MSc Economics

Parameter estimation of a mixed frequency vector autoregressive model of order 1

A Master's Thesis submitted for the degree of
"Master of Science"

supervised by

Leopold Sögner

Peter Horvath

1325889

Vienna, 08.06.2015

INSTITUT FÜR HÖHERE STUDIEN
INSTITUTE FOR ADVANCED STUDIES
Vienna

TU WIEN CONTINUING EDUCATION CENTER

MSc Economics

Affidavit

I, Peter Horvath

hereby declare

that I am the sole author of the present Master's Thesis,

Parameter estimation of a mixed frequency vector autoregressive model of order 1

33 pages, bound, and that I have not used any source or tool other than those referenced or any other illicit aid or tool, and that I have not prior to this date submitted this Master's Thesis as an examination paper in any form in Austria or abroad.

Vienna, 08.06.2015

Signature

Contents

List of Tables

Abstract

For multivariate time series, different time series might be sampled at different frequencies. A mixed frequency vector autoregressive model (MF–VAR) can manage the problems arising from the mixed sampling of variables. In my thesis, I implement an estimation method for the parameters of an MF–VAR(1) model, described in Anderson et al. (2012), and investigate how the accuracy of the parameter estimation changes if the slow frequency variable is sampled less and less often compared to the fast frequency variable and if the innovation variance matrix increases. I find that the larger the distance between the observations of the slow process, the worse the estimation is.

I would like to express my thanks to my advisor, Leopold Soegner for his countless helpful comments and guidance during the writing of this thesis.

1 Introduction

In several economic applications, the variables, which are desired to be used, are not available at the same frequency. One way to cope with this issue is to use mixed frequency data and mixed frequency methods. For example, one of the most important economic indicators, gross domestic product (GDP) is published only quarterly with a 3–month lag, while inflation or the unemployment rate are available for every month. In finance, when one wants to forecast the daily prices of commodities, e.g. price of crude oil, the inventory would be a proper explanatory factor, however, it is only available at monthly frequency with a 6–month lag (see (Ye et al., 2006)).

To cope with the problems emerging from the mixed frequency data sampling, several methods have been developed and being used. To nowcast (due to the delay of publishing) and forecast GDP, Baffigi et al. (2004) use Bridge models which link the fast frequency variables to the slow frequency variables.

Ghysels et al. (2007) introduce the Mixed Data Sampling method which uses the fast frequency variable to explain the slow frequency one. Anderson et al. (2015) provide identification of the parameters in mixed frequency vector autoregressive model (MF–VAR) which includes both the slow and fast frequency data as explanatory variables.

In this study I apply the MF–VAR(1) model, described in Anderson et al. (2012) and Anderson et al. (2015). I estimate the system and noise parameters to investigate how the distance between two observations of the slow frequency process affect the accuracy of the parameter estimation. The parameters are estimated by ordinary least square (OLS) and general method of moments (GMM). To test the finite sample performance of the estimation methods, a Monte Carlo study with different set–ups is performed.

The further content of the thesis is the followings: section 2 provides a short literature review of methods which are used to estimate and forecast including mixed frequency data. Section 3 describes the mixed frequency vector autoregressive model, which is in the center of my thesis and the methods applied to estimate the model parameters. Section 4 describes the simulation design and the results of the parameter estimation on a particular example. Section 5 concludes.

2 Literature review

This section provides a summary on mixed–frequency data forecasting and/or the estimation of parameters from mixed frequency data. It also covers empirical examples from the literature.

Foroni and Marcellino (2013) provide a comprehensive description of the existing mixed frequency methodologies. They thoroughly describe the existing models concerning estimation and forecasting with mixed frequency data.

According to them, one of the simplest methods to forecast with mixed frequency data without aggregation of the high frequency variable is the bridge models (BM). The main point of BM is that first a forecast of the high frequency variable is performed over a period when low frequency variable can be observed, and then the aggregated forecast of the first step is used to forecast the low frequency variable. Baffigi et al. (2004) test whether BM can give better nowcasts and forecasts of the 1– and 2–quarter ahead the French, German and Italian GDP, than benchmark models (ARIMA, VAR). According to their results, BM performs particularly good in nowcasting the GDP, when the new GDP data are not published but indicators are already available and in forecasting the GDP, it also gives good results if at least some indicators are already known in the forecasting period comparing to the benchmark models.

Estimating parameters from mixed frequency data demands for different methods. One of the most famous methods which is used in the literature is the Mixed Data Sampling (MIDAS).

Ghysels et al. (2007) fully describe the structure and the opportunities of MIDAS. With MIDAS, one can use a variable sampled at high(er) frequency, e.g. daily to predict/explain a variable which is sampled only at low(er) frequency, e.g. monthly. The advantage of this technique is that data does not have to be pre–filtered. That is, taking their example, the daily data does not have to be transformed into monthly so the possibility of information loss can be avoided. Ghysels et al. (2007) provide several extensions of the basic MIDAS such as: explanatory variables can be sampled at different frequencies, multivariate MIDAS, tick–by–tick data sampled at unequally spaced intervals. To use high frequency data as a predictor of lower frequency data, one can use different polynomial specifications in MIDAS which also control the number of lags used in a regressions.

To demonstrate its applicability in empirical data Ghysels et al. (2007) provide two examples. First, they take the ICAPM model proposed by Merton (1973) to predict the Dow Jones index returns on weekly (5 days), 2–weekly (10 days), 3–weekly (15 days) and monthly (22 days) horizons using five different model specifications. As explanatory variables they include daily squared returns, the daily absolute returns, the daily ranges, realized volatility (RV) and "realized power" to capture the volatility. According to their findings, different polynomial specifications might lead to various results, however, they claim that these differences are not so considerable. In their second empirical example, they forecast the daily volatility of Alcoa Inc. and Microsoft stocks using adjusted and unadjusted RV sampled at every 5 and every 30 minutes.

An other way to handle mixed frequency data is using mixed frequency vector autoregressive models (MF–VAR). A main difference between MF–VAR and MIDAS is that using MIDAS, one can only include the variable sampled at high frequency as explanatory variable and the low frequency variable must be the dependent variable. However, if someone wants to use e.g. inventory data (which is available only monthly) to predict daily prices, it is not possible with MIDAS but it becomes possible with MF–VAR.

Since the low frequency data is not as often observed as the high frequency, the low frequency data has to be estimated in the unobserved periods before it would be used to forecast the high frequency variable. To do so, the identifiability of the system and noise parameters of the MF–VAR has to be proven. Anderson et al. (2012) provides two ways which guarantee generic identifiability. First, they use extended Yule Walker equations which provides enough observable second moments to identify the parameters. Second, they show in a simple case, when there is only one high and only one low frequency variable in a MF–VAR(1) system and the slow frequency variable is sampled at every 2^{nd} period, that substitution method also works. That is, they write the slow and fast frequency variables as a function of known, observed variable, therefore they manage to avoid that the slow frequency process is not alway observable.

Anderson et al. (2015) also investigate the problem of generic identifiability of an MF–VAR system. In their work, they impose only three restrictions regarding the VAR system: order of the system, stability and rank of the innovation variance matrix. In addition to their previous paper, they also show generic identifiability in a case of a slow flow variable.

In empirical research, it is not obvious whether MIDAS or MF–VAR can give more accurate forecasts (or better parameter estimation). To give an answer to that question which method can provide more precise forecasts Kuzin et al. (2011) use AR–MIDAS (a MIDAS regression with lagged dependent term), MIDAS and MF–VAR to nowcast and forecast quarterly Euro GDP data using monthly available variables from 1992 to 2008. They apply the Bayesian information criteria to choose the proper lagged term. To compare the outcomes of the nowcasts and forecasts they use relative mean squared error (relative to an univariate AR model). On short horizon (3 months or less) the AR–MIDAS model has smaller MSE than benchmark AR, however, on longer horizon (8–9 months) MF–VAR performs better than the benchmark and the other two model specifications. That is, they show with an empirical dataset that none of the aforementioned methods is superior to another.

3 Model and Parameter Estimation

This section describes the model which is in the center of my thesis and the estimation method which is used to estimate the system parameters. The original MF–VAR model and the proofs of generic identifiability of the parameters can be found in Anderson et al. (2012) and Anderson et al. (2015).

3.1 Model

I will consider the case where both the high and low frequency variable are scalars. Following Anderson et al. (2012), I will use a multivariate autoregressive model of order 1:

$$\begin{pmatrix} x_t^f \\ x_t^s \end{pmatrix} = \begin{pmatrix} c^f \\ c^s \end{pmatrix} + \begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \end{pmatrix} \begin{pmatrix} x_{t-1}^f \\ x_{t-1}^s \end{pmatrix} + \begin{pmatrix} \nu_t^f \\ \nu_t^s \end{pmatrix} \qquad t \in \mathbb{Z} \tag{1}$$

$x_t^f \in \mathbb{R}$ denotes the variable observed at high/fast (e.g. daily) frequency, while $x_t^s \in \mathbb{R}$ is observed only at low/slow (e.g. weekly/monthly) frequency. That is, we can observe x_t^f at every $t \in \mathbb{Z}$, but we can observe x_t^s only at every $t \in N\mathbb{Z}$. In certain cases, e.g. monthly data are used to forecast daily data, then N is not a constant number but $N \in \{20, 21, 22, 23\}$ since the number of trading/business day in a month is changing.

The innovation term is normally distributed:

$$\boldsymbol{\nu} = \begin{pmatrix} \nu_t^f \\ \nu_t^s \end{pmatrix} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\nu}}),$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{\nu}} = \mathbb{E}\left(\begin{pmatrix} \nu_t^f \\ \nu_t^s \end{pmatrix} \begin{pmatrix} \nu_\tau^f & \nu_\tau^s \end{pmatrix} \right) = \begin{cases} \begin{pmatrix} \Sigma_{ff} & \Sigma_{fs} \\ \Sigma_{sf} & \Sigma_{ss} \end{pmatrix} & \text{if } t = \tau, \\ \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} & \text{if } t \neq \tau. \end{cases}$$

In order to get equation (1) in the MF–VAR system form as Anderson et al. (2012) and Anderson et al. (2015) define the MF–VAR system in their papers, we have to subtract the mean $\mathbb{E}(x_t^i) = \mu^i$, where $i \in \{f, s\}$. Then we can define

$y_t^i := x_t^i - \mu^i, \qquad i \in \{f, s\}$. So we can rewrite (1) in "demeaned" form (which is similar to (Anderson et al., 2012, p. 184(1)):

$$\begin{pmatrix} y_t^f \\ y_t^s \end{pmatrix} = \begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \end{pmatrix} \begin{pmatrix} y_{t-1}^f \\ y_{t-1}^s \end{pmatrix} + \begin{pmatrix} \nu_t^f \\ \nu_t^s \end{pmatrix} \tag{2}$$

which allows to obtain (see e.g. Hamilton, 1994, p. 258):

$$\begin{pmatrix} c^f \\ c^s \end{pmatrix} = \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \end{pmatrix} \right) \begin{pmatrix} \mu^f \\ \mu^s \end{pmatrix}.$$

3.2 Estimation of the Parameters

3.2.1 Estimation of System Parameters

Following Anderson et al. (2012) the matrix equation (2) can be rewritten in the following way:

$$y_t^f = \begin{pmatrix} a_{ff} & a_{fs} \end{pmatrix} \begin{pmatrix} y_{t-1}^f \\ y_{t-1}^s \end{pmatrix} + \nu_t^f, \tag{3}$$

$$y_t^s = \begin{pmatrix} a_{sf} & a_{ss} \end{pmatrix} \begin{pmatrix} y_{t-1}^f \\ y_{t-1}^s \end{pmatrix} + \nu_t^s. \tag{4}$$

Similar to Anderson et al. (2012), in equation (3), we can postmultiply both sides with $\begin{pmatrix} y_{t-1}^{f'} & y_{t-1}^{s'} \end{pmatrix}$, where $y_{t-1}^{i'}, \quad i \in \{f, s\}$ denote those periods when both variables can be observed and use ordinary least squares to estimate a_{ff} and a_{fs}.

Let $\mathbb{T}$ stand for the set of time indices where both processes can be observed. That is, $\{1, 1+N, 1+2N, \ldots, T'\}$, where T' is the last period when both processes can be observed. If N is 2, 5, 10 or 20 as it is in this study, then $T' = 501$. In addition, let $x_t := \begin{pmatrix} y_{t-1}^f & y_{t-1}^s \end{pmatrix}$. Then, the parameters a_{ff} and a_{fs} can be estimated by means of ordinary least squares (see e.g. Hansen, 2015, p. 88):

$$\begin{pmatrix} \hat{a}_{ff} \\ \hat{a}_{sf} \end{pmatrix} = \left(\sum_{t \in \mathbb{T}} x_t x_t^T \right)^{-1} \sum_{t \in \mathbb{T}} x_t y_{t+1}^f, \tag{5}$$

In equation (4), y_t^s and y_{t-1}^s cannot be simultaneously observed. (If $N \geq 2$, then for certain t none of them is observed.) Therefore, we cannot use OLS to estimate the remaining system parameters (a_{sf}, a_{ss}). However, we can use the general method of moments (GMM) to estimate them. First, we have to express

y_t^f and y_t^f as a function of only observable variables, y_k^s and $y_t^f, \ldots, y_{t-k}^f$ (for more details see Appendix A):

$$y_t^f = a_{ff} y_{t-1}^f + a_{fs} a_{sf} \sum_{i=1}^{k-1} a_{ss}^{i-1} y_{t-1-i}^f + a_{fs} a_{ss}^{k-1} y_{t-k}^s + a_{fs} \sum_{i=1}^{k-1} a_{ss}^{i-1} \nu_{t-i}^s + \nu_t^f, \quad (6)$$

$$y_t^s = a_{sf} \sum_{i=1}^{k} a_{ss}^{i-1} y_{t-i}^f + a_{ss}^k y_{t-k}^s + \sum_{i=1}^{k} a_{ss}^{i-1} \nu_{t+1-i}^s. \quad (7)$$

From the equations (6), (7) and the assumption that $\mathbb{E}(\boldsymbol{\nu}_t) = \mathbf{0}$ as well as $\mathbb{E}(\nu_t^i y_\tau^j) = 0$ if $t \neq \tau$ where $i, j \in \{f, s\}$, we can calculate the following second moments which are also the moment conditions for the GMM estimator:

$$\mathbb{E}(y_t^s y_{t-k}^f) = \mathbb{E}(a_{sf} y_{t-k}^f \sum_{i=1}^{k} a_{ss}^{i-1} y_{t-i}^f + a_{ss}^k y_{t-k}^f y_{t-k}^s), \quad (8)$$

$$\mathbb{E}(y_t^s y_{t-k}^s) = \mathbb{E}(a_{sf} y_{t-k}^s \sum_{i=1}^{k} a_{ss}^{i-1} y_{t-i}^f + a_{ss}^k y_{t-k}^s y_{t-k}^s). \quad (9)$$

To estimate the remaining unknown system parameters $\boldsymbol{\theta} = \begin{pmatrix} a_{sf}, a_{ss} \end{pmatrix}^T$, the GMM estimator (Hansen, 2015, pp. 278–279) minimizes the following quadratic form on the set of stable system matrices (Lütkepohl, 2005, p. 16):

$$\begin{aligned} \hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \ & \boldsymbol{g_T(\theta)}^T \ I_2 \ \boldsymbol{g_T(\theta)} \\ \text{subject to } & \det(I_2 - \hat{\mathbf{A}} z) \neq 0 \quad \text{if } |z| \leq 1 \end{aligned} \quad (10)$$

where $\boldsymbol{g_T(\theta)}$ is the finite sample analog of

$$\boldsymbol{g(\theta)} := \begin{pmatrix} \mathbb{E}(y_t^s y_{t-k}^f - a_{sf} y_{t-k}^f \sum_{i=1}^{k} a_{ss}^{i-1} y_{t-i}^f - a_{ss}^k y_{t-k}^f y_{t-k}^s) \\ \mathbb{E}(y_t^s y_{t-k}^s - a_{sf} y_{t-k}^s \sum_{i=1}^{k} a_{ss}^{i-1} y_{t-i}^f - a_{ss}^k y_{t-k}^s y_{t-k}^s) \end{pmatrix}$$

and I_2 stands for $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

It is important to mention the reason behind the choice of these two particular second moments. First of all, if we consider the following four second moments and their finite sample analogs:

$$\widehat{\mathbb{E}(y_t^f y_{t-k}^f)} = \frac{1}{T-1} \sum y_i^f y_{i-k}^f, \quad (11)$$

$$\widehat{\mathbb{E}(y_t^f y_{t-k}^{s'})} = \frac{1}{T-1} \sum y_i^f y_{i-k}^{s'}, \quad (12)$$

$$\widehat{\mathbb{E}(y_t^{s'} y_{t-k}^{f'})} = \frac{1}{\lfloor \frac{T}{N} \rfloor} \sum y_i^{s'} y_{i-k}^{f'}, \tag{13}$$

$$\widehat{\mathbb{E}(y_t^{s'} y_{t-k}^{s'})} = \frac{1}{\lfloor \frac{T}{N} \rfloor} \sum y_i^{s'} y_{i-k}^{s'}, \tag{14}$$

where N is the number of periods between two observations of the slow frequency variable, T is the number of periods, $y_i^{j'}, j \in \{f, s\}$ denotes those observations when both process can be observed and $\lfloor a \rfloor$ is the integer part of number a. $k \in \mathbb{N}$ since if $k = 0$, then $\mathbf{g}(\boldsymbol{\theta})$ is independent of a_{sf}, a_{ss} which have to be estimated.

As we can see from the equations (11), (12), (13), (14), the first two second moments (equations (11), (12)) are sums of $T-1$ elements (the sum goes from $t = 2$ since $k \in \mathbb{N}$), while the second two moments (equations (13), (14)), which I use in the GMM estimation, is a sum of $\lfloor \frac{T}{N} \rfloor$ elements, which is the number of observable slow frequency variable(s). However, if we used the first two moments in the GMM estimator, then the estimated a_{ff} and a_{sf} should be included in the estimation, while the last two second moments does not include the OLS estimates of those system parameters.

Moreover, it is also important to mention that k can be not a constant in the equations (13), (14). Hence, even if N is not constant, e.g. when one wants to estimate the daily variables using monthly variable, then $N \in \{20, 21, 22, 23\}$, we can use OLS, equation (5) to estimate a_{ff}, a_{fs} and the above defined GMM estimator, equation (10) to estimate a_{sf}, a_{ss}. Furthermore, if N is not constant, then k is also not constant in equations (13) and (14), but always the smallest $N \in \mathbb{N}$ for which both variable can be observed.

Therefore, this estimation method makes it feasible to work with unevenly spaced slow frequency data.

3.2.2 Estimation of the Innovation Variance Matrix

To estimate the innovation covariance matrix, the errors need to be estimated. That is why, first y_t^s has to be estimated for $\forall t$ when it is not observed. Using the estimated system parameters $(\hat{a}_{ff}, \hat{a}_{fs}, \hat{a}_{sf}, \hat{a}_{ss})$, the y_t^s can be estimated between two period when y_t^s is observed.

Denote the estimated slow frequency series by $\tilde{y}_t^s$, where

- $\tilde{y}_t^s = y_t^s$ if at t, y_t^s is observed,
- $\tilde{y}_t^s = \hat{a}_{sf} y_{t-1}^f + \hat{a}_{ss} \tilde{y}_{t-1}^s$, if y_t^s is not observed. ($\tilde{y}_{t-1}^s$ is already estimated, that is why, it can be used.)

Using the estimated $\tilde{y}_t^s$ and the estimated system parameters, we can estimate both series by equation (2). By estimates of $\hat{y}_t^f, \hat{y}_t^s$, we can calculate $\hat{\nu}_t^f = y_t^f - \hat{y}_t^f$, $\hat{\nu}_t^s = \tilde{y}_t^s - \hat{y}_t^s (= y_t^s - \hat{y}_t^s$, if at t the slow process can be observed) and therefore

$$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\nu}} = \begin{pmatrix} \mathrm{Var}(\hat{\nu}_t^f) & \mathrm{Cov}(\hat{\nu}_t^f, \hat{\nu}_t^s) \\ \mathrm{Cov}(\hat{\nu}_t^s, \hat{\nu}_t^f) & \mathrm{Var}(\hat{\nu}_t^s) \end{pmatrix}, \tag{15}$$

and

$$\widehat{\boldsymbol{\Sigma}}'_{\boldsymbol{\nu}} = \begin{pmatrix} \mathrm{Var}(\hat{\nu}_t^f) & \mathrm{Cov}(\hat{\nu}_t^{f'}, \hat{\nu}_t^{s'}) \\ \mathrm{Cov}(\hat{\nu}_t^{s'}, \hat{\nu}_t^{f'}) & \mathrm{Var}(\hat{\nu}_t^{s'}) \end{pmatrix} \tag{16}$$

can be obtained, where $\hat{\nu}_t^{i'}, i \in \{f, s\}$ is the estimated error term when both series can be observed.

That is, I will also test which estimation is better for the innovation covariance matrix. The first estimation of $\boldsymbol{\Sigma}_{\boldsymbol{\nu}}$ uses the estimated $\tilde{y}_t^s$ as the true value since y_t^s is not always observed. While the second estimation only includes those periods when y_t^s can be observed.

4 Simulation and Results

To test the finite sample performance of the estimation method, I run simulations. Throughout my thesis, I use MATLAB R2013a version 8.1.0.604. I run 1000 simulations and calculate the maximum, minimum, median, mean and standard deviations of the estimated parameters. The time series dimension is $T = 501$.

The steps of the simulation are:

1. Choose N, which is, the number of periods between the realization of the slow process and the time series dimension T = 501.

2. Define $\mathbf{A} = \begin{pmatrix} a_{ff} & a_{fs} \\ a_{sf} & a_{ss} \end{pmatrix}$.

3. Define $\mathbf{\Sigma}_{\boldsymbol{\nu}}$.

4. Draw the initial values of y_1^f, and y_1^s from normal distribution. That is, we draw a starting value from a standard normal distribution: $\begin{pmatrix} y_1^f \\ y_1^s \end{pmatrix} \sim N(0, \mathbf{\Sigma}_{\boldsymbol{y}}(0))$, where $vec(\mathbf{\Sigma}_{\boldsymbol{y}}(0)) = (I_4 - \mathbf{A} \otimes \mathbf{A})^{-1} vec(\mathbf{\Sigma}_{\boldsymbol{\nu}})$[1]. (Anderson et al., 2012, p. 187)

5. Generate ν_t^f and ν_t^s for $\forall t$.

6. Simulate y_t^f and y_t^s for $t = 2, \ldots, T$, using $\mathbf{A}$ and $\begin{pmatrix} \nu_t^f \\ \nu_t^s \end{pmatrix}$.

7. Since y_t^s is observable only for $\forall t \in N\mathbb{Z}$, we set every non–observable elements of y^s to NaN (Not–A–Number).

The steps of estimating the system parameters $(a_{ff}, a_{fs}, a_{sf}, a_{ss})$ are:

1. Estimate the a_{ff}, a_{fs} with OLS using equation (3).

2. Estimate a_{sf}, a_{ss} by the described GMM estimator using $\hat{a}_{ff}, \hat{a}_{fs}$. To obtain the GMM estimates by means of (10), use MATLAB fminsearch function, which finds a local minimum of a given function (equation (10)) with respect to initial values. I set these initial values to $\begin{pmatrix} a_{sf} & a_{ss} \end{pmatrix}$, which are the true values of system parameters.

---

[1] *vec* is defined as: $\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \\ d_{31} & d_{32} \end{pmatrix}$ then $vec(\boldsymbol{D}) = \begin{pmatrix} d_{11} \\ d_{12} \\ d_{21} \\ d_{22} \\ d_{31} \\ d_{32} \end{pmatrix}$

The steps to estimate $\boldsymbol{\Sigma_\nu}$ are:

1. Since y_t^s is not observed $\forall t$, use $\hat{\mathbf{A}}$ to calculate a $\tilde{y}_t^s$ series, where
   - $\tilde{y}_t^s = y_t^s$ if at t, y_t^s is observed,
   - $\tilde{y}_t^s = \hat{a}_{sf} y_{t-1}^f + \hat{a}_{ss} \tilde{y}_{t-1}^s$, if y_t^s is not observed. ($\tilde{y}_{t-1}^s$ is already estimated, that is why, it can be used.)
2. Using the estimated $\tilde{y}_t^s$ series and $\hat{\mathbf{A}}$, estimate $\hat{y}_t^f, \hat{y}_t^s$
3. Calculate the residuals: $\hat{\nu}_t^f = y_t^f - \hat{y}_t^f$, $\hat{\nu}_t^s = y_t^s - \hat{y}_t^s$
4. From $\hat{\nu}_t^f, \hat{\nu}_t^s$, get $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\nu}} = \begin{pmatrix} \mathrm{Var}(\hat{\nu}_t^f) & \mathrm{Cov}(\hat{\nu}_t^f, \hat{\nu}_t^s) \\ \mathrm{Cov}(\hat{\nu}_t^s, \hat{\nu}_t^f) & \mathrm{Var}(\hat{\nu}_t^s) \end{pmatrix}$ (see equation (15)), and $\widehat{\boldsymbol{\Sigma}}'_{\boldsymbol{\nu}} = \begin{pmatrix} \mathrm{Var}(\hat{\nu}_t^f) & \mathrm{Cov}(\hat{\nu}_t^{f'}, \hat{\nu}_t^{s'}) \\ \mathrm{Cov}(\hat{\nu}_t^{s'}, \hat{\nu}_t^{f'}) & \mathrm{Var}(\hat{\nu}_t^{s'}) \end{pmatrix}$ (see equation (16))

4.1 Impact of N on the Estimates

A further goal of this simulation analysis is to investigate the impact of N (the number of periods between two observation of the slow frequency process) on the estimation of $\mathbf{A}$ and $\boldsymbol{\Sigma_\nu}$.

If $N > 1$, then a VAR–model, where $N = 1$ becomes an MF–VAR model. We claim, the larger the N, the more information is lost, therefore one can expect the estimate of the parameters becomes less and less accurate as the N is increasing. That is why, I run simulations with $N = 2, 5, 10, 20$ to see in a my simulation runs how the mean, median, maximum, minimum and standard deviation of the estimated parameters changes with increasing N. In practice if monthly and quarterly data are used, then $N = 3$ is relevant, since quarterly (slow frequency) data can be observed at every third month (fast frequency data). While if daily and monthly data are used, then $N = \{20, 21, 22, 23\}$ can be relevant.

To test the effect of N, the true values of $\mathbf{A}$ and $\boldsymbol{\Sigma_\nu}$ have to be given. In the simulations I choose an $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ with eigenvalues within the unit circle, that is, the system is stable, and a positive definite $\boldsymbol{\Sigma_\nu}$:

$$\mathbf{A}_1 = \begin{pmatrix} 0.1 & 0.02 \\ 0.05 & 0.07 \end{pmatrix},$$

$$\mathbf{A}_2 = \begin{pmatrix} 0.5 & -0.5 \\ 0.5 & 0.7 \end{pmatrix},$$

$$\mathbf{A}_3 = \begin{pmatrix} 0.95 & 0.1 \\ -0.2 & 0.85 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\nu}} = \begin{pmatrix} 1 & -0.2 \\ -0.2 & 1 \end{pmatrix},$$

where the absolute values of the eigenvalues of $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ are (0.1200, 0.0500), (0.7746, 0.7746), (0.9097, 0.9097) respectively. We expect the higher the persistence of the system, that is, the higher the absolute value of the eigenvalues, the more accurate the estimations.

4.2 Estimation results

4.2.1 Estimation of System Parameters $(a_{ff}, a_{fs}, a_{sf}, a_{ss})$

Tables 1, 2 and 3 include the results of the system parameter estimations for $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ after 1000 simulations.The true system parameters are in the first row. The N which sets the number of periods between two observations of the slow process is in the very first column. The blocks, which are separated by lines, correspond to different Ns.

As it can be expected with higher Ns, the estimation becomes worse. That is, the range and the standard deviation of the estimations increases for larger N, as more value of the slow process can be observed if N is smaller, we have to estimate less unobserved values. Therefore, we can expect that the estimation will be also better if N is smaller, as in Tables 1, 2, 3.

Moreover, in the estimation of the system parameters of $\mathbf{A}_1$, which has the smallest eigenvalues, the standard deviation of the estimations are larger for every N than in case of $\mathbf{A}_2$, $\mathbf{A}_3$ as expected. To understand this result, we should take a closer look at general formula of the OLS parameter variance matrix (see e.g.

Hansen, 2015, p. 90):

$$\begin{pmatrix} Var(\hat{a}_{ff}) \\ Var(\hat{a}_{fs}) \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\nu_f^2$$

where $\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{\mathbb{T}} \end{pmatrix}$ and x_i are the component of the OLS estimation (equation (5)). Moreover, $\mathbf{X}^T\mathbf{X} = \mathbf{\Sigma}_{\boldsymbol{y}}(0)$ can be calculated by the following formula: $vec(\mathbf{\Sigma}_{\boldsymbol{y}}(0)) = (I_4 - \mathbf{A} \otimes \mathbf{A})^{-1}vec(\mathbf{\Sigma}_{\boldsymbol{\nu}})$ (Anderson et al., 2012, p. 187). So it can be seen that if the system is more persistent, that is, the absolute value of the eigenvalues of the system matrix is larger, then the variance of the estimates are smaller. That is, the more persistent the system, the smaller the standard deviation of the system parameter estimates.

It is also important to mention that different methods have been used to estimate a_{ff}, a_{fs} and a_{sf}, a_{ss}, which affects the results. For every N, the mean and the median are closer to the true value for a_{ff}, a_{fs} which are estimated by OLS. Similarly, the range of the estimated parameters is a smaller for a_{ff}, a_{fs}, estimated by OLS, then for the a_{sf}, a_{ss}, estimated by GMM. These differences can be explained by the different ways to compute the estimates. The MATLAB fminsearch function uses a numerical approximation while for OLS the explicit formula is used. That is, while we can use an analytical formula to estimate a_{ff}, a_{fs}, the estimation of a_{sf}, a_{ss} is performed by tremendous numerical calculations, which may affect the results.

Table 1: System parameter estimation with mixed frequency data for matrix $\mathbf{A}_1$

		$a_{ff} = 0.1$	$a_{fs} = 0.02$	$a_{sf} = 0.05$	$a_{ss} = 0.07$
N = 2	Mean	0.0967	0.0201	0.0883	0.1054
	Median	0.0949	0.0215	0.0613	0.0888
	Maximum	0.2884	0.2271	24.1045	0.4352
	Minimum	-0.0878	-0.1890	-11.3504	-0.2987
	Std Dev	0.0631	0.0673	1.5627	0.1360
N = 5	Mean	0.0997	0.0234	0.1402	0.0866
	Median	0.1038	0.0216	0.1409	0.1771
	Maximum	0.4399	0.3677	16.9357	0.7189
	Minimum	-0.2212	-0.2982	-12.7485	-0.7920
	Std Dev	0.1074	0.1061	1.6124	0.4066
N = 10	Mean	0.0996	0.0213	0.0654	0.0845
	Median	0.0952	0.0233	0.1292	0.1306
	Maximum	0.5671	0.4514	20.8074	0.9116
	Minimum	-0.3244	-0.4520	-16.5008	-0.8849
	Std Dev	0.1467	0.1506	1.9143	0.5332
N = 20	Mean	0.1025	0.0215	0.0827	0.0659
	Median	0.1019	0.0200	0.1539	0.1213
	Maximum	0.9444	0.9533	37.9083	0.9690
	Minimum	-0.5766	-0.7740	-16.5420	-1.0129
	Std Dev	0.2162	0.2195	1.9760	0.5631

Notes: The true value of the system parameters is in the first row. The first column specifies the different Ns which is the number of periods between the realization of slow process. Each block (corresponding to different N) includes the mean, median, maximum, minimum and standard deviation of the estimated system parameters after 1000 simulations. The number of periods is 500.

Table 2: System parameter estimation with mixed frequency data for matrix $\mathbf{A}_2$

		a_{ff} = 0.5	a_{fs} = -0.5	a_{sf} = 0.5	a_{ss} = 0.7
N = 2	Mean	0.4984	-0.4995	0.4990	0.6986
	Median	0.4963	-0.5001	0.4986	0.7004
	Maximum	0.6225	-0.3652	0.6239	0.8250
	Minimum	0.3809	-0.6515	0.3753	0.5829
	Std Dev	0.0402	0.0424	0.0357	0.0365
N = 5	Mean	0.4975	-0.4979	0.5346	0.6817
	Median	0.4939	-0.5009	0.5254	0.7033
	Maximum	0.7403	-0.2986	1.0997	0.9137
	Minimum	0.2522	-0.6920	0.1979	-0.5823
	Std Dev	0.0689	0.0660	0.1145	0.1278
N = 10	Mean	0.4982	-0.5001	0.5087	0.5279
	Median	0.4932	-0.4996	0.4605	0.6842
	Maximum	0.8226	-0.1282	3.0319	1.0045
	Minimum	0.1954	-0.8085	-1.7517	-0.8685
	Std Dev	0.0980	0.0963	0.5829	0.4230
N = 20	Mean	0.4964	-0.5026	0.5383	0.4911
	Median	0.4931	-0.5026	0.4725	0.6978
	Maximum	0.8981	0.0378	4.6286	1.0424
	Minimum	0.1115	-0.9857	-2.0127	-0.9675
	Std Dev	0.1400	0.1434	0.6353	0.5123

Notes: The true value of the system parameters is in the first row. The first column specifies the different Ns which is the number of periods between the realization of slow process. Each block (corresponding to different N) includes the mean, median, maximum, minimum and standard deviation of the estimated system parameters after 1000 simulations. The number of periods is 500.

Table 3: System parameter estimation with mixed frequency data for matrix $\mathbf{A}_3$

		a_{ff} = 0.95	a_{fs} = 0.1	a_{sf} = -0.2	a_{ss} = 0.85
N = 2	Mean	0.9489	0.1011	-0.2030	0.8462
	Median	0.9496	0.1007	-0.2022	0.8475
	Maximum	1.0339	0.1797	-0.1375	0.8987
	Minimum	0.8344	0.0103	-0.2961	0.7595
	Std Dev	0.0309	0.0259	0.0231	0.0197
N = 5	Mean	0.9492	0.1018	-0.2038	0.8454
	Median	0.9500	0.1013	-0.2026	0.8474
	Maximum	1.0780	0.2391	-0.1075	0.9036
	Minimum	0.7858	-0.0487	-0.3448	0.7281
	Std Dev	0.0465	0.0388	0.0297	0.0242
N = 10	Mean	0.9486	0.0968	-0.2100	0.8415
	Median	0.9482	0.0965	-0.2037	0.8494
	Maximum	1.1719	0.3254	-0.0482	0.9168
	Minimum	0.7529	-0.0849	-0.7329	-0.7687
	Std Dev	0.0682	0.0572	0.0622	0.0640
N = 20	Mean	0.9481	0.0976	-0.3416	0.7373
	Median	0.9474	0.1008	-0.2288	0.8500
	Maximum	1.3736	0.3902	2.2425	0.9758
	Minimum	0.6381	-0.2282	-5.3750	-0.9416
	Std Dev	0.0974	0.0858	0.4691	0.3427

Notes: The true value of the system parameters is in the first row. The first column specifies the different Ns which is the number of periods between the realization of slow process. Each block (corresponding to different N) includes the mean, median, maximum, minimum and standard deviation of the estimated system parameters after 1000 simulations. The number of periods is 500.

4.2.2 Estimation of Noise Parameters $(\Sigma_{ff}, \Sigma_{fs}, \Sigma_{sf}, \Sigma_{ss})$

Tables 4, 5 and 6 include the results of the noise parameter estimations after 1000 simulation runs. The block columns belong to the different estimation methods of the noise parameters which specified in the first row of the tables ($\widehat{\mathbf{\Sigma}}_{\boldsymbol{\nu}}$ equation (15) and $\widehat{\mathbf{\Sigma}}'_{\boldsymbol{\nu}}$ equation (16)). The true parameter values are in the second row. The number of periods between two observations of the slow process is in the very first column. The blocks correspond to different Ns and estimation methods.

Similar phenomenon can be observed in the estimation of the noise parameters as of the system parameter. That is, the larger the N, the worse the estimation since both the range and the standard deviation of the estimates increases in N.

An other eye–catching phenomenon is the accuracy of the estimations between Σ_{ff} and the other noise parameters $(\Sigma_{fs}, \Sigma_{sf}, \Sigma_{ss})$ when we do not take $\tilde{y}^s$ as true value but use only the $y^{s'}$ to estimate the parameters. To understand these differences, the estimations of the different noise parameters have to be investigated. As known, $\widehat{\Sigma}_{ff} = \mathrm{Var}(\hat{\nu}_t^f)$, where $\hat{\nu}_t^f = y_t^f - \hat{y}_t^f$, while $\widehat{\Sigma}'_{fs} = \widehat{\Sigma}'_{sf} = \mathrm{Cov}(\hat{\nu}_t^{s'}, \hat{\nu}_t^{f'})$ and $\widehat{\Sigma}'_{ss} = \mathrm{Var}(\hat{\nu}_t^{s'})$, where $\hat{\nu}_t^{s'} = y_t^{s'} - \hat{y}_t^{s'}$ (as before $x_t^{i'}$, $x \in \{y, \nu\}, i \in \{f, s\}$ is the value of the parameters when both process can be observed).

Hence, when Σ_{ff} is estimated, the whole time period is taken into consideration, however, when the other noise parameters $(\Sigma_{fs}, \Sigma_{sf}, \Sigma_{ss})$ are estimated, only every N time period is considered. Of course, it would not be problem if there were two high frequency processes. But with mixed frequency data, when the estimated values of the slow process is used to estimate the processes, and particularly for the estimation of $\hat{y}_t^{s'} = a_{sf} y_{t-1}^f + a_{ss} \tilde{y}_{t-1}^s$ an estimated value is always used, the estimations are bad.

However, when I estimate the noise parameters using the whole time period, the estimations are better for the $\Sigma_{fs}, \Sigma_{sf}, \Sigma_{ss}$ than for the other estimation method since the standard deviation and the range of the estimations are smaller. The estimation of Σ_{ff} remains the same but the estimation of the other noise parameters $(\Sigma_{fs}, \Sigma_{sf}, \Sigma_{ss})$ is changed to $\widehat{\Sigma}_{fs} = \widehat{\Sigma}_{sf} = \mathrm{Cov}(\hat{\nu}_t^s, \hat{\nu}_t^f)$ and $\widehat{\Sigma}_{ss} = \mathrm{Var}(\hat{\nu}_t^s)$, where $\hat{\nu}_t^s = \tilde{y}_t^s - \hat{y}_t^s$. So it is assumed that the estimated values are the true values of the unobserved of the slow frequency process.

Table 4: Noise parameter estimation with mixed frequency data for matrix $\mathbf{A}_1$

		$\widehat{\boldsymbol{\Sigma}}'_{\boldsymbol{\nu}}$				$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\nu}}$			
		$\boldsymbol{\Sigma_{ff}} = 1$	$\boldsymbol{\Sigma_{fs}} = -0.2$	$\boldsymbol{\Sigma_{sf}} = -0.2$	$\boldsymbol{\Sigma_{ss}} = 1$	$\boldsymbol{\Sigma_{ff}} = 1$	$\boldsymbol{\Sigma_{fs}} = -0.2$	$\boldsymbol{\Sigma_{sf}} = -0.2$	$\boldsymbol{\Sigma_{ss}} = 1$
N = 2	Mean	0.9962	-0.2014	-0.2014	3.4873	0.9962	-0.1007	-0.1007	1.7493
	Median	0.9942	-0.2040	-0.2040	1.0945	0.9942	-0.1018	-0.1018	0.5488
	Maximum	1.2070	0.9388	0.9388	624.4652	1.2070	0.5709	0.5709	314.5604
	Minimum	0.8107	-1.3776	-1.3776	0.7465	0.8107	-0.6870	-0.6870	0.3733
	Std Dev	0.0655	0.1313	0.1313	22.0201	0.0655	0.0666	0.0666	11.0820
N = 5	Mean	1.0261	-0.1931	-0.1931	3.8526	1.0261	-0.0387	-0.0387	0.7768
	Median	1.0105	-0.1921	-0.1921	1.3832	1.0105	-0.0387	-0.0387	0.2793
	Maximum	2.1474	4.7905	4.7905	240.8068	2.1474	0.9547	0.9547	48.1869
	Minimum	0.8082	-7.1944	-7.1944	0.5806	0.8082	-1.4362	-1.4362	0.1162
	Std Dev	0.1193	0.4266	0.4266	11.7843	0.1193	0.0856	0.0856	2.3686
N = 10	Mean	1.0690	-0.1757	-0.1757	5.7575	1.069	-0.0176	-0.0176	0.5839
	Median	1.0361	-0.1938	-0.1938	1.6718	1.0361	-0.0195	-0.0195	0.1712
	Maximum	2.4330	14.3190	14.3190	631.9262	2.433	1.4474	1.4474	63.214
	Minimum	0.8119	-9.2224	-9.2224	0.5107	0.8119	-0.906	-0.906	0.0511
	Std Dev	0.1732	0.8939	0.8939	25.9373	0.1732	0.0904	0.0904	2.6033
N = 20	Mean	1.1241	-0.1671	-0.1671	5.1392	1.1241	-0.0086	-0.0086	0.2712
	Median	1.0700	-0.1506	-0.1506	1.5343	1.07	-0.0083	-0.0083	0.0788
	Maximum	3.4278	9.0843	9.0843	1262.7794	3.4278	0.4581	0.4581	70.1367
	Minimum	0.8332	-6.8827	-6.8827	0.3810	0.8332	-0.3468	-0.3468	0.0191
	Std Dev	0.2298	0.8583	0.8583	41.2464	0.2298	0.0442	0.0442	2.2794

Notes: The first row specifies the estimation method used to estimate the innovation covariance matrix. The true value of the system parameters is in the second row. The first column specifies the different Ns which is the number of periods between the realization of slow process. Each block (corresponding to different N) includes the mean, median, maximum, minimum and standard deviation of the estimated system parameters after 1000 simulations. The number of periods is 500.

Table 5: Noise parameter estimation with mixed frequency data for matrix $\mathbf{A}_2$

		$\widehat{\boldsymbol{\Sigma}}'_{\nu}$				$\widehat{\boldsymbol{\Sigma}}_{\nu}$			
		$\boldsymbol{\Sigma}_{ff} = 1$	$\boldsymbol{\Sigma}_{fs} = -0.2$	$\boldsymbol{\Sigma}_{sf} = -0.2$	$\boldsymbol{\Sigma}_{ss} = 1$	$\boldsymbol{\Sigma}_{ff} = 1$	$\boldsymbol{\Sigma}_{fs} = -0.2$	$\boldsymbol{\Sigma}_{sf} = -0.2$	$\boldsymbol{\Sigma}_{ss} = 1$
N = 2	Mean	1.1201	-0.5439	-0.5439	1.4712	1.1201	-0.2724	-0.2724	0.7371
	Median	1.1204	-0.5403	-0.5403	1.4675	1.1204	-0.2708	-0.2708	0.7358
	Maximum	1.3662	-0.2791	-0.2791	1.9279	1.3662	-0.1396	-0.1396	0.9640
	Minimum	0.8965	-0.8630	-0.8630	1.0075	0.8965	-0.4315	-0.4315	0.5038
	Std Dev	0.0747	0.0965	0.0965	0.1349	0.0747	0.0483	0.0483	0.0675
N = 5	Mean	1.3337	-0.9012	-0.9012	2.0467	1.3337	-0.1811	-0.1811	0.4126
	Median	1.3209	-0.8687	-0.8687	1.9699	1.3209	-0.1747	-0.1747	0.3966
	Maximum	1.8671	-0.2549	-0.2549	5.3487	1.8671	-0.0479	-0.0479	1.0702
	Minimum	1.0086	-2.3156	-2.3156	1.1745	1.0086	-0.4627	-0.4627	0.2357
	Std Dev	0.1154	0.2708	0.2708	0.4718	0.1154	0.0543	0.0543	0.0949
N = 10	Mean	1.7739	-1.3640	-1.3640	3.4379	1.7739	-0.1389	-0.1389	0.3498
	Median	1.6132	-1.0819	-1.0819	2.6107	1.6132	-0.1097	-0.1097	0.2643
	Maximum	4.1763	0.9129	0.9129	30.8545	4.1763	0.1196	0.1196	3.0903
	Minimum	1.0933	-10.7936	-10.7936	0.9018	1.0933	-1.0798	-1.0798	0.0948
	Std Dev	0.4862	1.0160	1.0160	2.5971	0.4862	0.1031	0.1031	0.2635
N = 20	Mean	1.9821	-1.5082	-1.5082	3.9808	1.9821	-0.0775	-0.0775	0.2056
	Median	1.7358	-1.0568	-1.0568	2.6183	1.7358	-0.0541	-0.0541	0.1363
	Maximum	7.5778	11.4032	11.4032	88.5508	7.5778	0.5786	0.5786	4.4665
	Minimum	1.1765	-20.1236	-20.1236	0.6461	1.1765	-0.9862	-0.9862	0.0328
	Std Dev	0.7431	1.7966	1.7966	4.9273	0.7431	0.0914	0.0914	0.2531

Notes: The first row specifies the estimation method used to estimate the innovation covariance matrix. The true value of the system parameters is in the second row. The first column specifies the different Ns which is the number of periods between the realization of slow process. Each block (corresponding to different N) includes the mean, median, maximum, minimum and standard deviation of the estimated system parameters after 1000 simulations. The number of periods is 500.

Table 6: Noise parameter estimation with mixed frequency data for matrix $\mathbf{A}_3$

		$\widehat{\boldsymbol{\Sigma}}'_{\boldsymbol{\nu}}$				$\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\nu}}$			
		$\boldsymbol{\Sigma_{ff}}=1$	$\boldsymbol{\Sigma_{fs}}=-0.2$	$\boldsymbol{\Sigma_{sf}}=-0.2$	$\boldsymbol{\Sigma_{ss}}=1$	$\boldsymbol{\Sigma_{ff}}=1$	$\boldsymbol{\Sigma_{fs}}=-0.2$	$\boldsymbol{\Sigma_{sf}}=-0.2$	$\boldsymbol{\Sigma_{ss}}=1$
N = 2	Mean	1.0022	-0.1153	-0.1153	1.6982	1.0022	-0.0576	-0.0576	0.8508
	Median	1.0009	-0.1190	-0.1190	1.6916	1.0009	-0.0597	-0.0597	0.8483
	Maximum	1.2031	0.1316	0.1316	2.2299	1.2031	0.0657	0.0657	1.1152
	Minimum	0.8051	-0.3748	-0.3748	1.1562	0.8051	-0.1897	-0.1897	0.5781
	Std Dev	0.0654	0.0843	0.0843	0.1548	0.0654	0.0422	0.0422	0.0775
N = 5	Mean	1.0220	0.0191	0.0191	2.8292	1.0220	0.0040	0.0040	0.5704
	Median	1.0235	0.0180	0.0180	2.8156	1.0235	0.0035	0.0035	0.568
	Maximum	1.2484	0.5415	0.5415	4.2259	1.2484	0.1086	0.1086	0.848
	Minimum	0.8154	-0.6492	-0.6492	1.7950	0.8154	-0.1303	-0.1303	0.3609
	Std Dev	0.0687	0.1726	0.1726	0.3964	0.0687	0.0345	0.0345	0.0797
N = 10	Mean	1.0482	0.0883	0.0883	3.4173	1.0482	0.0091	0.0091	0.348
	Median	1.0432	0.0889	0.0889	3.3791	1.0432	0.0090	0.0090	0.3436
	Maximum	1.4029	1.1900	1.1900	10.1940	1.4029	0.1211	0.1211	1.0422
	Minimum	0.8290	-1.2071	-1.2071	1.5367	0.8290	-0.1227	-0.1227	0.155
	Std Dev	0.0770	0.2793	0.2793	0.7494	0.0770	0.0279	0.0279	0.0757
N = 20	Mean	1.1537	0.7362	0.7362	11.8173	1.1537	0.0371	0.0371	0.6143
	Median	1.0952	0.2024	0.2024	4.3336	1.0952	0.0111	0.0111	0.2229
	Maximum	7.5599	154.5911	154.5911	3553.3702	7.5599	7.5118	7.5118	182.8047
	Minimum	0.8322	-2.6791	-2.6791	1.1417	0.8322	-0.1765	-0.1765	0.0587
	Std Dev	0.3929	5.6741	5.6741	114.6614	0.3929	0.2790	0.2790	5.9075

Notes: The first row specifies the estimation method used to estimate the innovation covariance matrix. The true value of the system parameters is in the second row. The first column specifies the different Ns which is the number of periods between the realization of slow process. Each block (corresponding to different N) includes the mean, median, maximum, minimum and standard deviation of the estimated system parameters after 1000 simulations. The number of periods is 500.

5 Conclusions and further research

In this study, I investigated mixed frequency vector autoregressive models of order 1 (MF–VAR(1)), described in Anderson et al. (2012).

First, the system parameters were estimated by ordinary least squares and the general method of moments. The estimation methods which are used in this study work for cases, when N, which is the number of periods between two observations of the slow frequency process, is not constant[2]. That is, it can be also applied if the high frequency variable is daily and the low frequency variable is monthly.

Then I investigated how the periods, N between two observations of the slow frequency process affect the properties of the estimations and how the persistence of the system has an impact on the estimation. To test the finite sample performance of the estimation methods, a Monte Carlo study with different set–ups was performed.

According to my findings, the larger the number of periods between two observations of the slow frequency process, the worse the estimation since the both the range and the standard deviation of the estimates are larger. Moreover, the persistence of the system has a positive impact on the estimates, as could be expected.

An obvious extension would be to derive similar estimation methods, which allow of variable N, when the processes are multidimensional, that is, instead of scalars, vector variables would be used. Furthermore, we considered only the MF–VAR model of order 1, which usually would not be enough to do careful empirical research. Therefore, it should be augmented to the general case when the order can be any $p > 1$.

[2]If N is not constant, then it could take different values. E.g. if there is a monthly and a daily variable, then N would be 20, 21, 22, 23 depending on the number of business days.

References

Anderson, B., M. Deistler, E. Felsenstein, B. Funovits, L. Koelbl, and M. Zamani (2015): "Multivariate AR systems and mixed frequency data: g–identifiability and estimation," *Econometric Theory*, 1–34.

Anderson, B., M. Deistler, E. Felsenstein, B. Funovits, P. Zadrozny, M. Eichler, W. Chen, and M. Zamani (2012): "Identifiability of regular and singular multivariate autoregressive models from mixed frequency data," in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, 184–189.

Baffigi, A., R. Golinelli, and G. Parigi (2004): "Bridge models to forecast the euro area GDP," *International Journal of forecasting*, 20, 447–460.

Foroni, C. and M. G. Marcellino (2013): "A survey of econometric methods for mixed-frequency data," *Available at SSRN 2268912*.

Ghysels, E., A. Sinko, and R. Valkanov (2007): "MIDAS regressions: Further results and new directions," *Econometric Reviews*, 26, 53–90.

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

Hansen, B. E. (2015): "Econometrics," Manuscript.

Kuzin, V., M. Marcellino, and C. Schumacher (2011): "MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro Area," *International Journal of Forecasting*, 27, 529–542.

Lütkepohl, H. (2005): *New Introduction to Multiple Time Series Analysis*, Springer.

Merton, R. C. (1973): "An intertemporal capital asset pricing model," *Econometrica: Journal of the Econometric Society*, 867–887.

Ye, M., J. Zyren, and J. Shore (2006): "Forecasting short-run crude oil price using high-and low-inventory variables," *Energy Policy*, 34, 2736–2743.

A Appendix: calculations

A.1 Calculation of y_t^s as a function of y_k^s and $y_t^f, \dots, y_{t-k}^f$

By recursive calculation, from (4) we get:

$$\begin{aligned}
y_t^s &= a_{sf} y_{t-1}^f + a_{ss} y_{t-1}^s + \nu_t^s \\
&= a_{sf} y_{t-1}^f + a_{ss}(a_{sf} y_{t-2}^f + a_{ss} y_{t-2}^s + \nu_{t-1}^s) + \nu_t^s \\
&= a_{sf} y_{t-1}^f + a_{ss} a_{sf} y_{t-2}^f + a_{ss}^2 y_{t-2}^s + a_{ss} \nu_{t-1}^s + \nu_t^s \\
&= a_{sf}(y_{t-1}^f + a_{ss} y_{t-2}^f) + a_{ss}^2 y_{t-2}^s + a_{ss} \nu_{t-1}^s + \nu_t^s \\
&= a_{sf}(y_{t-1}^f + a_{ss} y_{t-2}^f) + a_{ss}^2(a_{sf} y_{t-3}^f + a_{ss} y_{t-3}^s + \nu_{t-2}^s) + a_{ss} \nu_{t-1}^s + \nu_t^s \\
&= a_{sf}(y_{t-1}^f + a_{ss} y_{t-2}^f + a_{ss}^2 y_{t-3}^f) + a_{ss}^3 y_{t-3}^s + a_{ss}^2 \nu_{t-2}^s + a_{ss} \nu_{t-1}^s + \nu_t^s \\
&= \cdots = \\
&= a_{sf} \sum_{i=1}^{k} a_{ss}^{i-1} y_{t-i}^f + a_{ss}^k y_{t-k}^s + \sum_{i=1}^{k} a_{ss}^{i-1} \nu_{t+1-i}^s
\end{aligned}$$

A.2 Calculation of y_t^f as a function of y_k^s and $y_t^f, \dots, y_{t-k}^f$

From (3):

$$y_t^f = a_{ff} y_{t-1}^f + a_{fs} y_{t-1}^s + \nu_t^f$$

We also know from the previous calculation, that $y_{t-1}^s = a_{sf} \sum_{i=1}^{k-1} a_{ss}^{i-1} y_{t-1-i}^f + a_{ss}^{k-1} y_{t-k}^s + \sum_{i=1}^{k-1} a_{ss}^{i-1} \nu_{t-i}^s$. Substituting this into $y_t^f = \dots$ we get:

$$\begin{aligned}
y_t^f &= a_{ff} y_{t-1}^f + a_{fs}(a_{sf} \sum_{i=1}^{k-1} a_{ss}^{i-1} y_{t-1-i}^f + a_{ss}^{k-1} y_{t-k}^s + \sum_{i=1}^{k-1} a_{ss}^{i-1} \nu_{t-i}^s) + \nu_t^f \\
&= a_{ff} y_{t-1}^f + a_{fs} a_{sf} \sum_{i=1}^{k-1} a_{ss}^{i-1} y_{t-1-i}^f + a_{fs} a_{ss}^{k-1} y_{t-k}^s + a_{fs} \sum_{i=1}^{k-1} a_{ss}^{i-1} \nu_{t-i}^s + \nu_t^f .
\end{aligned}$$

B Appendix: Code

B.1 Simulation

```
clear; clc; close all;
global numberofday y_f y_s y_daynumber
global nDays1 nSlow y_s1 y_f1 nFast t_slow

% The dimension of the time series
T = 501;

```

```
% The frequency of slow process with respect to the fast process (N in the written thesis).
% So slow process exists at t = 1, 1+t_slow, 1+2*t_slow ...
t_slowseq = [2 5 10 20];

% To save results, for 3 dimensional matrix are constructed.
A_table = zeros(5, 4, 4);
Sigma_nu_table1 = zeros(5, 4, 4);
Sigma_nu_table2 = zeros(5, 4, 4);

% Number of simulation runs
numsim = 1000;

% To save the results of each simulation runs
A_par = zeros(numsim, 4, 4);
Sigma_nu_par1 = zeros(numsim, 4, 4);
Sigma_nu_par2 = zeros(numsim, 4, 4);

% % %  The true innovation variance matrix
nu_sigma = [1 -0.2; -0.2 1];

% The true A matrix
A_true = zeros(2,2);
A_true = [0.95 0.1; -0.2 0.85];

for kiter=1:4

% "First, initialize the random number generator to make the results in
% this example repeatable."
% http://www.mathworks.com/help/matlab/math/random-integers.html
rng(0,'twister');
t_slow = t_slowseq(kiter);

% We need a vector whose entry claims the length from the previous
% realization of a slow process (positive integers)
k = zeros(t_slow, 1);
for i = 1:t_slow;
    k(i) = i;
end;

k = repmat(k, floor(T/t_slow), 1);
```

```
k = [t_slow; k]; %% For the first entry t_slow would belong, but it doesn't matter since we start from t = 1+k to have t-k(=1).



%%% Create a vector with 1 if the slow process has a value, and NaN if
% only fast process exists.
y_snan= NaN(T, 1);
for i=1:t_slow:T;
    y_snan(i) = 1;
end;
%% Simulation starts here
for sim_i = 1:numsim  % The number of simulations
% Initialization of the two processes
% The first column is the value of the process
% The second column says the length of that entry from the previous
% realization of the slow process.
y_f = zeros(T, 2);
y_s = zeros(T, 2);

y_f(:, 2) = k;
y_s(:, 2) = k;



%%% First we generate the two initial value for y^f_1 and y^s_1. These
% also exist in the empirical part.
% Random number generator
% http://www.mathworks.com/help/matlab/random-number-generation.html
% Draw a number from standard normal distribution
vec_sigma_y = inv(eye(4) - kron(A_true, A_true)) * reshape(nu_sigma, 4, 1);
sigma_y = reshape(vec_sigma_y, 2, 2);
y_mu = [0, 0];
R_y = chol(sigma_y);
y_1 = repmat(y_mu,1,1) + randn(1,2)*R_y;
y_f(1, 1) = y_1(1);
y_s(1, 1) = y_1(2);

 %% Simulating error terms (e, which is \nu in the paper) y_f
```

```
% Bivariate Normal Random Numbers: http://www.mathworks.com/help/matlab/ref/randn.html#buf2cft-1
nu_mu = [0 0];
R = chol(nu_sigma);
% First column is e^f, second column is e^s
nu = repmat(nu_mu,T,1) + randn(T,2)*R;


%%% Simulating the series
for i=2:T;
    y_f(i,1) = A_true(1, 1) * y_f(i-1, 1) + A_true(1, 2) * y_s(i-1, 1) + nu(i,1);
    y_s(i, 1) = A_true(2, 1) * y_f(i-1, 1) + A_true(2, 2) * y_s(i-1, 1) + nu(i,2);
end;

% y_s1 is the y_s vector without NaN values:
y_s1 = y_s(:, 1).*y_snan;
y_s1(isnan(y_s1)) = [];

nFast = T; % by definition
nSlow = length(y_s1);


% We have to find out at which point of time slow process exists
% First we create a sequence from 1 to nFast with step size 1.
numbers = zeros(nFast, 1);
for i = 1:nFast;
    numbers(i) = i;
end;

% The days when slow and fast process exist.
y_daynumber = numbers.*y_snan;
y_daynumber(isnan(y_daynumber)) = [];


% The length of different periods is also necessary for the calculations (right now it is always t_slow)
numberofday = zeros(length(y_daynumber), 1);
for i = 1:(length(numberofday)-1);
    numberofday(i) = y_daynumber(i+1) - y_daynumber(i);
end;

y_s1 = [y_s1 numberofday];

```

```
%%% I create a "fast" frequency vector which has value only at those t when there is a
% value in the slow process. Basically, this is the pair of y_s1
y_f1 = y_f(:, 1).*y_snan;
y_f1(isnan(y_f1)) = [];

Ahat = zeros(2, 2);

%%% OLS estimation for a_ff and a_fs
X_ols = [y_f1(1:(end-1)) y_s1((1:end-1), 1)];
y_ols = y_f(2:end, 1).*y_snan(1:end-1);
y_ols(isnan(y_ols)) = [];

Ahat(1, :) = ((X_ols'*X_ols)\(X_ols'*y_ols))';
fDays = t_slow*ones(T,1);

global a_ff a_fs
a_ff = Ahat(1, 1);
a_fs = Ahat(1, 2);

nDays1 = fDays.*y_snan;
nDays1(isnan(nDays1)) = [];

Ahat(2, :) = fminsearch(@GMMmin, A_true(2, :));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%% Sigma_nu_hat %%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%% First we need to create a "new" slow process, where every element is
% observed. When y_s is known, then we leave it like that, but unkown y_s
% has to be estimated using the A_hat.

newy_s = y_s(:, 1).*y_snan;
for i = 2:nFast
    if isnan(newy_s(i)) == 1
     newy_s(i) = Ahat(2, 1) * y_f(i-1, 1) + Ahat(2, 2) * newy_s(i-1);
    else
     newy_s(i) = newy_s(i);
    end
end

y_fhat = zeros(T, 1);
```

```
y_shat = zeros(T, 1);

y_fhat(1) = y_f(1, 1);
y_shat(1) = newy_s(1); % Same as y_s(1, 1)
for i=2:T;
    y_fhat(i) = Ahat(1, 1) * y_f(i-1, 1) + Ahat(1, 2) *
       newy_s(i-1);
    y_shat(i) = Ahat(2, 1) * y_f(i-1, 1) + Ahat(2, 2) *
       newy_s(i-1);
end;

%%% We drop those y_shat terms which cannot be observed
y_shat1 = y_shat(:, 1).*y_snan;
y_shat1(isnan(y_shat1)) = [];

%%% y_shat1's counterpart
y_fhat1 = y_fhat(:, 1).*y_snan;
y_fhat1(isnan(y_fhat1)) = [];




nu_fhat = y_f(:, 1) - y_fhat;
nu_f1hat = y_f1 - y_fhat1;
nu_s1hat = y_s1(:, 1) - y_shat1;

nu_shat = newy_s - y_shat;


COV_nu_f1_nu_s1 = cov(nu_f1hat, nu_s1hat);
COV_nu_f_nu_s = cov(nu_fhat, nu_shat);



Sigma_nu_par1(sim_i, 1, kiter) = var(nu_fhat);
Sigma_nu_par1(sim_i, 2, kiter) = COV_nu_f1_nu_s1(1, 2);
Sigma_nu_par1(sim_i, 3, kiter) = COV_nu_f1_nu_s1(2, 1);
Sigma_nu_par1(sim_i, 4, kiter) = var(nu_s1hat);

Sigma_nu_par2(sim_i, 1, kiter) = var(nu_fhat);
Sigma_nu_par2(sim_i, 2, kiter) = COV_nu_f_nu_s(1, 2);
Sigma_nu_par2(sim_i, 3, kiter) = COV_nu_f_nu_s(2, 1);
Sigma_nu_par2(sim_i, 4, kiter) = var(nu_shat);

A_par(sim_i, 1, kiter) = Ahat(1, 1);
A_par(sim_i, 2, kiter) = Ahat(1, 2);
A_par(sim_i, 3, kiter) = Ahat(2, 1);
```

```
A_par(sim_i, 4, kiter) = Ahat(2, 2);
end

%%% Filling up A_table, and Sigma_nu_table
for i=1:4
    A_table(:, i, kiter) = [mean(A_par(:, i, kiter)); median(A_par(:, i, kiter)); max(A_par(:, i, kiter)); min(A_par(:, i, kiter)); std(A_par(:, i, kiter))];
    Sigma_nu_table1(:, i, kiter) = [mean(Sigma_nu_par1(:, i, kiter)); median(Sigma_nu_par1(:, i, kiter)); max(Sigma_nu_par1(:, i, kiter)); min(Sigma_nu_par1(:, i, kiter)); std(Sigma_nu_par1(:, i, kiter))];
    Sigma_nu_table2(:, i, kiter) = [mean(Sigma_nu_par2(:, i, kiter)); median(Sigma_nu_par2(:, i, kiter)); max(Sigma_nu_par2(:, i, kiter)); min(Sigma_nu_par2(:, i, kiter)); std(Sigma_nu_par2(:, i, kiter))];
end

end
```

B.2 GMM estimator

```
function [g] = GMM(A)
global numberofday y_f y_daynumber
global nDays1 nSlow y_s1 y_f1
%%%%% A(1, 1) = a_{ff}, A(1, 2) = a_{fs}, A(2, 1) = a_{sf}, A(2, 2) = a_{ss}

% A vector with a_{ss} = A(2,2) powers. a_ss_powers = [a_{ss}^{k-1}, a_{ss}^{k-2}, ..., a_{ss}^{1}, a_{ss}^{0}]
a_ss_powers = zeros(max(numberofday), 1);
for i = 1:length(a_ss_powers);
    a_ss_powers(i) = A(1, 2)^(length(a_ss_powers)-i);
end;

E22_1 = 0;
E22_2_1 = 0;
E22_2_2 = 0;
for i = 2:nSlow
    E22_1 = E22_1 + (y_s1(i-1, 1)*y_s1(i, 1))/length(2:nSlow);
    E22_2_1 = E22_2_1 + (y_s1(i-1, 1)*A(1, 1)*a_ss_powers(1:y_s1(i-1, 2))'*y_f(y_daynumber(i-1):(y_daynumber(i)-1), 1))/length(2:nSlow);
    E22_2_2 = E22_2_2 + (y_s1(i-1, 1)*(A(1,2)^nDays1(i-1))*y_s1(i-1, 1))/length(2:nSlow);
end
```

```
E21_1 = 0;
E21_2_1 = 0;
E21_2_2 = 0;
for i = 2:nSlow
      E21_1 = E21_1 + (y_f1(i-1)*y_s1(i, 1))/length(2:nSlow);
      E21_2_1 = E21_2_1 + (y_f1(i-1)*A(1, 1)*a_ss_powers(1:y_s1(i-1, 2))'*y_f(y_daynumber(i-1):(y_daynumber(i-1)-1), 1))/length(2:nSlow);
      E21_2_2 = E21_2_2 + (y_f1(i-1)*(A(1,2)^nDays1(i-1))*y_s1(i-1, 1))/length(2:nSlow);
end

%%%%% The "g" function
g(1) = E21_1 - (E21_2_1 + E21_2_2);
g(2) = E22_1 - (E22_2_1 + E22_2_2);

end
```

```
function [a] = GMMmin(A)
     global a_ff a_fs
     Aprime = [a_ff a_fs; A];
     eigval = abs(eig(Aprime));
     if max(eigval) > 1-10^(-4)
         a = 10^8;
     else
           a = GMM(A)*eye(2)*GMM(A)';
     end
end
```