *Proceedings of the 10th International PhD workshop on Systems and Control*, 2006.

[13] S. Shreve, *Stochastic Calculus for Finance II., Continuous-Time Models*. Springer, 2004.

[14] P. Billingsley, *Probability and Measure*, 3rd ed. Wiley, 1986.

[15] M. Kárný, J. Böhm, T. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař, *Optimized Bayesian Dynamic Advising, Theory and Algorithms*. Springer, 2005.

[16] T. Cipra, *Finančnï Ekonometrie*. EKOPRESS, 2008, in Czech.

[17] G. Box and G. Jenkins, *Time Series Analysis*. San Francisco, U.S.A.: Holden-Day, 1970.

[18] P. Brockwell and R. Davis, *Introduction to Time Series and Forecasting*. Springer, 1996.

[19] J. Anděl, *Základy matematické statistiky*. Matfyz-Press, 2005, in Czech.

[20] I. Nagy, L. Pavelková, E. Suzdaleva, J. Homolová, and M. Kárný, *Bayesian Decision Making*. Czech Academy of Sciences, 2005.

[21] S. Shreve, *Stochastic Calculus for Finance I., The Binomial Asset Pricing Model*. Springer, 2004.

[22] I. Karatzas and S. Shreve, *Methods of Mathematical Finance*. Springer, 1998.

[23] H. Fölmer and A. Schied, *Stochastic Finance, An Introduction in Discrete Time*. Walter de Gruyter, 2002.

[24] P. Billingsley, *Convergence of Probability Measures*, 2nd ed. Wiley, 1999.

[25] M. Kárný and R. Kulhavý, "Structure determination of regression-type models for adaptive prediction and control," in *Bayesian Analysis of Time Series and Dynamic Models*, J. C. Spall, Ed. Marcel Dekker, 1988, pp. 313–345.

[26] C. Robert, *The Bayesian Choice, From Decision Theoretic Foundations to Computational Implementation*. Springer, 2007.

[27] J. Ghosh, M. Delampady, and T. Samanta, *An Introduction to Bayesian Analysis, Theory and Methods*. Springer, 2006.

[28] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[29] A. Votava, "*Estimation of forgetting factor in the frame of dynamic decision making* Bachelor's thesis, Faculty of Mathematics and Physics, Charles University," Prague, 2009.

[30] S. Kotz, B. N., and N. Johnson, *Continuous Multivariate Distributions*, 2nd ed. Wiley, 2000.

[31] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 2nd ed. Athena Scientific, 2000, vol. 1.

[32] ——, *Dynamic Programming and Optimal Control*, 2nd ed. Athena Scientific, 2001, vol. 2.

[33] G. Constantinides, "Multiperiod consumption and investment behavior with convex transaction costs," *Management Science*, vol. 25, pp. 1127–1137, 1979.

[34] G. Genotte and A. Jung, "Investment strategies under transaction costs: The finite horizon case," *Management Science*, vol. 38, pp. 385–404, 1994.

[35] J. Zeman, "Futures trading:design of a strategy," accepted to ICORFE 2009 Conference, Venice, Italy, October 2009.