



UPPSALA
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Social Sciences 170*

VAR Models, Cointegration and Mixed-Frequency Data

SEBASTIAN ANKARGREN



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2019

ISSN 1652-9030
ISBN 978-91-513-0734-3
urn:nbn:se:uu:diva-391500

Dissertation presented at Uppsala University to be publicly examined in Hörsal 2, Kyrkogårdsgatan 10, Uppsala, Friday, 11 October 2019 at 13:15 for the degree of Doctor of Philosophy. The examination will be conducted in English. Faculty examiner: Dr. techn. Gregor Kastner (Vienna University of Economics and Business).

Abstract

Ankargren, S. 2019. VAR Models, Cointegration and Mixed-Frequency Data. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences* 170. 45 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-0734-3.

This thesis consists of five papers that study two aspects of vector autoregressive (VAR) modeling: cointegration and mixed-frequency data.

Paper I develops a method for estimating a cointegrated VAR model under restrictions implied by the economy under study being a small open economy. Small open economies have no influence on surrounding large economies. The method suggested by Paper I provides a way to enforce the implied restrictions in the model. The method is illustrated in two applications using Swedish data, and we find that differences in impulse responses resulting from failure to impose the restrictions can be considerable.

Paper II considers a Bayesian VAR model that is specified using a prior distribution on the unconditional means of the variables in the model. We extend the model to allow for the possibility of mixed-frequency data with variables observed either monthly or quarterly. Using real-time data for the US, we find that the accuracy of the forecasts is generally improved by leveraging mixed-frequency data, steady-state information, and a more flexible volatility specification.

The mixed-frequency VAR in Paper II is estimated using a state-space formulation of the model. Paper III studies this step of the estimation algorithm in more detail as the state-space step becomes prohibitive for larger models when the model is employed in real-time situations. We therefore propose an improvement of the existing sampling algorithm. Our suggested algorithm is adaptive and provides considerable improvements when the size of the model is large. The described approach makes the use of large mixed-frequency VARs more feasible for nowcasting.

Paper IV studies the estimation of large mixed-frequency VARs with stochastic volatility. We employ a factor stochastic volatility model for the error term and demonstrate that this allows us to improve upon the algorithm for the state-space step further. In addition, regression parameters can be sampled independently in parallel. We draw from the literature on large VARs estimated on single-frequency data and estimate mixed-frequency models with 20, 34 and 119 variables.

Paper V provides an R package for estimating mixed-frequency VARs. The package includes the models discussed in Paper II and IV as well as additional alternatives. The package has been designed with the intent to make the process of specification, estimation and processing simple and easy to use. The key functions of the package are implemented in C++ and are available for other packages to use and build their own mixed-frequency VARs.

Keywords: vector error correction, small open economy, mixed-frequency data, Bayesian, steady state, nowcasting, state-space model, large VARs, simulation smoothing, factor stochastic volatility, R

Sebastian Ankargren, Department of Statistics, Uppsala University, SE-75120 Uppsala, Sweden.

© Sebastian Ankargren 2019

ISSN 1652-9030

ISBN 978-91-513-0734-3

urn:nbn:se:uu:diva-391500 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-391500>)

*In theory, there is no difference
between theory and practice.
In practice, there is.*

Benjamin Brewster, 1882

List of papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

- I Ankargren, S. and J. Lyhagen (2019) Estimating a VECM for a Small Open Economy.
- II Ankargren, S., Unosson, M. and Y. Yang (2019) A Flexible Mixed-Frequency Vector Autoregression with a Steady-State Prior.
- III Ankargren, S. and P. Jonéus (2019) Simulation Smoothing for Nowcasting with Large Mixed-Frequency VARs.
- IV Ankargren, S. and P. Jonéus (2019) Estimating Large Mixed-Frequency Bayesian VAR Models.
- V Ankargren, S. and Y. Yang (2019) Mixed-Frequency Bayesian VAR Models in R: The `mfbvar` Package.

Contents

1	Introduction	9
2	Research goals	11
3	Background	12
3.1	VAR models	12
3.2	Cointegrated VAR models	14
3.3	Bayesian VAR models	18
3.4	Mixed-frequency VAR models	23
3.5	Statistical computations in the R programming language	28
4	Summary of papers	32
4.1	Paper I	32
4.2	Paper II	32
4.3	Paper III	33
4.4	Paper IV	34
4.5	Paper V	35
5	Acknowledgments	37
	References	39

1. Introduction

What will happen in the future? If it were possible to answer that question with complete certainty, many statisticians and econometricians would be out of work. Prediction and forecasting are two of the general themes in statistics and related disciplines that are the reason why many researchers as well as people in industry to go to work every day. The nature of the predictions may vary greatly—from what the next trend in music will be to when emergency rooms will need to be fully staffed—but in the end all of them attempt to make an educated guess about an unknown value.

Forecasting plays a central role in many economic decisions that affect everyone. On the fiscal side, knowing where the economy is headed is crucial to the government when planning its budget and possible changes to taxes and expenditures. On the monetary side, whether the repo rate should be altered or not is highly dependent on the economic outlook, what is happening to inflation, and imminent economic threats. For these reasons the Swedish Ministry of Finance and Sveriges Riksbank forecast key economic variables so that the decision-makers can make informed decisions with as few adverse effects as possible.

A central stylized fact in statistics, a science of information, is that *the more you know the better*. In statistics jargon, this translates into *the more data you have the better*. Typically, in data sets used in traditional statistical analyses there will be missing values, which means that for some of the observed units—individuals, countries, companies—one or more of the variables—such as age or educational level—have not been recorded. Arguably, the data that exist contain valuable information, but statistical analysis is not straight-forward owing to the aforementioned missing values. However, to squeeze every drop of information out of the data statisticians may sometimes go to great lengths to capture what is in the values that are observed through what is known as imputation.

Surprisingly, however, the idea that more information is better and should be preferred is not at all as predominant in the field of time series econometrics, a subject at the intersection of statistical analysis of time series and economics—in short, the study of economic series over time. Indeed, the main issue that four out of five papers collected in this thesis deal with is the fundamental fact that popular economic measures, such as the growth rate of gross domestic product (GDP) and the inflation rate, are typically observed at different frequencies. In particular, the GDP growth rate is observed once a quarter, whereas the inflation rate is based on the consumer price index which, in turn, is updated and published once a month.

Standard practice in applied analyses is to aggregate to the lowest common frequency, which means that the monthly inflation rate in the preceding example is transformed into a quarterly inflation rate. The transformation is often performed by taking the average of the three months within every quarter, but other aggregations (such as taking the last value within the quarter) are also frequently used.

As is argued in the papers in this thesis, going from a monthly to a quarterly time series incurs an unnecessary loss of information. The reason why the conversion is often carried out is because it simplifies the subsequent step of estimating a model substantially. The methods developed in this thesis are admittedly more involved than standard methods, but they rely on data to a larger extent. From a purely intuitive perspective, keeping the different types of data at their original frequencies without lumping parts of the data together just to simplify estimation of a model is a more sensible approach. In fact, when non-statisticians hear about this issue, the response is often a somewhat confused "I would have assumed you already did that?". The work in this thesis is a step toward doing what others expect us to have been doing all along.

The first paper in the thesis differs slightly from the remaining four as it deals with cointegration, the situation when certain relations between variables are stable while the variables themselves may not be. Nevertheless, the particular issue under consideration fits into the more general idea of using more information, as previously discussed. In particular, Sweden, and many other countries, fall into the category of small open economies. When small open economies are modeled, it is often the case that variables capturing relevant large economies are included in the model. Conceptually, there is in such situations only feedback from the large economy to the small and not vice versa almost by definition. Imposing this one-sided feedback is not standard in the literature, but by being able to do so more information is effectively leveraged in the model.

2. Research goals

Prior to commencing my PhD studies, I worked with VAR models at the Ministry of Finance and have constantly been engaged in more applied work (Ankargren et al., 2017, 2018; Ankargren and Shahnazarian, 2019). These experiences have influenced my thesis work in several ways. First, I want my research to be of practical use. Second, it became clear to me that interpretability is essential if a method is to stand any chance of being used. Third, I need to make my research accessible if people are to use it.

The first point is part of the reason for Paper I–IV. Paper I answers the question: How can we estimate a cointegration model for a small open economy such as Sweden? While cointegration is well-studied, the question of how to do it in a way that resonates with the notion of a small open economy has not been addressed. Paper II answers a different question, namely: How can we estimate one of the most common macroeconomic models used at Swedish government institutions when the frequencies of the data are mixed? This extension is of high practical relevance as it improves on a commonly used model by making it more in line with the nature of the data. Paper III–IV address the question: How can we bring the mixed-frequency models into the high-dimensional regime? "Big data" are everywhere, and so too in macroeconomic forecasting. Paper III makes existing methodology for mixed-frequency models feasible under high-dimensional settings, whereas Paper IV makes further improvements that enable us to estimate mixed-frequency models of dimensions previously unseen in the literature.

As for the second point, one could make the argument that with an abundance of (possibly mixed-frequency) data, why not just use methods from, e.g., the machine learning literature? The drawback is that if macroeconomic forecasters are unable to explain their forecasts, they will not use them. The implication is that a method may be superior in terms of predictive ability, but that is all in vain if you are incapable of putting the forecast into a larger story. VAR models are far from perfect, but the advantage is that people are used to them and know how to analyze them. It is for this reason that Paper II–IV develop methods for mixed-frequency forecasting using VAR models.

The third point is the rationale for including Paper V in the thesis. Few forecasters have the time—or the experience—to implement econometric models. Paper V therefore simplifies the issue of implementation by providing an R package with user-friendly functions for estimating the mixed-frequency VAR models. The goal is to provide an accessible way for anyone interested to try this class of models. R is an open-source programming language that is freely distributed, making the package available to virtually anyone.

3. Background

3.1 VAR models

Modern time series econometrics is largely influenced by work that was carried out almost half a century ago by Box and Jenkins (1970), who popularized the use of time series models with autoregressive and moving average terms. The idea that many economic time series can be well approximated by finite-order linear models with autoregressive and moving average components is still prevalent today. Needless to say, the field has evolved dramatically, but linear models with autoregressive and moving average components are still central foundations in modern time series econometrics. In part because models with moving average components are more difficult to estimate, purely autoregressive models are even more popular. Not only are they ubiquitous in mainstream time series analysis, but assuming an autoregressive structure as an approximation is also common in other fields.

Before the 1980s, macroeconomists mainly used large, structural models. In a seminal paper, Sims (1980) criticized this practice and advocated the use of vector autoregressions (VARs). The argument is that VARs provide a more data-dependent alternative to models driven heavily by theory. The nature of VARs as a complement to theory-heavy models still remains today when dynamic stochastic general equilibrium (DSGE) models constitute the main workhorse in structural empirical macroeconomics.

VARs thus still have a natural role today as a complement to more structural models. In fact, it is common to use both types of models in day-to-day work. For example, in preparing its forecasts for the repo rate decisions the Riksbank uses VARs with the steady-state prior developed by Villani (2009), which is the focus of Paper II, and the DSGE model RAMSES presented in Adolfson et al. (2007), which is currently in its third installment. Iversen et al. (2016) discussed the forecasting round at the Riksbank in depth and it is interesting to note that, historically, the VAR shows better forecasting performance than both the published forecasts and the DSGE forecasts (Reslow and Lindé, 2016).

A standard VAR model is in its simplest form a multivariate regression model and has been thoroughly discussed by, e.g., Hamilton (1994) and Lütkepohl (2005). It can be formulated as

$$x_t = c + \sum_{i=1}^p \Phi_i x_{t-i} + \varepsilon_t, \quad (3.1)$$

where x_t is an $n \times 1$ vector of variables, c an $n \times 1$ vector of intercepts, Φ_i are $n \times n$ regression parameter matrices and ε_t is an error term for which it is

typically assumed that $\varepsilon_t \sim N(0, \Sigma)$, where Σ is a positive-definite covariance matrix and the sequence $\{\varepsilon_t\}$ is independent over time. The assumption of a constant Σ is relaxed in Paper II and IV.

An equivalent formulation of the VAR that will be useful later is obtained by writing the model as a VAR(1) with restrictions. This restricted VAR(1) form is called the companion form (Hamilton, 1994):

$$\begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p} \end{pmatrix} = \underbrace{\begin{pmatrix} \Phi_1 & \Phi_2 & \dots & \Phi_p \\ & I_{n(p-1)} & & 0 \end{pmatrix}}_{F(\Phi)} \begin{pmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Apart from forecasting, VARs are often employed as policy tools (see Stock and Watson (2001) for an accessible introduction). Impulse responses, historical decompositions and scenario analyses are some of the most common goals of VAR modeling.

Impulse response analysis attempts to answer the question of what the effect of a shock is on current or future values of the variables in the model. If we assume that the VAR in (3.1) is stable so that the lag polynomial is invertible, then the VAR permits a representation as an infinite-order vector moving average (VMA) process¹:

$$x_t = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i},$$

where Ψ_i are moving average weights. These moving average weights can be calculated as $\Psi_i = F(\Phi)_{11}^{(i)}$, where $F(\Phi)_{11}^{(i)}$ is the upper-left $n \times n$ block of $F(\Phi)$ raised to the power i , and $\Psi_0 = I_n$.

From the VMA representation, it is easy to find the response of variable i at time $t+h$ to a one-unit change in the j th shock at time t for $h = 0, 1, \dots$ as:

$$\frac{\partial x_{i,t+h}}{\varepsilon_{j,t}} = \Psi_h^{(i,j)}.$$

That is, the full matrix Ψ_h provides the responses of all variables to all shocks. The results are usually presented for $\Psi_h^{(i,j)}$ with (i, j) fixed and as a function of h over, say, a couple of years.

What is different for models intended to provide e.g. impulse responses is that to be economically meaningful, a necessary first step is to structurally identify the model to obtain a structural VAR (SVAR). The SVAR is

$$A_0 x_t = d + \sum_{i=1}^p A_i x_{t-i} + u_t,$$

¹The result is a multivariate version of Wold's decomposition theorem (Wold, 1938), see Hamilton (1994).

where $d = A_0 c$, A_0 is invertible, $A_i = A_0 \Phi_i$, $A_0 \varepsilon_t = u_t \sim N(0, I_n)$ and $\Sigma = A_0^{-1} (A_0^{-1})'$; for a textbook treatment, see Kilian and Lütkepohl (2017).

The reason for moving to a SVAR is that, to provide responses to interpretable shocks, the shocks must be disentangled from one another—i.e., orthogonalized. For example, to study the effect of monetary policy shocks on, e.g., inflation, the monetary policy shock must first be identified and separated from other shocks.² Exact identification of the structural model means that A_0 is uniquely identified. Common ways of achieving identification is by means of Cholesky decomposing $\Sigma = PP'$ and letting $A_0 = P^{-1}$. It is also possible to set-identify the structural model by imposing sign or zero restrictions (see Uhlig, 2005 for a seminal contribution, Fry and Pagan, 2011 for a review and Arias et al., 2018 for a recent important methodological advancement), where A_0 is obtained by rotating P^{-1} by an orthogonal matrix³ Q in such a way that certain restrictions imposing no or positive (negative) responses are satisfied. It is also possible to exploit model heteroskedasticity (Lütkepohl and Velinov, 2016) and external data such as high-frequency data and external instruments. See Kilian and Lütkepohl (2017), Chap. 15.

3.2 Cointegrated VAR models

An important concept for VAR modeling is that of *cointegration*. Loosely speaking, cointegration is to be understood as a phenomenon that restricts certain linear combinations of variables from drifting away from each other while the variables themselves may drift arbitrarily. Central to cointegration is that the stochastic process under consideration is integrated of some order larger than zero. To this end, we first define a process that is integrated of order zero.

Definition 1 (Johansen, 1995) *A stochastic process x_t that satisfies $x_t - E(x_t) = \sum_{i=0}^{\infty} C_i \varepsilon_{t-i}$ is called $I(0)$ —integrated of order zero—if $C = \sum_{i=0}^{\infty} C_i \neq 0$ and $C(z) = \sum_{i=0}^{\infty} C_i z^i$ is convergent for $|z| \leq 1 + \delta$, where $\delta > 0$.*

The definition, from Johansen (1995), establishes that the stochastic part of x_t can be described by a linear process where the infinite-dimensional moving average weights do not sum to zero. For example, a random walk is not $I(0)$ since $C_i = 1$, whereby $C(z)$ is not convergent. However, $I(0)$ is not directly interchangeable with (weak) stationarity. If $x_t = \varepsilon_t - \theta \varepsilon_{t-1}$, then $C_0 = 1$, $C_1 = -\theta$ and $C_i = 0$ for $i = 2, 3, \dots$. Thus, $C = 1 - \theta$ and if $\theta = 1$ we obtain $C = 0$. The interpretation is that even an MA(1)—which is weakly stationary—can fail to be $I(0)$.

²Christiano et al. (1999) provided an overview of monetary policy analysis in SVARs that discusses the issue of identification in detail; the book by Kilian and Lütkepohl (2017) goes deeper into the issue of SVARs more generally.

³An orthogonal matrix Q satisfies $Q'Q = QQ' = I$.

The *order of integration* is the number of times a process must be differenced to be $I(0)$. Let Δ^d represent the difference operator with the property that $\Delta x_t = x_t - x_{t-1}$ and $\Delta^d x_t = \Delta^{d-1}(\Delta x_t)$. Then the following definition, also from Johansen (1995), defines the order of integration of a stochastic process.

Definition 2 (Johansen, 1995) *A stochastic process x_t is called integrated of order d , denoted by $I(d)$ for $d = 0, 1, 2, \dots$ if $\Delta^d[x_t - E(x_t)]$ is $I(0)$.*

The $I(0)$ and $I(1)$ cases dominate in applications.⁴ In the following, x_t is restricted to be integrated of at most order one. We can then define cointegration as follows.

Definition 3 (Johansen, 1995) *Let x_t be integrated of order one. We call x_t cointegrated with cointegrating vector $\beta \neq 0$ if $\beta'x_t$ is integrated of order zero. The cointegrating rank is the number of linearly independent cointegrating relations.*

The defining property of cointegration is that two processes may be individually $I(1)$, but certain linear combinations thereof are $I(0)$.⁵ To provide a concrete example, suppose that x_1 and x_2 are governed by the same random walk:

$$\begin{aligned} x_{1,t} &= \sum_{s=1}^t \varepsilon_{3,s} + \varepsilon_{1,t} \\ x_{2,t} &= \sum_{s=1}^t \varepsilon_{3,s} + \varepsilon_{2,t}, \end{aligned}$$

where $(\varepsilon_1, \varepsilon_2, \varepsilon_3)' \sim N(0, I_3)$. Both x_1 and x_2 are $I(1)$, but the difference

$$x_{1,t} - x_{2,t} = \varepsilon_{1,t} + \varepsilon_{2,t} \sim N(0, 2)$$

is integrated of order zero. Figure 3.1 plots the two variables as well as the difference between them. It illustrates an intuitive interpretation of cointegrating behavior: While the variables individually appear to possibly drift off arbitrarily, their difference is stable and stationary.

The error-correction formulation of the VAR is often employed to model cointegrating series. Work in this field was pioneered by Søren Johansen (see in particular Johansen, 1988; Johansen and Juselius, 1990; Johansen, 1991 and Johansen, 1995) following the prize-winning seminal papers Granger (1981); Engle and Granger (1987).

⁴Strictly speaking, d can take any non-negative fractional value and not necessarily only integers. A non-integer value gives rise to so-called fractional integration.

⁵The definition used here is purposely somewhat restrictive for ease of presentation, as cointegration could equally well occur if e.g. $x_t \sim I(2)$ and $\beta'x_t \sim I(1)$. The central feature is that the order of integration of $\beta'x_t$ is reduced, not that it is zero.

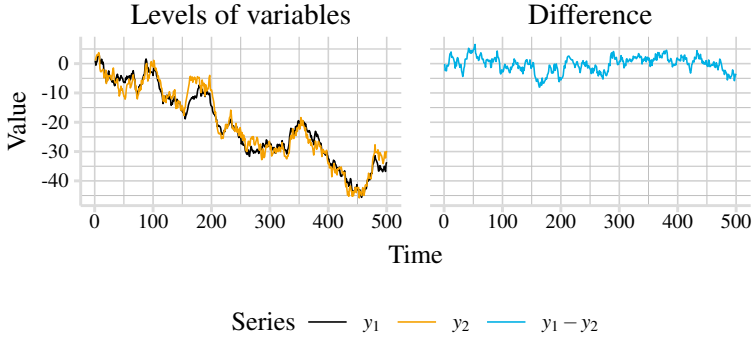


Figure 3.1. Illustration of cointegration. The series $x_{1,t}$ and $x_{2,t}$ individually exhibit drifting behavior (left panel), but the difference $x_{1,t} - x_{2,t}$ is stable around zero (right panel).

The vector error-correction model (VECM) formulation of a VAR model is simply a rearrangement of terms resulting in the representation

$$\Delta x_t = \Phi x_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta x_{t-i} + \varepsilon_t,$$

where any constant terms such as intercept and trends are omitted for simplicity. The new parameter matrices are related to (3.1) through $\Phi = \sum_{i=1}^p \Phi_i - I_n$ and $\Gamma_i = \sum_{j=i+1}^p \Phi_j$.

The existence of cointegration has certain implications for Φ , namely that it is of reduced rank. Let $r \leq n$ denote the *cointegrating rank*. Interchangeably, we say that there are r cointegrating relations. By this property, we can decompose

$$\Phi = \alpha \beta',$$

where α and β are $n \times r$ matrices of full rank. Two special cases are helpful in understanding the concept. The case when there are no cointegrating relations, i.e. $r = 0$, implies that $\sum_{i=1}^p \Phi_i = I_n$. There is, in other words, no linear combination of the variables that is $I(0)$, and there is no cointegration. Conversely, if $r = n$ then $\beta = I_n$ (up to a rotation) and x_t is in fact $I(0)$.

It can be shown that x_t has the representation

$$x_t = C \sum_{s=1}^t \varepsilon_s + C^*(L) \varepsilon_t + C x_0, \quad (3.2)$$

where $C = \beta_{\perp} (\alpha'_{\perp} \Gamma \beta_{\perp})^{-1} \alpha'_{\perp}$, α_{\perp} (β_{\perp}) is the orthogonal complement of α (β), $\Gamma = I_n - \sum_{i=1}^{p-1} \Gamma_i$, $C^*(L) \varepsilon_t = \sum_{i=0}^{\infty} C_i^* \varepsilon_{t-i}$ is an $I(0)$ process and x_0 is the initial value. While the representation, known as Granger's representation theorem,

is somewhat involved, it allows for a compelling mathematical explanation for the concept of cointegration. Recall from Definition 3 that cointegration is present if x_t is $I(1)$, but $\beta'x_t$ is $I(0)$. Heuristically, (3.2) contains a random walk component— $\sum_{s=1}^t \varepsilon_s$ —and x_t is $I(1)$. Premultiplying by β' , however, leaves only

$$\beta'x_t = \beta'C^*(L)\varepsilon_t,$$

as $\beta'C = 0$. Because $C^*(L)\varepsilon_t$ is a stationary process, so is $\beta'C^*(L)\varepsilon_t$.

The parameters of the model can be estimated by maximum likelihood using reduced rank regression (Anderson, 1951). Let

$$\begin{aligned} z_{0t} &= \Delta x_t \\ z_{1t} &= x_{t-1} \\ z_{2t} &= (\Delta x'_{t-1} \quad \cdots \quad \Delta x'_{t-p+1})' \\ \Psi &= (\Gamma_1 \quad \cdots \quad \Gamma_{p-1}) \end{aligned}$$

The model is, using the new notation,

$$z_{0t} = \alpha\beta'z_{1t} + \Psi z_{2t} + \varepsilon_t.$$

The estimation procedure, at the high level, consists of: 1) partialing out z_{2t} , 2) estimating β , and given β 3) estimating also α and Ψ .

Let R_{jt} , $j = 0, 1$ denote the residuals from regressing z_{jt} on z_{2t} . Define also the product matrices

$$S_{ij} = \frac{1}{T} \sum_{t=1}^T R_{it}R'_{jt}, \quad i, j = 0, 1.$$

The estimator of β is obtained as the eigenvectors for the r first (largest) eigenvalues obtained as the solution to the eigenvalue equation

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0.$$

Standard multivariate regression is used to estimate α and Γ_i given knowledge of β by regressing z_{0t} on $\hat{\beta}'z_{1t}$ and z_{2t} .

In Section 3.1, impulse responses were obtained as the moving average matrices in the VMA representation of the VAR. However, integrated VARs do not permit such representations. Nevertheless, impulse responses can still be obtained and analyzed in the cointegrated case. From Granger's representation in (3.2), the $I(1)$ process x_t is formulated in terms of its errors and

$$\frac{\partial x_{t+h}}{\partial \varepsilon_t} = C + C_h^*.$$

Alternatively, one can also think of the system as being in equilibrium with $x_s = 0$ for $s < 0$ and then trace the evolution after a unit shock puts the system

in disequilibrium. Consequently, the impulse response has a natural interpretation as the h -step ahead forecast, where the initial shock is the only source of deviation from equilibrium. The latter way of viewing impulse responses as forecasts is discussed in depth by Lütkepohl (2005), whereas structural analysis of the VECM based on Granger's representation theorem is covered by Juselius (2006).

3.3 Bayesian VAR models

An important issue associated with VAR models is the curse of dimensionality. The number of regression parameters to estimate in each equation is $np + 1$. At the same time, the sample size, denoted by T , is typically limited as many applications deal with quarterly data. Standard macro VARs, e.g. Christiano et al. (2005), typically include 5–10 endogenous variables and, say, 4 lags. With 20 years of data, roughly 20–40 parameters per equation need to be estimated with 80 observations. Inevitably, maximum likelihood estimation is imprecise and highly variable.

To deal with the curse of dimensionality, the prevailing way to estimate VARs today is to use Bayesian methods. This choice is typically justified as a form of shrinkage device rather than as a philosophical stance. By altering the way in which the prior distributions for the parameters in the model are specified, vastly different estimators can be obtained as each implies a unique way of enforcing shrinkage into the estimation of the model.

A seminal contribution and building block for many more recent proposals is the Minnesota prior developed by Litterman (1979). The Minnesota prior acknowledges that because of the large number of parameters in the model, it is challenging to with care and precision put a prior on each individual parameter explicitly. Instead, the way a full prior for the parameters can be specified is by means of a low-dimensional set of hyperparameters.

Let us first be concerned with placing a prior on the dynamic regression parameters, i.e. being explicit about what $p(\Phi_1, \dots, \Phi_p)$ is. A common assumption is that the prior distribution family is normal and that there is prior independence among the parameters. Such an assumption reduces the task of specifying the prior distribution to specifying a prior mean and variance for each parameter.

The key idea of the Minnesota prior is to let the prior have a structure that is in line with three stylized facts in macroeconomics:

1. many series can be approximated by random walks
2. lags that are closer in time are more important than distant lags
3. in explaining a certain variable, lags of the variable itself are more important than lags of other variables.

The first point suggests that the prior mean should be set such that the model reduces to a set of n random walks under the prior, i.e. $E(\Phi_1) = I_n$ while

$E(\Phi_2) = \dots = E(\Phi_p) = 0$. The second and third points suggest letting the prior variances be tighter for parameters related to 1) lags of other variables, and 2) more distant lags. More specifically, the way the Minnesota prior operationalizes this idea is by the following equations:

$$\begin{aligned} E(\phi_k^{(i,j)}) &= \begin{cases} 1, & \text{if } k = 1 \text{ and } i = j, \\ 0, & \text{otherwise} \end{cases} \\ \sqrt{V(\phi_k^{(i,j)})} &= \begin{cases} \frac{\lambda_1}{k^{\lambda_3}}, & \text{if } i = j, \\ \frac{\lambda_1 \lambda_2}{k^{\lambda_3}}, & \text{otherwise,} \end{cases} \end{aligned} \quad (3.3)$$

where $\phi_k^{(i,j)}$ is element (i, j) of Φ_k . The prior is fully specified given the three hyperparameters λ_1 , λ_2 and λ_3 . The overall tightness, i.e. the degree of shrinkage, is set by λ_1 , whereas λ_2 determines the additional penalization that should be made for lags of other variables. The final hyperparameter, λ_3 , specifies the degree to which more distant lags should be penalized. Typical values for the hyperparameters are $\lambda_1 = 0.2$, $\lambda_2 = 0.5$ and $\lambda_3 = 1$ (see Doan, 1992; Canova, 2007; Carriero et al., 2015a).⁶

There are many extensions of the basic Minnesota prior where additional hyperparameters provide other features. The review by Karlsson (2013) provides a thorough tour through many of these.

One prior distribution of particular interest in this thesis is the steady-state prior proposed by Villani (2009). The idea is as elegant as it is simple: In the VAR in (3.1), it is typically difficult to elicit a prior for the intercept. For this reason, it is customary to assign it a loose prior such as $c \sim N(0, 100^2 I_n)$. However, by a reparametrization of the model one obtains

$$x_t - \mu = \sum_{i=1}^p \Phi_i(x_{t-i} - \mu) + \varepsilon_t.$$

While the likelihood remains unchanged, the intercept is replaced in the equation by the unconditional mean

$$E(x_t | c, \Phi) = \mu = (I_n - \Phi_1 - \dots - \Phi_p)^{-1} c.$$

The unconditional mean is interchangeably referred to as the *steady state*, the reason for which becomes obvious when considering the long-term forecast in the model.⁷ If the lag polynomial $\Phi(L) = (I_n - \Phi_1 L - \dots - \Phi_p L^p)$ is stable with largest root smaller than one in absolute value, the forecast $E(x_{t+h} | x_t, \dots, x_{t-p+1}, \mu, \Phi)$ converges toward μ as $h \rightarrow \infty$. Thus, a prior distribution for μ can relatively effortlessly be elicited by stipulating a prior belief concerning what the long-term forecast should be. The concept of a steady

⁶For ease of exposition, a term accounting for different scales of variables is omitted from (3.3).

⁷*Unconditional* here refers to the fact that it is not conditional on previous values of x_t .

state is ubiquitous in economics so it is often possible to formulate a prior for μ .

One of the most natural applications of a model with a steady-state prior is when modeling inflation in a country where the central bank operates with an inflation target. The Swedish inflation target is 2 % and it is therefore natural to have a prior belief that inflation in the long run should be close to 2 %. Because of this perk, the steady-state BVAR is—and has been—used extensively by the Riksbank as documented in e.g. Adolfson et al. (2007); Iversen et al. (2016). Its use is, however, not limited to the Riksbank; the National Institute of Economic Research has used the model in several studies (see e.g. Raoufina, 2016; Stockhammar and Österholm, 2016; Lindholm et al., 2018) and the Financial Supervisory Authority used it to analyze household debt (Financial Supervisory Authority, 2015). The Ministry of Finance also make frequent use of the model as demonstrated by e.g. Ankargren et al. (2017); Shahnazarian et al. (2015, 2017). Other interesting uses of the steady-state prior include those of Clark (2011), Wright (2013) and Louzis (2016, 2019).

Among the papers mentioned above, almost all feature low-dimensional models with constant parameters. In the current VAR literature, there is one tendency that can be discerned: an increased use of large and more flexible models.

Traditional VAR models are relatively modest in size, with the number of variables usually kept in single digits. Bańbura et al. (2010) in particular was central in moving the literature toward larger dimensions, where the number of variables is usually around 20–50, sometimes even in the hundreds. The large-dimensional situation has traditionally been the domain of factor models, but VARs tend to outperform such methods (Koop, 2013). There is currently considerable interest in developing scalable methods for VARs, and the high-dimensional literature is making its entry into the VAR literature; for example, Koop et al. (2019) used compressed regression, Gefang et al. (2019) developed variational inference methods for VARs, and Follett and Yu (2019); Kastner and Huber (2018) used global-local shrinkage priors.

In terms of more flexible modeling, VARs now frequently feature either time-varying regression parameters or a time-varying error covariance matrix, where the latter usually goes by the name of stochastic volatility. The seminal papers by Primiceri (2005) and Cogley and Sargent (2005) include both sources of time variation. Several subsequent studies have noted that there are often improvements in forecasting ability (see, among others, Clark, 2011; Clark and Ravazzolo, 2015; D’Agostino et al., 2013). Carriero et al. (2015b) arrived at a similar conclusion in a univariate mixed-frequency regression model.

VAR models estimated by Bayesian methods first require the specification of a full prior distribution. Given the prior, the posterior distribution is ob-

tained as

$$p(\Phi, \Sigma | Y) \propto L(Y | \Phi, \Sigma) p(\Phi, \Sigma), \quad (3.4)$$

where p denotes the prior and posterior distributions and L the likelihood function. I will let Θ generally denote “the parameters” (which should be clear from the context) and upper-case letters represent the full history of the lower-case variable; i.e., Y represents the set $\{y_1, \dots, y_T\}$.⁸

For most problems, $p(\Phi, \Sigma | Y)$ is not available analytically. The main tool of Bayesian statistics is Markov Chain Monte Carlo (MCMC), which is an algorithm for sampling from non-standard and possibly high-dimensional probability distributions. The idea is to create a Markov chain that converges to the distribution of interest. Because the stationary distribution of the Markov chain is, by construction, the target distribution, any desired number of draws can be obtained from the distribution once the chain has converged.⁹ Estimation of most Bayesian VAR models employs a certain type of MCMC algorithm known as Gibbs sampling. Gibbs sampling numerically approximates a joint posterior distribution by breaking down the act of sampling from the joint posterior distribution into smaller tasks consisting of drawing from the conditional posterior distributions. Early seminal work on Gibbs sampling include studies by Geman and Geman (1984); Gelfand and Smith (1990). For an introduction to Gibbs sampling and MCMC more generally, see Geyer (2011).

To offer a concrete example, suppose we want to sample from a bivariate normal distribution with mean zero, unit variances and correlation ρ . It is easy to sample from this joint distribution directly, but using Gibbs sampling the algorithm would be:

1. Set initial values $x^{(0)}, y^{(0)}$
2. For $i = 1, \dots, R$, sample:

$$\begin{aligned} x^{(i)} | y^{(i-1)} &\sim N(\rho y^{(i-1)}, 1 - \rho^2) \\ y^{(i)} | x^{(i)} &\sim N(\rho x^{(i)}, 1 - \rho^2). \end{aligned}$$

Precisely the same conceptual idea is used for estimating VAR models.

Returning to (3.4), in many cases $p(\Phi, \Sigma | Y)$ is intractable. Exceptions do of course exist, and some overly simplistic priors (such as the original Minnesota prior) are available in closed form; see also Kadiyala and Karlsson (1993, 1997) for a discussion of numerical methods for other standard prior distributions. However, the analytical tractability of the full posterior distribution vanishes when the prior is made more flexible. For instance, Villani (2009)

⁸In the preceding sections, the VAR model is described using the letter X . The current section denotes the data by Y and therefore appears to make an unwarranted change in notation. The reason for this shift will be made clear in the following sections, where observed data are denoted by Y , but the VAR model is specified on a latent variable X .

⁹Whether the Markov chain has converged is a separate issue that itself has spawned a large literature. See Gelman and Shirley (2011) for an overview.

used a normal prior for μ and the normal-diffuse prior for (Φ, Σ) . The joint posterior distribution $p(\mu, \Phi, \Sigma | Y)$ is not tractable—but because $p(\Phi | \Sigma, \mu, Y)$, $p(\Sigma | \Phi, \mu, Y)$ and $p(\mu | \Phi, \Sigma, Y)$ are, a Gibbs sampler based on

$$\begin{aligned}\mu^{(i)} &\sim p(\mu | \Phi^{(i-1)}, \Sigma^{(i-1)}, Y) \\ \Phi^{(i)} &\sim p(\Phi^{(i)} | \Sigma^{(i-1)}, \mu^{(i)}, Y) \\ \Sigma^{(i)} &\sim p(\Sigma^{(i)} | \Phi^{(i)}, \mu^{(i)}, Y)\end{aligned}$$

can be constructed. All three of the above conditional posterior distributions are easy to sample from and thus one can obtain samples from the joint posterior distribution.

When forecasting is the objective, the ultimate object of interest is the predictive density defined as

$$f(y_{T+1:T+h} | Y) = \int f(y_{T+1:T+h} | Y, \Theta) p(\Theta | Y) d\Theta. \quad (3.5)$$

The predictive density is more rarely available analytically, but fortunately the structure of the integral immediately suggests a sampling-based solution. Given a draw $\Theta^{(i)}$ from the posterior $p(\Theta | Y)$, generate $y_{T+1:T+h}$ from

$$y_{T+1:T+h} \sim f(y_{T+1:T+h} | Y, \Theta^{(i)}) = \prod_{i=1}^h f(y_{t+i} | y_{t+i-1}, Y, \Theta^{(i)}).$$

Generating from $f(y_{T+1:T+h} | Y, \Theta^{(i)})$ is simple as it amounts to generating forecasts from the model with the parameters known. The samples $y_{T+1:T+h}^{(i)}$ are a set of R draws from the predictive density (3.5). Because the samples describe the full distribution of the forecasts, they can be processed accordingly to yield summaries thereof (e.g. point or interval forecasts).

For modeling stochastic volatility, estimation usually follows the approach presented by Kim et al. (1998), who introduced mixture indicators. Conditional on the mixture indicators, the stochastic volatility model is a linear and normal state-space model and a standard simulation smoothing procedure can be employed. Recent advances on estimating stochastic volatility models was made by Kastner and Frühwirth-Schnatter (2014), who used the ancillarity-sufficiency interweaving strategy proposed by Yu and Meng (2011) to boost the efficiency. The standard stochastic volatility model is a univariate model and various multivariate constructions can be used to transfer the stochastic volatility concept into VAR models. See also Carriero et al. (2016, 2019) for modeling stochastic volatility in large VARs.

An alternative route for handling stochastic volatilities when the number of variables is high is to use a factor stochastic volatility model. Based on previous work by Kastner and Frühwirth-Schnatter (2014), Kastner et al. (2017) developed an efficient MCMC algorithm for estimating the factor stochastic

volatility model and Kastner and Huber (2018) employed the factor stochastic volatility model in a VAR with 215 variables. The factor stochastic volatility model estimated with shrinkage priors on the factor loadings was considered by Kastner (2019).

3.4 Mixed-frequency VAR models

The standard textbook description of multivariate time series is that a vector of values $y_t = (y_{1,t} \ y_{2,t} \ \dots \ y_{n,t})'$ is observed. In practice, the situation is more complex. There are three important issues that complicate this description: 1) Series typically start at different points in time, 2) series often end at different time points, and 3) series may be sampled at different frequencies. The first point is usually not a major concern, given that all series are “long enough.”¹⁰ The second and third points are important concerns for real-time macroeconomic forecasters. The use of mixed-frequency methods is largely driven by these two points.

To get a sense of the issue at hand, consider the following example. On February 12, 2019, the executive board of the Riksbank decided to leave the repo rate unchanged at -0.25. For the board to make an informed decision, the staff prepared a report with analyses and forecasts as it does for every monetary policy decision. It is of the utmost importance that the board has an accurate assessment of the current economic conditions—particularly inflation and economic activity. However, what makes such an assessment difficult is that variables are published with lags. To be more specific, inflation for January was published on February 19, the unemployment rate for January was published on February 21 and GDP growth for the fourth quarter of 2018 was published on February 28. The staggered nature of the publications is commonly referred to as *ragged edges*. Therefore, when the staff attempts to make an assessment of the current state of the economy, an assessment must first be made of where the economy was. In addition to forecasting the current state—a so-called *nowcast*—they must also make “forecasts” of the past, often called *backcasts*. A thorough description of the issue was presented by Bańbura et al. (2011).

Because datasets are rarely balanced, standard off-the-shelf methods will face problems. For example, if we want to estimate a VAR and forecast inflation, GDP growth and unemployment, we must first tackle two issues. First, GDP growth is sampled on a quarterly basis, and inflation and unemployment are sampled monthly. A standard application would aggregate inflation and

¹⁰Even if all series were observed for a long time, say since the beginning of the 20th century, the best approach is not guaranteed to be using all the data. The economy has changed dramatically over the last 120 years and any statistical model estimated on long time spans would likely be subject to structural breaks as economic conditions, regulations and definitions have shifted as well.

unemployment to the quarterly frequency. This is, for example, the procedure used for the VAR that the Riksbank employs for forecasting (see Iversen et al., 2016). As a consequence, the most recent quarter for which all variables are observed is the third quarter of 2018. Hence, an estimation procedure that requires balanced data neglects the monthly nature of two of the variables as well as the observations of these in the fourth quarter. The second issue is how to make use of all information when forecasting. At the beginning of February, forecasts can be made by: 1) estimating the quarterly model on data through 2018Q3, and 2) making forecasts for 2018Q4, 2019Q1 and so on conditional on inflation and unemployment in 2018Q4. In a wider sense, this approach uses all available information, although it can be argued that the aggregation into quarterly frequency in the data preparation stage already incurs a loss of information. At the end of February, however, the situation becomes more complicated. The balanced part of the sample now ends in 2018Q4 and we have two additional monthly observations of inflation and unemployment for January 2019. How the two additional observations can be leveraged in making forecasts is now not as clear-cut. There are, of course, suggestions in the literature,¹¹ but at the end of the day incorporation of new observations of monthly variables is not seamless and often requires a two-step approach.

Mixed-frequency methods are statistical approaches that attempt to make use of the full set of information in a more principled way. Largely speaking, there are three main strands within this set of methods: univariate regressions, factor models and VARs. Univariate regressions include methods like bridge equations and mixed-data sampling (MIDAS) regressions, with early important contributions made by Baffigi et al. (2004) and Ghysels et al. (2007). A comprehensive overview of bridge equation and MIDAS approaches and the various tweaks of the latter that allow for more flexible modeling was provided in the review by Forni and Marcellino (2013). Kuzin et al. (2011) offered an early comparison of MIDAS and VARs for forecasting Euro area GDP, finding that MIDAS performs better for shorter horizons and the VAR for longer-term forecasts. In terms of mixed-frequency factor models, Mariano and Murasawa (2003) proposed a factor model for a coincident index of business cycles based on mixed-frequency data and important extensions have thereafter been made by Camacho and Perez-Quiros (2010), Mariano and Murasawa (2010) and Marcellino et al. (2016).

The topic of Paper II–V is the third category: mixed-frequency Bayesian VARs. The central work that Paper II–V build on is Schorfheide and Song (2015), who developed a mixed-frequency VAR with a Minnesota-style normal prior for real-time forecasting of US macroeconomic time series. Other important contributions that are closely related include Eraker et al. (2015),

¹¹One way is to use an auxiliary model for forecasting the February and March observations and then computing the observation in the first quarter of 2019, which would be partially based on forecasts. That is a bridge equation approach, see for example Baffigi et al. (2004) and Itkonen and Juvenon (2017).

who also proposed a mixed-frequency Bayesian VAR, albeit with a different sampling strategy. Ghysels (2016) presented a MIDAS-VAR, which employs ideas from the specification of MIDAS regressions in estimating multivariate models, and Ghysels and Miller (2015) discussed testing for cointegration in the presence of mixed-frequency data.

The MIDAS-VAR is fundamentally different from the approach taken in Schorfheide and Song (2015); Eraker et al. (2015) and Paper II–V in the sense that the latter frame the problem as a missing data problem—if we had observed quarterly variables at a monthly frequency, estimation would have been straight-forward. The solution is thus to use ideas that can be traced back to the Expectation-Maximization (EM) algorithm (Dempster et al., 1977) and Bayesian data augmentation (Tanner and Wong, 1987), by alternating between filling in missing values and estimating parameters. Precisely how this is achieved will be made clear in the following.

Let $x_t = (x'_{m,t} \quad x'_{q,t})'$ be an n -dimensional vector with n_m monthly and n_q quarterly variables ($x_{m,t}$ and $x_{q,t}$, respectively). The time index t refers to the monthly frequency. In the following description as well as in Paper II–V, I will focus exclusively on data consisting of monthly and quarterly series.

The inherent problem with mixed-frequency data is that $x_{m,t}$ is fully observed (up to the ragged edge), whereas $x_{q,t}$ is not. What we assume is that the observation for the quarterly series that we do obtain—one every three months—is a linear combination of an underlying, unobserved monthly series. This link between observations and an underlying process is the central device that allows the model to be estimated and handle mixed frequencies.

The linear combination employed may vary and be different for different types of variables (stock and flow variables, for example). I will refer to the linear combination more generally as the *aggregation scheme*, i.e. the way we stipulate that our observation is aggregated from an (unobserved) underlying process.

To distinguish between observed and unobserved variables, we let $y_t = (y'_{m,t} \quad y'_{q,t})'$ denote the observed variables at time t . Its dimension is n_t , where $n_t \leq n$. The time-varying dimension is to reflect the fact that we do not observe all variables every month. Two aggregation schemes are common in the literature: intra-quarterly averaging, and triangular weighting.

Intra-quarterly averaging is typically used for data in log-levels and assumes that quarterly observations are averages of the constituent months. The relation between observations and the underlying process can therefore be summarized by

$$y_{q,t} = \begin{cases} \frac{1}{3} (x_{q,t} + x_{q,t-1} + x_{q,t-2}), & \text{if } t \in \{\text{Mar, Jun, Sep, Dec}\} \\ \emptyset, & \text{otherwise.} \end{cases} \quad (3.6)$$

For data that is also differenced, Mariano and Murasawa (2003) showed how the intra-quarterly average for log-levels implies a triangular weighting

for differenced data. Let $y_{q,t}^* = y_{q,t} - y_{q,t-3}$ be the log-differenced quarterly series that is defined as

$$\begin{aligned} y_{q,t}^* &= \frac{1}{3} (x_{q,t} + x_{q,t-1} + x_{q,t-2}) - \frac{1}{3} (x_{q,t-3} + x_{q,t-4} + x_{q,t-5}) \\ &= \frac{1}{3} [(x_{q,t} - x_{q,t-3}) + (x_{q,t-1} - x_{q,t-4}) + (x_{q,t-2} - x_{q,t-5})]. \end{aligned}$$

Because $x_{q,t} - x_{q,t-3} = \Delta x_{q,t} + \Delta x_{q,t-1} + \Delta x_{q,t-2}$, the log-differenced observation $y_{q,t}^*$ can be written in terms of the log-differenced latent series $x_{q,t}^* = \Delta x_{q,t}$ as

$$y_{q,t}^* = \begin{cases} \frac{1}{3}(x_{q,t}^* + 2x_{q,t-1}^* + 3x_{q,t-2}^* + 2x_{q,t-3}^* + x_{q,t-4}^*), & t \in \{\text{Mar, Jun, Sep, Dec}\} \\ \emptyset, & \text{otherwise.} \end{cases} \quad (3.7)$$

The weighted average in (3.7) defines the triangular weighting scheme.

The relation between $y_{q,t}$ and $x_{q,t}$ can more succinctly be written using a selection matrix, $S_{q,t}$, and an aggregation matrix, Λ_{qq} . Both of these are fully known and require no estimation, but simply allow us to formulate the previous equation as a matrix product. Considering now the intra-quarterly average weighting scheme for simplicity, let

$$y_{q,t} = S_{q,t} \Lambda_{qq} \begin{pmatrix} x_{q,t} \\ x_{q,t-1} \\ x_{q,t-2} \end{pmatrix},$$

where $S_{q,t}$ is an $n_q \times n_q$ identity matrix. Rows corresponding to missing elements of $y_{q,t}$ are removed to facilitate appropriate inclusion when variables are observed. The aggregation matrix Λ_{qq} is

$$\Lambda_{qq} = \left(\frac{1}{3} I_{n_q} \quad \frac{1}{3} I_{n_q} \quad \frac{1}{3} I_{n_q} \right).$$

For the monthly variables, the relation can similarly be written as:

$$y_{m,t} = S_{m,t} \Lambda_{mm} \begin{pmatrix} x_{m,t} \\ x_{m,t-1} \\ x_{m,t-2} \end{pmatrix}. \quad (3.8)$$

This relation is, however, simpler than what one might suspect at first glance. In this case, $S_{m,t}$ is the n_m identity matrix with no rows deleted for the balanced part of the sample, as no monthly variable is missing during this period. Only in the ragged edge part are rows of $S_{m,t}$ deleted to account for missingness. Moreover, the aggregation here is simply $\Lambda_{mm} = I_{n_m}$. The reason for including Λ_{mm} is purely for the purpose of exposition. Collecting (3.6) and (3.8) yields

$$y_t = S_t \Lambda \begin{pmatrix} x_t \\ x_{t-1} \\ x_{t-2} \end{pmatrix}, \quad (3.9)$$

where

$$S_t = \begin{pmatrix} S_{m,t} & 0 \\ 0 & S_{q,t} \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \Lambda_{mm} & 0 \\ 0 & \Lambda_q \end{pmatrix}$$

and Λ_{qq} is Λ_q but with zero-only columns deleted.

The accompanying VAR is specified on the monthly frequency. Intuitively, we are specifying the VAR that we would like to, but are unable to use owing to the mixed frequencies. This sentiment is reflected in, e.g., Cimadomo and D'Agostino (2016), who used the same mixed-frequency approach for handling a change in frequency. The authors studied the effect of government spending on economic output, but faced a challenge in that data on government spending are available quarterly only after 1999. They therefore proceeded with a mixed-frequency construction to enable use of longer time series.

The VAR model is precisely (3.1), the complicating issue being that x_t is now partially observed. Equations (3.1) and (3.9) together form a state-space model, where (3.1) is the observation equation and (3.9) is the transition (or state) equation.

The objective of estimating the mixed-frequency VAR is to be able to characterize the posterior distribution as this is the foundation for the predictive density. The key to estimation is that given X , estimation is standard. To reiterate, this strongly connects with the ideas of the EM algorithm and data augmentation (Dempster et al., 1977; Tanner and Wong, 1987) and can be viewed from the perspective of imputation, where imputation and estimation occur jointly. The posterior distribution, augmented with the underlying variable X , is $p(X, \Phi, \Sigma | Y)$, which is intractable and not available in closed form. Fortunately, the structure of the problem lends itself well to Gibbs sampling.

For the mixed-frequency VAR under a normal-inverse Wishart prior, the sampling algorithm consists of repeating the following:

$$\begin{aligned} (\Phi^{(i)}, \Sigma^{(i)}) &\sim p(\Phi, \Sigma | X^{(i-1)}) \\ X^{(i)} &\sim p(X | \Phi^{(i)}, \Sigma^{(i)}, Y). \end{aligned}$$

The resulting set of draws $\{X^{(i)}, \Phi^{(i)}, \Sigma^{(i)}\}_{i=1}^R$ is a (correlated) set of R draws from the posterior $p(X, \Phi, \Sigma | Y)$.

The first step is standard in estimating Bayesian VARs and thoroughly described in e.g. Karlsson (2013); in brief, $\Sigma^{(i)}$ is drawn from an inverse Wishart distribution where the scale matrix is a function of $X^{(i-1)}$, and $\Phi^{(i)}$ is drawn from a normal distribution with moments that depend on $\Sigma^{(i)}$ and $X^{(i-1)}$.

The distinguishing feature of the mixed-frequency model is the final step. As has been demonstrated by Frühwirth-Schnatter (1994); Carter and Kohn (1994); De Jong and Shephard (1995) and later Durbin and Koopman (2002), a draw from $p(X | \Phi, \Sigma, Y)$ can be obtained using a forward-filtering, backward-smoothing (FFBS) algorithm. Algorithms that produce draws from the posterior $p(X | \Phi, \Sigma, Y)$ in a state-space model are often referred to as simulation

smoothers. Aspects of state-space models, including simulation smoothing, have been discussed in depth by Durbin and Koopman (2012).

Looking beyond the use of mixed-frequency models for forecasting, it is interesting to note that they are largely absent in the structural VAR literature. A number of highly influential papers in the monetary policy literature (including Leeper et al., 1996; Bernanke and Mihov, 1998; Uhlig, 2005; Sims and Zha, 2006) have used VARs with structural identification to study various aspects of monetary policy in the United States. What is common to all of the aforementioned papers is that they have estimated monthly VAR models including a monthly GDP series, which is interpolated using the Chow and Lin (1971) procedure. This gives rise to a two-step approach, where the uncertainty of the first step (interpolation) is unaccounted for in the second (impulse response analysis in the structural VAR). The issues associated with the so-called *generated regressors* problem are well-known, see e.g. Pagan (1984).

On the other hand, Ghysels (2016) criticizes the use of mixed-frequency VARs based on state-space models owing to their nature of being formulated in terms of latent variables, and hence in terms of high-frequency latent shocks, claiming that they do not have the same structural interpretation. While this criticism may be warranted in many situations, the frequent use of interpolation to some degree invalidates the critique, as economists evidently are interested in the high-frequency shocks and attribute them meaning. Given their interest in the high-frequency shocks, avoiding the interpolation step in favor of joint inference in the mixed-frequency model is compelling and offers a more econometrically sound approach. Comparing the results from a mixed-frequency model with those obtained in the key monetary policy papers based on interpolation would be an interesting and illuminating exercise. For some of the work in the direction of employing mixed-frequency VARs also for structural questions, see Foroni et al. (2013); Foroni and Marcellino (2014, 2016); Bluwstein and Canova (2016).

3.5 Statistical computations in the R programming language

Paper V is slightly unorthodox in that it does not present any new statistical theory or methods, but an R package implementing existing mixed-frequency VAR methods. One of the early insights was that the target audience of the mixed-frequency work is mainly central bankers and other forecasters, primarily located at government agencies and institutes. These people would generally not implement standard Bayesian VARs on their own due to time constraints, and would be much less inclined to implement mixed-frequency VARs, which require more work. Moreover, forecasting rounds can be fast and if the models are too slow, they will likely not be relevant forecasting

tools. For this reason, a considerable amount of time has been invested in the implementations to provide a fast and user-friendly modeling experience. In this section, I will give a simple example to illustrate the implementations in the package.

The package is available for the R programming language (R Core Team, 2019), an open-source software for statistics and related computational problems. R is famous for its large number of user-contributed packages, but also infamous for its slowness:

R is not a fast language. This is not an accident. R was purposely designed to make data analysis and statistics easier for you to do. It was not designed to make life easier for your computer. While R is slow compared to other programming languages, for most purposes, it's fast enough. (Wickham, 2015, p. 331)

The “for most purposes” caveat is, unfortunately, not applicable to the mixed-frequency VARs. The problem, as with any MCMC-based approach, is that costly computations—such as generating numbers from high-dimensional multivariate normal distributions, or filtering and smoothing using the Kalman filter and smoother—need to be repeated a large number of times. Even if the computations can be carried out in a fraction of a second, when they need to be repeated tens of thousands of times, the computational costs of every piece pile up.

With this in mind, the approach taken is therefore to let the costly parts of the MCMC algorithms be implemented in C++, a much faster and stricter programming language, and use R mostly as an interface. Use of C++ is facilitated by the extensive work carried out by the Rcpp team (Eddelbuettel and François, 2011; Eddelbuettel, 2013). In addition, the RcppArmadillo package (Eddelbuettel and Sanderson, 2014) implements a port to the Armadillo library, developed by Sanderson and Curtin (2016). The Armadillo library enables easy use of fast linear algebra routines.

Figure 3.2 shows the time it takes using R or C++ to produce a draw from the multivariate normal distribution $N(\mu, \Sigma)$. The procedure is short and consists of:

1. Generate a vector z of independent $N(0, 1)$ variates
2. Compute the lower Cholesky decomposition $\Sigma = LL'$
3. Compute $y = \mu + Lz$.

The body of the functions in the example contain three lines of code. However, despite the C++ implementation requiring little additional effort, it is notably faster.

To further appreciate the gains of moving from R for the heavy computations, consider the following state-space model:

$$\begin{aligned}y_t &= \alpha_t + \varepsilon_t, & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2) \\ \alpha_t &= \alpha_{t-1} + \eta_t, & \eta_t &\sim N(0, \sigma_\eta^2).\end{aligned}$$

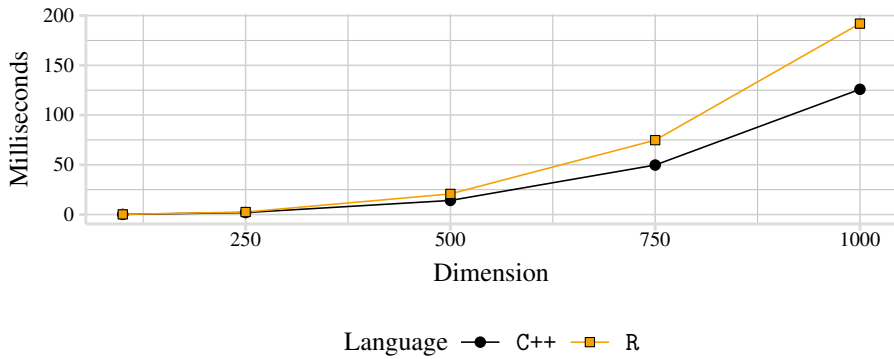


Figure 3.2. Computational cost of sampling from a multivariate normal distribution

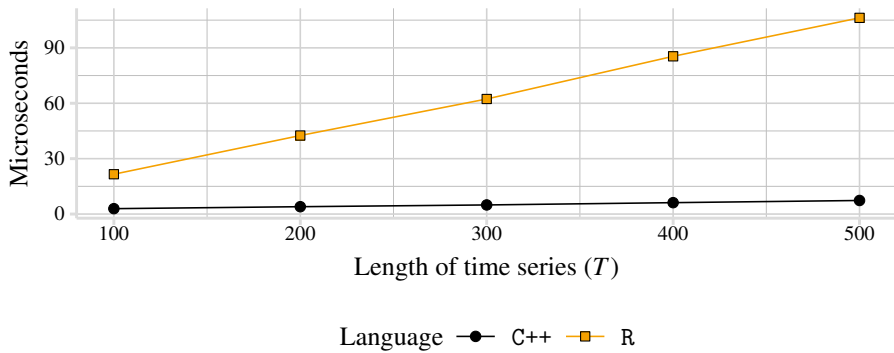


Figure 3.3. Computational cost of the Kalman filter for the local-level model

The model is known as a local-level model and is discussed in detail in chapter 2 of Durbin and Koopman (2012).

Computing $\alpha_{t|t} = E(\alpha_t | y_t, y_{t-1}, \dots)$ is achieved by means of the celebrated Kalman filter originally developed by Kalman (1960). The filter consists of the following equations for recursively computing $a_{t|t}$:

$$\begin{aligned}
 v_t &= y_t - a_t, & F_t &= P_t - \sigma_\varepsilon^2 \\
 a_{t|t} &= a_t + K_t v_t, & P_{t+1} &= P_t(1 - K_t) + \sigma_\eta^2 \\
 K_t &= P_t / F_t, & t &= 1, \dots, T.
 \end{aligned}$$

Implementing the Kalman filter for the local-level model also requires little effort, where the main part is a loop over t containing six lines of code.

Figure 3.3 shows the computational burden of the Kalman filter for the local-level model for various lengths of the time series y_t . The difference

between the R and C++ implementations are large and the C++ function scales better in T than its R counterpart.

What Figure 3.2 and 3.3 illustrate is that for these simple demonstrations, implementing the functions in C++ comes with a substantial speed improvement. Admittedly, the implementations required for the mixed-frequency models are more involved than these examples, but there are still large (if not larger) gains associated with moving the main computations to C++. With pure implementations in R, none of Paper II–V would have been feasible.

4. Summary of papers

4.1 Paper I

The first paper in the thesis deals with the issue of cointegration when the model is used to model a small open economy. Small open economies engage in international affairs and trade, but are too small to affect global economic conditions and variables. Such a description applies to the Swedish economy. However, Sweden is largely influenced by the rest of the world and global developments. Because of this, it is common for macroeconomic models for small open economies to include a set of foreign variables as a proxy for the global economy. The Riksbank VAR (Iversen et al., 2016) therefore includes three foreign variables constructed as weighted averages of Sweden's largest trading partners. Similarly, the DSGE model used by the Riksbank (RAMSES, Adolfson et al., 2008, 2013) contains a domestic and foreign block to allow modeling of spillovers from the global economy.

The contribution of the paper is a proposed method for incorporating the restrictions implied by the notion of a small open economy when estimating a VECM that includes a domestic and a foreign block of variables. The estimation procedure allows for imposing the small open economy property and the implied restrictions on adjustment parameters α , long-run parameters β and short-run parameters Γ simultaneously. To this end, the iterative estimation method presented in Boswijk (1995) and Groen and Kleibergen (2003) is used.

The paper presents Monte Carlo results showing that impulse responses are more accurate if the restrictions are used in full. In two applications using Swedish data, we estimate the impulse responses with and without restrictions. The results show that the impulse responses can exhibit notable differences whether restrictions are enforced or not, thereby demonstrating the usefulness of the proposed method as these restrictions are in many cases uncontroversial.

4.2 Paper II

Paper II develops a Bayesian mixed-frequency VAR using the steady-state prior proposed by Villani (2009). As is discussed in Section 3.3, the steady-state BVAR is frequently used for forecasting and economic analyses, particularly for modeling the Swedish economy. The contribution of the paper is to present the necessary methodology for estimating the steady-state BVAR on

mixed-frequency data. To this end, we build upon the work of Schorfheide and Song (2015).

Several variables included in the common macroeconomic models for which the steady-state prior is employed are in fact sampled on a monthly basis, inflation and unemployment being the two leading examples. The crux of the matter, however, is that these models typically also include GDP growth—a quarterly variable. The mismatch in frequency is usually handled by aggregating the monthly variables so that a quarterly dataset is obtained. The proposed method allows users of the steady-state BVAR to continue using their familiar models, but make better use of their data and incorporate the monthly data directly into the model.

We improve the flexibility of the model by using the hierarchical steady-state prior proposed by Louzis (2019), and the common stochastic volatility model put forward by Carriero et al. (2016). The hierarchical steady-state prior has the benefit that it requires only elicitation of prior means for the steady-state parameters as opposed to the original steady-state prior, which needs also prior variances to be specified. Common stochastic volatility is a parsimonious way of accounting for heteroskedasticity, where a single time-varying factor is used to scale a constant error covariance matrix.

The methodology is employed in a medium-scale VAR using real-time US data with ten monthly and three quarterly variables. Overall, the results show that the quality of the forecasts is improved when mixed-frequency data, steady-state information, and stochastic volatility are incorporated. Comparing the original steady-state prior with the hierarchical specification, we find that the latter tends to perform equally as well. Using a hierarchical structure therefore provides an alternative that simplifies the incorporation of prior information with no cost in terms of performance.

4.3 Paper III

Paper III sets out to adapt the mixed-frequency framework put forward by Schorfheide and Song (2015) to the high-dimensional setting when the data contain ragged edges. We improve upon the computational aspects of the simulation smoothing algorithm and provide a new adaptive procedure that is faster than the Schorfheide and Song (2015) algorithm.

Schorfheide and Song (2015) provided a simulation smoothing algorithm that uses an alternative representation of the model for the balanced part of the sample, in which the dimension of the state vector is $n_q(p + 1)$ instead of np . The reduced state dimension ameliorates the computational efficiency substantially. For the unbalanced part of the sample, the algorithm makes use of the companion form with state dimension $n(p + 1)$. When dimensions increase, even if the companion form is only used for one or two time points

(as opposed to several hundreds, as for the balanced part), it still dominates in terms of computational time.

We develop a blocked filtering and an adaptive filtering algorithm. The blocked filtering algorithm improves the computational efficiency by exploiting the structures and sub-blocks of many of the large matrices, thereby avoiding costly matrix operations. A similar approach but for DSGE models was taken by Strid and Walentin (2009). The adaptive filtering algorithm instead utilizes the nature of the data and its observational structure, only including in the state vector what is necessary. By doing so, the costly matrix operations do not occur to begin with as the flaw of the Schorfheide and Song (2015) procedure in large models is that it includes unnecessary terms in the state vector.

We find that the adaptive procedure works better than the blocked filtering algorithm. The adaptive procedure makes considerable improvements compared to the Schorfheide and Song (2015) algorithm. The size of the gains increases with both the number of variables and the number of lags, thereby showing that our adaptive procedure scales better. The largest model that we consider in our comparison of computational efficiencies makes use of 120 variables and 12 lags and is close in size to the large VARs used by Bańbura et al. (2010) and Carriero et al. (2019). Using our adaptive algorithm requires less than 10 % of the computational effort. On a standard desktop computer, the implication is that the mixed-frequency of block of the model needs less than 3 hours to yield 10,000 draws using the adaptive algorithm, whereas over 30 hours is needed otherwise. The algorithm therefore provides an essential building block for developing large-dimensional VARs for nowcasting in the presence of data with ragged edges.

4.4 Paper IV

Paper IV provides further contributions to making estimation of large mixed-frequency VARs feasible for nowcasting. We use a factor stochastic volatility model along the lines of the model employed by Kastner et al. (2017) to capture the time-varying error variances in the model. The use of a factor stochastic volatility model makes the equations in the model conditionally independent. We exploit the conditional independence to provide a high-dimensional model with stochastic volatility estimated on mixed-frequency data that can be estimated in a relatively short amount of time.

The factor stochastic volatility model decomposes the error term in the model into a common component and an idiosyncratic term. Because the idiosyncratic terms are independent across equations, the equations in the model are independent given the common component. Furthermore, when the model features a large number of monthly variables and only a single or a few quarterly variables, the dimension of the state equation in the state-space model

used in the estimation is much smaller than the dimension of the observation equation. Consequently, the situation lends itself well to the Koopman and Durbin (2000) univariate approach for filtering and smoothing. In brief, the method avoids expensive matrix inversions and multiplications in favor of a larger number of scalar and vector operations. Coupling the univariate filtering procedure with the adaptive filtering algorithm from Paper III, we obtain even larger computational improvements that are more pronounced the larger the model. In addition, an important aspect of the model is that the conditional independence between equations allows us to sample the regression parameters for each equation in parallel.

The univariate simulation smoothing procedure and the parallel sampling of regression parameters are essential for making estimation of large-dimensional mixed-frequency VARs possible. We demonstrate the computational benefits by retrieving data from the FRED database (McCracken and Ng, 2016) and estimating three models with 20, 34, and 119 variables, where each model contains monthly data and quarterly GDP growth. By the construction of the model and our routines, the additional computation induced by the mixed-frequency nature of the data is responsible for a much smaller share of the computational burden than before. As such, the methodology provides an accessible way to combine large VARs with mixed-frequency data.

4.5 Paper V

Estimating mixed-frequency VARs can be time-consuming, especially when moving into more high-dimensional regimes as discussed in Paper III–IV. Not only is the mixed-frequency step in the estimation a demanding step, but drawing from high-dimensional multivariate normal distributions, which arise as the posterior distributions of the regression parameters, is challenging. Paper V therefore presents a free and open source package for the R programming language (R Core Team, 2019). The package is called `mfvar`, is licensed under the GNU GPL-3 license and is available for download from the Comprehensive R Archive Network (CRAN) at <https://cran.r-project.org/package=mfvar>.

The package implements several versions of the mixed-frequency VAR, including the normal-inverse Wishart and normal-diffuse priors, and the steady-state prior (possibly with the hierarchical specification used in Paper II). Additionally, common or factor stochastic volatility can be used for the error covariance matrix as discussed in Paper II and Paper IV. The aim of the package has been to make these models easy to use, the goal being to promote mixed-frequency VARs. In Paper V, we document the key features of the package and provide examples of how the package can be used.

Because computational time has been a central issue in the papers collected in this thesis, all functions used for estimating models are implemented in

C++ via the Rcpp package (Eddelbuettel and François, 2011; Eddelbuettel, 2013) and the Armadillo library for linear algebra (Eddelbuettel and Sanderson, 2014). A second contribution of the package is not only the package itself, but its functions, which can be reused by other functions. Key functions implemented in C++, such as the various simulation smoothers discussed in Papers III–IV, are available as header files and can therefore be imported by other packages. Paper IV also discusses samplers tailored for multivariate normal posterior distributions; these are also available as header files. Much as Papers III–IV largely aim to inspire and provide building blocks for further developments of large-dimensional mixed-frequency VARs, so too does the `mfbvar` package aim to provide tools for further research into this field by letting experienced users cherry-pick among functions to build their own models.

5. Acknowledgments

They say that writing a thesis is an individual endeavor, but I strongly believe that, in reality, it is the collective effort of many. A large number of people have helped me stay afloat during these years and for that I am forever grateful.

First and foremost, I wish to thank my supervisor, Johan Lyhagen. A couple of weeks into my first year, I remember Ronnie saying in a speech after his defense that while it hadn't been easy all of the time, he very much appreciated the scientific freedom you had given him. I didn't get it then, but I do now—and I strongly concur. You have never uttered a word of restraint (despite the fact that four out of five papers are Bayesian!), and I am humbled by your trust in me and my own abilities when I doubted myself.

I also want to thank my assistant supervisor, Yukai Yang. Your enthusiasm and dedication are truly contagious—you always leave me feeling a little bit more cheerful and inspired after we talk. I have enjoyed our many joint teaching ventures, and I know I will not forget your letting me mooch off of your UPPMAX project before I took the time to get my own.

In the initial phase of my PhD work, I was also fortunate enough to work with Shaobo Jin. I am profoundly impressed by your ability to keep on going when most people would stop. I recall us running into some really nasty expressions in the afternoon one day. I went home with a foggy brain, only to later receive an e-mail where you told me you had solved it—still containing plenty of nasty expressions, but that never bothered you. I learned a lot from working with you, and to keep on pushing is one of the key things I take with me.

To Thommy Perlinger, my fellow vänersborgare. Traveling to conferences with you is nothing but a pleasure. Having someone in your company whose best friend is TripAdvisor makes everything so much easier. It was also you, and Lisbeth Hansson, who once brought me in as a teaching assistant in my first semester as a master student, which first exposed me to the act of teaching (and my future wife!). Both of you have since placed a great deal of confidence in me and my teaching, and I am very grateful for all of your support. I also wish to thank Ronnie Pingel and Lars Forsberg—you are largely responsible for making me continue studying statistics as a Bachelor's student.

To my PhD student peers, I wish all of you the best. At times you will feel like you're running up a hill that never ends. I now know that it *does* actually end—an overwhelming insight. I have found great comfort in our PhD meetings and I'm very happy that there's now quite a few of us (you?). To everyone else in and around the department, thank you for always making me feel at home.

Outside of the department, there are two people who have had a tremendous impact on the course of my career. Hovick Shahnazarian and Mårten Bjellerup: You let me write my Bachelor's thesis with you and then took me in for three summers at the Ministry of Finance. The world you opened up for me is largely why I decided to pursue a PhD in the first place. Our applied work has been indispensable to my thesis because it has revealed important parts of the literature to me. I am greatly indebted to you.

I would also like to extend my gratitude to Sveriges Riksbank and the Research Division for the opportunity to partake in the internship program, as well as to everyone at the Modeling Division for hosting me. Seeing how models and forecasts are used in practice was eye-opening and incredibly stimulating.

To my dear friend, Oscar. It always felt weird that I was a PhD student and you were not—I'm pleased now that you finally came to your senses and that the situation is about to be reversed. Our many talks about lots of things—particularly music and statistics—have been welcome breaks when debugging malfunctioning code. I don't think any of us expected this when we went to the outskirts of Helsinki for our Bayesian bootcamp almost six years ago.

Matilda, you have been the one to boost my spirits when my motivation has been at its lowest and self-doubt at its highest. You have put up with this job occasionally occupying virtually all of my mental power (and often also evenings). It would definitely not have been possible without you and your never-ending support. I don't think I could've asked for more.

References

- Adolfson, M., Laséen, S., Christiano, L., Trabandt, M., and Walentin, K. (2013). Ramses II — Model Description. Occasional Paper No. 12, Sveriges Riksbank.
- Adolfson, M., Laséen, S., Lindé, J., and Villani, M. (2007). Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through. *Journal of International Economics*, 72(2):481–511, doi:10.1016/j.jinteco.2007.01.003.
- Adolfson, M., Laséen, S., Lindé, J., and Villani, M. (2008). Evaluating an Estimated New Keynesian Small Open Economy Model. *Journal of Economic Dynamics and Control*, 32(8):2690–2721, doi:10.1016/j.jedc.2007.09.012.
- Anderson, T. W. (1951). Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions. *Annals of Mathematical Statistics*, 22(1):327–351, doi:10.1214/aoms/1177729580.
- Ankargren, S., Bjellerup, M., and Shahnazarian, H. (2017). The Importance of the Financial System for the Real Economy. *Empirical Economics*, 53(4):1553–1586, doi:10.1007/s00181-016-1175-4.
- Ankargren, S., Grip, J., and Shahnazarian, H. (2018). The importance of the balance sheet channel. Unpublished mimeo.
- Ankargren, S. and Shahnazarian, H. (2019). The Interaction Between Fiscal and Monetary Policies: Evidence from Sweden. Working Paper No. 365, Sveriges Riksbank.
- Arias, J. E., Rubio-Ramírez, J. F., and Waggoner, D. F. (2018). Inference Based on Structural Vector Autoregressions Identified With Sign and Zero Restrictions: Theory and Applications. *Econometrica*, 86(2):685–720, doi:10.3982/ECTA14468.
- Baffigi, A., Golinelli, R., and Parigi, G. (2004). Bridge Models to Forecast the Euro Area GDP. *International Journal of Forecasting*, 20(3):447–460, doi:10.1016/S0169-2070(03)00067-0.
- Bañbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian Vector Autoregressions. *Journal of Applied Econometrics*, 25(1):71–92, doi:10.1002/jae.1137.
- Bañbura, M., Giannone, D., and Reichlin, L. (2011). Nowcasting. In Clements, M. P. and Hendry, D. F., editors, *The Oxford Handbook of Economic Forecasting*, chapter 8. Oxford University Press, doi:10.1093/oxfordhb/9780195398649.013.0008.
- Bernanke, B. S. and Mihov, I. (1998). Measuring Monetary Policy. *The Quarterly Journal of Economics*, 113(3):869–902, doi:10.1162/003355398555775.
- Bluwstein, K. and Canova, F. (2016). Beggar-Thy-Neighbor? The International Effects of ECB Unconventional Monetary Policy Measures. *International Journal of Central Banking*, 12(3):69–120.
- Boswijk, P. (1995). Identifiability of Cointegrated Systems. Discussion Paper No. 78, Tinbergen Institute.

- Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Camacho, M. and Perez-Quiros, G. (2010). Introducing the Euro-STING: Short-Term Indicator of Euro Area Growth. *Journal of Applied Econometrics*, 25:663–694, doi:10.1002/jae.
- Canova, F. (2007). *Methods for Applied Macroeconomic Research*. Princeton University Press, New Jersey.
- Carriero, A., Clark, T. E., and Marcellino, M. (2015a). Bayesian VARs: Specification Choices and Forecast Accuracy. *Journal of Applied Econometrics*, 30(4):46–73, doi:10.1002/jae.2315.
- Carriero, A., Clark, T. E., and Marcellino, M. (2015b). Realtime Nowcasting with a Bayesian Mixed Frequency Model with Stochastic Volatility. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 178(4):837–862, doi:10.1111/rssa.12092.
- Carriero, A., Clark, T. E., and Marcellino, M. (2016). Common Drifting Volatility in Large Bayesian VARs. *Journal of Business & Economic Statistics*, 34(3):375–390, doi:10.1080/07350015.2015.1040116.
- Carriero, A., Clark, T. E., and Marcellino, M. (2019). Large Vector Autoregressions with Stochastic Volatility and Non-Conjugate Priors. *Journal of Econometrics*, doi:10.1016/j.jeconom.2019.04.024. Advance online publication.
- Carter, C. K. and Kohn, R. (1994). On Gibbs Sampling for State Space Models. *Biometrika*, 81(3):541–553, doi:10.1093/biomet/81.3.541.
- Chow, G. C. and Lin, A.-I. (1971). Best Linear Unbiased Interpolation, Distribution, and Extrapolation of Time Series by Related Series. *The Review of Economics and Statistics*, 53(4):372–375, doi:10.2307/1928739.
- Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). Monetary Policy Shocks: What Have We Learned and to What End? In *Handbook of Macroeconomics*, volume 1, pages 65–148. Elsevier, doi:10.1016/S1574-0048(99)01005-8.
- Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy*, 113(1):1–45, doi:10.1086/426038.
- Cimadomo, J. and D’Agostino, A. (2016). Combining Time Variation and Mixed Frequencies: An Analysis of Government Spending Multipliers in Italy. *Journal of Applied Econometrics*, 31:1276–1290, doi:10.1002/jae.2489.
- Clark, T. E. (2011). Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility. *Journal of Business & Economic Statistics*, 29(3):327–341, doi:10.1198/jbes.2010.09248.
- Clark, T. E. and Ravazzolo, F. (2015). Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics*, 30(4):551–575, doi:10.1002/jae.2379.
- Cogley, T. and Sargent, T. J. (2005). Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US. *Review of Economic Dynamics*, 8(2):262–302, doi:https://doi.org/10.1016/j.red.2004.10.009.
- D’Agostino, A., Gambetti, L., and Giannone, D. (2013). Macroeconomic Forecasting and Structural Change. *Journal of Applied Econometrics*, 28(1):82–101, doi:10.1002/jae.1257.

- De Jong, P. and Shephard, N. (1995). The Simulation Smoother for Time Series Models. *Biometrika*, 82(2):339–350, doi : 10 . 1093/biomet/82 . 2 . 339.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, doi : 10 . 1111/j . 2517-6161 . 1977 . tb01600 . x.
- Doan, T. A. (1992). *RATS User's Manual, Version 4*. Estima, Evanston, IL.
- Durbin, J. and Koopman, S. J. (2002). A Simple and Efficient Simulation Smoother for State Space Time Series Analysis. *Biometrika*, 89(3):603–615.
- Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, UK, second edition, doi : 10 . 1093/acprof :oso/9780199641178 . 001 . 0001.
- Eddelbuettel, D. (2013). *Seamless R and C++ Integration with Rcpp*. Springer, New York, doi : 10 . 1007/978-1-4614-6868-4.
- Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ Integration. *Journal of Statistical Software*, 40(8):1–18, doi : 10 . 18637/jss . v040 . i08.
- Eddelbuettel, D. and Sanderson, C. (2014). RcppArmadillo: Accelerating R with High-Performance C++ Linear Algebra. *Computational Statistics and Data Analysis*, 71:1054–1063, doi : 10 . 1016/j . csda . 2013 . 02 . 005.
- Engle, R. F. and Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation, and Testing. *Econometrica*, 55(2):251–276, doi : 10 . 2307/1913236.
- Eraker, B., Chiu, C. W., Foerster, A. T., Kim, T. B., and Seoane, H. D. (2015). Bayesian Mixed Frequency VARs. *Journal of Financial Econometrics*, 13(3):698–721, doi : 10 . 1093/jjfinec/nbu027.
- Financial Supervisory Authority (2015). A Model for Household Debt. FI Analysis No. 4, Financial Supervisory Authority.
- Follett, L. and Yu, C. (2019). Achieving Parsimony in Bayesian Vector Autoregressions with the Horseshoe Prior. *Econometrics and Statistics*, doi : 10 . 1016/j . ecosta . 2018 . 12 . 004. Advance online publication.
- Forni, C., Ghysels, E., and Marcellino, M. (2013). Mixed-Frequency Vector Autoregressive Models. *Advances in Econometrics*, 32:247–272, doi : 10 . 1108/s0731-905320130000031012.
- Forni, C. and Marcellino, M. (2013). A Survey of Econometric Methods for Mixed-Frequency Data, doi : 10 . 2139/ssrn . 2268912.
- Forni, C. and Marcellino, M. (2014). Mixed-Frequency Structural Models: Identification, Estimation, and Policy Analysis. *Journal of Applied Econometrics*, 29:1118–1144, doi : 10 . 1002/jae.
- Forni, C. and Marcellino, M. (2016). Mixed Frequency Structural Vector Auto-Regressive Models. *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 179(2):403–425, doi : 10 . 1111/rssa . 12120.
- Frühwirth-Schnatter, S. (1994). Data Augmentation and Dynamic Linear Models. *Journal of Time Series Analysis*, 15(2):183–202, doi : 10 . 1111/j . 1467-9892 . 1994 . tb00184 . x.
- Fry, R. and Pagan, A. (2011). Sign Restrictions in Structural Vector Autoregressions: A Critical Review. *Journal of Economic Literature*, 49(4):938–60, doi : 10 . 1257/jel . 49 . 4 . 938.

- Gefang, D., Koop, G., and Poon, A. (2019). Variational Bayesian Inference in Large Vector Autoregressions with Hierarchical Shrinkage. Manuscript available at <https://sites.google.com/site/garykoop/research>.
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410):398–409, doi:10.1080/01621459.1990.10476213.
- Gelman, A. and Shirley, K. (2011). Inference and Monitoring Convergence. In Brooks, S., Gelman, A., Jones, G. L., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*, pages 163–174. Chapman and Hall/CRC.
- Geman, S. and Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, doi:10.1109/TPAMI.1984.4767596.
- Geyer, C. (2011). Introduction to MCMC. In Brooks, S., Gelman, A., Jones, G. L., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*, pages 3–48. Chapman and Hall/CRC.
- Ghysels, E. (2016). Macroeconomics and the Reality of Mixed Frequency Data. *Journal of Econometrics*, 193(2):294–314, doi:10.1016/j.jeconom.2016.04.008.
- Ghysels, E. and Miller, J. I. (2015). Testing for Cointegration with Temporally Aggregated and Mixed-Frequency Time Series. *Journal of Time Series Analysis*, 36(6):797–816, doi:10.1111/jtsa.12129.
- Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS Regressions: Further Results and New Directions. *Econometric Reviews*, 26(1):53–90, doi:10.1080/07474930600972467.
- Granger, C. W. J. (1981). Some Properties of Time Series Data and Their Use in Econometric Model Specification. *Journal of Econometrics*, 16(1):121–130, doi:10.1016/0304-4076(81)90079-8.
- Groen, J. J. J. and Kleibergen, F. (2003). Likelihood-Based Cointegration Analysis in Panels of Vector Error-Correction Models. *Journal of Business & Economic Statistics*, 21(2):295–318, doi:10.1198/073500103288618972.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, New Jersey.
- Itkonen, J. and Juvonen, P. (2017). Nowcasting the Finnish Economy with a Large Bayesian Vector Autoregressive Model. *Economics Review* 6, Bank of Finland.
- Iversen, J., Laséen, S., Lundvall, H., and Söderström, U. (2016). Real-Time Forecasting for Monetary Policy Analysis: The Case of Sveriges Riksbank. Working Paper No. 318, Sveriges Riksbank.
- Johansen, S. (1988). Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, 12(2):231–254, doi:10.1016/0165-1889(88)90041-3.
- Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, 59(6):1551–1580, doi:10.2307/2938278.
- Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.
- Johansen, S. and Juselius, K. (1990). Maximum Likelihood Estimation and Inference on Cointegration — With Application to the Demand for Money. *Oxford Bulletin*

- of *Economics and Statistics*, 52(2):169–210,
doi:10.1111/j.1468-0084.1990.mp52002003.x.
- Juselius, K. (2006). *The Cointegrated VAR Model: Methodology and Applications*. Oxford University Press USA, Oxford.
- Kadiyala, K. R. and Karlsson, S. (1993). Forecasting with Generalized Bayesian Vector Auto Regressions. *Journal of Forecasting*, 12(3–4):365–378,
doi:10.1002/for.3980120314.
- Kadiyala, K. R. and Karlsson, S. (1997). Numerical Methods for Estimation and Inference in Bayesian VAR-Models. *Journal of Applied Econometrics*, 12(2):99–132.
- Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, doi:10.1115/1.3662552.
- Karlsson, S. (2013). Forecasting with Bayesian Vector Autoregression. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2, chapter 15, pages 791–897. Elsevier B.V.,
doi:10.1016/B978-0-444-62731-5.00015-4.
- Kastner, G. (2019). Sparse Bayesian Time-Varying Covariance Estimation in Many Dimensions. *Journal of Econometrics*,
doi:10.1016/j.jeconom.2018.11.007. Advance online publication.
- Kastner, G. and Frühwirth-Schnatter, S. (2014). Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Estimation of Stochastic Volatility Models. *Computational Statistics and Data Analysis*, 76:408–423,
doi:10.1016/j.csda.2013.01.002.
- Kastner, G., Frühwirth-Schnatter, S., and Freitas Lopes, H. (2017). Efficient Bayesian Inference for Multivariate Factor Stochastic Volatility Models. *Journal of Computational and Graphical Statistics*, 26(4):905–917,
doi:10.1080/10618600.2017.1322091.
- Kastner, G. and Huber, F. (2018). Sparse Bayesian Vector Autoregressions in Huge Dimensions. *Preprint available on arXiv*, arXiv:1704.03239.
- Kilian, L. and Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*. Themes in Modern Econometrics. Cambridge University Press, Cambridge,
doi:10.1017/9781108164818.
- Kim, S., Shephard, N., and Chib, S. (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *The Review of Economic Studies*, 65(3):361–393, doi:10.1111/1467-937X.00050.
- Koop, G., Korobilis, D., and Pettenuzzo, D. (2019). Bayesian Compressed Vector Autoregressions. *Journal of Econometrics*,
doi:10.1016/j.jeconom.2018.11.009. Advance online publication.
- Koop, G. M. (2013). Forecasting with Medium and Large Bayesian VARs. *Journal of Applied Econometrics*, 28(2):177–203, doi:10.1002/jae.1270.
- Koopman, S. J. and Durbin, J. (2000). Fast Filtering and Smoothing for Multivariate State Space Models. *Journal of Time Series Analysis*, 21(3):281–296,
doi:10.1111/1467-9892.00186.
- Kuzin, V., Marcellino, M., and Schumacher, C. (2011). MIDAS vs. Mixed-Frequency VAR: Nowcasting GDP in the Euro Area. *International Journal of Forecasting*, 27(2):529–542, doi:10.1016/j.ijforecast.2010.02.006.

- Leeper, E. M., Sims, C. A., and Zha, T. (1996). What Does Monetary Policy Do? *Brookings Papers on Economic Activity*, 27(2):1–78, doi:10.1257/000282806776157678.
- Lindholm, U., Mossfeldt, M., and Stockhammar, P. (2018). Forecasting Inflation in Sweden. Working Paper No. 152, National Institute of Economic Research.
- Litterman, R. B. (1979). Techniques of Forecasting Using Vector Autoregressions. Working Paper No. 115, Federal Reserve Bank of Minneapolis.
- Louzis, D. P. (2016). Steady-State Priors and Bayesian Variable Selection in VAR Forecasting. *Studies in Nonlinear Dynamics & Econometrics*, 20(5):495–527, doi:10.1515/snde-2015-0048.
- Louzis, D. P. (2019). Steady-State Modeling and Macroeconomic Forecasting Quality. *Journal of Applied Econometrics*, 34(2):285–314, doi:10.1002/jae.2657.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin.
- Lütkepohl, H. and Velinov, A. (2016). Structural Vector Autoregressions: Checking Identifying Long-Run Restrictions via Heteroskedasticity. *Journal of Economic Surveys*, 30:377–392, doi:10.1111/joes.12100.
- Marcellino, M., Porqueddu, M., and Venditti, F. (2016). Short-Term GDP Forecasting With a Mixed-Frequency Dynamic Factor Model With Stochastic Volatility. *Journal of Business & Economic Statistics*, 34, doi:10.1080/07350015.2015.1006773.
- Mariano, R. S. and Murasawa, Y. (2003). A New Coincident Index of Business Cycles Based on Monthly and Quarterly Series. *Journal of Applied Econometrics*, 18(4):427–443, doi:10.1002/jae.695.
- Mariano, R. S. and Murasawa, Y. (2010). A Coincident Index, Common Factors, and Monthly Real GDP. *Oxford Bulletin of Economics and Statistics*, 72(1):27–46, doi:10.1111/j.1468-0084.2009.00567.x.
- McCracken, M. W. and Ng, S. (2016). FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics*, 34(4):574–589, doi:10.1080/07350015.2015.108.
- Pagan, A. (1984). Econometric Issues in the Analysis of Regressions with Generated Regressors. *International Economic Review*, 25(1):221–247, doi:10.2307/2648877.
- Primiceri, G. E. (2005). Time Varying Structural Vector Autoregressions and Monetary Policy. *The Review of Economic Studies*, 72(3):821–852, doi:10.1111/j.1467-937X.2005.00353.x.
- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>.
- Raoufina, K. (2016). Forecasting Employment Growth in Sweden Using a Bayesian VAR Model. Working Paper No. 144, National Institute of Economic Research.
- Reslow, A. and Lindé, J. (2016). En myt att Riksbankens prognoser styrts av modeller. *Ekonomisk debatt*, 44(8):18–30.
- Sanderson, C. and Curtin, R. (2016). Armadillo: A Template-Based C++ Library for Linear Algebra. *Journal of Open Source Software*, 1(2):1–26, doi:10.21105/joss.00026.

- Schorfheide, F. and Song, D. (2015). Real-Time Forecasting with a Mixed-Frequency VAR. *Journal of Business & Economic Statistics*, 33(3):366–380, doi : 10 . 1080/07350015 . 2014 . 954707.
- Shahnazarian, H., Eisensee, T., and Almerud, J. (2015). Bytesbalans och finansiellt sparande. Report from the Economic Affairs Department, Ministry of Finance.
- Shahnazarian, H., Solberger, M., and Spånberg, E. (2017). Forecasting and Analysing Corporate Tax Revenues in Sweden Using Bayesian VAR Models. *Finnish Economic Papers*, 28(1):50–74.
- Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1):1–48, doi : 10 . 2307/1912017.
- Sims, C. A. and Zha, T. (2006). Were There Regime Switches in U.S. Monetary Policy? *American Economic Review*, 96(1):54–81, doi : 10 . 1257/000282806776157678.
- Stock, J. H. and Watson, M. W. (2001). Vector Autoregressions. *Journal of Economic Perspectives*, 15(4):101–115, doi : 10 . 1257/jep . 15 . 4 . 101.
- Stockhammar, P. and Österholm, P. (2016). Do Inflation Expectations Granger Cause Inflation? Working Paper No. 145, National Institute of Economic Research.
- Strid, I. and Walentin, K. (2009). Block Kalman Filtering for Large-Scale DSGE Models. *Computational Economics*, 33(3):277–304, doi : 10 . 1007/s10614-008-9160-4.
- Tanner, M. A. and Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, 82(398):528–540, doi : 10 . 1080/01621459 . 1987 . 10478458.
- Uhlig, H. (2005). What Are the Effects of Monetary Policy on Output? Results From an Agnostic Identification Procedure. *Journal of Monetary Economics*, 52(2):381–419, doi : 10 . 1016/j . jmoneco . 2004 . 05 . 007.
- Villani, M. (2009). Steady-state Priors for Vector Autoregressions. *Journal of Applied Econometrics*, 24(4):630–650, doi : 10 . 1002/jae . 1065.
- Wickham, H. (2015). *R Packages*. O’Reilly, Sebastopol, CA.
- Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksell, Uppsala.
- Wright, J. H. (2013). Evaluating Real-Time VAR Forecasts with an Informative Democratic Prior. *Journal of Applied Econometrics*, 28(5):762–776, doi : 10 . 1002/jae . 2268.
- Yu, Y. and Meng, X. L. (2011). To Center or Not to Center: That is Not the Question—An Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency. *Journal of Computational and Graphical Statistics*, 20(3):531–570, doi : 10 . 1198/jcgs . 2011 . 203main.

Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Social Sciences 170*

Editor: The Dean of the Faculty of Social Sciences

A doctoral dissertation from the Faculty of Social Sciences, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-391500



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2019