Högn, Ralph; Czado, Claudia

**Working Paper**

# Multiresolution analysis of long time series with applications to finance

Discussion Paper, No. 497

**Provided in Cooperation with:**
Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

*Suggested Citation:* Högn, Ralph; Czado, Claudia (2005) : Multiresolution analysis of long time series with applications to finance, Discussion Paper, No. 497, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München, https://doi.org/10.5282/ubm/epub.1864

This Version is available at:
https://hdl.handle.net/10419/31156

# Multiresolution Analysis of Long Time Series With Applications to Finance

Ralph Högn, Claudia Czado

Zentrum Mathematik, Technische Universität München

August 17, 2005

**Abstract**

We consider multi-resolution time series models and their application to high-frequency financial data. An individual transaction share price of a specific firm is subject to market microstructure noise. Therefore, we propose trading duration time weighted averages over given time intervals. Averages over long intervals lead to a coarse resolution and averaging over shorter intervals lead to a finer resolution. Arranging sub-intervals of given lengths on scales with coarse to fine resolution imply a structure which can be represented as a directed acyclic graph. Time series models are then formulated using this graph structure. It is shown that these models have a linear state space representation which allows for efficient computation of the likelihood needed in parameter estimation and for a straightforward treatment of missing observations. Application of these models to the log transaction prices of the IBM shares traded at the New York Stock Exchange from February until October 2002 show that the corresponding one-step prediction errors are heavy tailed and therefore a specific variance term is allowed to follow a fiEGARCH specification, improving the tail behavior and leading to a better fit.

**Keywords:** multiresolution, time series, state space representation, colored transition noise, directed acyclic graphs.

# 1 Introduction

In this paper we introduce and discuss multi-resolution time series models and their application to high-frequency financial data. In particular, we consider models for all transactions data of company over several months traded at the New York Stock Exchange (NYSE). Such tick-by-tick financial data give rise to very large data sets. Each tick represents direct information from the market. The price process for an equity of one individual company is strongly determined by the specific behavior of this firm and general trends in the market, the economy and the branch. A detailed discussion of the properties of high-frequency financial data can be found for example in Campbell et al. (1997), Dacorogna et al. (2001), Goodhart and O'Hara (1997), Engle (2000) and in Hautsch and Pohlmeier (2002). We focus in this paper on models for transaction prices allowing for duration effects and market micro structures. Models for transaction prices are the building block for assessing financial market volatility (see for example Shephard (2005)), which is of paramount importance in risk management and option pricing.

First, transaction prices are quoted in discrete units, so-called ticks. Current tick-size at the NYSE is $\frac{1}{100}\$$. Further, trade direction (buy or sell from the market maker) has an impact on the trading price. Unfortunately, the trade direction is not recorded in the Trade-and-Quote (TAQ) database of the NYSE. Observations are sampled at irregularly spaced random time points and are unlikely to be independently distributed. Within one trading day some observations are separated only by a few seconds while others are separated by several minutes resulting in an inhomogeneous time series. An individual value measured at a specific time point is subject to market microstructures such as discreteness in the price observations, the trade direction and bid-ask bounces. As a result, an individual value is severely distorted and of little significance. Therefore we are more interested in determining average values over intervals in order to reduce such effects. In Section 2 we propose to use the trading duration time weighted average of the log-prices over a given time interval to transform the inhomogeneous time series into a homogeneous one.

The remainder is organized as follows: In Section 3 we discuss a multi-resolution approach for averaging, where on scales with a coarse (fine) resolution averaging is conducted over long (short) intervals. This structure is represented by a directed acyclic graph. In Section 4 we formulate time series models for normally distributed observations using a given directed acyclic graph. In Section 5 a linear state space representation for these models is derived which allows for efficient evaluation of the likelihood and the treatment of missing observations. In Section 6 we apply

these models to trading duration time weighted averages of log-prices of IBM shares from the NYSE. Here time invariant variance terms are not appropriate and therefore we allow a specific variance term to follow a fiEGARCH specification as introduced by Bollerslev and Mikkelsen (1996) which improves the fit. The final section provides conclusions and model extensions.

# 2 Trading Duration Time Weighted Averages

Regular trading days and hours at the NYSE are week-days from Monday until Friday from $9:30$ a.m. until $4:00$ p.m. Eastern Time. In after-hours less trading activity, less market participants and less liquidity are observed. Therefore we neglect after-hours trading. Since periods of trading interchange with periods of non-trading (or after-hours trading) in calendar time, the calendar time scale is transformed into a trading time scale, which omits the periods of non-trading or that of after-hours trading. Starting at zero we extend for each full trading day the trading time scale by $23.400$ seconds the length of a regular trading day. Week-ends and holidays are omitted. The complete observation interval on the trading time scale we denote by $I_O$.

For a given interval $I := [t_l^I, t_r^I] \subset I_O$ we define the trading duration time weighted average of log-prices as follows: Let $N_I$ denote the number of trades occurring within $I$, $t_i$ the time point of the $i^{th}$ trade, $i = 1, \ldots, N_I$ and $p(t_i)$ the observed log-price at time $t_i$, $t_i \in I$. For the special case of several transactions recorded at time $t_i$ we call the number of trades at $t_i$ by $n_i$, the volume of the $l^{th}$ trade occurring at $t_i$ by $v_{i,l}$, and the log-price for the $l^{th}$ trade at $t_i$ by $p_{i,l}$, $l = 1, \ldots, n_i$. In this case we define the log-price at $t_i$ to be $p(t_i) := \frac{1}{v_i} \sum_{l=1}^{n_i} p_{i,l} v_{i,l}$ where $v_i := \sum_{l=1}^{n_i} v_{i,l}$. For $t_0 := t_l^I$ we define $p(t_0)$ to be the last observed log-price before $t_0$ if no trading is recorded at $t_0$. The trading duration time weighted average of log-prices over $I$ is then given by

$$p_I := \frac{1}{|I|} \sum_{i=0}^{N_I} p(t_i)[t_{i+1} - t_i], \tag{1}$$

where $|I|$ denotes the interval length. If there is no trade in $I_{j,k}$ then (1) implies $p_I = p(t_0)$. Definition (1) allows that the price process $p(t),\ t \in I$, can be interpreted as a step-function in continuous time: $p(t) := p(t_i),\ t_i \leq t < t_{i+1},\ i = 1, \ldots, N_I$. Therefore, we may view the trading duration time weighted average of log-prices as a normalized Riemann integral

$$p(I) := \frac{1}{|I|} \int_I p(s)\, ds.$$

# 3 Interval Arrangement and Directed Acyclic Graphs

It is an open questions how to choose the interval length of $I$. As discussed in Section 3.3 of Dacorogna et al. (2001), averaging over long intervals leads to slow reaction to changes in the underlying time series but has less noise, while averages over shorter intervals yield a faster reaction to changes but have more noise. Therefore, we do not consider just one interval length but we combine different interval lengths into one model using a multi-resolution approach. For example, on a first scale we consider averages over weeks, on a second scale averages over days, on a third scale averages over hours and on the finest scale averages over five minute intervals. This multi-resolution approach allows to formulate models that characterize the long-term behavior, relating to coarse scales (weeks), and the short term behavior, relating to finer scales (five minute intervals). More formally, let $J$ denote the number of scales. For a family of natural numbers $\{N_j, j = 1, \ldots, J\}$ with $N_{j+1} > N_j$ we consider a collection of sub-intervals $\{I_{j,k} \in I_O : j = 1, \ldots, J, k = 1, \ldots, N_j\}$ for which the following assumptions hold for $j = 1, \ldots, J$:

(2) $$|I_{j,k}| > 0,\ k = 1, \ldots, N_j,$$

(3) $$\bigcup_{k=1}^{N_j} I_{j,k} = I_O,$$

(4) $$I_{j,k} \cap I_{j,k+2} = \emptyset,\ k = 1, \ldots, N_j - 2,$$

For $k = 1, \ldots, N_j - 1$ there exists a $k' \in \{1, \ldots, N_{j-1}\}$

(5) such that $(I_{j,k} \cup I_{j,k+1}) \subset (I_{j-1,k'} \cup I_{j-1,k'+1})$

From the interval arrangement defined in (2) - (5) we obtain an undirected graph $\mathcal{G}^U := (\mathcal{V}, \mathcal{E})$ with the set of nodes defined $\mathcal{V} := \{(j,k) \in \mathbb{N} \times \mathbb{N} : j = 1, \ldots, J,\ k = 1, \ldots, N_j\}$ and the set of edges $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ defined by

(6) $$I_{j-1,k} \cap I_{j,k'} \neq \emptyset \Leftrightarrow \big((j-1,k),(j,k')\big) \in \mathcal{E}.$$

A directed version of $\mathcal{G}^U$, denoted by $\mathcal{G}^S$, is obtained by representing the edges as arrows pointing from $(j-1,k)$ to $(j,k')$, if $((j-1,k),(j,k')) \in \mathcal{E}$. Conditions (5) and (6) imply that a node can have up to two parents. Since arrows only point from a node on scale $j$ to a node on scale $j+1$, $j = 0, \ldots, J-1$, we call Graph $\mathcal{G}^S$ a **two-parent scale-to-scale graph**. Figure 1 illustrates an example for an interval arrangement defined by (2) - (5) and gives a graphical representation of the corresponding two-parent scale-to-scale Graph $\mathcal{G}^S$.

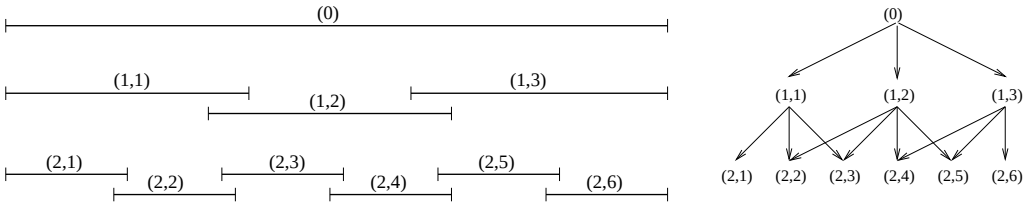


Figure 1: Example for an Interval Arrangement defined by (2) - (5)
and the Representation of the Corresponding Graph $\mathcal{G}^S$

We need some basic graph theory, as discussed for example in Huang and Cressie (2001) and Lauritzen (1996). A path of length $n$ from node $(j, k_j)$ to node $(i, k_i)$, $i := j + n - 1$, is a sequence $(j, k_j), (j+1, k_{j+1}), \ldots, (i, k_i)$ of distinct nodes such that $((l, k_l), (l + 1, k_{l+1})) \in \mathcal{E}$ for $l = j, \ldots, i - 1$. Node $(j, k_j)$ is said to be an ancestor of node $(i, k_i)$ if there is a path from $(j, k_j)$ to $(i, k_i)$, for $i > j$ and node $(i, k_i)$ is then said to be a descendant of node $(j, k_j)$. We denote the set of descendants on scale $i$ of node $(j, k_j)$ by $de(i, j, k)$. If $((j - 1, k_{j-1}), (j, k_j)) \in \mathcal{E}$ then node $(j - 1, k_{j-1})$ is said to be a parent of node $(j, k_j)$. The set of parents on scale $j - 1$ of a node $(j, k_j)$ is denoted by $pa(j - 1, j, k_j)$. We say that a node $(j - 1, k_{j-1}) \in pa(j - 1, j, k_j)$ is a left parent of $(j, k_j)$ if $(j - 1, k_{j-1})$ is the unique parent of $(j, k_j)$ or if both nodes $(j - 1, k_{j-1})$ and $(j - 1, k_{j-1} + 1)$ are parents of $(j, k_j)$. In a similar manner, we say that a node $(j - 1, k_{j-1}) \in pa(j - 1, j, k_j)$ is a right parent of $(j, k_j)$ if $(j - 1, k_{j-1})$ is the unique parent of $(j, k_j)$ or if both nodes $(j - 1, k_{j-1} - 1)$ and $(j - 1, k_{j-1})$ are parents of $(j, k_j)$. Since $(j, k_j)$ can have up to two parents, it has a unique left parent and a unique right parent, $j = 2, \ldots, J$. We denote the left parent of $(j, k_j)$ on scale $j - 1$ by $lpa(j - 1, j, k_j)$ and the right parent by $rpa(j - 1, j, k_j)$, respectively. A cycle is a path which begins and ends with the same node. In this paper we consider directed acyclic graphs only.

# 4 Time Series Models Using Directed Acyclic Graphs

We formulate now time series models using a given directed acyclic graph $\mathcal{G}^S$ for trading duration time weighted averages of log-prices. Huang and Cressie (2001) and Huang et al. (2002) discussed stochastic models over an acyclic directed graph for spatial data. Modification of these models to time series is straightforward. However, Huang and Cressie (2001) and Huang et al. (2002) considered only a deterministic trend. Therefore, we relax their assumptions to allow for

a stochastic trend which may be more appropriate for time series. Since in our application in Section 6 we require only models for univariate time series we restrict ourselves to this case. An extension to multivariate time series is straightforward and is discussed in Högn (2005).

As Huang and Cressie (2001) we define a univariate stochastic process $\{Y_{j,k}, j = 0, \ldots, J,\ k = 1, \ldots, N_j\}$ indexed by the nodes of a given Graph $\mathcal{G}^S$ with univariate or multivariate transition noise $\{\boldsymbol{W}_{j,k}\ k = 1, \ldots, N_j\}$, $j = 1, \ldots, J$, as follows:

$$Y_{1,k} = \boldsymbol{G}_1 \boldsymbol{W}_{1,k}\ , k = 1, \ldots, N_1, \tag{7}$$

and for $j = 2, \ldots, J$

$$Y_{j,k} = \alpha_{j,k} Y_{lpa(j-1,j,k)} + \beta_{j,k} Y_{rpa(j-1,j,k)} + \boldsymbol{G}_j \boldsymbol{W}_{j,k},\ k = 1, \ldots, N_j, \tag{8}$$

where $\alpha_{j,k}, \beta_{j,k} \in [0, 1]$ and $\alpha_{j,k} + \beta_{j,k} = 1$. On scale $j$, $\boldsymbol{W}_{j,k} \in \mathbb{R}^{m_j^W}$ is a random vector and $\boldsymbol{G}_j \in \mathbb{R}^{1 \times m_j^W}$ is a known selection matrix. We assume that observations $Z_k$ are directly related to the finest scale only:

$$Z_k := Y_{J,k}. \tag{9}$$

This is not restrictive, since we can always find a graph $\mathcal{G}^S$ such that (9) is satisfied. We assume that the transition noise $\{\boldsymbol{W}_{j,k}, j = 1, \ldots, J, k = 1, \ldots, N_j\}$ are jointly normally distributed and the transition relating to different scales is mutually independent, i.e.

$$\{\boldsymbol{W}_{j,k}\ k = 1, \ldots, N_j\} \perp \{\boldsymbol{W}_{i,l}\ l = 1, \ldots, N_i\},\ i, j = 1, \ldots, J,\ i \neq j, \tag{10}$$

where $\boldsymbol{U} \perp \boldsymbol{V}$ denotes that both random vectors $\boldsymbol{U}, \boldsymbol{V}$ are independent. Let $\boldsymbol{W}^j := \{\boldsymbol{W}_{i,k},\ i = 1, \ldots, j,\ k = 1, \ldots, N_i\}$ denote the transition noise relating to scales $i = 1, \ldots, j$. From (7) and (8) it follows that $Y_{j,k}$, $j = 1, \ldots, J$, $k = 1, \ldots, N_j$, is a linear function of a subset of $\boldsymbol{W}^j$.

## 4.1 Physically Overlapping Parameters

We interpret $\alpha_{j,k}$ and $\beta_{j,k}$ for $j = 1, \ldots, J$, $k = 1, \ldots, N_j$, as physically overlapping parameters. In particular, if $I_{j,k} \subset I_{lpa(j-1,j,k)}$ then the node $(j, k)$ of $\mathcal{G}^S$ has a unique parent on scale $j-1$ and therefore $lpa(j-1, j, k) = rpa(j-1, j, k)$. In this case we set either $\alpha_{j,k} := 1$ or $\beta_{j,k} := 1$. When $(j, k)$ has two different parents on scale $j - 1$ we have that $I_{j,k} \subset (I_{lpa(j-1,j,k)} \cup I_{rpa(j-1,j,k)})$

and we partition $I_{j,k}$ into three disjoint sub-intervals:

$$\begin{aligned} I_{j,k}^{(1)} &:= \left(I_{j,k} \setminus I_{rpa(j-1,j,k)}\right) && \subset I_{lpa(j-1,j,k)}, \\ I_{j,k}^{(2)} &:= \left(I_{j,k} \cap I_{lpa(j-1,j,k)} \cap I_{rpa(j-1,j,k)}\right) && \subset \left(I_{lpa(j-1,j,k)} \cap I_{rpa(j-1,j,k)}\right), \\ I_{j,k}^{(3)} &:= \left(I_{j,k} \setminus I_{lpa(j-1,j,k)}\right) && \subset I_{rpa(j-1,j,k)}. \end{aligned}$$

It follows that $I_{j,k} = I_{j,k}^{(1)} \cup I_{j,k}^{(2)} \cup I_{j,k}^{(3)}$. The fraction of $Y_{j,k}$ relating to $I_{j,k}^{(1)}$ may then depend only on $Y_{lpa(j-1,j,k)}$, the fraction of $Y_{j,k}$ relating to $I_{j,k}^{(2)}$ equally on both $Y_{lpa(j-1,j,k)}$ and $Y_{lpa(j-1,j,k)}$, and the fraction of $Y_{j,k}$ relating to $I_{j,k}^{(3)}$ only on $Y_{rpa(j-1,j,k)}$. Therefore, we define

$$\alpha_{j,k} := \frac{|I_{j,k}^{(1)}| + \frac{1}{2}|I_{j,k}^{(2)}|}{|I_{j,k}|}, \quad \beta_{j,k} := \frac{\frac{1}{2}|I_{j,k}^{(2)}| + |I_{j,k}^{(3)}|}{|I_{j,k}|}. \tag{11}$$

It is straightforward to see that $\alpha_{j,k} + \beta_{j,k} = 1$ as required.

## 4.2 Transition Noise Specifications

For an individual scale $j$, the transition noise $\{\boldsymbol{W}_{j,k}, k = 1, \dots N_j\}$ is specified as a linear Gaussian state space model with the following representation, where $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the (multivariate) normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$:

$$\begin{aligned} \boldsymbol{W}_{j,1} &\sim N\left(\boldsymbol{\mu}_{j,1}^{W}, \boldsymbol{\Sigma}_{j,1}^{W}\right), \\ \boldsymbol{W}_{j,k+1} &= \boldsymbol{B}_{j,k}\boldsymbol{W}_{j,k} + \boldsymbol{\Gamma}_{j,k}\boldsymbol{V}_{j,k},\ k = 1, \dots, N_j - 1, \\ \boldsymbol{V}_{j,k} &\sim N\left(\boldsymbol{0}, \boldsymbol{\Sigma}_{Vj}\right),\ k = 1, \dots, N_j - 1, \\ \boldsymbol{V}_{j,k} &\perp \boldsymbol{W}_{j,1}, \boldsymbol{V}_{j,k} \perp \boldsymbol{V}_{j,l},\ k, l = 1, \dots, N_j - 1,\ k \neq l, \end{aligned} \tag{12}$$

where $\boldsymbol{V}_{j,k} \in \mathbb{R}^{m_j^V}$ is a random vector. Matrices $\boldsymbol{\Gamma}_{j,k} \in \mathbb{R}^{m_j^W \times m_j^V}$ and $\boldsymbol{B}_{j,k} \in \mathbb{R}^{m_j^W \times m_j^W}$ may contain some unknown parameters to be estimated. Since $\{\boldsymbol{W}_{j,k}, k = 1, \dots N_j\}$ for fixed scale $j$ are not assumed to be serially independent we refer to the transition noise as colored transition noise. For reference in Section 5, where we discuss a linear state space representation of Model (7) - (12) on a given two-parent scale-to-scale Graph $\mathcal{G}^S$, we collect the following properties of the transition noise: Assumption (12) implies that $\boldsymbol{W}_{j,k}$ is a sum of independent normal random vectors $\boldsymbol{W}_{j,1}, \boldsymbol{V}_{j,1}, \dots, \boldsymbol{V}_{j,k-1}$ for $j = 1, \dots, J$, $k = 1, \dots, N_j$. Since $\boldsymbol{V}_{j,k} \perp \boldsymbol{W}_{j,1}, \boldsymbol{V}_{j,1}, \dots, \boldsymbol{V}_{j,k-1}$ it follows that

$$\boldsymbol{V}_{j,k} \perp \boldsymbol{W}_{j,1}, \dots, \boldsymbol{W}_{j,k}. \tag{13}$$

Assumptions (10) and (12) imply that

(14) $\{\boldsymbol{V}_{j,k}, j = 1, \ldots, J,\ k = 1, \ldots, N_j\}$ are serially independent.

From (10) and (14) it follows that

(15) $$\boldsymbol{V}_{j,k} \perp \boldsymbol{W}_{i,l},\ j, i = 1, \ldots, j,\ j \neq i,\ k = 1, \ldots, N_j - 1,\ l = 1, \ldots, N_i.$$

Since $Y_{j,k}$, $j = 1, \ldots, J$, $k = 1, \ldots, N_j$ is a linear function of a subset of $\boldsymbol{W}^j$ it follows from (15) that

(16) $$\boldsymbol{V}_{j,k} \perp Y_{i,l},\ i = 1, \ldots, j-1,\ l = 1, \ldots, N_i,\ k = 1, \ldots, N_j - 1.$$

Using (13) we get

(17) $$\boldsymbol{V}_{j,k} \perp Y_{j,1}, \ldots, Y_{j,k}.$$

The following specifications for $\{\boldsymbol{W}_{j,k}, k = 1, \ldots, N_j\}$, $j = 1, \ldots, J$ will be used:

### 4.2.1 Stationary Gaussian ARMA(p,q) process

In our applications we will see that stationary ARMA(p,q) models work quite well for $j = 2, \cdots, J-1$. The regression formulation of a Gaussian ARMA(p,q) model on scale $j \in \{1, \ldots, J\}$ is given by

(18) $$\mu_{j,k} = \sum_{i=1}^{p} a_{j,i}\mu_{j,k-i} + \zeta_{j,k} + \sum_{i=1}^{q} b_{j,i}\zeta_{j,k-i},\ \zeta_{j,k} \sim N(0, \sigma_j^2)\ iid.$$

for $k = 1, \ldots, N_j$. Let $\mathbf{a_j}(\mathbf{L}) := \mathbf{1} - \mathbf{a_{j,1}L} - \cdots - \mathbf{a_{j,p}L^p}$, $\mathbf{b_j}(\mathbf{L}) := \mathbf{1} + \mathbf{b_{j,1}L} + \cdots + \mathbf{b_{j,q}L^q}$, where $\mathbf{L}$ denotes the backshift operator. We assume that the polynomials $\mathbf{a_j}(\mathbf{x})$ and $\mathbf{b_j}(\mathbf{x})$ have no common zeros. Furthermore, the roots of $\mathbf{a_j}(\mathbf{x})$ lie outside the unique circle. As shown in Brockwell and Davis (1991) these are sufficient conditions for the ARMA(p,q) process to be causal. This property is useful for computation of the autocovariance function needed for initialization of the first state vector $\boldsymbol{W}_{j,1}$. Hamilton (1994) shows that a sufficient condition for the ARMA(p,q) process to be identifiable is that the roots of $\mathbf{b_j}(\mathbf{x})$ lie on or outside the unit circle. There are several linear state space versions of an ARMA model, see for example Durbin and Koopman (2001) and Brockwell and Davis (1991). Parameters to be estimated are

(19) $$\boldsymbol{\psi}_j := \left(a_1, \quad \ldots, \quad a_p, \quad b_1, \quad \ldots, \quad b_q, \quad \sigma_j^2\right)'.$$

For initialization we assume $\boldsymbol{W}_{j,1} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}_{j,1}^W\right)$, where the variance-covariance matrix $\boldsymbol{\Sigma}_{j,1}^W$ is obtained as discussed in Brockwell and Davis (1991).

### 4.2.2 I(1) Specification with stationary ARMA(p,q) noise

For $j = 1$ and $j = J$ we allow the transition noise to follow a trend modeled with an integrated model of order 1, denoted by $I(1)$, and stationary ARMA(p,q) noise:

$$\begin{aligned} \mu_{j,k+1} &= \mu_{j,k} + \nu_{j,k}, \\ \nu_{j,k} &= \sum_{i=1}^{p} a_i \nu_{j,k-i} + \zeta_{j,k} + \sum_{i=1}^{q} b_i \zeta_{j,k-i},\ \zeta_{j,k} \sim N(0, \sigma_j^2)\ iid. \end{aligned} \tag{20}$$

for $k = 1, \ldots, N_j$. The assumptions for $\mathbf{a_j}(\mathbf{x})$ and $\mathbf{b_j}(\mathbf{x})$ and the parameters to be estimated remain the same as for the ARMA(p,q) specification. For example, for $p = 1$ and $q = 1$ the corresponding state space representation of $\boldsymbol{W}_{j,k} = (\mu_{j,k}, \nu_{j,k}, b_{j,1}\zeta_{j,k})$ for $k = 1, \ldots, N_j$ is given by

$$\boldsymbol{W}_{j,k+1} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & a_{j,1} & 1 \\ 0 & 0 & 0 \end{pmatrix} \boldsymbol{W}_{j,k} + \begin{pmatrix} 0 \\ 1 \\ b_{j,1} \end{pmatrix} \zeta_{j,k+1} \text{ and } \zeta_{j,k} \sim N(0, \sigma_j^2)\ iid. \tag{21}$$

On the first scale $j = 1$ we consider a diffuse initialization of the form

$$\begin{aligned} \boldsymbol{W}_{1,1} &:= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \delta + \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \boldsymbol{\eta}_{1,0},\ \boldsymbol{\eta}_{1,0} \sim N_2\left(\mathbf{0}, var(\boldsymbol{\eta}_{1,0})\right), \\ \delta_\kappa &\sim N(\mathbf{0}, \kappa),\ \delta_\kappa \xrightarrow{D} \delta,\ \kappa \to \infty, \end{aligned} \tag{22}$$

where $\xrightarrow{D}$ denotes convergence in distribution. A detailed treatment of diffuse initialization is given in Durbin and Koopman (2001) and in Koopman and Durbin (2003). On scales $j = 2, \ldots, J$ we use the non-diffuse initialization for $\mu_{j,1}$ with $\mu_{j,k} \sim N\left(0, var(\nu_{j,k})\right)$ and where $\mu_{j,k}$ is assumed to be independent of the other elements of $\boldsymbol{W}_{j,1}$.

For both, the ARMA(p,q) and the I(1) specification with stationary ARMA(p,q) noise we define $\boldsymbol{G}_j$ so that $\mu_{j,k} = \boldsymbol{G}_j \boldsymbol{W}_{j,k}$ holds. It is shown in Chapter 6 of Högn (2005), that the parameters to be estimated are identifiable under certain assumptions which are satisfied for the graphs considered in our application.

# 5 Linear State Space Representation of Time Series Models using Directed Acyclic Graphs

We give now the main result of the paper:

**Theorem 1.** *Model* (7) - (12) *on a two-parents scale-to-scale graph* $\mathcal{G}^S$ *has a linear state space representation given by*

$$\text{(23)} \qquad \textit{Transition equation:} \qquad \boldsymbol{X}_{k+1} = \bar{\boldsymbol{A}}_k \boldsymbol{X}_k + \bar{\boldsymbol{\Gamma}}_k \boldsymbol{\eta}_k,$$

$$\text{(24)} \qquad \textit{Observation equation:} \qquad Z_k = \bar{\boldsymbol{C}}_k \boldsymbol{X}_k,$$

*for* $k = 1, \ldots, N_J$*, with the following assumptions for the system and the observation noise in* (23) - (24)*:*

$$\text{(25)} \qquad \begin{aligned} &\boldsymbol{X}_1 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \textit{ and } \boldsymbol{\eta}_k \sim N(\boldsymbol{0}, \boldsymbol{Q}_k), \\ &\boldsymbol{\eta}_k \perp \boldsymbol{X}_1, \boldsymbol{\eta}_k \perp \boldsymbol{\eta}_l \ (k \neq l). \end{aligned}$$

Proof: See appendix.

There is not a unique linear state space representation. We are not aware of standard software to capture this, therefore our aim was to find a linear state space representation, which is not too difficult to implement. The second aim was that computations have to be done efficiently and requiring little memory. The derivation of state space representation (23) - (24) is very technical and therefore given in the appendix. Once a state space representation has been found, we apply the Kalman filter and smoothing algorithms as discussed for example in Durbin and Koopman (2001) and Harvey (1987). Using the Kalman filter, the log-likelihood of the observations $\boldsymbol{Z}^{N_J}$ can then be computed by the so called prediction error decomposition. For $k = 1, \ldots, N_J$, let $E(Z_k|z_1, \ldots, z_{k-1})$ denote the conditional mean of $Z_k$ given the realizations of $Z_1, \ldots, Z_{k-1}$. Innovations $\theta_k$, $k = 1, \ldots, N_J$, are defined as $\theta_k := Z_k - E(Z_k|Z_1, \ldots, Z_{k-1})$ with $\Delta_k := var(\theta_k) = var(Z_k|Z_1, \ldots, Z_{k-1})$, which can be computed from the Kalman filter. The log-likelihood value for a specific value of the parameter vector $\boldsymbol{\psi}$ to be estimated is then given by

$$\text{(26)} \qquad \log L(\boldsymbol{\psi}; \boldsymbol{z}^N) = -\frac{N_J}{2} \log 2\pi - \frac{1}{2} \sum_{k=1}^{N_J} \log \Delta_k - \frac{1}{2} \sum_{k=1}^{N} \frac{\theta_k^2}{\Delta_k},$$

where $\theta_k$ and $\Delta_k$ depend on $\boldsymbol{\psi}$. Maximum likelihood (ML) estimation can then be conducted using iterative procedures. In our applications we have used for this the Broyden-Fletcher-Goldfarb-Shannon (BFGS) algorithm, see for example Fletcher (1987). As an alternative, one can conduct ML estimation using an EM algorithm as discussed in Chapter 5 of Högn (2005). Since its rate of convergence near the maximum is slower than numerical maximization we prefer numerical maximization. A further advantage of this representation is, that missing observations can easily be handled. For a missing observation $Z_k$, $k \in \{1, \ldots, N_J\}$, one sets $\theta_k := 0$ and $\Delta_k := 0$, while the system equation for the corresponding state vector $\boldsymbol{X}_k$ remains unchanged.

# 6 Applications to Financial Time Series

We discuss now the overlapping interval arrangement for the linear stochastic Model (7) - (12) over a two-parent scale-to-scale graph $\mathcal{G}^S$ used in our application. The sample period lasts from February 2002 to October 2002 with regular trading between $9:30$ a.m. and $4:00$ p.m. from Monday until Friday each week excluding holidays.

Huang et al. (2002) introduce the so-called mass balance. They show that the mass balance results in a very restrictive correlation structure of the data $Z_k$. Therefore we consider now overlapping intervals, where mass balance does not need to hold. For this we introduce some adjustments to weeks, trading days and hours as follows:

On scale $j = 1$ we consider **adjusted weeks** defined by: Monday until Friday, excluded holidays, plus half a regular trading day before ($12:45$ - $4:00$ p.m.) and the first half of the following trading day ($09:30$a.m. - $12:45$ p.m.). Therefore weeks overlap and we call them adjusted weeks resulting in $N_1 = 39$ adjusted weeks.

On scale $j = 2$ we consider **adjusted trading days** defined by: Actual trading plus one hour ($3:00$ - $4:00$ p.m.) the trading day before plus one hour ($09:30$ - $10:30$a.m.) the following trading day resulting in $N_2 = 189$ adjusted days.

On scale $j = 3$ we consider **adjusted hours** defined by: adjusted hour $:=$ one hour $+$ 10min. before $+$ 10min. after giving $N_3 = 1229$ adjusted hours.

On the finest scale $j = 4 =: J$ we consider **overlapping five minute intervals**: Each 5 minute interval overlaps with the consecutive 5 minute interval by 2 minutes, yielding $N_4 = 24570$ overlapping intervals. As discussed in Andersen et al. (2001) for shorter intervals the distortions related to market microstructure effects may become too severe.

As observations we use the trading duration time weighted averages of log-prices over the five minute intervals defined in (1). We disregard opening and closing effects by removing for each regular trading day the first 30 minutes and the last 15 minutes. The data corresponding to the five minute intervals that lie within this time span are then considered as missing observations. This leaves a sample size of 21933 observations, enumerated by $k \in K_O$. Figure 2 shows the corresponding averages over adjusted weeks and over overlapping five minute intervals. We see, that the main dynamics in the trend are already captured considering trading duration time weighted averages over adjusted weeks. Therefore, a diffuse initialization of the state vectors is used only on scale $j = 1$ and a non-diffuse initialization is used on the further scales.

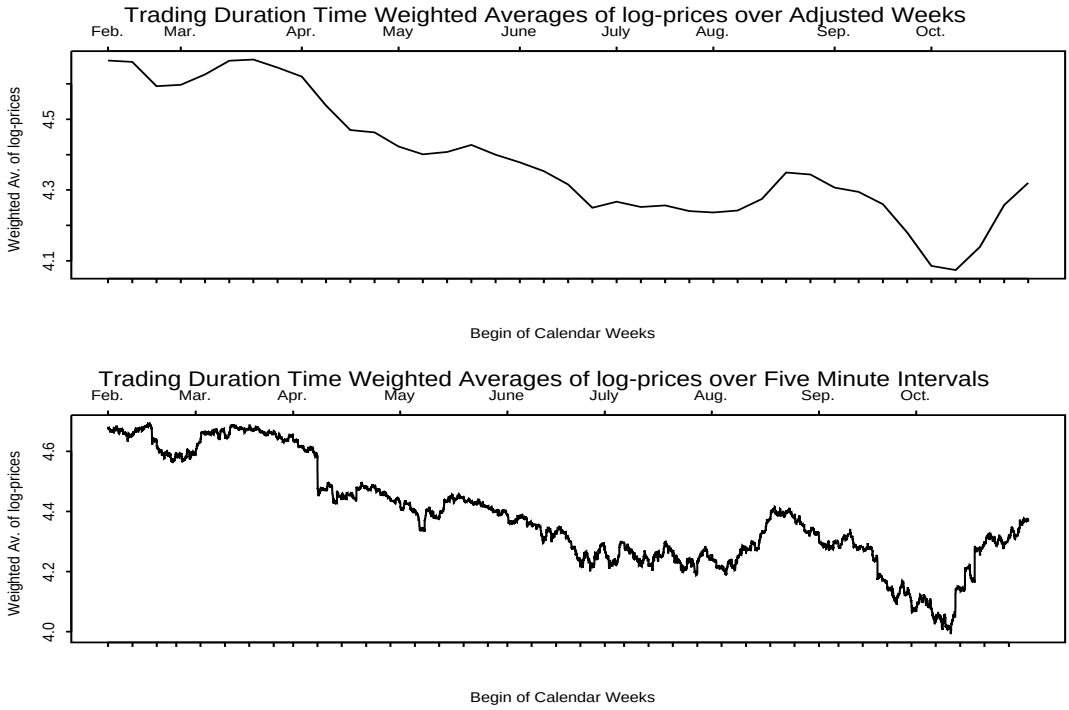


Figure 2: Trading Duration Time Weighted Averages of Log-prices over Adjusted Weeks and Overlapping five Minute Intervals

As a starting model we consider the following specifications:

- scale 1: I(1) model with stationary ARMA(1,1) noise and diffuse initialization as given in (22),
- scale 2: stationary ARMA(1,1) model,
- scale 3: stationary ARMA(1,1) model,
- scale 4: I(1) model with stationary ARMA(1,1) noise and a non-diffuse initialization.

The variance parameter $\sigma_3^2$ on the third scale corresponding to adjusted hours was estimated close to zero. This indicates that the third scale can be omitted, which was done for the models to follow. For model comparison we apply the Akaike information criterion (AIC) (see for example Harvey (1987)). Since our observations $z_k$, $k = 1, \ldots, N_J$, are univariate the AIC is given by AIC $:= -2 \log L(\hat{\boldsymbol{\psi}}; \boldsymbol{z}^{N_J}) + 2n_{\boldsymbol{\psi}}$ where $n_{\boldsymbol{\psi}}$ denotes the number of parameters to be estimated. Models with more parameters obtain a larger penalty. A model with a smaller value of the AIC is preferred. We summarize now the estimated models, where the I(1) with ARMA(1,1) noise on

scale 1 and the ARMA(1,1) specification on scale 2 are held fixed and on scale 4 I(1) models with ARMA(p,q) noise for several orders of $p$ and $q$ are considered. Here $AIC_i(n_{\boldsymbol{\psi}})$ denotes the AIC value of model $i$ with $n_{\boldsymbol{\psi}}$ parameters to be estimated:

- Model 2: I(1) model with ARMA(1,1) noise, $AIC_2(9) = -255330$.
- Model 3: I(1) model with ARMA(2,1) noise, $AIC_3(10) = -255660$.
- Model 4: I(1) model with ARMA(1,2) noise, $AIC_4(10) = -256654$.
  Since $b_{4,1}$ is estimated close to 1 we set for the following models $b_{4,1} := 1$.
- Model 5: I(1) model with ARMA(2,2) noise, $b_{4,1} := 1$, $AIC_5(10) = -256768$.
- Model 6: I(1) model with ARMA(3,2) noise, $b_{4,1} := 1$, $AIC_6(11) = -256830$.
- Model 7: I(1) model with ARMA(1,3) noise, $b_{4,1} := 1$, $AIC_7(10) = -256700$.
- Model 8: I(1) model with ARMA(2,3) noise, $b_{4,1} := 1$, $AIC_8(11) = -256768$.

The AIC values for the different fitted models do not vary very much. However, the model with the lowest AIC is Model 6 with $p = 3$ and $q = 2$. The estimated parameters for Model 6 are on scales $j = 1, 2$ :

$$\begin{aligned} &\hat{a}_{1,1} = 0.677, \qquad &&\hat{b}_{1,1} = 0.950, \qquad &&\hat{\sigma}_1^2 = 3.618 \cdot 10^{-9}, \\ &\hat{a}_{2,1} = -0.068, \qquad &&\hat{b}_{2,1} = -0.969 \qquad &&\hat{\sigma}_2^2 = 6.886 \cdot 10^{-7}. \end{aligned}$$

and on scale $j = 4$:

$$\hat{a}_{4,1} = 0.072,\ \hat{a}_{4,2} = 0.014,\ \hat{a}_{4,3} = 0.107,\ \hat{b}_{4,2} = 0.590,\ \hat{\sigma}_4^2 = 4.537 \cdot 10^{-7}.$$

For a diagnostic checking of Model 6 we use the estimated standardized innovations of the state space representation (23) - (25) defined as

$$\hat{\theta}_k^s := \frac{\hat{\theta}_k}{\sqrt{\hat{\boldsymbol{\Delta}}_k}},\ k \in K_O, \tag{27}$$

where $\hat{\theta}_k(\hat{\boldsymbol{\Delta}}_k)$ is the innovation $\theta_k$ ( the variance of $\theta_k$) evaluated at the estimated parameters. Using the independence of the system noise it follows that $\{\hat{\theta}_k^s, k \in K_O\} \sim N(0,1)$ i.i.d.. The diagnostic plots for Model 6 are given in Figure 3. In the time series panel (top left), a value greater than 2 or smaller than -2 (horizontal lines) indicates that this value is not a realization of the $N(0,1)$ distribution at the 95% confidence level and therefore the corresponding observation cannot be appropriately represented by the model under consideration. The normal

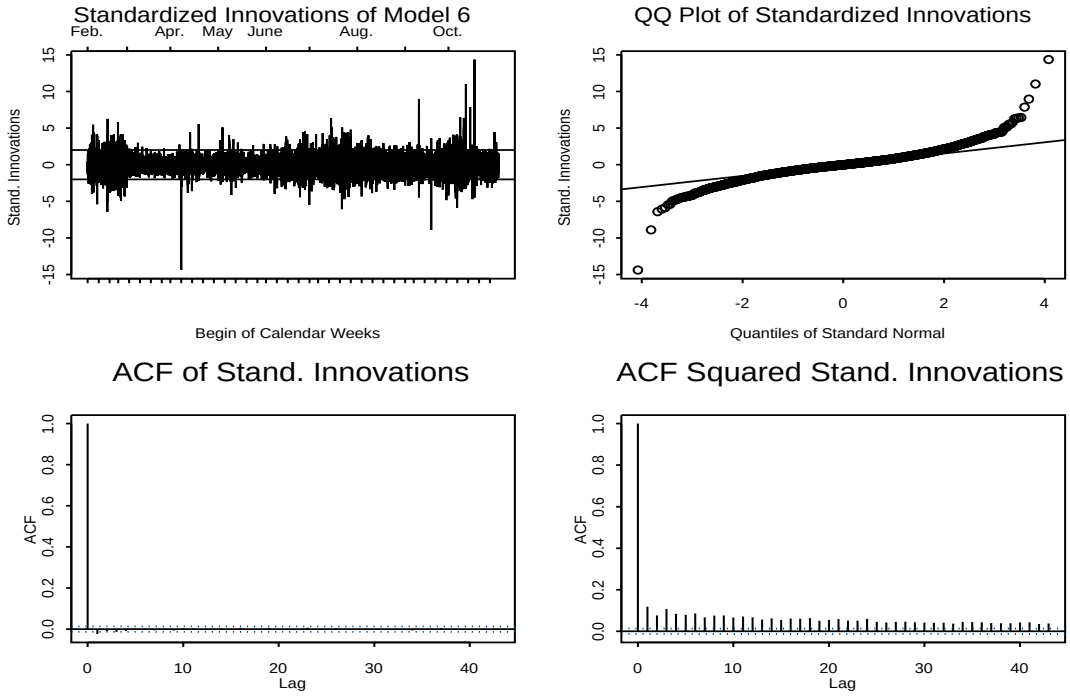


Figure 3: Diagnostic Plots for Standardized Innovations of Model 6

quantile-quantile (QQ) panel (top right) shows that the empirical distribution of the standardized innovations is symmetric but heavy tailed. The ACF panel (bottom left) with approximate 95 % confidence intervals (dotted lines) gives evidence that the empirical autocorrelation of the standardized innovations is negligible. The squared standardized innovations may be viewed as as a measure for the second order moment of the standardized innovations. An autocorrelation in the squared standardized innovations (bottom right) then indicates that the variance conditional on its past history may change over time and giving rise to volatility clustering. Moreover, the autocorrelation decays only slowly. Therefore the assumption of a time invariant variance on the finest scale $j = 4$ may not be appropriate and the following model extension will be considered.

## 6.1 fiEGARCH-specification of Transition Variance on finest Scale

Define $\Delta_k^* := \frac{\Delta_k}{\sigma_4^2}$. Following the approach in Koopman and Bos (2004) we parameterize the log-likelihood (26) for $N_O :=$ card $K_O$ as

$$\log L(\boldsymbol{\psi}; \theta_1, \ldots, \theta_{N_4}) = -\frac{N_O}{2}\log(2\pi) - \frac{1}{2}\sum_{k \in K_O}\left(\log \sigma_4^2 + \log \Delta_k^* + \frac{\theta_k^2}{\sigma_4^2 \Delta_k^*}\right).$$

Applying the concept of conditional Gaussian state space models discussed in Harvey (1987) to the state space representation (23) - (25) allows for dependence of the system noise $\boldsymbol{\epsilon}_k$ on observations $\boldsymbol{Z}^{k-1} := (Z_1, \ldots, Z_{k-1})'$. Using (21) we get that $\zeta_{4,k+1} \in \boldsymbol{V}_{4,k}$ and from (68) it follows that $\boldsymbol{V}_{4,k} \in \boldsymbol{\eta}_k$, $k = 1, \ldots, N_4 - 1$. Therefore we have that $\zeta_{4,k+1} \in \boldsymbol{\eta}_k$ and we assume that $\zeta_{4,k}|\boldsymbol{Z}^{k-1} \sim N\left(0, \boldsymbol{Q}_k(\boldsymbol{Z}^{k-1})\right)$, where the notation $\boldsymbol{Q}_k(\boldsymbol{Z}^{k-1})$ indicates dependence on $\boldsymbol{Z}^{k-1}$. Since the observations are assumed to be normally distributed the innovations $\theta_k$ is a linear function of $Z_1, \ldots, Z_k$ for $k \in K_O$ we can write equivalently $\zeta_{4,k}|\boldsymbol{\theta}^{k-1} \sim N\left(0, \sigma^2_{4,k}\right)$, $k \in K_O$, where $\sigma^2_{4,k} := Var(\zeta_{4,k}|\boldsymbol{Z}^{k-1})$ depends on $\boldsymbol{\theta}^{k-1} := (\theta_1, \ldots, \theta_{k-1})'$.

To allow for long memory indicated by the slow decay of the autocorrelation of the squared standardized innovations we consider a fiEGARCH(1,1) (fractionally integrated exponential GARCH) specification for $\sigma^2_{4,k}$, as introduced by Bollerslev and Mikkelsen (1996), where $h_k := log(\sigma^2_{4,k})$ is modeled in terms of a fractionally integrated ARMA(1,1) process. There are several ways to write a fiEGARCH model. We use the following parameterization given in Zivot and Wang (2003):

$$\theta_k = \sigma_{4,k}\sqrt{\Delta^*_k}\,\epsilon_k,\ \epsilon_k \sim N(0,1)\ iid.,$$
$$(1-\phi)(1-L)^d h_k = m + \varphi(|\epsilon_{k-1}| + \gamma\epsilon_{k-1}).$$

Here $\gamma \in \mathbb{R}$ is a leverage parameter to accommodate an asymmetric relation between innovations and volatility, $m$ is a parameter for the overall mean of $h_k$ and

$$(1-\phi)(1-L)^d h_k := (1-L)^d h_k - \phi(1-L)^d h_{k-1}.$$

For $d > -1$ the fractional integration operator $(1-L)^d$ is defined as

$$(1-L)^d := \sum_{l=0}^{\infty}(-1)^l \frac{\Gamma(d+1)}{\Gamma(l+1)\Gamma(d-l+1)}. \tag{28}$$

Here $\Gamma(\cdot)$ denotes the gamma function. With $\boldsymbol{\psi}_G := (\phi, \varphi, \gamma, d, m)'$ the parameter vector to be estimated is now given as $\boldsymbol{\psi}^* := (\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \boldsymbol{\psi}_4, \boldsymbol{\psi}_G)'$. For this we maximize the conditional likelihood given the initial conditions $\mathcal{F}_0$ discussed below:

$$\log L(\boldsymbol{\psi}^*; \theta^{N_J}|\mathcal{F}_0) = -\frac{N_0}{2}\log(2\pi) - \frac{1}{2}\sum_{k\in K_O}\left(\log\sigma^2_{4,k} + \log\Delta^*_k + \frac{\theta^2_k}{\sigma^2_{4,k}\Delta^*_k}\right). \tag{29}$$

For estimation the truncation of the infinite distributed lags in (28) is necessary. Since the fractional differencing operator is designed to capture the long-memory feature of the process, truncating at too low a lag may destroy important long-run dependencies. Bollerslev and Mikkelsen (1996) propose the truncation lag to be $1000$. $\mathcal{F}_0$ is then given by the pre-sample values $\epsilon_0 :=$

$0,\ h_k := \log(\hat{\sigma}_4^2),\ k = -1000, \ldots, 0$, where $\hat{\sigma}_4^2$ is obtained from the fitted Model 6. The parameters of Model 6 with a fiEGARCH(1,1) specification were estimated to be on scales 1,2:

$$\begin{aligned} \hat{a}_{1,1} &= 0.678, & \hat{b}_{1,1} &= 0.950, & \hat{\sigma}_1^2 &= 3.618 \cdot 10^{-9}, \\ \hat{a}_{2,1} &= -0.068, & \hat{b}_{2,1} &= -0.969, & \hat{\sigma}_2^2 &= 2.115 \cdot 10^{-6}, \end{aligned}$$

and on scale $j = 4$ with $b_{4,1} := 1$:

$$\hat{a}_{4,1} = 0.156,\ \hat{a}_{4,2} = -0.021,\ \hat{a}_{4,3} = 0.136,\ \hat{b}_{4,2} = 0.654,\ \hat{\sigma}_4^2 = 1.476 \cdot 10^{-7},$$

and the parameters for the fiEGARCH(1,1) extension:

$$\hat{m} = -0.353,\ \hat{\phi} = 0.092,\ \hat{\varphi} = 0.243,\ \hat{\boldsymbol{\gamma}} = -0.028,\ \hat{d} = 0.539.$$

The log-likelihood of the fitted model is $\log L = 131\,682.3$ giving an AIC value for 15 parameters of $AIC(15) = -263\,306.3$, which is clearly lower than the AIC value for Model 6. Comparing the parameter estimates for both models without and with the fiEGARCH(1,1) extension we see on scales 1,2 only minor changes, except for $\hat{\sigma}_2^2$. The estimated parameters relating to scale 4 change distinctly. In particular, $\hat{\sigma}_4^2$ decreases when considering the fiEGARCH(1,1) extension.

For diagnostics we utilize the standardized residuals given by

$$\hat{\theta}_k^G := \frac{\hat{\theta}_k}{\hat{\sigma}_{4,k}\sqrt{\hat{\boldsymbol{\Delta}}_k^*}},\ k \in K_O,$$

which should again be i.i.d. standard normally distributed. Figure 4 shows a time series plot of the standardized residuals and the estimated standard deviations on the finest scale $j = 4$, $\hat{\sigma}_{4,k}$, $k \in K_O$. A comparison to the normal QQ plot given in Figure 5 for $\hat{\theta}_k^S$ and $\hat{\theta}_k^G$, $k \in K_O$, exhibits an improvement in the tail behavior but the empirical distribution of $\hat{\theta}_k^G$ is still heavy tailed. Moreover, the distinct autocorrelation of the squared innovations almost vanishes after applying the fiEGARCH extension (see Figure 6).

# 7 Summary and Conclusions

In this paper we have introduced time series models with colored transition noise over a directed acyclic two-parents scale-to-scale graph. We have derived a linear Gaussian state space representation for these models which allows for efficient computation of the likelihood and for easy treatment of missing observations. We have applied these models to trading duration time weighted

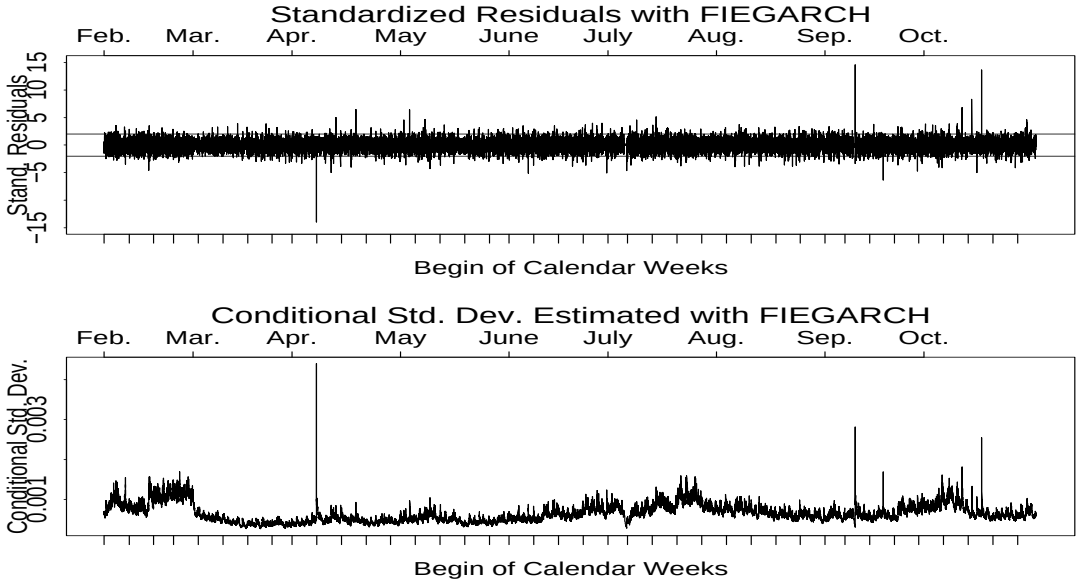


Figure 4: Time Series Plot of the Standardized Residuals $\hat{\theta}_k^G$ and the Estimated $\hat{\sigma}_{4,k}$

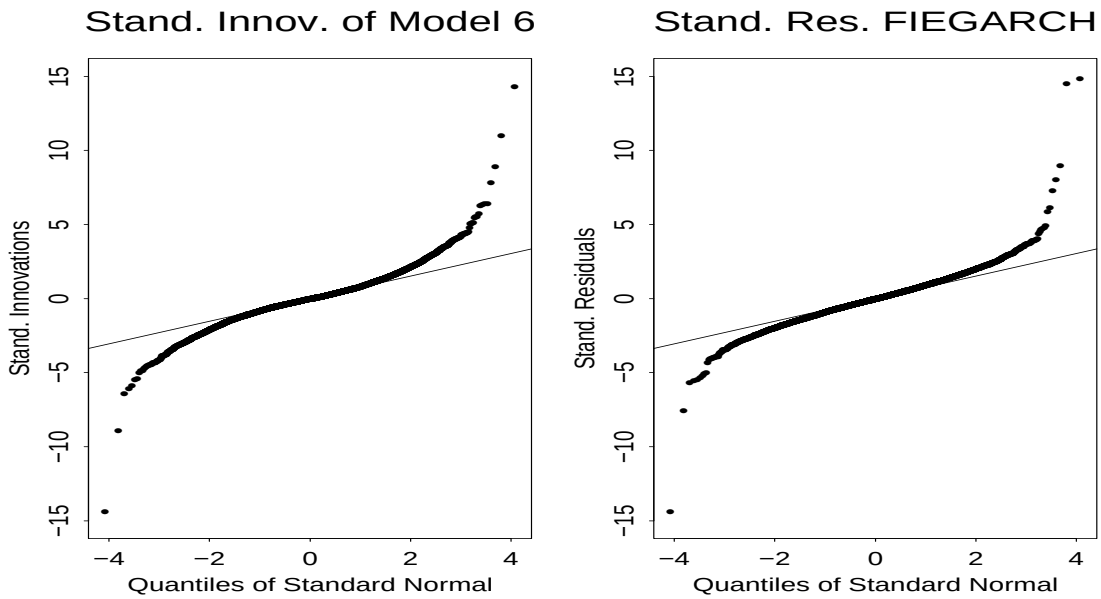


Figure 5: Normal QQ Plot of the Standardized Innovations $\hat{\theta}_k^S$ of Model 6 and Standardized Residuals $\hat{\theta}_k^G$ of Model 6 with fiEGARCH(1,1) Extension

averages of log-prices over overlapping five minute intervals for IBM shares traded at the New York Stock Exchange from February until October 2002. Since the standardized innovations cor-

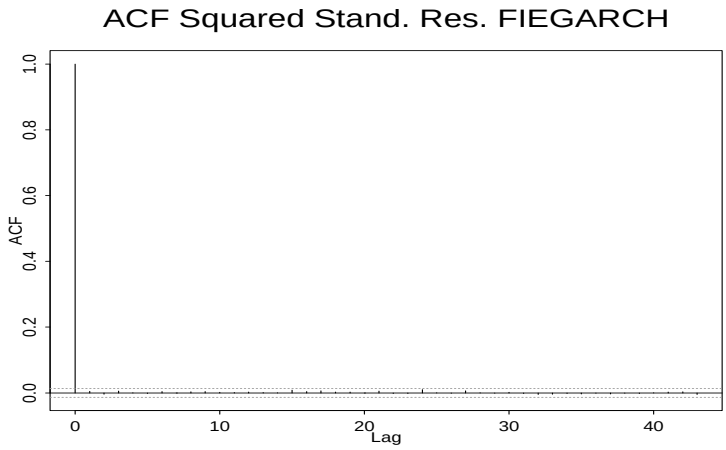


Figure 6: ACF of Squared Standardized Residuals $\hat{\theta}_k^G$ with fiEGARCH(1,1) Extension

responding to a linear Gaussian model are heavy tailed and the squared standardized innovations exhibit a distinct and slowly decaying autocorrelation, we have allowed the transition variance relating to the finest scale to follow a fiEGARCH-process. This leads to a conditionally Gaussian state space representation. The tail behavior of the standardized residuals was improved though they remain heavy tailed. Therefore, the standard normal error distribution of the fiEGARCH specification may be replaced by a more heavy tailed student-t distribution. However, for this case the corresponding state space model is not conditionally Gaussian. As a consequence, the prediction error decomposition given in (26) and (29) is not valid any more. One common approach is quasi maximum likelihood estimation, where the state space model is still taken to be Gaussian. Instead of a fiEGARCH specification a stochastic volatility model as introduced by Taylor (1986) may be used. The corresponding state space model is not linear. A possible approach for parameter estimation is then maximum likelihood evaluation using importance sampling as discussed in Durbin and Koopman (2001), Durbin and Koopman (2000) and Durbin and Koopman (1997). As an alternative, a Bayesian point of view may be considered, leading to estimation by Markov Chain Monte Carlo (MCMC) methods (see for example Shephard and Pitt (1997) and Bos and Shephard (2004)).

Moreover, explanatory variables can be incorporated in the state space specification of the transition noise. On the finest scale one may consider explanatory variables relating to the market microstructure, such as functions of the bid-offer spread, the trading volume and the trading duration time. On a coarser scale other explanatory variables, relating to the specific company or the

branch to which the specific firm corresponds to, as well as indicators for the general economical development may be used.

In our applications, we considered one specific interval arrangement, but there are far more possibilities. For example, the length of sub-intervals on an individual scale can be chosen in many different ways and the scale-wise structure allows for different ways of combinations. Therefore, the question arises whether there is an optimal interval arrangement and a corresponding graph in terms of model fit. The model fit will also depend on the state space specifications used for the transition noise related to an individual scale.

An open question is, whether the parameter estimates are asymptotically normally distributed. As discussed in Harvey (1987), Subsection 4.5.1, a first assumption is that the state space representation has time invariant system matrices. The state space representation derived in Section 5, however, has time varying system matrices. It may not be straightforward, to find a time invariant state space representation.

Concerning our applications we have restricted ourselves to a small class of state space models on individual scales. The general formulation of the model over a given graph $\mathcal{G}^S$ allows for a much broader class of models and a wide variety of model extensions.

# Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 386, Statistical Analysis of Discrete Structures.

# References

Andersen, T., T. Bollerslev, F. Diebold, and P. Labys (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

Bollerslev, T. and H. Mikkelsen (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics 73*, 151–184.

Bos, C. and N. Shephard (2004). Inference for adaptive time series models: Stochastic volatility and conditionally Gaussian state space form. Research Report, Nuffield College, University of Oxford. Brighton, U.K.

Brockwell, P. and R. Davis (1991). *Time Series: Theory and Methods* (2 ed.). Springer.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Dacorogna, M., R. Gencay, U. A. Müller, R. B. Olsen, and O. V. Pictet (2001). *An Introduction to High Frequency Finance*. Academic Press.

Durbin, J. and S. Koopman (1997). Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika 84*(3), 669–684.

Durbin, J. and S. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of Royal Statistical Society, Series B 62*, 3–56.

Durbin, J. and S. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

Engle, R. (2000). The econometrics of ultra-high-frequency data. *Econometrica 68*(1), 1–22.

Fletcher, R. (1987). *Practical Methods of Optimization*. Wiley.

Goodhart, C. and M. O'Hara (1997). High frequency data in financial markets: Issues and applications. *Journal of Empirical Fincance 4*, 73–114.

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

Harvey, A. C. (1987). *Forecasting,Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hautsch, N. and W. Pohlmeier (2002). Econometric analysis of financial transaction data: Pitfalls and opportunities. *Allgemeines Statistisches Archiv 86*, 5–30.

Högn, R. (2005). *Multiresolution of Long Time Series With Applications to Finance*. Ph. D. thesis, University of Technology Munich.

Huang, H. and N. Cressie (2001). Multiscale graphical modeling in space: Applications to command and control. In M. Moore (Ed.), *Spatial Statistics: Methological Aspects and Applications*, pp. 83–113. Springer.

Huang, H., N. Cressie, and J. Gabrosek (2002). Fast, resolution-consistent spatial prediction of global processes from satellite data. *Journal of Computational and Graphical Statistics 11*(1), 63–88.

Koopman, S. and C. Bos (2004). Time series models with a common stochastic variance for analysing economic time series. *Journal of Business and Economic Statistics 22*(3), 346–357.

Koopman, S. J. and J. Durbin (2003). Filtering and smoothing of state vector for diffuse state space models. *Journal of Time Series Analysis 24*(1), 85–98.

Lauritzen, S. (1996). *Graphical Models*. Clarendon Press.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*(3), 653–667.

Shephard, N. E. (2005). *Stochastic Volatility: Selected Readings*. Oxford University Press.

Taylor, S. (1986). *Modeling Financial Time Series*. John Wiley.

Zivot, E. and J. Wang (2003). *Modeling Financial Time Series with S-Plus*. Springer Verlag.

# A Proof of Theorem 1

We discuss a quite general procedure to transform Model (7) - (12) on a two-parents scale-to-scale graph $\mathcal{G}^S$ into linear state space form given by (23) - (25). For simplicity we restrict our proof to $J = 4$, since in our application this is the maximum number of scales. Application of the following methods for $J < 4$ and $J > 4$ is straightforward. To ease notation we make two assumptions for $\mathcal{G}^S$ concerning the construction of the first state vector $\boldsymbol{X}_1$: The node $(j, 1)$ has the unique parent $(j-1, 1)$, $j = 2, 3, 4$. Therefore (8) can be written for $k = 1$ as

$$Y_{j,1} = Y_{j-1,1} + \boldsymbol{G}_j \boldsymbol{W}_{j,1},\ j = 2, \ldots, 4. \tag{30}$$

(31) $(j, 1)$ has at least three children $(j+1, 1), (j+1, 2), (j+1, 3)$ for $j = 1, 2, 3$.

Restrictions (30) and (31) are satisfied for the graph $\mathcal{G}^S$ considered in our application but relaxations are possible.

## A.1 Initializations

We construct now the first state vector $\boldsymbol{X}_1$ and its variance-covariance matrix $\boldsymbol{\Sigma}_1$ for known $\boldsymbol{\Sigma}_{j,1}^W$, $j = 1, \ldots, 4$. Let $\mathbf{0}_n$ denote a zero column vector of dimension $n$, $n \in \mathbb{N}$. To simplify the construction of $\boldsymbol{\Sigma}_1$ we define

$$\boldsymbol{X}_{-2} := \begin{pmatrix} \boldsymbol{W}_{1,1} \\ \mathbf{0}_{m_1^W} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \boldsymbol{W}_{2,1} \\ \boldsymbol{W}_{3,1} \\ \boldsymbol{W}_{4,1} \end{pmatrix}, \quad \boldsymbol{X}_{-1} := \begin{pmatrix} \boldsymbol{W}_{1,1} \\ \boldsymbol{W}_{1,2} \\ Y_{2,1} \\ 0 \\ 0 \\ 0 \\ 0 \\ \boldsymbol{W}_{2,2} \\ \boldsymbol{W}_{3,1} \\ \boldsymbol{W}_{4,1} \end{pmatrix}, \quad \boldsymbol{X}_0 := \begin{pmatrix} \boldsymbol{W}_{1,1} \\ \boldsymbol{W}_{1,2} \\ Y_{2,1} \\ Y_{2,2} \\ Y_{3,1} \\ 0 \\ 0 \\ \boldsymbol{W}_{2,3} \\ \boldsymbol{W}_{3,2} \\ \boldsymbol{W}_{4,1} \end{pmatrix}, \quad \boldsymbol{X}_1 := \begin{pmatrix} \boldsymbol{W}_{1,1} \\ \boldsymbol{W}_{1,2} \\ Y_{2,1} \\ Y_{2,2} \\ Y_{3,1} \\ Y_{3,2} \\ Y_{4,1} \\ \boldsymbol{W}_{2,3} \\ \boldsymbol{W}_{3,3} \\ \boldsymbol{W}_{4,2} \end{pmatrix}. \tag{32}$$

For the remainder of this paper we omit the dimensions of matrices $\mathbf{0}$ and $\boldsymbol{I}$ to ease notation. Using (10) the variance-covariance matrix of $\boldsymbol{X}_{-2}$ has a block diagonal form given by

$$\boldsymbol{\Sigma}_{-2}:=var(\boldsymbol{X}_{-2})=\begin{pmatrix} \boldsymbol{\Sigma}_{1,1}^{W} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{\Sigma}_{2,1}^{W} & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{\Sigma}_{3,1}^{W} & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{\Sigma}_{4,1}^{W} \end{pmatrix}. \tag{33}$$

Since $Z_1 = Y_{4,1}$ the first observation $Z_1$ is directly related to state vector $\boldsymbol{X}_1$ by

$$Z_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \boldsymbol{X}_1 =: \bar{\boldsymbol{C}}\boldsymbol{X}_1. \tag{34}$$

We show that transition from $\boldsymbol{X}_k$ to $\boldsymbol{X}_{k+1}$, $k = -2, -1, 0$ can be expressed in the form $\boldsymbol{X}_{k+1} = \bar{\boldsymbol{A}}_k\boldsymbol{X}_k + \bar{\boldsymbol{\Gamma}}_k\boldsymbol{\eta}_k$. Consider first the transition from $\boldsymbol{X}_{-2}$ to $\boldsymbol{X}_{-1}$. Using (12) we have

$$\boldsymbol{W}_{1,2} = \boldsymbol{B}_{1,1}\boldsymbol{W}_{1,1} + \boldsymbol{\Gamma}_{1,1}\boldsymbol{V}_{1,1}. \tag{35}$$

From (8) together with (30) it follows with $\alpha_{2,1} := 1$ and $Y_{1,1} = \boldsymbol{G}_1\boldsymbol{W}_{1,1}$ that

$$Y_{2,1} = Y_{1,1} + \boldsymbol{G}_2\boldsymbol{W}_{2,1} = \boldsymbol{G}_1\boldsymbol{W}_{1,1} + \boldsymbol{G}_2\boldsymbol{W}_{2,1}. \tag{36}$$

Using (12) gives

$$\boldsymbol{W}_{2,2} = \boldsymbol{B}_{2,1}\boldsymbol{W}_{2,1} + \boldsymbol{\Gamma}_{2,1}\boldsymbol{V}_{2,1}. \tag{37}$$

Combining Equations (35) - (37) and noting that $\boldsymbol{W}_{1,1}$, $\boldsymbol{W}_{3,1}$ and $\boldsymbol{W}_{4,1}$ are elements of $\boldsymbol{X}_{-2}$ and $\boldsymbol{X}_{-1}$ then gives the matrix equation

$$\boldsymbol{X}_{-1} =: \bar{\boldsymbol{A}}_{-2}\boldsymbol{X}_{-2} + \bar{\boldsymbol{\Gamma}}_{-2}\boldsymbol{\eta}_{-2} \text{ with explicit form} \tag{38}$$

$$\begin{pmatrix} \boldsymbol{W}_{1,1} \\ \boldsymbol{W}_{1,2} \\ Y_{2,1} \\ \mathbf{0} \\ \boldsymbol{W}_{2,2} \\ \boldsymbol{W}_{3,1} \\ \boldsymbol{W}_{4,1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & 0 & 0 & 0 & 0 & \\ \boldsymbol{B}_{1,1} & 0 & 0 & 0 & 0 & \\ \boldsymbol{G}_1 & 0 & \boldsymbol{G}_2 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & & \boldsymbol{B}_{2,1} & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{I} & 0 & \\ 0 & 0 & 0 & 0 & \boldsymbol{I} & \end{pmatrix} \begin{pmatrix} \boldsymbol{W}_{1,1} \\ \mathbf{0} \\ \mathbf{0} \\ \boldsymbol{W}_{2,1} \\ \boldsymbol{W}_{3,1} \\ \boldsymbol{W}_{4,1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \boldsymbol{\Gamma}_{1,1} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & \boldsymbol{\Gamma}_{2,1} \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{V}_{1,1} \\ \boldsymbol{V}_{2,1} \end{pmatrix}.$$

Transition from $\boldsymbol{X}_{-1}$ to $\boldsymbol{X}_0$ is constructed in a similar manner. Note that $\boldsymbol{W}_{1,1}$, $\boldsymbol{W}_{1,2}$ and $\boldsymbol{W}_{2,2}$ are elements of state vector $\boldsymbol{X}_{-1}$. Restriction (31) then implies $(1,1) = lpa(1,2,2)$ and it follows that $(Y_{1,1} \cup Y_{1,2}) \supseteq (Y_{lpa(1,2,2)} \cup Y_{rpa(1,2,2)})$, where we set $\beta_{2,2} := 0$ if node $(1,2)$ has the unique parent $(1,1)$. Using (8) then yields

$$Y_{2,2} = \alpha_{2,2} Y_{1,1} + \beta_{2,2} Y_{1,2} + \boldsymbol{G}_2 \boldsymbol{W}_{2,2} \tag{39}$$

Using (30) yields

$$Y_{3,1} = Y_{2,1} + \boldsymbol{G}_3 \boldsymbol{W}_{3,1}, \tag{40}$$

where $Y_{2,1}$ and $\boldsymbol{W}_{3,1}$ are elements of $\boldsymbol{X}_{-1}$. Applying (12) we get

$$\boldsymbol{W}_{2,3} = \boldsymbol{B}_{2,2} \boldsymbol{W}_{2,2} + \boldsymbol{\Gamma}_{2,2} \boldsymbol{V}_{2,2}, \boldsymbol{W}_{3,2} = \boldsymbol{B}_{3,1} \boldsymbol{W}_{3,1} + \boldsymbol{\Gamma}_{3,1} \boldsymbol{V}_{3,1}, \tag{41}$$

where $\boldsymbol{W}_{3,1} \in \boldsymbol{X}_{-1}$. Combining (39) - (41) and noting that $\boldsymbol{W}_{4,1}$ is element of $\boldsymbol{X}_{-1}$ and $\boldsymbol{X}_0$ we get the matrix equation

$$\boldsymbol{X}_0 = \bar{\boldsymbol{A}}_{-1} \boldsymbol{X}_{-1} + \bar{\boldsymbol{\Gamma}}_{-1} \boldsymbol{\eta}_{-1} \text{ with explicit form} \tag{42}$$

$$\begin{pmatrix} \boldsymbol{W}_{1,1} \\ \boldsymbol{W}_{1,2} \\ Y_{2,1} \\ Y_{2,2} \\ Y_{3,1} \\ 0 \\ \boldsymbol{W}_{2,3} \\ \boldsymbol{W}_{3,2} \\ \boldsymbol{W}_{4,1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \boldsymbol{I} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{I} & 0 & 0 & 0 & 0 \\ \alpha_{2,2} & \beta_{2,2} & 0 & 0 & \boldsymbol{G}_2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \boldsymbol{G}_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{B}_{2,2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \boldsymbol{B}_{3,1} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{W}_{1,1} \\ \boldsymbol{W}_{1,2} \\ Y_{2,1} \\ 0 \\ 0 \\ 0 \\ \boldsymbol{W}_{2,2} \\ \boldsymbol{W}_{3,1} \\ \boldsymbol{W}_{4,1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \boldsymbol{\Gamma}_{2,2} & 0 \\ 0 & \boldsymbol{\Gamma}_{3,1} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{V}_{2,2} \\ \boldsymbol{V}_{3,1} \end{pmatrix}.$$

For transition from $\boldsymbol{X}_0$ to $\boldsymbol{X}_1$ we note that (30) gives $Y_{4,1} = Y_{3,1} + \boldsymbol{G}_4 \boldsymbol{W}_{4,1}$. In a similar manner as before we get the matrix equation

$$\boldsymbol{X}_1 = \bar{\boldsymbol{A}}_0 \boldsymbol{X}_0 + \bar{\boldsymbol{\Gamma}}_0 \boldsymbol{\eta}_0 \text{ with explicit form} \tag{43}$$

$$\begin{pmatrix} \boldsymbol{W}_{1,1} \\ \boldsymbol{W}_{1,2} \\ Y_{2,1} \\ Y_{2,2} \\ Y_{3,1} \\ Y_{3,2} \\ Y_{4,1} \\ \boldsymbol{W}_{2,3} \\ \boldsymbol{W}_{3,3} \\ \boldsymbol{W}_{4,2} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \boldsymbol{I} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \boldsymbol{I} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{I} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{I} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_{3,2} & \beta_{3,2} & 0 & 0 & 0 & 0 & \boldsymbol{G}_3 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & \boldsymbol{G}_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{I} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{B}_{3,2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \boldsymbol{B}_{4,1} \end{pmatrix} \begin{pmatrix} \boldsymbol{W}_{1,1} \\ \boldsymbol{W}_{1,2} \\ Y_{2,1} \\ Y_{2,2} \\ Y_{3,1} \\ 0 \\ 0 \\ \boldsymbol{W}_{2,3} \\ \boldsymbol{W}_{3,2} \\ \boldsymbol{W}_{4,1} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \boldsymbol{\Gamma}_{3,2} & 0 \\ 0 & \boldsymbol{\Gamma}_{4,1} \end{pmatrix} \begin{pmatrix} \boldsymbol{V}_{3,2} \\ \boldsymbol{V}_{4,1} \end{pmatrix}.$$

Now we consider the variance-covariance matrix of $\boldsymbol{X}_{-1}$, $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$. From (38), (42) and (43) we have

$$\boldsymbol{\eta}_{-2} := \begin{pmatrix} \boldsymbol{V}_{1,1} \\ \boldsymbol{V}_{2,1} \end{pmatrix}, \ \boldsymbol{\eta}_{-1} := \begin{pmatrix} \boldsymbol{V}_{2,2} \\ \boldsymbol{V}_{3,1} \end{pmatrix}, \ \boldsymbol{\eta}_0 := \begin{pmatrix} \boldsymbol{V}_{3,2} \\ \boldsymbol{V}_{4,1} \end{pmatrix}.$$

Since $\boldsymbol{V}_{j,k} \perp W_{j,1}, \ldots, W_{j,k}$ for $j = 1, \ldots, 4$ and since the transition noise on different scales is mutually independent it follows that $\boldsymbol{\eta}_k \perp \boldsymbol{X}_k$, $k = -2, -1, 0$, and $\boldsymbol{\eta}_{-2}$, $\boldsymbol{\eta}_{-1}$, $\boldsymbol{\eta}_0$ are mutually independent. Applying (38), (42) and (43) we therefore get

$$\begin{aligned} \boldsymbol{\Sigma}_{-1} &:= var(\boldsymbol{X}_{-1}) = \bar{\boldsymbol{A}}_{-2}\bar{\boldsymbol{\Sigma}}_{-2}\bar{\boldsymbol{A}}'_{-2} + \bar{\boldsymbol{\Gamma}}_{-2}var(\boldsymbol{\eta}_{-2})\bar{\boldsymbol{\Gamma}}'_{-2}, \\ \boldsymbol{\Sigma}_0 &:= var(\boldsymbol{X}_0) = \bar{\boldsymbol{A}}_{-1}\bar{\boldsymbol{\Sigma}}_{-1}\bar{\boldsymbol{A}}'_{-1} + \bar{\boldsymbol{\Gamma}}_{-1}var(\boldsymbol{\eta}_{-1})\bar{\boldsymbol{\Gamma}}'_{-1}, \\ \boldsymbol{\Sigma}_1 &:= var(\boldsymbol{X}_1) = \bar{\boldsymbol{A}}_0\bar{\boldsymbol{\Sigma}}_0\bar{\boldsymbol{A}}'_0 + \bar{\boldsymbol{\Gamma}}_0 var(\boldsymbol{\eta}_0)\bar{\boldsymbol{\Gamma}}'_0, \end{aligned} \tag{44}$$

where

$$var(\boldsymbol{\eta}_{-2}) = \begin{pmatrix} \boldsymbol{\Sigma}_{V1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{V2} \end{pmatrix}, \ var(\boldsymbol{\eta}_{-1}) = \begin{pmatrix} \boldsymbol{\Sigma}_{V2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{V3} \end{pmatrix}, \ var(\boldsymbol{\eta}_0) = \begin{pmatrix} \boldsymbol{\Sigma}_{V3} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{V4} \end{pmatrix}.$$

In the remainder of this section we consider at first the structure of the further state vectors for the linear Gaussian state space representation. In Subsection A.3 we construct system matrices $\bar{\boldsymbol{A}}_n$, $\bar{\boldsymbol{\Gamma}}_n$ and system noise $\boldsymbol{\eta}_n$, $n = 1, \ldots, N_4 - 1$, so that we can apply transition equations (7), (8) and (12) to the stochastic model over a given two-parent scale-to-scale Graph $\mathcal{G}^S$. In Subsection A.4 we show that the independence assumptions given in (25) for state space models are also fulfilled for the derived state space representation of Model (7) - (12) over a given two-parent scale-to-scale Graph $\mathcal{G}^S$.

## A.2 Structure of State Vectors of the Linear Gaussian State Space Representation

For further construction of the state space representation of Model (7) - (12) we proceed scale by scale, starting with $j = 4$. At first we require

$$Y_{4,k} \in \boldsymbol{X}_k, \;\; k = 1, \ldots, N_4. \tag{45}$$

Therefore (45) determines the number of state vectors to be equal to $N_4$. The state vectors $\boldsymbol{X}_2, \ldots, \boldsymbol{X}_{N_4}$ will have a similar structure as $\boldsymbol{X}_1$ defined in (32). For a precise definition we define general relationships between the set $\{1, \ldots, N_4\}$ and set $\{1, \ldots, N_j\}$, $j = 1, \ldots, 4$, as follows: For each $n \in \{1, \ldots, N_4\}$ there is a number

$$\begin{aligned}
&k_1^W(n) \in \{1, \ldots, N_1 - 1\} \text{ such that } \boldsymbol{W}_{1,k_j^W(n)}, \boldsymbol{W}_{1,k_j^W(n)+1} \in \boldsymbol{X}_n, \\
&k_j^Y(n) \in \{1, \ldots, N_j - 1\} \text{ such that } Y_{j,k_j^Y(n)}, Y_{j,k_j^Y(n)+1} \in \boldsymbol{X}_n, \;\; j = 2, 3, \\
&k_4^Y(n) \in \{1, \ldots, N_4\} \text{ such that } Y_{4,k_4^Y(n)} \in \boldsymbol{X}_n, \\
&k_j^W(n) \in \{1, \ldots, N_j\} \text{ such that } \boldsymbol{W}_{j,k_j^W(n)} \in \boldsymbol{X}_n, \;\; j = 2, 3, 4.
\end{aligned} \tag{46}$$

Using (45) it follows in particular, that $k_4^Y(n) = n$, $n = 1, \ldots, N_4$. For $n = 1, \ldots, N_4$ the exact values for $k_j^Y(n)$, $j = 1, 2, 3$ and $k_j^W(n)$, $j = 1, \ldots, 4$ will be given later. Note that for a given $j$, $k_j^Y(n)$ and $k_j^W(n)$ depend on $n$. For given $n \in \{1, \ldots, N_4\}$ we define the nth state vector $\boldsymbol{X}_n$ by

$$\boldsymbol{X}_n := \left( \boldsymbol{W}_{1,k_1^W(n)}, \boldsymbol{W}_{1,k_1^W(n)+1}, Y_{2,k_2^Y(n)}, Y_{2,k_2^Y(n)+1}, \ldots, Y_{4,k_4^Y(n)}, \boldsymbol{W}_{2,k_2^W(n)}, \ldots, \boldsymbol{W}_{4,k_4^W(n)} \right)'. \tag{47}$$

Note that all $\boldsymbol{X}_n$ have the same dimensionality. For example, we have for $n = 1$

$$k_1^W(1) = k_2^W(1) = k_3^Y(1) = k_4^Y(1) = 1, \;\; k_2^W(1) = k_3^W(1) = 3, \;\; k_4^W(1) = 2.$$

Conversely, the inverse relationships between set $\{1, \ldots, N_j\}$, $j = 1, \ldots, 4$, and set $\{1, \ldots, N_4\}$ are defined for $k = 1, \ldots, N_j$ by the index sets

$$\begin{aligned}
n_j^Y(k) &:= \left\{ n \in \{1, \ldots, N_4\} : Y_{j,k} \in \boldsymbol{X}_n \right\}, \\
n_j^W(k) &:= \left\{ n \in \{1, \ldots, N_4\} : \boldsymbol{W}_{j,k} \in \boldsymbol{X}_n \right\}.
\end{aligned} \tag{48}$$

Since the state vector $\boldsymbol{X}_n$ is defined by sub-vectors, we need to be able to identify individual components. For a matrix $\boldsymbol{M}$ with $m$ rows and an index set $S \subset \{1, \ldots, m\}$ let $\boldsymbol{M}[S]$ denote the rows of $\boldsymbol{M}$ with indices $i \in S$. We define index sets $i_1^Y(j)$, $i_2^Y(j)$ and $i^W(j)$ of row indices of

sub-vectors of $\boldsymbol{X}_n$, $n = 1, \ldots, N_4$, by

$$
\begin{aligned}
\boldsymbol{X}_n[i_1^Y(j)] &:= Y_{j,k_j^Y(n)},\ j = 2,3,4, \\
\boldsymbol{X}_n[i_2^Y(j)] &:= Y_{j,k_j^Y(n)+1},\ j = 2,3, \\
\boldsymbol{X}_n[i^W(j)] &:= \boldsymbol{W}_{j,k_j^W(n)},\ j = 1,\ldots,4.
\end{aligned} \tag{49}
$$

For example, $i_1^Y(2)$ is the row index of $Y_{2,k_2^Y(n)}$ in state vector $\boldsymbol{X}_n$, $n = 1, \ldots, N_4$. The index sets in (49) can be computed recursively. Recall that $m_j^W$ denotes the dimension of random vector $\boldsymbol{W}_{j,k}$ in (7), (8) for $j = 1, \ldots, 4$ and $k = 1, \ldots, N_j$. Therefore $i_1^Y(j)$, $i_2^Y(j)$ and $i^W(j)$ do not depend on $n$. Let $\max[i_1^Y(j)]$ denote the maximum of index set $i_1^Y(j)$, $\max[i_2^Y(j)]$ denote the maximum of $i_2^Y(j)$ and $\max[i^W(j)]$ denote the maximum of $i^W(j)$, respectively. Using (47) we get

$$
\begin{aligned}
& i_1^W(1) = \{1, \ldots, m_1^W\},\ i_2^W(1) := \{m_1^W + 1, \ldots, 2m_1^W\}, \\
& i_1^Y(2) = 2m_1^W + 1,\ i_2^Y(2) = i_1^Y(2) + 1, \\
& i_1^Y(3) = i_2^Y(2) + 1,\ i_2^Y(3) = i_1^Y(3) + 1, \\
& i_1^Y(4) = i_2^Y(3) + 1, \\
& i^W(2) = \{i_1^Y(4) + 1, \ldots, i_1^Y(4) + m_2^W\}, \\
& i^W(j) = \{\max[i^W(j-1)] + 1, \ldots, \max[i^W(j-1)] + m_j^W\},\ j = 3,4.
\end{aligned} \tag{50}
$$

We want to incorporate (7), (8) into the state space representation. Therefore, if $Y_{j,k} \in \boldsymbol{X}_n$ and $Y_{j,k} \notin \boldsymbol{X}_{n-1}$ then we require for $j \in \{2, \ldots, 4\}$, $k \in \{1, \ldots, N_j\}$ and $n \in \{2, \ldots, N_4\}$ that

$$
\boldsymbol{W}_{j,k} \in \boldsymbol{X}_{n-1}, \tag{51}
$$

$$
Y_{lpa(j-1,j,k)}, Y_{rpa(j-1,j,k)} \in \boldsymbol{X}_{n-1}. \tag{52}
$$

In the remainder of this section we consider how to specify sets $n_j^Y(k)$ and $n_j^W(k)$, $j = 1, \ldots, 4$, $k = 1, \ldots, N_j$. This is equivalent to the specification of $k_j^Y(n)$ and $k_j^W(n)$ for $j = 1, \ldots, 4$, $n = 1, \ldots, N_4$. We start with $j = 4$ and proceed scale by scale for $j = 3, 2, 1$.

**Scale 4:**

The index set $n_4^Y(k)$, $k = 1, \ldots, N_4$ is specified through (45), i.e. $n_4^Y(k) = \{k\}$. Note that $\boldsymbol{W}_{4,2} \in \boldsymbol{X}_1$. Using (45) together with (51) then gives

$$
\boldsymbol{X}_{n-1}[i^W(4)] := \boldsymbol{W}_{4,n},\ \ n = 2, \ldots, N_4. \tag{53}
$$

Let $\boldsymbol{W}_{4,N_4} \in \boldsymbol{X}_{N_4}$ then we have $n_4^W(k) = k - 1,\ k = 1, \ldots, N_4 - 1,\ n_4^W(N_4) = N_4$. **Scales 3,2,1:**

For the scales $j = 3, 2, 1$ we proceed in a recursive manner. We need some additional notation and definitions. For any set $\mathcal{M}$ let card $\mathcal{M}$ denote the cardinal number of $\mathcal{M}$. We define for $i = 2, 3, 4$ and $j = 1, \ldots, i-1$

$$c(i,j,0) := 0, \;\; c(i,j,k) := \text{card}\left\{\bigcup_{l=1}^{k} de(i,j,l)\right\}, \; k = 1, \ldots, N_j. \tag{54}$$

Recall that $de(i,j,l)$ denotes the set of descendants on scale $i$ of node $(j,l)$ in Graph $\mathcal{G}^S$. Therefore $c(i,j,k)$ is the cumulative number of descendants on scale $i$ of nodes $(j,1), \ldots, (j,k)$. Definition (54) implies

$$c(i,j,k) = max\,\{l : (i,l) \in de(i,j,k)\}, \; j = 1, \ldots, 3, \; i = j+1, \ldots, 4, \; k = 1, \ldots, N_j.$$

Moreover,

$$m[j,k] := \min\left\{n \in \{1, \ldots, N_4\} : n \in n_{j+1}^{Y}(k)\right\}, \; k = 1, \ldots, N_j. \tag{55}$$

This means, $m[j,k]$ gives index $n$ of the first state vector $\boldsymbol{X}_n$ with $Y_{j,k} \in \boldsymbol{X}_n$, i.e.

$$Y_{j,k} \in \boldsymbol{X}_n \Rightarrow n \geq m[j,k]. \tag{56}$$

Note, that it follows from (45) that $m[4,k] = k$, $k = 1, \ldots, N_4$. Now we consider the specification of $n_j^Y(k)$, $k = 1, \ldots, N_j$. Condition (52) requires in particular that if $Y_{j+1,k} \in \boldsymbol{X}_n$ and $Y_{j+1,k} \notin \boldsymbol{X}_{n-1}$ then

$$Y_{lpa(j,j+1,k)}, \; Y_{rpa(j,j+1,k)} \in \boldsymbol{X}_{n-1}, \; k = 1, \ldots, N_{j+1}, \; n = 2, \ldots, N_4. \tag{57}$$

We consider a sufficient condition for (57). Note that (54) implies

$$(Y_{j,1} \cup Y_{j,2}) \supseteq (Y_{lpa(j,j+1,k)} \cup Y_{rpa(j,j+1,k)}), \; k = 1, \ldots, c(j+1,j,1).$$

Further $lpa(j,j+1,k) = rpa(j,j+1,k)$ for $k = 1$ and may hold for some $k \in \{2, \ldots, c(j+1,j,1)\}$. Using (54) we have $(j+1,k) \notin de(j+1,j,1), \; k = c(j+1,j,1)+1, \ldots, N_{j+1}$. Note that $(j,2) = lpa(j,j+1,k)$ for $(j+1,k) \in de(j+1,j,2) \setminus de(j+1,j,1)$. Since a node can have up to two parents we get that

$$(Y_{j,2} \cup Y_{j,3}) \supseteq (Y_{lpa(j,j+1,k)} \cup Y_{rpa(j,j+1,k)}), \; k = c(j+1,j,1)+1, \ldots, c(j+1,j,2).$$

More generally, we write for $k_j = 2, \ldots, N_j - 1$

$$\begin{aligned}
&(j+1,k_{j+1}) \notin de(j+1,j,k_j-1), \; k_{j+1} = c(j+1,j,k_j-1)+1, \ldots, N_{j+1}, \\
&\quad (j,k_j) = lpa(j,j+1,k_{j+1}) \text{ for } (j+1,k_{j+1}) \in de(j+1,j,k_j) \setminus de(j+1,j,k_j-1), \\
&(Y_{j,k_j} \cup Y_{j,k_j+1}) \supseteq (Y_{lpa(j,j+1,k_{j+1})} \cup Y_{rpa(j,j+1,k_{j+1})}) \text{ for} \\
&\quad k_{j+1} = c(j+1,j,k_j-1)+1, \ldots, c(j+1,j,k_j).
\end{aligned} \tag{58}$$

Recall that node $(j+1, k_{j+1})$ with $k_{j+1} = c(j+1, j, k_j)$ is the child of node $(j, k_j)$, $k_j = 1, \ldots, N_j$, such that

$$(j+1, k) \in de(j+1, j, k_j) \Rightarrow k \leq c(j+1, j, k_j)$$

holds. Therefore, we may say that node $(j+1, k_{j+1})$ with $k_{j+1} = c(j+1, j, k_j)$ is the last child of node $(j, k_j)$. Note that $m[j+1, c(j+1, j, k_j)]$ denotes the lowest index $n$ of state vector $\boldsymbol{X}_n$ with $Y_{j+1,c(j+1,j,k_j)} \in \boldsymbol{X}_n$. More formally, $Y_{j+1,c(j+1,j,k_j)} \in \boldsymbol{X}_n \Rightarrow n \geq m[j+1, c(j+1, j, k_j)]$. To meet Condition (57) we set therefore

$$\begin{aligned}
&\boldsymbol{X}_{n-1}[i_1^Y(j)] := Y_{j,1},\ \boldsymbol{X}_{n-1}[i_2^Y(j)] := Y_{j,2},\ , n = 2, \ldots, m[j+1, c(j+1, j, 1)], \\
&\boldsymbol{X}_{n-1}[i_1^Y(j)] := Y_{j,k_j},\ \boldsymbol{X}_{n-1}[i_2^Y(j)] := Y_{j,k_j+1}, \\
&\qquad n = m[j+1, c(j+1, j, k_j - 1)] + 1, \ldots, m[j+1, c(j+1, j, k_j)], \\
&\qquad k_j = 2, \ldots, N_2 - 2, \\
&\boldsymbol{X}_{n-1}[i_1^Y(j)] := Y_{j,N_j-1},\ \boldsymbol{X}_{n-1}[i_2^Y(j)] := Y_{j,N_j}, \\
&\qquad n = m[j+1, c(j+1, j, N_j - 2)] + 1, \ldots, N_4 + 1.
\end{aligned} \tag{59}$$

Note that (59) specifies $n_j^Y(k)$, $k = 1, \ldots, N_j$ and therefore $m[j, k]$. $m[j, k]$ then gives index $n$ of the first state vector $\boldsymbol{X}_n$ with $Y_{j,k} \in \boldsymbol{X}_n$, i.e.

$$Y_{j,k} \in \boldsymbol{X}_n \Rightarrow n \geq m[j, k]. \tag{60}$$

In particular, we have for $n = m[j, k]$, $k = 3, \ldots, N_j$,

$$\begin{aligned}
\boldsymbol{X}_{n-1}[i_1^Y(j)] = Y_{j,k-j},\ \boldsymbol{X}_{n-1}[i_2^Y(j)] = Y_{j,k-1}, \\
\boldsymbol{X}_{n}[i_1^Y(j)] = Y_{j,k-1},\ \boldsymbol{X}_{n}[i_2^Y(j)] = Y_{j,k}.
\end{aligned} \tag{61}$$

Using (58) and (59) it follows that

$$\left\{\boldsymbol{X}_{n-1}[i_1^Y(j)] \cup \boldsymbol{X}_{n-1}[i_2^Y(j)]\right\} \supset \left(Y_{lpa(j,j+1,k_{j+1})}, Y_{rpa(j,j+1,k_{j+1})}\right),\ n = 2, \ldots, N_4. \tag{62}$$

By (51) and (55) $\boldsymbol{W}_{j,k} \in \boldsymbol{X}_{n-1}$ has to hold for $n = m[j, k]$, $k = 3, \ldots, N_j$. This is ensured by setting

$$\begin{aligned}
&\boldsymbol{X}_{n-1}[i^W(j)] := \boldsymbol{W}_{j,3},\ \ n = 2, \ldots, m[j, 4] - 1, \\
&\boldsymbol{X}_{n-1}[i^W(j)] := \boldsymbol{W}_{j,k},\ \ n = m[j, k], \ldots, m[j, k+1] - 1,\ \ k = 4, \ldots, N_j - 1, \\
&\boldsymbol{X}_{n-1}[i^W(j)] := \boldsymbol{W}_{j,N_j},\ \ n = m[j, N_j], \ldots, N_4.
\end{aligned} \tag{63}$$

Note that (63) specifies $n_j^W(k)$, $k = 3, \ldots, N_j$.

## A.3 Construction of System Matrices and System Noise for the Linear Gaussian State Space Representation

In this subsection we consider the construction of the system matrices $\bar{\boldsymbol{A}}_n$, $n = 1, \ldots, N_4$. With $\bar{\boldsymbol{A}}_n$ given, system noise $\boldsymbol{\eta}_n$ and matrix $\bar{\boldsymbol{\Gamma}}_n$ are also specified. Similar to the previous subsection we proceed scale by scale, starting with scale $j = 4$.

**Scale 4:**

Using (45) and (57) we have $Y_{4,n} \in \boldsymbol{X}_n$ and $Y_{lpa(3,4,n)}, Y_{rpa(3,4,n)} \in \boldsymbol{X}_{n-1},\ n = 2, \ldots, N_4$. If $lpa(3,4,n) \neq rpa(3,4,n)$ then $\alpha_{4,n}, \beta_{4,n}$ are given by (11), $n = 1, \ldots, N_4$.

If $lpa(3,4,n) = rpa(3,4,n)$ we set

$$\begin{aligned} &\beta_{4,n} := 0 \text{ if } lpa(3,4,n) = rpa(3,4,n) \text{ and } n \leq c(4,3,N_3-1), \\ &\beta_{4,n} := 1 \text{ if } lpa(3,4,n) = rpa(3,4,n) \text{ and } n > c(4,3,N_3-1). \end{aligned} \tag{64}$$

Note that for $n > c(4,3,N_3-1)$ it follows using (8) that node $(3,N_3)$ is the only parent on scale $j = 3$ of node $(4,n)$. From (53) it follows that $\boldsymbol{W}_{4,n} \in \boldsymbol{X}_{n-1}$, $n = 2, \ldots, N_4$. Using (62), (64) and defining for $n = 2, \ldots, N_4$

$$\bar{\boldsymbol{A}}_{n-1}[i_1^Y(4)] := \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \alpha_{4,n} & \beta_{4,n} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{G}_4 \end{pmatrix} \tag{65}$$

then gives

$$\bar{\boldsymbol{A}}_{n-1}[i_1^Y(4)] \cdot \boldsymbol{X}_{n-1} = \alpha_{4,n} Y_{3,k_3^Y(n-1)} + \beta_{4,n} Y_{3,k_3^Y(n-1)+1} + \boldsymbol{G}_4 \boldsymbol{W}_{4,n} = Y_{4,n}. \tag{66}$$

Using (12) with $j = 4$ we have $\boldsymbol{W}_{4,n} = \boldsymbol{B}_{4,n-1}\boldsymbol{W}_{4,n-1} + \boldsymbol{\Gamma}_{4,n-1}\boldsymbol{V}_{4,n-1}$. Therefore (53) implies for $n = 1, \ldots, N_4 - 1$

$$\bar{\boldsymbol{A}}_n[i^W(4)] = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{B}_{4,n} \end{pmatrix}, \tag{67}$$

$$\boldsymbol{V}_{4,n} \in \boldsymbol{\eta}_n, \tag{68}$$

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{4,n} \end{pmatrix}' \in \bar{\boldsymbol{\Gamma}}_n. \tag{69}$$

**Scale 3:**

We consider now the construction of the sub-matrices $\bar{\boldsymbol{A}}_{n-1}[i_1^Y(3)]$ and $\bar{\boldsymbol{A}}_{n-1}[i_2^Y(3)]$ of the system matrices $\bar{\boldsymbol{A}}_{n-1}$. Note that (32) together with (55) imply $m[3,1] = m[3,2] = 1$. Using (59) we get for $n = 2, \ldots, m[3,k] - 1$ and $k = 3, \ldots, N_3$ that $Y_{3,k} \notin \boldsymbol{X}_n$ and in particular $\boldsymbol{X}_{n-1}[i_1^Y(3)] = \boldsymbol{X}_n[i_1^Y(3)]$, $\boldsymbol{X}_{n-1}[i_2^Y(3)] = \boldsymbol{X}_n[i_2^Y(3)]$. Therefore we set for $n =$

$2, \ldots, m[3,k]-1$ and $k = 3, \ldots, N_3$

$$\begin{aligned} \bar{\boldsymbol{A}}_{n-1}[i_1^Y(3)] &:= \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \\ \bar{\boldsymbol{A}}_{n-1}[i_2^Y(3)] &:= \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}. \end{aligned} \tag{70}$$

Since $n = m[3,k]$ is defined as the first index of state vector $\boldsymbol{X}_n$ with $Y_{3,k} \in \boldsymbol{X}_n$, (59) and (55) imply for $n = m[3,k]$

$$\begin{aligned} \boldsymbol{X}_{n-1}[i_1^Y(3)] &= Y_{3,k-2}, \qquad & \boldsymbol{X}_{n-1}[i_2^Y(3)] &= Y_{3,k-1}, \\ \boldsymbol{X}_n[i_1^Y(3)] &= Y_{3,k-1}, \qquad & \boldsymbol{X}_n[i_2^Y(3)] &= Y_{3,k} \end{aligned}$$

and we set for $n = m[3,k]$ and $k = 3, \ldots, N_3$

$$\begin{aligned} \bar{\boldsymbol{A}}_{n-1}[i_1^Y(3)] &:= \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \\ \bar{\boldsymbol{A}}_{n-1}[i_2^Y(3)] &:= \begin{pmatrix} \mathbf{0} & \mathbf{0} & \alpha_{3,k} & \beta_{3,k} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{G}_3 & \mathbf{0} \end{pmatrix}. \end{aligned} \tag{71}$$

If $lpa(2,3,k) \neq rpa(2,3,k)$ then $\alpha_{3,k}$ and $\beta_{3,k}$ are given as in (11). If $lpa(2,3,k) = rpa(2,3,k)$ in $\mathcal{G}^S$ then we set

$$\begin{aligned} &\beta_{3,k} := 0 \text{ if } lpa(2,3,k) = rpa(2,3,k) \text{ and } k \leq c(3,2,N_2-1), \\ &\beta_{3,k} := 1 \text{ if } k > c(3,2,N_2-1). \end{aligned} \tag{72}$$

Note that for $k > c(3,2,N_2-1)$ it follows using (8) that node $(2,N_2)$ is the only parent on scale $j=2$ of node $(3,k)$. For $n = m[3,N_3], \ldots, N_4$ we have using (59) $Y_{3,N_3-1}, Y_{3,N_3} \in \boldsymbol{X}_n$. Therefore $\bar{\boldsymbol{A}}_{n-1}[i_1^Y(3)]$ and $\bar{\boldsymbol{A}}_{n-1}[i_2^Y(3)]$ for $n = m[3,N_3]+1, \ldots, N_4$ are given by (70).

Now we consider the construction of sub-matrix $\bar{\boldsymbol{A}}_{n-1}[i^W(3)]$. Equation (63) requires for $n = m[3,k]-1$, $k = 4, \ldots, N_3$,

$$\bar{\boldsymbol{A}}_{n-1}[i^W(3)] := \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{B}_{3,k-1} & \mathbf{0} \end{pmatrix}. \tag{73}$$

Equation (73) also requires for $n = m[3,k]-1$, $k = 4, \ldots, N_3$,

$$\boldsymbol{V}_{3,k-1} \in \boldsymbol{\eta}_{n-1}, \tag{74}$$

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Gamma}_{3,k-1} & \mathbf{0} \end{pmatrix}' \in \bar{\boldsymbol{\Gamma}}_{n-1}. \tag{75}$$

This gives for $n = m[3,k]-1$, $k = 4, \ldots, N_3$,

$$\bar{\boldsymbol{A}}_{n-1}[i_3^W]\boldsymbol{X}_{n-1} + \bar{\boldsymbol{\Gamma}}_{n-1}\boldsymbol{\eta}_{n-1} = \boldsymbol{B}_{3,k-1}\boldsymbol{W}_{3,k-1} + \boldsymbol{\Gamma}_{3,k-1}\boldsymbol{V}_{3,k-1} = \boldsymbol{W}_{3,k}$$

and therefore $\boldsymbol{X}_n[i^W(3)] = \boldsymbol{W}_{3,k},\ n = m[3,k] - 1,\ k = 4,\ldots,N_3$, as required by (63). For $n \neq m[3,k] - 1$, $k = 4,\ldots,N_3$, we then set

$$\bar{\boldsymbol{A}}_{n-1}[i^W(3)] := \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I} & \mathbf{0} \end{pmatrix}. \tag{76}$$

Modifications for the scales $j = 2, 1$ are obvious and are not given explicitly.

## A.4 Independence Assumptions of the Linear Gaussian State Space Representation

So far we have shown that the transition from $\boldsymbol{X}_n$ to $\boldsymbol{X}_{n+1}$ has the form of (23). Since Model (7)-(12) over a two-parent scale-to-scale Graph $\mathcal{G}^S$ is Gaussian, the state space representation derived in this section is also Gaussian. Recalling that $Y_{4,k} = Z_k$ for $k = 1,\ldots,4$ we get

$$Z_k = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \boldsymbol{X}_k,$$

which has the form of (24). We discuss now that Assumptions (25) for the linear Gaussian state space model are also satisfied. Since system noise vector $\boldsymbol{\eta}_n$, $n = 1,\ldots,N_4 - 1$, in the state space representation contains only $\boldsymbol{V}_{j,k}$ of Model (7) - (12) over a two-parent scale-to-scale Graph $\mathcal{G}^S$ we get using (14), that $\{\boldsymbol{\eta}_n,\ n = 1,\ldots,N_4 - 1\}$ is serially independent. Furthermore, since $\{\boldsymbol{V}_{j,k}, j = 1,\ldots,4, k = 1,\ldots,N_j\}$ is jointly normally distributed the family $\{\boldsymbol{\eta}_n,\ n = 1,\ldots,N_4 - 1\}$ is also jointly normally distributed. From (13), (15), (16) and (17) it follows that $\boldsymbol{\eta}_n \perp \boldsymbol{X}_1,\ n = 1,\ldots,N_4 - 1$. Therefore, Assumptions (25) are satisfied.