IWQW

Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung

Diskussionspapier
Discussion Papers

No. 12/2009

Constructing a Quasilinear Moving Average Using the Scaling Function

Stefan Schlüter
University of Erlangen-Nuremberg

Constructing a Quasilinear Moving Average Using the Scaling Function

STEPHAN SCHLÜTER

Department of Statistics and Econometrics, University of Erlangen-Nürnberg, Lange Gasse 20,

D-90403 Nürnberg

Email: stephan.schlueter@wiso.uni-erlangen.de

ABSTRACT

The scaling function ϕ from multiresolution analysis can be used to constuct a smoothing tool in the context of time series analysis. We give a ϕ-based time series smoothing function for which we show the properties of a quasilinear moving average. Furthermore, we discuss its features and especially derive the distributional properties of our ϕ-based quasilinear moving average given some simple underlying stochastic processes. Eventually we compare it to existing smoothing methods in order to motivate its application.

Keywords: Scaling function; Quasilinear moving average; Influence function.

1   Introduction

There are different ways of approaching the scaling function, but it is generally considered to be a part of multiresolution analysis (see Percival, 2000), i.e. the analysis of a stochastic process on different frequency layers. Therefore, the process is split up into a certain number of frequency components which give the amount of total oszillation explained by the respective frequency. In signal processing, where this concept was developed first, the Fourier transform was an adequate tool for identifying regular patterns. However, there are various fields of application, e.g. financial time series analysis, where patterns with time-varying intensity and period (e.g. yearly temperature-dependent oszillations) do occure. The Fourier transform lacks the flexibility to capture that. A generalisation of Fourier transform is the wavelet transform which is a mapping from time space into time-frequency space that is – in contrary to the latter – capable of identifying regular patterns

with time-varying intensity and frequency. In addition to that, it is possible to aggregate the time-dependent influence of all frequencies smaller than a specific one. This is done based on the scaling function ϕ, also called father wavelet (see Mallat, 2003).

Within this article we use the scaling function to construct a smoothing tool by excluding high-frequency patterns. However, when doing this, several questions arise, e.g.: What properties does this smoothing function have? Can we derive the specific distributional properties given a known underlying process? How robust is our smoothing function, i.e. how big is the influence of a temporary shock? Within this paper we answer these questions. In this context we compare the ϕ-based quasilinear moving average to existing smoothing methods in order to justify its application.

The paper is structured as follows: In Section 2 we define the scaling function ϕ and show that the ϕ-based quasilinear moving average has the required properties. Therebye we find that our formula contains the linear moving average as a special case. Subsequently, in Section 3 we discuss the distributional properties given some simple stochastic processes. In Section 4 we benchmark our formula mainly by deriving its influence function and respective formulas for the linear moving average, the exponential smoothing method and the Kalman filter. Section 5 concludes the paper.

2 Using the scaling function to construct a quasilinear moving average

Before constructing a moving average formula we give a brief definition of a scaling function ϕ and show how to construct a ϕ-based representation of a process $(X_t)_{t \in \mathbb{Z}}$. Subsequently we give in Section 2.2 a ϕ-based smoothing formula and prove it to be a quasilinear moving average. Some properties that follow from this proof are included as well.

2.1 Scaling functions

In literature there are various definitions for the scaling function, some are more restrictive than others. Kaiser (1994), for example, mentiones orthogonality as a core feature, whereas Ahuja et al. (2005) give non-orthogonal functions as well and state that finite support is a desirable requirement. The definition below is mainly based on Percival & Walden (2000), Mallat (2003) and Farge (1992). We omit the orthonormality condition, because it is not required for our purpose and thus merely restricts the set of scaling functions. Our theory is developed for a stochastic processes on a discrete time grid with a constant $\Delta t = 1$.

Definition 2.1 (Scaling Function)**.**
Let $\phi(\cdot) \in \mathcal{L}^2(\mathbb{R})$ with $\phi_{0,b}(x) \equiv \phi(x - b)$ and $\phi_{a,b}(x) \equiv 2^{-a/2}\phi\left(2^{-a}x - b\right), a, b \in \mathbb{Z}$, be a function such that the following properties hold:

Figure 1: The Haar scaling function at different scales
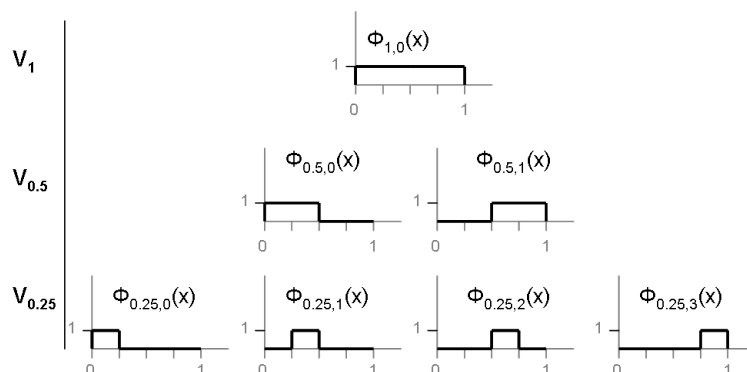
1. $\{\phi_{a,b} : b \in \mathbb{Z}\}$ *is a basis of the closed subspace* $V_a \subset \mathcal{L}^2(\mathbb{R})$ *with* $V_a \equiv \overline{span\{\phi_{a,b}(\cdot) : b \in \mathbb{Z}\}}$.

2. *Given a set of indicies* $\{(a_1, a_2, \ldots, a_k) \in \mathbb{Z}^k \mid a_1 < a_2 < \ldots < a_k, k \in \mathbb{Z}\}$ *the corresponding subspaces have the property*

$$V_0 \supset V_{a_1} \supset V_{a_2} \supset V_{a_3} \supset \ldots \supset V_{a_k} \text{ and } \overline{\bigcup_{a \in \mathbb{Z}} V_a} = \mathcal{L}^2(\mathbb{R}), \quad \bigcap_{a \in \mathbb{Z}} V_a = \{0\},$$

where the overline denotes the closure and the zero represents the null function.

3. $\|\phi\|_1 = 1$.

Then the function ϕ *is called scaling function.*

From Definition 2.1 instantely follows $\|\phi_{a,b}\| = \langle \phi_{a,b}, \phi_{a,b} \rangle = 1, \ \forall a, b \in \mathbb{Z}$. This is because while translating ϕ by (a, b) the energy is preserved by the factor $2^{-a/2}$. Further properties can be found e.g. in Percival (2000) or Farge (2003). We see that b serves as a translation parameter and a is a scaling parameter that is indirect proportional to the frequency.

There is a vast amount of suggestions for scaling functions. For an overview see e.g. Mallat (2003), Daubechies (1992) or Kaiser (1994). Ahuja et al. (2005) also discuss relevant functional properties and favour the B-Spline function (see Appendix B). As a simple example we plot three consecutive subspaces of the Haar scaling function (see Appendix A) in Figure 2.1 to show the effect of shifting the scaling parameter a from Definition 2.1.

We can use ϕ to represent any real-valued function x with finite energy (i.e. $\|x\|_2 < \infty$) as a linear combination of bases of V_0. The coefficients are computed via the orthogonal projection of x on

3

$\phi_{0,b}$, i.e.

$$x(t) = \sum_{b \in \mathbb{Z}} \langle x, \phi_{0,b} \rangle \, \phi_{0,b}(t). \tag{2.1}$$

By increasing a we can generate a more and more smoothed version of $x(t)$, denoted by $h(t)$,

$$h_a(t) = \sum_{b \in \mathbb{Z}} \langle x, \phi_{a,b} \rangle \, \phi_{a,b}(t), \tag{2.2}$$

which excludes the effect of all scales smaller than 2^a (see Percival & Walden, 2000). For efficient computation of $h_a(t)$ it is quite useful to deal with a scaling function that has a finite support $\mathcal{S} = [s_l, s_u] \subset \overline{\mathbb{R}}$. Given that, (2.1) and (2.2) can be reduced to

$$
\begin{aligned}
x(t) &= \sum_{b=t-\lceil s_u \rceil}^{t-\lfloor s_l \rfloor} \phi_{0,b}(t) \sum_{j=\lceil s_l \rceil + b}^{\lfloor s_u \rfloor + b} x(j)\phi_{0,b}(j), \\
h_a(t) &= \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \sum_{j=\lceil (s_l+b)2^a \rceil}^{\lfloor (s_u+b)2^a \rfloor} x(j)\phi_{a,b}(j)
\end{aligned}
\tag{2.3}
$$

whereby $a > 0$ and (2.1) resp. (2.2) is contained as limit for $s_l \to -\infty, s_u \to \infty$.

Having $\mathcal{S}$ we eventually specify for every time t the cone of incluence (COI), which is the set $\left\{ (k,a) \in \mathbb{Z}^2 \mid (\lceil (s_l - s_u)2^a \rceil \le k - t \le \lfloor (s_u - s_l)2^a \rfloor) \right\}$ that denotes – given a scale a – the area around t for which information is required in order to be able to exactly compute (2.3). If not all of the required information is available – which is the case for the edges of a data sample – the coefficients and therefore (2.3) are skewed. Meyers et al. (1993) and Torrence & Compo (1998), for example, give some methods to reduce this skewing, but no method allows to quanitfy how much. Moreover applying these methods can introduce other problems like break points or artificial patterns. A sufficient solution has – to our best knowledge – not yet been found.

2.2 A ϕ-based quasilinear moving average

We now first give the definition of a quasilinear mean and then show how to apply this concept to an autocorrelated time series. A quasilinear mean for an independent identically distributed (i.i.d.) sample $(x_1, \ldots, x_n)$ is defined as

$$M_u = u^{-1} \left(\sum_{i=1}^{n} u(x_i)g_i \right), \quad g_i \ge 0, \quad \sum_{i=1}^{n} g_i = 1, \tag{2.4}$$

with $g_1, \ldots, g_n \in \mathbb{R}$ and $u\,[a,b] \to \mathbb{R}$ continuous, strictly monotone with existing inverse u^{-1} (see Nagumo, 1930, and Kolmogorov, 1930). To identify a function M as a quasilinear mean, we have to show that it can be split up into the components given in (2.4). Alternatively to that, the below theorem can be applied.

Theorem 1 (A quasilinear mean).
The real-valued function M is a quasilinear mean, if and only if it has the following properties:

1. *M is continuous,*

2. *M is monotonic increasing, i.e. $\mathbb{R}^n \ni x \le y \in \mathbb{R}^n \Rightarrow M(x) \le M(y)$,*

3. *M is reflexive, i.e. $M(x, x, \ldots, x) = x$,*

4. *M is associative, i.e. $M(x_1, \ldots, x_{n+1}) = M(x, \ldots, x, x_{n+1})$ with $x = M(x_1, \ldots, x_n)$.*

Proof: See De Finetti (1931), Kolmogorov (1930).

$\square$

The above proof was done for an i.i.d. sample, but we can also apply this concept to an autocorrelated time series and call it moving average. Therefor we interpret at every time t a certain area around it as an i.i.d. data sample on which the quasilinear mean formula is applied. According to this interpration various forms of moving averages can be computed (for examples, see e.g. Hamilton, 1994). Within this section we propose a new candidate by proving that the smoothed time series $h(t)$ from (2.3) is a quasilinear moving average:

Theorem 2 (A ϕ-based quasilinear moving average).
Let $\mathbf{X} = (X_t)_{t \in \mathbb{Z}} \in L^2(\mathbb{R})$ be a real-valued stochastic process with discrete time steps of length $\Delta t = 1$. Let ϕ be a continuous scaling function according to Definition 2.1 with $\sum_b \phi_{a,b}(t)\phi_{a,b}(k) \ge 0$, for any $t, k \in \mathbb{Z}$. Then

$$M(\mathbf{X}; t, a) = \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \sum_{s=\lceil (s_l+b)2^a \rceil}^{\lfloor (s_u+b)2^a \rfloor} X_s \phi_{a,b}(s) \tag{2.5}$$

is a quasilinear moving average of $\mathbf{X}$ at time t for which the properties from Theorem 1 hold.

Proof: *We check if $M(\mathbf{X}; t, a)$ has the properties given in Theorem 1.*

1. Continuity: *The function ϕ is continuous with $\|\phi\| = 1$ and X_t is square-integrable. The convolution of ϕ and X_t is therefore finite and $M(\mathbf{X}; t, a)$ as a linear combination of ϕ does exist and is continuous.*

2. The function $M(\mathbf{X}; t, a)$ is monotonic increasing: *Let $\mathbf{X}'$ be equal to $\mathbf{X}$ but with modified*

5

component $X_k' = X_k + \delta, k \in \mathbb{Z}, \delta > 0$. The difference

$$M(\mathbf{X}'; t, a) - M(\mathbf{X}; t, a) = \delta \sum_{b = \lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \cdot \phi_{a,b}(k) \tag{2.6}$$

is greater or equal zero as demanded in Theorem 2.

3. The function $M(\mathbf{X}; t, a)$ is reflexive: *This follows from $\sum \phi_{a,b} = 1$ (Property 3 of Definition 2.1), because with $X_t = x \in \mathbb{R} \;\; \forall t \in \mathbb{Z}$ we can derive:*

$$
\begin{aligned}
M(\mathbf{X}; t, a) \;&=\; \sum_{b = \lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \sum_{s = \lceil (s_l + b) 2^a \rceil}^{\lfloor (s_u + b) 2^a \rfloor} x \cdot \phi_{a,b}(s) = x \sum_{b = \lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \underbrace{\sum_{s = \lceil (s_l + b) 2^a \rceil}^{\lfloor (s_u + b) 2^a \rfloor} \phi_{a,b}(s)}_{=1} \\
&=\; x \sum_{b = \lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) = x.
\end{aligned}
$$

4. The function $M(\mathbf{X}; t, a)$ is associative: *For the sake of clarity we leave out the exact summation bounds given above and sum the translation parameter b over $\mathbb{Z}$. Let t be such that all data points within its COI are known and set $x = M(X_1, \ldots, X_n; t, a)$*

$$
\begin{aligned}
M(X_1, \ldots, X_n, X_{n+1}; t, a) \;&=\; M(x, \ldots, x, X_{n+1}; t, a) \\
\Leftrightarrow \sum_{b \in \mathbb{Z}} \sum_{s \neq n+1} X_s \phi_{a,b}(t) \phi_{a,b}(s) + X_{n+1} \sum_{b \in \mathbb{Z}} \phi_{a,b}(s) \phi_{a,b}(t) \;&=\; \sum_{b \in \mathbb{Z}} \sum_{s \neq n+1} x \cdot \phi_{a,b}(t) \phi_{a,b}(s) + X_{n+1} \sum_{b \in \mathbb{Z}} \phi_{a,b}(s) \phi_{a,b}(t) \\
\Leftrightarrow M(X_1, \ldots, X_n; t, a) \;&=\; \sum_{b \in \mathbb{Z}} \sum_{s \neq n+1} x \cdot \phi_{a,b}(t) \phi_{a,b}(s) = x = M(X_1, \ldots, X_n; t, a)
\end{aligned}
$$

In the second equation we subtract the last addend on both sides and the last equation holds because $M(x; t, a)$ is reflexive.

$\square$

Using Property 3 and Property 2 of Theorem 1 we can instantly show the internality of $M(\mathbf{X}; t, a)$. I.e. if $X_1, \ldots, X_n$ denote the data points within the cone of influence for time t and scale a, than holds

$$\min(x_1, \ldots, x_n) \leq M(x_1, \ldots, x_n; t, a) \leq \max(x_1, \ldots, x_n).$$

Eventually, having defined and constructed a quasilinear moving average, we want to remark that (2.5) is reduced to the well-known linear moving average when choosing the Haar scaling function for ϕ in Theorem 2. This shows that linear smoothing functions are included in this concept.

3 Distributional properties of a ϕ-based quasilinear moving average

Transforming a process via (2.5) gives rise to the issue of how the distributional properties are modified. This question is of interest when using the modified process within a larger analytical discussion like e.g. trend analysis or forecasting. In this section we therefore discuss this matter based on two simple underlying stochastic processes $(X_t)_{t\in\mathbb{Z}}$, namely an AR(1)- and a GARCH(1,1)-process (if definitions are unknown, see Appendix C). The properties of any linear combination of these processes resp. any generalisation of the lag-order can be derived analogously. For an autoregressive process of lag-order one we can show:

Theorem 3 (The ϕ-based quasilinear moving average of an AR(1)-process)**.**
Let $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ with $X_t = \mu + \theta X_{t-1} + \epsilon_t$ and $\mu, \theta \in \mathbb{R}, \epsilon_t \sim \mathcal{N}(0,\sigma^2), 0 \leq \sigma < \infty$. W.r.t.g. set $X_{-\infty} = 0$. Then $M(\mathbf{X};t,a)$ from (2.5) is again Gaussian distributed at time t with mean and variance

$$
\begin{aligned}
E\left(M(\mathbf{X};t,a)\right) &= \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \left[\sum_{j=\lceil (s_l+b)2^a \rceil}^{t-1} X_j \phi_{a,b}(j) \right.\\
&\quad + \left. \sum_{j=t}^{\lfloor (s_u+b)2^a \rfloor} \phi_{a,b}(j) \left[\mu \sum_{i=0}^{j-t} \theta^i + \theta^{j-t+1} X_{t-1} \right] \right] \\
Var\left(M(\mathbf{X};t,a)\right) &= \sigma^2 \sum_{p=t}^{t+\lfloor 2^a(s_u - s_L)\rfloor} \left[\sum_{q=\lfloor p/2^a - s_u \rfloor}^{\lfloor t/2^a - s_L \rfloor} k_{p,q} \right]^2,
\end{aligned}
\tag{3.1}
$$

with

$$
k_{p,q} = \phi_{a,q}(t) \sum_{j=t}^{\lfloor s_u + q/2^a \rfloor} \theta^{j-t} \phi_{a,q}(j).
$$

Proof: *With $\mathbf{X}$ defined as an autoregressive process and (2.5) we write*

$$
\begin{aligned}
M(\mathbf{X};t,a) &= \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \sum_{s=\lceil (s_l+b)2^a \rceil}^{\lfloor (s_u+b)2^a \rfloor} X_j \phi_{a,b}(j) \\
&= \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \left[\sum_{j=\lceil (s_l+b)2^a \rceil}^{t-1} X_j \phi_{a,b}(j) + \sum_{j=t}^{\lfloor (s_u+b)2^a \rfloor} \phi_{a,b}(j) \left[\mu + \theta X_{j-1} + \epsilon_j \right] \right],
\end{aligned}
$$

because up to time t the realizations of the process are known. Using

$$X_{t+k} = \mu \sum_{i=0}^{k} \theta^i + \theta^{k+1} X_{t-1} + \sum_{i=0}^{k} \theta^i \epsilon_{t+k-i},$$

and $X_{-\infty} = 0$ as well as the autoregressivity of $\mathbf{X}$ resp. $E(e_t) = 0 \; \forall t$ we can simplify the expression to

$$
\begin{aligned}
M(\mathbf{X};t,a) \;=\;& \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \sum_{j=\lceil (s_l+b)2^a \rceil}^{t-1} X_j \phi_{a,b}(j) \\
&+ \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \sum_{j=t}^{\lfloor (s_u+b)2^a \rfloor} \phi_{a,b}(j)\phi_{a,b}(t) \left[\mu \sum_{i=0}^{j-t} \theta^i + \theta^{j-t+1} X_{t-1} + \sum_{i=0}^{j-t} \theta^i \epsilon_{j-i} \right] \\
=\;& \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \sum_{j=\lceil (s_l+b)2^a \rceil}^{t-1} X_j \phi_{a,b}(j) \\
&+ \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \sum_{j=t}^{\lfloor (s_u+b)2^a \rfloor} \phi_{a,b}(j)\phi_{a,b}(t) \left[\mu \sum_{i=0}^{j-t} \theta^i + \theta^{j-t+1} X_{t-1} \right] \\
&+ \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \sum_{j=t}^{\lfloor (s_u+b)2^a \rfloor} \phi_{a,b}(j)\phi_{a,b}(t) \sum_{i=0}^{j-t} \theta^i \epsilon_{j-i} \qquad (3.2)
\end{aligned}
$$

The third factor of (3.2) containing the error terms can be written as

$$\sum_{p=t}^{t+\lfloor 2^a(s_u - s_L) \rfloor} \epsilon_p \left[\sum_{q=\lfloor p/2^a - s_u \rfloor}^{\lfloor t/2^a - s_L \rfloor} k_{p,q} \right] \qquad (3.3)$$

with

$$k_{p,q} = \phi_{a,q}(t) \sum_{j=t}^{\lfloor s_u + q/2^a \rfloor} \theta^{j-t} \phi_{a,q}(j). \qquad (3.4)$$

As $\epsilon_t \sim \mathcal{N}(0,\sigma^2)$ is i.i.d., any sum of innovations is again Gaussian distributed and the (3.1) follows instantely.

$\square$

With the above theorem we see, that $M(\mathbf{X};t,a)$ is stationary if $\mathbf{X}$ is a stationary AR(1)-process, i.e. if $|\theta| < 1$.

In Theorem 3 we fix volatility to be time-invariant. This restriction is omitted in the GARCH(p,q)-model developed by Engle (1982) and Bollerslev (1986), which computes volatility as an autore-

gressive process based on the last p observed errors and last q computed volatility values. We now can show:

Theorem 4 (The ϕ-based quasilinear moving average of a GARCH(1,1)-process).
Let $\mathbf{X} = (X_t)_{t \in \mathbb{Z}}$ be an GARCH(1,1)-process, i.e $X_t = \sigma_t \epsilon_t$ with $\epsilon_t \sim \mathcal{N}(0,1)$ and $\sigma_t^2 = \omega + \alpha X_t^2 + \beta \sigma_{t-1}^2$. For the parameter holds: $\mu \in \mathbb{R}$, $\omega > 0$, $\alpha, \beta \geq 0$. W.r.t.g. set $X_{-\infty} = 0$. Then $M(\mathbf{X}; t, a)$ from (2.5) is again Gaussian distributed with mean and variance

$$E\left(M(\mathbf{X}; t, a)\right) = \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \sum_{j=\lceil (s_l + b)2^a \rceil}^{t-1} X_j \phi_{a,b}(j)$$

$$Var\left(M(\mathbf{X}; t, a)\right) = \sum_{p=t}^{t+\lfloor 2^a(s_u - s_L) \rfloor} \left(\alpha_0 \sum_{i=0}^{p-t} (\alpha_1 + \beta)^i + (\alpha_1 + \beta)^{p-t}(\alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2) \right) \left[\sum_{q=\lfloor p/2^a - s_u \rfloor}^{\lfloor t/2^a - s_L \rfloor} k_{p,q} \right]^2$$

with

$$k_{p,q} = \phi_{a,q}(t) \sum_{j=t}^{\lfloor s_u + q/2^a \rfloor} \theta^{j-t} \phi_{a,q}(j).$$

Proof: *With $\mathbf{X}$ defined as an autoregressive process and (2.5) we write*

$$
\begin{aligned}
M(\mathbf{X}; t, a) &= \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \sum_{j=\lceil (s_l + b)2^a \rceil}^{\lfloor (s_u + b)2^a \rfloor} X_j \phi_{a,b}(j) \\
&= \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \left[\sum_{j=\lceil (s_l + b)2^a \rceil}^{t-1} X_j \phi_{a,b}(j) + \sum_{j=t}^{\lfloor (s_u + b)2^a \rfloor} \phi_{a,b}(j) \sigma_t \epsilon_t \right],
\end{aligned}
$$
$$(3.5)$$

because, again, up till time t the realisations of the process are known. So, as the mean is additive and $E(\epsilon_t) = 0$, we get

$$E\left(M(\mathbf{X}; t, a)\right) = \sum_{b=\lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \sum_{j=\lceil (s_l + b)2^a \rceil}^{t-1} X_j \phi_{a,b}(j).$$

The process $\mathbf{X}$ is uncorrelated. So its sum is again Gaussian distributed. The mean is already given above and for the variance we use a result from (see McNeil et al., 2005) that reads as

$$Var(X_{t+h}|\mathcal{F}_t) = \alpha_0 \sum_{i=0}^{h} (\alpha_1 + \beta)^i + (\alpha_1 + \beta)^h (\alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2), \qquad (3.6)$$

9

whereby $\mathcal{F}_t$ is the Filtration adapted to $\mathbf{X}$. With (3.3) and (3.6) we can conclude

$$
\begin{aligned}
Var\left(M(\mathbf{X}; t, a)\right) &= \sum_{p=t}^{t+\lfloor 2^a(s_u-s_L)\rfloor} Var(\epsilon_p|\mathcal{F}_t) \left[\sum_{q=\lfloor p/2^a-s_u\rfloor}^{\lfloor t/2^a-s_L\rfloor} k_{p,q}\right]^2 \\
&= \sum_{p=t}^{t+\lfloor 2^a(s_u-s_L)\rfloor} \left(\alpha_0 \sum_{i=0}^{p-t} (\alpha_1+\beta)^i + (\alpha_1+\beta)^{p-t}(\alpha_1\epsilon_{t-1}^2+\beta\sigma_{t-1}^2)\right) \left[\sum_{q=\lfloor p/2^a-s_u\rfloor}^{\lfloor t/2^a-s_L\rfloor} k_{p,q}\right]^2
\end{aligned}
$$

$\square$

4 Comparing the ϕ-based quasilinear moving average to other smoothing functions

Having introduced a new moving average formula for time series smoothing, we now compare it to existing ones in order to justify and motivate its application. Therefore we first pick and define three widely used smoothing methods, namely linear moving average, exponential smoothing and Kalman filter. In a second step, we compare them to our formula. The focus of the analysis lies thereby on the influence of shifts in the time series and and the calibration options.

4.1 A variety of smoothing functions

The definitions given in this section are mainly based on Hamilton (1994) and Provost (1994). For the computation of the influence function we define a shock as a value shift of size $\delta_k \in \mathbb{R}$ at time $k \in \mathbb{Z}$. Moreover, let $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$ be a stochastic process.

The ϕ-based quasilinear moving average $M(\mathbf{X}; t, a)$ reads – according to (2.5) – as

$$
M(\mathbf{X}; t, a) = \sum_{b=\lceil t/2^a-s_u\rceil}^{\lfloor t/2^a-s_l\rfloor} \phi_{a,b}(t) \sum_{s=\lceil (s_l+b)2^a\rceil}^{\lfloor (s_u+b)2^a\rfloor} X_s \phi_{a,b}(s).
$$

We see that $M(\mathbf{X}; a, t)$ is only influenced, if k lies within the COI of t given a specific ϕ. The COI's size is controlled by the scale a which has - as it is indirect proportional to the frequency- a fundamental meaning. The smaller a, the more high-frequency oszillations are captured in (2.5) and vice versa, i.e. (2.5) tends to be smoother with increasing size of a. Moreover the smaller the

distance $|k - t|$, the larger the influence of δ_k. The influence is given by

$$I^\phi(\delta_k; t, a) = \delta \sum_{b = \lceil t/2^a - s_u \rceil}^{\lfloor t/2^a - s_l \rfloor} \phi_{a,b}(t) \cdot \phi_{a,b}(k).$$ (4.1)

where ϕ represents the chosen scaling function (see (2.6)). The linear moving average (LMA) at time t is defined as

$$LMA(\mathbf{X}; t, n) = \frac{1}{2n + 1} \left(X_{t-n} + X_{t-1} + \ldots + X_t + X_{t+1} + X_{t+n} \right),$$ (4.2)

whereby $n \in \mathbb{N}_0$ and $t \in \mathbb{Z}$. The choice of n decides about the influence of outliners and the smootheness of LMA: the larger n the smaller the influence of a single value and vice versa. However, at the same time: the larger n, the bigger the probability that an outliner lies in the domain of the $LMA(\mathbf{X}; t, n)$. From (4.2) we derive for the influence function:

$$I^{LMA}(\delta_k; n, k) = \begin{cases} \delta_k/n & \text{if } t - n \leq k \leq t + n, \\ \\ 0 & \text{if } |t - k| > n. \end{cases}$$

Exponential smoothing (ES), again, is characterized by

$$ES(\mathbf{X}; t, \alpha) = ES(t - 1; \alpha) + \alpha \left(X_t - ES(t - 1; \alpha) \right) = \alpha X_t + (1 - \alpha)ES(t - 1; \alpha),$$ (4.3)

with $0 < \alpha < 1$. The smaller α the less sensitive is $ES(\mathbf{X}; t, \alpha)$ and vice versa (see Provost, 1994). As ES is iterative, we can rewrite $ES(\mathbf{X}; t, \alpha)$ from (4.3) as

$$ES(t; \alpha) = \alpha \sum_{i=0}^{h} (1 - \alpha)^i X_{t-i} + (1 - \alpha)^h ES(t - h; \alpha) \to \alpha \sum_{i=0}^{\infty} (1 - \alpha)^i X_{t-i} \text{ for } h \to \infty.$$

So the influence function reads as

$$I^{ES}(\delta_k; \alpha, k) = \begin{cases} \delta_k \cdot \alpha(1 - \alpha)^{t-k} & \text{if } k \leq t, \\ \\ 0 & \text{if } k > t. \end{cases}$$

For the Kalman filter, eventually, we refer to the simplified recursive scheme given by Provost (1994). The smoothed time series ST_t at time t is computed as follows: Let δ_t be the difference

11

between the actual obeservation at time t and the predicted observation based on the estimates at time $t-1$. Let SSL_T be the smoothed slope at time t. Then, the Kalman filter reads as

$$
\begin{aligned}
\delta_t &= X_t - (ST_{t-1} + SSL_{t-1}), \\
ST_t &= (ST_{t-1} + SSL_{t-1}) + \alpha_1 \cdot \delta_t, \;\; 0 < \alpha_1 < 1, \\
SSL_t &= SSL_{t-1} + \alpha_2 \cdot \delta_t, \;\; 0 < \alpha_2 < 1.
\end{aligned} \tag{4.4}
$$

The parameter α_1 measures the influence of the prediction error on the smoothed time series, α_2 measures its influence on the smoothed slope. Both allow to tune the smootheness of the Kalman filter generated trend. As starting values for time $t = t_0$, it is suggested to use $ST_{t_0} = X_{t_0}$, $SSL_{t_0} = 0$. To get the respective influence function, we rewrite (4.4) according to the standard Kalman state equation:

$$
\begin{pmatrix} ST_t \\ SSL_t \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} X_t + \begin{pmatrix} 1 - \alpha_1 & 1 - \alpha_1 \\ -\alpha_2 & 1 - \alpha_2 \end{pmatrix} \begin{pmatrix} ST_{t-1} \\ SSL_{t-1} \end{pmatrix} + w_t \tag{4.5}
$$

where w_t is distributed according to the specified prior distribution. Then follows for the two-dimensional influence function (ST_t, SSL_t):

$$
I^{Kalman}(\delta_k; k, \alpha_1, \alpha_2) = \begin{cases} \begin{pmatrix} 1 - \alpha_1 & 1 - \alpha_1 \\ \alpha_2 & 1 - \alpha_2 \end{pmatrix}^{t-k} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \delta_k & \text{if } k \leq t, \\ \\ 0 & \text{if } k > t. \end{cases} \tag{4.6}
$$

From (4.6) we see – like Provost (1994) – that we have to impose restrictions on the combination possibilities of (α_1, α_2). The matrix-product in (4.6) has to be strictly monotonous decreasing with increasing power, because otherwise a shock would have the more influence, the more far away in the past it is.

4.2 Comparing the smoothing functions

In section 4.1 we give the functional form of four smoothing methods as well as the respective influence functions. Based on this information we are now able to compare them. As a visual aid we plot the four influence functions derived above exemplarily for time $t = 15$ and $\delta_k = 1$ in Figure 2.

Concerning the parameter calibration, all functions allow to reduce the influence of local process

12

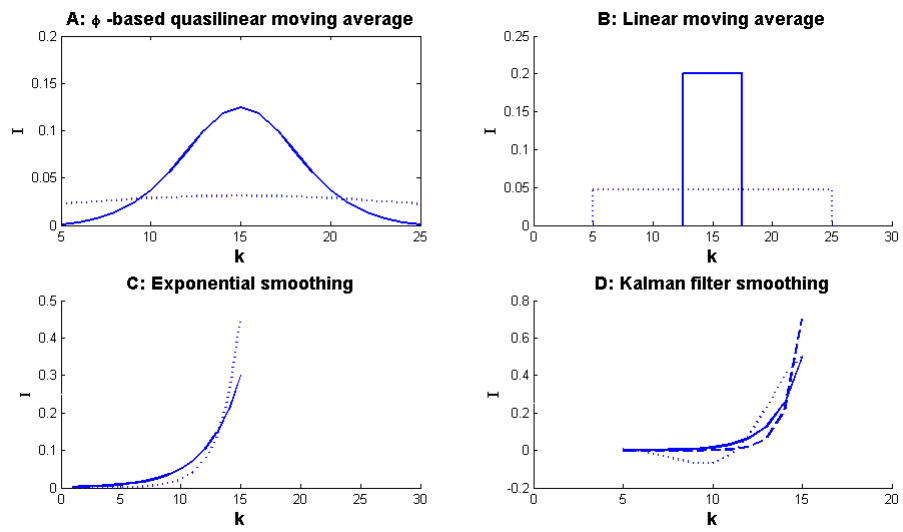Figure 2: The four influence functions at time $t = 15$

Figure A: I^ϕ using B-spline scaling function with $L = 4$. The dotted shows the influence function for scale $a = 4$, the solid line is the same function for $a = 2$. Figure B: I^{LMA} for varios k and $n = 2.5$ (solide line) and $n = 10$ (dotted line). Figure C: I^{ES} for varios k and $\alpha = 0.3$ (solide line) and $\alpha = 0.45$ (dotted line). Figure D: I^{Kalman} for varios k and $(\alpha_1, \alpha_2) = (0.5, 0.002)$ (solide line), $(\alpha_1, \alpha_2) = (0.5, 0.3)$ (dotted line) and $(\alpha_1, \alpha_2) = (0.7, 0.002)$ (dashed line).

realisations to the cost of increasing influence of more distant values. So increasing the smootheness of the transformed functions means at the same time increasing the the influence of shocks from the past. This is especially significant when computing the LMA for a time t, as within its cone of influence all values have the same wheight. In all other methods the influence of a value is super-linear decreasing with increasing distance to t. ES and Kalman filter have the additional advantage of future shocks being excluded as their influence functions equal zero for time index $k > t$.

However, the Kalman filter has one significant drawback, which is parameter calibration. It has a bivariate parameter with no fundamental meaning. Thus makes it complicated to justify any parameter-choice. The same holds for ES - in contrary to LMA where n can be chosen such that a weekly or monthly average, for example, is the result. In case of a ϕ-based quasilinear mean, the parameter a represents a certain scale resp. frequency (depending on ϕ). All oszillations with lower scale resp. higher frequency are excluded. If we want to analyze, for example, the time series development on a weekly basis, the a is chosen such that all oszillations with period lower than seven are excluded.

Basically, there are two main application scenarios for smoothing functions: they are used as a running average (computed every time step based on historical data), or they are applied to smooth out high-frequent oszillations (i.e. noise) in order to identify any structures within a given time series (e.g. seasonal patterns). In the first scenario, LMA and (2.5) are skewed or show the drawback of lagging behind in case of a ramp does exist. The reason is, that both methods require information about the future development of the process. As this is not given, the data have to be either forecasted, ignored in the computation process or the time for which the respective formula is applied is shifted in the past such that all required information is given. In case of LMA at time t this would mean that not $LMA(\mathbf{X}; t, n)$ is computed, but $LMA(\mathbf{X}; t - n, n)$.

ES needs only historical data, but shows the same drawback (see Provost, 1994) – in contrary to the Kalman filter, which includes the ramp explictely (see (4.4)). Provost (1994) adds some more advantages of the Kalman filter to this statement, but still its main disadvantage, the parameter calibration, remains.

If a fixed data set is given, (2.5) and the LMA enjoy the benefit of having a cone if influence, i.e. only a limited data-set influences the smoothed function. Outside this area, shocks have no effect. However, this cone of influence is a disadvantage for the data set's edge regions as there the smoothed function is skewed (see Section 2). At the same time, ES still lags behind in case of a ramp, which makes this method inadequate for both scenarios. In toto, there are four main arguments for using the ϕ-based quasilinear moving average:

- In contrary, to ES and Kalman filtering we are able to give the scaling parameter a a content-related interpretation. This is essential for motivating any parameter choice.

- There is a cone of influence for our formula, i.e. outside a certain region, shocks do not

influence (2.5) any more – in contrary to ES and Kalman filter where due to recursiveness the whole past has an influence on the current value (although with decreasing size).

- Formula (2.5) introduces a rank system within its cone of influence. The greater the distance to time t, the smaller the influence of the corresponding value on (2.5). Thus the smoothed function adaptd the motion of the underlying time series – more than the LMA.

- In contrary to the Kalman filter, no prior distribution or starting values (as it is also demanded for ES) are required. Formula (2.5) only requires a function ϕ and therefor various selection criterions exist (see e.g. Ahuja et al., 2005).

Due to these points, we consider (2.5) it as an adequate tool for smoothing a time series, especially when seasonal patterns are included.

5 Conclusion

Within this paper we have constructed a moving average based on the scaling function and prove it to be quasilinear. We have discussed further features and have especially derived its distributional properties given some simple underlying stochastic processes. This helps to answers some questions that arise when using the smoothing function in practice.

Eventually we have compared our function to other widely-used smoothing functions. This benchmark motivates the use of the ϕ-based quasilinear moving average. In contrary to exponential smoothing or Kalman filter, the parameter has a content-related interpretation (it is inverse proportional to the frequency) and no starting values resp. a prior-distribution (Kalman filter) are required. Moreover, (2.5) is not recursive like ES and Kalman filter. So, the smoothed function is not influenced by the whole past but only within a certain area arount time t (the cone of influence). Shocks from the past do not necessarily have a influence on the recent smoothed values. Due to these facts we consider the ϕ-based quasilinear mean as a powerful and flexible time series smoothing tool.

References

[1] Ahuja N, Lertrattanapanich S, Bose N K. Properties determining choice of mother wavelet. IEEE Proceedings - Vision, Image & Signal Processing 2005; 152(5); 659–664.

[2] Bollerslev T. Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics 1986; 31; 307–327.

[3] Daubechies I. Tel Lectures on Wavelets. Society for Industrial and Applied Mathematics: Philadelphia; 1992.

[4] De Finetti B. Sul Concetto di Media. Giornale dell' Instituto Italiano degli Attuari 1931; 2; 369–396.

[5] Engle R F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica 1982; 50(4); 987–1007.

[6] Farge M. Wavelet transforms and their applications to turbulence. Annual Review Fluid Mechanics 1992; 24; 395–457.

[7] Hamilton J D. Time Series Analysis. Princeton University Press: Princeton; 1994.

[8] Kaiser G. A Friendly Guide to Wavelets. Birkhäuser: Boston; 1994.

[9] Kolmogorov A. Sur la Notion de la Moyenne. Rendiconti Accademia dei Lincei 1930; 12(6); 388–391.

[10] Mallat S. A wavelet tour of signal processing, 2nd edition. Academic Press: Manchester; 2003.

[11] McNeil A J, Frey R, Embrechts P. Quantitative risk management: concepts, techniques, and tools. Princeton University Press, Princeton; 2005.

[12] Meyers S D, Kelly B G, O'Brien J J. An Introduction to Wavelet Analysis in Oceanography and Meteorology: With Application to the Dispersion of Yanai Waves. Monthly Weather Review 1993; 121(10); 2858-2866.

[13] Nagumo. Über eine Klasse der Mittelwerte. Japan Journal of Mathematics 1930; 76; 71–79.

[14] Percival D, Walden A. Wavelet methods for time series analysis. Cambridge University Press, Cambridge; 2000.

[15] Provost, M J. THe Use if Optimal Estimation Techniques in the Analysis of Gas Turbines. PhD Thesis, Cranfield University, Cranfield; 1994.

[16] Stein E.M.; Shakarchi, R. Fourier Analysis: An Introduction. Princeton University Press: Princeton; 2003.

[17] Stollnitz E, DeRose T D, Salesin D H. Wavelets for computer graphics: A primer, part 1. IEEE Computer Graphics and Applications 1995, 15(3):76–84.

[18] Unser M. Ten good reasonst for using spline wavelets. IEEE Signal Processing 1999; 16; 22–38.

A   The Haar scaling function

Let $x \in \mathbb{R}$, then the Haar scaling function is defined as (see Mallat, 2003):

$$\phi = \begin{cases} 1 & \text{for } 0 \leq t < 1 \\ 0 & \text{otherwise.} \end{cases}$$

B The B-spline scaling function

Let $L \in \mathbb{N} \setminus \{1\}$, then the function ϕ with

$$\phi(x) = \frac{1}{(L-1)!} \sum_{k=0}^{L} (-1)^k \begin{pmatrix} L \\ k \end{pmatrix} (x-k)_+^{L-1}$$

is called B-spline scaling function (see Unser, 1999).

C Autoregressivity in a time series and in its volatility

A process $(X_t)_{t \in \mathbb{Z}}$ is called autoregressive process of order p $(p \in \mathbb{N})$ or short AR(p)-process, if it has the form

$$X_t = \theta_0 + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots + \theta_p X_{t-p} + \epsilon_t, \tag{C.1}$$

where ϵ_t is the error term (i.e. a random variable itself) and $\theta_0, \ldots, \theta_p \in \mathbb{R}$ (see Hamilton, 1994).

Autoregressive effects in volatility are captured by the so-called GARCH(p,q)-model for $p, q \in \mathbb{N}$ (see Engle, 1982, and Bollerslev, 1986). GARCH is the acronym for "Generalized Autoregressive Conditional Heteroscedacity" and the standard model reads as

$$\begin{aligned} X_t &= \sigma_t \epsilon_t \quad \epsilon_t \sim \mathcal{N}(0,1), \\ \sigma_t^2 &= \omega_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \end{aligned}$$

with $\omega_0 > 0$, $\alpha_i, i = 1, \ldots, p, \geq 0$, $\beta_j, j = 1, \ldots, q, \geq 0$. For further properties and a detailed discussion we refer to McNeil et al. (2005).

Diskussionspapiere 2009
Discussion Papers 2009

01/2009   **Addison, John T. and Claus Schnabel:** Worker Directors: A German Product that Didn't Export?

02/2009   **Uhde, André and Ulrich Heimeshoff:** Consolidation in banking and financial stability in Europe: Empirical evidence

03/2009   **Gu, Yiquan and Tobias Wenzel:** Product Variety, Price Elasticity of Demand and Fixed Cost in Spatial Models

04/2009   **Schlüter, Stephan:** A Two-Factor Model for Electricity Prices with Dynamic Volatility

05/2009   **Schlüter, Stephan and Fischer, Matthias**: A Tail Quantile Approximation Formula for the Student t and the Symmetric Generalized Hyperbolic Distribution

06/2009   **Ardelean, Vlad:** The impacts of outliers on different estimators for GARCH processes: an empirical study

07/2009   **Herrmann, Klaus:** Non-Extensivity versus Informative Moments for Financial Models - A Unifying Framework and Empirical Results

08/2009   **Herr, Annika:** Product differentiation and welfare in a mixed duopoly with regulated prices: The case of a public and a private hospital

09/2009   **Dewenter**, Ralf, **Haucap**, Justus and **Wenzel**, Tobias: Indirect Network Effects with Two Salop Circles: The Example of the Music Industry

10/2009   **Stuehmeier**, Torben and **Wenzel**, Tobias: Getting Beer During Commercials: Adverse Effects of Ad-Avoidance

11/2009   **Klein**, Ingo, **Köck**, Christian and **Tinkl**, Fabian, Spatial-serial dependency in multivariate GARCH models and dynamic copulas: A simulation study

_____

Friedrich-Alexander-Universität
IWQW
Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung

Diskussionspapiere 2008
Discussion Papers 2008

01/2008 **Grimm, Veronika and Gregor Zoettl**: Strategic Capacity Choice under Uncertainty: The Impact of Market Structure on Investment and Welfare

02/2008 **Grimm, Veronika and Gregor Zoettl**: Production under Uncertainty: A Characterization of Welfare Enhancing and Optimal Price Caps

03/2008 **Engelmann, Dirk and Veronika Grimm**: Mechanisms for Efficient Voting with Private Information about Preferences

04/2008 **Schnabel, Claus and Joachim Wagner**: The Aging of the Unions in West Germany, 1980-2006

05/2008 **Wenzel, Tobias**: On the Incentives to Form Strategic Coalitions in ATM Markets

06/2008 **Herrmann, Klaus**: Models for Time-varying Moments Using Maximum Entropy Applied to a Generalized Measure of Volatility

07/2008 **Klein, Ingo and Michael Grottke:** On J.M. Keynes' "The Principal Averages and the Laws of Error which Lead to Them" - Refinement and Generalisation

_____