# Mixed Frequency Vector Autoregressive Models

**Eric Ghysels***

First Draft: July 2011
This Draft: March 14, 2012

## ABSTRACT

Many time series are sampled at different frequencies. When we study co-movements between such series we usually analyze the joint process sampled at a common low frequency. This has consequences in terms of potentially mis-specifying the co-movements and hence the analysis of impulse response functions - a commonly used tool for economic policy analysis. We introduce a class of mixed frequency VAR models that allows us the measure the impact of high frequency data on low frequency and vice versa. Our approach does not rely on latent processes/shocks representations. As a consequence, the mixed frequency VAR is an alternative to commonly used state space models for mixed frequency data. State space models involve latent processes, and therefore rely on filtering to extract hidden states that are used in order to predict future outcomes. We also explicitly characterize the mis-specification of a traditional common low frequency VAR and its implied mis-specified impulse response functions. The class of mixed frequency VAR models can also characterize the timing of information releases for a mixture of sampling frequencies and the real-time updating of predictions caused by the flow of high frequency information. Hence, they are parameter-driven models whereas mixed frequency VAR models are observation-driven models as they are formulated exclusively in terms of observable data and do not involve latent processes and thus avoid the need to formulate measurement equations, filtering etc. We also propose various parsimonious parameterizations, in part inspired by recent work on MIDAS regressions. Various estimation procedures for mixed frequency VAR models are also proposed, both classical and Bayesian. Numerical and empirical examples quantify the consequences of ignoring mixed frequency data.

*Department of Economics, University of North Carolina, Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, USA, and Department of Finance, Kenan-Flagler Business School, e-mail: `eghysels@unc.edu`. Parts of the paper were written while the author was Resident Scholar at the Federal Reserve Bank of New York and subsequently Wim Duisenberg Senior Research Fellow at the European Central Bank. The financial support and stimulating intellectual environment of both institutions is gratefully acknowledged. We would like to thank Gianni Amisano, Mike Chernov, Domenico Giannone, Jim Hamilton, Massimiliano Marcellino, Isaac Miller, Michael Owyang and participants at the EABCN Training School on The Econometric Analysis of Mixed Frequency Data with Macro/Finance Applications, held at the European University Institute - October 2011 for helpful comments. We also thank Sunjin Park for excellent research assistance.

# 1 Introduction

It is simply a fact of life that time series observations are sampled at different frequencies. Some data series - such as financial ones - are easy to collect and readily available, while others are costly to record and therefore not frequently sampled. When we study co-movements between such series we usually analyze the joint process sampled at a common low sampling frequency. A typical example, following the seminal work of Sims (1980), is a vector autoregressive (VAR) model with both real and financial time series sampled quarterly - even though financial series are observed more frequently. We introduce a mixed frequency data VAR model and analyze the consequences of ignoring the availability of high frequency data. Take a simple example: GDP growth observed quarterly and non-farm payroll published monthly. We could look at the dynamics between the two series at a quarterly frequency - ignoring the fact that we do have monthly data for the second series. How does the shock to non-farm payroll and its impact on future GDP growth produced by standard VAR model analysis relate to the monthly surprises in the series? The quarterly VAR model shocks will be some mixture of the innovations in the underlying series. What type of mixture would this be? What are the costs in terms of impulse response analysis when we mis-align the data by ignoring the high frequency data? How does the flow of high frequency data allow us to update predictions of future low and high frequency data? We provide formal answers to all of these types of questions.

We introduce a relatively simple mixed sampling frequency VAR model. By simple we mean, (1) parsimonious, (2) one that can track the proper timing of low and high frequency data - that may include releases of quarterly data in the middle of the next quarter along with the releases of monthly data or daily data, (3) a specification that allows us to measure the impact of high frequency data onto low frequency ones and vice versa and perhaps most subtle (4) a specification that does not involve latent shocks.

We characterize the mapping between the mixed frequency VAR model and a traditional VAR model where all the data are sampled at a common low frequency. This mapping allows us to study the mis-specification of impulse response functions of traditional VAR models. The VAR models we propose can also handle time-varying mixed frequencies. Not all months have the same number of trading days, not all quarters have the same number of weeks, etc. Assuming a deterministic calendar effect, which makes all variation in changing mixed frequencies perfectly predictable, we are able to write a VAR with time-varying high frequency data structures.

The mixed frequency VAR provides an alternative to commonly used state space models involving mixed frequency data.[1] State space models involve latent processes, and therefore rely on filtering to extract hidden states that are used in order to predict future outcomes. State space models are, using the terminology of Cox (1981), parameter-driven models. The mixed frequency VAR models are, using again the same terminology, observation-driven models as they are formulated exclusively in terms of

[1] See for example, Harvey and Pierse (1984), Harvey (1989), Bernanke, Gertler, and Watson (1997), Zadrozny (1990), Mariano and Murasawa (2003), Mittnik and Zadrozny (2004), and more recently Aruoba, Diebold, and Scotti (2009), Ghysels and Wright (2009), Kuzin, Marcellino, and Schumacher (2009), Marcellino and Schumacher (2010), among others.

observable data. The fact we rely only on observable shocks has implications with respect to impulse response functions. Namely, we formulate impulse response functions in terms of observable data - high and low frequency - instead of shocks to some latent processes. Finally, mixed frequency VAR models, like MIDAS regressions, may be relatively frugal in terms of parameterization.

Technically speaking we adapt techniques typically used to study seasonal time series with hidden periodic structures, to multiple time series that have different sampling frequencies. The techniques we adapt relate to work by Gladyshev (1961), Pagano (1978), Tiao and Grupe (1980), Hansen and Sargent (1990, Chap. 17), Hansen and Sargent (1993), Ghysels (1994), Franses (1996), among others. In addition, the mixed frequency VAR model is a multivariate extension of MIDAS regressions proposed in recent work by Ghysels, Santa-Clara, and Valkanov (2006), Ghysels and Wright (2009), Andreou, Ghysels, and Kourtellos (2010) and Chen and Ghysels (2011), among others.

We study two classes of estimation procedures, classical and Bayesian, for mixed frequency VAR models. For the former we characterize how the mis-specification of traditional VAR models translates into pseudo-true VAR parameter and impulse response estimates. Parameter proliferation is an issue in both mixed frequency and traditional VAR models. We therefore also cover a Bayesian approach which easily accommodates the potentially large set of parameters to be estimated.

The paper is organized as follows. Section 2 introduces the structure of mixed frequency VAR models, discusses parsimony and impulse response functions. Section 3 elaborates on structural VAR models in the context of real-time updating of predictions and policy analysis. Section 4 covers the (mis-specified) traditional low frequency VAR process dynamics and impulse response functions implied by a mixed frequency VAR and also characterizes the loss of information due to ignoring high frequency data. Section 5 discusses classical and Bayesian estimation procedures. Section 6 provides numerical illustrative examples and finally Section 7 reports empirical findings with conclusions appearing in Section 8.

## 2 Mixed Frequency Vector Autoregressive Models

Since the work of Sims (1980), it is now standard to characterize the co-movements of macroeconomic time series as a VAR model. This typically involves some real activity series (i.e. GDP growth), some price series (i.e. inflation) and some monetary policy instrument (i.e. short term interest rates). This means we actually do have a mixture of respectively quarterly, monthly and daily series. Usually the sampling frequencies are aligned, for example inflation is computed quarterly and only end-of-the-quarter interest rates are sampled. Since the purpose of VAR models is to capture time series dynamics, it is natural to wonder how much harm is done both in terms of specification errors and prediction inaccuracy. Specification errors affect policy impulse response analysis and also have consequences as far as the asymptotic properties of estimators goes.

When we think of mixed frequencies, we need to distinguish situations where the high frequency data are sampled $m(\tau_L)$ times more often than the low frequency series where either $m(\tau_L) = m$, a constant or $m(\tau_L)$ has a deterministic time path. For example quarterly/annual, monthly/quarterly, daily/hourly amount to fixed $m$, whereas of daily/quarterly or weekly/quarterly involve $m(\tau_L)$ featuring pre-determined calendar effects. We start with the case of fixed $m$, namely:

**Assumption 2.1.** *We consider a K-dimensional process with the first $K_L < K$ elements, collected in the vector process $x_L(\tau_L)$, are only observed every $m$ fixed periods. The remaining $K_H = K$ - $K_L$ series, represented by double-indexed vector process $x_H(\tau_L, k_H)$ which is observed at the (high) frequency periods $k_H = 1, \ldots, m$ during period $\tau_L$.*

We will often refer to $x_L(\tau_L)$ as the low frequency (multivariate) process, and the $x_H(\tau_L, k_H)$ process as the high frequency (multivariate) one. Note that, for the sake of simplicity we consider the combination of two sampling frequencies. More than two sampling frequencies would amount to more complex notation, but would be conceptually similar to the analysis with a combination of two frequencies (see also section 4 for further discussion).

## 2.1 Shocks: Latent versus observable

So far attempts to accommodate mixed frequency data involve latent processes and therefore latent shocks. Zadrozny (1990) starts with a joint high frequency VAR(MA) model as if high frequency observations for $x_L(\tau_L)$ were available. A state space representation is then used to match the latent process with the mixture of data observed. This approach has recently been generalized by Chiu, Eraker, Foerster, Kim, and Seoane (2011) who develop a Bayesian approach to such mixed frequency VAR models where the missing data are drawn via a Gibbs sampler. Note that in such an approach the fundamental shocks are with respect to the hidden high frequency VAR. Factor models are also commonly used to handle mixed frequency data. For example, Mariano and Murasawa (2003) extract a coincident factor using quarterly and monthly time series (see also Nunes (2005)). Along similar lines, Aruoba, Diebold, and Scotti (2009) describe a dynamic one-factor model evolving on a daily basis to construct a coincident business index. Here too, the system is driven by latent shocks - not shocks to a high frequency VAR, but instead shocks that drive the latent factor that is measured with error through repeated high and low frequency data observations.

Our approach does not involve latent shocks. This means there is no need for filtering and the impulse response functions are based on observable shocks. To analyze mixed frequency vector processes we use insights from periodic models and construct stacked skip-sampled processes. We will start with an example where all the low frequency $\tau_L$ series appear at the end of the (low frequency) period. Namely, consider the following finite order VAR representation of a stacked vector:

$$\begin{bmatrix} x_H(\tau_L,1) \\ \vdots \\ x_H(\tau_L,m) \\ x_L(\tau_L) \end{bmatrix} = A_0 + \sum_{j=1}^{P} A_j \begin{bmatrix} x_H(\tau_L-j,1) \\ \vdots \\ x_H(\tau_L-j,m) \\ x_L(\tau_L-j) \end{bmatrix} + \varepsilon(\tau_L) \tag{2.1}$$

which is $K_L + m * K_H$ dimensional VAR model with $P$ lags.[2] Hence, with quarterly data we stack for example the months of January, February and March together with the first quarter low frequency data. Similarly we stack April, May and June with the second quarter, etc. For the moment, we focus on predicting next quarter's high and low frequency data *given* previous quarter's high and low frequency observations. Note however, that one may think of a specification similar to structural VAR models where we pre-multiply the vector $[x_H(\tau_L,1)',\ldots,x_H(\tau_L,m)',x_L(\tau_L)']'$ with a matrix $A_c$ :

$$A_c \begin{bmatrix} x_H(\tau_L,1) \\ \vdots \\ x_H(\tau_L,m) \\ x_L(\tau_L) \end{bmatrix} = A_0 + \sum_{j=1}^{P} A_j \begin{bmatrix} x_H(\tau_L-j,1) \\ \vdots \\ x_H(\tau_L-j,m) \\ x_L(\tau_L-j) \end{bmatrix} + \varepsilon(\tau_L) \tag{2.2}$$

where the matrix $A_c$ pertains to contemporaneous (in this case within quarter) relationships. Writing the matrix $A_c$ explicitly, we have the left hand side of (2.2) as:

$$\begin{bmatrix} I_{K_H} & \ldots & A_c^{1,m} & A_c^{1,m+1} \\ \vdots & \ldots & \vdots & \vdots \\ A_c^{m,1} & \ldots & I_{K_H} & A_c^{m,m+1} \\ A_c^{m+1,1} & \ldots & A_c^{m+1,m} & I_{K_L} \end{bmatrix} \begin{bmatrix} x_H(\tau_L,1) \\ \vdots \\ x_H(\tau_L,m) \\ x_L(\tau_L) \end{bmatrix} \tag{2.3}$$

Hence, elements below the diagonal pertain to downstream impacts, i.e. high frequency data affect subsequent within-$\tau_L$ period observations. This will be relevant notably for intra-$\tau_L$ period prediction updating - a topic discussed in section 3.1. In contrast, elements above the diagonal will be relevant notably when we will discuss policy rules in section 3.2. Obviously, with $A_c$ invertible we can always view equation (2.1) as one obtained after pre-multiplying both sides of (2.2) by $A_c^{-1}$. Hence, for the moment we will ignore the presence of $A_c$.

One might think that the stacked system appearing in (2.1) could be prone to parameter proliferation. That may not actually be the case as we will show later in the section. While we do not address parameter proliferation issues for the moment, it is worth pointing out the relationship with MIDAS regressions, in particular, by looking at a special case with $K_L$ = $K_H$ =1. The last equation in the

[2] The assumption of a finite order VAR may appear somewhat restrictive. It is worth noting that much of our analysis could be extended to VARMA models. Since VAR models are more widely used and considering VARMA models significantly complicates the estimation we forego the generalization of adding MA terms.

system then reads:

$$x_L(\tau_L) = A_0^{m+1,1} + \sum_{j=1}^{P} A_j^{m+1,m+1} x_L(\tau_L - j) + \sum_{j=1}^{P}\sum_{k=1}^{m} A_j^{m+1,k} x_H(\tau_L - j, k) + \varepsilon(\tau_L)^{m+1,1} \quad (2.4)$$

which is the ADL MIDAS regression model discussed in Andreou, Ghysels, and Kourtellos (2010). There are various parsimonious parameterizations suggested for such regressions, see e.g. Ghysels, Sinko, and Valkanov (2006), Andreou, Ghysels, and Kourtellos (2010) and Sinko, Sockin, and Ghysels (2010), that will be discussed later.

Note that the aforementioned VAR model contains, besides MIDAS regressions, also the impact of what one might call the low frequency shock $\varepsilon(\tau_L)^{m+1,1}$ (the last element of the innovation vector in this particular example) onto both future high *and* low frequency series as well as high frequency shocks $\varepsilon(\tau_L)^{i,1}$ ($i$ = 1, ..., $m$ again in this particular example) onto future high and low frequency series.

## 2.2 The constituents of the stacked vector

We adopt a general approach, and therefore analyze a generic stacked vector systems. Yet, we also need to keep in mind that the observations we stack into vectors may differ from application to application and in particular may depend on the focus of the application.

For example, let us consider two different scenarios involving a mixture monthly and quarterly data. The first scenario, one could refer to as economic time, seeks to study the fundamental dynamics of the economy. Namely, there is a number of people employed during the month of January, another number for February, a third for March, and then there is a GDP number for the first quarter. This yields four numbers, three monthly employment figures and one GDP, which would logically be collected in a single stacked vector. An alternative scenario is news-release time. For example, on January 6 the Bureau of Labor Statistics (BLS) releases the December employment report, on January 27 the Bureau of Economic Analysis (BEA) will release the GDP number for the fourth quarter of the previous year, on February 3 the BLS will release the employment report for January and a revised value for the employment number for December, on February 29 the BEA will release a revised estimate of previous fourth quarter of GDP, on March 2 the BLS will release the employment report for February (and revisions of the December and January counts), and on March 29 the BEA will release yet another estimate of the previous year GDP. Perhaps we want to collect all eight of these numbers in the vector for the first quarter. Note also that in the first scenario we would take final data, not the real-time series.

Clearly, both scenarios are of interest and can be covered by our generic mixed frequency VAR model. While throughout the paper we will try to provide a general discussion, it will be clear that some parts of our analysis will be more relevant for specific applications. For example, the mapping from mixed frequency to traditional low frequency VAR models and the analysis of potentially mis-specified impulse response functions appearing in section 4 is clearly more relevant for aforementioned economic time structural dynamic analysis. Likewise, the distinction between mixed and periodic stacked VAR representations appearing in subsection 2.3 will also mostly pertain to the first scenario type of research.

On the other hand, if one is interested in a real-time forecasting exercise, then we clearly consider the second approach. For example, assume all low frequency data are release at the same time and compare:

$$\underline{\mathrm{x}}(\tau_L) = \begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ \vdots \\ x_H(\tau_L, m) \\ x_L(\tau_L) \end{bmatrix} \quad versus \quad \begin{bmatrix} x_L(\tau_L) \\ x_H(\tau_L, 1) \\ \vdots \\ \vdots \\ x_H(\tau_L, m) \end{bmatrix} \quad or \quad \begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ x_L(\tau_L) \\ \vdots \\ x_H(\tau_L, m) \end{bmatrix} \tag{2.5}$$

where the release of low frequency appears at the end, beginning or some time in the middle of $\tau_L$. The order of appearance in the vector therefore determines the timing of intra-$\tau_L$ period releases and that will be important later to understand the impact and timing of shocks as well as the updating of predictions as new intra-$\tau_L$ period shocks occur. The high frequency releases of low frequency data can be scattered at various $k_H$ throughout period $\tau_L$ and therefore impact the structure of shocks and responses. More specifically the $K_L$ low frequency series $K_L^{k_H}$ in $x_L(\tau_L)$ are released at time $k_H$ in period $\tau_L$ for $k_H = 1, \ldots, m$, with $\sum_{i=1}^{m} K_L^i = K_L$.[3] When we need to keep track of the high frequency releases on low frequency data we use $x_L(\tau_L, k_H)$, for the sub-vector released at $k_H$. All $x_L(\tau_L, k_H)$ combined for $k_H = 1, \ldots, m$, yield the time-stamped low frequency process. Hence, when all the low frequency data are released at the end of period $\tau_L$ then $\underline{\mathrm{x}}(\tau_L) \equiv (x_H(\tau_L, 1)', \ldots, x_H(\tau_L, m)', x_L(\tau_L)')'$, otherwise it contains $(x_L(\tau_L, k_H)', x_H(\tau_L, k_H)')'$ for the sequence $k_H = 1, \ldots, m$.[4]

For many parts of the paper the details about the specific constituents of the stacked vector will be irrelevant, and we will put all the high frequency data first followed by the low frequency data. However, when the focus is real-time analysis, as in subsections 2.5 and 3.1, we will deal more explicitly with the specific order of the elements in the stacked vector.

[3] Most releases are on a fixed schedule, with notable exceptions such as some FOMC announcements. In addition to the extensive academic literature, mostly studying the phenomenon of financial market impact of announcements - one can find many details regarding announcement schedules on financial news sites such as `http://www.nasdaq.com/markets/us-economic-calendar.aspx` or `http://biz.yahoo.com/c/e.html`, among many others. The framework presented in this paper can, with some modification handle announcements that may occur at random - the technicalities of it are clarified in subsection 3.3.

[4] If $x_H(\tau_L, k_H)$ is empty for some $k_H$, we only stack the high frequency data.

## 2.3 Mixed and Periodic Stacked VAR Representations

It will be convenient to use a more compact notation for the $K_L + m * K_H$ dimensional vector $\underline{\mathrm{x}}(\tau_L)$, namely we will write equation (2.1) as:

$$\mathbb{A}(\mathcal{L}_L)(\underline{\mathrm{x}}(\tau_L) - \mu_{\underline{\mathrm{x}}}) = \varepsilon(\tau_L) \tag{2.6}$$

where $\mathcal{L}_L$ is the low frequency lag operator, i.e. $\mathcal{L}_L\underline{\mathrm{x}}(\tau_L) = \underline{\mathrm{x}}(\tau_L - 1)$, and:

$$\begin{aligned} \mathbb{A}(\mathcal{L}_L) &= I - \sum_{j=1}^{P} A_j \mathcal{L}_L^j \\ \mu_{\underline{\mathrm{x}}} &= (I - \sum_{j=1}^{P} A_j)^{-1} A_0 \end{aligned} \tag{2.7}$$

where we assume that the VAR is covariance stationary to be able to write the above equations (see Assumption 2.2 below) and we let $E[\varepsilon(\tau_L)\varepsilon(\tau_L)'] = \mathbb{C}\mathbb{C}'$.

We are also interested in a second representation which will be useful for studying the relationship between mixed frequency and traditional VAR models which ignores the availability of $x_H(\tau_L, k_H)$. To this end, we will introduce a joint process $\overline{x}(\tau_L) \equiv (x_H(\tau_L)', x_L(\tau_L)')'$, where the first sub-vector of low frequency observations is left unspecified for the moment - i.e. we are not going to be explicit until the next section about how the high frequency data aggregate to low frequency observations. We are interested in the VAR model:

$$\mathbb{B}(\mathcal{L}_L)(\overline{x}(\tau_L) - \mu_{\overline{x}}) = \overline{\varepsilon}(\tau_L) \tag{2.8}$$

where $\mathbb{B}(\mathcal{L}_L) = I - \sum_{j=1}^{\mathcal{P}} B_j \mathcal{L}_L^j$, and $E[\overline{\varepsilon}(\tau_L)\overline{\varepsilon}(\tau_L)'] = \overline{\mathbb{C}\mathbb{C}}'$. Note that the lag length of the VAR may not be finite, i.e. $\mathcal{P}$ may be infinite.[5] Obviously, what also interests us is the relationships between the (traditional) VAR characterized by $\mathbb{B}(\mathcal{L}_L)$ and $\overline{\mathbb{C}}$ and the original mixed frequency dynamics $\mathcal{L}_L$ and $\mathbb{C}$. It is one of our goals to characterize this relationship.

While the VAR model appearing in equation (2.6) looks standard, its simple appearance is deceiving. Inherently, its structure shares features with so called periodic time series models originated by Gladyshev (1961). Yet, it does not quite fully resemble periodic models, and we will need an augmented version of (2.6) to achieve this. We will consider a $(K * m) \times (K_L + m * K_H)$ matrix $H$ such that:

$$\ddot{x}(\tau_L) = H\underline{\mathrm{x}}(\tau_L) \tag{2.9}$$

One prominent example is $\ddot{x}(\tau_L) \equiv (x_H(\tau_L, 1)', x_L(\tau_L)', \ldots, x_H(\tau_L, m)', x_L(\tau_L))'$. Note that the low frequency series is repeated, capturing every high frequency period the relationship between how and

[5] Incorrectly assuming one has a finite VAR will have consequences on the asymptotic properties of the parameter estimators - a topic that will be discussed later.

high frequency data. Note that this is simply a reshuffling of the original vector - we are not constructing a mapping involving a latent vector. Recall that we discussed two different scenarios in subsection 2.2. The first scenario seeks to study the fundamental dynamics of the economy, whereas the second is oriented towards real-time analysis. It is mostly for the first type of analysis that we will need the $\ddot{x}(\tau_L)$ representation. As discussed later, this will allow us to examine how mixed frequency impulse responses will get scrambled.[6]

Formally stated, we assume the following data generating process for the mixture of high and low frequency data:

**Assumption 2.2.** *The vector* $\ddot{x}(\tau_L) \equiv (x_H(\tau_L, 1)', x_L(\tau_L, 1)', \ldots, x_H(\tau_L, m)', x_L(\tau_L, m))'$ *is of dimension* $m * K$ *and has a finite order covariance stationary VAR representation:*

$$\ddot{\mathbb{A}}(\mathcal{L}_L)(\ddot{x}(\tau_L) - \mu_{\ddot{x}}) = \ddot{\varepsilon}(\tau_L) \tag{2.10}$$

*where* $\ddot{\mathbb{A}}(\mathcal{L}_L) = I - \sum_{j=1}^{P} \ddot{A}_j \mathcal{L}_L^j$, $\mu_{\ddot{x}} = \ddot{A}_0(I - \sum_{j=1}^{P} \ddot{A}_j)^{-1}$ *and* $E[\ddot{\varepsilon}(\tau_L)\ddot{\varepsilon}(\tau_L)'] = \ddot{\mathbb{C}}\ddot{\mathbb{C}}'$.

In the remainder of this section we will work mostly with the vector $\underline{\mathrm{x}}(\tau_L)$ appearing in (2.6) rather than $\ddot{x}(\tau_L)$. The latter will be useful when we derive the mapping between traditional and mixed frequency VAR models, a topic which we will cover in section 4.

## 2.4 Parsimony

The question of parsimony in VAR models has been much discussed as it is an issue that is particularly acute for large dimensional models and/or models involving many lags. One might think that the acuteness of parameter proliferation is likely to be even more an issue with a mixture of sampling frequencies. It is the purpose of this section to show that this may not be as severe as one might think. There are mainly two reasons why there may not be a parameter proliferation problem despite the potentially large dimensional VAR systems. First, the stacking of high frequency data typically involves repeating the same parametric structure across all $m$ replicas (unlike the periodic models which inspired the structure of mixed frequency VAR models). Second, the key insights of MIDAS regressions also play a key role in keeping the parameter space low dimensional. We develop a few examples showing how one could potentially write sparsely parameterized mixed frequency VAR models. These are not per se *the* specifications, but they provide a few leads on how one may go about writing conveniently a parametric structure. The common theme, however, is that we aim for specifications with the appealing feature that the number of parameters does not depend on $m$, i.e. the number of high frequency observations per low frequency time period.

[6] If we were to consider the second scenario involving real-time applications, the low frequency components of the vector $\ddot{x}(\tau_L)$ will gradually update the data release throughout $\tau_L$, together with the high frequency observations $x_H(\tau_L, k_H)$, in chronological order. Hence, the difference between the vector $\underline{\mathrm{x}}(\tau_L)$ and $\ddot{x}(\tau_L)$ is that the former simply stacks all the timed releases of low and high frequency data, whereas the latter repeats all the elements of the low frequency process with stale values until intra$-\tau_L$ updates happen. For the sake of brevity we skip the details of such analysis.

For the purpose of streamlining the exposition we will start again with an example where all the low frequency $\tau_L$ series appear at the end of the stacked vector as in equation (2.1). In addition, we set $K_L = K_H = 1$ and assume that all the series are either demeaned or are assumed mean zero.[7] Therefore, we rewrite equation (2.1) as:

$$\begin{bmatrix} x_H(\tau_L,1) \\ \vdots \\ x_H(\tau_L,m) \\ x_L(\tau_L) \end{bmatrix} = \sum_{j=1}^{P} \begin{bmatrix} A_j^{1,1} & \dots & A_j^{1,m} & A_j^{1,m+1} \\ \vdots & \dots & \vdots & \vdots \\ A_j^{m,1} & \dots & A_j^{m,m} & A_j^{m,m+1} \\ A_j^{m+1,1} & \dots & A_j^{m+1,m} & A_j^{m+1,m+1} \end{bmatrix} \begin{bmatrix} x_H(\tau_L-j,1) \\ \vdots \\ x_H(\tau_L-j,m) \\ x_L(\tau_L-j) \end{bmatrix} + \varepsilon(\tau_L)$$

which is $m+1$-dimensional VAR model with $P$ lags. When we assume that the high frequency process is ARX(1) with the impact of the low frequency series constant throughout the period, we have:

$$\begin{aligned} \begin{bmatrix} x_H(\tau_L,1) \\ \vdots \\ x_H(\tau_L,m) \\ x_L(\tau_L) \end{bmatrix} &= \begin{bmatrix} 0 & \dots & \rho & a \\ \vdots & \dots & \vdots & \vdots \\ 0 & \dots & \rho^m & a(1+\sum_{j=0}^{m-1}\rho^j) \\ w(\gamma)_m & \dots & w(\gamma)_1 & \alpha_1 \end{bmatrix} \begin{bmatrix} x_H(\tau_L-1,1) \\ \vdots \\ x_H(\tau_L-1,m) \\ x_L(\tau_L-1) \end{bmatrix} \\ &+ \sum_{j=2}^{P} \begin{bmatrix} 0 & \dots & 0 & 0 \\ \vdots & \dots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \\ w(\gamma)_{jm} & \dots & w(\gamma)_{(j-1)m+1} & \alpha_j \end{bmatrix} \begin{bmatrix} x_H(\tau_L-j,1) \\ \vdots \\ x_H(\tau_L-j,m) \\ x_L(\tau_L-j) \end{bmatrix} + \varepsilon(\tau_L) \end{aligned} \tag{2.11}$$

which involves $P$ parameters $\alpha_j$, two parameters $\rho$ and $a$ and a low dimensional MIDAS polynomial parameter vector $\gamma$. When all $\alpha_j = 0$ for $j > 1$, and the dimension of $\gamma$ is 2, which is not unreasonable (see Appendix A for details), we end up with 5 parameters regardless of the value of $m$. Admittedly, this is a tightly constrained model, yet it is not an unreasonable starting point. Continuing with the system in (2.11), the innovation covariance matrix may also be sparsely specified. Continuing with the above specification, we can write:

$$E[\varepsilon(\tau_L)\varepsilon(\tau_L)'] = \begin{bmatrix} \sigma_{HH} & \rho\sigma_{HH} & \dots & \rho^{m-1}\sigma_{HH} & \sigma_{HL} \\ \rho\sigma_{HH} & (1+\rho^2)\sigma_{HH} & 0 & \vdots & \sigma_{HL} \\ \vdots & \vdots & \ddots & \rho\sigma_{HH} & \vdots \\ \rho^{m-1}\sigma_{HH} & \rho^{m-2}\sigma_{HH} & \dots & (1+\sum_{i=1}^{m-1}\rho^2)\sigma_{HH} & \vdots \\ \sigma_{HL} & \sigma_{HL} & \dots & \dots & \sigma_{LL} \end{bmatrix} \tag{2.12}$$

adding another three parameters, and therefore a total of eight again independent of $m$. Obviously, for some of the high frequency applications, one may consider adding ARCH-type dynamics to the innovations, or add announcement effects to some of the elements of the covariances - which would entail a richer, yet still moderate and independent of $m$, parameter structure.

[7] In subsection 5.2 we cover the cases with $K_L$ low frequency and $K_H$ high frequency series.

The specification of the MIDAS regressions, when $K_L > 1$, deserves some attention as well. Namely, consider the following:

$$[A_1^{m+1,1} \dots A_1^{m+1,m} A_2^{m+1,1} \dots A_P^{m+1,m}] = B \otimes \left[ \sum_{i=1}^{K_H \times P} (w(\gamma)_i) \right] \tag{2.13}$$

with $B$ a $K_L \times K_H$ matrix and $\sum_{i=1}^{K_H \times P} (w(\gamma)_i)$ is a *scalar* MIDAS polynomial. Hence, we impose a common decay pattern with a single polynomial lag structure with $B$ containing the collection of slope parameters identified as the sum of the polynomial lag weights add up to one. As noted before, there are various parsimonious parameterizations suggested for the weights $\omega(\gamma)_i$, that are briefly reviewed Appendix A. The above specification has the virtue of reducing $(K_L \times K_H) \times P \times m$ parameters to just $K_L \times K_H$ + the dimension of $\gamma$ which is 2 in many of the examples discussed in the aforementioned Appendix. Needless to say that this characterization of the polynomials may be too restrictive - yet as in the previous case, it may be a reasonable starting point in many practical settings. Along the same lines, one can consider a less parsimonious specification inspired by the so called multiplicative MIDAS (see equation (A.8) in Appendix A):

$$[A_i^{m+1,1} \dots A_i^{m+1,m}] = B_i \otimes \left[ \sum_{i=1}^{m} (w(\gamma)_i) \right] \qquad i = 1, \dots, P \tag{2.14}$$

meaning that within-$\tau_L$ period high frequency weights remain invariant and yield a low frequency parameterized process $x_H(\tau_L - j)(\gamma)$ with lag coefficients $B_i$. The advantage of this specification is that the impact of high frequency data on low frequency ones nests specifications with ad hoc linear time aggregation such as time averaging - taking the last within-$\tau_L$ period high frequency observation. Bai, Ghysels, and Wright (2009) show that the above specification matches a steady state Kalman filter prediction equation obtained from a single factor state space model and provides a good approximation for many more complex state space model specifications. Note that, at least for the block of low frequency series, the above specification is quite similar to a traditional VAR with lag coefficients $B_i$, augmented by a small number of parameters used in the filtering scheme. Here again, the number of parameters does not augment with $m$.

We adopt in the remainder of the paper a generic setting where all the parameters are collected into a vector $\Psi$. The above sparsely parameterized mixed frequency VAR model is a frugal example, while more richly specified structures obviously will involve higher dimensional parameter vectors. In general, we will write the finitely parameterized mixed frequency VAR models appearing in equations (2.6) and (2.10) respectively as:

$$\begin{aligned} \mathbb{A}_\Psi(\mathcal{L}_L)(\underline{\mathrm{x}}(\tau_L) - \mu_{\underline{\mathrm{x}}}^{\Psi}) &= \varepsilon(\tau_L) \\ \ddot{\mathbb{A}}_\Psi(\mathcal{L}_L)(\ddot{x}(\tau_L) - \mu_{\ddot{x}}^{\Psi}) &= \ddot{\varepsilon}(\tau_L) \end{aligned} \tag{2.15}$$

with $E[\varepsilon(\tau_L)\varepsilon(\tau_L)']$ = $\mathbb{C}(\Psi)\mathbb{C}(\Psi)'$, and $E[\ddot{\varepsilon}(\tau_L)\ddot{\varepsilon}(\tau_L)']$ = $\ddot{\mathbb{C}}(\Psi)\ddot{\mathbb{C}}(\Psi)'$. Note that the parameter vector

$\Psi$ governs both the stacked and periodic representations and that we do not count the unconditional mean as part of the parameter space since we used demeaned series in our analysis.

To streamline the notation, we will drop the parameter vector $\Psi$ for the remainder of this section, although one has to keep in mind that the material we will present is subject to potential specification errors resulting from parsimonious parameterizations - a subject we will address in the next section.

## 2.5 Shocks and Choleski factorization

Much has been written about impulse response functions in VAR models, in particular with regards to the interpretation of shocks. The class of mixed frequency VAR models sheds new light on this topic. First of all, let us recall that the vectors $\underline{\mathrm{x}}(\tau_L)$ and $\ddot{x}(\tau_L)$ have a natural order for the intra$-\tau_L$ period timing of shocks since their elements represent a sequence of time events. If more than one series is released at a specific time, then the order of associated shocks is subject to the same considerations as in traditional VAR models - or perhaps not. For example if during a day, or a week, or month both financial and macro series are released, we do not necessarily know how to order them - except that macro data are released before financial markets open, so there is again a natural order despite the contemporaneous time stamp in the vector $\underline{\mathrm{x}}(\tau_L)$.

It is important to note that for the purpose of information accounting we will work with $\underline{\mathrm{x}}(\tau_L)$ appearing in (2.6) rather than $\ddot{x}(\tau_L)$, as the latter contains repetitive strings of stale low frequency data. The stacked mixed frequency VAR model implies an impulse response function:

$$
\begin{aligned}
(\underline{\mathrm{x}}(\tau_L) - \mu_{\underline{\mathrm{x}}}) &= (I - \sum_{j=1}^{P} A_j \mathcal{L}_L^j)^{-1} \varepsilon(\tau_L) \\
&= \sum_{j=0}^{\infty} F_j \varepsilon(\tau_L - j) \equiv F(\mathcal{L}_L)\varepsilon(\tau_L) \qquad (2.16)
\end{aligned}
$$

where $I = (\mathbb{A}(\mathcal{L}_L))F(\mathcal{L}_L)$, which allows us to study the intra$-\tau_L$ period timing of shocks, both high frequency as well as low frequency.

This means that shocks $\varepsilon(\tau_L)$ tell us something about the timed surprise of either type of series, and therefore the impulse responses tell us what is the impact of say a macroeconomic announcement of a low frequency series onto future low and high frequency ones, and surprises in high frequency series on both future low and high frequency series. Compared to the impulse responses from the VAR in equation (2.8), namely: $(\overline{x}(\tau_L) - \mu_{\overline{x}}) = (\mathbb{B}(\mathcal{L}_L))^{-1}\overline{\varepsilon}(\tau_L)$ we can see how intra$-\tau_L$ period shocks are scrambled - something we will be more explicit about in the next section.

Since the order of the series is no long arbitrary, it is also the case that the Choleski factorization of the innovations is no longer arbitrary. In particular consider:

$$
E[\varepsilon(\tau_L)\varepsilon(\tau_L)'] = \mathbb{C}\mathbb{C}' = \mathcal{M}_{[m]}\Omega\mathcal{M}_{[m]}' \qquad (2.17)
$$

where $\Omega$ is a diagonal matrix and $\mathcal{M}_{[m]}$ is a lower triangular matrix. We add the index $m$ to the latter as it will be relevant for the material presented in the next subsection. Since the inverse of a lower triangular matrix is again a lower triangular one, consider $(\mathcal{M}_{[m]})^{-1} = \mathcal{N}_{[m]}$, and:

$$
\begin{aligned}
\mathbb{A}(\mathcal{L}_L)(\underline{\mathrm{x}}(\tau_L) - \mu_{\underline{\mathrm{x}}}) &= \varepsilon(\tau_L) \\
&= \mathcal{M}_{[m]}\eta(\tau_L) \\
\mathcal{N}_{[m]}\mathbb{A}(\mathcal{L}_L)(\underline{\mathrm{x}}(\tau_L) - \mu_{\underline{\mathrm{x}}}) &= \eta(\tau_L)
\end{aligned} \tag{2.18}
$$

where $E[\eta(\tau_L)\eta(\tau_L)'] = \Omega$, a diagonal matrix.

When we turn our attention again to the parsimonious examples in the previous subsection, and in particular equation (2.12) we realize that the parameters governing the covariance matrix $E[\varepsilon(\tau_L)\varepsilon(\tau_L)']$ and thus its Choleski factorization, are tied to the parameters governing the VAR dynamics, in particular the parameter $\rho$ in equation (2.11). This leaves us with the choice of either (1) estimate the factorization unconstrained, or (2) explore the common parameter restrictions and aim for a more efficient estimation of the impulse response functions. This issue is reminiscent of structural VAR models as alluded to in equation (2.2). We will revisit the connection with traditional structural VAR models in the next subsections. To summarize: while Choleski factorizations are typically ambiguous in terms impulse response analysis in traditional VAR models, they are a more natural tool for impulse response analysis for time-stamped mixed frequency VAR systems. In addition, there are potential gains to be made from considering common parameter restrictions between the mixed frequency VAR dynamics and the lower triangular factorization.

# 3 Structural Mixed Frequency VAR Models

We turn our attention now to structural VAR models and consider various specifications for the $A_c$ matrix appearing in equation (2.2). We will focus on two particular applications, namely real-time prediction updating and policy analysis. A subsection is devoted to each topic. A final subsection will deal with a generalization of mixed frequency VAR models relevant for both real-time and policy analysis.

## 3.1 Real-time predictions

The potential mis-specification of shocks due to aggregation of mixed frequency data also leads us to the question how much is lost by ignoring the real-time stream of high frequency data as one foregoes the possibility to engage in within-$\tau_L$ updates of forecasts. It turns out this will be an example of using certain types of structural VAR matrices to update within-$\tau_L$ information.

Continuing with the example in equations (2.6) and (2.18) consider the following transformations for $i = 1, \ldots, m-1 : \mathcal{N}_{[i]}\mathbb{A}(\mathcal{L}_L)(\underline{\mathrm{x}}(\tau_L) - \mu_{\underline{\mathrm{x}}}) = \mathcal{N}_{[i]}\varepsilon(\tau_L)$ or:

$$\mathcal{N}_{[i]}\underline{\mathrm{x}}(\tau_L) = \mathcal{N}_{[i]}A_0 + \sum_{j=1}^{P}\mathcal{N}_{[i]}A_j\underline{\mathrm{x}}(\tau_L - j) + \mathcal{N}_{[i]}\varepsilon(\tau_L) \tag{3.19}$$

involving the matrices, $\mathcal{N}_{[i]}$, $i = 1, \ldots, m-1$, which can be written as:

$$\mathcal{N}_{[i]} = \begin{bmatrix} I & 0 & \cdots & \cdots & 0 & 0 \\ \mathcal{N}_{[i]}^{2,1} & I & 0 & \cdots & 0 & 0 \\ \vdots & & & & \vdots & \vdots \\ \mathcal{N}_{[i]}^{i+1,1} & \cdots & \mathcal{N}_{[i]}^{i+1,i} & I & \vdots & 0 \\ \vdots & & \vdots & 0 & \ddots & \vdots \\ \mathcal{N}_{[i]}^{m+1,1} & \cdots & \mathcal{N}_{[i]}^{m+1,i} & 0 & \cdots & I \end{bmatrix} \tag{3.20}$$

where the matrices $\mathcal{N}_{[i]}^{a,b}$ are of dimension $K_H \times K_H$ except for $a = m+1$. Matrices $\mathcal{N}_{[i]}^{m+1,b}$ are of dimension $K_H \times K_H$. These matrices are related to the inverse of the Choleski lower triangular decomposition, namely recall from equation (2.18) that $(\mathcal{M}_{[m]})^{-1} = \mathcal{N}_{[m]}$, and define the matrices, $\mathcal{N}_{[i]}$ as the partial triangular decompositions orthogonalizing only the first $i$ shocks.

To clarify the role played by the transformation appearing in (3.20), let us for instance take a look at $\mathcal{N}_{[1]}$, which applies to a first high frequency data point becoming available, and the special case considered before of $K_L = K_H = 1$, i.e.:

$$\mathcal{N}_{[1]} = \begin{bmatrix} I & 0 & \cdots & 0 \\ \mathcal{N}_{[1]}^{2,1} & I & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ \mathcal{N}_{[1]}^{m+1,1} & 0 & \cdots & I \end{bmatrix}$$

Then the last equation in the system reads:

$$\begin{aligned} x_L(\tau_L) &= A_0^{m+1,1} - \mathcal{N}_{[1]}^{m+1,1}\underbrace{x_H((\tau)_L, 1)}_{\text{New Info}} + \sum_{j=1}^{P} A_j^{m+1,m+1}x_L(\tau_L - j) \\ &\quad + \sum_{j=1}^{P}\sum_{k=1}^{m} A_j^{m+1,k}x_H(\tau_L - j, k) + \varepsilon(\tau_L)^{m+1,1} \end{aligned} \tag{3.21}$$

which is the ADL MIDAS regression model with (one) lead(s) discussed in Andreou, Ghysels, and Kourtellos (2010). Alternatively, we can also write the last equation, based on the inversion of the $\mathcal{N}_{[1]}$

matrix as:

$$
\begin{aligned}
x_L(\tau_L) &= (A_0^{m+1,1} - \mathcal{N}_{[1]}^{m+1,1} A_0^{1,1}) - \mathcal{N}_{[1]}^{m+1,1} \varepsilon(\tau_L)^{1,1} \\
&+ \sum_{j=1}^{P} (A_j^{m+1,m+1} - \mathcal{N}_{[1]}^{m+1,1} A_j^{1,1}) x_L(\tau_L - j) \\
&+ \sum_{j=1}^{P} \sum_{k=1}^{m} (A_j^{m+1,k} - \mathcal{N}_{[1]}^{m+1,1} A_j^{1,k}) x_H(\tau_L - j, k) + \varepsilon(\tau_L)^{m+1,1}
\end{aligned} \tag{3.22}
$$

The latter representation is closer to a Kalman filter approach as it adds the information innovation $\varepsilon(\tau_L)^{1,1}$, which equals $x_H((\tau)_L, 1)$ - $E_{\tau_L}[x_H((\tau)_L, 1)]$, to the equation and re-weights all the old information accordingly.

Note, the simplicity of the updating scheme: (1) we estimate a mixed frequency VAR model, (2) compute the Choleski factorization of the errors and then take the $m-1$ lower triangular truncations of the original factorization. It is also worth recalling that we may or may not impose common parameter restrictions between the parameters of the mixed frequency VAR and the covariance matrix of the full system as noted at the end of the previous section.

## 3.2 Policy response functions

The analysis in the previous subsection is one example of mixed frequency VAR models with a particular choice of $A_c$ matrix appearing in equation (2.2). In the present subsection we study structural VAR models with mixed frequency data for the purpose of studying policy analysis. To do so, we consider a high frequency vector that contains some monetary policy instrument, such as the Federal funds rate (henceforth FFR). In fact, to simplify the presentation, let us only focus on FFR in combination with some low frequency series. In particular:

$$
A_c \begin{bmatrix} FFR(\tau_L, 1) \\ \vdots \\ FFR(\tau_L, k) \\ \vdots \\ FFR(\tau_L, m) \\ x_L(\tau_L) \end{bmatrix} = A_0 + \sum_{j=1}^{P} A_j \begin{bmatrix} FFR(\tau_L - j, 1) \\ \vdots \\ FFR(\tau_L - j, k) \\ \vdots \\ FFR(\tau_L - j, m) \\ x_L(\tau_L - j) \end{bmatrix} + \varepsilon(\tau_L) \tag{3.23}
$$

with

$$A_c = \begin{bmatrix} I_{K_H} & \dots & \dots & \dots & \dots & A_c^{1,m} & A_c^{1,m+1} \\ \vdots & \ddots & & & & \vdots & \vdots \\ \vdots & & \ddots & & & \vdots & \vdots \\ A_c^{k,1} & \dots & A_c^{k,k-1} & I_{K_H} & & A_c^{k,m} & A_c^{k,m+1} \\ \vdots & \vdots & & & \ddots & \vdots & \vdots \\ A_c^{m,1} & \dots & & & \dots & I_{K_H} & A_c^{m,m+1} \\ A_c^{m+1,1} & \dots & & & \dots & A_c^{m+1,m} & I_{K_L} \end{bmatrix} \tag{3.24}$$

Let us focus on the equation for $FFR(\tau_L, k)$. For simplicity, we set $A_c^{k,j}$ = 0 for $j < k-1$ and $k < j \leq m$. Moreover, we leave unspecified the regressors appearing on the right hand side of the above equation, and therefore we have:

$$\begin{aligned} FFR(\tau_L, k) &= A_0^{k,1} - A_c^{k,k-1} FFR(\tau_L, k-1) - A_c^{k,m+1} x_L(\tau_L) \\ &\quad + \text{regressors prior to } \tau_L + \varepsilon(\tau_L)^{k,1} \end{aligned} \tag{3.25}$$

Note that the above equation for $FFR(\tau_L, k)$ features the low frequency $x_L(\tau_L)$ (as well as lagged low and high frequency data). This means that policy may respond to current conditions - although $x_L(\tau_L)$ may not yet be observed at period $k$ of $\tau_L$. This raised some interesting issues. To address these, let us define the information set $I(\tau_L, k)$ as all the information available at period $k$ of $\tau_L$. Therefore, one may interpret equation (3.25) as:

$$FFR(\tau_L, k) = A_0^{k,1} - A_c^{k,k-1} FFR(\tau_L, k-1) - A_c^{k,m+1} E[x_L(\tau_L)|I(\tau_L, k)] + \dots \tag{3.26}$$

involving real-time estimates of $x_L(\tau_L)$. Therefore, we may think of cross-equation restrictions since $E[x_L(\tau_L)|I(\tau_L, k)]$ involves the rows of $\mathcal{N}_{[k]}\underline{\mathrm{x}}(\tau_L)$ pertaining to the concurrent estimates of $x_L(\tau_L)$. Recall that in equation (2.12) we noted that the parameters governing the covariance matrix $E[\varepsilon(\tau_L)\varepsilon(\tau_L)']$ and thus its Choleski factorization, are tied to the parameters governing the VAR dynamics. Imposing such restrictions - while feasible - may be convoluted. Fortunately, there is an easy shortcut. It is worth recalling that the instruments used in the estimation of (3.25), and all $FFR$ equations across all $k$, are orthogonal to the error $x_L(\tau_L)$ - $E[x_L(\tau_L)|I(\tau_L, k)]$. Therefore, using an argument often invoked in the estimation of rational expectations models (see e.g. McCallum (1976)), we can obtain consistent estimates of $A_c^{k,m+1}$ in equation (3.25) using ex post realizations of low frequency series to analyze the real-time policy decision rules.

### 3.3 Time-varying mixed frequencies and randomly timed events

We started the section with Assumption 2.1 where assumed a fixed $m$ for the balance between low and high frequency data. In many applications this is not the case. For example, the number of trading days varies from month to month, and therefore also from quarter to quarter. Most often, however, this variation is deterministic and driven by pure calendar effects that are perfectly predictable. In this subsection we relax Assumption 2.1 and replace it by:

**Assumption 3.1.** *We consider a K-dimensional process with the first $K_L < K$ elements, collected in the vector process $x_L(\tau_L)$, only observed every $m_\tau$ periods. The remaining $K_H$ = K - $K_L$ series, represented by double-indexed vector process $x_H(\tau_L, k_H)$ which is observed at the (high) frequency periods $k_H$ = $1, \ldots, m_\tau$ during period $\tau_L$. The sequence $\{m_{\tau_L}\}$, $\tau$ = $1, \ldots,$ is deterministic and takes values in a finite set $M(m)$.*

The above assumption deals with perfectly predictable calendar effects which makes $\{m_\tau\}$ typically vary over a small set of values. For example, the number of trading days in a month can be between 20 and 23, depending on the month and holidays. Since the mixed frequencies scheme is time varying we also have time variation in the system dynamics. Namely, we consider:

$$\mathbb{A}_{\Psi}^{\tau_L}(\mathcal{L}_L)(\underline{\mathrm{x}}(\tau_L) - \mu_{\underline{\mathbf{X}}}^{\Psi,\tau_L}) = \varepsilon(\tau_L) \tag{3.27}$$

with $E[\varepsilon(\tau_L)\varepsilon(\tau_L)']$ = $\mathbb{C}^{\tau_L}(\Psi)\mathbb{C}^{\tau_L}(\Psi)'$, which is of dimension $\tilde{m}_\tau^2$, with $\tilde{m}_\tau \equiv K_L$ + $m_\tau * K_H$, and the matrix polynomial $\mathbb{A}_{\Psi}^{\tau}(\mathcal{L}_L)$ is of dimension $\tilde{m}_{\tau_L} \times \tilde{m}_{(\tau-1)_L}$, and the vector $\mu_{\underline{\mathbf{X}}}^{\Psi,\tau_L}$ is of dimension $\tilde{m}_\tau$. Note that these matrices are no longer the typical squared ones encountered in traditional VAR models or the mixed frequency VAR models with fixed $m$ we have seen so far. Note that we use the same parameter vector $\Psi$ as before. Indeed, one might think that we need to enlarge again the parameter space as we deal with time varying mixed frequencies. Say we that $M(m)$ is of dimension four, to take again the daily trading day example. Now we have potentially sixteen system matrices, four covariance matrices - and thus associated Choleski factorizations for real-time updates - as well as four vectors for the mean. Yet, it is easy to see that equations such as (2.11) and (2.12) remain tightly parameterized if we do two things: (1) replace $m$ by $m_{\tau_L}$, and (2) accommodate time varying mixed frequencies with the same MIDAS polynomials. On the latter subject we refer to the Matlab MIDAS Toolbox (Sinko, Sockin, and Ghysels (2010, Sec. 2.8)) where various schemes are discussed for MIDAS polynomials that handle time varying mixed frequencies within the same framework as fixed $m$ specifications that are characterized by unequal number of MIDAS lags over time that cover the same time span - say a month. Hence, we typically still have the same small number of parameters, despite the time variation in mixed frequencies. As long as the mixture is perfectly predictable, we can use the right dimensions of system matrices as well as the suitable Choleski factorizations to do estimation and updating.

Next, we consider randomly timed events. In footnote 3 we touched on the fact that data release schemes or events like FOMC meetings may change across $\tau_L$. In particular, let us consider a sequence $E \equiv [(\tau_L^e, k_H^e), e = 1, \ldots, T^e]$. Hence, the sequence is a set of time stamps for events. This may be in the context of fixed $m$, as in Assumption Assumption 2.1, or time varying mixed frequencies as in Assumption 3.1. To keep notation simple, it will be easier to look at the fixed $m$ case. Associated with the sequence $E$ is a sequence $E^-$ which gives all the time stamps prior to the events in $E$. Typically, $(\tau_L^{e^-}, k_H^{e^-})$ will be $(\tau_L^e, k_H^e - 1)$, but it may be $(\tau_L^e - 1, m)$ if $k_H^e$ = 1. We will take advantage of the real-time updating to handle randomly timed events. In particular, for the sequence $E$ we pre-multiply $\underline{\mathrm{x}}(\tau_L^{e-})$ by $\mathcal{N}_{[k_H^{e-}]}$ - to have the pre-event predictions - and pre-multiply $\underline{\mathrm{x}}(\tau_L^{e})$ by $\mathcal{N}_{[k_H^{e}]}$ using the suitable matrices appearing in equation (3.20). These computations will measure the impact of the event sequence $E$ on predictions of both low and high frequency data. As noted before, this is reminiscent of a Kalman filter update without going through a latent process specification, measurement equations, etc. An example is the timing of FOMC meetings. Francis, Ghysels, and Owyang (2011) use arguments that are conceptually similar in the context of a single MIDAS regression - hence not a complete mixed frequency VAR - to study the low frequency impact of monetary policy shocks identified via the occurrence of FOMC meetings.

# 4 Implied Low Frequency VAR Models

In this section we characterize the relationship between the mixed frequency data VAR appearing in (2.6) and the aggregated series VAR in (2.8). Recall that in the previous section we mostly worked with vector $\underline{\mathrm{x}}(\tau_L)$ appearing in (2.6) rather than $\ddot{x}(\tau_L)$, while in this section we will work with the latter.

We need to be more specific about how the low frequency VAR model is obtained in terms of aggregation. Throughout the section we will work with fixed mixed sampling frequencies, as stated in Assumption 2.1. Moreover, we will start with a simple skip-sampling scheme where the low frequency VAR model is obtained from picking every last high frequency observation of the low frequency time period. The latter will be relaxed in subsequent analysis. We noted before that we assume that all the processes are covariance stationary and therefore have a spectral representation. Indeed, it will be convenient to characterize the mapping between the mixed frequency VAR model in (2.6) and the aggregated series VAR in (2.8) via their spectral domain representation. Then the following result holds:

**Theorem 4.1.** *Let the process $\ddot{x}(\tau_L)$ satisfy Assumptions 2.1 and 2.2. Moreover, let $\ddot{x}(\tau_L)$ and $\overline{x}(\tau_L)$ have spectral densities equal to respectively $\ddot{S}(z)$ and $\overline{S}(z)$, for $z = \exp(-i\omega)$ with $\omega \in [0, \pi]$, which can be written as:*

$$\begin{aligned} \ddot{S}(z) &= \ddot{\mathbb{A}}(z)^{-1}\ddot{\mathbb{C}}\ddot{\mathbb{C}}'(\ddot{\mathbb{A}}(z^{-1})^{-1})' \\ \overline{S}(z) &= \mathbb{B}(z)^{-1}\overline{\mathbb{C}}\overline{\mathbb{C}}'(\mathbb{B}(z^{-1})^{-1})' \end{aligned} \tag{4.1}$$

*then the low frequency VAR model is determined by the following relationship:*

$$\overline{S}(z) = \frac{1}{m}\sum_{j=0}^{m-1}\frac{1}{m}Q(\frac{z+2\pi j}{m})\ddot{S}((\frac{z+2\pi j}{m})^m)Q((\frac{z+2\pi j}{m})^{-1}) \tag{4.2}$$

*where $Q(z) = \begin{bmatrix} I & zI \ldots z^{m-1}I \end{bmatrix}$*

Proof: See Appendix B

The result in equation (4.2) is a combination of two operations: (1) the averaging of high frequency covariances as in a typical Tiao and Grupe (1980) formula and (2) the skip-sampling of the aforementioned process. Intuitively, what drives the result is the following: the process $\ddot{x}(\tau_L)$ contains all the covariance relationships among the high frequency series and between high and low frequency series. The formula in (4.2) averages these covariances within $\tau_L$ to produce a low frequency covariance structure of the repeated stacked vector process. The latter provides the key ingredients of the traditional VAR impulse response functions.

To handle more general aggregation schemes, we introduce a $K$-dimensional latent high frequency process $z_H(\tau_L, k_H)$, used to construct $\ddot{x}(\tau_L)$. We will focus on linear aggregation schemes - that includes the two most common cases, stock and flow aggregation. In general, we will consider the $m * K$-dimensional stacked vector $[z_H(\tau_L, 1)' \ldots z_H(\tau_L, m)']'$, and generically denote the aggregation filter as:

$$
\begin{aligned}
\ddot{x}(\tau_L) &= \mathbb{D}(\mathcal{L}_H) \begin{bmatrix} z_H(\tau_L, 1) \\ \vdots \\ z_H(\tau_L, m) \end{bmatrix} \\
\mathbb{D}(\mathcal{L}_H) &\equiv diag(\mathbb{D}_1(\mathcal{L}_H), \ldots, \mathbb{D}_m(\mathcal{L}_H))
\end{aligned} \tag{4.3}
$$

where the aggregation scheme may involve long spans, i.e. $P_a$ may be larger than $m$. For example, the $K_L^i$ elements of the low frequency vector released at different times may all pertain to $(\tau - 1)_L$ realizations of $z_H$. Stock and flow sampling schemes are special cases.[8] In section 6 we will provide some specific examples of filters. For general aggregation schemes we can characterize the result in equation (4.2) in terms of $\ddot{S}_{z_H}$ which is the spectral density of $[z_H(\tau_L, 1)' \ldots z_H(\tau_L, m)']'$, namely:

$$
\overline{S}(z) = \frac{1}{m} \sum_{j=0}^{m-1} \frac{1}{m} Q_{\mathbb{D}}(\frac{z + 2\pi j}{m}) \ddot{S}_{z_H}((\frac{z + 2\pi j}{m})^m) Q_{\mathbb{D}}((\frac{z + 2\pi j}{m})^{-1}) \tag{4.4}
$$

where $Q_{\mathbb{D}}(z) = Q(z)\mathbb{D}(z)$. Hence, one can think of the low frequency VAR model in terms of skip-sampled filtered $z_H$, which may represent a combination of flow and stock variables - through filtering with $\mathbb{D}(\mathcal{L}_H)$, which also may capture a mixture of releases involving publication delays (as noted in the discussion below equation (4.3)). Naturally, the aggregation scheme will affect how the low frequency VAR model will look like. Publication delays, i.e. one of the low frequency series pertains to $\tau_L - j$, will shape differently the low frequency VAR dynamics. In section 6 we will provide numerical illustrative examples of such delay effects.

[8] For more discussion of general linear aggregation schemes, see e.g. Lütkepohl (1987).

## 5 Specification and Estimation

Empirical work involves critical choices of model specification and parametrization. In the context of VAR models this amounts to selecting: (1) the variables that are included in the VAR, (2) the sampling frequency of the model, (3) the number of lags to be included and (4) restrictions on the parameter space. Choices of the second type - namely sampling frequency - are not much discussed in the literature and are the focus of this section. Obviously, the choice of sampling frequency is not detached from all the other aforementioned model selection choices. For instance, lag selection is very much related to sampling frequency and so are the parameterizations of the VAR.

To formulate a maximum likelihood based estimator of mixed frequency VAR models, consider the conditional density of the $\tau_L^{th}$ observation:

$$\begin{aligned} f(\underline{\mathrm{x}}(\tau_L)|\underline{\mathrm{x}}(\tau_L-1),\ldots,\underline{\mathrm{x}}(\tau_L-P);\Psi) &= (2\pi)^{\tilde{m}_{\tau_L}}|(\mathbb{C}^{\tau_L}(\Psi)\mathbb{C}^{\tau_L}(\Psi)')^{-1}|^{1/2}\times \\ &\quad \exp\left(\varepsilon(\tau_L)'(\mathbb{C}^{\tau_L}(\Psi)\mathbb{C}^{\tau_L}(\Psi)')^{-1}\varepsilon(\tau_L)\right) \end{aligned}$$

which yields the sample log likelihood function for a sample of size $T_L$ :

$$\begin{aligned} \mathbb{L}(\underline{\mathrm{x}}(\tau_L)_1^{T_L}|\Psi) &= (-1/2)\sum_{\tau_L=1}^{T_L}\tilde{m}_{\tau_L}\log(2\pi)+(1/2)\sum_{\tau_L=1}^{T_L}\log|(\mathbb{C}^{\tau_L}(\Psi)\mathbb{C}^{\tau_L}(\Psi)')^{-1}| \\ &\quad -(1/2)\sum_{\tau_L=1}^{T_L}[\varepsilon(\tau_L)'(\mathbb{C}^{\tau_L}(\Psi)\mathbb{C}^{\tau_L}(\Psi)')^{-1}\varepsilon(\tau_L)] \end{aligned} \tag{5.1}$$

which for $m$ fixed, i.e. under Assumption 2.1, specializes to the usual sample log likelihood function:

$$\begin{aligned} \mathbb{L}(\underline{\mathrm{x}}(\tau_L)_1^{T_L}|\Psi) &= (-T_L(K_L+m*K_H)/2)\log(2\pi)+(T_L/2)\log|(\mathbb{C}(\Psi)\mathbb{C}(\Psi)')^{-1}| \\ &\quad -(1/2)\sum_{\tau_L=1}^{T_L}[\varepsilon(\tau_L)'(\mathbb{C}(\Psi)\mathbb{C}(\Psi)')^{-1}\varepsilon(\tau_L)] \end{aligned} \tag{5.2}$$

The asymptotic analysis of VAR models is well known, see e.g. Hamilton (1994), and applies in the current setting without any modifications. This also covers the case of time varying mixed sampling frequencies under Assumption 3.1.

In a first subsection we cover the asymptotic properties of mis-specified VAR models with an emphasis on mixed versus low frequency specifications. A final subsection covers Bayesian mixed frequency VAR estimation.

## 5.1 Asymptotic Properties of Mis-specified Low Frequency Data VAR Model Estimators

Having specified some potentially parsimonious mixed frequency specifications, we now turn our attention to the comparison of a low frequency VAR model with a finite number of lags and with parameter vector $\Phi$ compared with a mixed frequency VAR model with finite number of lags and parameter vector $\Psi$. More specifically, we assume the DGP is the $m*K$ dimensional vector $\ddot{x}(\tau_L)$ described by equation (2.10):

$$\ddot{\mathbb{A}}(\mathcal{L}_L)(\ddot{x}(\tau_L) - \mu_{\ddot{x}}) = \ddot{\varepsilon}(\tau_L)$$

with $E[\ddot{\varepsilon}(\tau_L)\ddot{\varepsilon}(\tau_L)']$ = $\ddot{\mathbb{C}}\ddot{\mathbb{C}}'$. Against the backdrop of this DGP we have on the one hand the mixed frequency VAR specification appearing in (2.15) parameterized by $\Psi$ :

$$\ddot{\mathbb{A}}_\Psi(\mathcal{L}_L)(\ddot{x}(\tau_L) - \mu_{\ddot{x}}^{\Psi}) = \ddot{\varepsilon}(\tau_L)$$

with $E[\varepsilon(\tau_L)\varepsilon(\tau_L)']$ = $\mathbb{C}(\Psi)\mathbb{C}(\Psi)'$, and on the other hand the $K$ dimensional traditional low frequency VAR parameterized by $\Phi$ :

$$\mathbb{B}_\Phi(\mathcal{L}_L)(\overline{x}(\tau_L) - \mu_{\overline{x}}^{\Phi}) = \overline{\varepsilon}(\tau_L) \tag{5.3}$$

where $\mathbb{B}(\mathcal{L}_L)$ = $I$ - $\sum_{j=1}^{\mathcal{P}} B_j(\Phi)\mathcal{L}_L^j$, and $E[\overline{\varepsilon}(\tau_L)\overline{\varepsilon}(\tau_L)']$ = $\overline{\mathbb{C}}(\Phi)\overline{\mathbb{C}}(\Phi)'$. Hence, we look at a researcher who ignores the high frequency data, picks a finite set of lags and possibly imposes parameter restrictions on the VAR, versus a researcher who looks at the high frequency data, picks a finite set of lags - not necessarily the right number - and possibly imposes restrictions to tackle parameter proliferation.

The use of the DGP in equation (2.10) as the benchmark against which to assess approximation errors may require some explanation. We think of the DGP as a description of the data series sampled at their primitive sampling frequencies. It is against this backdrop that we compare parsimoniously parameterized mixed frequency VAR models and traditional low frequency ones. Since the discussion here essentially revolves around the estimation of mis-specified linear Gaussian processes, we will be using a notion of *relative* (rather than absolute) entropy - that is the Kullback and Leibler (1951) measure to assess approximation errors. For the latter, the penalty for high dimensional systems will not appear as we have chosen it to be the benchmark against which all other models are compared.

Analogous to the equation (5.2) we also have the sample log likelihood function:

$$\begin{aligned}\mathbb{L}(\overline{x}(\tau_L)_1^{T_L}|\Phi) \quad = \quad & (-T_L(K_L + K_H)/2)\log(2\pi) + (T_L/2)\log|(\overline{\mathbb{C}}(\Phi)\overline{\mathbb{C}}(\Phi)')^{-1}| \\ & -(1/2)\sum_{\tau_L=1}^{T_L}[\varepsilon(\tau_L)'(\overline{\mathbb{C}}(\Phi)\overline{\mathbb{C}}(\Phi)')^{-1}\varepsilon(\tau_L)]\end{aligned} \tag{5.4}$$

Using results from Hansen and Sargent (1993) we obtain the following:

**Proposition 5.1.** *Let Assumptions 2.1 and C.1 through C.4 hold and the DGP is described by (2.10). Moreover, assume the low frequency process is constructed as a skip sampled sequence analogous - as* *in Theorem 4.1. Then the maximum likelihood estimator appearing in (5.1), denoted $\hat{\Psi}$, minimizes*

$$\begin{aligned}
\hat{\Psi} &= \underset{\Psi}{\text{Argmin}} \Big[E_1(\ddot{S}(\Psi), \ddot{S}) + E_2(\ddot{S}(\Psi), \ddot{S}) + E_3(\ddot{S}(\Psi), \ddot{S})\Big] \\
E_1(\ddot{S}(\Psi), \ddot{S}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log(\det \ddot{S}(\omega, \Psi))d\omega \\
E_2(\ddot{S}(\Psi), \ddot{S}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} trace(\ddot{S}(\omega, \Psi)^{-1}\ddot{S}(\omega))d\omega \\
E_3(\ddot{S}(\Psi), \ddot{S}) &= (\mu_{\ddot{x}} - \mu_{\ddot{x}}^{\Psi})'\ddot{S}(0, \Psi)(1)^{-1}(\mu_{\ddot{x}} - \mu_{\ddot{x}}^{\Psi})
\end{aligned} \tag{5.5}$$

*whereas the maximum likelihood estimator appearing (5.4), denoted $\hat{\Phi}$, minimizes*

$$\begin{aligned}
\hat{\Psi} &= \underset{\Phi}{\text{Argmin}} \Big[E_1(\overline{S}(\Phi), \ddot{S}) + E_2(\overline{S}(\Phi), \ddot{S}) + E_3(\overline{S}(\Phi), \ddot{S})\Big] \\
E_1(\overline{S}(\Phi), \ddot{S}) &= \frac{m}{2\pi}\int_{-\pi}^{\pi} \log(\det \overline{S}(\omega, \Phi))d\omega \\
E_2(\overline{S}(\Phi), \ddot{S}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} trace(\overline{S}(\omega, \Phi)^{-1}Q(\exp(i\omega))'\ddot{S}(\omega)Q(\exp(-i\omega)))d\omega \\
E_3(\overline{S}(\Phi), \ddot{S}) &= (\mu_{\ddot{x}} - 1_m \otimes \mu_{\overline{x}}^{\Phi})'\overline{S}(0, \Phi)(1)^{-1}(\mu_{\ddot{x}} - \otimes\mu_{\overline{x}}^{\Phi})
\end{aligned} \tag{5.6}$$

Proof see Appendix C

Note that if the mixed frequency VAR model is correctly specified, then the terms $E_2$ and $E_3$ in equation (5.5) disappear and one has a standard MLE. Likewise, again assuming the mixed frequency VAR model is correctly specified, one can replace all the terms involving the DGP in equation (2.10) appearing in $E_2$ and $E_3$ of equation (5.6) with the parameterized mixed frequency VAR one. Note also that the term $E_3$ in both cases refers to the mis-specification of the mean. The term is important when it comes to approximation errors in the context of periodic models, as emphasized by Hansen and Sargent (1993). In our analysis, this play less of a role and we will typically handle cases without specification errors of the overall mean of the process - whether it is sampled at low or high frequency.

Finally, we far we assumed a skip sampling scheme in the Proposition 5.1. It would be easy to adopt the general aggregation schemes we discussed in the previous section. It suffices to replace $Q(.)$ with $Q_{\mathbb{D}}(z)$, and make the appropriate changes to the spectral representations. We refrain here from providing the details, as they are relatively straightforward.

## 5.2 Bayesian Mixed Frequency VAR

Recent work on MIDAS regressions includes Bayesian estimation approaches, see notably Rodriguez and Puggioni (2010) and Ghysels and Owyang (2011). It is the purpose of this section to expand these recent developments to a mixed frequency VAR framework. We do so with the objective of staying as close as possible to the standard Bayesian VAR approach.

It will be convenient to start again with a simplified example. Namely, consider the case where all the low frequency $\tau_L$ series are release at the end of the period as in equation (2.1) with zero mean series, with $K_L$ and $K_H$ of any dimension (hence not necessarily one). To highlight the role of the MIDAS regression parameters we write the last set of equations as a function of $\gamma$ :

$$\begin{bmatrix} x_H(\tau_L,1) \\ \vdots \\ x_H(\tau_L,m) \\ x_L(\tau_L) \end{bmatrix} = \sum_{j=1}^{P} \begin{bmatrix} A_j^{1,1} & \dots & A_j^{1,m} & A_j^{1,m+1} \\ \vdots & \dots & \vdots & \vdots \\ A_j^{m,1} & \dots & A_j^{m,m} & A_j^{m,m+1} \\ A_j^{m+1,1} & \dots & A_j^{m+1,m} & A_j^{m+1,m+1} \end{bmatrix} \begin{bmatrix} x_H(\tau_L - j,1) \\ \vdots \\ x_H(\tau_L - j,m) \\ x_L(\tau_L - j) \end{bmatrix} + \varepsilon(\tau_L)$$

where $dim(A_j^{m+1,m+1}) = K_L^2$, $dim(A_j^{i,m+1}) = K_H \times K_L$ for $i = 1,\ldots,m$ and finally $dim(A_j^{a,b}, a, b = 1,\ldots,m) = K_H^2$.

Since the MIDAS part of the VAR is novel in terms of Bayesian estimation, we focus first on its formulation. We therefore start with the matrices $A_j^{m+1,\cdot}(\gamma)$ and make them explicitly functions of the MIDAS polynomial parameters. We will suggest two approaches - one for general MIDAS polynomial specifications and a second using the MIDAS with step functions of Ghysels, Sinko, and Valkanov (2006) and U-MIDAS (unrestricted MIDAS polynomial) approach suggested by Foroni, Marcellino, and Schumacher (2011) (see also Appendix A). We cover the general case first.

Recall from subsection 2.4 that we considered two schemes, appearing in equations (2.13) and (2.14):

$$[A_1^{m+1,1}(\gamma)\ldots A_1^{m+1,m}(\gamma)A_2^{m+1,1}(\gamma)\ldots A_P^{m+1,m}(\gamma)] = \begin{cases} B \otimes \left[\sum_{i=1}^{K_H \times P} (w(\gamma)_i)\right] \\ \\ \left[B_i \otimes \left[\sum_{i=1}^{m} (w(\gamma)_i)\right], i = 1,\ldots,P\right] \end{cases} \tag{5.7}$$

Recall also that $B$ a $K_L \times K_H$ matrix and in subsection 2.4 we assumed $\sum_{i=1}^{K_H \times P} (w(\gamma)_i)$ is a *scalar* MIDAS polynomial such that the weighting schemes are the same across the different low frequency equations. We can easily relax this, by assuming a scheme where all the MIDAS polynomials are driven by a common prior, namely $[A_1^{m+1,1}(\gamma)\ldots A_1^{m+1,m}(\gamma)A_2^{m+1,1}(\gamma)\ldots A_P^{m+1,m}(\gamma)]$ can be expressed as:

$$\begin{cases} \left(B^{a,b}\left[\sum_{i=1}^{K_H \times P}\left(w(\gamma^{a,b})_i\right)\right], a = 1,\ldots,K_L; b = 1,\ldots,K_L\right) \\ \\ \left[B_i^{a,b}\left[\sum_{i=1}^{m}\left(w(\gamma^{a,b})_i\right)\right], i = 1,\ldots,P, a = 1,\ldots,K_L; b = 1,\ldots,K_L\right] \end{cases} \tag{5.8}$$

We will consider the case of MIDAS Beta polynomials (see Appendix A), the other cases are similar and therefore not covered. The prior both in the case of a single MIDAS polynomial (5.7) or the common prior in the case of many single MIDAS polynomials as in (5.8), is a Gamma distribution. Since the MIDAS Beta polynomial involves two parameters, we draw each parameter from an independent Gamma. In the case of (5.8) the $K_L \times K_H$ MIDAS polynomials each involve two parameters and they also have two independent Gamma distributions. We use a Gamma distribution as the values of the Beta polynomial take on positive values. For simplicity we cover the single MIDAS polynomial, then the prior for $\gamma \equiv (\gamma_1, \gamma_2)$ is:

$$\gamma_i \sim \Gamma\left(\mathbf{f}_0, \mathbf{F}_0\right) \quad i = 1, 2 \tag{5.9}$$

where $\mathbf{f}_0 = \mathbf{F}_0 = 1$. This prior amounts to a flat weighting scheme that put equal weight on all high frequency data. Yet, there are several variations that put further restrictions. They are: (a) downward sloping weights: $\gamma \equiv (1, \gamma_2)$ with $\gamma_2 \sim \Gamma\left(\mathbf{f}_0, \mathbf{F}_0\right)$ and $\mathbf{f}_0 = \mathbf{F}_0 = 1$, (b) hump-shaped weights $\gamma \equiv (1 + \gamma_1, 1 + \gamma_1 + \gamma_2)$, among others. The downward sloping scheme is particularly appealing as it amount to a single parameter weighting scheme.

Following Ghysels and Owyang (2011), we utilize a Metropolis-in-Gibbs step (as in Chib and Greenberg (1995)) to sample the MIDAS hyperparameters. The Metropolis step is an accept-reject step which requires a candidate draw, $\gamma^*$, from a proposal density, $q\left(\gamma^*|\gamma^{[i]}\right)$, where $\gamma^{[i]}$ is the last accepted draw. The draw is then accepted with a probability that depends on both the likelihood and parameters' prior distribution. In this case, the functional form of the weighting polynomial motivates our choice of the proposal density. Because we have chosen the beta weighting polynomial, a Gamma proposal distribution provides a suitable candidate.

To formalize, for the $(i+1)$ iteration, we can draw a candidate $\gamma^* = (\gamma_1^*, \gamma_2^*)'$ from

$$\gamma_j^* \sim \Gamma\left(c\left(\gamma_j^{[i]}\right)^2, c\gamma_j^{[i]}\right),$$

where $c$ is a tuning parameter chosen to achieve a reasonable acceptance rate. The candidate draw is then accepted with probability $a = \min\{\alpha, 1\}$, where

$$\alpha = \frac{\mathbb{L}(\underline{\mathrm{x}}(\tau_L)_1^{T_L}|\Psi_{-\gamma}, \gamma^*)}{\mathbb{L}(\underline{\mathrm{x}}(\tau_L)_1^{T_L}|\Psi_{-\gamma}, \gamma^{[i]})} \frac{d\Gamma\left(\gamma^*|\mathbf{f}_0, \mathbf{F}_0\right)}{d\Gamma\left(\gamma^{[i]}|\mathbf{f}_0, \mathbf{F}_0\right)} \frac{d\Gamma(\gamma^{[i]}|c\left(\gamma^*\right)^2, c\gamma^*)}{d\Gamma(\gamma^*|c\left(\gamma^{[i]}\right)^2, c\gamma^{[i]})},$$

where $\mathbb{L}(\underline{\mathrm{x}}(\tau_L)_1^{T_L}|\Psi_{-\gamma}, \gamma^*)$ is the conditional likelihood given the parameters $\Psi_{-\gamma}$ - which are all the parameters in $\Psi$ excluding $\gamma$ and $d\Gamma\left(.|.,.\right)$ is the Gamma density function. Obviously, whenever there are multiple MIDAS polynomials the aforementioned Metropolis step is repeated for each weighting scheme separately. Hence, in such case we essentially draw various weighting profiles. For convenience we will keep using the notation $\gamma$ for a single as well as multiple MIDAS polynomial weighting schemes to avoid further complicating the notation.

The Bayesian analysis of mixed frequency VAR and traditional VAR models becomes quite similar once the parameter draws for the MIDAS polynomial are given. First we write the VAR as a first order system (again assuming unconditional mean zero processes):

$$\underline{\mathrm{x}}(\tau_L) = [\underline{\mathrm{x}}(\tau_L - 1) \dots \underline{\mathrm{x}}(\tau_L - P)] \times \begin{pmatrix} A_1(\Psi_{-\gamma}, \gamma) \\ A_2(\Psi_{-\gamma}, \gamma) \\ \vdots \\ A_P(\Psi_{-\gamma}, \gamma) \end{pmatrix} + \varepsilon(\tau_L) \tag{5.10}$$

where we acknowledge that the lag matrices $A_j(\Psi_{-\gamma}, \gamma)$ depend on the parameter vector driving the MIDAS polynomial weights as well as the remaining parameters in $\Psi_{-\gamma}$.

Equation (5.10) looks like a 'regular' regression framework associated with VAR models. Hence, from here on we can follow Doan, Litterman, and Sims (1984), Litterman (1986), Kadiyala and Karlsson (1997), Sims and Zha (1998), among others, for the formulation of priors regarding $\Psi_{-\gamma}$, which is partitioned into three blocks (dropping the dependence on $\gamma$ etc. for convenience):

- $\Psi_{-\gamma,H} = ((A_j^{a,b}, a = 1, \dots, m, b = 1, \dots, m+1), j = 1, \dots, P)'$, the set parameters pertaining to the high frequency components of the vector.
- $\Psi_{-\gamma,L} = ((B \quad or \quad B_i, i = 1, \dots, P)'$, the slope parameters pertaining to the MIDAS regressions.
- $\Psi_{-\gamma,V} = ((\mathbb{C}\mathbb{C}')_{a,b}; a = 1, \dots, m+1, b = 1, \dots, m+1)'$, the parameters pertaining to the covariance matrix of the errors.

In particular, the means and variances for the priors in $\Psi_{-\gamma,H}$ are (dropping again the dependence on $\gamma$ etc.):

$$\begin{array}{lll}
\mathbb{E}[A_j^{a,b}] = \mathbf{0}_{K_H^2}, & \mathbb{V}[A_j^{a,b}] = \frac{\lambda^2}{[(j-1)m+(m-b+a)]^2}\mathbf{1}_{K_H^2} & a = 1, \dots, m, b = 1, \dots, m-1 \\
\mathbb{E}[A_1^{a,m}] = diag(\rho^a)_{K_H^2} & \mathbb{V}[A_1^{a,m}] = \frac{\lambda^2}{[(a)]^2}\mathbf{1}_{K_H^2} & a = 1, \dots, m \\
\mathbb{E}[A_j^{a,m}] = \mathbf{0}_{K_H^2}, & \mathbb{V}[A_j^{a,m}] = \frac{\lambda^2}{[(j-1)a+a]^2}\mathbf{1}_{K_H^2} & j > 1, a = 1, \dots, m \\
\mathbb{E}[A_j^{a,m+1}] = \mathbf{0}_{K_H \times K_L}, & \mathbb{V}[A_j^{a,m+1}] = \vartheta_{HL}\frac{\lambda^2}{[(j-1)m]^2}\mathbb{S}_{HL} & a = 1, \dots, m
\end{array} \tag{5.11}$$

where the notation $\mathbb{V}[\quad]$ stands for a matrix of variances, $\mathbf{0}$ and $\mathbf{1}$ are matrices respectively of zeros and ones, with the dimension as subscript, $diag(x)$ is a diagonal matrix with elements $x$ and again the dimension as subscript, and finally $\mathbb{S}_{HL} \equiv [\sigma_{i,H}^2/\sigma_{j,L}^2; i = 1, \dots, K_H, j = 1, \dots, K_L]$. The latter captures the difference in scaling between high and low frequency data, as is typically done in Bayesian VAR models (see above references). The hyperparameter $\lambda$ governs the overall tightness of the prior distributions around the AR(1) (including white noise) specification for the high frequency process. The hyperparameter $\vartheta_{HL} \in (0, 1)$ governs the extent to which the low frequency data affect high frequency data. Note that we leave within low frequency series prior distribution uniform. Namely, since we write $\mathbb{V}[A_j^{a,b}]$ is only scaled by $j$, $a$ and $b$ we essentially treat the dependence within the vector of high frequency data as uniform. We can change this by replacing $\mathbf{1}_{K_H^2}$ with a matrix that would involve another set of hyperparameters that would govern the extent to which low frequency series are mutually affected. This is an easy generalization which we do not consider for the sake of simplicity.

The prior in (5.11) is much inspired by the parsimonious representation appearing (2.11). Namely, it essentially says that all the high frequency processes are AR(1) with autoregressive parameter $\rho$ that is common among all the series. We typically set $\rho = 0$, i.e. all high frequency processes are white noise, or put $\rho$ equal to some value between zero and one, and possibly equal to one. The variances of the prior tell us that the precision on the parameters is tighter as lags increase. This is typically done in traditional Bayesian VAR models, and is shrinking at a rate that is the square of the lag length as in Litterman (1986). Note, however, that the decay is not only governed by $j$, but also by $(a, b)$ as they represent the intra-$\tau_L$ period lag structure. Regarding the MIDAS regressions in $\Psi_{-\gamma,L}$, we have the following priors for the slope coefficients:

$$\begin{aligned} \mathbb{E}[B] &= \mathbf{0}_{K_L \times K_H}, \quad \mathbb{V}[B] = \vartheta_{LH}\lambda^2 \mathbf{1}_{K_L \times K_H}\mathbb{S}_{LH} \\ \mathbb{E}[B_j] &= \mathbf{0}_{K_L \times K_H}, \quad \mathbb{V}[B_j] = \vartheta_{LH}\tfrac{\lambda^2}{j^2} \mathbf{1}_{K_L \times K_H}\mathbb{S}_{LH} \end{aligned} \tag{5.12}$$

with $\vartheta_{LH}$ and $\mathbb{S}_{LH}$ having interpretations similar to the ones considered for the high frequency data regressions in the VAR. Not that the prior in (5.12) implies that we typically start from a VAR that has flat weights for the MIDAS polynomial and the high frequency data do not have an impact on the low frequency data. Note also that the reverse is also true since we put the prior that $\mathbb{E}[A_j^{a,m+1}] = \mathbf{0}_{K_H \times K_L}$.

Last but not least, we also need to formulate priors for $\Psi_{-\gamma,V}$. Here we refer to Kadiyala and Karlsson (1997) who consider so called Minnesota priors with fixed residual covariance matrices, or the Normal-Wishart and Diffuse priors, the Normal-Diffuse and Extended Natural Conjugate priors. In all cases they derive the posterior distributions which are summarized in Kadiyala and Karlsson (1997, Table 1). The MCMC procedure therefore can rely on the explicitly derived posterior distributions - conditional of draws of $\gamma$ for the MIDAS weights.

It is important to note that one can also implement the estimation of mixed frequency VAR models using exclusively standard Bayesian VAR methods, that is avoid the use of the Metropolis step described earlier in the subsection. This involves the step functions approach to MIDAS of Ghysels, Sinko, and Valkanov (2006) and the completely unrestricted specification where each weight is estimated separately, i.e. the U-MIDAS approach suggested by Foroni, Marcellino, and Schumacher (2011) shown to work for small values of $m$. Using the step function example, we can think of replacing equation (5.7) with:

$$[A_1^{m+1,1} \ldots A_1^{m+1,m} A_2^{m+1,1} \ldots A_P^{m+1,m}] = \Big[\underbrace{B_1 \ldots B_1}\underbrace{B_2 \ldots B_2} \ldots \underbrace{B_s \ldots B_s}\Big] \tag{5.13}$$

with $s$ step matrices that imply fixed lag effects across subsets that span $P \times m$ lags.[9] The U MIDAS case is one where $s = P \times m$, and therefore each matrix is unrestricted. Now, we are left with specifying priors for the step functions. The easy case is where the steps are multiples of $m$, as is usually the case. Then we can use a prior similar to that appearing in (5.12) for $\mathbb{E}[B_j]$ and $\mathbb{V}[B_j]$. In particular in the U MIDAS case one may also think of using the shrinkage methods for large dimensional VAR models. However, large VAR models in the traditional sense means a lot of individual series (sampled at the same frequency), whereas here large dimensions result from the stacking of the same series. There is already a great amount of shrinkage explicitly taking into account via the priors appearing in equations (5.11) and (5.12).

# 6 Numerical Examples

In this section we provide some numerical illustrative examples to compare the behavior of impulse response functions in mixed frequency and traditional VAR models. We look at bivariate systems, i.e. we study cases with a single series of each type - low and high frequency. We do this to simplify the study of impulse response functions. We use the setup in equation (2.11). We consider $m$ =3, corresponding to a monthly/quarterly data mixture and $m$ = 12, representing a weekly/quarterly mixture. For the former - which we refer to as Case 1, we have: $a = 0$, $\rho = .5$, $\alpha_1 = .5$, with $[w(\gamma)_3, w(\gamma)_2, \quad w(\gamma)_1] = [.2, .5, .9]$ and set the innovation covariance matrix to an identity matrix. Therefore, Case 1 corresponds to mildly persistent high and low frequency series (since $\rho = \alpha_1 = .5$), without any impact of low frequency onto high frequency series. Moreover, the MIDAS weights feature a typical downward decay pattern. Given the mixed frequency VAR data generating process we characterize the corresponding traditional VAR(1) model - with point-sampling of the high frequency data - which is obtained from minimizing the criterion appearing in Proposition 5.1. Hence, we view the traditional VAR as a misspecified model estimated via standard MLE.[10]

Since $a$ is zero, there is no causality between the low frequency series. Hence, we expect that a shock to $x_L(\tau_L)$ will not affect future $x_H(\tau_L + k, i)$ for $i = 1, \ldots, 3$, for $k > 0$. Recall that we are thinking of a quarterly model, hence the impulse responses are in terms of quarterly time ticks. The results appearing in Figure 1. The left panel shows three impulse response functions: (1) the impact of a shock to $x_H(\tau_L, 1)$ on future quarterly $x_L(\tau_L + k)$, for $k = 1, \ldots, 24$ (quarters) as determined by the mixed frequency VAR, (2) the impact of a shock to $x_H(\tau_L, 3)$ on future quarterly $x_L(\tau_L + k)$, and (3) the impact of a shock to aggregate *high* frequency series in the VAR(1) for $\overline{x}(\tau_L)$ on future quarterly low frequency data. The impulse response functions (1) and (2) feature $+$ whereas (3) is marked by $o$. It is interesting to note that while high frequency shocks affect the low frequency series up to 10 quarters in the future, the standard frequency VAR hardly shows any impact of shocks to the aggregated (skip sampled) high frequency series on the low frequency process.

[9]We assume here for simplicity that all the high frequency series involve the same step sizes.

[10]In the Internet Appendix we provide the Matlab code used to compute the numerical results. Some parts of the code were inspired by code reported in Hansen and Sargent (1990).

It is instructive to look at the numerical coefficients of the mis-specified VAR(1) model which appear in Table 1. Clearly, the implied dynamics have little resemblance with the original mixed frequency dynamics. The high frequency process features no persistence but is driven by the low frequency process. In contrast, the low frequency process picks up more persistence (with $0.889$) than the DGP. The right panel also shows three impulse responses, namely: (1) the impact of a shock to $x_L(\tau_L)$ on future quarterly $x_H(\tau_L + k, 1)$, for $k = 1, \ldots, 24$ as determined by the mixed frequency VAR, (2) the impact of a shock to $x_L(\tau_L)$ on future quarterly $x_H(\tau_L + k, 3)$, and (3) the impact of a shock to aggregate *low* frequency series in the VAR(1) for $\overline{x}(\tau_L)$ on future quarterly high frequency data. Since $a$ is zero, the first two impulse response functions are flat and equal to zero. In great contrast, and as implied by the dynamics in (**??**), we find that the low frequency process greatly impacts the high frequency process. Obviously, this numerical example touches on the topic of Granger causality being altered due to aggregation.[11]

For the second case we reverse the causal relationship between high and low frequency data. In addition we also change the persistence of the series. Namely, we set $a = 0.4$, $\rho = \alpha_1 = .1$ and $[w(\gamma)_3, w(\gamma)_2, \quad w(\gamma)_1] = [0., 0., 0.]$, implying that the high frequency data do not affect the low frequency (which is an exogenous process). The results appear respectively in Figure 2 and Table 1. We observe a similar spurious phenomenon: the impact of high frequency series - which is zero - becomes significant after we estimate a standard VAR as we can see from the left hand side of Figure 2. The right hand side plot in the figure also tells us that the actual impact of low frequency data onto high frequency series is severely mis-specified in the standard VAR as well. As one can read from Table we see that the VAR dynamics become non-trivial with a feedback effect from low frequency to high frequency equal to $0.3805$ together with the spurious impact of $x_H(\tau_L)$ onto $x_L(\tau_L)$ equal to $0.1095$.

The third example involves a mixture of weekly and quarterly series. We consider somewhat different parameter values, namely $a = .01$, $\rho = .1$, $\alpha_1 = .5$, $[w(\gamma)_3, w(\gamma)_2, w(\gamma)_1] = [.5, .4, .3]$ and all other remaining 9 MIDAS polynomial weights equal to zero. Hence, this example causality runs in both directions, albeit of small magnitude. The impulse response functions appear in Figure 3, which has again two panels with the same type of three impulse response functions as in Figure 1. We clearly misread the dynamics of the impulse responses again both in terms of magnitude and timing. We provided only two numerical examples. One obviously can think of many alternative scenarios. The Matlab code provided enables the reader to experiment with alternative specifications.

# 7 Empirical Examples

The empirical application is tailored after Chiu, Eraker, Foerster, Kim, and Seoane (2011) who develop a Bayesian approach to such mixed frequency VAR models where the missing data are drawn via

[11]Much has been written about the spurious effects temporal aggregation may have on testing for Granger causality, see e.g. Granger (1980), Granger (1988), Lütkepohl (1993), Granger (1995), Renault, Sekkat, and Szafarz (1998), Breitung and Swanson (2002), McCrorie and Chambers (2006), among others. In concurrent research we study the topic of testing for Granger causality in a mixed frequency setting - see Ghysels, Motegi, and Valkanov (2011).

a Gibbs sampler. Their primary objective is to formulate a model that allows analysis of GDP at a frequency higher than the quarterly data readily available. Their analysis also has the virtue of keeping the empirical exercise simple and transparent. We want to do the same and therefore replicate their setting for the purpose of comparison. In particular they consider a latent VAR(1) model involving industrial production, inflation, and unemployment rate, and GDP for the US. All but the last series are observed monthly. The data are the twelve-month change in industrial production (denoted IP) and inflation (denoted INFL), the four-quarter change in real GDP (denoted GDP), and the unemployment rate (denoted UNEMP), all expressed as percentage points. Chiu, Eraker, Foerster, Kim, and Seoane (2011) assume - like we do - that every month, the monthly data are observed, and the quarterly data are observed only during the last month of each quarter. We compare the mixed frequency VAR with a traditional quarterly VAR model. Therefore, we study the co-movements of:

$$\begin{bmatrix} IP(\tau_L) \\ INFL(\tau_L) \\ UNEMP(\tau_L) \\ GDP(\tau_L) \end{bmatrix} \quad vs \quad \begin{bmatrix} x_H(\tau_L, 1) \\ x_H(\tau_L, 2) \\ x_H(\tau_L, 3) \\ GDP(\tau_L) \end{bmatrix} \quad with \quad x_H(\tau_L, j) = \begin{bmatrix} IP(\tau_L, j) \\ INFL(\tau_L, j) \\ UNEMP(\tau_L, j) \end{bmatrix} \tag{7.1}$$

**INCOMPLETE**

# 8 Conclusions

## References


AKAIKE, H. (1973): "Information theory and an extension of the maximum likelihood principle," in *Second international symposium on information theory*, vol. 1, pp. 267–281. Springer Verlag.

ANDREOU, E., E. GHYSELS, AND A. KOURTELLOS (2010): "Regression Models With Mixed Sampling Frequencies," *Journal of Econometrics*, 158, 246–261.

ARUOBA, S. B., F. X. DIEBOLD, AND C. SCOTTI (2009): "Real-time measurement of business conditions," *Journal of Business and Economic Statistics*, 27, 417–427.

BAI, J., E. GHYSELS, AND J. WRIGHT (2009): "State space models and MIDAS regressions," *Econometric Reviews*, forthcoming.

BERNANKE, B., M. GERTLER, AND M. WATSON (1997): "Systematic Monetary Policy and the Effects of Oil Price Shocks," *Brookings Papers on Economic Activity*, 1, 91–157.

BREITUNG, J., AND N. SWANSON (2002): "Temporal aggregation and spurious instantaneous causality in multiple time series models," *Journal of Time Series Analysis*, 23, 651–665.

CHEN, X., AND E. GHYSELS (2011): "NewsGood or Badand Its Impact on Volatility Predictions over Multiple Horizons," *Review of Financial Studies*, 24-81, 46.

CHIB, S., AND E. GREENBERG (1995): "Understanding the metropolis-hastings algorithm," *American Statistician*, pp. 327–335.

CHIU, C., B. ERAKER, A. FOERSTER, T. KIM, AND H. SEOANE (2011): "Estimating VARs Sampled at Mixed or Irregular Spaced Frequencies: A Bayesian Approach," Federal Reserve Bank of Kansas City, RWP 11-11.

CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (1996): "The effects of monetary policy shocks: some evidence from the flow of funds," *Review of Economics and Statistics*, 78, 16–34.

CHRISTIANO, L., M. EICHENBAUM, AND C. EVANS (1997): "Sticky price and limited participation models of money: A comparison," *European Economic Review*, 41, 1201–1249.

COX, D. (1981): "Statistical Analysis of Time Series: Some Recent Developments [with Discussion and Reply]," *Scandinavian Journal of Statistics*, 8, 93–115.

DOAN, T., R. LITTERMAN, AND C. SIMS (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3, 1–100.

ENGLE, R. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, pp. 987–1007.

FORONI, C., M. MARCELLINO, AND C. SCHUMACHER (2011): "U-MIDAS: MIDAS regressions with unrestricted lag polynomials," Discussion Paper Deutsche Bundesbank and European University Institute.

FRANCIS, N., E. GHYSELS, AND M. OWYANG (2011): "The Low-Frequency Impact of Daily Monetary Policy Shocks," Discussion paper, Federal Reserve Bank of St Louis Working Paper Series 2011-009B, Available at SSRN: http://ssrn. com/abstract= 1793358.

FRANSES, P. (1996): *Periodicity and stochastic trends in economic time series*. Oxford University Press, USA.

GHYSELS, E. (1994): "On the periodic structure of the business cycle," *Journal of Business and Economic Statistics*, pp. 289–298.

GHYSELS, E., A. HARVEY, AND E. RENAULT (1996): "Stochastic volatility," *Handbook of Statistics*, 14, 119–191.

GHYSELS, E., K. MOTEGI, AND R. VALKANOV (2011): "Granger Causality Tests with Mixed Data Frequencies," Work in Progress.

GHYSELS, E., AND M. OWYANG (2011): "Taylor Monetary Policy Rules with Financial Market Expectations," Discussion Paper, Federal Reserve Bank of Saint-Louis and UNC.

GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2006): "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics*, 131, 59–95.

GHYSELS, E., A. SINKO, AND R. VALKANOV (2006): "MIDAS regressions: Further results and new directions," *Econometric Reviews*, 26, 53–90.

GHYSELS, E., AND J. WRIGHT (2009): "Forecasting professional forecasters," *Journal of Business and Economic Statistics*, 27, 504–516.

GLADYSHEV, E. (1961): "Periodically correlated random sequences," *Soviet Mathematics*, 2, 385–388.

GRANGER, C. (1980): "Testing for Causality: A Personal Viewpoint," *Journal of Economic Dynamics and Control*, 2, 329–352.

——— (1988): "Some Recent Developments in A Concept of Causality," *Journal of Econometrics*, 39, 199–211.

——— (1995): "Causality in the Long Run," *Econometric Theory*, 11, 530–536.

HAMILTON, J. (1994): *Time series analysis*. Princeton University Press, Princeton, New Jersey.

HANSEN, L., AND T. SARGENT (1983): "The dimensionality of the aliasing problem in models with rational spectral densities," *Econometrica*, 51, 377–387.

——— (1990): *Recursive linear models of dynamic economies*. NYU and University of Chicago.

——— (1993): "Seasonality and approximation errors in rational expectations models* 1," *Journal of Econometrics*, 55(1-2), 21–55.

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

HARVEY, A. C., AND R. G. PIERSE (1984): "Estimating missing observations in economic time series," *Journal of the American Statistical Association*, 79, 125–131.

KADIYALA, K., AND S. KARLSSON (1997): "Numerical Methods for Estimation and Inference in Bayesian VAR-models," *Journal of Applied Econometrics*, 12, 99–132.

KULLBACK, S., AND R. LEIBLER (1951): "On information and sufficiency," *Annals of Mathematical Statistics*, 22, 79–86.

KUZIN, V., M. MARCELLINO, AND C. SCHUMACHER (2009): "MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area," Discussion Paper 07/2009 Deutsche Bundesbank.

LITTERMAN, R. (1986): "Forecasting with Bayesian vector autoregressions: Five years of experience," *Journal of Business and Economic Statistics*, 4, 25–38.

LJUNG, L. (1978): "Convergence analysis of parametric identification methods," *Automatic Control, IEEE Transactions on*, 23, 770–783.

LÜTKEPOHL, H. (1987): *Forecasting Aggregated Vector ARMA Processes*. Springer-Verlag, Berlin.

——— (1993): "Testing for Causation Between Two Variables in Higher Dimensional VAR Models," in *Studies in Applied Econometrics*, ed. by H. Schneeweiss, and K. Zimmerman, p. 75. Springer-Verlag, Heidelberg.

MARCELLINO, M., AND C. SCHUMACHER (2010): "Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP," *Oxford Bulletin of Economics and Statistics*, 72, 518–550.

MARIANO, R. S., AND Y. MURASAWA (2003): "A new coincident index of business cycles based on monthly and quarterly series," *Journal of Applied Econometrics*, 18, 427–443.

MCCALLUM, B. (1976): "Rational expectations and the estimation of econometric models: an alternative procedure," *International Economic Review*, 17, 484–490.

MCCRORIE, J., AND M. CHAMBERS (2006): "Granger causality and the sampling of economic processes," *Journal of Econometrics*, 132(2), 311–336.

MILLER, J. (2011a): "Cointegrating MiDaS Regressions and a MiDaS Test," Department of Economics, University of Missouri, Discussion Paper.

——— (2011b): "Conditionally Efficient Estimation of Long-run Relationships Using Mixed-frequency Time Series," *Department of Economics, University of Missouri, mimeo*.

MITTNIK, S., AND P. A. ZADROZNY (2004): *Forecasting quarterly German GDP at monthly intervals using monthly Ifo business conditions data*. Springer.

NUNES, L. (2005): "Nowcasting quarterly GDP growth in a monthly coincident indicator model," *Journal of Forecasting*, 24, 575–592.

PAGANO, M. (1978): "On periodic and multiple autoregressions," *The Annals of Statistics*, pp. 1310–1317.

PHILLIPS, P. (1973): "The problem of identification in finite parameter continuous time models," *Journal of Econometrics*, 1, 351–362.

PÖTSCHER, B. (1987): "Convergence results for maximum likelihood type estimators in multivariable ARMA models," *Journal of Multivariate Analysis*, 21, 29–52.

RENAULT, E., K. SEKKAT, AND A. SZAFARZ (1998): "Testing for spurious causality in exchange rates," *Journal of Empirical Finance*, 5, 47–66.

RODRIGUEZ, A., AND G. PUGGIONI (2010): "Mixed frequency models: Bayesian approaches to estimation and prediction," *International Journal of Forecasting*, 26, 293–311.

RUDEBUSCH, G. (1998a): "Do measures of monetary policy in a VAR make sense?," *International Economic Review*, 39(4), 907–931.

——— (1998b): "Do measures of monetary policy in a VAR make sense? A reply to Christopher A. Sims," *International Economic Review*, 39, 943–948.

SIMS, C. (1980): "Macroeconomics and reality," *Econometrica*, 48(1), 1–48.

——— (1998): "Comment on Glenn Rudebusch's "Do Measures of Monetary Policy in a Var Make Sense?"," *International Economic Review*, 39, 933–941.

SIMS, C., AND T. ZHA (1998): "Bayesian methods for dynamic multivariate models," *International Economic Review*, 39, 949–968.

SINKO, A., M. SOCKIN, AND E. GHYSELS (2010): "Matlab Toolbox for MIDAS Regressions," Available at `http://www.unc.edu/~eghysels/Software_datasets.html`.

TIAO, G., AND M. GRUPE (1980): "Hidden periodic autoregressiye-moving average models in time series data," *Biometrika*, 67(2), 365.

WHITE, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, pp. 1–25.

ZADROZNY, P. A. (1990): "Forecasting US GNP at monthly intervals with an estimated bivariate time series model," *Federal Reserve Bank of Atlanta Economic Review*, 75, 2–15.

# Table 1: Approximate Vector Autoregressive Models

The table reports the mis-specified VAR(1) using the setup in equation (2.11). Given the mixed frequency VAR data generating process we characterize the corresponding traditional VAR(1) model - with point-sampling of the high frequency data - which is obtained from minimizing the criterion appearing in Proposition 5.1.In the Internet Appendix we provide the Matlab code used to compute the numerical results. Case 1 has $m = 3$, with $a = 0$, $\rho = \alpha_1 = .5$, with $[w(\gamma)_3, w(\gamma)_2, w(\gamma)_1] = [.2, .5, .9]$ and set the innovation covariance matrix to an identity matrix. Case 2 has $m$ 3, $a = 0.4$, $\rho = \alpha_1 = .1$, and $[w(\gamma)_3, w(\gamma)_2, \quad w(\gamma)_1] = [.0, .0, .0]$. Case 3 has $m = 12$, $a = .05$, $\rho = .1$, $\alpha_1 = .5$, $[w(\gamma)_3, w(\gamma)_2, w(\gamma)_1] = [.5, .4, .3]$ with all other remaining 9 MIDAS polynomial weights equal to zero and set the innovation covariance matrix to an identity matrix.

Case 1

$$\begin{bmatrix} x_H(\tau_L) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} -0.0064 & 0.1807 \\ 0.0500 & 0.8890 \end{bmatrix} \begin{bmatrix} x_H(\tau_L - 1) \\ x_L(\tau_L - 1) \end{bmatrix} + \bar{\varepsilon}(\tau_L) \qquad E[\bar{\varepsilon}(\tau_L)\bar{\varepsilon}(\tau_L)'] = \begin{bmatrix} 0.3405 & 0.0005 \\ 0.0005 & 0.3377 \end{bmatrix}$$

Case 2

$$\begin{bmatrix} x_H(\tau_L) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0.0582 & 0.3805 \\ 0.1095 & 0.6203 \end{bmatrix} \begin{bmatrix} x_H(\tau_L - 1) \\ x_L(\tau_L - 1) \end{bmatrix} + \bar{\varepsilon}(\tau_L) \qquad E[\bar{\varepsilon}(\tau_L)\bar{\varepsilon}(\tau_L)'] = \begin{bmatrix} 0.3437 & 0.0003 \\ 0.0003 & 0.3368 \end{bmatrix}$$

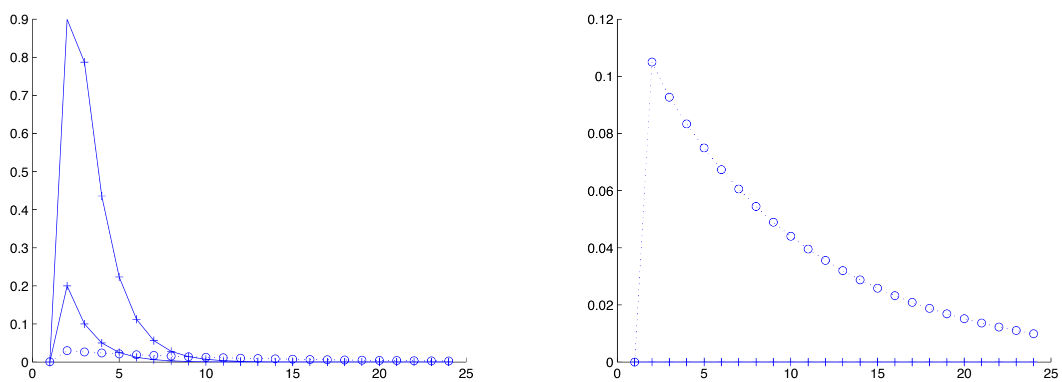
Case 3

$$\begin{bmatrix} x_H(\tau_L) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0.7020 & 0.0193 \\ -0.0330 & 0.8972 \end{bmatrix} \begin{bmatrix} x_H(\tau_L - 1) \\ x_L(\tau_L - 1) \end{bmatrix} + \bar{\varepsilon}(\tau_L) \qquad E[\bar{\varepsilon}(\tau_L)\bar{\varepsilon}(\tau_L)'] = \begin{bmatrix} 0.8350 & 0.0174 \\ 0.0174 & 1.0096 \end{bmatrix}$$

## Figure 1: Impulse Response Functions Monthly/Quarterly

The left panel shows the impulse response function of a mixed frequency VAR and a mis-specified VAR(1) using the setup in equation (2.11) with $a$ = 0, $\rho$ = $\alpha_1$ = .5, with $[w(\gamma)_3, w(\gamma)_2, w(\gamma)_1]$ = $[.2, .5, .9]$ and set the innovation covariance matrix to an identity matrix. We consider $m$ =3, corresponding to a monthly/quarterly data mixture with point-sampling of the high frequency data - which is obtained from minimizing the criterion appearing in Proposition 5.1. The left panel shows three impulse response functions: (1) the impact of a shock to $x_H(\tau_L, 1)$ on future quarterly $x_L(\tau_L + k)$, for $k$ = $1, \ldots, 24$ (quarters) as determined by the mixed frequency VAR, (2) the impact of a shock to $x_H(\tau_L, 3)$ on future quarterly $x_L(\tau_L + k)$, and (3) the impact of a shock to aggregate *high* frequency series in the VAR(1) for $\overline{x}(\tau_L)$ on future quarterly low frequency data. The right panel also shows three impulse responses, namely: (1) the impact of a shock to $x_L(\tau_L)$ on future quarterly $x_H(\tau_L + k, 1)$, for $k$ = $1, \ldots, 24$ as determined by the mixed frequency VAR, (2) the impact of a shock to $x_L(\tau_L)$ on future quarterly $x_H(\tau_L + k, 3)$, and (3) the impact of a shock to aggregate *low* frequency series in the VAR(1) for $\overline{x}(\tau_L)$ on future quarterly high frequency data.

Impulse Responses Case 1 - m = 3



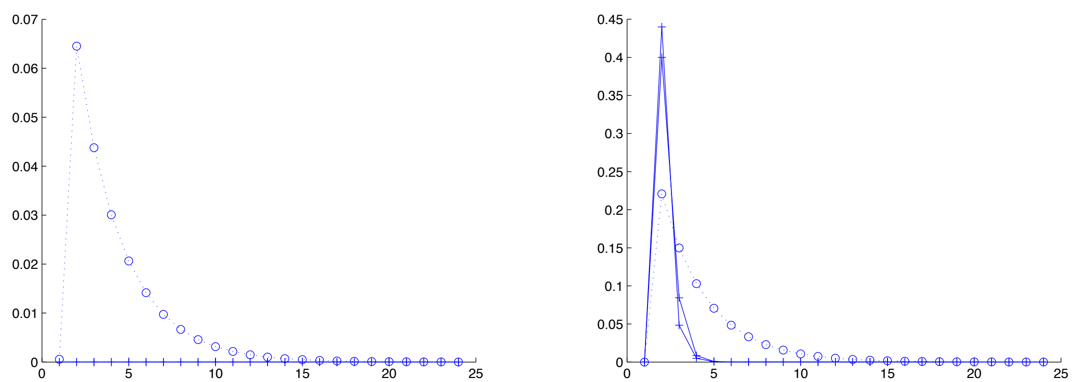

(a) High Frequency Shocks

(b) Low Frequency Shocks

Figure 2: Impulse Response Functions Monthly/Quarterly

The left panel shows the impulse response function of a mixed frequency VAR and a mis-specified VAR(1) using the setup in equation (2.11) with $a = .4$, $\rho = \alpha_1 = .1$, $[w(\gamma)_3, w(\gamma)_2, w(\gamma)_1] = [0., 0., 0.]$ and set the innovation covariance matrix to an identity matrix. We consider $m$ =3, corresponding to a monthly/quarterly data mixture with point-sampling of the high frequency data - which is obtained from minimizing the criterion appearing in Proposition 5.1. The left panel shows three impulse response functions: (1) the impact of a shock to $x_H(\tau_L, 1)$ on future quarterly $x_L(\tau_L + k)$, for $k = 1, \ldots, 24$ (quarters) as determined by the mixed frequency VAR, (2) the impact of a shock to $x_H(\tau_L, 3)$ on future quarterly $x_L(\tau_L + k)$, and (3) the impact of a shock to aggregate *high* frequency series in the VAR(1) for $\overline{x}(\tau_L)$ on future quarterly low frequency data. The right panel also shows three impulse responses, namely: (1) the impact of a shock to $x_L(\tau_L)$ on future quarterly $x_H(\tau_L + k, 1)$, for $k = 1, \ldots, 24$ as determined by the mixed frequency VAR, (2) the impact of a shock to $x_L(\tau_L)$ on future quarterly $x_H(\tau_L + k, 3)$, and (3) the impact of a shock to aggregate *low* frequency series in the VAR(1) for $\overline{x}(\tau_L)$ on future quarterly high frequency data.

Impulse Responses Case 2 - m = 3



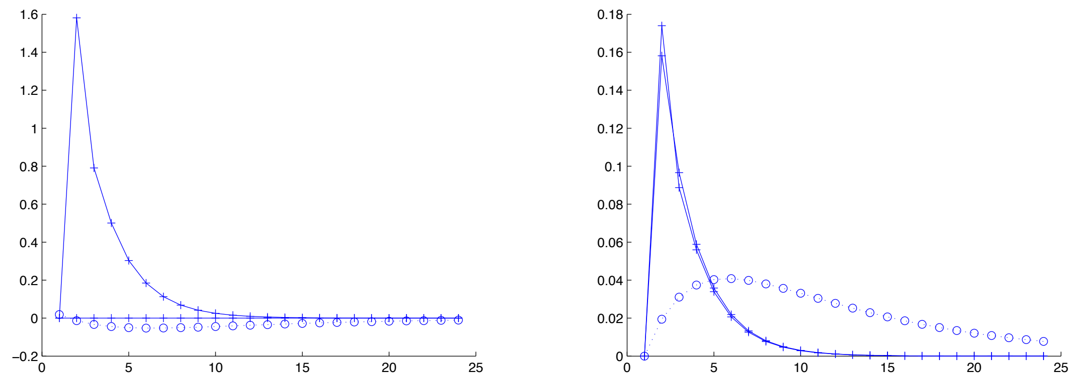

(a) High Frequency Shocks

(b) Low Frequency Shocks

## Figure 3: Impulse Response Functions Weekly/Quarterly

The left panel shows the impulse response function of a mixed frequency VAR and a mis-specified VAR(1) using the setup in equation (2.11) with $a = .05$, $\rho = .1$, $\alpha_1 = .5$, $[w(\gamma)_3, w(\gamma)_2, w(\gamma)_1] = [.5, .4, .3]$ with all other remaining 9 MIDAS polynomial weights equal to zero and set the innovation covariance matrix to an identity matrix. We consider $m$ =3, corresponding to a monthly/quarterly data mixture with point-sampling of the high frequency data - which is obtained from minimizing the criterion appearing in Proposition 5.1. The left panel shows three impulse response functions: (1) the impact of a shock to $x_H(\tau_L, 1)$ on future quarterly $x_L(\tau_L + k)$, for $k = 1, \ldots, 24$ (quarters) as determined by the mixed frequency VAR, (2) the impact of a shock to $x_H(\tau_L, 3)$ on future quarterly $x_L(\tau_L + k)$, and (3) the impact of a shock to aggregate *high* frequency series in the VAR(1) for $\overline{x}(\tau_L)$ on future quarterly low frequency data. The right panel also shows three impulse responses, namely: (1) the impact of a shock to $x_L(\tau_L)$ on future quarterly $x_H(\tau_L + k, 1)$, for $k = 1, \ldots, 24$ as determined by the mixed frequency VAR, (2) the impact of a shock to $x_L(\tau_L)$ on future quarterly $x_H(\tau_L + k, 3)$, and (3) the impact of a shock to aggregate *low* frequency series in the VAR(1) for $\overline{x}(\tau_L)$ on future quarterly high frequency data.

Impulse Responses Case 3 - m = 12




(a) High Frequency Shocks (b) Low Frequency Shocks

**Technical Appendix**

# A Parsimony - Details

We proceed with a brief discussion of univariate MIDAS regression polynomial specifications. A detailed description appears in Sinko, Sockin, and Ghysels (2010). The most commonly used parameterizations (some involving restrictions denoted by superscript $r$) are:

1. Normalized beta probability density function,

$$w_i(\gamma) = w_i(\gamma_1, \gamma_2) = \frac{x_i^{\gamma_1 - 1}(1 - x_i)^{\gamma_2 - 1}}{\sum_{i=1}^{N} x_i^{\gamma_1 - 1}(1 - x_i)^{\gamma_2 - 1}} \tag{A.1}$$

$$w_i^r(\gamma) = w_i(1, \gamma_2) \tag{A.2}$$

   where $x_i = (i-1)/(N-1)$ and one often sets the first parameter equal to one as in (A.2).

2. Normalized exponential Almon lag polynomial

$$w_i(\gamma) = w_i(\gamma_1, \gamma_2) = \frac{e^{\gamma_1 i + \gamma_2 i^2}}{\sum_{i=1}^{N} e^{\gamma_1 i + \gamma_2 i^2}} \tag{A.3}$$

$$w_i^r = w_i(\gamma_1, 0) \tag{A.4}$$

3. Almon lag polynomial specification of order $P$ (not normalized, i.e. sum of individual weights is not equal to 1).

$$\beta w_i(\gamma_0, \ldots, \gamma_P) = \sum_{p=0}^{P} \gamma_p i^p \tag{A.5}$$

   Note that this can also be written in matrix form:

$$\begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2^2 & \cdots & 2^P \\ 1 & 3 & 3^2 & \cdots & 3^P \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & N & N^2 & \cdots & N^P \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_P \end{bmatrix} \tag{A.6}$$

   Therefore the use of Almon lags in MIDAS models can be achieved via OLS estimation with properly transformed high frequency data regressors using the matrix representation appearing in the above equation. Once the weights are estimated via OLS, one can always rescale them to obtain a slope coefficient (assuming the weights do not sum up to zero).

4. Polynomial specification with step functions (not normalized)

$$\beta w_i(\gamma_1, \ldots, \gamma_P) = \gamma_1 I_{i \in [a_0, a_1]} + \sum_{p=2}^{P} \gamma_p I_{i \in (a_{p-1}, a_p]}$$
$$a_0 = 1 < a_1 < \ldots < a_P = N \tag{A.7}$$
$$I_{i \in [a_{p-1}, a_p]} = \begin{cases} 1, & a_{p-1} \le i \le a_p \\ 0, & otherwise \end{cases}$$

where $a_0 = 1 < a_1 < \ldots < a_P = N$. The step functions approach to MIDAS appeared in Ghysels, Sinko, and Valkanov (2006). A special case is a completely unrestricted specification where each weight is estimated separately. This so called U-MIDAS (unrestricted MIDAS polynomial) approach suggested by Foroni, Marcellino, and Schumacher (2011) is shown to work for small values of $m$.

A so called *multiplicative* ADL MIDAS regression specifications is also suggested in Andreou, Ghysels, and Kourtellos (2010). Taking the last equation in (2.11) we have:

$$\begin{aligned} x_L(\tau_L) &= \sum_{j=1}^{P} \alpha_j + \sum_{j=1}^{P} b_j x_H(\tau_L - j)(\gamma) + \varepsilon(\tau_L, m+1) \\ x_H(\tau_L - j)(\gamma) &\equiv \sum_{k=1}^{m} w(\gamma)_k x_H(\tau_L - j, k) \end{aligned} \tag{A.8}$$

hence, the within-$\tau_L$ period high frequency weights remain invariant and yield a low frequency parameterized process $x_H(\tau_L - j)(\gamma)$.

# B Proof of Theorem 4.1

We start be defining the covariance generating functions, based on equations (**??**):

$$\begin{aligned} \ddot{S}(z) &= \ddot{c}(0) + \sum_{k=1}^{\infty} \left[\ddot{c}(-k) z^{-k} + \ddot{c}(-k)' z^{k}\right] \\ \overline{S}(z) &= \overline{c}(0) + \sum_{k=1}^{\infty} \left[\overline{c}(-k) z^{-k} + \overline{c}(-k)' z^{k}\right] \end{aligned} \tag{B.1}$$

Since the former is derived from a periodic linear model, we can apply what is known as the Tiao and Grupe (1980) formula which yields a *high* frequency non-periodic process $\underline{\overline{\mathrm{x}}}(\tau_L, k_H)$, with spectral density equal to:

$$S_{\underline{\overline{\mathrm{x}}}}(z) = \frac{1}{m} Q(z) \ddot{S}(z^m) Q(z^{-1}) \tag{B.2}$$

where $Q(z) = \left[I zI \ldots z^{m-1} I\right]$.

The resulting process is the high frequency aperiodic representation which inherits the high frequency sampling of the periodic models. To obtain the implied low frequency process, we need to address the aliasing problem (see e.g. Hansen and Sargent (1983) and Phillips (1973)) associated with sampling this process only at low frequency. More specifically, $\overline{S}(z)$ is a skipped sampled version of $S_{\underline{\mathrm{x}}}$, and therefore:

$$\overline{S}(z) = \frac{1}{m}\sum_{j=0}^{m-1} S_{\underline{\mathrm{x}}}(\frac{z+2\pi j}{m}) \tag{B.3}$$

which can be written as:

$$\overline{S}(z) = \frac{1}{m}\sum_{j=0}^{m-1} \frac{1}{m}Q(\frac{z+2\pi j}{m})\ddot{S}((\frac{z+2\pi j}{m})^m)Q((\frac{z+2\pi j}{m})^{-1}) \tag{B.4}$$

Using the spectral density representations appearing in equations (4.1) yields:

$$\overline{S}(z) = \frac{1}{m}\sum_{j=0}^{m-1} \frac{1}{m}Q(\frac{z+2\pi j}{m})\ddot{S}((\frac{z+2\pi j}{m})^m)Q((\frac{z+2\pi j}{m})^{-1}) \tag{B.5}$$

□

# C Proof of Proposition 5.1

We start with listing the regularity conditions. We assume the DGP is the $K_L + m * K_H$ dimensional vector $\underline{\mathrm{x}}(\tau_L)$ described by equation (2.6):

$$\mathbb{A}(\mathcal{L}_L)(\underline{\mathrm{x}}(\tau_L) - \mu_{\underline{\mathrm{x}}}) = \varepsilon(\tau_L)$$

with $E[\varepsilon(\tau_L)\varepsilon(\tau_L)'] = \mathbb{C}\mathbb{C}'$. Similar to Theorem 4.1 we assume all the processes are covariance stationary and therefore have a spectral representation. In particular:

**Assumption C.1.** *The process $\ddot{x}(\tau_L)$ satisfy Assumptions 2.1 and 2.2 and has spectral density $\ddot{S}(z)$ for $z = \exp(-i\omega)$ with $\omega \in [0,\pi]$, which can be written as:*

$$\ddot{S}(z) = \ddot{\mathbb{A}}(z)^{-1}\ddot{\mathbb{C}}\ddot{\mathbb{C}}'(\ddot{\mathbb{A}}(z^{-1})^{-1})'$$

*The same process is parameterized as:*

$$\ddot{\mathbb{A}}_\Psi(\mathcal{L}_L)(\ddot{x}(\tau_L) - \mu_{\ddot{x}}^{\Psi}) = \ddot{\varepsilon}(\tau_L)$$

*with spectral density:*

$$\ddot{S}(z,\Psi) = \ddot{\mathbb{A}}_\Psi(z)^{-1}\ddot{\mathbb{C}}_\Psi\ddot{\mathbb{C}}_\Psi'(\ddot{\mathbb{A}}_\Psi(z^{-1})^{-1})'$$

*Furthermore, a simple skip-sampled or the general aggregation scheme as in (4.3) yields a low frequency VAR model with spectral density as in (4.1) parameterized by $\Phi$ :*

$$\overline{S}(z,\Phi) = \mathbb{B}_\Phi(z)^{-1}\overline{\mathbb{C}}_\Phi\overline{\mathbb{C}}_\Phi'(\mathbb{B}_\Phi(z^{-1})^{-1})'$$

Moreover, the parameter vector spaces are respectively $\Phi \in \Delta_\Phi$, $\Psi \in \Delta_\Psi$ and

**Assumption C.2.** *The parameter spaces $\Delta_\Phi$ and $\Delta_\Psi$ are compact subsets of a Eucledian space*

**Assumption C.3.** *The spectral densities* $\ddot{S}(z,\Psi)$ *and* $\overline{S}(z,\Phi)$ *for* $z = \exp(-i\omega)$ *are continuous mappings mapping respectively* $[-\pi,\pi]\times\Delta_\Psi$ *and* $[-\pi,\pi]\times\Delta_\Phi$ *into the space of positive definite Hermitian matrices such that for some* $0<\varepsilon_l<\varepsilon_u : \varepsilon_l I \leq \ddot{S}(\exp(-i\omega),\Psi) \leq \varepsilon_u I$ *and* $\varepsilon_l I \leq \overline{S}(\exp(-i\omega),\Phi) \leq \varepsilon_u I$ *for respectively each* $(\omega,\Psi)\in[-\pi,\pi]\times\Delta_\Psi$ *and* $(\omega,\Phi)\in[-\pi,\pi]\times\Delta_\Phi$, $\ddot{S}(\exp(i\omega),\Psi)$ *is the complex conjugate of* $\ddot{S}(\exp(-i\omega),\Psi)$ *and* $\overline{S}(\exp(i\omega),\Phi)$ *is the complex conjugate of* $\overline{S}(\exp(-i\omega),\Phi)$.

**Assumption C.4.** $\mu_{\overline{x}}$ *is a continuous function on the domain of* $\Delta_\Phi$. $\mu_{\ddot{x}}$ *is a continuous function on the domain of* $\Delta_\Psi$.

Under the above assumptions, Hansen and Sargent (1993) show for a generic potentially mis-specified model characterized by spectral density $G(\delta)$ involving parameter vector $\delta$ against DGP with spectral density $\ddot{S}$, that the maximum likelihood estimator minimizes the Kullback and Leibler (1951) information criterion which can be written as:

$$\begin{aligned}
E(G(\delta),\ddot{S}) &= E_1(G(\delta),\ddot{S}) + E_2(G(\delta),\ddot{S}) + E_3(G(\delta),\ddot{S}) \\
E_1(G(\delta),\ddot{S}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\log(detG(\exp(-i\omega)))d\omega \\
E_2(G(\delta),\ddot{S}) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}trace(detG(\delta,\exp(-i\omega))^{-1}\ddot{S}(\exp(-i\omega)))d\omega \\
E_3(G(\delta),\ddot{S}) &= (\mu_{\ddot{x}}-\nu(\delta))'G(1)^{-1}(\mu_{\ddot{x}}-\nu(\delta))
\end{aligned} \tag{C.6}$$

using results from Akaike (1973), Ljung (1978), White (1982) and Pötscher (1987). The results in equations (5.5) follows by substituting $G(\delta)$ with $\ddot{S}(\Psi)$ and the corresponding mean in $E_3$. For the result in equation (5.6) we need to add one operation, namely we need to handle fact that the low frequency process needs to be stacked and skip sampled. This results in the appearance of (1) the multiplicative term $m$ in the first term $E_1$, (2) the use of $Q(.)$ in $E_2$ and (3) the use of $1_m$, a $m$-dimensional vector of ones in the term $E_3$.

## Internet Appendix - Matlab Code

This program computes the spectral density of traditional low frequency VAR model, given a mixed frequency VAR(1) model:

$$\underline{\mathrm{x}}(\tau_L) = A * \underline{\mathrm{x}}(\tau_L) + \varepsilon_\tau$$

with $E[\varepsilon_\tau \varepsilon \tau'] = B$.

The stacked periodic representation is obtained via the transformation

$$\ddot{x}(\tau_L) = H * \underline{\mathrm{x}}(\tau_L)$$

The program assumes that the matrices A, B and H are in memory.

```
%
% Computing the impulse response of low frequency shocks onto
% high frequency series

%
% Computing the impulse response of first high frequency shock
%
[IMPULSE_Mixeds] = dimpulse(A,BChol',H,D,1,25);
%
% Computing the impulse response of last high frequency shock
%
[IMPULSE_Mixede] = dimpulse(A,BChol',H,D,m,25);
%
% Computing the impulse response of low frequency shock
%
[IMPULSE_Mixedl] = dimpulse(A,BChol',H,D,m+1,25);
%
% Computing the traditional VAR implied spectral density
%
I1 = eye(size(A));
i=sqrt(-1);
ngx=max(size(A));
ngy=max(size(H));
ngy2=max(size(H))./m;
%
% T is the number of coordinates for the discrete Fourier transforms
%
T=2^7;
%
%
% STY is the spectral density of the periodic mixed frequency VAR
%
STY=zeros(ngy*(T/2+1),ngy);
```

```
STy2=zeros(ngy2*(T/2+1),ngy2);
%
% STLF is the spectral density of the traditional low frequency VAR
%
STLF=zeros(T/2+1,ngy2);
%
for j=0:T/2;
  ex=exp(2*pi*i*j/T);
  exinv=exp(-2*pi*i*j/T);
  Q=zeros(ngy2,ngy);
  Qinv = Q;
 for h=1:m
   nh=int2str(h);
   Q(:,(h-1)*ngy2+1:h*ngy2)=m^(-.5)*ex^(h-1)*eye(ngy2);
   Qinv(:,(h-1)*ngy2+1:h*ngy2)=m^(-.5)*exinv^(h-1)*eye(ngy2);
 end
STY(j*ngy+1:(j+1)*ngy,:) = (H/(I1-A*ex^m))*CC*((H/(I1-A*ex^(-m))))';
STy2(j*ngy2+1:(j+1)*ngy2,:)= Q*(STY(j*ngy+1:(j+1)*ngy,:))*Qinv';
for k=1:ngy2;
   STLF(j+1,k)=STy2(j*ngy2+k,k);
   end
end
%
%  Now plot the spectral densities
%
%spec1 = real(STLF(1,1:T/2+1))');
%f=[0:T/2].*2*pi/T;
% plot(f',real(STLF(1,1:T/2+1))),title('spectr of series 1'),pause
% plot(f',real(STLF(2,1:T/2+1))),title('spectr of series 2'),pause
% size(STLF)
% STLFifft = ifft(STLF);
% size(STLFifft)
%
vecSTY = reshape(STY,1,ngy*(T/2+1)*ngy);
otherparams = [ngy ngy2 T m vecSTY];
% Need to generalize this here the -1 in the next line
param0 = [.1 .1 .1 .1 1. 2. 3.];
f = @(param0)klicobjspeczero(param0,otherparams);
options.MaxFunEvals = 2000;
[param_stV] = fminunc(f,param0,options);
param0 = param_stV;
f = @(param0)klicobjspec(param0,otherparams);
[param_stV] = fminunc(f,param0,options);
ngy22 = ngy2*ngy2;
AstV = reshape(param_stV(1:ngy22),ngy2,ngy2);
BstV = tril([param_stV(1,ngy22+1) param_stV(1,ngy22+2); ...
    param_stV(1,ngy22+2) param_stV(1,ngy22+3)])*...
   triu([param_stV(1,ngy22+1) param_stV(1,ngy22+2); ...
```

```
    param_stV(1,ngy22+2) param_stV(1,ngy22+3)]) ;
BstV = (chol(BstV))';
CstV = eye(ngy2);
DstV = zeros(ngy2,ngy2);
%
%
% Computing the impulse response traditional VAR to aggregated HF series
%
[IMPULSE_MixedstVHF] = dimpulse(AstV,BstV,CstV,DstV,1,25);
%
% Computing the impulse response traditional VAR to aggregated LF series
%
[IMPULSE_MixedstVLF] = dimpulse(AstV,BstV,CstV,DstV,2,25);
%
%plot(IMPULSE_Mixeds(:,4),'+', IMPULSE_Mixede(:,4),'o',...,
%    IMPULSE_MixedstVHF(:,2),'d')
hold on;
%plot(IMPULSE_Mixeds(2:25,4),'+-')
%plot(IMPULSE_Mixede(2:25,4),'+-')
%plot(IMPULSE_MixedstVHF(2:25,2),'o:')
%
plot(IMPULSE_Mixedl(2:25,1), '+-')
plot(IMPULSE_Mixedl(2:25,3), '+-')
plot(IMPULSE_MixedstVLF(2:25,1),'o:')
hold off
```

The next function provides the parameters of a traditional low frequency VAR(1) model that minimizes the Kullblack-Leibler distance (in terms of spectral representation) with respect to the mixed frequency VAR. It is assumed that the following is in memory:

- STY is the spectral density of the periodic mixed frequency VAR
- m the number of high frequency series
- the parameters and structure of the mixed frequency VAR

We impose unconditional mean zero.

```
function fun = klicobjspec(param0,otherparams)
ngy = otherparams(1,1);
ngy2 = otherparams(1,2);
T  = otherparams(1,3);
m = otherparams(1,4);
STY = reshape(otherparams(1,5:end),ngy*(T/2+1),ngy);
ngy22 = ngy2*ngy2;
AstV = real(reshape(param0(1:ngy22),ngy2,ngy2));
% Need to generalize this
```

```
BstV =  real(tril([param0(1,ngy22+1) param0(1,ngy22+2); ...
    param0(1,ngy22+2) param0(1,ngy22+3)])*...
   triu([param0(1,ngy22+1) param0(1,ngy22+2); ...
    param0(1,ngy22+2) param0(1,ngy22+3)])) ;
E2 = 0.;
E3 = 0.;
 fun = realmax;
SstV = zeros(T/2+1,ngy2);
I1 = eye(ngy2);
for j=0:T/2;
  ex=exp(2*pi*i*j/T);
  exinv=exp(-2*pi*i*j/T);
  Q=zeros(ngy2,ngy);
  Qinv = Q;
 for h=1:m
   nh=int2str(h);
   Q(:,(h-1)*ngy2+1:h*ngy2)=m^(-.5)*ex^(h-1)*eye(ngy2);
   Qinv(:,(h-1)*ngy2+1:h*ngy2)=m^(-.5)*exinv^(h-1)*eye(ngy2);
 end
SstV(j*ngy2+1:(j+1)*ngy2,:) = ...
    (I1/(I1-AstV*ex^m))*BstV*((I1/(I1-AstV*ex^(-m))))';
E2 = E2 + log(det(SstV(j*ngy2+1:(j+1)*ngy2,:)));
E3 = E3 + trace((Q*(STY(j*ngy+1:(j+1)*ngy,:))*Qinv')/...
    SstV(j*ngy2+1:(j+1)*ngy2,:));
end;
fun = real(E2 + E3);
```

```
function fun = klicobjspeczero(param0,otherparams)
% This function provides a first guess of the parameters by focusing on the
% zero frequency before running the klicobjspec
%
ngy = otherparams(1,1);
ngy2 = otherparams(1,2);
T  = otherparams(1,3);
m = otherparams(1,4);
STY = reshape(otherparams(1,5:end),ngy*(T/2+1),ngy);
ngy22 = ngy2*ngy2;
AstV = real(reshape(param0(1:ngy22),ngy2,ngy2));
% Need to generalize this
BstV =  real(tril([param0(1,ngy22+1) param0(1,ngy22+2); ...
    param0(1,ngy22+2) param0(1,ngy22+3)])*...
   triu([param0(1,ngy22+1) param0(1,ngy22+2); ...
    param0(1,ngy22+2) param0(1,ngy22+3)])) ;
E2 = 0.;
E3 = 0.;
 fun = realmax;
SstV = zeros(ngy2,ngy2);
```

```
I1 = eye(ngy2);
  ex=1.;
  exinv=1.;
  Q=zeros(ngy2,ngy);
  Qinv = Q;
 for h=1:m
   nh=int2str(h);
   Q(:,(h-1)*ngy2+1:h*ngy2)=m^(-.5)*ex^(h-1)*eye(ngy2);
   Qinv(:,(h-1)*ngy2+1:h*ngy2)=m^(-.5)*exinv^(h-1)*eye(ngy2);
 end
SstV = (I1/(I1-AstV*ex^m))*BstV*((I1/(I1-AstV*ex^(-m))))';
E2 = E2 + log(det(SstV));
E3 = E3 + trace((Q*(STY(1:ngy,:))*Qinv')/SstV);
fun = real(E2 + E3);
```

The next programss sets up the structures for Case 1 covered in section 6. All other cases can be derived similarly.

```
a = 0;
rho = .5;
gamma = .5;
delta1 = .9;
delta2 = .5;
delta3 = .2;
A = [0 0 rho a ;  0 0 rho^2 a*(1 + rho);...
     0 0 rho^3 a*(1 + rho + rho^2);
   delta3 delta2 delta1 gamma];
% Input of the parameters for the B matrix
%
B = eye(4);
%
% m = the number of high frequency periods
%
m = 3;
%
BChol = chol(B);
CC=B;
%
% m = the number of high frequency periods
%
% Design of the H matrix
%
H = [1 0 0 0; 0 0 0 1; 0 1 0 0;...
    0 0 0 1;0 0 1 0; 0 0 0 1];
%
D = 0.*H;
```