Macroeconomics and the reality of mixed frequency data

Eric Ghysels [*],[1]

Department of Economics, University of North Carolina, Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, USA
Department of Finance, Kenan-Flagler Business School, USA
CEPR, USA

ARTICLE INFO

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 to 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 are parameter-driven 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 as well as shocks 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. 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. 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.

© 2016 Elsevier B.V. All rights reserved.

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) a specification that does not involve latent shocks, (2) a specification that allows us to measure the impact of high frequency data onto low frequency ones and vice versa, (3) as far as VAR models go parsimonious, (4) a specification that can be estimated and analyzed with standard VAR analysis tools – such as impulse response analysis, and can be estimated

---

* Correspondence to: Department of Economics, University of North Carolina, Gardner Hall CB 3305, Chapel Hill, NC 27599-3305, USA.
  E-mail address: eghysels@unc.edu.
[1] 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.

with standard VAR estimation procedure (5) 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.

The mixed frequency VAR provides an alternative to commonly used state space models involving mixed frequency data.[2] 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 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 et al. (2006), Ghysels and Wright (2009), Andreou et al. (2010) and Chen and Ghysels (2011), among others.

We also characterize the mapping between the mixed frequency VAR model and (1) a traditional VAR model where all the data are sampled at a common low frequency as well as (2) a hidden state high frequency VAR commonly used in a state space model setting. This mapping allows us to study the mis-specification of impulse response functions of traditional VAR models.[3]

Finally, 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] See for example, Harvey and Pierse (1984), Bernanke et al. (1997), Zadrozny (1990), Mariano and Murasawa (2003), Mittnik and Zadrozny (2004), and more recently Aruoba et al. (2009), Ghysels and Wright (2009), Kuzin et al. (2009), Marcellino and Schumacher (2010), Foroni and Marcellino (2014) and Schorfheide and Song (2013), among others.

[3] 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. For the sake of keeping the paper concise we skip the details. They appear in a web version of the paper.

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 (random mixed frequency sampling is not considered). For example quarterly/annual, monthly/quarterly, hourly/daily 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 τ_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 et al. (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 et al. (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 τ_L series appear at the end of the (low frequency) period. Namely, consider the following finite order VAR representation of a stacked vector:

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

which is $K_L + m * K_H$ dimensional VAR model with P lags. 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:

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

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:

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

Hence, elements below the diagonal pertain to downstream impacts, i.e. high frequency data affect subsequent within-τ_L period observations. This will be relevant notably for intra-τ_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 Eq. (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.

Note that VAR models with latent shocks yield non-Markovian processes in observables, or by implication infinite order VAR models, in contrast to the stacked system appearing in (2.1) which is by definition a finite order VAR. Nevertheless, 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 system then reads:

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

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

which is the ADL MIDAS regression model discussed in Andreou et al. (2010). There are various parsimonious parameterizations

suggested for such regressions, see e.g. Ghysels et al. (2006), Andreou et al. (2010) and Sinko et al. (2010), which 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 $\varepsilon(\tau_L)^{i,1}$ ($i = 1, \ldots, m$ again in this particular example) onto future high and low frequency series.

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

$$
\mathbb{A}(\mathcal{L}_L)(\underline{x}(\tau_L) - \mu_{\underline{x}}) = \underline{\varepsilon}(\tau_L) \tag{2.5}
$$

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

$$
\mathbb{A}(\mathcal{L}_L) = I - \sum_{j=1}^{P} A_j \mathcal{L}_L^j
$$

$$
\mu_{\underline{x}} = \left(I - \sum_{j=1}^{P} A_j \right)^{-1} A_0 \tag{2.6}
$$

where we assume that the VAR is covariance stationary to be able to write the above equations (see Assumption 4.1 below) and we let $E[\underline{\varepsilon}(\tau_L)\underline{\varepsilon}(\tau_L)'] = \mathbb{C}\mathbb{C}'$.

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 second sub-vector of low frequency observations is left unspecified for the moment—i.e. we are not going to be explicit until Section 4 about how the high frequency data aggregate to low frequency observations. We will denote the low frequency VAR model as:

$$
\mathbb{B}(\mathcal{L}_L)(\overline{x}(\tau_L) - \mu_{\overline{x}}) = \overline{\varepsilon}(\tau_L) \tag{2.7}
$$

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. 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.

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 of 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 the aforementioned economic time structural dynamic analysis.

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:

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

$$\begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ x_L(\tau_L) \\ \vdots \\ x_H(\tau_L, m) \end{bmatrix} \tag{2.8}$$

where the release of low frequency appears at the end, beginning or some time in the middle of τ_L. The order of appearance in the vector therefore determines the timing of intra-τ_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-τ_L period shocks occur. Moreover, the high frequency releases of low frequency data can occur at various k_H throughout period τ_L and therefore impact the structure of shocks and responses. More specifically, K_L low frequency series $K_L^{k_H}$ in $x_L(\tau_L)$ are released at time k_H in period τ_L for $k_H = 1, \ldots, m$, with $\sum_{i=1}^{m} K_L^i = K_L$.[4] When we need to keep track of the high frequency releases of low frequency data we use $x_L(\tau_L, k_H)$, for the subvector 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 τ_L then $\underline{x}(\tau_L) \equiv (x_H(\tau_L, 1)', \ldots, x_H(\tau_L, m)', x_L(\tau_L)')'$, otherwise it contains $(x_L(\tau_L, k_H)', x_H(\tau_L, k_H)')'$ for the sequence $k_H = 1, \ldots, m$.[5]

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

Sections 2.5 and 3.1, we will deal more explicitly with the specific order of the elements in the stacked vector.

2.3. 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 it 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. 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. The sparsely specified VAR models introduced here will also serve another purpose. In Section 5.2 we cover Bayesian estimation of mixed frequency VAR models. The priors for the Bayesian VAR models will be inspired by the examples we present next.

For the purpose of streamlining the exposition we will start again with an example where all the low frequency τ_L series appear at the end of the stacked vector as in Eq. (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.[6] Therefore, we rewrite Eq. (2.1) as:

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

$$\times \begin{bmatrix} x_H(\tau_L - j, 1) \\ \vdots \\ x_H(\tau_L - j, m) \\ x_L(\tau_L - j) \end{bmatrix} + \underline{\varepsilon}(\tau_L)$$

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:

$$\begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ x_H(\tau_L, m) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0 & \cdots & \rho & a \\ \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & \rho^m & a\left(1 + \sum_{j=0}^{m-1} \rho^j\right) \\ w(\gamma)_m & \cdots & w(\gamma)_1 & \alpha_1 \end{bmatrix}$$

$$\times \begin{bmatrix} x_H(\tau_L - 1, 1) \\ \vdots \\ x_H(\tau_L - 1, m) \\ x_L(\tau_L - 1) \end{bmatrix}$$

---

[4] 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—technical details are omitted here.

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

[6] In Section 5.2 we will cover the cases with K_L low frequency and K_H high frequency series.

$$+ \sum_{j=2}^{P} \begin{bmatrix} 0 & \dots & 0 & 0 \\ \vdots & \dots & \vdots & \vdots \\ 0 & \dots & 0 & 0 \\ w(\gamma)_{jm} & \dots & w(\gamma)_{(j-1)m+1} & \alpha_j \end{bmatrix}$$

$$\times \begin{bmatrix} x_H(\tau_L - j, 1) \\ \vdots \\ x_H(\tau_L - j, m) \\ x_L(\tau_L - j) \end{bmatrix} + \underline{\varepsilon}(\tau_L) \qquad (2.9)$$

which involves P parameters α_j, two parameters ρ and a and a low dimensional MIDAS polynomial parameter vector γ. When all $\alpha_j = 0$ for $j > 1$, and the dimension of γ 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 (2.9), the innovation covariance matrix may also be sparsely specified. Continuing with the above specification, we can write $E[\underline{\varepsilon}(\tau_L)\underline{\varepsilon}(\tau_L)']$ as in Box I, adding another three parameters, and therefore a total of eight again independent of m.[7]

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

$$[A_1^{m+1,1} \dots A_1^{m+1,m} A_2^{m+1,1} \dots A_P^{m+1,m}]$$

$$= \left[\sum_{i=1}^{K_H \times m \times P} (w(\gamma)_i) \right] \otimes B \qquad (2.11)$$

with B a $K_L \times K_H$ matrix and a *scalar* MIDAS polynomial with $\sum_{i=1}^{K_H \times m \times P} (w(\gamma)_i) = 1$.[8] 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 $K_L \times K_H$ plus the dimension of γ which is two 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 Eq. (A.8) in Appendix A):

$$[A_j^{m+1,1} \dots A_j^{m+1,m}] = \left[\sum_{i=1}^{K_H \times m} (w(\gamma)_i) \right] \otimes B_j \quad j = 1, \dots, P \quad (2.12)$$

meaning that within-τ_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-τ_L period high frequency observation. Bai et al. (2013) 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.

The use of MIDAS polynomial permits to reduce the number of parameters to be estimated, and therefore achieves a necessary degree of parsimony that is often critical in empirical applications. However, parsimony achieved through zero restrictions and MIDAS polynomials clearly impose a specific structure on the data, and therefore one looses the rich dynamics that is allowed for by standard (unrestricted) VAR models. This may be especially problematic in the context of structural studies. A more appealing way to proceed in this case is to rely on Bayesian methods to handle parsimony—as discussed in Section 5.2. Foroni and Marcellino (2014) and Schorfheide and Song (2013) are some recent examples of identifying and estimating DSGE models using mixed frequency data. While they formulate VAR models imbedded in a state space models which feature latent shocks, the use of Bayesian and structural parameter-restricted Kalman filter estimation methods readily apply to the mixed frequency VAR models discussed in this paper. Formulating Bayesian priors for stacked mixed frequency VAR models is a particular case of the priors discussed later in the paper.

We adopt in the remainder of the paper a generic setting where all the parameters are collected into a vector Ψ. 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 Eqs. (2.5) as:

$$\mathbb{A}_\Psi(\mathcal{L}_L)(\underline{x}(\tau_L) - \mu_{\underline{x}}^\Psi) = \underline{\varepsilon}(\tau_L) \qquad (2.13)$$

with $E[\underline{\varepsilon}(\tau_L)\underline{\varepsilon}(\tau_L)'] = \mathbb{C}(\Psi)\mathbb{C}(\Psi)'$.

To streamline the notation, we will drop the parameter vector Ψ 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.

2.4. State space models revisited

It was noted in the Introduction that state space models are often used to handle mixed frequency data. Building on the early work of Harvey and Pierse (1984), Bernanke et al. (1997) and Zadrozny (1990), recent papers have further explored their use, see Aruoba et al. (2009), Ghysels and Wright (2009), Kuzin et al. (2009), Marcellino and Schumacher (2010), Foroni and Marcellino (2014) and Schorfheide and Song (2013), among others. They are a solution to the parsimony issue discussed in the previous subsection. To set the scene, we consider a state space model for the stacked vector we have been considering so far:

$$\begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ x_H(\tau_L, m) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} G_{11} & 0 & \dots & 0 \\ 0 & G_{22} & \dots & 0 \\ \vdots & & & \vdots \\ 0 & & \dots & G_{mm} \\ 0 & & \dots & G_L \end{bmatrix} \begin{bmatrix} s(\tau_L, 1) \\ \vdots \\ s(\tau_L, m) \end{bmatrix}$$

$$+ \underline{\varepsilon}(\tau_L)^s \qquad (2.14)$$

where $s(\tau, i)$ is a high frequency state process. To streamline the presentation, without loss of generality, let us proceed with the special case considered before of $K_L = K_H = 1$. In addition, we also assume that the latent state process is univariate. Hence, we consider a single factor process which drives respectively one low

---

[7] 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 manageable parameter structure independent of m.

[8] It is worth noting here that the multiplicative MIDAS specification is known to suffer from the well known Davies (1987) problem.

$$
\begin{bmatrix}
\sigma_{HH}^2 & \rho\sigma_{HH}^2 & \rho^2\sigma_{HH}^2 & \rho^3\sigma_{HH}^2 & \sigma_{HL} \\
(1+\rho^2)\sigma_{HH}^2 & \rho(1+\rho^2)\sigma_{HH}^2 & \rho^2(1+\rho^2)\sigma_{HH}^2 & (1+\rho)\sigma_{HL} \\
(1+\rho^2+\rho^4)\sigma_{HH}^2 & \rho(1+\rho^2+\rho^4)\sigma_{HH}^2 & (1+\rho+\rho^2)\sigma_{HL} \\
(1+\rho^2+\rho^4+\rho^6)\sigma_{HH}^2 & (1+\rho+\rho^2+\rho^3)\sigma_{HL} \\
\sigma_{LL}
\end{bmatrix}
\tag{2.10}
$$

Box I.

and one high frequency process. Moreover, we also assume that the state process is a mean zero $AR(1)$ process with persistence parameter ρ_s and innovation variance σ_s^2. Then following Bai et al. (2013, equation (2.21)), one can write the (low frequency) one-step ahead prediction conditional on past $\underline{x}(\tau_L - j), j = 0, 1, \ldots$ can be written as:

$$
x_L(\tau_L + 1) = E[x_L(\tau_L + 1)|\underline{x}(\tau_L - j)]
$$

$$
= \sum_{j=0}^{\infty} \vartheta^j
\begin{bmatrix}
\kappa_1 \prod_{k=2}^{m}(\rho_s - \rho_s\kappa_k) \\
\kappa_2 \prod_{k=3}^{m}(\rho_s - \rho_s\kappa_k) \\
\vdots \\
\rho_s^m \kappa_{m,2} \rho_s^{mj} \\
\rho_s^m \kappa_{m,1} \rho_s^{mj}
\end{bmatrix}'
\begin{bmatrix}
x_H(\tau_L - j, 1) \\
\vdots \\
x_H(\tau_L - j, m) \\
x_L(\tau_L - j)
\end{bmatrix}
\tag{2.15}
$$

where κ_k and $\kappa_{m,k}$ are the periodic Kalman gains defined in Bai et al. (2013) (above equation (2.21)). The above equation pertains to one-step ahead prediction conditional on past $\underline{x}(\tau_L - j), j$ of the low frequency process. Similar derivations apply to the high frequency process as well. Taken together, this implies, not surprisingly that the latent factor Markov order one state space model leads to an infinite order stacked VAR model. Note that the VAR is indeed tightly parameterized, but of infinite order. The discussion highlights the differences between traditional VAR and state space models in a mixed sampling frequency setting. As emphasized earlier, the innovations to the infinite VAR do not have the same (structural) interpretation as do those in the stacked VAR model.

2.5. Shocks and Cholesky 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 vector $\underline{x}(\tau_L)$ has a natural order for the intra-τ_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 $\underline{x}(\tau_L)$.

The stacked mixed frequency VAR model implies an impulse response function:

$$
(\underline{x}(\tau_L) - \mu_{\underline{x}}) = \left(I - \sum_{j=1}^{P} A_j \mathcal{L}_L^j \right)^{-1} \underline{\varepsilon}(\tau_L)
$$

$$
= \sum_{j=0}^{\infty} F_j \varepsilon(\tau_L - j) \equiv F(\mathcal{L}_L)\underline{\varepsilon}(\tau_L)
\tag{2.16}
$$

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

Therefore the impulse responses potentially tell us something about 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 Eq. (2.7), namely: $(\bar{x}(\tau_L) - \mu_{\bar{x}}) = (\mathbb{B}(\mathcal{L}_L))^{-1}\bar{\varepsilon}(\tau_L)$ we can see how intra-τ_L period shocks are scrambled — something we will be more explicit about in Section 4.

Since the order of the entries into the stacked vector is no long arbitrary, it is also the case that the Cholesky factorization of the innovations is no longer arbitrary. In particular consider:

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

where Ω 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:

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

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 Eq. (2.10) we realize that the parameters governing the covariance matrix $E[\varepsilon(\tau_L)\varepsilon(\tau_L)']$ and thus its Cholesky factorization, are tied to the parameters governing the VAR dynamics, in particular the parameter ρ in Eq. (2.9). 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 Eq. (2.2). We will revisit the connection with traditional structural VAR models in the next subsections. To summarize: while Cholesky 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 Eq. (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-τ_L updates of forecasts. It turns out this will be an example of using certain types of structural VAR matrices to update within-τ_L information.

Continuing with the example in Eqs. (2.5) and (2.18) consider the following transformations for $i = 1, \ldots, m - 1$: $\mathcal{N}_{[i]}\mathbb{A}(\mathcal{L}_L)(\underline{x}(\tau_L) - \mu_{\underline{x}}) = \mathcal{N}_{[i]}\varepsilon(\tau_L)$ or:

$$\mathcal{N}_{[i]}\underline{x}(\tau_L) = \mathcal{N}_{[i]}A_0 + \sum_{j=1}^{P}\mathcal{N}_{[i]}A_j\underline{x}(\tau_L - j) + \mathcal{N}_{[i]}\varepsilon(\tau_L) \tag{3.1}$$

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

$$\mathcal{N}_{[i]} = \begin{bmatrix} I & 0 & \cdots & \cdots & 0 & 0 \\ \mathcal{N}_{[i]}^{2,1} & I & 0 & \cdots & 0 & 0 \\ \vdots & & & & \vdots & \vdots \\ \mathcal{N}_{[i]}^{i+1,1} & \cdots & \mathcal{N}_{[i]}^{i+1,i} & I & \vdots & 0 \\ \vdots & & \vdots & 0 & \ddots & \vdots \\ \mathcal{N}_{[i]}^{m+1,1} & \cdots & \mathcal{N}_{[i]}^{m+1,i} & 0 & \cdots & I \end{bmatrix} \tag{3.2}$$

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 $K_L \times K_H$. These matrices are related to the inverse of the Cholesky lower triangular decomposition, namely recall from Eq. (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 (3.2), 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:

$$x_L(\tau_L) + \mathcal{N}_{[1]}^{m+1,1}x_H((\tau)_L, 1) = A_0^{m+1,1} + \sum_{j=1}^{P}A_j^{m+1,m+1}x_L(\tau_L - j)$$

$$+ \sum_{j=1}^{P}\sum_{k=1}^{m}A_j^{m+1,k}x_H(\tau_L - j, k) + \varepsilon(\tau_L)^{m+1,1} \tag{3.3}$$

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

$$x_L(\tau_L) = (A_0^{m+1,1} - \mathcal{N}_{[1]}^{m+1,1}A_0^{1,1}) - \mathcal{N}_{[1]}^{m+1,1}\underline{\varepsilon}(\tau_L)^{1,1}$$

$$+ \sum_{j=1}^{P}(A_j^{m+1,m+1} - \mathcal{N}_{[1]}^{m+1,1}A_j^{1,1})x_L(\tau_L - j)$$

$$+ \sum_{j=1}^{P}\sum_{k=1}^{m}(A_j^{m+1,k} - \mathcal{N}_{[1]}^{m+1,1}A_j^{1,k})x_H(\tau_L - j, k)$$

$$+ \underline{\varepsilon}(\tau_L)^{m+1,1}. \tag{3.4}$$

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 Cholesky 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 Eq. (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 particular, to simplify the presentation, let us only focus on FFR in combination with some low frequency series:

$$A_c\begin{bmatrix} FFR(\tau_L, 1) \\ \vdots \\ FFR(\tau_L, k) \\ \vdots \\ FFR(\tau_L, m) \\ x_L(\tau_L) \end{bmatrix} = A_0 + \sum_{j=1}^{P}A_j\begin{bmatrix} FFR(\tau_L - j, 1) \\ \vdots \\ FFR(\tau_L - j, k) \\ \vdots \\ FFR(\tau_L - j, m) \\ x_L(\tau_L - j) \end{bmatrix} + \underline{\varepsilon}(\tau_L) \tag{3.5}$$

with

$$A_c = \begin{bmatrix} I_{K_H} & \cdots & \cdots & \cdots & \cdots & A_c^{1,m} & A_c^{1,m+1} \\ \vdots & \ddots & & & \vdots & \vdots \\ \vdots & & \ddots & & \vdots & \vdots \\ A_c^{k,1} & \cdots & A_c^{k,k-1} & I_{K_H} & & A_c^{k,m} & A_c^{k,m+1} \\ \vdots & \vdots & & & \ddots & \vdots & \vdots \\ A_c^{m,1} & \cdots & & & \cdots & I_{K_H} & A_c^{m,m+1} \\ A_c^{m+1,1} & \cdots & & & \cdots & A_c^{m+1,m} & I_{K_L} \end{bmatrix}. \tag{3.6}$$

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 $k < j \le m$. Moreover, we leave unspecified the regressors appearing on the right hand side of the above equation, and therefore we have:

$$FFR(\tau_L, k) = A_0^{k,1} - A_c^{k,k-1}FFR(\tau_L, k - 1) - A_c^{k,m+1}x_L(\tau_L)$$

$$+ \text{regressors prior to } \tau_L + \underline{\varepsilon}(\tau_L)^{k,1}. \tag{3.7}$$

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 variable may respond to current (low frequency) conditions—although $x_L(\tau_L)$ may not yet be observed at period k of τ_L. This obviously raises 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 τ_L. Therefore, one may interpret Eq. (3.7) as:

$$FFR(\tau_L, k) = A_0^{k,1} - A_c^{k,k-1}FFR(\tau_L, k - 1)$$

$$- A_c^{k,m+1}E[x_L(\tau_L)|I(\tau_L, k)] + \cdots \tag{3.8}$$

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 $\mathcal{N}_{[k]}\underline{x}(\tau_L)$ pertaining to the concurrent estimates of $x_L(\tau_L)$. Recall that in Eq. (2.10) we noted that parameters

governing the covariance matrix $E[\underline{\varepsilon}(\tau_L)\underline{\varepsilon}(\tau_L)']$ and thus its Cholesky 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 (3.7), 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 Eq. (3.7) using ex post realizations of low frequency series to analyze the real-time policy decision rules.

4. High, mixed and low frequency VAR models

We noted in the Introduction that mixed frequency typically involve latent processes. In this section we discuss the relationships between VAR models involving exclusively high frequency data – and therefore latent processes – and respectively mixed and low frequency processes. We adopt a high frequency VAR setting similar to Zadrozny (1990) which was recently generalized by Chiu et al. (2011). We use a slightly more general setting in fact, as will become clear shortly. It should also be noted that we could also start from a hidden state space model with a common factor structure, as in Mariano and Murasawa (2003), Aruoba et al. (2009), among many others.

Throughout the section we will work with fixed mixed sampling frequencies, as stated in Assumption 2.1. Recall from Assumption 2.1 that we started with a K-dimensional process with the first $K_L < K$ elements, collected in the vector process $x_L(\tau_L)$, observed every m fixed periods while the remaining $K_H = K - K_L$ series are observed as $x_H(\tau_L, k_H)$. To build a high frequency process, we consider a K-dimensional partially latent high frequency process $z_H(\tau_L, k_H)$. Moreover, we will also consider the $m * K$-dimensional stacked version of the vector, which we will denote $\ddot{x}(\tau_L) \equiv [z_H(\tau_L, 1)' \ldots z_H(\tau_L, m)']'$, and characterize it via the linear aggregation scheme:

$$\begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ x_H(\tau_L, m) \\ x_L(\tau_L) \end{bmatrix} = \underline{D}(\mathcal{L}_H) \begin{bmatrix} z_H(\tau_L, 1) \\ \vdots \\ z_H(\tau_L, m) \end{bmatrix} \quad (4.1)$$

such as for example:

$$\underline{D}(\mathcal{L}_H) \equiv diag \left(\begin{bmatrix} 0_{K_H \times K_L} & I_{K_H} \end{bmatrix}, \ldots, \begin{bmatrix} 0_{K_H \times K_L} & I_{K_H} \end{bmatrix}, \right.$$
$$\left. \begin{bmatrix} \underline{D}_L(\mathcal{L}_H) & 0_{K_L \times K_H} \end{bmatrix} \right)$$

in which case the high frequency series $x_H(\tau_L, k_H)$ are exact copies of the $K_L + 1, \ldots, K$ elements of $Z_H(\tau_L, k_H)$, whereas the low frequency series are obtained via a linear aggregation scheme, where the aggregation scheme may involve long spans, i.e. that may involve lags larger than m. Hence, the low frequency vector may pertain to $\tau_L - 1$ realizations of z_H.[9]

Note that the above specification relates to what we referred to in Section 2.2 as economic time. When real-time series are the constituents for the stacked vector, one can also construct a relation similar to Eq. (4.1) with a more general structure for $\underline{D}(\mathcal{L}_H)$. Namely, one can think of the mixed frequency VAR model in terms of filtered z_H, which may represent a combination of flow and stock variables and may capture a mixture of releases involving

publication delays. Therefore, we will in general setting replace the scheme in Eq. (4.1) by the generic scheme:

$$\underline{x}(\tau_L) = \underline{D}(\mathcal{L}_H)\ddot{x}(\tau_L) = \sum_{j=0}^{P_a} \sum_{k_H=1}^{m} \underline{D}_{j*m+k_H}(\mathcal{L}_H^{j*m+k_H})\ddot{x}(\tau_L). \quad (4.2)$$

By the same token, we also can characterize the mapping from $\ddot{x}(\tau_L)$ and the aggregated process $\bar{x}(\tau_L)$ via:

$$\bar{x}(\tau_L) = \overline{D}(\mathcal{L}_H)\ddot{x}(\tau_L) \quad (4.3)$$

as well as:

$$\bar{x}(\tau_L) = \overline{F}(\mathcal{L}_H)\underline{x}(\tau_L). \quad (4.4)$$

Building on the example appearing in (4.1), the latter would be:

$$\begin{bmatrix} x_H(\tau_L) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0_{K_H} & \ldots & \overline{F}_H(\mathcal{L}_H) & 0_{K_L} \\ 0_{K_H} & \ldots & 0_{K_H} & I_{K_L} \end{bmatrix} \begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ x_H(\tau_L, m) \\ x_L(\tau_L) \end{bmatrix}. \quad (4.5)$$

Next, we need to specify the data generating process for the high frequency process. It will convenient to write the process in a stacked version as we did in the above Eqs. (4.1) and (4.2). In particular, we assume that:

Assumption 4.1. The vector $\ddot{x}(\tau_L) \equiv (z_H(\tau_L, 1)', \ldots, \ldots, z_H(\tau_L, m)')'$ is of dimension $m * K$ and has a finite order covariance stationary VAR representation:

$$\ddot{\mathbb{A}}(\mathcal{L}_L)(\ddot{x}(\tau_L) - \mu_{\ddot{x}}) = \ddot{\varepsilon}(\tau_L) \quad (4.6)$$

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

The above finite order VAR assumption for $\ddot{x}(\tau_L)$ requires some discussion. In particular, the finite order VAR structure for the stacked mixed frequency vector $\underline{x}(\tau_L)$ in Eq. (2.1) combined with the aggregation scheme in (4.1) may not be compatible with Assumption 4.1.

In addition, the DGP in Assumption 4.1 includes a class of periodic models studied in Gladyshev (1961) and in particular Tiao and Grupe (1980), Hansen and Sargent (1990, Chap. 17), among others noted earlier. For example, consider the zero mean periodic VAR(1) model:

$$\begin{bmatrix} I_K & \ldots & \ldots & \ldots & 0_K \\ -A_c^{2,1} & I_K & \ldots & \ldots & 0_K \\ 0_K & \ddots & \ldots & \ldots & 0_K \\ \vdots & & \ddots & \ddots & \vdots \\ 0_K & \ldots & \ldots & -A_c^{m,m-1} & I_K \end{bmatrix} \ddot{x}(\tau_L)$$
$$= \begin{bmatrix} 0_K & \ldots & \ldots & \ldots & A^{1,m} \\ 0_K & & \ldots & \ldots & 0_K \\ \vdots & \ddots & \ldots & \ldots & 0_K \\ \vdots & & \ldots & & \vdots \\ 0_K & \ldots & \ldots & 0_K & 0_K \end{bmatrix} \ddot{x}(\tau_L - 1) + \ddot{\varepsilon}(\tau_L). \quad (4.7)$$

We emphasize the inclusion of periodic models, in part because such models imply that impulse response functions vary with m, as stressed in a seasonal context by Hansen and Sargent (1990, Chap. 17). If in the above periodic VAR(1) model all the matrices $A_c^{i,i-1}$, for $i = 2, \ldots, m$, and $A^{1,m}$ are equal then we have a regular high frequency (non-periodic) VAR(1) model notably used by Chiu et al. (2011). Note that in such a case, all impulse response functions will be the same across time.

---

[9] For more discussion of general linear aggregation schemes, see e.g. Lütkepohl (1987). Note that stock and flow sampling schemes are of course special cases.

We are interested in the relationship between the stacked process $\ddot{x}(\tau_L)$ and respectively the mixed frequency process $\underline{x}(\tau_L)$ and the aggregated process $\bar{x}(\tau_L)$. Since we assume that all the processes are covariance stationary we can operate via their spectral representations. In particular, let $\ddot{x}(\tau_L)$, $\underline{x}(\tau_L)$ and $\bar{x}(\tau_L)$ have spectral densities equal to respectively $\ddot{S}(z)$, $\underline{S}(z)$ and $\bar{S}(z)$, for $z = \exp(-i\omega)$ with $\omega \in [0, \pi]$. Then the low frequency VAR model is determined by the following relationship:

$$\underline{S}(z) = \underline{D}(z)\ddot{S}(z)\underline{D}(z^{-1})'$$
$$\bar{S}(z) = \bar{D}(z)\ddot{S}(z)\bar{D}(z^{-1})'$$
$$\bar{S}(z) = \bar{F}(z)\underline{S}(z)\bar{F}(z^{-1})'. \tag{4.8}$$

The result follows straightforwardly from spectral density relationships via transfer functions (see e.g. Priestley, 1981). Note that the spectral densities $\underline{S}(z)$ and $\bar{S}(z)$ in (4.8) typically do *not* correspond to finite order VAR specifications. The mapping to such specifications is what will be covered in the next section.

5. Specification and estimation

Empirical work involves critical choices of model specification and parameterization. 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 τ_Lth observation:

$$f(\underline{x}(\tau_L)|\underline{x}(\tau_L - 1), \ldots, \underline{x}(\tau_L - P); \Psi)$$
$$= (2\pi)^{\frac{\tilde{m}_{\tau_L}}{2}}|(\mathbb{C}^{\tau_L}(\Psi)\mathbb{C}^{\tau_L}(\Psi)')^{-1}|^{1/2}$$
$$\times \exp -\frac{1}{2}(\underline{\varepsilon}(\tau_L)'(\mathbb{C}^{\tau_L}(\Psi)\mathbb{C}^{\tau_L}(\Psi)')^{-1}\underline{\varepsilon}(\tau_L))$$

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

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

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

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

The asymptotic analysis of VAR models is well known, see e.g. Hamilton (1994), and applies in the current setting without any modifications.[10]

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 VAR model estimators

Having specified some potentially parsimonious mixed frequency specifications, we now turn our attention to the study of a standard low frequency VAR model with a finite number of lags and parameter vector Φ as well as a mixed frequency VAR model with a finite number of lags and parameter vector Ψ. To so, we assume the DGP is the $m * K$ dimensional vector $\ddot{x}(\tau_L)$ described by Eq. (4.6):

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

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 (2.13) parameterized by Ψ, and on the other hand the K dimensional traditional low frequency VAR parameterized by Φ.

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.

In terms of spectral representations, we are looking at two parameterized spectral densities: (a) $\underline{S}(z, \Phi)$ and (b) $\bar{S}(z, \Psi)$ against the DGP $\ddot{S}(z)$. 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.

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

$$\mathbb{L}(\bar{x}(\tau_L)_1^{T_L}|\Phi) = (-T_L(K_L + K_H)/2) \log(2\pi)$$
$$+ (T_L/2) \log|(\overline{\mathbb{C}}(\Phi)\overline{\mathbb{C}}(\Phi)')^{-1}|$$
$$- (1/2)\sum_{\tau_L=1}^{T_L}[\bar{\varepsilon}(\tau_L)'(\overline{\mathbb{C}}(\Phi)\overline{\mathbb{C}}(\Phi)')^{-1}\bar{\varepsilon}(\tau_L)]. \tag{5.3}$$

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

Proposition 5.1. *Let Assumptions 2.1 and B.1 through B.4 hold and the DGP is described by (4.6) with aggregation schemes appearing in (4.8) characterized by $\underline{D}$ and $\bar{D}$. Then the maximum likelihood estimator appearing in (5.1), denoted $\hat{\Psi}$, minimizes*

$$\hat{\Psi} = \underset{\Psi}{\text{Argmin}} \left[E_1(\underline{S}(\Psi)) + E_2(\underline{S}(\Psi), \ddot{S}, \underline{D}) + E_3(\underline{S}(\Psi), \ddot{S})\right]$$

$$E_1(\underline{S}(\Psi)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(\det \underline{S}(\omega, \Psi))d\omega$$

$$E_2(\underline{S}(\Psi), \ddot{S}, \underline{D}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} trace(\underline{S}(\omega, \Psi)^{-1}\underline{D}(\omega)\ddot{S}(\omega)\overline{D}(\omega^{-1})')d\omega$$

---

[10] In a classical framework, maximum likelihood may prove to be intractable for large systems. In a context similar to ours with daily and monthly data, Chauvet et al. (2013) propose a two-step procedure to estimate a mixed-frequency VAR model based on SUR regressions.

$$E_3(\underline{S}(\Psi), \ddot{S}) = (\mu_{\underline{x}} - \mu_{\underline{x}}^{\Psi})' \underline{S}(0, \Psi)^{-1}(\mu_{\underline{x}} - \mu_{\underline{x}}^{\Psi}) \qquad (5.4)$$

where $\mu_{\underline{x}}^{\Psi}$ is the mean for the high frequency process implied by the parametric specification of the mixed frequency VAR whereas $\mu_{\underline{x}}$ is the mean implied by the DGP. Likewise, the maximum likelihood estimator appearing (5.3), denoted $\hat{\Phi}$, minimizes

$$\hat{\Phi} = \underset{\Phi}{\text{Argmin}} \left[E_1(\overline{S}(\Phi)) + E_2(\overline{S}(\Phi), \ddot{S}, \overline{D}) + E_3(\overline{S}(\Phi), \ddot{S}) \right]$$

$$E_1(\overline{S}(\Phi)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(\det \overline{S}(\omega, \Psi)) d\omega$$

$$E_2(\overline{S}(\Phi), \ddot{S}, \overline{D}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} trace(\overline{S}(\omega, \Phi)^{-1} \overline{D}(\omega) \ddot{S}(\omega) \overline{D}(\omega^{-1})') d\omega$$

$$E_3(\underline{S}(\Phi), \ddot{S}) = (\mu_{\overline{x}} - \mu_{\overline{x}}^{\Phi})' \overline{S}(0, \Psi)^{-1}(\mu_{\overline{x}} - \mu_{\overline{x}}^{\Psi}). \qquad (5.5)$$

Proof see Appendix B.

Not surprisingly the asymptotic properties depend not only on the tension between the assumed and actual DGP, but also the aggregation scheme, through the filters $\underline{D}$ and $\overline{D}$ appearing in Eq. (4.8). The term E_1 relates to the likelihood function in the absence of mis-specification. The terms E_2 and E_3 represent the impact of mis-specification, the latter pertaining to the zero frequency only.

5.2. Bayesian mixed frequency VAR

It was noted in the Introduction that the new class of models in the paper have a few virtues, including a specification that can be estimated and analyzed with standard VAR analysis tools—including Bayesian estimation. We therefore start with adopting standard BVAR tools to a mixed frequency setting.

Recent work on MIDAS regressions includes Bayesian estimation approaches, see notably Rodriguez and Puggioni (2010), Ghysels and Owyang (2011), Ghysels et al. (2014), Pettenuzzo et al. (2014), among others. It is also the purpose of this section to expand these recent developments to a mixed frequency VAR framework.

5.2.1. Standard Bayesian VAR and mixed frequency data

Standard priors in Bayesian VAR models involve univariate AR(1) models, or even more restrictively a random walk, see e.g. Doan et al. (1984), Litterman (1986), Kadiyala and Karlsson (1997), Sims and Zha (1998), among others. We will therefore also proceed with AR(1) priors, adapted to a mixed frequency setting.

It will be convenient to start again with a simplified example. Namely, consider the case where all the low frequency τ_L series are release at the end of the period as in Eq. (2.1) with zero mean series, with K_L and K_H of any dimension (hence not necessarily one). In particular, let us assume that all high frequency data are mean zero and follow a univariate AR(1) model with the same persistence featuring parameter ρ_H and the low frequency data are also mean zero and follow a univariate AR(1) model with the same persistence with parameter ρ_L. Note that a special case of course is $\rho_H = \rho_L$. Separating ρ_H from ρ_L in the prior allows for the flexibility of having different persistence. In particular, we typically set $\rho_H = 0$, i.e. all high frequency processes are white noise, and put ρ_H equal to some value between zero and one, and possibly equal to one. Under the univariate AR(1) the stacked system, similar to Eq. (2.9), looks as follows:

$$\begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ x_H(\tau_L, m) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0 & \cdots & diag(\rho_H) & 0 \\ \vdots & \cdots & \vdots & \vdots \\ 0 & \cdots & diag(\rho_H^m) & 0 \\ 0 & \cdots & 0 & diag(\rho_L^m) \end{bmatrix}$$

$$\times \begin{bmatrix} x_H(\tau_L - 1, 1) \\ \vdots \\ x_H(\tau_L - 1, m) \\ x_L(\tau_L - 1) \end{bmatrix} + \underline{\varepsilon}(\tau_L). \qquad (5.6)$$

The Bayesian priors will be tailored according to the above specification. To do so, it will be useful to first write again the general VAR as:

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

$$\times \begin{bmatrix} x_H(\tau_L - j, 1) \\ \vdots \\ x_H(\tau_L - j, m) \\ x_L(\tau_L - j) \end{bmatrix} + \underline{\varepsilon}(\tau_L)$$

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 $dim(A_j^{a,b}, a, b = 1, \ldots, m) = K_H^2$.

Using the aforementioned univariate AR of order one priors, we have:

$$\mathbb{E}[A_1^{a,m}] = diag(\rho_H^a)_{K_H^2} \qquad \mathbb{V}[A_1^{a,m}] = \frac{\lambda^2}{[(a)]^2} \mathbf{1}_{K_H^2} \quad a = 1, \ldots, m$$

$$\mathbb{E}[A_1^{m+1,m+1}] = diag(\rho_L^m)_{K_L^2} \qquad \mathbb{V}[A_1^{m+1,m+1}] = \frac{\lambda^2}{[(m)]^2} \mathbf{1}_{K_L^2} \qquad (5.7)$$

and

$$\mathbb{E}[A_j^{a,b}] = \mathbf{0}_{K_H^2}, \qquad \mathbb{V}[A_j^{a,b}] = \frac{\lambda^2}{[(j-1)m + (m-b+a)]^2} \mathbf{1}_{K_H^2}$$

$$1 \le a \le m, 1 \le b \le m - 1$$

$$\mathbb{E}[A_j^{m+1,b}] = \mathbf{0}_{K_H^2},$$

$$\mathbb{V}[A_j^{m+1,b}] = \frac{\lambda^2}{[(j-1)m + (m-b+1)]^2} \mathbf{1}_{K_H^2} \quad b = 1, \ldots, m$$

$$\mathbb{E}[A_j^{a,m}] = \mathbf{0}_{K_H^2}, \qquad (5.8)$$

$$\mathbb{V}[A_j^{a,m}] = \frac{\lambda^2}{((j-1)m + a)^2} \mathbf{1}_{K_H, K_H} \quad j > 1, a = 1, \ldots, m$$

$$\mathbb{E}[A_j^{a,m+1}] = \mathbf{0}_{K_H \times K_L},$$

$$\mathbb{V}[A_j^{a,m+1}] = \vartheta_{HL} \frac{\lambda^2}{[(j-1)m + a]^2} \mathbb{S}_{HL} \quad a = 1, \ldots, m$$

where the notation $\mathbb{V}[\;\;]$ 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 $\mathbb{S}_{HL} \equiv [\sigma_{i,H}^2 / \sigma_{j,L}^2; i = 1, \ldots, K_H, j = 1, \ldots, K_L]$. 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 λ 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.[11]

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-τ_L period lag structure.

Last but not least, we also need to formulate priors for the innovation variance covariance matrix. 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).

5.3. Extensions of Bayesian MIDAS to VAR

While there have been many discussions regarding priors in standard Bayesian VAR models, it is worth commenting a bit further on the specification of the MIDAS related part of the VAR.

We start with the framework of the previous subsection and revisit the prior (mean) $\mathbb{E}[A_j^{m+1,b}] = \mathbf{0}_{K_H^2}$ in (5.8). One can view the completely unrestricted specification where each weight is estimated separately as inspired by the U-MIDAS approach suggested by Foroni et al. (2015) shown to work for small values of m.

Instead of a U-MIDAS setup one can also implement the estimation of mixed frequency VAR models using the step functions approach to MIDAS of Ghysels et al. (2006). Namely, we can think of prior means for step functions:

$$[A_1^{m+1,1} \ldots A_1^{m+1,m} A_2^{m+1,1} \ldots A_P^{m+1,m}]$$
$$= \left[\underbrace{B_1 \ldots B_1}_{} \underbrace{B_2 \ldots B_2}_{} \ldots \underbrace{B_s \ldots B_s}_{} \right] \tag{5.9}$$

with s step matrices that imply fixed lag effects across subsets that span $P \times m$ lags. Now, we are left with specifying priors for the step functions. Then we can use a non-zero decaying prior similar to that appearing in (5.13) for $\mathbb{E}[B_j]$ and $\mathbb{V}[B_j]$.

In the remaining of this subsection we deal with modifying the specification of priors when we use MIDAS polynomials in the context of mixed frequency VAR models. Hence, we need to focus on the priors regarding $A_j^{m+1,b}$ appearing in (5.7). Recall from Section 2.3 that we considered two schemes, appearing in Eqs. (2.11) and (2.12):

$$[A_1^{m+1,1}(\gamma) \ldots A_1^{m+1,m}(\gamma) A_2^{m+1,1}(\gamma) \ldots A_P^{m+1,m}(\gamma)]$$
$$= \begin{cases} B \otimes \left[\sum_{i=1}^{K_H \times P} (w(\gamma)_i) \right] \\ B_i \otimes \left[\sum_{i=1}^{m} (w(\gamma)_i) \right], i = 1, \ldots, P. \end{cases} \tag{5.10}$$

Recall also that B a $K_L \times K_H$ matrix and in Section 2.3 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:

$$\begin{cases} \left(B^{a,b} \left[\sum_{i=1}^{K_H \times P} (w(\gamma^{a,b})_i) \right], a = 1, \ldots, K_L; b = 1, \ldots, K_L \right) \\ B_i^{a,b} \left[\sum_{i=1}^{m} (w(\gamma^{a,b})_i) \right], i = 1, \ldots, P, a = 1, \ldots, K_L; \\ b = 1, \ldots, K_L. \end{cases} \tag{5.11}$$

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 (5.10) or the common prior in the case of many single MIDAS polynomials as in (5.11), is a Gamma distribution. Since the MIDAS Beta polynomial involves two parameters, we draw each parameter from an independent Gamma. In the case of (5.11) 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:

$$\gamma_i \sim \Gamma(\mathbf{f}_0, \mathbf{F}_0) \quad i = 1, 2 \tag{5.12}$$

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 $\gamma_2 \sim \Gamma(\mathbf{f}_0, \mathbf{F}_0)$ 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, γ^*, from a proposal density, $q(\gamma^*|\gamma^{[i]})$, 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 $\boldsymbol{\gamma}^* = (\gamma_1^*, \gamma_2^*)'$ 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

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

where $\mathbb{L}(\underline{\mathbf{x}}(\tau_L)_1^{T_L}|\Psi_{-\gamma}, \boldsymbol{\gamma}^*)$ is the conditional likelihood given the parameters $\Psi_{-\gamma}$—which are all the parameters in Ψ excluding γ and $d\Gamma(.|.,.)$ 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 γ for a single as well as multiple MIDAS polynomial weighting schemes to avoid further complicating the notation.

The subsequent steps are similar to those appearing in the previous subsection with a slight modification. In particular, we should think of the matrices $A_j(\Psi_{-\gamma}, \gamma)$ as depending on the

---

[11] Another simplification is that we do not discuss the overall mean of the VAR and its Bayesian estimation—since this is widely covered in the aforementioned references.

parameter vector driving the MIDAS polynomial weights as well as the remaining parameters in $\Psi_{-\gamma}$. Hence, from here on we can again follow Doan et al. (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 γ etc. for convenience):

- $\Psi_{-\gamma,H} = ((A_j^{a,b}, a = 1, \ldots, m, b = 1, \ldots, m + 1), j = 1, \ldots, P)'$, the set parameters pertaining to the high frequency components of the vector.
- $\Psi_{-\gamma,V} = ((\mathbb{C}\mathbb{C}')_{a,b}; a = 1, \ldots, m + 1, b = 1, \ldots, m + 1)'$, the parameters pertaining to the covariance matrix of the errors,

which can be directly taken from (a) the specifications in (5.7) and (5.8) – except the second entry in the latter – and (b) the priors for the innovation variance covariance matrix mentioned in the previous subsection. Finally:

- $\Psi_{-\gamma,L} = (B \text{ or } B_i, i = 1, \ldots, P)'$, the slope parameters pertaining to the MIDAS regressions.

Regarding the MIDAS regressions in $\Psi_{-\gamma,L}$, we have the following priors for the slope coefficients:

$$\mathbb{E}[B] = \mathbf{0}_{K_L \times K_H}, \quad \mathbb{V}[B] = \vartheta_{LH} \lambda^2 \mathbf{1}_{K_L \times K_H} \mathbb{S}_{LH}$$
$$\mathbb{E}[B_j] = \mathbf{0}_{K_L \times K_H}, \quad \mathbb{V}[B_j] = \vartheta_{LH} \frac{\lambda^2}{j^2} \mathbf{1}_{K_L \times K_H} \mathbb{S}_{LH} \tag{5.13}$$

with ϑ_{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 (5.13) 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 $\mathbb{E}[A_j^{a,m+1}] = \mathbf{0}_{K_H \times K_L}$.

6. Numerical examples

In this section we provide some numerical illustrative examples to compare the behavior of impulse response functions (IFRs) in mixed frequency and traditional VAR models. The analysis in the section pertains to *population* properties. We need to rely on numerical computations to obtain these properties. 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 IRFs.

The DGP is based on Eq. (4.7), and in particular:

$$\begin{bmatrix} I_2 & 0_2 & 0_2 \\ -A_c^{2,1} & I_2 & 0_2 \\ 0_2 & -A_c^{3,2} & I_2 \end{bmatrix} \ddot{x}(\tau_L) = \begin{bmatrix} 0_2 & 0_2 & A^{1,3} \\ 0_2 & 0_2 & 0_2 \\ 0_2 & 0_2 & 0_2 \end{bmatrix} \ddot{x}(\tau_L - 1)$$
$$+ \ddot{\varepsilon}(\tau_L) \tag{6.1}$$

which is a periodic VAR(1) model with $m = 3$. We will mostly work with $a_{ij}^k = a_{ij}^l$ for $k, l = 1, 2$ and 3. Hence, we focus primarily on a regular high frequency VAR(1) model. Hence, we assume a data generating process which consists of a monthly bivariate system, and therefore the stacked system is of dimension 6 when we sample quarterly. The data structure following Eqs. (4.2) and (4.3) is as follows:

$$\bar{x}(\tau_L) \equiv \begin{bmatrix} x_H(\tau_L) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \ddot{x}(\tau_L, 1) \\ \ddot{x}(\tau_L, 2) \\ \ddot{x}(\tau_L, 3) \\ \ddot{x}(\tau_L, 4) \\ \ddot{x}(\tau_L, 5) \\ \ddot{x}(\tau_L, 6) \end{bmatrix} \tag{6.2}$$

for the low frequency VAR model and:

$$\underline{x}(\tau_L) \equiv \begin{bmatrix} x_H(\tau_L, 1) \\ x_H(\tau_L, 2) \\ x_H(\tau_L, 3) \\ x_L(\tau_L) \end{bmatrix}$$
$$= \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \ddot{x}(\tau_L, 1) \\ \ddot{x}(\tau_L, 2) \\ \ddot{x}(\tau_L, 3) \\ \ddot{x}(\tau_L, 4) \\ \ddot{x}(\tau_L, 5) \\ \ddot{x}(\tau_L, 6) \end{bmatrix}. \tag{6.3}$$

The above equation implies that the first series is available every month, whereas the second series is aggregated as a flow variable. The low frequency VAR involves both series being flows. We fit two mis-specified VAR models with Gaussian innovations: (1) the mixed frequency VAR(1) and (2) a low frequency VAR(1). In both cases we compute population pseudo-true parameter values which are obtained from minimizing the criterion appearing in Proposition 5.1. Hence, we view the mixed frequency and traditional VAR models as a mis-specified models estimated via standard MLE. The low frequency VAR(1) involves four parameters for the autoregressive dynamics, and three parameters for the innovation covariance matrix. The mixed frequency VAR is parameterized as follows:

$$\underline{x}(\tau_L) = \begin{bmatrix} 0 & 0 & \rho_H & a_1 \\ 0 & 0 & \rho_H^2 & a_2 \\ 0 & 0 & \rho_H^3 & a_3 \\ b_1 & b_2 & b_3 & \rho_L \end{bmatrix} \underline{x}(\tau_L - 1) + \underline{\varepsilon}(\tau_L) \tag{6.4}$$

hence the MIDAS regression coefficients are unconstrained, and so is the impact of low frequency onto high frequency. The total number of parameters is therefore equal to ten for the innovation covariance matrix plus eight in the above specification of the dynamics, which adds up to eighteen parameters. Hence, the mixed frequency VAR has eleven more parameters, four related to the conditional mean dynamics.

We consider several parameter settings for the DGP in Eq. (6.1). In particular, we consider three cases. Cases I through III have the setting:

$$A_c^{2,1} = A_c^{3,2} = A^{1,3} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{6.5}$$

where we have the following parameter configurations:

- Case I: $a_{ii} = .5$, for $i = 1, 2$; $a_{12} = 0$, $a_{21} = .1$.
- Case II: $a_{ii} = .5$, for $i = 1, 2$; $a_{21} = 0$, $a_{12} = .1$.
- Case III: $a_{ii} = .5$, for $i = 1, 2$; $a_{12} = .1$ $a_{21} = .1$.

Hence, for Case I, Granger causality runs from high to low frequency data, Case II is the opposite and Case III features bi-directional Granger causality. We took .5 for the persistence, which is neither an extremely persistent nor an i.i.d. sequence. The strength of the Granger causality was chosen accordingly. While one could choose many other examples, our small set of cases highlights the most important issues.

It is important to clarify the nature of the shocks in our analysis. There are essentially three types of shock configurations, which are given as (6.6) in Box II. The six-dimensional error process $\ddot{\varepsilon}(\tau_L)$ is latent since we assume never to see the high frequency bivariate process, and therefore the shocks pertain to a data filtration we do not observe. The four-dimensional $\underline{\varepsilon}(\tau_L)$ mixed frequency process is the best we can achieve—given the assumed information technology. Finally, the two-dimensional process $\bar{\varepsilon}(\tau_L)$ is obtained after aggregation. We will therefore compare three types of IFRs.

$$\underbrace{\begin{bmatrix} \ddot{x}_H(\tau_L, 1) \\ \ddot{x}_L(\tau_L, 1) \\ \ddot{x}_H(\tau_L, 2) \\ \ddot{x}_L(\tau_L, 2) \\ \ddot{x}_H(\tau_L, 3) \\ \ddot{x}_H(\tau_L, 3) \end{bmatrix} \rightsquigarrow \begin{bmatrix} \ddot{\varepsilon}(\tau_L, 1) \\ \ddot{\varepsilon}(\tau_L, 2) \\ \ddot{\varepsilon}(\tau_L, 3) \\ \ddot{\varepsilon}(\tau_L, 4) \\ \ddot{\varepsilon}(\tau_L, 5) \\ \ddot{\varepsilon}(\tau_L, 6) \end{bmatrix}}_{\text{Latent}} \qquad \underbrace{\begin{bmatrix} x_H(\tau_L, 1) \\ x_H(\tau_L, 2) \\ x_H(\tau_L, 3) \\ x_L(\tau_L) \end{bmatrix} \rightsquigarrow \begin{bmatrix} \underline{\varepsilon}(\tau_L, 1) \\ \underline{\varepsilon}(\tau_L, 2) \\ \underline{\varepsilon}(\tau_L, 3) \\ \underline{\varepsilon}(\tau_L, 4) \end{bmatrix}}_{\text{Mixed}}$$

$$\underbrace{\begin{bmatrix} x_H(\tau_L) \\ x_L(\tau_L) \end{bmatrix} \rightsquigarrow \begin{bmatrix} \overline{\varepsilon}(\tau_L, 1) \\ \overline{\varepsilon}(\tau_L, 2) \end{bmatrix}}_{\text{Low}}$$

(6.6)

Box II.

- **HI → Lo** : Responses of $x_L(\tau_L)$, i.e. the aggregated series, to shocks $\ddot{\varepsilon}(\tau_L, 1)$ and $\ddot{\varepsilon}(\tau_L, 5)$, that is the first and last period high frequency series shocks using the latent high frequency VAR appearing in (6.1), and compare them to:
  – Responses of $x_L(\tau_L)$ to shocks $\underline{\varepsilon}(\tau_L, 1)$ and $\underline{\varepsilon}(\tau_L, 3)$, again the first and last period high frequency series shocks given the mixed frequency filtration,
  – Responses of $x_L(\tau_L)$ to shocks $\overline{\varepsilon}(\tau_L, 1)$ in the low frequency bivariate system.
- **Lo → HI**: Responses of $\ddot{x}_H(\tau_L, 1)$ and $\ddot{x}_H(\tau_L, 3)$ to shocks $\ddot{\varepsilon}(\tau_L, 2)$ and $\ddot{\varepsilon}(\tau_L, 6)$, that is the first and last period latent low frequency series shocks, and compare them to:
  – Responses of $\underline{x}_H(\tau_L, 1)$ and $\underline{x}_H(\tau_L, 3)$ to shocks $\underline{\varepsilon}(\tau_L, 4)$, i.e. the observed low frequency series shocks
  – Responses of $\overline{x}_H(\tau_L)$ to shocks $\overline{\varepsilon}(\tau_L, 2)$ in the low frequency VAR.

Hence, our strategy is to study the impact of high frequency shocks onto low frequency shocks and vice versa, and to compare them across the three VAR specifications.

The IRFs appear in Figs. 1 through 3. Each of the figures have the same structure. The top panel shows the IRFs appearing above as **HI → Lo**. It contains two plots—the left one are the IRFs for a shock in the first month of the quarter, the right one contains the IRFs for the last month. The lower panel in each figure has the **Lo → HI** results again with two plots for first and last month. Note that for the low frequency VAR the left and right plots should contain the same IRF, since there is no distinction possible for first and last month of the quarter, hence the IRFs are simply repeated for the purpose of comparison. The plots also share the same line characteristics: circles for the high frequency IRFs, pluses for the low frequency ones, and finally stars for the mixed frequency ones. Finally, the time scale is quarterly.

The results for Case I appear in Fig. 1. The top panel shows how the low frequency IRF is very different from the high and mixed frequency ones. The latter are obviously not identical, but remarkably close. They both die out quickly — around three to four quarters, whereas the low frequency IRFS spikes at two quarters and tapers off much more slowly. It is also important to note that the IRFs are not the same for the first and last month high frequency shocks (recall that the low frequency IRF is repeated and therefore identical). Although this is not quite exactly the case, one notices that the IRFs depicted in the right plot looks almost like a shifted. This is expected. A shock in the first month affect the second and third month high frequency series, therefore the total impact on the aggregated low frequency process is different from the IRF for the last month where such an accumulation effect does not occur. We will observe similar patterns in the next section where we report empirical IRFs. It is therefore not an indication that the DGP has a periodic structure, i.e. within quarter variation of the VAR parameters as allowed for in Eq. (6.1). Obviously, as emphasized

by Hansen and Sargent (1990), in the case of genuine period models one might obtain far more diverse patterns of IRFs.

The lower panel in Fig. 1 is arguably more interesting. Recall that there is no Granger causality from the low frequency to the high frequency process. Hence, the IRFs with circles are flat. Neither the mixed nor the low frequency VAR models feature such a pattern. The former show positive responses, and the latter the opposite. Both appear to be poor approximations to the actual IRF.

Fig. 2 covers Case II, where the low frequency process Granger causes the high frequency one, i.e. the reverse of the previous case. Therefore, one observes now the flat IRFs in the top panel instead. We notice again that the mixed frequency VAR model IRFs—while not entirely flat on the x-axis, show a pattern to zero. In contrast, the low frequency IRF shows a pattern almost identical to that appearing in the top panel of Fig. 1. The lower panel of Fig. 2 now features non-trivial IRFs. Yet, the conclusions one can draw are similar to those we noted in Fig. 1. Case III—a mixture of the previous two show similar features: high frequency shocks are well identified in mixed frequency VAR models, but the reverse remains harder.

Table 1 displays the numerical values that were obtained from minimizing the criterion appearing in Proposition 5.1. Hence, the table shows the degree of mis-specification one obtains in large samples. Of particular interest are the off-diagonal elements which reveal that the Granger causality patterns in the DGPs are not maintained after aggregation, something which is known and also revealed by the IRFs.[12] Table 2 does the same for the mis-specified mixed frequency VAR models. While somewhat harder to identify, we also see that either the MIDAS coefficients, or the last column do not feature the desired Granger causality patterns.

Case IV is not fully covered. It is meant to show that the IRFs for low frequency shocks may not always be so poorly approximated in a mixed frequency VAR model. Case IV is identical to Case II, except that the persistence of the high frequency process is turned off, i.e. $a_{11} = 0$ while all other parameters are kept the same. We only report the IRFs to low frequency shocks in Fig. 4 which indicate that the mixed frequency VAR model captures the true DGP better.

7. Empirical examples

In the previous section we showed several empirical examples that revealed the discrepancies that may occur in IRFs due to mixed frequency sampling versus temporal aggregation. We now turn

---

[12] Much has been written about the spurious effects temporal aggregation may have on testing for Granger causality, see e.g. Granger (1980, 1988), Lütkepohl (1993), Granger (1995), Renault et al. (1998), Breitung and Swanson (2002), McCrorie and Chambers (2006), among others.
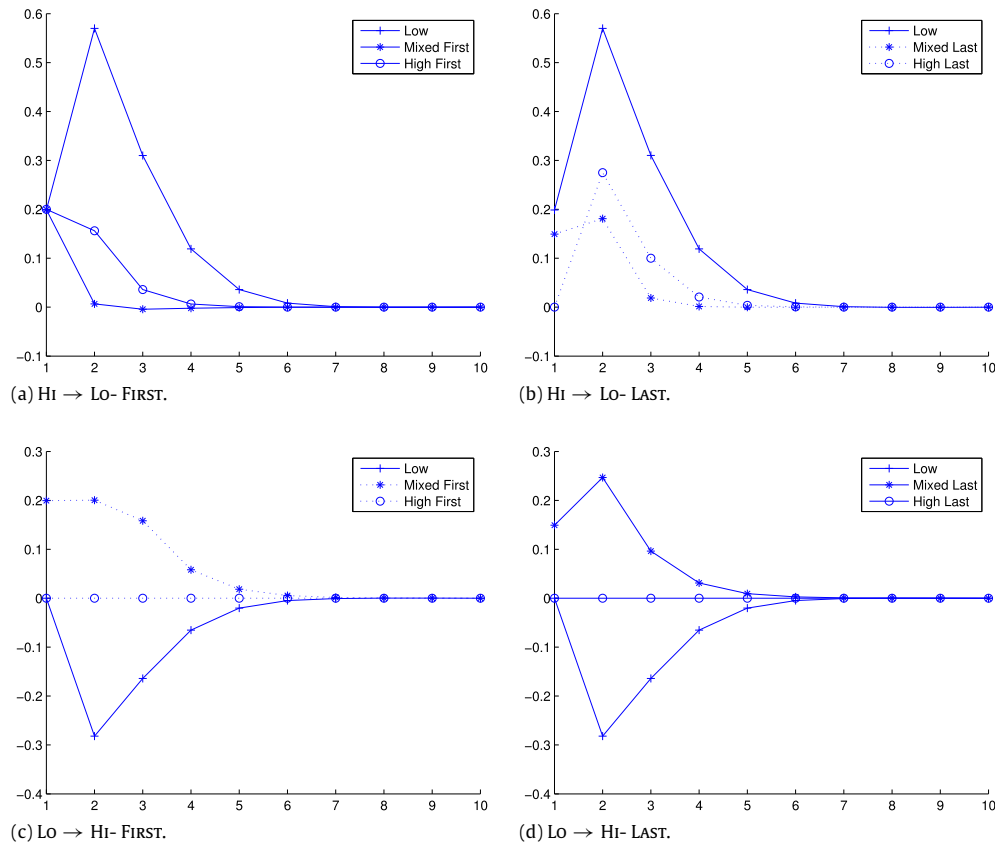
(a) Hɪ → Lo- Fɪʀsᴛ.

(b) Hɪ → Lo- Lᴀsᴛ.

(c) Lo → Hɪ- Fɪʀsᴛ.

(d) Lo → Hɪ- Lᴀsᴛ.

Fig. 1. Impulse response functions monthly/quarterly—Case I. The DGP appears in Eq. (6.1) with $m = 3$ and a parameter setting labeled Case I, namely with $a_{ii} = .5$, for $i = 1$, 2; $a_{12} = 0$, $a_{21} = .1$ in Eq. (6.5). Low and mixed frequency VAR(1) models are estimated using the aggregations schemes appearing in (6.2) and (6.3). Two mis-specified VAR models are considered: (1) the mixed frequency VAR(1) and (2) a low frequency VAR(1). In both cases population pseudo-true parameter values are obtained from minimizing the criterion appearing in Proposition 5.1. The top panel shows the impulse response functions (IRFs) of a high frequency series shock compared with the IRFs obtained from the mis-specified mixed frequency VAR and low frequency VAR(1) models. The top panel contains two plots—the left one is the IRF for a shock in the first month of the quarter, the right contains the IRF for the last month. The lower panel shows the IRFs of a low frequency series shock compared with the IRFs obtained from the mis-specified mixed frequency VAR and low frequency VAR(1) models. The lower panel also contains two plots—the left one is the IRF for a shock in the first month of the quarter, the right contains the IRF for the last month.

Table 1

Approximate low frequency vector autoregressive models. The table reports the mis-specified low frequency VAR(1) given the high frequency VAR data generating process appearing in Eq. (6.1) and the aggregation scheme appearing in (6.2). The entries to the Table contain the pseudo-true parameter values which are obtained from minimizing the criterion appearing in Proposition 5.1. We report the conditional mean dynamics as well as the innovation covariance matrix. The DGPs, Cases I through IV—are described in Section 6.

Case I

$$\begin{bmatrix} x_H(\tau_L) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0.30 & -0.11 \\ 0.20 & 0.28 \end{bmatrix} \begin{bmatrix} x_H(\tau_L - 1) \\ x_L(\tau_L - 1) \end{bmatrix} + \overline{\varepsilon}(\tau_L) \qquad E[\overline{\varepsilon}(\tau_L)\overline{\varepsilon}(\tau_L)'] = \begin{bmatrix} 6.31 & 0.50 \\ 0.50 & 6.36 \end{bmatrix}$$

Case II

$$\begin{bmatrix} x_H(\tau_L) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0.31 & -0.10 \\ 0.19 & 0.27 \end{bmatrix} \begin{bmatrix} x_H(\tau_L - 1) \\ x_L(\tau_L - 1) \end{bmatrix} + \overline{\varepsilon}(\tau_L) \qquad E[\overline{\varepsilon}(\tau_L)\overline{\varepsilon}(\tau_L)'] = \begin{bmatrix} 6.36 & 0.50 \\ 0.50 & 6.31 \end{bmatrix}$$

Case 3

$$\begin{bmatrix} x_H(\tau_L) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0.32 & -0.023 \\ 0.20 & 0.27 \end{bmatrix} \begin{bmatrix} x_H(\tau_L - 1) \\ x_L(\tau_L - 1) \end{bmatrix} + \overline{\varepsilon}(\tau_L) \qquad E[\overline{\varepsilon}(\tau_L)\overline{\varepsilon}(\tau_L)'] = \begin{bmatrix} 6.40 & 1.00 \\ 1.00 & 6.39 \end{bmatrix}$$

our attention to IRFs with real data. To do so we implement the Bayesian estimation procedure discussed in Section 5.

The empirical application is tailored after Chiu et al. (2011) who develop a Bayesian approach to such mixed frequency VAR models where the missing data are drawn via 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 et al. (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
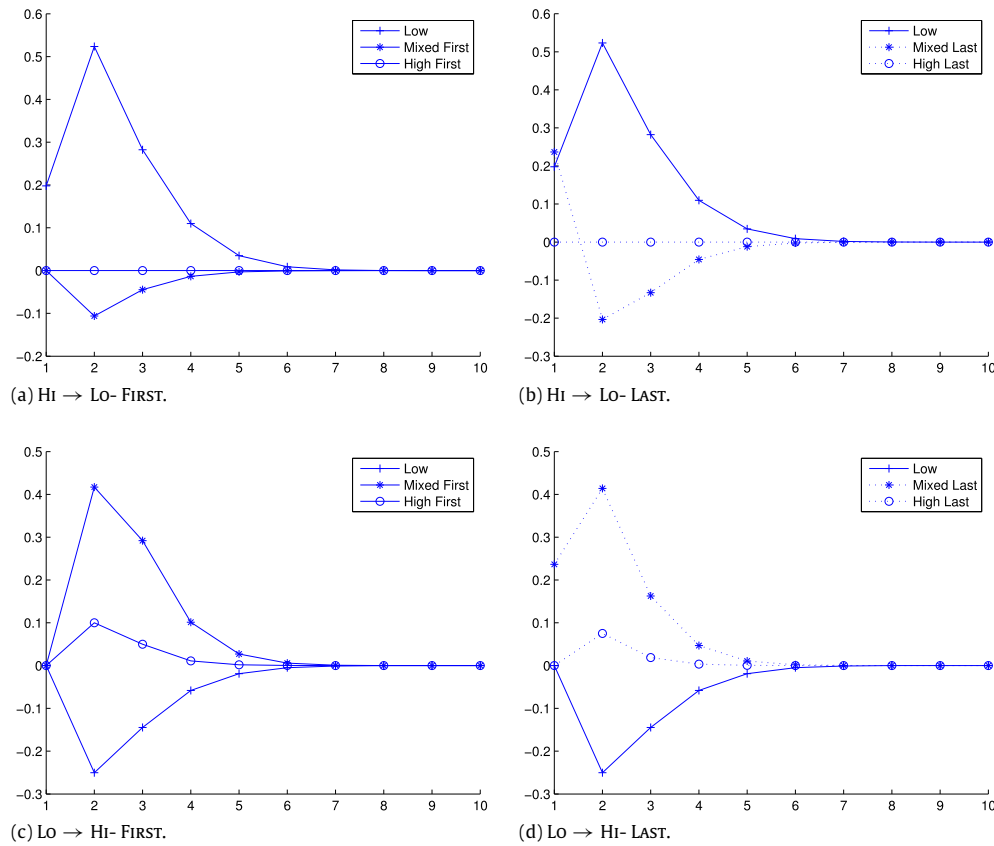
(a) Hɪ → Lo- Fɪʀsᴛ.

(b) Hɪ → Lo- Lᴀsᴛ.

(c) Lo → Hɪ- Fɪʀsᴛ.

(d) Lo → Hɪ- Lᴀsᴛ.

Fig. 2. Impulse response functions monthly/quarterly—Case II. The DGP appears in Eq. (6.1) with $m = 3$ and a parameter setting labeled Case II, namely with $a_{ii} = .5$, for $i = 1, 2$; $a_{21} = 0$, $a_{12} = .1$. in Eq. (6.5). Low and mixed frequency VAR(1) models are estimated using the aggregations schemes appearing in (6.2) and (6.3). Two mis-specified VAR models are considered: (1) the mixed frequency VAR(1) and (2) a low frequency VAR(1). In both cases population pseudo-true parameter values are obtained from minimizing the criterion appearing in Proposition 5.1. The top panel shows the impulse response functions (IRFs) of a high frequency series shock compared with the IRFs obtained from the mis-specified mixed frequency VAR and low frequency VAR(1) models. The top panel contains two plots—the left one is the IRF for a shock in the first month of the quarter, the right contains the IRF for the last month. The lower panel shows the IRFs of a low frequency series shock compared with the IRFs obtained from the mis-specified mixed frequency VAR and low frequency VAR(1) models. The lower panel also contains two plots — the left one is the IRF for a shock in the first month of the quarter, the right contains the IRF for the last month.

Table 2

Approximate mixed frequency vector autoregressive models. The table reports the mis-specified low frequency VAR(1) given the high frequency VAR data generating process appearing in Eq. (6.1) and the aggregation scheme appearing in (6.3) The entries to the Table contain the pseudo-true parameter values which are obtained from minimizing the criterion appearing in Proposition 5.1. We report the conditional mean dynamics as well as the innovation covariance matrix. The DGPs, Cases I through IV—are described in Section 6.

	Case I	
$\begin{bmatrix} x_H(\tau_L, 1) \\ x_H(\tau_L, 2) \\ x_H(\tau_L, 3) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0.49 & 0.02 \\ 0 & 0 & 0.24 & 0.02 \\ 0 & 0 & 0.12 & 0.04 \\ -0.13 & 0.12 & 0.07 & 0.29 \end{bmatrix} \begin{bmatrix} x_H(\tau_L - 1, 1) \\ x_H(\tau_L - 1, 2) \\ x_H(\tau_L - 1, 3) \\ x_L(\tau_L - 1) \end{bmatrix} + \underline{\varepsilon}(\tau_L)$		$E[\underline{\varepsilon}(\tau_L)\underline{\varepsilon}(\tau_L)'] = \begin{bmatrix} 1.00 & 0.50 & 0.25 & 0.20 \\ 0.50 & 1.25 & 0.62 & 0.30 \\ 0.25 & 0.63 & 1.31 & 0.15 \\ 0.20 & 0.30 & 0.15 & 6.40 \end{bmatrix}$

	Case II	
$\begin{bmatrix} x_H(\tau_L, 1) \\ x_H(\tau_L, 2) \\ x_H(\tau_L, 3) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0.49 & 0.05 \\ 0 & 0 & 0.24 & 0.06 \\ 0 & 0 & 0.12 & 0.06 \\ -0.04 & -0.05 & -0.18 & 0.30 \end{bmatrix} \begin{bmatrix} x_H(\tau_L - 1, 1) \\ x_H(\tau_L - 1, 2) \\ x_H(\tau_L - 1, 3) \\ x_L(\tau_L - 1) \end{bmatrix} + \underline{\varepsilon}(\tau_L)$		$E[\underline{\varepsilon}(\tau_L)\underline{\varepsilon}(\tau_L)'] = \begin{bmatrix} 1.00 & 0.50 & 0.25 & 0.00 \\ 0.50 & 1.26 & 0.64 & 0.00 \\ 0.25 & 0.64 & 1.33 & 0.24 \\ 0.00 & 0.00 & 0.24 & 6.28 \end{bmatrix}$

	Case III	
$\begin{bmatrix} x_H(\tau_L, 1) \\ x_H(\tau_L, 2) \\ x_H(\tau_L, 3) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0.49 & 0.05 \\ 0 & 0 & 0.24 & 0.06 \\ 0 & 0 & 0.12 & 0.07 \\ -0.09 & 0.03 & 0.06 & 0.31 \end{bmatrix} \begin{bmatrix} x_H(\tau_L - 1, 1) \\ x_H(\tau_L - 1, 2) \\ x_H(\tau_L - 1, 3) \\ x_L(\tau_L - 1) \end{bmatrix} + \underline{\varepsilon}(\tau_L)$		$E[\underline{\varepsilon}(\tau_L)\underline{\varepsilon}(\tau_L)'] = \begin{bmatrix} 1.00 & 0.50 & 0.26 & 0.20 \\ 0.50 & 1.26 & 0.64 & 0.30 \\ 0.26 & 0.64 & 1.34 & 0.39 \\ 0.20 & 0.30 & 0.39 & 6.36 \end{bmatrix}$

quarterly VAR model. Therefore, we study the co-movements of:

$$\begin{bmatrix} IP(\tau_L) \\ INFL(\tau_L) \\ UNEMP(\tau_L) \\ GDP(\tau_L) \end{bmatrix} \text{ vs } \begin{bmatrix} x_H(\tau_L, 1) \\ x_H(\tau_L, 2) \\ x_H(\tau_L, 3) \\ GDP(\tau_L) \end{bmatrix}$$

$$\text{with } x_H(\tau_L, j) = \begin{bmatrix} IP(\tau_L, j) \\ INFL(\tau_L, j) \\ UNEMP(\tau_L, j) \end{bmatrix}. \tag{7.1}$$

The data set used in the empirical application runs from January 1948 to December 2011, for a total of 256 quarters and therefore

Fig. 3. Impulse response functions monthly/quarterly—Case III. The DGP appears in Eq. (6.1) with $m = 3$ and a parameter setting labeled Case III, namely with $a_{ii} = .5$, for $i = 1, 2$; $a_{12} = .1$, $a_{21} = .1$. in Eq. (6.5). Low and mixed frequency VAR(1) models are estimated using the aggregations schemes appearing in (6.2) and (6.3). Two mis-specified VAR models are considered: (1) the mixed frequency VAR(1) and (2) a low frequency VAR(1). In both cases population pseudo-true parameter values are obtained from minimizing the criterion appearing in Proposition 5.1. The top panel shows the impulse response functions (IRFs) of a high frequency series shock compared with the IRFs obtained from the mis-specified mixed frequency VAR and low frequency VAR(1) models. The top panel contains two plots—the left one is the IRF for a shock in the first month of the quarter, the right contains the IRF for the last month. The lower panel shows the IRFs of a low frequency series shock compared with the IRFs obtained from the mis-specified mixed frequency VAR and low frequency VAR(1) models. The lower panel also contains two plots—the left one is the IRF for a shock in the first month of the quarter, the right contains the IRF for the last month.
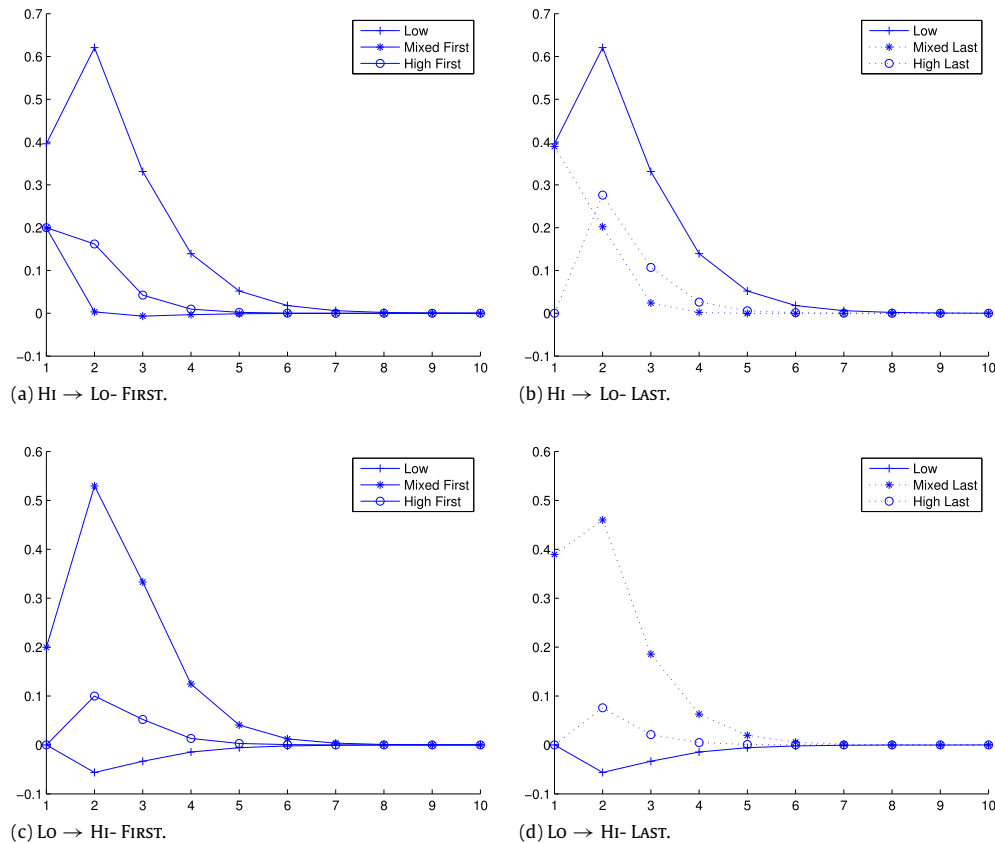

Fig. 4. Impulse response functions monthly/quarterly—Case IV. The DGP appears in Eq. (6.1) with $m = 3$ and a parameter setting labeled Case IV, namely with $a_{11} = 0$. and $a_{22} = .5$; $a_{21} = 0$, $a_{12} = .1$. in Eq. (6.5). Low and mixed frequency VAR(1) models are estimated using the aggregations schemes appearing in (6.2) and (6.3). Two mis-specified VAR models are considered: (1) the mixed frequency VAR(1) and (2) a low frequency VAR(1). In both cases population pseudo-true parameter values are obtained from minimizing the criterion appearing in Proposition 5.1. The top panel shows the impulse response functions (IRFs) of a high frequency series shock compared with the IRFs obtained from the mis-specified mixed frequency VAR and low frequency VAR(1) models. The top panel contains two plots—the left one is the IRF for a shock in the first month of the quarter, the right contains the IRF for the last month. The lower panel shows the IRFs of a low frequency series shock compared with the IRFs obtained from the mis-specified mixed frequency VAR and low frequency VAR(1) models. The lower panel also contains two plots—the left one is the IRF for a shock in the first month of the quarter, the right contains the IRF for the last month.
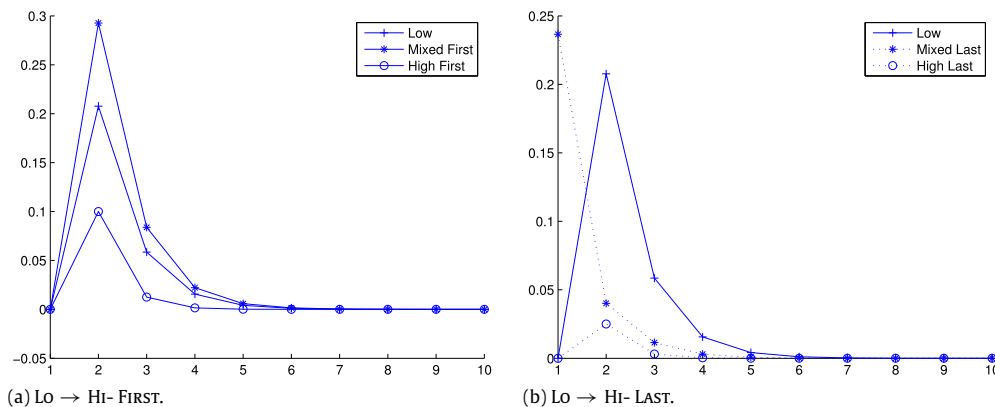
768 months.[13] Summary statistics for the variables are presented in Table 3 (which is comparable to Chiu et al. (2011, Table 2)).

---

[13] Our sample is slightly different from Chiu et al. (2011) who use data from January 1948 to June 2011, for a total of 762 months or 254 quarters.

We report the posterior means and standard deviations for the standard low frequency VAR model in Table 4. While our sample is not the same as in Chiu et al. (2011), we do find similar parameter estimates. As a result, we also find similar impulse response functions—which will be discussed later. The mixed frequency

Table 3

Summary statistics data. The table reports summary statistics for our sample data of industrial production, inflation, and unemployment rate, and GDP for the US. All but the last series are observed monthly. The data sources are as follows:

- Industrial Production Index (SA),   http://research.stlouisfed.org/fred2/series/INDPRO/.
- Inflation: CPI All Urban, http://www.bls.gov/cpi/tables.htm.
- Unemployment rate − 16yrs and older (SA),   http://www.bls.gov/webapps/legacy/cpsatab1.htm.
- Real GDP (SA and chained 2005 dollars), http://www.bea.gov/national/.

The series are transformed as follows: 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. The sample covers January 1948 until December 2011.

	Mean	Standard Dev.	Autocorrelation
IP	2.976	5.8253	0.9660
INFL	3.549	2.8095	0.9879
UNEMP	5.801	1.6504	0.9892
GDP	3.130	2.6830	0.8507

counterpart appears in Table 5. Here the parameter estimates display far more heterogeneous patterns, as expected.

Figs. 5 through 7 display the IRFs of GDP to shocks in the three high frequency series, respectively IP, INFL and UNEMP. The figures have the same structure. The top three plots display the IRFs separately for each month of the quarter, compared in each case with the IRF obtained from the quarterly aggregate VAR model. To avoid having too many lines on a single graph, we also show the

impulse response function – without confidence bands – together for all three months of the quarter again compared with the aggregate VAR impulse response function. The confidence bands are obtained from the Gibbs sampler iterations and represent the 5% and 95% percentiles obtained from the MCMC draws. Similarly, the impulse response functions are the median of the same draws. The legends for the plots appear in the plot without confidence bands. The same legend is used in all the plots, in particular the circled lines represent the aggregate VAR specification.

It is important to note the difference between our analysis and the latent high frequency VAR model of Chiu et al. (2011). The latter have only one response function per high frequency series, since the basic model is driven by a monthly VAR(1) and they do not analyze the impact of high frequency shocks to low frequency series, as we did with the numerical examples reported in the previous section. In their case the monthly estimates and the quarterly estimates converted to their monthly counterpart result in impulse response functions that, for the most part, are quite similar if not identical. Hence, their findings suggest that both the low frequency VAR and the latent high frequency one yield similar responses of the economy to the exogenous shocks. Most of the action in their comparison is with respect to the precision of the impulse response estimates improving with the monthly specification—exploiting the high frequency data. With the coarsely sampled estimation method, since the data are observed quarterly, co-movements of the variables, and hence the effects of shocks, are harder to disentangle. Hence, the latent high

Table 4

Posterior mean and standard error estimates VAR model. The table reports the posterior mean and variances of 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. The sample covers.

	IP($\tau_L - 1$)	INFL($\tau_L - 1$)	UNEMP($\tau_L - 1$)	GDP(τ_L)	A_0
IP(τ_L)	0.85	−0.35	0.34	−0.03	−0.07
	(0.06)	(0.07)	(0.06)	(0.12)	(0.19)
INFL(τ_L)	0.09	0.96	0.02	−0.07	0.01
	(0.02)	(0.02)	(0.03)	(0.04)	(0.16)
UNEMP(τ_L)	−0.02	0.03	0.94	−0.03	0.41
	(0.01)	(0.01)	(0.01)	(0.02)	(0.07)
GDP(τ_L)	0.13	−0.12	0.18	0.64	0.11
	(0.03)	(0.03)	(0.03)	(0.05)	(0.16)

Table 5

Posterior mean and standard error estimates mixed frequency VAR model. The table reports the posterior mean and variances of 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. The sample covers.

$\tau_L - 1$	IP(., 1)	INFL(., 1)	UNEMP(., 1)	IP(., 2)	INFL(., 2)	UNEMP(., 2)	IP(., 3)	INFL(., 3)	UNEMP(., 3)	GDP	A_0
IP(τ_L, 1)	−0.26	−0.17	−0.29	0.07	0.24	0.23	1.13	−0.12	0.11	0.07	−0.18
	(0.06)	(0.15)	(0.24)	(0.09)	(0.20)	(0.25)	(0.06)	(0.13)	(0.17)	(0.07)	(0.31)
INFL(τ_L, 1)	−0.01	−0.05	0.07	0.01	0.00	−0.11	0.03	1.04	0.06	−0.02	−0.09
	(0.02)	(0.07)	(0.13)	(0.04)	(0.10)	(0.15)	(0.02)	(0.07)	(0.12)	(0.03)	(0.13)
UNEMP(τ_L, 1)	0.01	0.02	0.05	0.03	−0.09	0.21	−0.05	0.06	0.71	−0.03	0.27
	(0.01)	(0.03)	(0.06)	(0.02)	(0.05)	(0.08)	(0.01)	(0.03)	(0.07)	(0.01)	(0.06)
IP(τ_L, 2)	−0.46	−0.01	−0.11	−0.02	−0.28	0.23	1.22	0.15	−0.01	0.29	−0.26
	(0.09)	(0.20)	(0.32)	(0.14)	(0.27)	(0.34)	(0.09)	(0.18)	(0.27)	(0.10)	(0.39)
INFL(τ_L, 2)	0.03	0.05	0.19	−0.13	−0.22	−0.55	0.16	1.13	0.38	−0.05	−0.03
	(0.03)	(0.10)	(0.18)	(0.06)	(0.15)	(0.21)	(0.04)	(0.10)	(0.17)	(0.04)	(0.19)
UNEMP(τ_L, 2)	0.03	0.03	0.10	0.06	−0.02	−0.02	−0.09	0.00	0.89	−0.05	0.33
	(0.01)	(0.04)	(0.07)	(0.02)	(0.06)	(0.10)	(0.01)	(0.04)	(0.09)	(0.01)	(0.07)
IP(τ_L, 3)	−0.54	−0.17	0.59	−0.12	−0.49	0.04	1.23	0.42	−0.41	0.43	−0.38
	(0.12)	(0.27)	(0.44)	(0.19)	(0.37)	(0.47)	(0.12)	(0.26)	(0.39)	(0.13)	(0.51)
INFL(τ_L, 3)	0.04	0.01	0.16	−0.12	−0.06	−0.33	0.18	1.01	0.18	−0.10	0.09
	(0.04)	(0.13)	(0.24)	(0.07)	(0.20)	(0.29)	(0.05)	(0.13)	(0.24)	(0.05)	(0.26)
UNEMP(τ_L, 3)	0.04	0.03	0.14	0.07	−0.04	−0.06	−0.10	0.03	0.87	−0.09	0.50
	(0.02)	(0.04)	(0.09)	(0.03)	(0.07)	(0.12)	(0.02)	(0.05)	(0.10)	(0.02)	(0.09)
GDP(τ_L)	−0.30	0.01	0.03	−0.04	0.00	0.01	0.35	−0.06	0.02	0.82	0.33
	(0.04)	(0.06)	(0.03)	(0.08)	(0.06)	(0.04)	(0.05)	(0.06)	(0.04)	(0.06)	(0.23)

(a) First month.
(b) Second month.
(c) Third month.

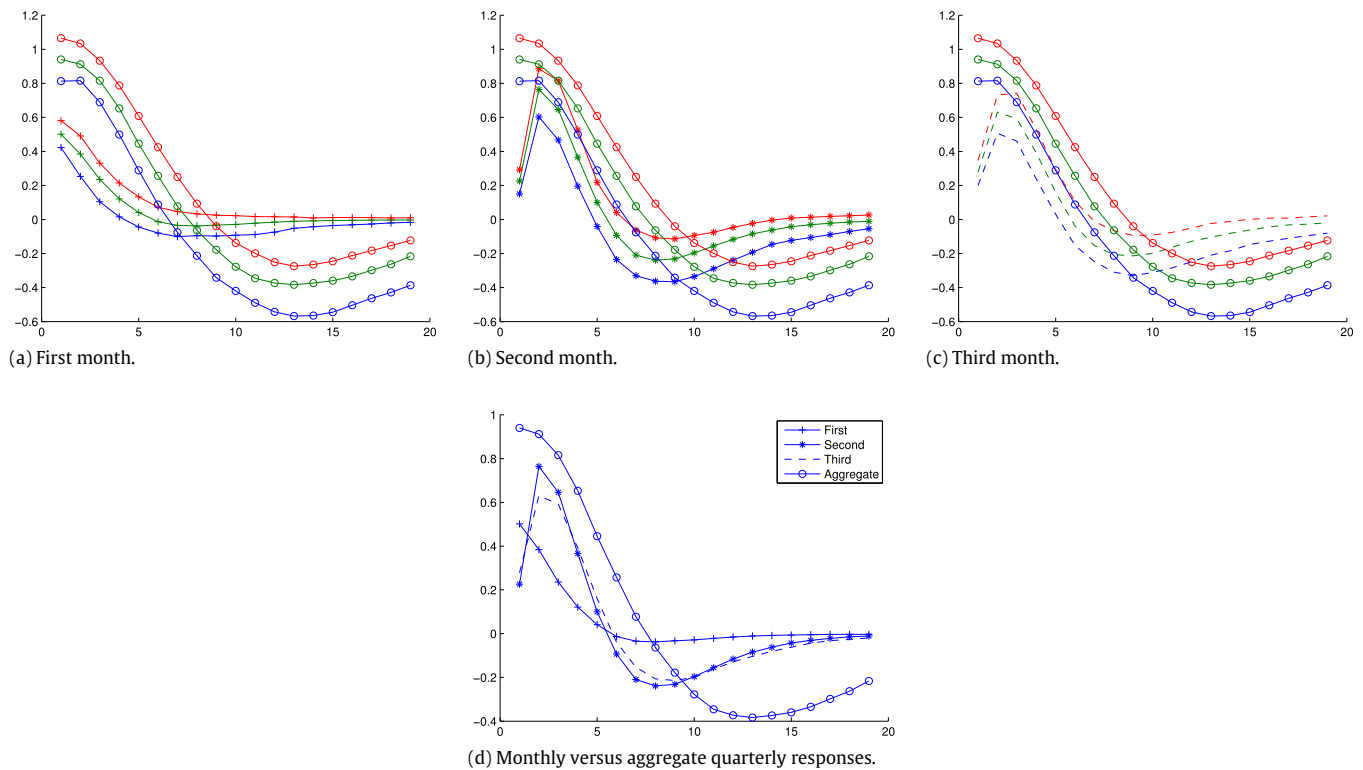(d) Monthly versus aggregate quarterly responses.

Fig. 5. Response function of GDP to IP shocks—BVAR and mixed frequency BVAR comparison. The plot displays a comparison of impulse response functions obtained from the BVAR model estimates reported in Table 4 with the corresponding three monthly impulse response functions obtained from the mixed frequency BVAR estimates reported in Table 5. Monthly versus aggregate quarterly impulse responses with confidence bands.

frequency VAR yields more precise inference—while the impulse responses remain the same as in the quarterly data VAR.

The numerical examples in the previous section suggest that this is not what we expect to happen in a mixed frequency VAR. Indeed, the numerical examples indicated that we may expect quite different impulse response functions with traditional VAR models and mixed frequency ones. Moreover, the numerical examples in the previous section told us that shocks occurring in different months will have different impacts, even in a non-periodic setting.

We observe all these patterns, different IRFs between low and mixed frequency VAR models and IRFs that differ across the month of the quarter, with our empirical results. This is indeed what we observe from Figs. 5 through 7. Consider first a shock of IP onto GDP growth appearing in Fig. 5. The mixed frequency impulse response functions look different from the quarterly data VAR specification. The latter displays a strong initial response of GDP followed by a negative impact after roughly two years. None of the mixed frequency impulse response functions show neither such a strong initial response, nor such a strong subsequent negative impact. The mixed frequency impulse response functions also die out more quickly. There appears to be a slightly different pattern for the first month's impulse response, but judging by the confidence bands this looks line it may not be significantly different from the other two months. The mixed frequency specification also imply a more delayed impact of shocks.

The case of inflation shocks reported in Fig. 6 is also quite interesting. The mixed frequency VAR model tells us that the impact of inflation shocks is close to zero—and they also look quite similar across the three months. The low frequency VAR model shows a very different, and arguably implausible pattern of a much stronger negative and long lasting impact on GDP. The only case were the low and mixed frequency VAR models appear to agree appears in Fig. 7 where unemployment shocks show a remarkable similar pattern across all specifications, and also across all months of the quarter.

8. Conclusions

In this paper we introduced a class of mixed frequency VAR models that are in may important ways very close to traditional VAR models. Unlike the bulk of the literature on mixed frequency models, we do not resort to latent variable/shock representations—but instead follow the standard observable shock paradigm. The use of many standard tools in VAR analysis easily apply to our setting, and in fact some tools—most prominently Cholesky factorizations—become even more suitable for policy analysis. The estimation of the new class of models also relies on well established procedures. In addition, the absence of hidden shocks avoids complications in terms of estimation and filtering.

Another merit of the paper is that it establishes a relationship between hitherto completely disconnected literatures: (1) the vast literature on VAR models and (2) the arguably more obscure literature on periodic models (for seasonality). One might indeed be tempted to characterize the connection as Hansen and Sargent meeting Sims, since the former introduced periodic models into the macro literature. It may also be appropriate to characterize the contributions in this paper as establishing a link between on the one hand MIDAS and on the other hand part of the literature on seasonal time series models.

Acknowledgments

(a) First month.   (b) Second month.   (c) Third month.

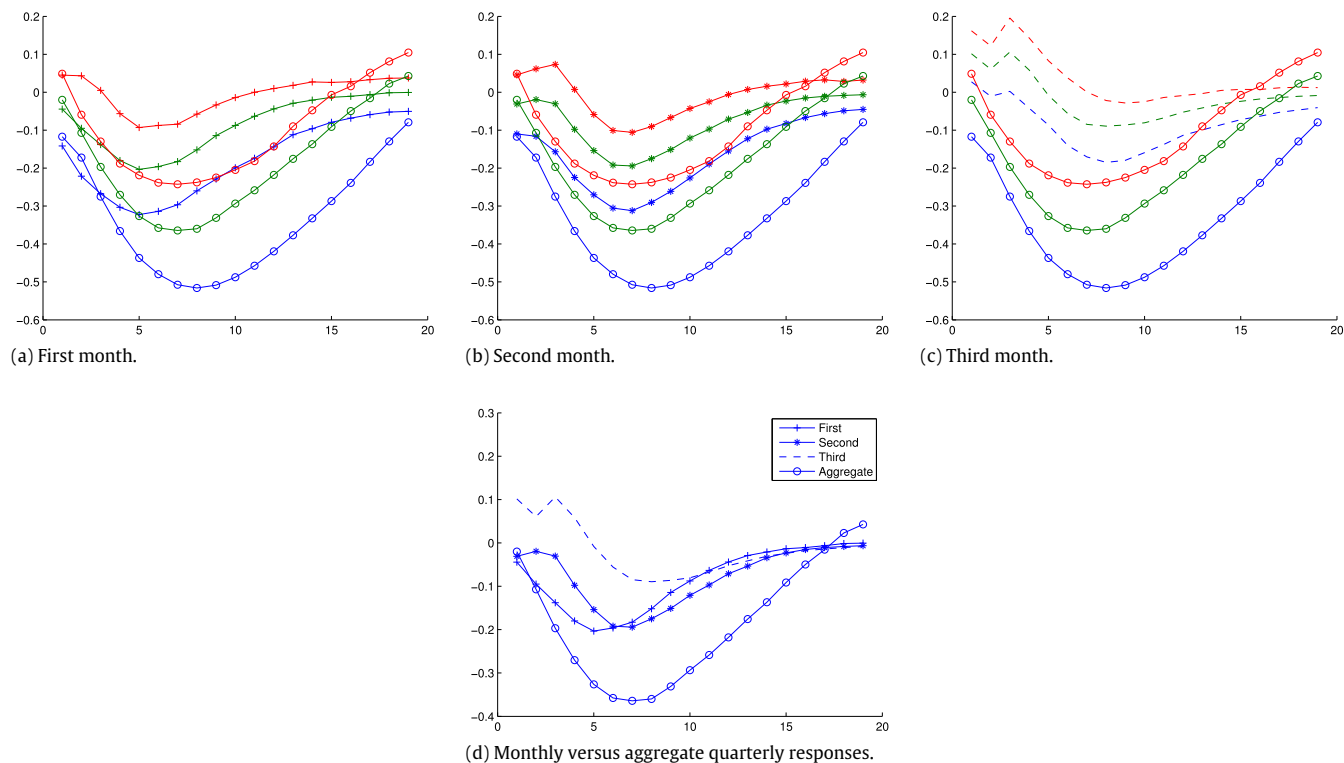(d) Monthly versus aggregate quarterly responses.

Fig. 6. Response function of GDP to INFL shocks—BVAR and mixed frequency BVAR comparison. The plot displays a comparison of impulse response functions obtained from the BVAR model estimates reported in Table 4 with the corresponding three monthly impulse response functions obtained from the mixed frequency BVAR estimates reported in Table 5. Monthly versus aggregate quarterly impulse responses with confidence bands.
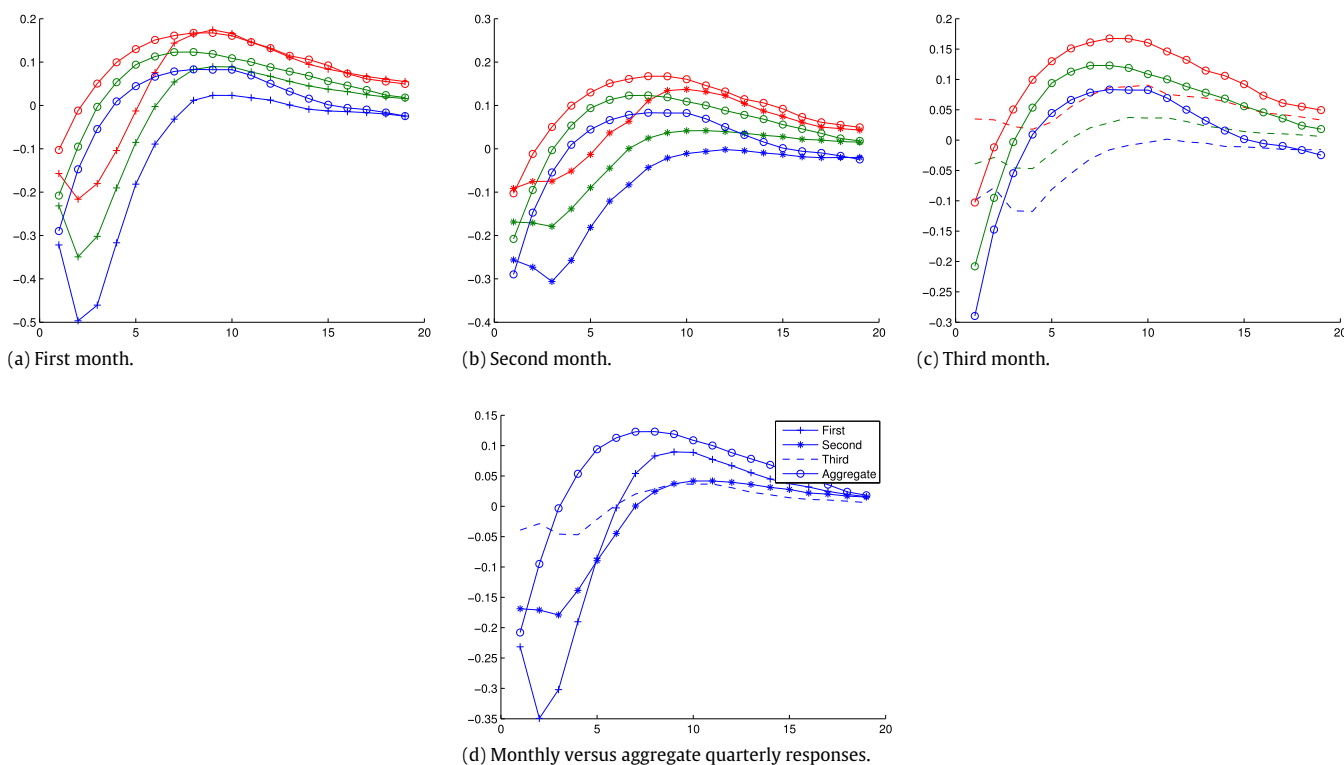

(a) First month.   (b) Second month.   (c) Third month.

(d) Monthly versus aggregate quarterly responses.

Fig. 7. Response function of GDP to UNEMP shocks—BVAR and mixed frequency BVAR comparison. The plot displays a comparison of impulse response functions obtained from the BVAR model estimates reported in Table 4 with the corresponding three monthly impulse response functions obtained from the mixed frequency BVAR estimates reported in Table 5. Monthly versus aggregate quarterly impulse responses with confidence bands.

Appendix

Technical appendix

Appendix A. Parsimony—Details

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

1. Normalized beta probability density function,

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

$$w_i^r(\gamma) = w_i(1, \gamma_2), \qquad (A.2)$$

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

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

$$w_i^r = w_i(\gamma_1, 0). \qquad (A.4)$$

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

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

Note that this can also be written in matrix form:

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

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)

$$\beta w_i(\gamma_1, \ldots, \gamma_P) = \gamma_1 I_{i \in [a_0, a_1]} + \sum_{p=2}^{P} \gamma_p I_{i \in (a_{p-1}, a_p]}$$

$$a_0 = 1 < a_1 < \cdots < a_P = N \qquad (A.7)$$

$$I_{i \in [a_{p-1}, a_p]} = \begin{cases} 1, & a_{p-1} \le i \le a_p \\ 0, & \text{otherwise} \end{cases}$$

where $a_0 = 1 < a_1 < \cdots < a_P = N$. The step functions approach to MIDAS appeared in Ghysels et al. (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 et al. (2015) is shown to work for small values of m.

A so called *multiplicative* ADL MIDAS regression specifications is also suggested in Andreou et al. (2010). Taking the last equation in (2.9) we have:

$$x_L(\tau_L) = \sum_{j=1}^{P} \alpha_j + \sum_{j=1}^{P} b_j x_H(\tau_L - j)(\gamma) + \varepsilon(\tau_L, m+1)$$

$$x_H(\tau_L - j)(\gamma) \equiv \sum_{k=1}^{m} w(\gamma)_k x_H(\tau_L - j, k) \qquad (A.8)$$

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

Appendix B. Proof of Proposition 5.1

We start with listing the regularity conditions. We assume the DGP is the $m * K$ dimensional vector $\ddot{x}(\tau_L)$ described by Eq. (4.6):

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

with $E[\ddot{\varepsilon}(\tau_L)\ddot{\varepsilon}(\tau_L)'] = \ddot{\mathbb{C}}\ddot{\mathbb{C}}'$.

Following Assumption 4.1, the above DGP is covariance stationary and therefore has a spectral representation. In particular:

Assumption B.1. The process $\ddot{x}(\tau_L)$ satisfy Assumptions 2.1 and 4.1 and has spectral density $\ddot{S}(z)$ for $z = \exp(-i\omega)$ with $\omega \in [0, \pi]$, which can be written as:

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

Against this DGP we consider:

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

with spectral density:

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

and:

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

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

Assumption B.2. The parameter spaces Δ_Φ and Δ_Ψ are compact subsets of a Euclidean space.

Assumption B.3. The spectral densities $\underline{S}(z, \Psi)$ and $\bar{S}(z, \Phi)$ 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 $0 < \varepsilon_l < \varepsilon_u$: $\varepsilon_l I \le \underline{S}(\exp(-i\omega), \Psi) \le \varepsilon_u I$ and $\varepsilon_l I \le \bar{S}(\exp(-i\omega), \Phi) \le \varepsilon_u I$ for respectively each $(\omega, \Psi) \in [-\pi, \pi] \times \Delta_\Psi$ and $(\omega, \Phi) \in [-\pi, \pi] \times \Delta_\Phi$, $\bar{S}(\exp(i\omega), \Psi)$ is the complex conjugate of $\underline{S}(\exp(-i\omega), \Psi)$ and $\bar{S}(\exp(i\omega), \Phi)$ is the complex conjugate of $\bar{S}(\exp(-i\omega), \Phi)$.

Assumption B.4. $\mu_{\bar{x}}$ is a continuous function on the domain of Δ_Φ. $\mu_{\ddot{x}}$ is a continuous function on the domain of Δ_Ψ.

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 δ against DGP with spectral density S, that the maximum likelihood estimator minimizes the Kullback and Leibler (1951) information criterion which can be written as:

$$E(G(\delta), S) = E_1(G(\delta), S) + E_2(G(\delta), S) + E_3(G(\delta), S)$$

$$E_1(G(\delta), S) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log(det G(\exp(-i\omega)))d\omega$$

$$E_2(G(\delta), S) = \frac{1}{2\pi} \int_{-\pi}^{\pi} trace(detG(\delta, \exp(-i\omega))^{-1}$$
$$\times S(\exp(-i\omega)))d\omega$$

$$E_3(G(\delta), S) = (\ddot{\mu}_{\ddot{x}} - \nu(\delta))'G(0)^{-1}(\ddot{\mu}_{\ddot{x}} - \nu(\delta)) \tag{B.9}$$

using results from Akaike (1973), Ljung (1978), White (1982) and Pötscher (1987). The results in Eqs. (5.4) follows by substituting $G(\delta)$ with $\underline{S}(\Psi)$ and the corresponding mean in E_3. For the result in Eq. (5.5) we need to the same for the standard low frequency VAR. Note that the spectral density S in each differs, as it is affected by the aggregation schemes appearing in Eqs. (4.8).

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Second International Symposium on Information Theory, vol. 1. Springer Verlag, pp. 267–281.

Andreou, E., Ghysels, E., Kourtellos, A., 2010. Regression models with mixed sampling frequencies. J. Econometrics 158, 246–261.

Aruoba, S.B., Diebold, F.X., Scotti, C., 2009. Real-time measurement of business conditions. J. Bus. Econom. Statist. 27, 417–427.

Bai, J., Ghysels, E., Wright, J., 2013. State space models and MIDAS regressions. Econometric Rev. 32, 779–813.

Bernanke, B., Gertler, M., Watson, M., 1997. Systematic monetary policy and the effects of oil price shocks. Brookings Pap. Econ. Activity 1, 91–157.

Breitung, J., Swanson, N., 2002. Temporal aggregation and spurious instantaneous causality in multiple time series models. J. Time Series Anal. 23, 651–665.

Chauvet, M., Goetz, T., Hecq, A., 2013. Realized Volatility and Business Cycle Fluctuations: A Mixed-Frequency VAR Approach. Discussion Paper. University of Maastricht.

Chen, X., Ghysels, E., 2011. NewsGood or bad and its impact on volatility predictions over multiple horizons. Rev. Financ. Stud. 24–81, 46.

Chib, S., Greenberg, E., 1995. Understanding the metropolis-hastings algorithm. Amer. Statist. 327–335.

Chiu, C., Eraker, B., Foerster, A., Kim, T., Seoane, H., 2011. Estimating VARs Sampled at Mixed or Irregular Spaced Frequencies: A Bayesian Approach. Federal Reserve Bank of Kansas City, RWP 11-11.

Cox, D., 1981. Statistical analysis of time series: Some Recent developments [with discussion and reply]. Scand. J. Stat. 8, 93–115.

Davies, R., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33–43.

Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projection using realistic prior distributions. Econometric Rev. 3, 1–100.

Foroni, C., Marcellino, M., 2014. Mixed-Frequency structural models: Identification, estimation, and policy analysis. J. Appl. Econometrics 29, 1118–1144.

Foroni, C., Marcellino, M., Schumacher, C., 2015. Unrestricted mixed data sampling (MIDAS) MIDAS regressions with unrestricted lag polynomials. J. R. Stat. Soc. Ser. A 178, 57–82.

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. J. Bus. Econom. Statist. 289–298.

Ghysels, E., Mira, A., Solgi, R., 2014. A General Bayesian MIDAS Regression Analysis with Application to Data Frequency Selection. Discussion Paper. UNC.

Ghysels, E., Owyang, M., 2011. Taylor Monetary Policy Rules with Financial Market Expectations. Discussion Paper. Federal Reserve Bank of Saint-Louis and UNC.

Ghysels, E., Santa-Clara, P., Valkanov, R., 2006. Predicting volatility: getting the most out of return data sampled at different frequencies. J. Econometrics 131, 59–95.

Ghysels, E., Sinko, A., Valkanov, R., 2006. MIDAS regressions: Further results and new directions. Econometric Rev. 26, 53–90.

Ghysels, E., Wright, J., 2009. Forecasting professional forecasters. J. Bus. Econom. Statist. 27, 504–516.

Gladyshev, E., 1961. Periodically correlated random sequences. Sov. Math. 2, 385–388.

Granger, C., 1980. Testing for causality: A personal viewpoint. J. Econ. Dyn. Control. 2, 329–352.

Granger, C., 1988. Some recent developments in a concept of causality. J. Econometrics 39, 199–211.

Granger, C., 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., Sargent, T., 1990. Recursive Linear Models of Dynamic Economies. NYU and University of Chicago.

Hansen, L., Sargent, T., 1993. Seasonality and approximation errors in rational expectations models* 1. J. Econometrics 55 (1–2), 21–55.

Harvey, A.C., Pierse, R.G., 1984. Estimating missing observations in economic time series. J. Amer. Statist. Assoc. 79, 125–131.

Kadiyala, K., Karlsson, S., 1997. Numerical methods for estimation and inference in Bayesian VAR-models. J. Appl. Econometrics 12, 99–132.

Kullback, S., Leibler, R., 1951. On information and sufficiency. Ann. Math. Stat. 22, 79–86.

Kuzin, V., Marcellino, M., Schumacher, C., 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. J. Bus. Econom. Statist. 4, 25–38.

Ljung, L., 1978. Convergence analysis of parametric identification methods. IEEE Trans. Automat. Control 23, 770–783.

Lütkepohl, H., 1987. Forecasting Aggregated Vector ARMA Processes. Springer-Verlag, Berlin.

Lütkepohl, H., 1993. Testing for causation between two variables in higher dimensional VAR models. In: Schneeweiss, H., Zimmerman, K. (Eds.), Studies in Applied Econometrics. Springer-Verlag, Heidelberg, p. 75.

Marcellino, M., Schumacher, C., 2010. Factor MIDAS for nowcasting and forecasting with ragged-edge data: A model comparison for German GDP. Oxf. Bull. Econ. Stat. 72, 518–550.

Mariano, R.S., Murasawa, Y., 2003. A new coincident index of business cycles based on monthly and quarterly series. J. Appl. Econometrics 18, 427–443.

McCallum, B., 1976. Rational expectations and the estimation of econometric models: an alternative procedure. Internat. Econom. Rev. 17, 484–490.

McCrorie, J., Chambers, M., 2006. Granger causality and the sampling of economic processes. J. Econometrics 132 (2), 311–336.

Mittnik, S., Zadrozny, P.A., 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. J. Forecast. 24, 575–592.

Pagano, M., 1978. On periodic and multiple autoregressions. Ann. Statist. 1310–1317.

Pettenuzzo, D., Timmermann, A.G., Valkanov, R.I., 2014. A Bayesian MIDAS Approach to Modeling First and Second Moment Dynamics. Available at SSRN 2471287.

Priestley, M., 1981. Spectral Analysis and Time Series. Academic Press, New York.

Pötscher, B., 1987. Convergence results for maximum likelihood type estimators in multivariable ARMA models. J. Multivariate Anal. 21, 29–52.

Renault, E., Sekkat, K., Szafarz, A., 1998. Testing for spurious causality in exchange rates. J. Empir. 5, 47–66.

Rodriguez, A., Puggioni, G., 2010. Mixed frequency models: Bayesian approaches to estimation and prediction. Int. J. Forecast. 26, 293–311.

Schorfheide, F., Song, D., 2013. Real-time Forecasting with a Mixed-frequency VAR. Discussion Paper. National Bureau of Economic Research.

Sims, C., 1980. Macroeconomics and reality. Econometrica 48 (1), 1–48.

Sims, C., Zha, T., 1998. Bayesian methods for dynamic multivariate models. Internat. Econom. Rev. 39, 949–968.

Sinko, A., Sockin, M., Ghysels, E., 2010. Matlab Toolbox for MIDAS Regressions. Available at http://www.unc.edu/~eghysels/Software_datasets.html.

Tiao, G., Grupe, M., 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 1–25.

Zadrozny, P.A., 1990. Forecasting US GNP at monthly intervals with an estimated bivariate time series model. Econ. Rev. Fed. Reserve Bank Atlanta 75, 2–15.