International Journal of Neural Systems, Vol. 8, No. 4 (August, 1997) 385–398
Special Issue on Data Mining in Finance

MODELING VOLATILITY USING STATE SPACE MODELS

JENS TIMMER*
Fakultät für Physik, Universität Freiburg,
Hermann-Herder Strasse 3, D-79104 Freiburg, Germany

ANDREAS S. WEIGEND†
Department of Information Systems, Leonard N. Stern School of Business,
New York University, 44 West Fourth Street, MEC 9-74, New York, NY 10012, USA

In time series problems, noise can be divided into two categories: dynamic noise which drives the process, and observational noise which is added in the measurement process, but does not influence future values of the system. In this framework, we show that empirical volatilities (the squared relative returns of prices) exhibit a significant amount of observational noise. To model and predict their time evolution adequately, we estimate state space models that explicitly include observational noise. We obtain relaxation times for shocks in the logarithm of volatility ranging from three weeks (for foreign exchange) to three to five months (for stock indices). In most cases, a two-dimensional hidden state is required to yield residuals that are consistent with white noise. We compare these results with ordinary autoregressive models (without a hidden state) and find that autoregressive models underestimate the relaxation times by about two orders of magnitude since they do not distinguish between observational and dynamic noise. This new interpretation of the dynamics of volatility in terms of relaxators in a state space model carries over to stochastic volatility models and to GARCH models, and is useful for several problems in finance, including risk management and the pricing of derivative securities.

Data sets used. Olsen & Associates high frequency DEM/USD foreign exchange rates (8 years). Nikkei 225 index (40 years). Dow Jones Industrial Average (25 years).

1. Introduction

Modeling and predicting the volatility of financial time series has become one of the central areas in finance and trading; examples range from pricing derivative securities to computing the risk of a portfolio. Volatility is usually predicted by using generalized autoregressive conditional heteroskedastic (GARCH) models; Bollerslev, Engle and Nelson (1995) guide through the GARCH literature, and Engle (1995) collects some of the key papers.

Here we present an alternative to GARCH that models the underlying dynamics using a state space model. This allows us to describe the hidden process in terms of variables natural for a dynamic system, such as decay times for shocks, its spectrum, and the dimensionality of the underlying process. Stochastic volatility models (see Shephard (1996) for a review) are a variant of the general state space approach presented here. They differ in that the mapping from the hidden variable to the observed variable is nonlinear. The interpretation developed in this article can also be helpful for understanding and characterizing stochastic volatility models.

This article is organized as follows: Section 2 discusses observational noise and dynamic noise and reviews intuitions and interpretations for linear systems, important for understanding the results in physical terms such as decay times of volatility

*E-mail: jeti@fdm.uni-freiburg.de, URL: http://www.fdm.uni-freiburg.de/~jeti
†E-mail: aweigend@stern.nyu.edu, URL: http://www.stern.nyu.edu/~aweigend

shocks. Section 3 defines and explains the formalism of state space models. Variations and interpretations that are typical in finance and in econometrics are given in Sec. 4. Section 5 describes the three data sets used for the empirical studies. The results are presented in Sec. 6, and the effect of ignoring existing observational noise on the model is discussed in Sec. 7. Section 8 summarizes the findings and discusses some of the applications of this approach for noisy time series in finance.

2. Some Background Concepts

2.1. *Observational noise and dynamic noise*

In time series modeling, one crucial question is whether or not observational noise is present in the data. Observational noise of a high level can pose a severe problem if it is not treated properly, leading to models that underestimate the functional relation between past and future values. A typical example of such *observational noise* is when an astronomer observes a star: fluctuations in the atmosphere, or a subway train passing by and shaking a telescope that points to the star, will not influence the dynamics of the star. In contrast, a noise component that does influence the dynamics of a system is called *dynamic noise.* For example, in an autoregressive process, the noise truly moves the state (sometimes also expressed as "the noise drives the system"), and subsequent values are derived from that moved state.

This article focuses on discrete time dynamics, typically modeled by difference equations or maps. The distinction between observational noise and dynamic noise is also important for continuous time dynamics, typically modeled by differential equations.

2.2. *Interpretations of linear systems*

To facilitate the interpretation of state space models (introduced in Sec. 3), we first review autoregressive processes without observational noise, and characterize them from several perspectives. A simple way of generating a time series is through an autoregressive (AR) process of order p, AR[p] (Yule, 1927; Priestley, 1981; Oppenheim and Schafer, 1989)

$$x(t) = \sum_{i=1}^{p} a_i x(t-i) + \varepsilon(t)\,, \tag{1}$$

where $\varepsilon(t)$ denotes an uncorrelated Gaussian distributed random variable with mean zero and constant variance σ^2, $\mathcal{N}(0, \sigma^2)$. Through the eyes of a physicist, such a process can be interpreted as a combination of *relaxators* and *damped oscillators* (Honerkamp, 1993). The simplest case is an AR[1] process

$$x(t) = ax(t-1) + \varepsilon(t)\,. \tag{2}$$

It can be characterized in the time domain as a relaxator by an exponentially decaying impulse response, proportional to $\exp(-t/\tau)$, with the relaxation time

$$\tau = -\frac{1}{\log\, a}\,. \tag{3}$$

After this time, the amplitude of an impulse will have decayed to 1/e or 37% of its initial value.

In the frequency domain, an AR process can be interpreted as a filter responding to white noise. The power spectrum of an AR[1] process drops off with

$$S(\omega) = \frac{\sigma^2}{|1 - ae^{-i\omega}|^2} = \frac{\sigma^2}{1 + a^2 - 2\cos\,\omega}\,. \tag{4}$$

For an AR[2] process, there are two qualitatively different cases, depending on the values of the parameters. We can always rewrite a univariate AR[2] model as a vector-valued AR[1] model using the transformation

$$\mathbf{A} = \begin{pmatrix} a_1 & a_2 \\ 1 & 0 \end{pmatrix}. \tag{5}$$

Its eigenvalues

$$\lambda_i = \frac{a_1}{2} \pm \sqrt{\frac{a_1^2}{4} + a_2} \tag{6}$$

characterize the behavior of the AR[2] process. If the eigenvalues are real ($a_1^2/4 + a_2 \geqslant 0$), the AR[2] process can be characterized as the superposition of two relaxators, and the spectrum drops off monotonically with increasing frequency. The corresponding decay constants are

$$\tau_i = -\frac{1}{\log\, \lambda_i} \quad (i = 1,\, 2)\,. \tag{7}$$

If the eigenvalues are complex, the AR[2] process describes a resonance, corresponding to a hump in the spectrum.[a] In both cases, the spectrum is given by

$$S(\omega) = \frac{\sigma^2}{|1 - a_1 e^{-i\omega} - a_2 e^{-2i\omega}|^2}\,. \tag{8}$$

By increasing the model order, an AR[3] process can combine a relaxator with an oscillator, and an AR[4] process can describe two oscillators, etc.

Despite the simplicity and multiple interpretability of AR models, not all processes in the world are linear autoregressive. Examples of generalizations without hidden states consist of including past q driving noise terms in the dynamics, yielding an autoregressive moving average ARMA[p, q] processes,[b] as well as including nonlinearities.[c] Here we extend autoregressive models in a different direction, by allowing for a hidden state.[d] The next section introduces the notation and gives the formalism of state space modeling.

3. Formalism of Linear State Space Models (LSSM)

In Eq. (1) the $x(t)$ served two roles: it was the variable that was observed, and it was the variable in which the dynamics was expressed. However, there are processes where the dynamics cannot be observed directly because it is masked by observational noise. Thus, no direct map exists from the observed data to the state. This requires the notion of a *hidden state*. In terms of notation, we keep the letter x as the variable that contains the dynamics, and use $y(t)$ for the observed variable. The state, characterized by the vector $\vec{x}(t)$, captures all the information needed to characterize the system at time t.

The key to state space modeling is to split the noise into two parts:

- dynamic noise $\vec{\varepsilon}(t)$ that drives the evolution of the hidden state, and
- observational noise $\eta(t)$ that is a non-explainable additive contribution to the measured $y(t)$.

These contributions have been discussed in intuitive terms in Sec. 2.1. Their formal role can be seen by observing how they enter the two equations that describe a linear state space model (LSSM):

$$\vec{x}(t) = \mathbf{A}\vec{x}(t-1) + \vec{\varepsilon}(t)\,, \quad \vec{\varepsilon}(t) \in \mathcal{N}(0, \mathbf{Q}) \tag{9}$$

$$y(t) = \mathbf{C}\vec{x}(t) + \eta(t)\,, \qquad \eta(t) \in \mathcal{N}(0, \mathbf{R})\,. \tag{10}$$

Equation (9) describes the dynamics. Equation (10) maps the dynamics to the observation and includes the observational noise $\eta(t)$.

As in the case of the observable linear autoregressive model, discussed in Sec. 2.2, describing the process via physical quantities can yield important insights. The spectrum of a LSSM is given by

$$S(\omega) = \mathbf{C}(\mathbf{1} - \mathbf{A}e^{-i\omega})^{-1}\mathbf{Q}((\mathbf{1} - \mathbf{A}e^{i\omega})^{-1})^T\mathbf{C}^T + R\,. \tag{11}$$

The superscript $(\cdot)^T$ denotes transposition. The spectra of AR processes, Eq. (8), are a subset of Eq. (11). Note that LSSM spectra include shapes that cannot be generated by AR processes. An important example of such a shape is a spectrum where for low frequencies the power drops similarly to an AR[1] process (see Eq. (4)), but for higher frequencies the power remains constant and does not continue to fall, as an AR model would require it to.

[a]For a damped oscillator (the case of complex eigenvalues), the parameters can be expressed through the characteristic period T and the relaxation time τ as

$$a_1 = 2\cos\left(\frac{2\pi}{T}\right)\exp(-1/\tau)$$
$$a_2 = -\exp(-2/\tau)\,.$$

[b]While for theoretical reasons ARMA[p, $p-1$] should be preferred to AR[p] processes for modeling of sampled continuous-time processes (Phadke and Wu, 1974), we find that in practice, differences in the results are small.

[c]The linear mapping given by Eq. (1) can be generalized to become a nonlinear mapping. Note that this is fully within the autoregressive framework and amounts to simple regression. Nonlinear approaches include radial basis functions (Casdagli, 1989; Moody and Darken 1989; Poggio and Girosi, 1990), neural networks (Lapedes and Farber, 1987; Weigend, Huberman and Rumelhart, 1990), and nonparametric kernel methods (Tjostheim and Auestad, 1994).

[d]This article explores the idea of a *continuous* hidden state, characterized by a scalar $x(t)$ or a vector $\vec{x}(t)$. The dynamics is expressed in terms of that unobserved state, and the state is subsequently mapped to the (conditional expectation of the) observed quantity. In contrast, Hidden Markov models (Rabiner, 1989; Fraser and Dimitriadis, 1994; Hamilton, 1994; Bengio and Frasconi, 1995; Shi and Weigend, 1997) assume the hidden state to be *discrete*: for each of these hidden states, there is an "agent" or "expert" (e.g. expressed as an autoregressive model) that generates the next data point. This introduces a second level of dynamics that is described by the transitions between the hidden states. This level of dynamics is absent in a pure autoregressive framework.

This can be interpreted as a low-frequency process whose spectral energy decreases as the frequency increases, until it is masked by a noise floor of a noise source with a flat spectrum. This low-frequency signal above a flat noise floor is the crucial *spectral signature* of a LSSM that cannot be emulated by an ordinary autoregressive model.

While parameter estimation in AR models is well established (e.g. by the Burg or the Durbin–Levinson algorithms), it is more cumbersome in the case of state space models. A standard approach uses the expectation maximization (EM) algorithm (Dempster, Laird and Rubin, 1977), a general iterative procedure for estimating parameters for models with hidden variables. In the E-step, it is assumed that the parameters of the model are known, and the hidden variables are estimated. In the M-step, the estimates of the hidden variables are taken literally and the values of the parameters are adjusted. This approach was first applied to LSSM by Shumway and Stoffer (1982).

Specifically for the case of the LSSM, the first E-step starts from the initial values of the parameters **A**, **Q**, **C**, R, and estimates the hidden dynamic variable $\vec{x}(t)$ using a Kalman filter. With the following definitions

- $z_{t|t'}$:= the predicted value of a quantity $z(t)$ based on the data $y(1), \ldots, y(t')$,
- $\Omega_{t|t'}$:= the covariance matrix of the estimated $\vec{x}(t)$, and
- $\Delta_{t|t'}$:= the variance of the prediction errors $(y(t) - y_{t|t'})$,

the equations for the Kalman filter are (Kalman, 1960; Gelb, 1974; Sorenson, 1985; Harvey, 1989; Aoki, 1990; Bomhoff, 1994; Hamilton, 1994, Mendel, 1995):

$$\Omega_{t|t-1} = \mathbf{A}\Omega_{t-1|t-1}\mathbf{A}^T + \mathbf{Q} \tag{12}$$

$$\Delta_{t|t-1} = \mathbf{C}\Omega_{t|t-1}\mathbf{C}^T + \mathrm{R} \tag{13}$$

$$K = \Omega_{t|t-1}\mathbf{C}^T\Delta_{t|t-1}^{-1} \tag{14}$$

$$\Omega_{t|t} = (1 - K\mathbf{C})\Omega_{t|t-1} \tag{15}$$

$$\vec{x}_{t|t-1} = \mathbf{A}\vec{x}_{t-1|t-1} \tag{16}$$

$$y_{t|t-1} = \mathbf{C}\vec{x}_{t|t-1} \tag{17}$$

$$\vec{x}_{t|t} = \vec{x}_{t|t-1} + K(y(t) - y_{t|t-1})\,. \tag{18}$$

There is a crucial difference between the first four equations and the last three. The first four equations, Eqs. (12)–(15), do not contain the data, they only describe relations between the parameters **A**, **Q**, **C**, R, Ω, Δ, and K. Their purpose is to find the value of K (the *Kalman gain*) that subsequently enters Eq. (18). K gives the appropriate weight to the added term originating in the error between the actual observation $y(t)$ and prediction $y_{t|t-1}$.

For true prediction, i.e. when $y(t)$ has not yet been observed, Eq. (16) has to be used for the unobserved state variable, and Eq. (17) for the observable. For model parameter estimation, on the other hand, the entire training data can be used, and an improved estimate of $\vec{x}_{t|N}$ can be obtained by the following three equations (Harvey, 1989):

$$\mathbf{B} = \Omega_{t|t}\mathbf{A}^T\Omega_{t+1|t}^{-1} \tag{19}$$

$$\vec{x}_{t|N} = \vec{x}_{t|t} + \mathbf{B}(\vec{x}_{t+1|N} - \mathbf{A}\vec{x}_{t|t}) \tag{20}$$

$$\Omega_{t|N} = \Omega_{t|t} + \mathbf{B}(\Omega_{t+1|N} - \Omega_{t+1|t})\mathbf{B}^T\,. \tag{21}$$

This concludes the E-step.

In the subsequent M-step, the parameters **A**, **Q**, **C**, R are updated; the derivation of the equations can be found in Honerkamp (1993). The iterative model fitting process ends when a convergence criterion is met. This concludes the description of how the model parameters are updated in the M-step.

Once a model has been built, its quality can be evaluated by several different criteria, including:

- **Predictive accuracy.** True out-of-sample predictions are generated using Eq. (17) on a test set that comes after the training period. The accuracy of the predictions can be compared to competing models by different evaluation criteria, such as squared errors or robust errors.
- **Whiteness of the prediction errors.** The model should explain all temporal correlations in the data: a perfect model takes the signal and turns it into white noise. Statistically, the question is whether we can reject (at a certain level of significance) the null hypothesis that the residuals are uncorrelated. Following Brockwell and Davis (1991), we use a Kolmogorov–Smirnov test to determine whether the periodogram of the residuals is consistent with a flat white noise spectrum.
- **Generating data from the model.** The distribution of a certain feature can be derived from

realizations of the model and compared with that feature as directly computed from the observed data.

For linear models, two additional criteria are useful:

- **Behavior in the time domain** (relaxation times). The parameters in linear models are related to relaxation times of the corresponding oscillators and relaxators. When the relaxation times are too small (of order of one time step), they usually only fit noise and this indicates that the order of the model is too large.
- **Behavior in the frequency domain** (spectrum). The spectrum of the linear process can be computed from the parameters of the estimated models through Eq. (11). Since the spectrum of the model should correspond to the expectation of the periodogram of the data, comparing the spectrum to the periodogram is another important qualitative criterion.

The suggestions listed here are just some of the useful general criteria that will be used in this article. For any specific problem, there are additional, more specific smoke alarms and sanity checks.

4. Applications of State Space Models to Finance

This section discusses two common applications of state space models in financial data, and compares them to our approach. For simplicity of notation, this discussion is written for the case of a scalar $x(t)$:

$$x(t) = ax(t-1) + \varepsilon(t) \tag{22}$$

$$y(t) = cx(t) + \eta(t)\,. \tag{23}$$

The dynamic equation, Eq. (22), is characterized by the single AR[1] coefficient a; $\varepsilon(t)$ is the dynamic noise that drives the dynamics. The observation equation, Eq. (23), maps the unobserved state $x(t)$ to the observed variable by scaling it with c. The added observational noise, $\eta(t)$, does not enter the dynamics.

4.1. *Smoothing*

The first approach decomposes the variance and results in a smoother series. It can be interpreted as a method for trend estimation. Here, parameter a is not estimated from the data to characterize the dynamics (as in our approach), but rather set to unity. Without loss of generality we can also set c to unity, yielding

$$x(t) - x(t-1) = \varepsilon(t) \tag{24}$$

$$y(t) = x(t) + \eta(t)\,. \tag{25}$$

Equation (24) interprets $\varepsilon(t)$ as the first difference of the series. Reducing the variance of $\varepsilon(t)$ by moving some of it onto $\eta(t)$ results in $x(t)$ as a smoothed version of $y(t)$. The variance of the original data $y(t)$ is thus decomposed into observational noise, $\eta(t)$, and a smoother signal, $x(t)$. This can be expressed in a Bayesian framework as a prior on the smoothness of the time series, as discussed by Kitagawa and Gersch (1996). Note that Eq. (24) resembles Brownian motion. However, it is not to be interpreted that way here, but as a smoothing constraint for the undisturbed signal instead. The smaller the $\varepsilon(t)$, the smoother the $x(t)$.

This smoothing approach is taken in most state space applications in finance. Bolland and Connor (1996) add to this approach a second non-constant part that is a linear function of the difference of the last two values of the state. This is effectively adding a constraint on the second differences (curvatures) of $x(t)$, in addition to the first differences. Moody and Wu (1996, 1997a, 1997b) use two variations of the simple smoothing model with $a = 1$, and use the term "true price" for the smoothed version of the observed prices.

4.2. *Variable parameter AR processes*

The second variation of the state space model also uses the state equation to model a slowly varying quantity as in Eq. (24), but the interpretation of the observation equation changes substantially. The constant c from Eq. (23) is replaced by $y(t-1)$. The equation then becomes

$$y(t) = x(t)y(t-1) + \eta(t)\,, \tag{26}$$

representing an AR[1] process. $x(t)$ has become an autoregressive parameter that slowly varies with time, and the former observational noise $\eta(t)$ now acts as dynamic noise (Wells, 1996), whereas we assume the parameters that characterize the system are constant over time.

4.3. *Modeling noisy linear systems*

The two cases above do not do justice to the dynamic structure of Eq. (22). In contrast, this article focuses on estimating the full hidden dynamics from the data. This allows us to characterize the process as a linear damped system of relaxators and oscillators, driven by dynamic noise, and observed through a veil of added observational noise.

In the econometric literature, *stochastic volatility models* have been used to describe the dynamic structure of returns, see Shephard (1996) for a recent review. In the notation of the present article, a stochastic volatility model can be expressed as

$$x(t) = a_0 + a_1 x(t-1) + \varepsilon(t) \tag{27}$$

$$y(t) = \eta(t) \exp(x(t)) \,. \tag{28}$$

The idea behind using $\exp(x(t))$ is to model the skewed distribution of squared returns found for the empirical data. Parameter estimation in this model is cumbersome due to the log-normal distribution of $\exp(x(t))$. It is usually based on the generalized method of moments, quasi-likelihood estimation or Markov chain Monte Carlo methods. In contrast to stochastic volatility models, we apply a static transformation to the data that will be introduced in the next section in order to make the distribution of squared return approximately normal. This allows us to use the standard maximum likelihood framework for the parameter estimation.

5. Data

This article reports results on the following data sets:

- High frequency DEM/USD foreign exchange rates.[e] We began with eight years of data (through June 29, 1995) spaced apart 30 minutes in ϑ-time (Theta-time). We dropped all points with missing values, and then took every fourth of the remaining points for our analysis, effectively downsampling to two hours in ϑ-time.[f] ϑ-time removes daily and weekly seasonality: times of day with a high mean volatility are expanded, and times of day and weekends with low volatility are contracted (Dacorogna, Gauvreau, Müller, Olsen and Pictet, 1996).
- Daily stock indices. We use two stock indices:
  — Nikkei 225 index (40 years of daily data, through October 15, 1996, 12 288 points total),[g]
  — Dow Jones Industrial Average (25 years of daily data, through October 16, 1987, 6252 points total).[h]

The top panel of Fig. 1 graphs the level of DEM/USD for the first half of 1995. Its periodogram, shown in the left panel of Fig. 2, drops to first approximation as the spectrum of a random walk whose $1/f^2$ line is also indicated. (f denotes the frequency.) The signature of observational noise — a noise floor masking the signal at high frequencies — is absent: the periodogram continues to drop to the highest time scale. The result is that *price levels* $p(t)$ of financial instruments do not exhibit significant observational noise; all the "noise" on prices is dynamic, i.e. it re-enters the dynamic equation.

The central panel of Fig. 1 shows the difference of the logarithm of the price levels

$$\log\, p(t) - \log\, p(t-1) = \log \frac{p(t)}{p(t-1)} \approx \frac{p(t)-p(t-1)}{p(t-1)} \,. \tag{29}$$

This quantity can be interpreted as the logarithm of the geometric growths, i.e. as the logarithm of the ratio of the prices. It is known in finance as continuously compounded returns. Using the fact that the logarithm Taylor expands around 1 as $\log \varepsilon \approx 1 + \varepsilon$, it can also be interpreted as the returns normalized by the levels, i.e. the *relative returns*. Note in the central panel of Fig. 1 that the width of the "band" varies over time; regimes with larger shocks (positive or negative) alternate with regimes with smaller widths. The corresponding periodogram of the relative returns is shown in the right panel of Fig. 2.

[e] We thank Michel Dacorogna (Olsen & Associates, Zurich) for the high frequency DEM/USD exchange rate data.

[f] Whether half-hour or two-hour intervals in ϑ-time are taken does not change the results reported here, since the time scale of the dynamics that we find is two orders of magnitude slower than the sampling interval. As an example for larger changes in the sampling interval, Brown (1990) shows that the estimated (unconditional) volatility decreases by 13% as the sampling interval of S&P 500 Index futures is changed from one minute to one hour.

[g] We thank Morio Yoda (Nikko Securities, Tokyo) for the Nikkei 225 stock index data.

[h] The Dow Jones Industrial Average data set is described in LeBaron and Weigend (1997) and available through `http://www.stern.nyu.edu/~aweigend/Research.`

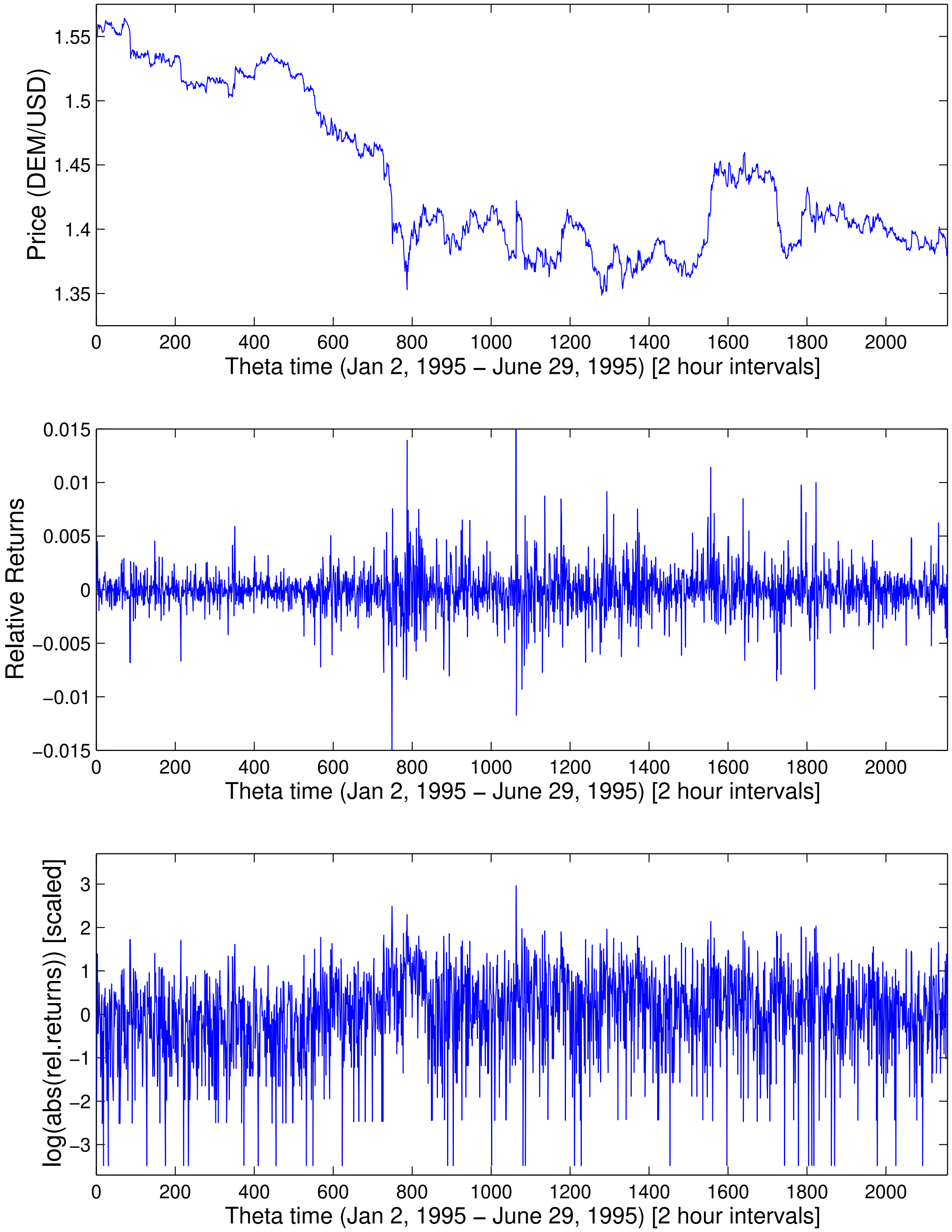

Fig. 1. This figure displays a six-month window of the high frequency foreign exchange data, sampled at two hour intervals in ϑ-time. The top panel shows the prices, the middle panel shows the relative returns, and the bottom panel shows the series used in our analysis, i.e. after applying the logarithm and scaling it to zero mean and unit variance.

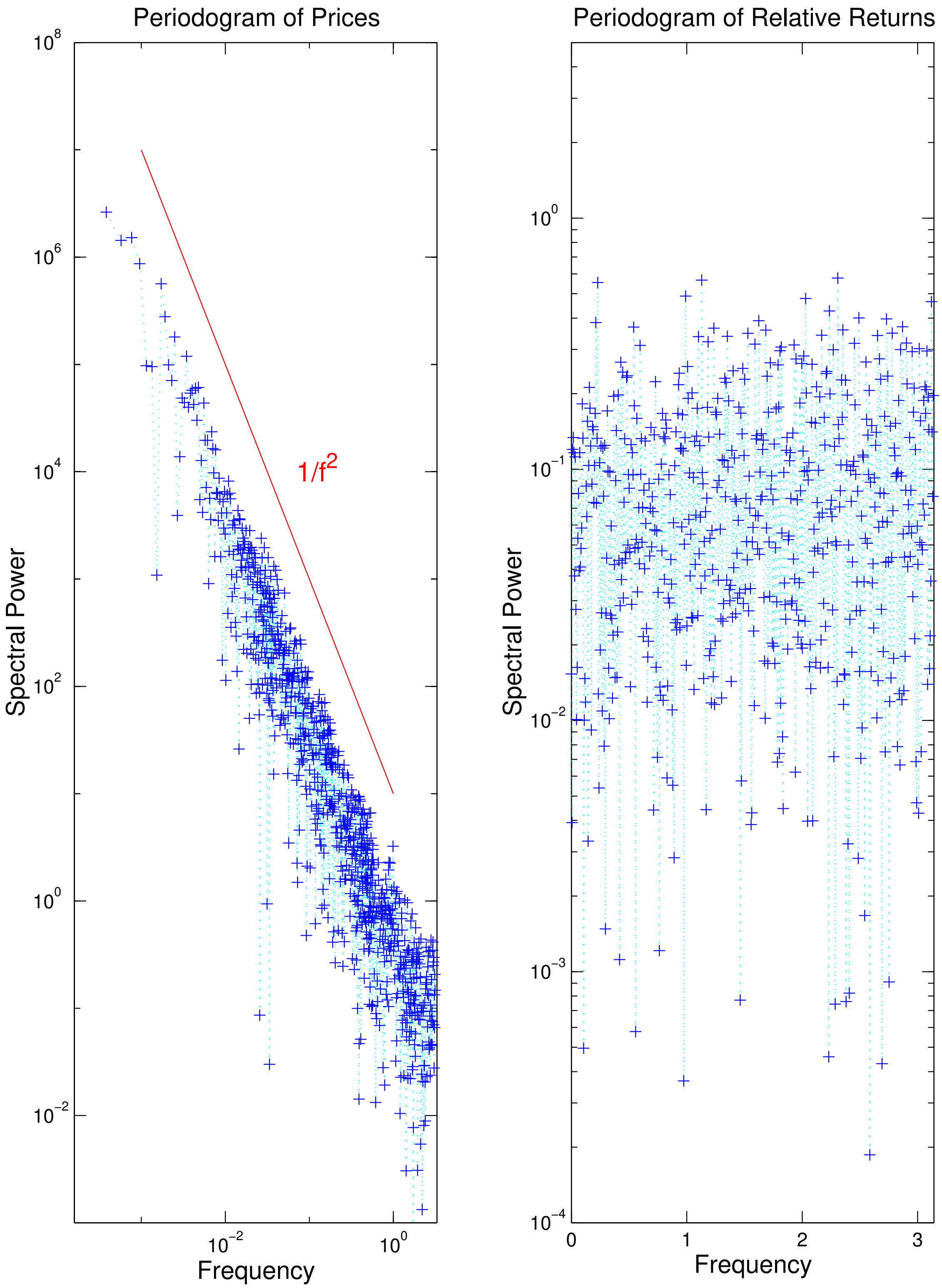

Fig. 2. Periodogram of the DEM/USD prices (left), and of the relative returns (right). Expressed in 1/time, the leftmost points correspond to 1/(8 years), 1/(4 years), 1/(2.6 years), 1/(2 years). To guide the eye, we also plotted the $1/f^2$ drop in spectral power of a random walk over six orders of magnitude. The periodogram of the returns on the right hand side is essentially flat. Neither the prices nor the returns indicate the presence of observational noise, in contrast to Fig. 3.

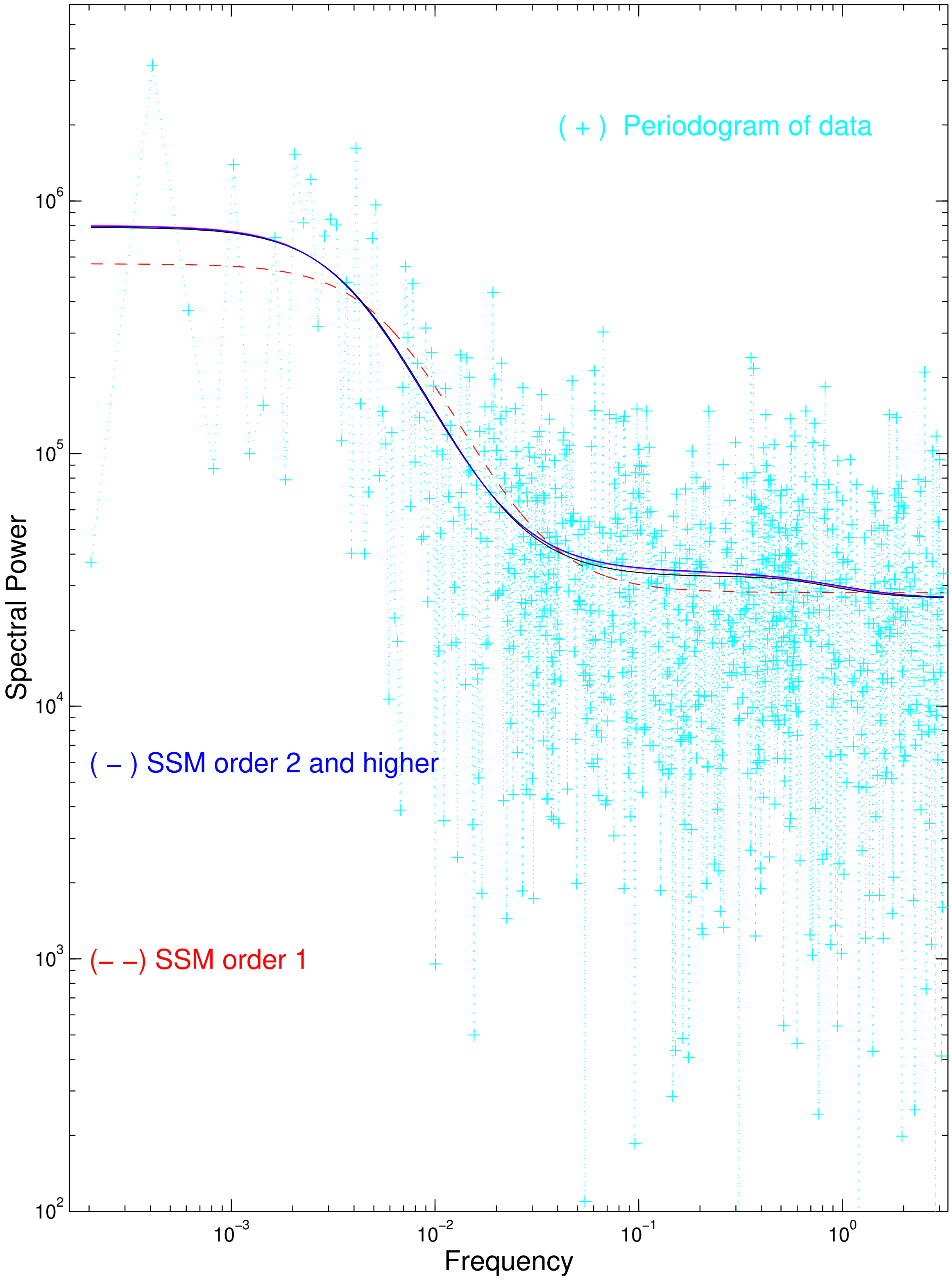

Fig. 3. Periodogram ("+") of the DEM/USD exchange rates, and spectra of the estimated state space models (SSM) of order one (dashed line) and higher orders (solid lines).

Note that it is essentially flat: the subsequent returns on the two-hour time scale in ϑ-time appear to be (linearly) uncorrelated.

To exploit this observed structure in the absolute values of the relative returns, we square the relative returns, i.e. ignoring their signs. The distribution of the squared returns is very skewed. To make it less skewed, we take their logarithm,

$$y(t) = \log\left[\log\frac{p(t)}{p(t-1)}\right]^2 . \tag{30}$$

The *logarithm of the squared relative returns*, $y(t)$, is shown for the DEM/USD data in the bottom panel of Fig. 1.

The squared relative returns can be interpreted as independent realizations of a random variable with a slowly changing mean. If the relative returns $\log p(t)/p(t-1)$ were normally distributed with unit variance, their squares would follow a χ_1^2 distribution. The variance of this χ^2 distribution is twice its mean, implying that the realizations are very noisy indeed! This is the source of the observational noise for volatility. On empirical data, it is well known that the relative returns $\log p(t)/p(t-1)$ are not normally distributed, but have fatter tails. However, the spirit of the explanation for the observational noise still applies; see also Diebold and Lopez (1995).

Figure 3 shows this effect. The periodogram of the data contains most of its power at low frequencies. Subsequently, as the frequency increases, it begins to drop. Finally, it flattens out as the signal gets masked by this "observational noise", stemming from the noisy realizations of the slowly changing means of the squared returns. Note the absence of a daily or weekly peak in this periodogram: while present for data in chronological time, it has been successfully removed by Olsen's projection of the data onto ϑ-time. This periodogram is similar to figures in Schnidrig and Würtz (1995) and in Andersen and Bollerslev (1997). However, neither of these papers interpret the signature as evidence for observational noise, nor do they use a state space model to explain the data.

The key features of the periodogram — a drop over many orders of magnitude for price levels, a roughly constant level for returns, and a low frequency signal disappearing into observational noise at higher frequencies for squared returns — hold for all the financial data sets we analyzed, including six other currencies on different time scales, as well as several stock indices. The next section gives detailed results for DEM/USD and Nikkei 225, as well as brief results for the Dow Jones industrial index.

Table 1. Results for the volatilities of the DEM/USD exchange rates. While linear state space models (LSSM) of order two and above fit the data well, ordinary AR models cannot explain the structure of the data.

Models DEM/USD	τ (decay times) 1 step = 2 hours	AR coefficients	Prob white noise	E_{NMS}
LSSM(1)	156	0.994	0	0.960
LSSM(2)	240, 1.09	$\begin{pmatrix} 0.996 & 0 \\ 0 & 0.399 \end{pmatrix}$	0.72	0.957
LSSM(3)	236 [$T = 17$ $\tau = 1.6$]	$\begin{pmatrix} 0.966 & 0 & 0 \\ 0 & 0.507 & -0.188 \\ 0 & 0.188 & 0.507 \end{pmatrix}$	0.57	0.957
LSSM(4)	243, 1.1 [$T = 8.5$ $\tau = 10$]	$\begin{pmatrix} 0.996 & 0 & 0 & 0 \\ 0 & 0.411 & 0 & 0 \\ 0 & 0 & 0.666 & -0.612 \\ 0 & 0 & 0.612 & 0.666 \end{pmatrix}$	0.70	0.957
AR(1)	0.45	0.107	0	0.988
AR(2)	0.85, 0.64	0.100 0.066	0	0.984
AR(3)	1.25 [$T = 6.5$ $\tau = 0.86$]	0.097 0.062 0.044	2e-6	0.982
AR(4)	1.8, 1.2 [$T = 4.1$ $\tau = 1.2$]	0.095 0.058 0.039 0.049	7e-4	0.979

6. Results

Table 1 summarizes the results for the high frequency DEM/USD data, comparing linear state space models with ordinary AR models. The linear state space models differ crucially from the AR models in the decay times τ: while the decay times of the state space models are significant, they are negligible for the AR models where the processes typically decay within one time step. Since the state space model is fitted to $y(t)$ as defined in Eq. (30), the decay times characterize when the logarithm of the squared relative returns has decayed to 37% of its initial value.

For first order models describing a single relaxator, there is a huge difference in decay time between 156 time steps for the LSSM in contrast to insignificant 0.45 time steps for the AR process. The eigenvalues of the second order models, given by Eq. (6), turn out to be real; the process thus corresponds to the superposition of two relaxators. The slower one of the two relaxators settles to around 240 of the 2-hour steps and corresponds to 20 days, whereas the slower AR relaxator still decays in a single time step. Using third and fourth order, oscillators emerge whose resonance frequencies $1/T$ correspond to about one day. They might indicate a tiny amount of periodicity left after the transformation of the raw data to ϑ-time, but they do not contribute significantly to the dynamics since their relaxation times are of the order of a few time steps only.

The decay constants presented here are defined for the logarithm of the squared relative returns. Nonlinear transformations do not allow for an amplitude-independent interpretations of decay times in general. However, fitting state space models directly to the absolute or squared relative returns (without taking the logarithm) yields similar decay constants. This implies that our characterization also hold for stochastic volatility models.

The fourth column in Table 1 shows that the residuals of the state space model of order one are not consistent with white noise, implying that a first order LSSM does not describe the data adequately. However, all higher order LSSMs produce residuals consistent with white noise at a significance level of 0.05 for the Kolmogorov–Smirnov test on the whiteness of the residuals (Brockwell and Davis, 1991). None of the residuals of the AR models are consistent with white noise. This is another indication that AR models are not an adequate model class for volatility.

The last column gives the normalized mean squared error, E_{NMS}, between the observed $y(t)$ and the predictions obtained via Eq. (17). Whereas for LSSM, the error drops quickly to a constant level of 0.957 at order 2, it decreases for AR models at a

Table 2. Results for the volatilities of the Nikkei 225 stock index. While linear state space models of order two and above fit the data well, ordinary AR models cannot explain the structure of the data.

Models Nikkei225	τ (decay times) 1 time step = 1 day	Prob of white noise	E_{NMS}
LSSM(1)	63.1	0.004	0.906
LSSM(2)	81.8, 1.45	0.56	0.905
LSSM(3)	81.2 [$T = 8.7 \quad \tau = 6.9$]	0.64	0.905
LSSM(4)	81.7, 1.46 [$T = 8.4 \quad \tau = 10$]	0.57	0.905
AR(1)	0.54	0	0.975
AR(2)	1.20, 0.82	0	0.959
AR(3)	1.85 [$T = 6.6 \quad \tau = 1.05$]	7e-7	0.951
AR(4)	2.93, 1.59 [$T = 4.15 \quad \tau = 1.61$]	0.002	0.940

much slower rate, and also remains at a higher level. In an AR(10) model, for example, E_{NMS} takes the value of 0.973, still significantly above the value of the second order LSSM.

We now turn to the power spectra. The curves in Fig. 3 are the power spectra of the state space models.[i] They are computed using Eq. (11). There is a clear difference between the first order spectrum and the higher order spectra. The higher orders ($\geqslant 2$) are very similar, indicating that the second order state space model is indeed sufficient. The spectra of the state space models correspond well to the periodogram of the data. Note that the spectra are not obtained by some direct smoothing of the periodogram in frequency space, but are the spectra of the state space models which were fitted in the time domain.

The results for the second data set, the logarithm of absolute values of the relative changes of the daily Nikkei 225 level, are summarized in Table 2. The key point is the large decay time of about 3.5 months, revealed by the state space models of order two and above, as well as the failure of AR models, very similar to the DEM/USD data set discussed.

The third data set, the logarithm of absolute values of the relative changes of the daily Dow Jones Industrial Index, reveals a decay time of 117 days or about five months. In that case, a one-dimensional hidden state already generates residuals that are consistent with white noise. As in the other two examples, no ordinary AR model in the observed variable explains the data. This effect will be clarified in the next section.

7. Ignoring Observational Noise

The failure of AR models shown in the previous section is a consequence of the observational noise that is present in the volatility data. Whereas linear state space models include the observational noise explicitly in the model, autoregressive models assume that the data is free from observational noise. We use a simple first order process to demonstrate the consequences of ignoring observational noise on the autoregressive parameter.

In an AR[1] model, $x(t) = ax(t-1) + \varepsilon(t)$, the parameter a can be estimated without bias as

$$\widehat{a} = \frac{\sum x(t-1)x(t)}{\sum x(t-1)x(t-1)} \,. \qquad (31)$$

If, however, the dynamics is covered by observational noise

$$y(t) = x(t) + \eta(t)\,, \qquad \eta \sim \mathcal{N}(0, \mathrm{R})\,, \qquad (32)$$

the expected value (denoted by $\langle\cdot\rangle$) of $\hat{a}$, estimated in analogy to Eq. (31) from $y(t)$, now becomes

$$\langle\hat{a}\rangle = \frac{\langle y(t-1) \quad y(t)\rangle}{\langle y(t-1) \quad y(t-1)\rangle} = \frac{a}{1+\mathrm{R}/\langle x(t)^2\rangle} \,. \qquad (33)$$

Thus, the larger the variance R of the observational noise, the worse the parameter a will be underestimated. This effect is known from linear regression as the problem of *errors-in-variables* (Fuller, 1987). It was first mentioned in time series context by Kostelich (1992), see also König and Timmer (1997). The underestimation of the functional relation between past and present values carries over to more general models, including nonlinear models (Carroll, Ruppert and Stefanski, 1995, Weigend, Zimmermann and Neuneier, 1996).

8. Summary and Applications

This article showed the important distinction between observational and dynamic noise. When observational noise is present, an autoregressive approach cannot model the data adequately — a state space approach is needed to capture the hidden dynamics. In finance, neither prices nor returns tend to have observational noise. However, volatilities do exhibit signature of observational noise in the periodogram: for low frequencies, there is structure above the noise floor of observational noise.

We showed on three representative financial data sets that a linear state space model with full dynamics can describe volatilities well. We also showed that the resulting models can be nicely interpreted, both from the perspective of physics as a superposition of two simple relaxators, and from the perspective of finance as volatility clustering with a decay time of about three weeks (for DEM/USD), 3.5 months (for Nikkei 225), and five months (for Dow Jones Industrial Average). These results are in strong contrast to AR models that ignore observational noise and consequently have a bias toward too small coefficients, as shown in Sec. 7. The more promising modeling approach using state space models over AR models for volatility suggests several applications in financial markets, including

[i] The spectra and the periodogram are normalized. For the lowest 200 frequencies, all periodogram points are plotted. Above this frequency, they are logarithmically thinned out for the sole reason to keep the files reasonably small for the on-line version. The visual impression in the printed version does not change.

- **Estimating risk.** Knowing the evolution of the volatility is important for determining the risk associated with a position on a financial instrument: the volatility can be interpreted as the conditional standard deviation of the returns.
- **Pricing derivative securities.** Using financial theory, discrepancies between the predicted volatility and the implied volatility can be translated into mispricings, which can in turn be exploited in trading.
- **Information for regime switching models.** The predicted volatility can be an important input for trading models based on the "gated experts" architecture (Weigend, Mangeas and Srivastava, 1995). In this case, the hidden state is offered as an additional input to the gate to help determine the current region.

In summary, we discussed the signature of observational noise in the frequency domain and showed on three data sets that volatilities exhibit that signature, but not the prices or returns. We showed that allowing for a hidden process with two or more degrees of freedom, and modeling the full dynamics of this process, gives interpretable results yielding residuals consistent with white noise. We are currently evaluating on several time horizons the performance for true volatility predictions of state space models in comparison to an approach using historic data (Figlewski, 1994), to GARCH (Bollerslev *et al.*, 1995), and to stochastic volatility models (Shephard, 1996).

Acknowledgments

Jens Timmer gratefully acknowledges the hospitality of the Information Systems Department of NYU's Leonard N. Stern School of Business. Andreas Weigend thanks Joel Hasbrouck and Steve Figlewski for their comments, as well as the participants of the NBER/NSF Forecasting Seminar for discussions at the NBER 1997 Summer Institute, and acknowledges support from the National Science Foundation (ECS-9309786) and the Air Force Office of Scientific Research (F49620-96-1-0240).

References

T. G. Andersen and T. Bollerslev 1997, "Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns," *J. Finance* **52**, 975–1005.

M. Aoki 1990, *State Space Modeling of Time Series* (Springer-Verlag, New York).

Y. Bengio and P. Frasconi 1995, "An input output HMM architecture," in *Advances in Neural Information Processing Systems 7 (NIPS*94)*, eds. G. Tesauro, D. S. Touretzky and T. K. Leen (MIT Press, Cambridge, MA), pp. 427–434.

P. J. Bolland and J. T. Connor 1996, "Identification of FX arbitrage opportunities with a non-linear multivariate Kalman filter, in *Neural Networks in Financial Engineering: Proceedings of the Third International Conference on Neural Networks in the Capital Markets (NNCM-95)*, eds. A.-P. N. Refenes, Y. Abu-Mostafa, J. Moody and A. Weigend (World Scientific, Singapore), pp. 122–134.

T. Bollerslev, R. F. Engle and D. B. Nelson, 1995, "ARCH models," in *Handbook of Econometrics*, Vol. 4, eds. R. F. Engle and D. L. McFadden (North-Holland, NY), Chapter 49.

E. J. Bomhoff 1994, *Financial Forecasting for Business and Economics* (Academic Press, London).

P. J. Brockwell and R. A. Davis 1991, *Time Series: Theory and Methods* (Springer-Verlag, New York).

S. Brown 1990, "Estimating volatility," in *Financial Options: From Theory to Practice*, eds. S. Figlewski, W. L. Silber and M. G. Subrahmanyam (Business One Irwin, Homewood, IL), pp. 516–537.

R. Carroll, D. Ruppert and L. Stefanski 1995, *Measurement Error in Nonlinear Models* (Chapman and Hall, London).

M. Casdagli 1989, "Nonlinear prediction of chaotic time series," *Physica* **D35**, 335–356.

M. M. Dacorogna, C. L. Gauvreau, U. A., Müller, R. B. Olsen and O. V. Pictet 1996, "Changing time scale for short-term forecasting in financial markets," *J. Forecasting* **15**, 203–227.

A. Dempster, N. Laird and D. Rubin 1977, "Maximum likelihood from incomplete data via EM algorithm," *J. Roy. Stat. Soc.* **39**, 1–38.

F. X. Diebold and J. A. Lopez 1995, "Modeling volatility dynamics," in *Macroeconometrics: Developments, Tensions, and Prospects*, ed. K. Hoover (Kluwer Academic Press, Boston), pp. 427–472.

R. F. Engle, ed. 1995, *ARCH: Selected Readings* (Oxford University Press, Oxford, UK).

S. Figlewski 1997, "Forecasting volatility," *J. Futures Markets* **6**, 1–88.

A. M. Fraser and A. Dimitriadis 1994, "Forecasting probability densities by using hidden Markov models, in *Time Series Prediction: Forecasting the Future and Understanding the Past*, eds. A. S. Weigend and N. A. Gershenfeld (Addison-Wesley, Reading, MA), pp. 265–282.

W. A. Fuller 1987 *Measurement Error Models* (John Wiley, New York).

A. Gelb 1974, *Applied Optimal Estimation* (MIT Press, Cambridge, MA).

J. D. Hamilton 1994, *Time Series Analysis* (Princeton University Press, Princeton, NJ).

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

J. Honerkamp 1993, *Stochastic Dynamical Systems* (VCH, New York).

R. E. Kalman 1960, "A new approach to linear filtering and prediction problems," *Trans. ASME J. Basic Eng. Series* **D82**, 35–45.

G. Kitagawa and W. Gersch 1996, *Smoothness Priors Analysis of Time Series* (Springer-Verlag, New York).

M. König and J. Timmer 1997, "Analyzing X-ray variability by linear state space models," *Astron. Astrophys. Suppl. Ser.* **124**, 589–596.

E. Kostelich 1992, "Problems in estimating dynamics from data," *Physica* **D58**, 138.

A. Lapedes and R. Farber 1987, "Nonlinear signal processing using neural networks," Working Paper LA-UR-87-2662, Los Alamos National Laboratory, Los Alamos, NM.

B. LeBaron and A. S. Weigend 1998, "A bootstrap evaluation of the effect of data splitting on financial time series," *IEEE Trans. Neural Networks* **9**, 213–220.

J. M. Mendel 1995, *Lessons in Estimation Theory for Signal Processing, Communications, and Control* (Prentice Hall, New Jersey).

J. Moody and C. Darken 1989, "Fast learning in networks of locally-tuned processing units," *Neural Comput.* **1**, 281–294.

J. E. Moody and L. Wu 1996, "What is the "true price"? — State space models for high frequency financial data," *Progress in Neural Information Processing (ICONIP'96)* (Springer, Berlin), pp. 697–704.

J. E. Moody and L. Wu 1997a, "What is the "true price"? — State space models for high frequency FX data, in *Decision Technologies for Financial Engineering, Proc. Fourth Int. Conf. Neural Networks in the Capital Markets (NNCM-96)*, eds. A. S. Weigend, Y. S. Abu-Mostafa and A.-P. N. Refenes (World Scientific, Singapore), pp. 346–358.

J. E. Moody and L. Wu 1997b, "What is the "true price"? — State space models for high frequency FX data," *Proc. IEEE/IAFE 1997 Conf. Computational Intelligence for Financial Engineering (CIFEr)*, IEEE Service Center, Piscataway, NJ, pp. 150–156.

A. V. Oppenheim and R. W. Schafer 1989, *Discrete-Time Signal Processing* (Prentice Hall, Englewood Cliffs, NJ).

M. Phadke and S. Wu 1974, "Modeling of continuous stochastic processes from discrete observations with application to sunspots data," *J. Am. Stat. Ass.* **69**, 325–329.

T. Poggio and F. Girosi 1990, "Networks for approximation and learning," *Proc. IEEE* **78**, 1481–1497.

M. Priestley 1981, *Spectral Analysis and Time Series* (Academic Press, London).

L. R. Rabiner 1989, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE* **77**, 257–286.

R. Schnidrig and D. Würtz 1995, "Investigation of the volatility and autocorrelation function of the USD/DEM exchange rate on operational time scales," *Proc. First Int. Conf. High Frequency Data in Finance (HFDF–I)*, Vol. 3, Zurich.

N. Shephard 1996, "Statistical aspects of ARCH and stochastic volatility," in *Time Series Models In Econometrics, Finance and Other Fields*, eds. D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen (Chapman and Hall, London), pp. 1–67.

S. Shi and A. S. Weigend 1997, "Taking time seriously: Hidden Markov experts applied to financial engineering," *Proc. IEEE/IAFE 1997 Conf. Computational Intelligence for Financial Engineering (CIFEr)*, IEEE Service Center, Piscataway, NJ, pp. 244–252.

R. Shumway and D. Stoffer 1982, "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Ser. Anal.* **3**, 253–264.

H. W. Sorenson 1985, *Kalman Filtering: Theory and Application* (IEEE Press).

D. Tjostheim and B. Auestad 1994, "Nonparametric identification of nonlinear time series: Selecting significant lags," *J. Am. Stat. Ass.* **89**, 1410–1419.

A. S. Weigend, B. A. Huberman and D. E. Rumelhart 1990, "Predicting the future: A connectionist approach," *Int. J. Neural Systems* **1**, 193–209.

A. S. Weigend, M. Mangeas and A. N. Srivastava 1995, "Nonlinear gated experts for time series: Discovering regimes and avoiding overfitting," *Int. J. Neural Systems* **6**, 373–399.

A. S. Weigend, H. G. Zimmermann and R. Neuneier 1996, "Clearning," in *Neural Networks in Financial Engineering, Proc. Third Int. Conf. Neural Networks in the Capital Markets (NNCM-95)*, eds. A.-P. N. Refenes, Y. Abu-Mostafa, J. Moody and A. Weigend (World Scientific, Singapore), pp. 511–522.

C. Wells 1996, *The Kalman Filter in Finance* (Kluwer Academic Publishers, Dordrecht).

G. Yule 1927, "On a method of investigating periodicity in disturbed series with special reference to Wolfer's sunspot numbers," *Phil. Trans. Roy. Soc. London* **A226**, 267–298.