# Bayesian Macroeconometrics

Marco Del Negro
*Federal Reserve Bank of New York*

Frank Schorfheide*
*University of Pennsylvania*
*CEPR and NBER*

July 6, 2009

Prepared for
**Handbook of Bayesian Econometrics**

Preliminary and Incomplete – Comments are Welcome

*Correspondence: Marco Del Negro: Research Department, Federal Reserve Bank of New York, 33 Liberty Street, New York NY 10045: marco.delnegro@ny.frb.org. Frank Schorfheide: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu. The views expressed in this chapter do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System.

# Contents

# 1 Bayesian Methods in Macroeconomics

intro

One of the goals of econometric analysis is to provide quantitative answers to substantive economic questions. Examples of such questions are:[c1] (i) What are the main driving forces behind business cycles? Are business cycles correlated across regions and countries? Will GDP decline over the next year and for what reason? (ii) What happens to aggregate output, consumption, investment, and prices in response to unanticipated changes in monetary or fiscal policy? (iii) Was it good luck or good monetary and fiscal policy that caused the decline in output volatility ("Great Moderation") in the mid 1980s? (iv) What is the optimal level of inflation? Should a fiscal authority tax labor income, capital income, or consumption expenditures? (v) Does economic growth lead to democracy or do countries with a democratic government grow faster?

Quantitative answers to some of these questions, e.g., will GDP decline over the next year, require very little economic theory and can be generated, for instance, by simply exploiting serial correlations. Other questions, such as what is the optimal level of inflation, require a sophisticated theoretical model that generates predictions about how agents change their behavior in response to economic policy changes. Thus macroeconometric analysis relies on a large range of model specifications, ranging from simple univariate autoregressions to elaborate dynamic stochastic general equilibrium models.

## 1.1 What are some of the key challenges for econometric analysis?

**Identification:** What patterns in the data allow us to identify an unanticipated change in monetary policy and its effects? How can one infer an aggregate labor supply elasticity from movements in aggregate output and hours worked? Identification has two dimensions: (i) with an infinite amount of data, is the quantity of interest in principle identifiable? (ii) Does a particular data set contain enough information to provide a precise measurement?

**Providing Measures of Uncertainty:** Unfortunately, macroeconomists always face a shortage of observations that are necessary to provide a reliable answers to a particular research question. For instance, even to fairly elementary and widely

[c1] fs: Relate to subsequent empirical illustrations.

studied questions such as what fractions of employment variation is due to technology shocks or what is the relative response of output and inflation to a 25 basis point reduction in the Federal Funds rate, precise answers remain elusive. Answers are sensitive to the choice of theoretical frameworks, empirical models, and data definitions. Documenting this uncertainty in scientific reporting is very important.

**Empirical Fit versus Theoretical Coherence:** Many macroeconomists have a strong preference for models with a high degree of theoretical coherence. This means that rather than postulating functional forms for decision rules of economic agents, these decision rules are derived from some deeper principles such as intertemporal optimization, rational expectations, competitive equilibrium from a specification of preferences and production technologies. Theoretical coherence goes typically hand-in-hand with tight cross-equation restrictions for empirical models that try to capture the joint dynamics of a vector of macroeconomic time series and tends to lead to some form of misspecification.[c1]

**High-Dimensional Models:** Some questions involve high-dimensional empirical models. The analysis domestic business cycles might involve processing information from a large cross section of macroeconomic and financial variables. The study of international comovements is often based on high-dimensional vector autoregressive models, which, if left unrestricted, cannot be reliably estimated with the available data.

## 1.2 How can Bayesian analysis help?

**Identification Issues:** Lack of identification manifests itself in "flat" likelihood functions, both in Bayesian and frequentist analysis. Frequentist analysis, which conditions on a "true" parameter tends to be very sensitive to the strength of identification because it affects the sampling distribution of estimators and test statistics. Much of frequentist analysis relies on either large sample or re-sampling based approximations of small sample distributions of estimators and test statistics. Approximations that are uniformly valid over the parameter space tend to be difficult to obtain if models are only weakly or partially identified.

In Bayesian analysis lack of identification manifests itself in a lack of updating when turning prior distributions into posterior distributions (See Poirier (1998,

[c1] fs: Does this hurt short and long-term forecasting performance?

ADD) or Moon and Schorfheide (2009, ADD)[c2] As long as the prior distributions are proper, so are the posterior distributions, which means that lack of identification poses no conceptual problems. However, as in frequentist analysis it does cause some practical challenges: it becomes more important to document which aspects of the prior distribution do not get updated by the likelihood function and more care is required in the choice of the prior distribution. In models in which Bayesian inference is implemented via numerical methods, the researcher has to insure that these methods work well despite potential lack of identification.

[HvD: Near-boundary features of criterion functions: examining market efficiency and long-run growth patterns. Distinguishing between deterministic and stochastic trend components and between cyclical versus trend components.]

**Combining Different Sources of Observations.** Many macroeconomic models have implications about price and quantity dynamics at the micro level. For instance, the popular Calvo mechanism for generating nominal rigidities in a dynamic stochastic general equilibrium (DSGE) model has implications about the frequency at which firms change their prices.[c1] Thus, micro data on price changes are in principle informative about key parameters of a model that is designed to describe aggregate output, inflation, and interest rate dynamics. In principle, the micro-level information could be incorporated into the estimation of the DSGE model by constructing a likelihood function for both aggregate and disaggregate data. In practice, this approach is not particularly useful for the following reason. While, strictly speaking the Calvo model introduces some heterogeneity across firms, this heterogeneity is not sufficient to explain price dynamics at the micro level. The Bayesian framework offers a coherent alternative of combining information from different sources, namely through prior distributions. In our example, one could use the micro-level data to construct a prior distribution for the parameter associated with the Calvo mechanism. The weight placed on the micro evidence can be controlled through the prior variance.

**Accounting for Uncertainty:** The fact that Bayesian analysis conditions on observed data rather than an unobserved "true" data generating process, and that macroeconomic forecasts and policy decisions are made conditional on the observed empirical evidence, implies that Bayesian methods are well suited for macroeconometric analysis. To the extent that the substantive analysis requires a researcher

[c2] fs: Refine statements.

[c1] fs: Link to discussion of identification.

to consider multiple theoretical and empirical frameworks, Bayesian analysis allows the researcher to assign probabilities to competing specifications and update these probabilities in view of the data.[c2] For instance, to study the effect of monetary policy shocks one could average across vector autoregressions (VARs) or DSGE models that employ different identification schemes or emphasize different propagation channels for monetary policy. Forecasting future inflation or output requires to account for uncertainty about realizations of structural shocks as well as uncertainty associated with parameter estimates. Since shocks and parameters are treated symmetrically in a Bayesian framework, namely as random variables, accounting for these two sources of uncertainty is conceptually straightforward in a Bayesian framework.

**Potential Misspecification:** Studying questions that involve counterfactuals of how agents would behave under economic policies that have not been observed before require models with a high degree of theoretical coherence. In turn these models are likely to suffer from misspecification. Bayesian methods offer a surprisingly rich tool kit for coping with misspecification issues arising in structural modelling. For instance, one can use a more densely parameterized reference model to assess predictions associated with a tightly parameterized structural model. Alternatively, one could use the restrictions associated with the theoretically coherent model only loosely, to center a prior distribution on a richer reference models.

**Too Many Parameters, Too Few Observations:** Consider studying the spill-overs of technology shocks among OECD countries. A natural framework to study this problem is a multi-country VAR model.[c1] However, if lagged variables from each foreign country affect the macroeconomic outcomes in the domestic country the ratio of observations to free parameters is likely to be very small. Of course, one can restrict the dimensionality of this multi-country VAR by simply setting most of the coefficients that capture effects of foreign variables to zero. Unfortunately, such a set of "hard" restrictions rules out the existence of these spill-over effects. Conceptually more appealing is the use of "soft" restrictions, which can be easily incorporated through prior distributions for these coefficients that are centered at zero but have a small, yet non-zero variance. Alternatively, and related to the previous discussion of misspecification, these "soft" restrictions embodied in prior

[c2] fs: In many instances, hypothesis testing is not very useful.

[c1] fs: The subsequent discussion is vague.

distributions could be used to tilt the estimates, of say, a vector autoregression, toward cross-equation restrictions implied by economic theory.

**Efficient Computations:** Many models used to analyze macroeconomic data include latent variables, which might capture the state of the business cycle, e.g., expansion versus contraction, in a univariate regime-switching model for output growth, or state variables of the aggregate economy, such as technology or capital, in a DSGE model. In a Bayesian framework there is no conceptual difference between parameters and latent variables – both are random variables – and data augmentation and Gibbs sampling algorithms provide an efficient way of analyzing models with latent variables.

## 1.3 Outline of this Chapter

Throughout this chapter we will emphasize multivariate models, that can capture comovements of macroeconomic time series. We will begin with a discussion of vector autoregressive models in Section 2, distinguishing between reduced form and structural VARs. Reduced form VARs essentially summarize autocovariance properties of vector time series and can also be used to generate multivariate forecasts. Unfortunately, VARs are not particularly parsimonious. The number of parameters in a VAR increases as a function of the number of endogenous variables and included lags, which in practice often leads to small observation-to-parameter ratios.[c1] Priors are useful tools to sharpen the inference and we discuss various methods of constructing prior distributions. More useful for substantive empirical work in macroeconomics are so-called structural VARs, in which the innovations do not correspond to one-step-ahead forecast errors, but instead are interpreted as structural shocks. Much of the structural VAR literature has focused on studying the propagation of monetary policy shocks, that is changes in monetary policy that are unanticipated by the public. We discuss various identification schemes and their implementation in a Bayesian framework. The remainder of Section 2 is devoted to a discussion of advanced topics such as the estimation of restricted and overidentified VARs. As an empirical illustration, we estimate the effects of a monetary policy, using a 4-variable VAR.

Section 3 is devoted to VARs with explicit restrictions on the long-run dynamics. While many macroeconomic time series are well described by stochastic trend

[c1] fs: Give an example of small.

models, these stochastic trends are often common to several time series. For example, while aggregate consumption and output (in logs) in many countries are well described as random walks with drifts, implying non-stationary behavior of the univariate time series, the ratio (or log difference) of consumption and investment is typically stationary. This observation is consistent with a standard neoclassical growth model (King, Plosser, and Rebelo, 1989 ADD), in which the exogenous technology process follows a random walk. It turns out we can impose such common trends in a VAR by restricting some of the eigenvalues of the characteristic polynomial to unity. VARs with eigenvalue restrictions, written as so-called vector error correction models (VECM) have been widely used in applied work after Engle and Granger (1987 ADD) popularized the concept of cointegration. While frequentist analysis of non-stationary time series models requires a different set of statistical tools, the shape of the likelihood function is largely unaffected by the presence of unit roots in autoregressive models. Nonetheless, the Bayesian literature has experienced a lively debate about how to best analyze VECMs. Most of the controversies are related to the elicitation of prior distributions. In most applications there is uncertainty about the number of cointegration relationships as well as the appropriate number of lags to include in the vector autoregressive model. Hence, our discussion will focus on prior elicitation, posterior inference, and an empirical illustration.

Modern dynamic macroeconomic theory implies fairly tight cross-equation restrictions for vector autoregressive processes and in Section 4 we turn to the discussion of methods suitable to estimated such restricted processes. We refer to these models as dynamic stochastic general equilibrium (DSGE) models. The term DSGE model is often used to refer to a broad class of dynamic macroeconomic models that spans the standard neoclassical growth model discussed in King, Plosser, and Rebelo (1988) as well as the monetary model with numerous real and nominal frictions developed by Christiano, Eichenbaum, and Evans (2005 ADD). A common feature of these models is that decision rules of economic agents are derived from assumptions about preferences and technologies by solving intertemporal optimization problems. Moreover, agents potentially face uncertainty with respect to, for instance, total factor productivity or the nominal interest rate set by a central bank. This uncertainty is generated by exogenous stochastic processes or shocks that shift technology or generate unanticipated deviations from a central bank's interest-rate feedback rule. Conditional on distributional assumptions for the exogenous shocks, the DSGE model generates a joint probability distribution for the

endogenous model variables such as output, consumption, investment, and inflation. Much of the econometric work related to DSGE models employs Bayesian methods. Section 4 discusses the estimation of linearized as well as nonlinear DSGE models and reviews various approaches to evaluate the empirical fit of DSGE models. Some of the methods are illustrated by estimating a simple stochastic growth model based on U.S. labor productivity and hours worked data.

The dynamics of macroeconomic variables tend to change over time. These changes might be a reflection of inherent nonlinearities of the business cycle or they might be caused by the introduction of new economic policies or the formation of new institutions. Such changes can be captured by econometric models with time-varying coefficients. Thus, we augment the VAR models of Section 2 and the DSGE models of Section 4 with time-varying parameters. We distinguish between models in which parameters evolve according to a potentially non-stationary autoregressive law of motion and model in which parameters evolve according to a finite-state Markov-switching process. If time-varying coefficients are introduced in a DSGE model, an additional layer of complication arises. When solving for the equilibrium law of motion one has to take into account that agents are aware that parameters are not constant over time and hence adjust their decision rules accordingly.

Due to the rapid advances in information technologies, macroeconomists have now access to and the ability to process data sets with a large cross-sectional as well as a large time series dimension. The key challenge for econometric modelling is to avoid the proliferation of parameters. Parsimonious empirical models for large data sets can be obtained in several ways. We consider restricted large-dimensional vector autoregressive models as well as factor models. The latter class of models assumes that the comovement between variables is due to a relatively small number of common factors, which in the context of a DSGE model could be interpreted as the most important economic state variables. These factors are typically unobserved and follow some vector autoregressive law of motion. We study empirical models for so-called data-rich environments in Section 6.

Throughout the various sections of the chapter we will encounter uncertainty about model specifications, e.g., the number of lags in a VAR, the importance of certain types of propagation mechanisms in DSGE models, the presence of time-variation in coefficients, the number of factors in a dynamic factor model. A treatment of Bayesian model selection and, more generally, decision-making under model

uncertainty is provided in Section 7. Finally, Section 8 offers a few concluding remarks.

Finally, a word on notation.[c1] We generally summarize our observations $y_1, \ldots, y_T$ as $Y$ or $Y^T$ if it is important to indicate the length $T$ of the sample. $\theta$ serves as generic parameter vector, $p(\theta)$ is the density associated with the prior distribution, $p(Y|\theta)$ is the likelihood function, and $p(\theta|Y)$ the posterior density. With respect to notation for probability distributions, we tend to follow Bauwens, Lubrano, and Richard [ADD, Appendix A]. $\mathcal{U}(a,b)$ denotes a uniform distribution on the interval $(a,b)$. $G(\alpha,\beta)$ is the Gamma distribution with density $p(x|\alpha,\beta) \propto x^{\alpha-1}\exp(-x/\beta)$, where $\propto$ denotes proportionality. The Chi-squared distribution is denoted by $\chi^2(\nu)$ and identical to $G(\nu/2,2)$. We distinguish between the Inverted Gamma-2 and Inverted Gamma-1 distribution: $X \sim IG_2(s,\nu)$ if and only if $\sqrt{X} \sim IG_1(s,\nu)$ and $X^{-1} \sim G(\nu/2, 2/s)$. The inverted gamma-2 distribution is often used for residual variances: $\sigma^2 \sim IG(s,\nu)$ implies that $p(\sigma^2|s,\nu) \propto (\sigma^2)^{-(\nu+2)/2}\exp(-s/(2\sigma^2))$. A random variable $0 \leq X \leq 1$ has a $B(\alpha,\beta)$ distribution if its density is of the form $p(x|\alpha,\beta) \propto x^{\alpha-1}(1-x)^{\beta-1}$. The normal distribution is denoted by $N(\mu,\sigma^2)$ and $X \sim t(\mu,s,m,\nu)$ if its density is $p(x|\mu,s,m,\nu) \propto [s+m(x-\mu)^2]^{-(\nu+1)/2}$.

A $p$-variate normal distribution is denoted by $N_p(\mu,\Sigma)$. We say that a $p \times q$ matrix $X$ is matrix-variate normal $MN_{p\times q}(M, Q\otimes P)$, meaning that $vec(X) \sim N_{pq}(vec(M), Q\otimes P)$ if $p(X|M,Q\otimes P) \propto \exp\{-\frac{1}{2}tr[Q^{-1}(X-M)'P^{-1}(X-M)]\}$, where $\otimes$ is the Kronecker product, $vec(\cdot)$ stacks the columns of a matrix, and $tr[\cdot]$ is the trace operator. The Inverted Wishart distribution is a multivariate generalization of the $IG_2$ distribution: a $q\times q$ matrix $\Sigma$ has $IW_q(S,\nu)$ distribution if $p(\Sigma|S,\nu) \propto |\Sigma|^{-(\nu+q+1)/2}\exp\left\{-\frac{1}{2}tr[\Sigma^{-1}S]\right\}$. Finally, if $X|\Sigma \sim MN_{p\times q}(M,\Sigma\otimes P)$ and $\Sigma \sim IW_q(S,\nu)$, we say that $(X,\Sigma) \sim MNIW(M,P,S,\nu)$. If there is no ambiguity about the dimension of the random vectors and matrices we drop the subscripts that signify dimensions, e.g. we write $N(\mu,\Sigma)$ instead of $N_p(\mu,\Sigma)$.

[c1] fs: To be removed.

# 2 Vector Autoregressions

var1

At first glance, VARs appear to be straightforward multivariate generalizations of univariate autoregressive models. At second sight, they turn out to be one of the key empirical tools in modern macroeconomics. Sims (1980) proposed that VARs should be used to replace large-scale macroeconometric models inherited from the 1960s, because the latter imposed *incredible* restrictions, which were largely inconsistent with the notion that economic agents take the effect of today's choices on tomorrow's utility into account. Since then, VARs have been used for macroeconomic forecasting and policy analysis, to investigate the sources of business cycle fluctuations, and to provide a benchmark against which modern dynamic macroeconomic theories can be evaluated. In fact, in Section 4 it will become evident that the equilibrium law of motion of many dynamic stochastic equilibrium models can be well approximated by a VAR. The remainder of this section is organized as follows. We derive the likelihood function of a reduced-form VAR in Section 2.1. Section 2.2 reviews popular prior distributions for VAR coefficients as well as posterior inference. Section 2.3 is devoted to structural VARs in which innovations are expressed as functions of structural shocks with a particular economic interpretation, e.g., an unanticipated change in monetary policy. Finally, Section 2.4 provides some suggestions for further reading.

Insert Figure Here

## 2.1 Preliminaries

preliminaries

Figure 1 depicts the evolution of three important quarterly macroeconomic time series for the U.S. over the period from 1964:Q1 to 2006:Q4: percentage deviations of GDP from a linear time trend, annualized inflation rates computed from the GDP deflator, and the effective Federal Funds interest rate. Vector autoregressions are linear time series models, designed to capture the joint dynamics of multiple time series. We will illustrate the VAR analysis using the three series plotted in Figure 1. Let $y_t$ be a $n \times 1$ random vector that takes values in $\mathbb{R}^n$, where $n = 3$ in our empirical illustration. The evolution of $y_t$ is described by the $p$'th order difference equation:

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_c + u_t. \tag{1}$$

We refer to (1) as reduced form representation of the VAR, because the $u_t$'s are simply one-step ahead forecast errors and do not have a specific economic interpretation. In order to characterize the conditional distribution of $y_t$ given its history, one has to make a distributional assumption for $u_t$. We shall proceed under the assumption that $u_t \sim N(0, \Sigma)$ and independent over time. We are now in a position to characterize the joint distribution of a sequence of observations. Let $Y_{t_0,t_1}$ denote the sequence $\{y_{t_0}, \ldots, y_{t_1}\}$, let $k = np + 1$, and define the $k \times n$ matrix $\Phi = [\Phi_1, \ldots, \Phi_p, \Phi_c]'$. The joint density of $Y_{1,T}$ conditional on $Y_{1-p,0}$ and the coefficient matrices $\Phi$ and $\Sigma$ is called (conditional) likelihood function and can be factorized as

$$p(Y_{t,T}|\Phi, \Sigma, Y_{1-p,0}) = \prod_{t=1}^{T} p(y_t|\Phi, \Sigma, Y_{1-p,t-1}). \tag{2}$$

This conditional likelihood function can be conveniently expressed, if the VAR is written as a multivariate linear regression model in matrix notation. Let $Y$ be a $T \times n$ matrix with rows $y_t'$, $x_t$ be the $k \times 1$ vector $x_t = [y_{t-1}', \ldots, y_{t-p}', 1]'$, and $X$ be the $T \times k$ matrix with rows $x_t'$. In slight abuse of notation we abbreviate $p(Y_{t,T}|\Phi, \Sigma, Y_{1-p,0})$ by $p(Y|\Phi, \Sigma)$:

$$\begin{aligned} p(Y|\Phi, \Sigma) \quad \propto \quad & |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}\hat{S}]\right\} \\ & \times \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(\Phi - \hat{\Phi})'X'X(\Phi - \hat{\Phi})]\right\}. \end{aligned} \tag{3}$$

Here,

$$\hat{\Phi} = (X'X)^{-1}X'Y, \quad \hat{S} = (Y - X\hat{\Phi})'(Y - X\hat{\Phi}), \tag{4}$$

that is, $\hat{\Phi}$ is the maximum-likelihood (MLE) estimator of $\Phi$ and $\hat{S}$ is a matrix with sums of squared residuals.

If we combine the likelihood function with the improper prior $p(\Phi, \Sigma) \propto |\Sigma|^{-(n+1)/2}$ we can deduce immediately that

$$\Phi, \Sigma|Y \sim MNIW\left(\hat{\Phi}, (X'X)^{-1}, \hat{S}, T - k\right). \tag{5}$$

Draws from this posterior can be easily obtained by sampling $\Sigma^{(s)}$ from a $IW(\hat{S}, T - k)$ distribution and $\Phi^{(s)}$ from the conditional distribution $MN(\hat{\Phi}, \Sigma^{(s)} \otimes (X'X)^{-1})$. An important challenge in practice is to cope with the dimensionality of the parameter matrix $\Phi$. Consider the data depicted in Figure 1. Our sample consists of 172 observations and each equation of a VAR with $p = 4$ lags has 13 coefficients,

which leaves roughly 13 observations to estimate each coefficient. If the sample is restricted to the post-1982 period, after the disinflation under Fed chairman Paul Volcker, the sample size shrinks to 96 observation and the observation-to-parameter ratio drops to about 7.4. Now imagine estimating a two-country VAR for the U.S. and the Euro Area on post-1982 data. This will double the number of parameters and leave less than 4 observations for each coefficient that needs to be estimated. Informative prior distribution can compensate for lack of sample information and we will subsequently discuss alternatives to the improper prior used so far.

## 2.2 Specific Priors

varpriors

Before discussing particular types of prior distributions two points are noteworthy. First, prior distributions can be conveniently represented by dummy observations. This insight dates back at least to Theil and Goldberger (1960 ADD). These dummy observations might be actual observations from other countries, observations generated by simulating a macroeconomic model, or observations generated from introspection. Suppose $T^*$ dummy observations are collected in matrices $Y^*$ and $X^*$ and we use the likelihood function associated with the VAR to relate the dummy observations to the parameters $\Phi$ and $\Sigma$. Using the same arguments that lead to (5), we deduce that $p(Y^*|\Phi,\Sigma)|\Sigma|^{-(n+1)/2}$ can be interpreted as a $MNIW(\underline{\Phi},(X^{*\prime}X^*)^{-1},\underline{S},T^*-k)$ prior for $\Phi$ and $\Sigma$, where $\underline{\Phi}^*$ and $\underline{S}$ are obtained from $\hat{\Phi}$ and $\hat{S}$ in (4) by replacing $Y$ and $X$ with $Y^*$ and $X^*$. Now let $\bar{T}=T+T^*$, $\bar{Y}=[Y^{*\prime},Y']'$, $\bar{X}=[X^{*\prime},X']'$, and $\bar{\Phi}$ and $\bar{S}$ the analogue of $\hat{\Phi}$ and $\hat{S}$ in (4), then we deduce that the posterior of $\Phi$ and $\Sigma$ is $MNIW(\bar{\Phi},(\bar{X}'\bar{X})^{-1},\bar{S},\bar{T}-k)$. Thus, the use of dummy observations leads to a conjugate prior.[1]

Second, any prior distribution that takes the $MNIW$ form, including the previously discussed dummy observation prior, preserves the Kronecker structure of the problem. This essentially implies that the posterior mean of $\Phi$ can be computed equation-by-equation and only involves the inversion of $k\times k$ matrices. If, on the other hand, the prior for $vec(\Phi)$ is normal with an unrestricted covariance matrix, then the computation of the (conditional) posterior mean requires the inversion of $(nk)\times(nk)$ matrices. In the two-country VAR(4) for the U.S. and Euro area mentioned previously it would boil down to inverting matrices of size $25\times 25$ versus

[1] Prior and likelihood are conjugate, if the posterior belongs to the same distributional family as the prior distribution.

$150 \times 150$. We will subsequently discuss two specific prior distributions: the so-called Minnesota prior (Section 2.2.1) as well as a prior for a VAR that is parameterized in terms of a trend and a cycle component (Section 2.2.2).

### 2.2.1 The Minnesota Prior

minnprior

The Minnesota Prior dates back to Litterman (1980, ADD) and Doan, Litterman, and Sims (1984). Our exposition follows the more recent description in Sims and Zha (1998), with the exception that for now we focus on a reduced form, rather than a structural VAR. Consider our lead example, in which $y_t$ is composed of output deviations, inflation, and interest rates, depicted in Figure 1. Notice that all three series are fairly persistent. In fact the univariate behavior of these series, maybe with the exception of post-1982 inflation rates, would be fairly well described by a random walk model of the form $y_{i,t} = y_{i,t-1} + \eta_{i,t}$. The idea behind the Minnesota prior is to center the distribution of $\Phi$ at a value that implies univariate random walk behavior of the components of $y_t$.[2] This prior can be implemented either by directly specifying a distribution for $\Phi$ or, alternatively, through dummy observations. We will pursue the latter route for the following reason. While it is fairly straightforward to choose prior means and variances for the elements of $\Phi$, it tends to be difficult to elicit beliefs about the correlation between elements of the $\Phi$ matrix. After all, there are $nk(nk+1)/2$ of them. At the same time, setting all these correlations to zero potentially leads to a prior that assigns a lot of probability mass to parameter combinations that imply quite unreasonable dynamics for the endogenous variables $y_t$. The use of dummy observations, provides a parsimonious way of introducing plausible correlations between parameters.

The Minnesota prior depends on several hyperparameters. Let $Y_0$ be a pre-sample and define $s = std(Y_0)$ and $\bar{y} = mean(Y_0)$. The remaining hyperparameters are stacked in the 5 vector $\lambda$ with elements $\lambda_i$. Suppose that $n = 2$ and $p = 2$. We begin with dummy observations that generate a prior distribution for $\Phi_1$. The

[2] The random walk approximation is taken for convenience and could be replaced by other statistical representations. For instance, if some series have very little serial correlation because they have been transformed to induce stationarity, e.g., log output has been converted into output growth, then an *iid* model might be preferable. In Section 4 we will discuss how DSGE model restrictions could be used to construct a prior.

hyperparameter $\lambda_1$ controls the overall tightness of the prior:

$$\begin{array}{rcl} Y^* & = & X^*\Phi + U \\ \left[\begin{array}{cc} \lambda_1 s_1 & 0 \\ 0 & \lambda_1 s_2 \end{array}\right] & = & \left[\begin{array}{ccccc} \lambda_1 s_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 s_2 & 0 & 0 & 0 \end{array}\right]\Phi + \left[\begin{array}{cc} u_{11} & u_{12} \\ u_{21} & u_{22} \end{array}\right]. \end{array}$$

Notice that the first observation implies

$$\lambda_1 s_1 = \lambda_1 s_1 \beta_{11} + u_{11}, \quad 0 = \lambda_1 s_1 \beta_{21} + u_{12},$$

which implies that $\beta_{11} \sim \mathcal{N}(1, \Sigma_{11}/(\lambda_1^2 s_1^2))$ and $\beta_{21} \sim \mathcal{N}(0, \Sigma_{22}/(\lambda_1^2 s_1^2))$. Here $\Sigma_{ij}$ denotes element $i, j$ of $\Sigma$.[3] The prior for $\Phi_2$ is implemented with the dummy observations

$$\left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \end{array}\right] = \left[\begin{array}{ccccc} 0 & 0 & \lambda_1 s_1 2^{\lambda_2} & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 s_2 2^{\lambda_2} & 0 \end{array}\right]\Phi + U,$$

where the hyperparameter $\lambda_2$ is used to scale the prior covariance matrix for coefficients associated with $y_{t-l}$ according to $l^{\lambda_2}$. A prior for the covariance matrix $\Sigma$, "centered" at a matrix that is diagonal with elements equal to the pre-sample variance of $y_t$, can be obtained by $\lambda_3$ replications of the observations

$$\left[\begin{array}{cc} s_1 & 0 \\ 0 & s_2 \end{array}\right] = \left[\begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]\Phi + U.$$

The remaining sets of dummy observations provide a prior for the intercept $\Phi_0$ and will generate some *a priori* correlation between the coefficients. They favor unit roots and cointegration, which is consistent with the beliefs of many applied macroeconomists, and they tend to improve VAR forecasting performance. The *sums-of-coefficients* dummy observations, introduced in Doan, Litterman, and Sims (1984), capture the view that when the average of lagged values of a variable is at some level $\bar{y}_i$, that same value $\bar{y}_i$ is likely to be a good forecast of $y_{i,t}$, regardless of the value of other variables:

$$\left[\begin{array}{cc} \lambda_4 \bar{y}_1 & 0 \\ 0 & \lambda_4 \bar{y}_2 \end{array}\right] = \left[\begin{array}{ccccc} \lambda_4 \bar{y}_1 & 0 & \lambda_4 \bar{y}_1 & 0 & 0 \\ 0 & \lambda_4 \bar{y}_2 & 0 & \lambda_4 \bar{y}_2 & 0 \end{array}\right]\Phi + U.$$

[3] Consider a regression $y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + u_t$ and suppose that the standard deviation of $x_{j,t}$ is $s_j$. If we define $\tilde{\beta}_j = \beta_j s_j$ and $\tilde{x}_{j,t} = x_{j,t}/s_j$ then the transformed parameters interact with regressors that have the same scale. Suppose we assume that $\tilde{\beta}_j \sim \mathcal{N}(0, \lambda^2)$, then $\beta_j \sim \mathcal{N}(0, \lambda^2/s_j^2)$. The $s_j$ terms that appear in the definition of the dummy observations achieve this scale adjustment.

The *co-persistence dummy observations*, proposed by Sims (1993, to be added) reflect the belief that when data on all $y$'s are stable at their initial levels, they tend to persist at that level:

$$\left[\begin{array}{cc} \lambda_5\bar{y}_1 & \lambda_5\bar{y}_2 \end{array}\right] = \left[\begin{array}{ccccc} \lambda_5\bar{y}_1 & \lambda_5\bar{y}_2 & \lambda_5\bar{y}_1 & \lambda_5\bar{y}_2 & \lambda_5 \end{array}\right]\Phi + U.$$

The strength of this belief is controlled by $\lambda_5$. This set of dummy observations introduces correlations in prior beliefs about all coefficients, including the intercept, in a given equation.

The VAR estimates tend to be sensitive to the choice of hyperparameter. If $\lambda = 0$ then all the dummy observations are zero and the VAR is estimated under an improper prior. The larger $\lambda$ the stronger more weight is placed on the Minnesota prior vis-a-vis the likelihood function. From a practitioners view, choosing $\lambda$ based on the marginal likelihood function

$$p_\lambda(Y) = \int p(Y|\Phi,\Sigma)p(\Phi,\Sigma|\lambda)d(\Phi,\Sigma)$$

tends to work well for inference as well as forecasting purposes.

The exact implementation of the Minnesota prior differs across studies and we will provide an illustration in the context of our output-inflation-interest VAR in Section 2.3. Many researchers do not use the co-persistence and own-persistence dummy variables. In fact, the prior is often implemented directly, without the use of dummy observations, by assuming that $vec(\Phi) \sim \mathcal{N}(vec(\underline{\Phi}), \underline{V})$, where the prior mean $vec(\underline{\Phi})$ captures the univariate random walk representations and $\underline{V}$ is a diagonal prior covariance matrix. Kadiyala and Karlsson (1997 ADD) consider different numerical approaches of implementing posterior inference and the computation of multi-step forecasts for VARs with various versions of the Minnesota prior as well as uninformative priors. Ni and Sun (2003) studies the frequentist risk of Bayes estimators in VARs under various popular informative and non-informative prior distributions.

### 2.2.2 A Prior for an Alternative VAR Parameterization

alternprior

The parameterization of econometric models is never unique.[c1] An attractive alternative to (1) is the following parameterization studied in Villani (2008):

$$y_t = \Gamma_0 + \Gamma_1 t + \widetilde{y}_t, \quad \widetilde{y}_t = \Phi_1\widetilde{y}_{t-1} + \ldots + \Phi_p\widetilde{y}_{t-p} + u_t. \tag{6}$$

[c1] fs: I was looking for the paper by Albert Marcet and Marek Jarocinski.

The first term, $\Gamma_0 + \Gamma_1 t$ captures the deterministic trend of $y_t$, whereas the second part, the law of motion of $\widetilde{y}_t$ captures stochastic fluctuations around the deterministic trend. These fluctuations could either be stationary or non-stationary. Thus, instead of imposing a prior distribution on $\Phi$ in (1) one can specify a prior for $\Gamma_j$, $j = 1, 2$, and $\Phi_j$, $j = 1, \ldots, p$ in (6). In the latter case, it is straightforward to separate beliefs about the deterministic trend from beliefs about the persistence of fluctuations around this trend.

The following univariate example is instructive. Consider an AR(1) model of the form

$$y_t = \phi_1 y_{t-1} + \phi_c + u_t, \quad u_t \sim \mathcal{N}(0, 1), \tag{7}$$

which could be applied to any of the three series depicted in Figure 1. Moreover, consider the following two prior distributions. Under Prior 1 $\phi_1$ and $\phi_c$ are independent: $\phi_1 \sim \mathcal{U}[0, 1-\xi]$ where $\xi > 0$ and $\phi_c \sim \mathcal{N}(\underline{\phi}_c, \lambda^2)$. This prior implies that conditional on $\phi_1$ the prior mean and variance for the expected value $\mathbb{E}[y_t] = \phi_c/(1-\phi_1)$ increase (in absolute value) as $\phi_1 \longrightarrow 1-\xi$. In turn, this generates a fairly diffuse marginal distribution of $y_t$ that might place little mass on values of $y_t$ that the researcher finds *a priori* plausible. Now suppose that under Prior 2 $\phi_1 \sim \mathcal{U}[0, 1-\xi]$ as before, but $\phi_c|\phi_1 \sim \mathcal{N}(\underline{\gamma}(1-\phi_1), \lambda^2(1-\theta_1)^2)$. This prior guarantees that the prior distribution of $\mathbb{E}[y_t]$ has mean $\underline{\gamma}$ and variance $\lambda^2$ for every value of $\theta_1$. This prior can be easily implemented by the following re-parameterization:

$$y_t = \gamma + \widetilde{y}_t, \quad \widetilde{y}_t = \phi_1 \widetilde{y}_{t-1} + u_t.$$

Now let $\phi_1 \sim \mathcal{U}[0, 1-\xi]$ and $\gamma \sim \mathcal{N}(\underline{\gamma}, \lambda^2)$. This prior has been used, for instance, in Schotman and Van Dijk (1991 ADD) in the context of unit-root testing.[c1]

A few remarks are in order: (i) Draws from the posterior of a VAR parameterized according to (6) can be obtained via Gibbs sampling, by iterating over the conditional distributions of $\Gamma_0$ and $\Gamma_1$ (MN), $\Phi = [\Phi_1, \ldots, \Phi_p]'$ (MN), and $\Sigma$ (IW). (ii) As the roots of the characteristic polynomial associated with $\Phi$ approach unity, some elements of $\Gamma_0$ may not be identifiable anymore. This is easily seen in the AR(1) example. If $\phi_1 = 1$, then it is impossible to distinguish $\gamma$ and $\widetilde{y}_1$. Thus, in practice, proper priors for the intercept and the initialization of the latent $\widetilde{y}_t$ process are advisable. (iii) The co-persistence dummy observations discussed in Section 2.2.1 have the purpose to control prior beliefs about the long-run mean of $y_t$. Translated into the AR(1) example, the co-persistence prior implies that $\phi_c|\phi_1 \sim \mathcal{N}(\bar{y}(1-\phi_1), 1/\lambda_4^2)$.

[c1] fs: Verify this statement.

## 2.3 Structural VARs

svar

The innovations $u_t$ in the VAR specified in (1) have the interpretation of one-step-ahead forecast errors. As long as $\Sigma$ is not diagonal, the forecast errors for the components of $y_t$ are correlated with each other. More importantly, the forecast errors do not have any specific economic interpretation. Hence, (1) is referred to as reduced form VAR. Dynamic macroeconomic theory suggests that the one-step ahead forecast errors are functions of some fundamental innovations, for instance to aggregate technology, preferences, or monetary policy. A structural VAR is an autoregressive model in which the forecast errors are explicitly linked to such fundamental innovations. We will provide an empirical illustration in which we fit a VAR(4) to our output, inflation, and interest rate data and estimate the dynamic effect of an unanticipated change in monetary policy.

A straightforward calculation shows that additional restrictions need to be imposed, for a structural VAR to be identified. Let $\epsilon_t$ be a vector of orthogonal structural shocks with unit variances. We now express the one-step ahead forecast errors as a linear combination of structural shocks

$$u_t = \Phi_\epsilon \epsilon_t = \Sigma_{tr} \Omega \epsilon_t. \tag{8}$$

Here $\Sigma_{tr}$ refers to the unique lower triangular Cholesky factor of $\Sigma$ and $\Omega$ is an arbitrary orthonormal matrix. The second equality ensures that the covariance matrix of $u_t$ is preserved, that is, $\Phi_\epsilon$ has to satisfy the restriction $\Sigma = \Phi_\epsilon \Phi_\epsilon'$. The fact that $\Omega$ can be *any* $n \times n$ orthonormal matrix creates an identification problem. In the remainder of Section 2.3 we will discuss the identification of structural shocks and posterior inference in structural VAR models in more detail.

### 2.3.1 (Lack of) Identification

It is instructive to examine the effect of the identification problem on the calculation of posterior distributions. Our structural VAR is parameterized in terms of the reduced form parameters $\Phi$ and $\Sigma$ and the orthonormal matrix $\Omega$ with the understanding that $\Sigma_{tr} = chol(\Sigma)$ is unique. The joint distribution of data and parameters is given by

$$p(Y, \Phi, \Sigma, \Omega) = p(Y|\Phi, \Sigma)\bar{p}(\Phi, \Sigma)p(\Omega|\Phi, \Sigma). \tag{9}$$

Here $p(Y|\Phi,\Sigma)$ is the likelihood function (3), which does not depend on $\Omega$. In other words, the absence of $\Omega$ from the likelihood function is a manifestation of the identification problem.

Without loss of generality, the joint prior density of $\Phi$, $\Sigma$, and $\Omega$ is factorized in a conditional density for $\Omega$, $p(\Omega|\Phi,\Sigma)$, and a marginal density for the reduced form parameters. The prior density for the reduced-form parameters is of the form

$$\bar{p}(\Phi,\Sigma) = \frac{p(\Phi,\Sigma)\mathcal{I}\{(\Phi,\Sigma)\in\mathcal{A}\}}{\int p(\Phi,\Sigma)\mathcal{I}\{(\Phi,\Sigma)\in\mathcal{A}\}d(\Phi,\Sigma)} \tag{10}$$

Here $p(\Phi,\Sigma)$ is a density function defined over the entire domain of the reduced-form parameter space, e.g. a $MNIW$ density, and $\mathcal{A}$ is a subset of the $\Phi-\Sigma$ domain. For instance, $\mathcal{A}$ could correspond to the set of all reduced form parameters for which the VAR is stationary or for which there exists a conditional distribution of $\Omega|\Phi,\Sigma$ and structural shocks are identifiable.[4] Finally, $\mathcal{I}\{A\in\mathcal{A}\}$ denotes the indicator function that is one if $A\in\mathcal{A}$ and zero otherwise.

As long as the conditional density of $\Omega$ is properly normalized for all $\Phi$ and $\Sigma$ such that $\mathcal{I}\{(\Phi,\Sigma)\in\mathcal{A}\}$, we deduce from integrating (9) with respect to $\Omega$ that

$$p(Y,\Phi,\Sigma) = p(Y|\Phi,\Sigma)\bar{p}(\Phi,\Sigma). \tag{11}$$

Thus, the calculation of the posterior distribution of the reduced form parameters is not affected by the presence of the non-identifiable matrix $\Omega$. The conditional posterior density of $\Omega$ can be calculated as follows:

$$p(\Omega|Y,\Phi,\Sigma) = \frac{p(Y|\Phi,\Sigma)\bar{p}(\Phi,\Sigma)p(\Omega|\Phi,\Sigma)}{\int p(Y|\Phi,\Sigma)\bar{p}(\Phi,\Sigma)p(\Omega|\Phi,\Sigma)d\Omega} = p(\Omega|\Phi,\Sigma). \tag{12}$$

Thus, the conditional distribution of the non-identifiable parameter $\Omega$ does not get updated in view of the data. This is a well-known property of Bayesian inference in partially identified models, see for instance Kadane (1974 ADD), Poirier (1998 ADD), and Moon and Schorfheide (2009 ADD). We can deduce immediately, that draws from the joint posterior distribution can be obtained in two steps. First, generate draws $\Phi^{(s)}$ and $\Sigma^{(s)}$, $s=1,\ldots,n_s$, from the posterior of the reduced form parameters, properly accounting for the truncation $\mathcal{I}\{(\Phi,\Sigma)\in\mathcal{A}\}$. Second, pair each reduced form parameter draw with a draw from the conditional prior distribution $p(\Omega|\Phi^{(s)},\Sigma^{(s)})$.

[4] If VARs are identified by sign-restrictions, see below, there might be reduced-form parameter values that are inconsistent with these sign-restrictions and $\mathcal{A}$ is a strict subset of the $\Phi-\Sigma$ domain.

Not surprisingly, much of the literature on structural VARs reduces to arguments about the appropriate choice of $p(\Omega|\Phi,\Sigma)$. Most authors use dogmatic priors for $\Omega$ such that the conditional distribution of $\Omega$ given the reduced form parameters reduces to a point mass. Priors for $\Omega$ are typically referred to as identification schemes because conditional on $\Omega$ the relationship between the forecast errors $u_t$ and the structural shocks $\epsilon_t$ is uniquely determined. The papers by Cochrane (1994), Christiano and Eichenbaum (1999, ADD), and Stock and Watson (2001) provide detailed surveys. Before exploring particular identification schemes that have been employed in the literature, we consider a simple bivariate illustrative example.

Suppose that $n = 2$, $p = 1$, $\Phi_c = 0$, and that the eigenvalues of $\Phi_1$ are all less than one in absolute value. The eigenvalue restriction guarantees that the VAR can be written as infinite-order moving average (MA($\infty$)):

$$y_t = \sum_{j=0}^{\infty} \Phi_1^j \Sigma_{tr} \Omega \epsilon_t. \tag{13}$$

We will refer to the sequence of partial derivatives

$$\frac{\partial y_{t+j}}{\partial \epsilon_t} = \Phi_1^j \Sigma_{tr} \Omega, \quad j = 0, \ldots \tag{14}$$

as impulse response function. The set of orthonormal matrices $\Omega$ can easily be characterized by an angle $\varphi$:

$$\Omega(\varphi) = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \tag{15}$$

where $\varphi \in (-\pi, \pi]$. Each column represents a vector of unit length in $\mathbb{R}^2$ and the two vectors are orthogonal. Notice that $\Omega(\varphi) = -\Omega(\varphi + \pi)$. Thus, rotating the two vectors by 180 degrees simply changes the sign of the impulse response function. We will now consider three different identification schemes, that restrict $\Omega$ conditional on $\Phi$ and $\Sigma$.

**Example 1:** Suppose $y_t$ is composed of output deviations from trend, $\tilde{y}_t$, and the Federal Funds rate, $R_t$, and $\epsilon_t$ consists of innovations to technology, $\epsilon_{z,t}$, and monetary policy, $\epsilon_{R,t}$. That is, $y_t = [\tilde{y}_t, R_t]'$ and $\epsilon_t = [\epsilon_{z,t}, \epsilon_{R,t}]'$. Identification can be achieved by imposing restrictions on the informational structure. For instance, Boivin and Giannoni (2006, ADD) assume in a slightly richer setting that the private sector does not respond to monetary policy shocks contemporaneously. This assumption can be formalized by setting $\varphi = 0$ in (15) conditional on all values of

$\Phi$ and $\Sigma$. Such a restriction on $\Omega$ is typically referred to as short-run identification scheme. A short-run identification scheme was used in the seminal work by Sims (1980).

**Example 2:** Assume that $y_t$ is composed of inflation rates, $\pi_t$, and output growth: $y_t = [\pi_t, \Delta\tilde{y}_t]'$. Moreover, $\epsilon_t = [\epsilon_{R,t}, \epsilon_{z,t}]'$. But now we use the following assumption to identify a monetary policy shock: unanticipated changes in monetary policy shocks do not raise output in the long-run. This identification scheme has been used, for instance, by Cogley and Nason (1994, ADD) and Schorfheide (2000). The long-run response is given by

$$[(I - \Phi_1)^{-1}\Sigma_{tr}]_{(2.)}\Omega_{(.1)}(\varphi) = 0, \tag{16}$$

where $A_{(.j)}$ ($A_{(j.)}$) is the $j$'th column (row) of a matrix $A$. Geometrically, we need to find a vector of unit length that is perpendicular to $[(I-\Phi_1)^{-1}\Sigma_{tr}]'_{(2.)}$. This implies that there exist exactly two values of $\varphi$, shifted by $\pi$ that solve (16). In practice it is common to normalize the direction of the impulse response function, that is, choose $\Omega$ given $\Phi$ and $\Sigma$ to solve (16) *and* such that the monetary policy shock raises prices in the long-run. A long-run identification scheme was initially used by Blanchard and Quah (1989) to identify supply and demand disturbances in a bivariate VAR. Since long-run effects of shocks in dynamic systems are intrinsically difficult to measure, structural VARs identified with long-run schemes often lead to imprecise impulse response function estimates and inference that is very sensitive to lag length choice and pre-filtering of the observations. Leeper and Faust (1999, ADD) provide a detailed discussion.

**Example 3:** The priors in the preceding examples were degenerate. Faust (1998), Canova and Nicolo (2002), and Uhlig (2005) propose to be more agnostic in the choice of $\Omega$. As in Example 2, let $y_t = [\pi_t, \Delta\tilde{y}_t]'$ and $\epsilon_t = [\epsilon_{R,t}, \epsilon_{z,t}]'$. Now assume that an expansionary monetary policy shock raises both prices and output upon impact. Formally, this implies that $\Sigma_{tr}\Omega_{(.1)} \geq 0$ and is referred to as a sign-restriction identification scheme. Since $[\Sigma_{tr}]_{(11)} > 0$, the first inequality implies that $\varphi \in (-\pi/2, \pi/2]$. Since $[\Sigma_{tr}]_{(22)} > 0$, the second inequality generates a lower bound for $\varphi$, such that the values for $\varphi$ that are consistent with the sign-restrictions lie in the interval[c1] $(-\underline{\varphi}(\Sigma), \pi/2]$. To implement Bayesian inference, the researcher now has to specify a prior distribution on the interval $(-\underline{\varphi}(\Sigma), \pi/2]$. More generally, $r$

[c1] fs: Verify the bound.

columns of $\Omega$ characterize subspaces of $\mathbb{R}^n$ and the problem of choosing a prior distribution for $\Omega$ can be viewed as placing probabilities on a collection of subspaces.[5] Uhlig (2005) proposes to use a distribution that is uniform on the relevant collection of subspaces, which in our bivariate example translates into a uniform distribution for $\varphi$, and discusses the extension to higher-dimensional VARs. Moreover, it is possible to impose sign-restrictions not just upon impact but also on responses at horizon $j > 0$. In that case, not all reduced form parameter values might be consistent with the sign restrictions and properly accounting for the truncation in (10) in the posterior simulator becomes important. Uhlig (2005) provides an acceptance sampling algorithm to do so.

Draws from the joint posterior distribution of $\Phi$, $\Sigma$, and $\Omega$ can be easily converted into impulse response functions or variance decompositions. A variance decomposition measures the fraction that each of the structural shock contributes to the overall variance of a particular element of $y_t$. In the stationary bivariate example the covariance matrix is given by

$$\Gamma_{yy} = \sum_{j=0}^{\infty} \Phi_1^j \Sigma_{tr} \Omega \Omega' \Sigma_{tr}' (\Phi^j)'.$$

Let $\mathcal{I}^i$ be matrix for which element $i, i$ is equal to one and all other elements are equal to zero. Then we can define the contribution of the $i$'th structural shock to the variance of $y_t$ as

$$\Gamma_{yy,0}^{(i)} = \sum_{j=0}^{\infty} \Phi_1^j \Sigma_{tr} \Omega \mathcal{I}^{(i)} \Omega' \Sigma_{tr}' (\Phi^j)'. \tag{17}$$

Thus the fraction of the variance of $y_{j,t}$ explained by shock $i$ is $[\Gamma_{yy,0}^{(i)}]_{(jj)} / [\Gamma_{yy,0}]_{(jj)}$. With impulse response and variance decomposition draws in hand, one can compute posterior summary statistics such as means, medians, standard deviations, or pointwise credible sets. Sims and Zha (1999) propose an alternative method to compute credible bands for impulse response functions, which relies on the first few principle components of the covariance matrix of the responses and aims to capture some of the correlation among responses at different horizons.

**Illustration:** We estimate a VAR(4) based on the output, inflation, and interest rate series depicted in Figure 1. In addition, we also include commodity price

[5] A similar problem arises when placing prior probabilities on cointegration spaces and we will provide a more extensive discussion in Section 3.2.3.

inflation as an observable. Our analysis follows Boivin and Giannoni (2006), but we use a Bayesian approach, starting from a Minnesota prior. [ADD: some blurb about the choice of hyperparameters, etc.]. Our identification follows Example 1. We assume that the private sector does not respond contemporaneously to monetary policy shocks. Hence, if the interest rate $R_t$ is the last element of the $4 \times 1$ vector $y_t$, then $\Omega$ is simply the identity matrix for all values of the reduced-form parameters. Impulse responses to a one-standard deviation monetary policy shock are depicted in Figure 2. [ADD: some blurb about the effects of a monetary policy shock.]

INSERT FIGURE HERE

### 2.3.2 An Alternative SVAR Parameterization

svaralter

We introduced structural VARs by expressing the one-step-ahead forecast errors of a reduced form VAR as a linear function of orthogonal structural shocks. Suppose we now pre-multiply both sides of (1) by $\Omega'\Sigma_{tr}^{-1}$ and define $A_0' = \Omega'\Sigma_{tr}^{-1}$, $A_j = \Omega'\Sigma_{tr}^{-1}\Phi_j$, $j = 1, \ldots, p$, and $A_c = \Omega'\Sigma_{tr}^{-1}\Phi_c$ then we obtain:

$$A_0 y_t = A_1 y_{t-1} + \ldots A_p y_{t-p} + A_c + \epsilon_t. \tag{18}$$

Much of the empirical analysis in the Bayesian SVAR literature is based on this alternative parameterization, see for instance, Sims and Zha (1998). The advantage of (18) is that the coefficients have direct behaviorial interpretations. For instance, one could impose identifying restrictions on $A_0$ such that the first equation in (18) corresponds to the monetary policy rule of the central bank. Accordingly, $\epsilon_{1,t}$ would correspond to unanticipated deviations from the expected policy.

A detailed discussion of the Bayesian analysis of (18) is provided in Sims and Zha (1998). As before, let $x_t' = [y_{t-1}', \ldots, y_{t-p}', 1]$, let $Y$, $X$, and $E$ be matrices with rows $y_t'$, $x_t'$, and $\epsilon_t'$ respectively. Moreover, define $A = [A_1, \ldots, A_p, A_c]'$ such that (18) can be expressed as multivariate regression of the form

$$YA_0' = XA + E \tag{19}$$

with likelihood function

$$p(Y|A_0, A) \propto |A_0|^T \exp\left\{-\frac{1}{2} tr[(YA_0 - XA)'(YA_0 - XA)]\right\}. \tag{20}$$

Notice that conditional on $A_0$ the likelihood function is quadratic in $A$, meaning that under a suitable choice of prior, the posterior of $A$ is matrix-variate normal. Sims and Zha (1998) propose prior distributions that share the Kronecker structure of the likelihood function and hence lead to posteriors distributions that can be evaluated with a high degree of numerical efficiency, that is, without having to invert matrices of the dimension $nk \times nk$.

It is convenient to factorize the joint prior density as $p(A_0)p(A|A_0)$ and to assume that the conditional prior distribution of $A$ takes the form[c1]

$$A|A_0 \sim MN\Big(\underline{A}(A_0), \lambda^{-1} I \otimes \underline{V}(A_0)\Big), \tag{21}$$

where the matrix of means $\underline{A}(A_0)$ and the covariance matrix $\underline{V}(A_0)$ are potentially functions of $A_0$. The matrices $\underline{A}(A_0)$ and $\underline{V}(A_0)$ can, for instance, be constructed from the dummy observations presented in Section 2.2.1:

$$\underline{A}(A_0) = (X^{*'}X^*)^{-1}X^{*'}Y^*A_0, \quad \underline{V}(A_0) = (X^{*'}X^*)^{-1}.$$

Combining the likelihood function (20) with the prior (21) leads to a posterior for $A$ that is conditionally matrix-variate normal:

$$A|A_0, Y \sim MN\Big(\bar{A}(A_0), I \otimes \bar{V}(A_0)\Big), \tag{22}$$

where

$$\begin{aligned} \bar{A}(A_0) &= \Big(\lambda\underline{V}^{-1}(A_0) + X'X\Big)^{-1}\Big(\lambda\underline{V}^{-1}(A_0)\underline{A}(A_0) + X'YA_0\Big) \\ \bar{V}(A_0) &= \Big(\lambda\underline{V}^{-1}(A_0) + X'X\Big)^{-1}. \end{aligned}$$

The specific form of the posterior for $A_0$ depends on the form of the prior density $p(A_0)$. The prior distribution typically includes normalization and identification restrictions.

**Example 4:** is based on a structural VAR analyzed by Robertson and Tallman (2001, ADD). The vector $y_t$ is composed of a price index for industrial commodities (PCOM), M2, the Federal Funds Rate (R), real GDP interpolated to monthly frequency ($\tilde{y}$), the consumer price index (CPI), and the unemployment rate (U). The exclusion restrictions on the matrix $A_0$ used by the authors are summarized in Table 1. The structural VAR here is over-identified, because the covariance matrix

[c1] fs: Be consistent in the use of $\lambda$ versus $\lambda^{-1}$.

of the one-step-ahead forecast errors of a VAR with $n = 6$ has in principle 21 free elements, whereas the matrix $A_0$ only has 18 free elements. The first equation represents an information market, the second equation is the monetary policy rule, the third equation describes money demand, and the remaining three equations characterize the production sector of the economy. Despite the fact we have imposed overidentifying restrictions, the system requires a further normalization. In principle we could multiply the coefficients for each equation $i = 1, \ldots, n$ by $-1$, without changing the distribution of the endogenous variables. A common normalization scheme is to impose that the diagonal elements of $A_0$ are all non-negative. For all practical purposes, however, this normalization can be imposed after the posterior draws from the un-normalized coefficient matrices have been generated.

INSERT TABLE HERE

Waggoner and Zha (2003) developed an efficient MCMC algorithm to generate draws from a restricted $A_0$ matrix. For expositional purposes assume that the prior for $A|A_0$ takes the form (21), with the restriction that $\underline{A}(A_0) = \underline{M}A_0$ for some matrix $\underline{M}$ and that $\underline{V}(A_0) = \underline{V}$ does not depend on $A_0$, as is the case for our dummy observation prior. Then the marginal likelihood function for $A_0$ is of the form

$$p(Y|A_0) = \int p(Y|A_0, A)p(A|A_0)dA \propto |A_0|^T \exp\left\{-\frac{1}{2}tr[A_0'\bar{S}A_0]\right\}, \qquad (23)$$

where $\bar{S}$ is a function of the data as well as $\underline{M}$ and $\underline{V}$. Waggoner and Zha (2003) write the restricted columns of $A_0$ as $A_{0(.i)} = U_i b_i$ where $b_i$ is a $q_i \times 1$ vector and $U_i$ an $n \times q_i$, composed of orthonormal column vectors. Under the assumption that $b_i \sim N(\underline{b}_i, \underline{\Omega}_i)$, independently across $i$, we obtain

$$\begin{aligned} p(b_1, \ldots, b_n|Y) &\propto |[U_1 b_1, \ldots, U_n b_n]|^T \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} b_i' U_i'(\bar{S} + \underline{\Omega}_i^{-1}]U_i b_i\right\} \qquad (24) \\ &\propto |[U_1 b_1, \ldots, U_n b_n]|^T \exp\left\{-\frac{T}{2}\sum_{i=1}^{n} b_i' S_i b_i\right\} \qquad (25) \end{aligned}$$

with the understanding that $A_0$ can be recovered from the $b_i$'s. Now consider the conditional density of $b_i|b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n$:

$$p(b_i|Y, b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n) \propto |[U_1 b_1, \ldots, U_n b_n]|^T \exp\left\{-\frac{T}{2} b_i' S_i b_i\right\}$$

Let $V_i$ be a $q_i \times q_i$ matrix such that $V_i' S_i V_i = I$. Moreover, let $w$ be an $n \times 1$ vector perpendicular to each vector $U_j b_j$, $j \neq i$ and define $w_1 = V_i' U_i' w / \|V_i' U_i' w\|$. Choose $w_2, \ldots, w_{q_i}$ such that $w_1, \ldots, w_{q_i}$ form an orthonormal basis for $\mathbb{R}^{q_i}$ and we can write

$$b_i = V_i \sum_{j=1}^{q_i} \beta_j w_j. \tag{26}$$

Now,

$$\begin{aligned} & p(\beta_1, \ldots, \beta_{q_i} | Y, b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n) \\ & \propto \left( \sum_{j=1}^{q_i} |[U_1 b_1, \ldots, \beta_j V_i w_j, \ldots, U_n b_n]| \right)^T \exp\left\{ -\frac{T}{2} \sum_{j=1}^{q_i} \beta_j^2 \right\} \\ & \propto |\beta_1|^T \exp\left\{ -\frac{T}{2} \sum_{j=1}^{q_i} \beta_j^2 \right\} \end{aligned} \tag{27}$$

The last line follows because $w_2, \ldots, w_{q_i}$ by construction fall in the space spanned by $U_j b_j$, $j \neq i$. Thus, $\beta_1$ has a Gamma distribution and $\beta_j$, $2 \leq j \leq q_i$, are normally distributed. Draws from the posterior of $A_0$ can be obtained by Gibbs sampling according to (26) and (27), and letting $A_{0(.i)} = U_i b_i$.

## 2.4 Further VAR Topics

furthervar

The literature on Bayesian analysis is by now extensive and our presentation is by no means exhaustive. Readers who are interested in using VARs for forecasting purposes can find efficient algorithms to efficiently compute such predictions, possibly conditional on the future path of a subset of variables, in Waggoner and Zha (1999). Rubio-Ramrez, Waggoner, and Zha (2008) provide conditions for the global identification of VARs of the form (18). Our exposition was based on the assumption that the VAR innovations are homoskedastic. Extensions to GARCH-type heteroskedasticity can be found, for instance, in Uhlig (1997, ADD) and Pelloni and Polasek (2003). We will discuss VAR models with stochastic volatility in Section 5.

# 3 VARs with Reduced Rank Restrictions

var2

It is well documented that many economic time series such as aggregate output, hours worked, unemployment, and interest rates tend to be very persistent. Sums of coefficients in autoregressive models often tend to be close to one. More formally, the dynamic behavior of an autoregressive process $\phi(L)y_t = u_t$, where $\phi(L) = 1 - \sum_{j=1}^{p} \phi_j L^p$ and $L$ is the lag operator, crucially depends on the roots of the characteristic polynomial $\phi(z)$. If the smallest root is unity and all other roots are outside the unit circle, then $y_t$ is non-stationary, whereas temporal differences $\Delta y_t = (1-L)y_t$ are stationary. Unit root processes are also said to be integrated of order one, I(1), because temporal differencing can induce stationarity. Since a unit root implies that $\phi(1) = 1 - \sum_{j=1}^{p} \phi_j = 0$, one can deduce immediately that the sum of autoregressive coefficients is one, if and only if the $y_t$ has a unit root.

At the same time, it has long been recognized that linear combinations of macroeconomic time series (potentially after a logarithmic transformation) appear to be stationary. An example are the so-called "Great Ratios," such as the consumption-output or investment-output ratio, see Klein and Kosobud (1961, ADD). The left panel of Figure 3 depicts log nominal GDP and nominal aggregate investment for the U.S. over the period from 1965 to 2006. Both series have a clear upward trend and if one were to compute deviations from a simple linear deterministic time trend, these deviations would look fairly persistent and exhibit unit-root-like features. The right panel of Figure 3 shows the log of the investment-output ratio. While the ratio is far from constant, it exhibits no apparent trend and the fluctuations look at first glance stationary, albeit quite persistent. The observation that particular linear combinations of non-stationary economic time series often appear to be stationary has triggered a large literature on cointegration in the mid 1980's, e.g., Engle and Granger (1987, ADD), Johansen (1988, 1991, ADD), and Phillips (1991, ADD).

Insert Figure Here

If a linear combination of I(1) time series is stationary, then these series are said to be cointegrated. Cointegration implies that the series have common stochastic trends that can be eliminated by taking suitable linear combinations. In Section 4 we will discuss how such cointegration relationships arise in a dynamic stochastic general equilibrium framework. For now, we will show that one can impose co-trending

restrictions in a VAR by restricting some of the eigenvalues of its characteristic polynomial to unity. This leads to a reduced rank regression or so-called vector error correction model (Section 3.1). It turns out that such restricted VARs have become an empirically successful as well as useful tool in applied macroeconomics. In Section 3.2 we discuss Bayesian inference in cointegration system under various types of prior distributions.

## 3.1 Cointegration Restrictions

cointrestr

The exposition in this section follows Johansen (1995, ADD). Starting point is the reduced form VAR, specified in (1). For concreteness, we could assume that $y_t$ is composed of log GDP and investment, plottend in Figure 3. Subtracting $y_{t-1}$ from both sides of the equality leads to

$$\Delta y_t = (\Phi_1 - I)y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \Phi_c + u_t. \tag{28}$$

For $j = 1, \ldots, p-1$ define $\Pi_j = -\sum_{i=j+1}^{p} \Phi_p$ and $\Pi_c = \Phi_c$. Then we can rewrite (28) as

$$\Delta y_t = -\Phi(1)y_{t-1} + \Pi_1 \Delta y_{t-1} + \ldots + \Pi_{p-1}\Delta y_{t-p+1} + \Phi_c + u_t, \tag{29}$$

where, in slight abuse of notation, $\Phi(z) = I - \sum_{j=1}^{p} \Phi_j z^j$ is the characteristic polynomial associated with the reduced form VAR. Notice that if the VAR has unit roots, that is, $|\Phi(1)| = 0$, then the matrix $\Phi(1)$ is of reduced rank. If $|\Phi(z)| = 0$ implies that $z = 1$, that is all roots of $\Phi(z)$ are equal to one, then $\Phi(1) = 0$ and $\Delta y_t$ follows a VAR in first differences.

Thus, imposing unit roots on the characteristic polynomial of the VAR is equivalent to parameterizing $\Phi(1)$ in terms of a reduced-rank matrix. This insight has lead to the so-called vector error correction or vector equilibrium correction (VECM) representation:

$$\Delta y_t = \alpha\beta' y_{t-1} + \Pi_1 \Delta y_{t-1} + \ldots + \Pi_{p-1}\Delta y_{t-p+1} + \Pi_c + u_t, \tag{30}$$

studied by Engle and Granger (1987, ADD). Here $\alpha$ and $\beta$ are both $n \times r$ matrices and we let $\Pi_* = \alpha\beta'$. A few remarks are in order. First, it can be easily verified that the parameterization of $\Pi_*$ in terms of $\alpha$ and $\beta$ is not unique: for any non-singular $r \times r$ matrix $A$ we can define $\tilde{\alpha}$ and $\tilde{\beta}$ such that $\Pi_* = \alpha A A^{-1}\beta' = \tilde{\alpha}\tilde{\beta}'$. In addition to the matrix $\alpha$ of dimension $n \times r$ it is useful to define a matrix $\alpha_\perp$ of full column rank and dimension $n \times (n-r)$ such that $\alpha'\alpha_\perp = 0$. The matrix $[\alpha, \alpha_\perp]$ has rank

$n$. The matrix $\alpha_\perp$ is not uniquely defined, but whenever it is used the conclusions depend only on the orthogonality property. Moreover, define $\beta_\perp$ in similar fashion. It can be verified that

$$\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp + \alpha(\beta'\alpha)^{-1}\beta' = I_{n\times n} \tag{31}$$

If $\beta'\alpha$ has full rank then any vector $v$ in $\mathbb{R}^n$ can be decomposed into a vector in the space spanned by $\beta_\perp$, and the space spanned by $\alpha$.

Second, if $|\Phi(z)| = 0$ implies that $z = 1$ or $|z| > 1$, the rank of $\Pi_* = \alpha\beta'$ is $r$ and $\alpha_\perp'\beta_\perp$ has full rank, then according to Granger's celebrated representation theorem (30) implies that $y_t$ can be expressed as

$$y_t = \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp'\sum_{\tau=1}^{t}(u_t + \Pi_c) + \Psi(L)(u_t + \Pi_c) + P_{\beta_\perp}y_0, \tag{32}$$

where $\Gamma = I - \sum_{j=1}^{p-1}\Pi_i$, $P_{\beta_\perp}$ is the matrix that projects onto the space spanned by $\beta_\perp$, and $\Psi(L)u_t = \sum_{j=0}^{\infty}\Psi_j u_{t-j}$ is a stationary linear process. It follows immediately that the $r$ linear combinations $\beta'y_t$ are stationary and that $y_t$ has $n - r$ common stochastic trends. Third, since $\Pi_*$ in (30) is rank deficient, the model is often called a reduced rank regression, first studied by Anderson (1958, ADD).

In the context of our GDP-investment example, visual inspection of Figure 3 suggests appears that the cointegration vector is close to $[1, -1]$. Thus, according to (30) the growth rates of output and investment should be modelled as functions of lagged growth rates as well as the log investment-output ratio. (32) highlights that output and investment have a common stochastic trend. The remainder of Section 3 focuses on the formal Bayesian analysis of the vector error correction model. We will examine various approaches of specifying a prior distribution for $\Pi_*$ and discuss Gibbs samplers to implement posterior inference. In practice the researcher faces uncertainty about the number of cointegration relationships as well as the number of lags that should be included. A discussion of model selection and averaging approaches is deferred to Section 7.

### 3.2 Bayesian Inference on $\Pi_*$

bayescointinf

Define $\Pi = [\Pi_1, \ldots, \Pi_{p-1}, \Pi_c]'$ and assume that $u_t \sim N(0, \Sigma)$. Inspection of (30) suggests that conditional on $\alpha$ and $\beta$, the VECM reduces to a multivariate linear Gaussian regression model. In particular if $\Pi, \Sigma|\alpha, \beta$ is MNIW, then we can deduce

immediately that the posterior $\Pi, \Sigma|Y, \alpha, \beta$ is also of the MNIW form and can easily be derived following the calculations in Section 2. Hence, throughout this subsection, we will simply focus on priors and posterior inference for $\Pi_* = \alpha\beta'$ conditional on $\Pi$ and $\Sigma$.[6] In particular, we assume

$$\Delta y_t = \Pi_* y_{t-1} + u_t, \quad \Pi_* = \alpha\beta', \quad u_t \sim N(0, \Sigma), \tag{33}$$

and tread $\Sigma$ as known. As before, it is convenient to write the regression in matrix form. Let $\Delta Y$, $X$, and U denote the $T \times n$ matrices with rows $\Delta y_t'$, $y_{t-1}'$, and $u_t'$, respectively, such that $\Delta Y = X\Pi_*' + U$.

### 3.2.1 Gaussian or Improper Priors for $\alpha$ and $\beta$

bayescointgew

We begin the analysis of the VECM with priors for $\alpha$ and $\beta$ that are either improper or Gaussian, that is $p(\alpha, \beta) \propto c$ or $\alpha \sim N(\underline{\alpha}, \underline{V}_\alpha)$ and $\beta \sim N(\underline{\beta}, \underline{V}_\beta)$. Geweke (1996, ADD) used such priors to study inference in the reduced rank regression model. Throughout this section we normalize $\beta' = [I_{r\times r}, B'_{r\times(n-r)}]$, where $B$ is to be estimated. This normalization requires that the elements of $y_t$ are ordered such that each of these variables appears in at least one cointegration relationship. Alternatively, one could normalize the length of each column of $\beta$ to one.

In the context of our output-investment illustration, one might find it attractive to center the prior for the cointegration coefficient $B$ at $-1$, reflecting either pre-sample evidence on the stability of the investment-output ratio or the belief in an economic theory that implies that industrialized economies evolve along a balanced growth path along which consumption and output grow at the same rate. We will encounter a DSGE model with such a balanced growth path property in Section 4. An informative prior for $\alpha$ could be constructed from beliefs about the speed at which the economy returns to its balanced growth path in the absence of shocks.

Conditional on an initial observation and the covariance matrix $\Sigma$ (both subsequently omitted from our notation), the likelihood function is of the form

$$p(Y|\alpha, \beta) \propto |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} tr[\Sigma^{-1}(\Delta Y - X\beta\alpha')'(\Delta Y - X\beta\alpha')] \right\}. \tag{34}$$

[6] A Gibbs sampler that iterates over $\Pi, \Sigma|Y, \Pi_*$ and $\Pi_*|Y, \Pi, \Sigma$ can be used to implement inference for the full model (30).

In turn we will derive conditional posterior distributions for $\alpha$ and $\beta$ based on the likelihood (34). We begin with the posterior of $\alpha$. Define $\tilde{X} = X\beta$. Then

$$p(\alpha|Y,\beta) \propto p(\alpha)\exp\left\{-\frac{1}{2}tr[\Sigma^{-1}(\alpha\tilde{X}'\tilde{X}\alpha' - 2\alpha\tilde{X}'\Delta Y)]\right\}. \tag{35}$$

Thus, as long as the prior of $vec(\alpha')$ is flat or normal, the posterior of $vec(\alpha')$ is multivariate normal. If the prior has the same Kronecker structure as the likelihood function, then the posterior is matrix-variate normal. In particular, if $p(\alpha) \propto c$ then

$$\alpha'|Y,\beta \sim MN\Big((\tilde{X}'\tilde{X})^{-1}\tilde{X}'\Delta Y, \Sigma, (\tilde{X}'\tilde{X})\Big).$$

The derivation of the conditional posterior of $\beta$ is more tedious. Partition $X = [X_1, X_2]$ such that the partitions of $X$ conform with the partitions of $\beta' = [I, B']$ and rewrite the reduced rank regression as

$$\Delta Y = X_1\alpha' + X_2B\alpha' + U.$$

Now define $Z = \Delta Y - X_1\alpha'$ and write

$$Z = X_2B\alpha' + U. \tag{36}$$

The fact that $B$ is right-multiplied by $\alpha'$ complicates the analysis. The following steps are designed to eliminate the $\alpha'$ term. Post-multiplying (36) by the matrix $C = [\alpha(\alpha'\alpha)^{-1}, \alpha_\perp]$, yields the seemingly unrelated regression

$$\big[\tilde{Z}_1,\ \tilde{Z}_2\big] = X_2\big[B,\ 0\big] + \big[\tilde{U}_1,\ \tilde{U}_2\big], \tag{37}$$

where

$$\tilde{Z}_1 = Z\alpha(\alpha'\alpha)^{-1}, \quad \tilde{Z}_2 = Z\alpha_\perp, \quad \tilde{U}_1 = U\alpha(\alpha'\alpha)^{-1}, \quad \tilde{U}_2 = U\alpha_\perp$$

Notice that we cannot simply drop the $\tilde{Z}_2$ equations. Through $\tilde{Z}_2$ we obtain information about $\tilde{U}_2$ and hence indirectly information on $\tilde{U}_1$, which sharpens the inference for $B$. Formally, let $\tilde{\Sigma} = C'\Sigma C$ and partition $\tilde{\Sigma}$ conforming with $\tilde{U} = [\tilde{U}_1, \tilde{U}_1]$. Let $\tilde{\Sigma}_{1|2} = \tilde{\Sigma}_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21}$ and $\tilde{Z}_{1|2} = \tilde{Z}_1 - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{Z}_2$. Then we can deduce

$$p(B|\alpha',Y) \propto p(\beta(B))\exp\left\{-\frac{1}{2}tr\left[\tilde{\Sigma}_{1|2}^{-1}(\tilde{Z}_{1|2} - X_2B)'(\tilde{Z}_{1|2} - X_2B)\right]\right\}. \tag{38}$$

As in the case of $\alpha'$, if $\beta(B)$ is combined with a flat or a Gaussian prior, the conditional posterior is normal. In particular, if $p(\beta) \propto c$ then

$$B|Y,\alpha' \sim MN\Big((X_2'X_2)^{-1}X_2'\tilde{Z}_{1|2}, \tilde{\Sigma}_{1|2}, (X_2'X_2)^{-1}\Big).$$

**Illustration:** We fit a VECM with $p = 4$ to the investment and GDP data depicted in Figure 3. Our prior distribution is informative with respect to the cointegration relationship, namely $B \sim N(-1, \lambda)$, and un-informative with respect to the remaining parameters. Some results are plotted in Figure 4.[c1]

INSERT FIGURE HERE

### 3.2.2 A Prior for $\Pi_*$ via Conditioning

bayescointkp

Kleibergen and van Dijk (1994, ADD) and, more recently, Kleibergen and Paap (2002) criticize the use of potentially non-informative priors for $\alpha'$. If the loadings $\alpha$ for the cointegration relationships $\beta' y_{t-1}$, then $\beta(B)$ becomes non-identifiable. Under a diffuse prior for $B$ the conditional posterior of $B$ given $\alpha = 0$ is improper and its density integrates to infinity. The marginal posterior density of $\alpha$ can be written as

$$p(\alpha|Y) \propto p(\alpha) \int p(Y|\alpha, B)dB.$$

Since $\int p(Y|B, \alpha = 0)dB$ determines the marginal density at $\alpha = 0$, the posterior of $\alpha$ tends to favor near-zero values for which the cointegration relationships are poorly identified.

Kleibergen and Paap (2002) propose the following alternative. Starting point is a singular value decomposition of $\Pi_*$, which takes the form $\Pi_* = VDW'$. Here both $V$ and $W$ are orthonormal matrices and $D$ is a diagonal matrix that contains the singular values. Suppose that $V$ is partitioned into

$$V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \tag{39}$$

and $D$ and $W$ are partitioned conformingly. We assume that the dimension of $D_{11}$ is $r \times r$. It can be verified[c1] that the matrix $\Pi_*$ can be expressed as

$$\Pi_* = \beta\alpha' + \beta_\perp \Lambda \alpha'_\perp, \tag{40}$$

where

$$\begin{aligned} \alpha' &= V_{11}D_{11}[W'_{11}, W'_{21}], \quad B = V_{21}V_{11}^{-1}, \quad \beta' = [I, B'], \quad \text{and} \\ \Lambda &= (V'_{22}V_{22})^{-1/2}V_{22}D_{22}W'_{22}(W_{22}W'_{22})^{-1/2}. \end{aligned}$$

[c1] fs: To be completed

[c1] fs: Verify!

For $\Lambda = 0$ the rank of $\Pi_*$ reduces to $r$ and we obtain the familiar $\Pi_* = \beta\alpha'$. Hence, the key idea behind the construction of the prior distribution for $\alpha$ and $\beta$ is to start from a distribution for $\Pi_*$ that ignores the reduced rank restriction, then derive a conditional distribution for $\Pi_*$ given $\Lambda = 0$, and finally to use a change of variables to obtain a distribution for the parameters of interest, $\alpha$ and $\beta$. Thus,

$$p(\alpha, B) \propto p(\Pi_*(\alpha, B, \Lambda)|\Lambda = 0)|J_{\Lambda=0}(\Pi_*(\alpha, B, \Lambda))|, \tag{41}$$

where $J_{\Lambda=0}(\Pi_*(\alpha, B, \Lambda))$ is the Jacobian associated with the mapping between $\Pi_*$ and $(\alpha, B, \Lambda)$. Kleibergen and Paap (2002) use a diffuse prior for $\Pi_*$ and show that the Jacobian has the form

$$|J_{\Lambda=0}(\Pi_*(\alpha, B, \Lambda))| = |\beta'\beta|^{(n-r)/2}|\alpha\alpha'|^{(n-r)/2}. \tag{42}$$

Thus, as $\alpha \longrightarrow 0$ the prior density vanishes and counteracts the divergence of $\int p(Y|\alpha, B)dB$. Details of the implementation of a posterior simulator are provided in Kleibergen and Paap (2002). Strachan (2003) uses a similar idea but argues that it is preferable to impose non-ordinal identifying restrictions on the cointegration vector, i.e. normalizing its length, rather than linear identifying restrictions, i.e. normalizing $\beta' = [I, B']$.

The philosophy behind the prior introduced in this subsection, is very different from the one underlying the empirical illustration in Section 3.2.1. In the empirical application our prior was deliberately chosen to be informative. The particular numerical choice was motivated in part by economic theory and in part by pre-sample observations. The goal of the prior proposed in Kleibergen and Paap (2002) is mainly to correct irregularities in the likelihood function of the VECM, caused by local non-identifiability of $\alpha$ and $\beta$, and otherwise to be agnostic about parameter values.

### 3.2.3 Priors on Cointegration Spaces

bayescointvil

Strachan and Inder (2004, ADD) and Villani (2005) point out that $\beta$ should be interpreted as a characterization of a subspace of $\mathbb{R}^n$. Hence, the problem of choosing a prior distribution is a problem of placing probabilities on a collection of subspaces. We will restrict our exposition to the case $n = 2$ and $r = 1$ and begin by parameterizing $\beta$ in terms of polar coordinates, normalizing its length to one:

$$\beta' = [\cos\theta, \sin\theta].$$

The one-dimensional subspace associated with $\beta(\theta)$ is given by $\lambda\beta(\theta)$, where $\lambda \in \mathbb{R}$. In general, the set of $r$-dimensional subspaces of $\mathbb{R}^n$ is called the Grassman manifold, $\mathcal{G}_{r,n-r}$. For our case of $n = 2$ and $r = 1$ the elements of $\mathcal{G}_{1,2}$ can be indexed by the angle $\theta \in (-\pi/2, \pi/2]$.

Villani (2005) proposes that the uniform distribution on the Grassman manifold can serve as a reference prior for the analysis of cointegration systems. This uniform distribution is defined to be the distribution that is invariant under the group of orthonormal transformations of $\mathbb{R}^n$. For $n = 2$ this group is given by

$$\Omega(\varphi) = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}, \quad \varphi \in (-\pi/2, \pi/2].$$

$\Omega(\varphi)$ leads simply to a rotation around the origin of the subspace spanned by $\beta(\theta)$, mapping $\beta(\theta)$ into $\beta(\theta - \varphi)$[c1]. Thus, the transformation can be represented as a shift of the angle $\theta$. The distribution that is invariant under this transformation is the distribution that is generated by assuming that $\theta \sim U(-\pi/2, \pi/2]$.

Villani (2005) works with the ordinal normalization of the cointegration vector $\beta' = [1, B]$. Thus, let $B = \sin\theta/\cos\theta$. Assuming that $\theta$ has a uniform distribution, a change of variable leads to the following prior for $B$:

$$p(B) = (1 + B^2)^{-1}. \tag{43}$$

Thus, the implied prior for $B$ is a Cauchy distribution. Since $B = \beta_2/\beta_1$ we can deduce that $\beta_1$ and $\beta_2$ are $N(0, 1)$ variables and therefore

$$p(\beta) \propto \exp\left\{-\frac{1}{2}\beta'\beta\right\}. \tag{44}$$

Villani (2005) provides a generalization of the preceding argument to the case $n > 2$ and $r > 1$ and shows that the prior[c2]

$$p(\alpha, \beta, \Sigma) \propto |\Sigma|^{-(p+r+q+1)/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(A + \nu\alpha\beta'\beta\alpha')]\right\}.$$

implies that the subspace spanned by $\beta$ is marginally uniformly distributed on the Grassman manifold. The author also discusses the implementation of a posterior simulator.

[c1] fs: Check that sign is correct

[c2] fs: Check definitions of nuisance parameters.

Strachan and Inder (2004, ADD), on the other hand, proceed with their analysis under the non-ordinal normalization of the cointegration vector and propose methods to induce informative prior distributions on the Grassman manifold. Consider our investment-output example. In Section 3.2.1 we used a prior for $\beta' = [I, B]$, where $B \sim N(-1, \lambda)$. This prior reflects the belief that, approximately, the log investment-output ratio is stationary. In terms of polar coordinates, this belief could alternatively be expressed by replacing the uniform distribution of $\theta \in (-\pi/2, \pi/2]$ with a (scaled) beta distribution centered at $-\pi/4$. A more extensive survey of this literature can be found in Koop, Strachan, van Dijk, and Villani (2004). Koop, Leon-Gonzalez, and Strachan (FC) proposes efficient posterior simulators for cointegrated models with priors on the cointegration space.

# 4 Dynamic Stochastic General Equilibrium Models

dsge

The term DSGE model is often used to refer to a broad class of dynamic macroeconomic models that spans the standard neoclassical growth model discussed in King, Plosser, and Rebelo (1988) as well as the monetary model with numerous real and nominal frictions developed by Christiano, Eichenbaum, and Evans (2005). A common feature of these models is that decision rules of economic agents are derived from assumptions about preferences and technologies by solving intertemporal optimization problems. Moreover, agents potentially face uncertainty with respect to, for instance, total factor productivity or the nominal interest rate set by a central bank. This uncertainty is generated by exogenous stochastic processes or shocks that shift technology or generate unanticipated deviations from a central bank's interest-rate feedback rule. Conditional on distributional assumptions for the exogenous shocks, the DSGE model generates a joint probability distribution for the endogenous model variables such as output, consumption, investment, and inflation. DSGE models can in principle used for a variety of tasks, including the study of sources and propagation of business cycle fluctuations, welfare analysis under counterfactual economic policies, as well as forecasts of the future path of key macroeconomic variables.

In principle, macroeconometric analysis could proceed as follows: specify a DSGE model that is sufficiently rich to address the substantive economic question of interest; derive its likelihood function and fit the model to historical data; answer the questions based on the estimated DSGE model. Unfortunately, this is easier said than done and a number of challenges have to be overcome: estimation and inference procedures have to be implemented efficiently, the theoretical coherence of the DSGE model often generates misspecifications that need to be accounted for in the econometric analysis, and despite the tight parameterization of DSGE models, they often suffer from identification problems.

The remainder of this section is organized as follows. We present a prototypical DSGE model in Section 4.1. The model solution and state-space representation is discussed in Section 4.2. Bayesian inference for the DSGE model parameters and extensions to model with indeterminate equilibria and heteroskedastic shocks, and models solved with nonlinear approximation techniques are discussed in Sections 4.3, 4.4, and 4.5, respectively. Section 4.6 discusses numerous methods of documenting the fit of DSGE models and comparing it to less restrictive models such as vector

autoregressions. Finally, we provide a brief discussion of some empirical applications in Section 4.7. A detailed survey of Bayesian techniques for the estimation and evaluation of DSGE models is provided in An and Schorfheide (2007).

INSERT FIGURE HERE

## 4.1 A Prototypical DSGE Model

dsgemodel

Figure 5 depicts post-war aggregate log output, hours worked, and log labor productivity for the U.S. Both output and labor productivity are plotted in terms of percentage deviations from a linear trend. The simplest DSGE model that tries to capture the dynamics of these series is the neoclassical stochastic growth model. According to this model, an important source of the observed fluctuations in the three series are exogenous changes in total factor productivity. We will illustrate the techniques discussed in this section with the estimation of the stochastic growth model based on observations on labor productivity and hours worked.

The model consists of a representative households and perfectly competitive firms. The representative household maximizes the expected discounted lifetime utility from consumption $C_t$ and hours worked $H_t$:

$$\mathbb{E}_t\left[\sum_{s=0}^{\infty}\beta^{t+s}\left(\ln C_{t+s} - \frac{(H_{t+s}/B_{t+s})^{1+1/\nu}}{1+1/\nu}\right)\right] \tag{45}$$

subject to a sequence of budget constraints

$$C_t + I_t \leq W_t H_t + R_t K_t.$$

The household owns the capital stock $K_t$ and rents it to the firms at the rate $R_t$. Capital accumulates according to

$$K_{t+1} = (1-\delta)K_t + I_t, \tag{46}$$

where $I_t$ is investment and $\delta$ is the depreciation rate. The household uses the discount rate $\beta$ and $B_t$ is an exogenous disturbance that can be interpreted as labor supply shock. If $B_t$ increases then the disutility associated with hours worked falls. Finally, $\nu$ is the so-called Frisch labor supply elasticity. The first-order conditions associated with the household's optimization problem are given by a consumption Euler equation and a labor supply condition:

$$\frac{1}{C_t} = \beta\mathbb{E}\left[\frac{1}{C_{t+1}}(R_{t+1}-(1-\delta))\right] \quad \text{and} \quad \frac{1}{C_t}W_t = \frac{1}{B_t}\left(\frac{H_t}{B_t}\right)^{1/\nu}. \tag{47}$$

The firms rent capital, hire labor services, and produce final goods according to the following Cobb-Douglas technology:

$$Y_t = (A_t H_t)^{\alpha} K_t^{1-\alpha}. \tag{48}$$

The stochastic process $A_t$ represents the exogenous labor augmenting technical progress. The firms solve a static profit maximization problem and choose labor and capital to equate the marginal products of labor and capital with wages and the rental rate of capital:

$$W_t = \alpha \frac{Y_t}{H_t}, \quad R_t = (1-\alpha)\frac{Y_t}{K_t}. \tag{49}$$

An equilibrium is a sequence of prices and quantities such that the representative household maximizes her utility and the firms maximize their profits taking the prices as given. Market clearing implies that

$$Y_t = C_t + I_t. \tag{50}$$

To close the model we will assume a law of motion for the two exogenous processes. Log production technology evolves according to a random walk with drift:

$$\ln A_t = \ln A_0 + (ln\gamma)t + \ln \widetilde{A}_t, \quad \ln \widetilde{A}_t = \rho_a \ln \widetilde{A}_{t-1} + \sigma_a \epsilon_{a,t}, \quad \epsilon_{a,t} \sim iidN(0,1). \tag{51}$$

If $|\rho_a| < 1$ the log of technology is trend stationary. If $|\rho_a| = 1$ then technology follows a random walk process with drift. The preference process is assumed to follow a stationary AR(1) process:

$$\ln B_t = (1-\rho_b)\ln B_* + \rho_b \ln B_{t-1} + \sigma_b \epsilon_{b,t}, \quad \epsilon_{b,t} \sim iidN(0,1) \tag{52}$$

The solution to the rational expectations difference equations (46) to (50) determines the law of motion for the endogenous variables $Y_t$, $C_t$, $I_t$, $K_t$, $H_t$, $W_t$, and $R_t$.

The technology process $\ln A_t$ induces a common trend in output, consumption, investment, capital, and wages. Hence, it is convenient to define the detrended variables as follows:

$$\widetilde{Y}_t = \frac{Y_t}{A_t}, \ \widetilde{C}_t = \frac{C_t}{A_t}, \ \widetilde{I}_t = \frac{I_t}{A_t}, \ \widetilde{K}_{t+1} = \frac{K_{t+1}}{A_t}, \ \widetilde{W}_t = \frac{W_t}{A_t} \tag{53}$$

According to our timing convention, $K_{t+1}$ refers to end of period $t$, beginning of $t+1$, and is a function of shocks dated $t$ and earlier. Hence, we are detrending

$K_{t+1}$ by $A_t$. It is straightforward, to rewrite (46) to (50) in terms of the detrended variables. First, notice that

$$\frac{A_t}{A_{t-1}} = \gamma \widetilde{A}_{t-1}^{\rho_a - 1} e^{\sigma_a \epsilon_{a,t}} \tag{54}$$

This ratio is stationary regardless whether $|\rho_a| < 1$ or $\rho_a = 1$. The transformed equilibrium conditions are

$$\begin{aligned}
\frac{1}{\widetilde{C}_t} &= \beta \mathbb{E}\left[\frac{1}{\widetilde{C}_{t+1}} \frac{A_t}{A_{t+1}} (R_{t+1} - (1-\delta))\right], \quad \frac{1}{\widetilde{C}_t}\widetilde{W}_t = \frac{1}{B_t}\left(\frac{H_t}{B_t}\right)^{1/\nu} \\
\widetilde{W}_t &= \alpha \frac{\widetilde{Y}_t}{H_t}, \quad R_t = (1-\alpha)\frac{\widetilde{Y}_t}{\widetilde{K}_t}\frac{A_t}{A_{t-1}} \\
\widetilde{Y}_t &= H_t^{\alpha}\left(K_t \frac{A_{t-1}}{A_t}\right)^{1-\alpha}, \quad \widetilde{Y}_t = \widetilde{C}_t + \widetilde{I}_t, \quad \widetilde{K}_{t+1} = (1-\delta)\widetilde{K}_t \frac{A_{t-1}}{A_t} + \widetilde{I}_t.
\end{aligned} \tag{55}$$

We will collect the parameters of the DSGE model in the vector $\theta$

$$\theta = [\alpha, \beta, \gamma, \delta, \nu, \rho_a, \sigma_a, \rho_b, \sigma_b]'.$$

If we set the standard deviations of the innovations $\epsilon_{a,t}$ and $\epsilon_{b,t}$ to zero, the model economy becomes deterministic and has a steady state in terms of the detrended variables. This steady state is a function of $\theta$. For instance, the rental rate of capital, the capital-output, and the investment-output ratios are given by:

$$R_* = \frac{\gamma}{\beta} + (1-\delta), \quad \frac{\widetilde{K}_*}{\widetilde{Y}_*} = \frac{(1-\alpha)\gamma}{R_*}, \quad \frac{\widetilde{I}_*}{\widetilde{Y}_*} = \frac{\widetilde{K}_*}{\widetilde{Y}_*}\left(1 - \frac{1-\delta}{\gamma}\right). \tag{56}$$

In a stochastic environment, the detrended variables follow a stationary law of motion, even if the underlying technology shock is non-stationary. Moreover, if $\rho_a = 1$, the model generates a number of cointegration relationships, which according to (53) are obtained by taking pair-wise differences of $\ln Y_t$, $\ln C_t$, $\ln I_t$, $\ln K_{t+1}$, and $\ln W_t$.

## 4.2 Model Solution and State-Space Form

dsgesolution

The solution to the equilibrium conditions (55) leads to a probability distribution for the endogenous model variables, indexed by the vector of structural parameters $\theta$. This likelihood function can be used for Bayesian inference. Before turning to the Bayesian analysis of DSGE models, a few remarks about the model solution are in order. In most DSGE models, the intertemporal optimization problems of economic agents can be written recursively, using Bellman equations. In general, the

value and policy functions associated with the optimization problems are nonlinear both in terms of the state and control variables and the solution of the optimization problems require numerical techniques. In general the solution of the DSGE model can be written as

$$s_t = \Phi(s_{t-1}, \epsilon_t; \theta), \tag{57}$$

where $s_t$ is a vector of suitably defined state variables and $\epsilon_t$ is a vector that stacks the innovations for the structural shocks.

For now, we proceed under the assumption that the DSGE model's equilibrium law of motion is approximated by log-linearization techniques, ignoring the discrepancy between the nonlinear model solution and the first-order approximation. We adopt the convention that if a variable $X_t$ ($\widetilde{X}_t$) has a steady state $X_*$ ($\widetilde{X}_t$), then $\widehat{X}_t = \ln X_t - \ln X_*$ ($\widehat{X}_t = \ln \widetilde{X}_t - \ln \widetilde{X}_*$). The log-linearized equilibrium conditions of the neoclassical growth model (55) are given by the following system of linear expectational difference equations[7]:

$$\begin{aligned}
\widehat{C}_t &= \beta \mathbb{E}_t\left[\widehat{C}_{t+1} - (1-\rho_a)\widehat{A}_{t+1} - \frac{\gamma/\beta}{\gamma/\beta + (1-\delta)}\widehat{R}_{t+1}\right] \\
\widehat{H}_t &= \nu\widehat{W}_t - \widehat{C}_t + (1+\nu)\widehat{B}_t, \quad \widehat{W}_t = \widehat{Y}_t - \widehat{H}_t, \\
\widehat{R}_t &= \widehat{Y}_t - \widehat{K}_t + \widehat{A}_t, \quad \widehat{K}_{t+1} = (1-\delta)\widehat{K}_t + \frac{\delta + 1 - \gamma}{\gamma}\widehat{I}_t + (1-\delta)(1-\rho_a)\widehat{A}_t, \\
\widehat{Y}_t &= \alpha\widehat{H}_t + (1-\alpha)\widehat{K}_t + (1-\alpha)(1-\rho_a)\widehat{A}_t, \quad \widehat{Y}_t = \widehat{C}_t + \widehat{I}_t, \\
\widehat{A}_t &= \rho_a\widehat{A}_{t-1} + \sigma_a\epsilon_{a,t}, \quad \widehat{B}_t = \rho_b\widehat{B}_{t-1} + \sigma_b\epsilon_{b,t}.
\end{aligned} \tag{58}$$

There are a multitude of techniques available to solve linear rational expectations models, for instance, Sims (2002, ADD). Economists focus on solutions that guarantee a stationary law of motion for the endogenous variables, with the loose justification that any non-stationary solution would violate the transversality conditions associated with the underlying dynamic optimization problems. For the neoclassical growth model, the solution takes the form

$$s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t. \tag{59}$$

The system matrices $\Phi_1$ and $\Phi_\epsilon$ are functions of the DSGE model parameters $\theta$ and $s_t$ is composed of three elements: the capital stock at the end of period $t$, $\widehat{K}_{t+1}$, as well as the two exogenous processes $\widehat{A}_t$ and $\widehat{B}_t$. The other endogenous variables, $\widehat{Y}_t$, $\widehat{C}_t$, $\widehat{I}_t$, $\widehat{H}_t$, $\widehat{W}_t$, and $\widehat{R}_t$ can be expressed as linear functions of $s_t$.

[7] Here $\widehat{A}_t = \ln \widetilde{A}_t$

As all DSGE models, the linearized neoclassical growth model has some apparent counterfactual implications. For instance, according to (58) the labor share $\widehat{lsh} = \widehat{H}_t + \widehat{W}_t - \widehat{Y}_t$ is constant, which is clearly at odds with the data. Thus, it is common to estimate DSGE models based on only a subset of the potentially observable variables that appear in the model. In our empirical illustration, we will consider the estimation of the neoclassical stochastic growth model on output and hours data. Fluctuations in our model are generated by two exogenous disturbances, $\widehat{A}_t$ and $\widehat{B}_t$. Thus, likelihood functions for more than two variables will be degenerate, because the model predicts that certain linear combinations of these variables are constant, which is clearly at odds with the data. To cope with this problem authors have either added so-called measurement errors, Sargent (1989, ADD), Altug (1989, ADD), Ireland (2004, ADD), or additional shocks as in Leeper and Sims (1994, ADD) and more recently Smets and Wouters (2003). For the subsequent discussion, we restrict the dimension of the vector of observables $y_t$ to $n = 2$ so that it matches the number of exogenous shocks. The measurement equation takes the form

$$y_t = \Psi_0(\theta) + \Psi_1(\theta)t + \Psi_2(\theta)s_t \tag{60}$$

Thus (59) and (60) provide a state-space representation for the linearized DSGE model. If the innovations $\epsilon_t$ are Gaussian, then the likelihood function can be obtained from the Kalman filter, described in Chapter [Time Series].

In our empirical illustration we define $y_t$ to be composed of log labor productivity and log hours worked. In this case $A(\theta)$ and $B(\theta)$ in (60) represent the following equations

$$\begin{bmatrix} \ln GDP_t/H_t \\ \ln H_t \end{bmatrix} = \begin{bmatrix} \ln \widetilde{Y}_* + \ln A_0 - \ln H_* \\ \ln H_* \end{bmatrix} + \begin{bmatrix} \ln \gamma \\ 0 \end{bmatrix} t + \begin{bmatrix} \widehat{A}_t + \widehat{Y}_t - \widehat{H}_t \\ \widehat{H}_t \end{bmatrix},$$

where $H_*$ is the steady state of hours worked and the variables $\widehat{A}_t$, $\widehat{Y}_t$, and $\widehat{H}_t$ are linear functions of $s_t$. From the measurement equation it is clear that we could have alternatively used log GDP instead of log labor productivity as an observable.

Although we focus on output and hours dynamics in this section, it is instructive to examine the measurement equations that the model yields for output and investment, the two series examined in Section 3. Suppose we use the GDP deflator to convert the series depicted in Figure 3 from nominal into real terms.[8] Then, we can

[8]This conversion is more delicate than it is made to appear here. Our model implies that the relative price of investment goods in terms of consumption goods is one, which is counterfactual.

write

$$\begin{bmatrix} \ln GDP_t \\ \ln I_t \end{bmatrix} = \begin{bmatrix} \ln \widetilde{Y}_* + \ln A_0 \\ \ln \widetilde{I}_* + \ln A_0 \end{bmatrix} + \begin{bmatrix} \ln \gamma \\ \ln \gamma \end{bmatrix} t + \begin{bmatrix} \widehat{A}_t + \widehat{Y}_t \\ \widehat{A}_t + \widehat{I}_t \end{bmatrix}.$$

This representation highlights the common trend, generated by the technology process in output and investment. If $\rho_a = 1$ then (59) implies that $\widehat{A}_t$ follows a random walk process and hence generates a stochastic trend component. Thus, the model generates the following cointegration relationship:

$$\begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \ln GDP_t \\ \ln I_t \end{bmatrix} = \frac{\gamma/\beta + 1 - \delta}{(1-\alpha)(\gamma - 1 + \delta)} + \widehat{Y}_t - \widehat{I}_t.$$

Recall that both $\widehat{Y}_t$ and $\widehat{I}_t$ are stationary, even if $\rho_a = 1$. We used this model implication in Sections 3.2.1 and 3.2.3 as one of the justifications of our informative prior for the cointegration vector.

## 4.3 Bayesian Inference

dsgeinference

We will now estimate the stochastic growth model based on quarterly data on labor productivity and hours worked ranging from 1955 to 2006. Unlike in Figure 5, we do not remove a deterministic trend from the two series prior to estimation. Our analysis begins with the specification of a prior distribution. Most of the literature on Bayesian estimation of DSGE models uses fairly informative prior distributions. However, this should not be interpreted as "cooking up" desired results based on almost dogmatic priors. To the contrary, the spirit behind the prior elicitation is to utilized other sources of information that does not directly enter the likelihood function. To the extent that this information is indeed precise, the use of a tight prior distribution is desirable. If the information is vague, it should translate into a more dispersed prior distribution. Most importantly, the choice of prior should be properly documented.

There are three important sources of information that could be used for the elicitation of prior distribution: (i) information from macroeconomic time series other than output and hours during the period 1955 to 2006; (ii) micro-level observations that are, for instance, informative about labor-supply decisions; (iii) macroeconomic data, including observations on output and hours worked, prior to 1955. Consider source (i). It is apparent from (56), that long-run averages of real interest rates, capital-output ratios, and investment-output ratios are informative about $\alpha$, $\beta$, and

$\delta$. Moreover, the parameter $\alpha$ equal the labor share in our model. Since none of these variables directly enter $Y^T$, it is sensible to incorporate this information through the prior distribution. The parameters $\rho_a$, $\rho_b$, $\sigma_a$, and $\sigma_b$ implicitly affect the persistence and volatility of output and hours worked. Hence, prior distributions for these parameters can be chosen such that the implied dynamics of output and hours are broadly in line with pre-sample evidence, that is, information from source (iii). Finally, micro-econometric estimates of labor supply elasticities could be used to set a prior for the Frisch elasticity $\nu$, accounting for the fact that most of the variation in hours worked at the aggregate level is due to the extensive margin, that is, individuals moving in and out of unemployment.

The prior distribution for our empirical illustration is summarized in Table 2. Based on NIPA data we choose the prior means for $\alpha$, $\beta$, and $\delta$ to be consistent with a labor share of 0.4, an investment-to-output ratio of 28%, and an annual interest rate of 4%. These choices yield values of $\underline{\alpha} = 0.34$, $\underline{\beta} = 0.99$, and $\underline{\delta} = 0.014$ in quarterly terms. We decided to use dogmatic priors for $\beta$ and $\delta$, i.e. these parameters are fixed, and a Beta distribution for $\alpha$ with a standard deviation of 0.02. An important parameter for the behavior of the model is the labor supply elasticity. As discussed in Rios-Rull *et al.* (2009, ADD) *a priori* plausible vary considerably. Micro-level estimates based on middle-age white males yield a value of 0.2, balanced growth considerations under slightly different household preferences suggest a value of 2.0, and Rogerson's (1988, ADD) model of hours variation along the extensive margin would lead to $\nu = \infty$. We use a Gamma distribution with parameters that imply a prior mean of 2 and a standard deviation of 1. Our prior for the technology shock parameters is fairly diffuse with respect to the average growth rate, it implies that the total factor productivity has a serial correlation between 0.91 and 0.99, and the standard deviation of the shocks is about 1% each quarter. Our prior implies that the preference shock is slightly less persistent than the technology shock. Finally, we define $\ln Y_0 = \ln Y_* + \ln A_0$ and use fairly agnostic priors on the location parameters $\ln Y_0$ and $\ln H_*$.

The distributions specified in the first columns of Table 2 are marginal distributions. A joint prior is typically obtained by taking the product of the marginals for all elements of $\theta$, which is what we will do in the empirical illustration. Alternatively, one could replace a subset of the structural parameters by, for instance, $R_*$, $lsh_*$, $\widetilde{I}_*/\widetilde{K}_*$, and $\widetilde{K}_*/\widetilde{Y}_*$, and then regard beliefs about these various steady states as independent. Del Negro and Schorfheide (2008 ADD) propose to augment an initial

prior $p(\theta)$ constructed from marginal distributions for the individual elements of $\theta$ by a quasi-likelihood function that reflects beliefs about steady-state relationships and autocovariances. In a nutshell this quasi-likelihood function is generated by interpreting long-run averages of variables that do not appear in the model and pre-sample autocovariances of $y_t$ as noisy measures of steady states and population autocovariances.

Due to the nonlinear relationship between the DSGE model parameters $\theta$ and system matrices $\Psi_0$, $\Psi_1$, $\Psi_2$, $\Phi_1$ and $\Phi_\epsilon$ it is not possible to obtain useful characterizations of marginal or conditional posterior distributions of $\theta$, despite the linear Gaussian state-space form of (59) and (60). Up to now the most commonly used procedures to generate draws from the posterior distribution of $\theta$ are the Random-Walk Metropolis (RWM) Algorithm described in Schorfheide (2000) and Otrok (2001) or Importance Sampler (IS) proposed in DeJong, Ingram, and Whiteman (2000). The basic RWM Algorithm takes the following form

*Random-Walk Metropolis (RWM) Algorithm*

1. Use a numerical optimization routine to maximize $\ln p(Y|\theta) + \ln p(\theta)$. Denote the posterior mode by $\tilde{\theta}$.
2. Let $\tilde{\Sigma}$ be the inverse of the Hessian computed at the posterior mode $\tilde{\theta}$.
3. Draw $\theta^{(0)}$ from $\mathcal{N}(\tilde{\theta}, c_0^2\tilde{\Sigma})$ or directly specify a starting value.
4. For $s = 1, \ldots, n_{sim}$, draw $\vartheta$ from the proposal distribution $\mathcal{N}(\theta^{(s-1)}, c^2\tilde{\Sigma})$. The jump from $\theta^{(s-1)}$ is accepted ($\theta^{(s)} = \vartheta$) with probability min $\{1, r(\theta^{(s-1)}, \vartheta|Y)\}$ and rejected ($\theta^{(s)} = \theta^{(s-1)}$) otherwise. Here
$$r(\theta^{(s-1)}, \vartheta|Y) = \frac{\mathcal{L}(\vartheta|Y)p(\vartheta)}{\mathcal{L}(\theta^{(s-1)}|Y)p(\theta^{(s-1)})}.$$
5. Approximate the posterior expected value of a function $h(\theta)$ by $\frac{1}{n_{sim}}\sum_{s=1}^{n_{sim}} h(\theta^{(s)})$.

The RWM Algorithm works well if the parameter space is of fairly low dimension and the posterior is unimodal. An and Schorfheide (2007) describe a hybrid MCMC algorithm with transition mixture to deal with a bimodal posterior distribution. Most recently, Chib and Ramamurthy (2009, ADD) have developed a multi-block Metropolis-within-Gibbs algorithm that randomly groups parameters in blocks and thereby dramatically reduces the persistence of the resulting Markov chain and

improves the efficiency of the posterior sampler compared to a single-block RWM algorithm. A detailed discussion can be found in Chapter [Bayesian Computation].

**Illustration:** We apply the RWM Algorithm to generate draws from the posterior distribution of the parameters of our stochastic growth model. Posterior means and 90% credible intervals are summarized in Table 2. We consider two versions of the model. In the deterministic trend version the autocorrelation parameter of the technology shock is estimated subject to the restriction that it lies in the interval $[0, 1)$, whereas it is fixed at 1 in the stochastic trend version. Due to the fairly tight prior, the distribution of $\alpha$ is essentially not updated in view of the data. The labor supply elasticity estimates are 0.44 and 0.69, respectively, which is in line with estimates reported in Rios-Rull *et al.* (2009, ADD). These relatively small values of $\nu$ imply that most of the fluctuations in hours worked are due to the labor supply shock. The estimated shock autocorrelations are around 0.97, and the innovation standard deviations of the shocks are 0.7% for the technology shock and 1.1% for the preference shock. The estimates of $\ln H_*$ and $\ln Y_0$ capture the level of the two series.

## 4.4 Indeterminacy and Stochastic Volatility

dsgeindetsv

Linear rational expectations systems can have multiple stable solutions, which is referred to as indeterminacy. DSGE models that allow for indeterminate equilibrium solutions have received a lot of attention in the literature, because this indeterminacy might arise if a central bank does not react forcefully enough to counteract deviations of inflation from its long-run target value. In an influential paper, Clarida, Gali, and Gertler (2000, ADD) estimated interest rate feedback rules based on U.S. post-war data and found that the policy rule estimated for pre-1979 data would lead to indeterminate equilibrium dynamics in a DSGE model with nominal price rigidities. The presence of indeterminacies raises a few complications for Bayesian inference, described in detail in Lubik and Schorfheide (2004).

Consider the following simple example. Suppose that $y_t$ is scalar and satisfies the expectational difference equation:

$$y_t = \frac{1}{\theta}\mathbb{E}_t[y_{t+1}] + \epsilon_t, \quad \epsilon_t \sim N(0,1), \quad \theta \in (0,2]. \tag{61}$$

It can be verified that if $\theta > 1$ the unique stable solution is $y_t = \epsilon_t$, If on the other hand $\theta \leq 1$, we obtain a much larger class of solutions that can be characterized by

the ARMA(1,1) process

$$y_t = \theta y_{t-1} + (1+M)\epsilon_t - \theta\epsilon_{t-1}.$$

Notice that one needs to introduce additional parameters, here the scalar $M$, to characterize the solutions under indeterminacy. Moreover, the structural parameter $\theta$ drops out of the equilibrium law of motion if $\theta > 1$. Thus, for certain parameterizations of the DSGE model, $\theta$ is only set-identified, meaning that one can only learn from the data that $\theta \in (1, \infty]$, but not the precise value. Thus, indeterminacy exacerbates identification problems in DSGE models. We saw in the analysis of VARs with sign restrictions that lack of identification implies that certain conditional distributions do not get updated through the likelihood function.

One of the most striking features of post-war U.S. GDP data is the reduction in the volatility of output growth around 1984. This phenomenon has be termed the Great Moderation and is also observable in many other industrialized countries. To investigate the sources of this volatility reduction Justiniano and Primiceri (2008) allow the volatility of the structural shocks $\epsilon_t$ in (59) vary over time. In the context of our stochastic growth model, consider for instance the technology shock $\epsilon_{a,t}$. We previously assumed that $\epsilon_{a,t} \sim N(0,1)$. Alternatively, suppose that

$$\epsilon_{a,t} \sim N(0, v_t^2), \quad \ln v_t = \rho_v \ln v_{t-1} + \eta_t, \quad \eta_t \sim N(0, \omega^2). \tag{62}$$

Justiniano and Primiceri (2008) first solved their log-linearized DSGE model with a standard solution technique, ignoring the stochastic volatilities in the structural shock processes, and then used the following Gibbs sampler to conduct inference:

1. Conditional on the sequence $v_{1,T} = \{v_t\}_{t=1}^T$ the likelihood function of the state-space model can still be evaluated with the Kalman filter. Consequently, the RWM step described in Section (4.3) can be used to generate draws from the conditional posterior of $\theta$ given $v_{1,T}$.

2. Draws from the conditional distribution of $\epsilon_{a,1,T}$ given $\theta$ and $v_{1,T}$ can be obtained by using the Kalman smoother as in Carter and Kohn (1994), described in Chapter [Time Series].

3. The distribution of $\rho_v$ and $\omega$ given $v_{1,T}$ is of the Normal-Inverse Gamma form because $\ln v_t$ in (62) evolves according to an AR(1) process.

4. To obtain draws from the distribution of $v_{1,T}$ given $\epsilon_{a,1,T}$, $\rho_v$, and $\omega$, notice that (62) can be interpreted as nonlinear state-space model, where $\epsilon_{a,t}$ is the "observable" and $v_t$ is the latent state. Smoothing algorithms that generate draws of the sequence of stochastic volatilities have been developed by Jacquier, Polson, and Rossi (1994) and Kim, Shephard, and Chib (1998, ADD) and are discussed in more detail in the Chapters [Time Series, Finance].

## 4.5 Estimation of Nonlinear DSGE Models

dsgenonlinear

DSGE models are inherently nonlinear, as can be seen from the equilibrium conditions (55) associated with our stochastic growth model. Nonetheless, given the magnitude of the business cycle fluctuations of a country like the U.S. or the Euro area, the equilibrium dynamics are quite well approximated by the linear system (59). However, this linear approximation becomes unreliable if economies are hit by large shocks, as is often the case for emerging market economies, or if the goal of the analysis is to study asset pricing implications or consumer welfare. It can be easily shown that for any asset $j$, yielding return $R_{j,t}$, the linearized consumption Euler equation takes the form:

$$\widehat{C}_t = \beta I\!E_t \left[\widehat{C}_{t+1} - (1-\rho_a)\widehat{A}_{t+1} - \widehat{R}_{j,t+1}\right], \tag{63}$$

implying that all assets yield the same expected return. Thus, risk premia disappear in log-linear approximation.

The use of nonlinear model solution techniques complicates the implementation of Bayesian estimation for two reasons. First, it is computationally more demanding to obtain the nonlinear solution. The most common approach in the literature on estimated DSGE model is to use second-order perturbation methods.[9] Second, the evaluation of the likelihood function becomes more costly because both the state transition equation as well as the measurement equation of the state-space model are nonlinear. Thus, (59) and (60) are replaced by (57) and

$$y_t = \Psi(s_t; \theta). \tag{64}$$

Fernandez-Villaverde and Rubio-Ramirez (2007a) and Fernandez-Villaverde and Rubio-Ramirez (2007b) show how a particle filter can be used to evaluate the likeli-

[9] A comparison of linear and nonlinear solution methods for DSGE models can be found in Aruoba, Fernandez-Villaverde, and Rubio-Ramirez (2004, ADD).

hood function associated with a DSGE model. A detailed description of the particle filter is provided in Chapter [Time Series].

For the particle filter to work in the context of the stochastic growth model described above, the researcher has to introduce measurement errors in (64). Suppose that $\{s_{t-1}^{(i)}\}_{i=1}^N$ is a swarm of particles that approximates $p(s_{t-1}|Y^T,\theta)$. Without errors in the measurement equation any proposed particle $\tilde{s}_t^{(i)}$ that does not satisfy the equations

$$y_t = \Psi(\tilde{s}_t^{(i)};\theta), \quad \tilde{s}_t^{(i)} = \Phi(s_{t-1}^{(i)},\epsilon_t^{(i)};\theta) \tag{65}$$

for some $\epsilon_t^{(i)}$ has probability zero. If $\tilde{s}_t^{(i)}$ is sampled from a continuous distribution, the probability that conditional (65) is satisfied is also zero. Thus, in the absence of measurement errors, one has to find *all* real solutions of (65) conditional on $y_t$ and $s_{t-1}^{(i)}$, $i = 1,\ldots,N$, which is a difficult computational task because the nonlinear equation might have multiple solutions. If errors $\eta_t \sim N(0,\Sigma_\eta)$ are added to the measurement equation (64), then (65) turns into

$$y_t = \Psi(\tilde{s}_t^{(i)};\theta) + \eta_t, \quad \tilde{s}_t^{(i)} = \Phi(s_{t-1}^{(i)},\epsilon_t^{(i)};\theta) \tag{66}$$

which can be solved for any $\tilde{s}_t^{(i)}$ by setting $\eta_t = y_t - \Psi(\tilde{s}_t^{(i)};\theta)$. An efficient implementation of the particle filter is one for which a large fraction of the $N$ $\tilde{s}_t^{(i)}$'s are associated with values of $\eta_t$ that are small relative to $\Omega_\eta$.

## 4.6 DSGE Model Evaluation

dsgeevaluation

An important aspect of empirical work with DSGE models is the evaluation of fit. We will distinguish three approaches. First, one could examine to what extent an estimated DSGE model is able to capture salient features of the data. For instance, in the context of the stochastic growth model we could examine whether the model is able to capture the correlation between output and hours worked that we observe in the data. This type of evaluation can be implemented with predictive checks. Second, the researcher might be interested in assessing whether the fit of the stochastic growth model improves if we allow for convex investment adjustment costs. Posterior odds of a model with versus a model without adjustment costs are useful for such an assessment. Finally, a researcher might want to compare the DSGE model to are more flexible reference model such a VAR and we consider three methods of doing so.

### 4.6.1 Predictive Checks

A general discussion of the role of predictive checks in Bayesian analysis can be found in Geweke (2005, ADD). Predictive checks can be implemented based on either the prior or the posterior distribution of the DSGE model parameters $\theta$. Let $Y^{rep}$ be a sample of observations of length $T$ that we could have observed in the past or that we might observe in the future. The predictive distribution for $Y^{rep}$ based on the time $t$ information set $\mathcal{I}_t$ is:

$$p(Y^{rep}|\mathcal{I}_t) = \int p(Y^{rep}|\theta)p(\theta|\mathcal{I}_t)d\theta. \tag{67}$$

We can then use $\mathcal{I}_0$ to denote the prior information and $\mathcal{I}_T$ to denote posterior information. Draws from the predictive distribution can be obtained in two steps. First, generate a parameter draw $\tilde{\theta}$ from $\mathcal{I}_t$. Second, simulate a trajectory of observations $Y^{rep}$ from the DSGE model conditional on $\tilde{\theta}$. The simulated trajectories can be converted into sample statistics of interest, $\mathcal{S}(Y^{rep})$, such as the sample correlation between output and hours worked, to obtain an approximation for predictive distributions of sample moments. Finally, one can compute the observed value of $\mathcal{S}$ based on the actual data and assess how far it lies in the tails of its predictive distribution. If $\mathcal{S}$ is located far in the tails, one concludes that the model has difficulties explaining the observed patterns in the data.

The goal of prior predictive checks is to determine whether the model is able to capture salient features of the data. Canova (1994, ADD) was the first author who used prior predictive checks to assess implications of a stochastic growth model that is solely driven by a technology shock. The advantage of prior predictive checks is that they do not require the implementation of a posterior sampler. Posterior predictive checks can be used to assess the "absolute" fit of an estimated model, similar to a frequentist specification test. Chang, Doh, and Schorfheide (2007) use posterior predictive checks to determine the extent to which a stochastic growth model, similar to the one analyzed in this section, with and without non-stationary labor supply shocks and labor adjustment costs is able to capture the observed persistence of hours worked.

### 4.6.2 Posterior Odds

dsgeodds

The Bayesian framework allows researchers to assign probabilities to various competing models. These probabilities are updated through marginal likelihoods ratios

according to

$$\frac{\pi_{i,T}}{\pi_{j,T}} = \frac{\pi_{i,0}}{\pi_{j,0}} \times \frac{p(Y|\mathcal{M}_i)}{p(Y|\mathcal{M}_j)} \tag{68}$$

Here, $\pi_{i,0}$ ($\pi_{i,0}$) is the prior (posterior) probability of model $\mathcal{M}_i$ and

$$p(Y|\mathcal{M}_i) = \int p(Y|\theta_{(i)}, \mathcal{M}_i) p(\theta_{(i)} d\theta_{(i)} \tag{69}$$

is the marginal likelihood function. The key challenge in posterior odds comparisons is the computation of the marginal likelihood which involves a high-dimensional integral. If posterior draws for the DSGE model parameters are generated with the RWM algorithm, the methods proposed by Geweke (1999, ADD) and Chib and Jeliazkov (2001, ADD) can be used to obtain numerical approximations of the marginal likelihood. Posterior odds-based model comparisons are fairly popular in the DSGE model literature. For instance, Rabanal and Rubio-Ramirez (2005) use posterior odds to assess the importance of price and wage stickiness in the context of a small-scale New Keynesian DSGE model and Smets and Wouters (2007) use odds to determine the importance of a variety of real and nominal frictions in a medium-scale New Keynesian DSGE model. A more detailed discussion of model selection and model averaging based on posterior probabilities will be provided in Section 7.

### 4.6.3 VARs as Reference Models

Vector autoregressions play an important role in the assessment of DSGE models, since they provide a more densely parameterized benchmark. We consider three approaches of using VARs for the assessment of DSGE models.

**Models of Moments:** Geweke (2007) points out that many DSGE models are too stylized to deliver a realistic distribution for the data $Y$, that is useable for likelihood-based inference. Instead these models are designed to capture certain underlying population moments, such as the volatilities of output growth, hours worked, and the correlation between these to variables. Suppose we collect these population moments in the vector $\varphi$, which in turn is a function of the DSGE model parameters $\theta$. Thus, a prior distribution for $\theta$ induces a model-specific distribution for the population characteristics, denoted by $p(\varphi|\mathcal{M}_i)$. At the same time the researcher considers a VAR as reference model, $\mathcal{M}_0$, that is meant to describe the data and at the same

time delivers predictions about $\varphi$. Let $p(\varphi|Y, \mathcal{M}_0)$ denote the posterior distribution of population characteristics as obtained from the VAR. Geweke (2007) shows that

$$\frac{\pi_{1,0} \in p(\varphi|\mathcal{M}_1)p(\varphi|Y, \mathcal{M}_0)d\varphi}{\pi_{2,0} \in p(\varphi|\mathcal{M}_2)p(\varphi|Y, \mathcal{M}_0)d\varphi} \tag{70}$$

can be interpreted as odds ratio of $\mathcal{M}_1$ versus $\mathcal{M}_2$ conditional on the reference model $\mathcal{M}_0$. The numerator in (70) is large, if there is a strong overlap between the predictive densities for $\varphi$ between DSGE model $\mathcal{M}_1$ and VAR $\mathcal{M}_0$. The ratio formalizes the confidence interval overlap criterion proposed by DeJong, ingram, and Whiteman (1996). It has been used by Geweke (2007) to examine asset pricing implications of DSGE models and by Kano and Nason (2009, ADD) to study the business cycle implications of internal consumption habit.

**Loss-Function-Based Evaluation** Schorfheide (2000) proposes a Bayesian framework for a loss function-based evaluation of DSGE models. As in Geweke (2007)'s framework, the research is interested in the relative ability of two DSGE model to capture a certain set of population moments $\varphi$. Unlike in Geweke (2007), the DSGE models are given a chance to explain the data $Y$. Suppose there are two DSGE models $\mathcal{M}_1$ and $\mathcal{M}_2$, and a VAR that serves as a third reference model $\mathcal{M}_0$. The first-step of the analysis consists of computing model-specific posterior predictive distributions $p(\varphi|Y, \mathcal{M}_i)$ and posterior model probabilities $\pi_{i,T}$, $i = 1, 2, 3$. Second, one can form a predictive density for $\varphi$ by averaging across the three models

$$p(\varphi|Y) = \sum_{i=1,2,3} \pi_{i,T} p(\varphi|Y, \mathcal{M}_i). \tag{71}$$

If, say, DSGE model $\mathcal{M}_1$ is well specified and attains a high posterior probability, then the predictive distribution is dominated by $\mathcal{M}_1$. If on the other hand, none of the DSGE models fits well, then the predictive density is dominated by the VARs. Third, one specifies a loss function $L(\hat{\varphi}_{(i)}, \varphi)$ under which DSGE model predictions $\hat{\varphi}_{(i)}$ of $\varphi$ are to be evaluated. Finally one can compare DSGE models $\mathcal{M}_1$ and $\mathcal{M}_2$ the posterior expected loss $\int L(\hat{\varphi}_{(i)}, \varphi)p(\varphi|Y)d\varphi$. This procedure has the feature that if the DSGE models are poorly specified, the evaluation is loss-function dependent, whereas the model ranking becomes effectively loss function independent if one of the DSGE models has a posterior probability that is close to one.

**DSGE-VARs** Building on work by Ingram and Whiteman (1994), Del Negro and Schorfheide (2004) link DSGE models and VARs by constructing families of prior

distributions that are more or less tightly concentrated in the vicinity of the restrictions that a DSGE model implies for the coefficients of a VAR. We will refer to such a model as DSGE-VAR. Recall the VAR from Section 2 that we specified in Equation (1):

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_c + u_t, \quad u_t \sim N(0, \Sigma),$$

which can be written in matrix form as $Y = X\Phi + U$. Assuming that the data have been transformed such that $y_t$ is stationary, let $\mathbb{E}_\theta^D[\cdot]$ be the expectation under DSGE model conditional on parameterization $\theta$ and define the autocovariance matrices

$$\Gamma_{XX}(\theta) = \mathbb{E}_\theta^D[x_t x_t'], \quad \Gamma_{XY}(\theta) = \mathbb{E}_\theta^D[x_t y_t'].$$

A VAR approximation of the DSGE model can be obtained from the following restriction functions that relate the DSGE model parameters to the VAR parameters

$$\Phi^*(\theta) = \Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta), \quad \Sigma^*(\theta) = \Gamma_{YY}(\theta) - \Gamma_{YX}(\theta)\Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta). \tag{72}$$

This approximation is typically not exact because the state-space representation of the linearized DSGE model generates moving average terms. In order to account for potential misspecification of the DSGE model, we now use a prior distribution that, while centered at $\Phi^*(\theta)$ and $\Sigma^*(\theta)$, allows for deviations of $\Phi$ and $\Sigma$ from the restriction functions:

$$\Phi, \Sigma|\theta \sim MNIW\Big(\Phi^*(\theta), [\lambda T \Gamma_{XX}(\theta)]^{-1}, \lambda T \Sigma^*(\theta), \lambda T - k\Big). \tag{73}$$

This prior distribution can be interpreted as a posterior calculated from a sample of $\lambda T$ artificial observations generated from the DSGE model with parameters $\theta$.

So far, we have specified a prior distribution for the reduced form parameters of a VAR conditional on the DSGE model parameters $\theta$. The next step is to turn the reduced form VAR into a structural VAR. According to the DSGE model, the one-step-ahead forecast errors $u_t$ are functions of the structural shocks $\epsilon_t$, which we represent by

$$u_t = \Sigma_{tr}\Omega\epsilon_t. \tag{74}$$

$\Sigma_{tr}$ is the Cholesky decomposition of $\Sigma$ and $\Omega$ is an orthonormal matrix that is not identifiable based on the likelihood function associated with (1). Let $A_0(\theta)$ be the contemporaneous impact of $\epsilon_t$ on $y_t$ according to the DSGE model. Using a

QR factorization, the initial response of $y_t$ to the structural shocks can be can be uniquely decomposed into

$$\left(\frac{\partial y_t}{\partial \epsilon_t'}\right)_{DSGE} = A_0(\theta) = \Sigma_{tr}^*(\theta)\Omega^*(\theta), \tag{75}$$

where $\Sigma_{tr}^*(\theta)$ is lower triangular and $\Omega^*(\theta)$ is orthonormal. The initial impact of $\epsilon_t$ on $y_t$ in the VAR, on the other hand, is given by

$$\left(\frac{\partial y_t}{\partial \epsilon_t'}\right)_{VAR} = \Sigma_{tr}\Omega. \tag{76}$$

To identify the DSGE-VAR, we maintain the triangularization of its covariance matrix $\Sigma$ and replace the rotation $\Omega$ in (76) with the function $\Omega^*(\theta)$ that appears in (75). The rotation matrix is chosen such that in absence of misspecification the DSGE's and the DSGE-VAR's impulse responses to all shocks approximately coincide. To the extent that misspecification is mainly in the dynamics, as opposed to the covariance matrix of innovations, the identification procedure can be interpreted as matching, at least qualitatively, the short-run responses of the VAR with those from the DSGE model. The estimation of the DSGE-VAR can be implemented as follows.

The final step is to specify a prior distribution for the DSGE model parameter $\theta$, which can follow the same elicitation procedure that was used when the DSGE model was estimated directly. Thus, we obtain the following hierarchical model

$$p_\lambda(Y, \Phi, \Sigma, \theta) = p(Y|\Phi, \Sigma)p_\lambda(\Phi, \Sigma, \Omega|\theta)p(\theta) \tag{77}$$

with the understanding that the distribution of $\Omega|\theta$ is a point mass at $\Omega^*(\theta)$. To implement posterior inference it is convenient to factorize the posterior distribution as follows:

$$p_\lambda(\Phi, \Sigma, \Omega, \theta|Y) = p_\lambda(\theta|Y)p_\lambda(\Phi, \Sigma|Y, \theta)p(\Omega|\theta). \tag{78}$$

The distribution of $\Phi, \Sigma|\theta$ is of the Inverse Wishart-Normal form:

$$\Phi, \Sigma|Y, \theta \sim MNIW\Big(\bar{\Phi}(\theta), [\lambda T\Gamma_{XX}(\theta)+X'X]^{-1}, (1+\lambda)T\bar{\Sigma}(\theta), (1+\lambda T)-k\Big). \tag{79}$$

where

$$\begin{aligned}
\bar{\Phi}(\theta) &= \left(\frac{\lambda}{1+\lambda}\Gamma_{XX}(\theta) + \frac{1}{1+\lambda}\frac{X'X}{T}\right)^{-1}\left(\frac{\lambda}{1+\lambda}\Gamma_{XY} + \frac{1}{1+\lambda}\frac{X'Y}{T}\right) \\
\bar{\Sigma}(\theta) &= \frac{1}{(1+\lambda)T}\Big[(\lambda T\Gamma_{YY}(\theta) + Y'Y) - (\lambda T\Gamma_{YX}(\theta) + Y'X) \\
&\quad \times(\lambda T\Gamma_{XX}(\theta) + X'X)^{-1}(\lambda T\Gamma_{XY}(\theta) + X'Y)\Big].
\end{aligned}$$

The marginal posterior density of $\theta$ can be obtained through the marginal likelihood

$$\begin{aligned} p_\lambda(Y|\theta) &= \frac{|\lambda T\Gamma_{XX}(\theta)+X'X|^{-\frac{n}{2}}|(1+\lambda)T\hat{\Sigma}_b(\theta)|^{-\frac{(1+\lambda)T-k}{2}}}{|\lambda T\Gamma_{XX}(\theta)|^{-\frac{n}{2}}|\lambda T\Sigma^*(\theta)|^{-\frac{\lambda T-k}{2}}} \\ &\times\frac{(2\pi)^{-nT/2}2^{\frac{n((1+\lambda)T-k)}{2}}\prod_{i=1}^n\Gamma[((1+\lambda)T-k+1-i)/2]}{2^{\frac{n(\lambda T-k)}{2}}\prod_{i=1}^n\Gamma[(\lambda T-k+1-i)/2]}. \end{aligned} \tag{80}$$

A derivation is provided in Del Negro and Schorfheide (2004). Draws from the marginal posterior of $\theta$ can be generated with the same algorithms that are used for the direct Bayesian estimation of DSGE models, described previously. One can show that in large samples the resulting estimator of $\theta$ can be interpreted as a Bayesian minimum distance estimator that projects the VAR coefficient estimates onto the subspace generated by the restriction functions (72).

Since the empirical performance of the DSGE-VAR procedure crucially depends on the weight placed on the DSGE model restrictions, it is important to consider a data-driven procedure to determine $\lambda$. A natural criterion for the choice of $\lambda$ in a Bayesian framework is the marginal data density

$$p_\lambda(Y)=\int p_\lambda(Y|\theta)p(\theta)d\theta. \tag{81}$$

For computational reasons we restrict the hyperparameter to a finite grid $\Lambda$. If one assigns equal prior probability to each grid point then the normalized $p_\lambda(Y)$'s can be interpreted as posterior probabilities for $\lambda$. Del Negro, Schorfheide, Smets, and Wouters (2007) emphasize that the posterior of $\lambda$ provides a measure of fit for the DSGE model: high posterior probabilities for large values of $\lambda$ indicate that the model is well specified and a lot of weight should be placed on its implied restrictions. Define

$$\hat{\lambda}=\text{argmax}_{\lambda\in\Lambda}\, p_\lambda(Y). \tag{82}$$

If $p_\lambda(Y)$ peaks at an intermediate value of $\lambda$, say between 0.5 and 2, then a comparison between DSGE-VAR($\hat{\lambda}$) and DSGE model impulse responses can potentially yield important insights about the misspecification of the DSGE model. The DSGE-VAR approach was designed to improve forecasting and monetary policy analysis with VARs. The framework has been used as a model evaluation tool in Del Negro, Schorfheide, Smets, and Wouters (2007) and for policy analysis with potentially misspecified DSGE models in Del Negro and Schorfheide (forthcoming).

## 4.7 DSGE Models in Applied Work

dsgemodelapps

DSGE models estimated with Bayesian methods are currently enjoying a moment of popularity in applied macroeconomic research. Such popularity is in large part due to the contribution of Smets and Wouters (2003), who find that the fit of a suitably enriched DSGE model is comparable to that of more heavily parameterized models, such as VARs. This finding had quite important implications, as many applied macroeconomists quickly realized. For twenty years after the publication of "Macroeconomics and Reality" (Sims (1980)) and "Time to Build and Aggregate Fluctuations" (Kydland and Prescott (1982)) there have been two competing approaches for addressing questions like: What shocks drive U.S. business cycles? What are the sources of the Great Moderation? What do impulse responses to monetary policy shocks look like? Structural VARs, which we discussed in Section 2.3, represent one approach. Their advantage is that their ability to fit the data is good. Their downside is that it is not always straightforward to discuss the underlying identification assumptions in terms of an explicit economic theory. Quantitative DSGE models following the Kydland and Prescott (1982) tradition represent the alternative approach. Its advantage is that the identifying assumptions are clear.[c1] Its disadvantage is that these models' fit of the data in a statistical sense is questionable.[10] Smets and Wouters (2003)'s results show that an estimated DSGE model, if cleverly enriched with adjustment frictions and a broad set of structural shocks, can deliver a fit that in a statistical sense is comparable to that of a VAR.[11]

The estimated DSGE literature focuses on a number of questions, and one of the most important ones is precisely "What shocks drive U.S. business cycles?" Even within the applied DSGE model literature there is little agreement on the answer, which depends not only on the model and the set of observables used (not surprisingly) but also on the way the data are constructed. Smets and Wouters (2007) find

[c1] fs: I would say that except for fairly simple models, we have no idea how the identification works. I would cite theoretical coherence as advantage.

[10] Proponents of this approach claim that this is not a problem, and argue for an alternative approach to assess a model's fit.

[11] This overly elegiac description of estimated DSGE models should not hide that fact that major issues lie ahead. For instance, model misspecification may result in some shocks being very persistent. To the extent that the dynamics of the exogenous shocks capture the model's misspecification, optimal monetary policy conducted using the DSGE model may be misleading. Moreover, the assumption that the shocks are uncorrelated may result in "incredible identifying restrictions" to the extent that, again, the shocks mostly capture misspecification.

that wage-markup shocks, that is, shocks to the degree of competition in labor markets that affect the economy because of nominal wage stickiness, play a dominant role.[12] Justiniano, Primiceri, and Tambalotti (2008), using the same model and a similar data set, find that on the contrary shocks to the marginal efficiency of investment are the dominant source of business cycles. Schmidth-Grohe and Uribe (2009, ADD) introduce anticipated technology shocks (so-called "news shocks") and argue that news shocks to technology play a key role. Finally, Christiano, Motto, and Rostagno (2006, ADD) estimate a DSGE model with financial frictions and claim that introducing shocks associated with these frictions diminishes the importance of shocks to the marginal efficiency of investment. Justiniano and Primiceri (2008) are the only ones who address the question using a model with time-varying volatilities, thereby allowing for the possibility that the sources of business cycle vary over time.

Other issues addressed with DSGE models estimated with Bayesian methods are: the importance of nominal rigidities in the U.S. economy, and the best way to introduce them (e.g., Rabanal and Rubio-Ramirez (2005), Laforte (2007), Benati (2008b), Del Negro and Schorfheide (2008)); the importance of informational frictions (e.g., Mankiw and Reis (2007), Reis (2008b)) and of learning (e.g., Milani (2006), Milani (2007) , Milani (2008)); the consequence of labor market frictions (e.g., Gertler, Sala, and Trigari (forthcoming), Krause and Lubik (2007), Sala, Sderstrm, and Trigari (2008), Krause, Lopez-Salido, and Lubik (forthcoming)) and credit market frictions (e.g., Iacoviello and Neri (2008)); the relationship between the yield curve and monetary policy (e.g., Doh (2007)); the impact of monetary policy shocks (Rabanal (2007)) and optimal monetary policy (e.g., Levin, Onatski, Williams, and Williams (2005), Justiniano and Preston (forthcoming), Reis (2008a)). There also exist several papers using open economy DSGE models estimated with Bayesian methods. A non-exhaustive list includes Adolfson, Linde, Laseen, and Villiani (2005), Justiniano and Preston (2006), Lubik and Schorfheide (2007), Adolfson, Laseen, Linde, and Villani (2007), Adolfson, Lasen, Lind, and Villani (2008), Aldofson, Laseen, Linde, and Villani (2008), Cristadoro, Gerali, Neri, and Pisani (2008).

Estimated DSGE models are also appealing to Central Banks, as they are a tool for forecasting, structural interpretation of economic data, and most importantly policy exercises. Smets and Wouters (2005a) and Adolfson, Andersson, Lind, Villani, and Vredin (2007) among others discuss the use of DSGE models at Central Banks, and

[12] Smets and Wouters (2003) is based on Euro Area data. Smets and Wouters (2005b) compare results for the U.S. with those for the Euro Area.

Adolfson, Lind, and Villani (2007) and Edge, Kiley, and Laforte (2009, ADD) assess the forecasting performance of both closed and open economy DSGE models.

# 5 Time-Varying Coefficient Models

tvcmodels

The coefficients in the models presented in Sections 2 to 4 were assumed to be time-invariant, implying that economic relationships are stable. In Figure 6 we plot quarterly U.S. GDP deflator inflation from 1960 to 2006. Suppose we adopt the view that the inflation rate can be decomposed in a target inflation, set by the central bank, and some stochastic fluctuations around this target. The figure offers three views of U.S. monetary history. First, it is conceivable that the target rate was essentially constant between 1960 to 2006, but there were times, for instance, the 1970s when the central bank let the actual inflation deviate substantially from the target. An alternative interpretation is that throughout the 1970s the Fed tried to exploit an apparent unemployment-inflation trade-off and gradually revised its target upwards. In the early 1980s, however, it realized that the long-run Phillips curve is essentially vertical and that the high inflation lead to a significant distortion of the economy. Volcker decided to disinflate, that is reduced the target inflation rate. This time-variation in the target rate could either be captured by a regime switching process that shifts from a 2.5% target to a 7% target and back to 2.5% or through a slowly varying process.

This section describes models that aim at capturing structural changes in the economy. We introduce models in which the coefficients either vary gradually over time according to a multivariate autoregressive process (Section 5.1), or they change abruptly as in Markov-switching or structural break models (Section 5.2). The models will take the form of state-space models and much of the technical apparatus needed for the estimation with Bayesian methods can be found in Chapter [Time Series] or in the monographs by Kim and Nelson (1999b) and Durbin and Koopman (2001). We shall focus placing these methods in context of the empirical macroeconomics literature and on providing an overview of the applications (Section 5.3). There are other important classes of nonlinear time series models such as threshold vector autoregressive models or smooth transition vector autoregressive models in which the parameter change is often linked directly to observables rather than latent state variables. However, due to space constraints we are unable to discuss these models in this chapter.

## 5.1 Models with Autoregressive Coefficients

arcoef

We will now take the models from Sections 2 and 4 and introduce time-varying coefficients that follow an autoregressive law of motion. Most of the subsequent discussion is devoted to VARs (Section 5.1.1). Whenever time-varying coefficients are introduced into a DSGE model, an additional complication arises. For the model to be theoretically coherent, one should assume that the agents in the model are aware of the time-variation, say, in the coefficients of a monetary policy rule and form their expectations and decision rules accordingly. Hence, the presence of time-varying coefficients significantly complicates the solution of the DSGE model's equilibrium law of motion, and requires the estimation of a nonlinear state-space model (Section 5.1.2).

### 5.1.1 Vector Autoregressions

vararcoef

While VARs with time-varying coefficients were estimated with Bayesian methods almost two decades ago (e.g., Sims (1993)), the current popularity of this approach in empirical macroeconomics is largely due to the contribution of Cogley and Sargent (2001), who took advantage of the developments in MCMC methods occurred in the 1990s. They estimate a VAR in which the coefficients follow unit-root autoregressive processes. The motivation for their work, as well as for the competing Markov-switching approach of Sims and Zha (2006) discussed in Section 5.2, arises from the interest in documenting the time-varying nature of business cycles in the US and other countries.

Cogley and Sargent (2001) set out to investigate time-variation in U.S. inflation persistence using a three-variable VAR with inflation, unemployment, and interest rates. One rationale underlying their reduced form specification is provided by models in which the policymaker and/or agents in the private sector gradually learn about the dynamics of the economy and consequently adapt their behavior, e.g., Sargent (1999). As discussed in the introduction to this section, the central bank might adjust its target inflation rate in view of changing beliefs about the effectiveness of monetary policy and the agents might slowly learn about the policy change. To the extent that this adjustment occurs gradually in every period, it can be best captured by reduced form models where the coefficients are allowed to vary in each period.

In addition to the evidence for changes in the conduct of monetary policy and the agents' perception of central bank behavior, studies by McConnell and Perez-Quiros (2000) and Kim and Nelson (1999a), for instance, had suggested the possibility of breaks in the volatility of U.S. GDP, potentially caused by heteroskedastic structural shocks. Cogley and Sargent's earlier work was criticized by Sims (2001), who pointed out that the lack of time-varying volatility in their VAR may well bias the results in favor of finding changes in the dynamics. Cogley and Sargent (2005) address this criticism by adding time-varying volatility to their model. Here we base out description of a VAR with autoregressive coefficients on their work. Thus, we also explain how to introduce heteroskedastic shocks in a VAR.

Consider the reduced-form VAR in Equation (1), which we are reproducing here for convenience:

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_c + u_t.$$

We defined $x_t = [y'_{t-1}, \ldots, y'_{t-p}, 1]'$ and $\Phi = [\Phi_1, \ldots, \Phi_p, \Phi_c]'$. Now let $X_t = I_n \otimes x_t$ and $\phi = vec(\Phi)$. Then we can write the VAR as

$$y_t = X'_t \phi_t + u_t, \tag{83}$$

where we replaced the vector of constant coefficients, $\phi$, by a vector of time-varying coefficients, $\phi_t$. We assume that the parameters evolve according to independent random walks:

$$\phi_t = \phi_{t-1} + \nu_t, \quad \nu_t \sim N(0, Q). \tag{84}$$

The innovations $\nu_t$ uncorrelated with the vector of VAR innovations, $u_t$. The $u_t$ innovations are also normally distributed, but unlike in Section 2, their variance now evolves over time:

$$u_t \sim N(0, R_t), \quad R_t = B^{-1} H_t B^{-1\,\prime}, \tag{85}$$

where $H_t$ is a diagonal matrix with elements $h^2_{i,t}$ following a geometric random walk:

$$\ln h_{i,t} = \ln h_{i,t-1} + \sigma_i \eta_{i,t}, \quad \eta_{i,t} \sim N(0,1). \tag{86}$$

$B$ is a lower triangular matrix with ones on the diagonal.[13] If the prior distributions for $\phi_0$, $Q$, $B$, and $\sigma_i$ are conjugate, then one can use a Gibbs sampler that iterates over the following five conditional distributions to implement posterior inference:

[13] Notice that this form of stochastic volatility was also used in Section 4.4 to make the innovation variances for shocks in DSGE models time varying.

1. Conditional on the matrices $H_{1,T}$, $B$, and $Q$, the posterior for $\phi_t$ can be obtained using the state-space representation given by equations (83) and (84), and applying the algorithm developed by Carter and Kohn (1994), described in Chapter [Time Series].

2. Conditional on the VAR parameters $\phi_t$, the innovations to equation (83) are known. According to (85) $Bu_t$ is normally distributed with variance $H_t$, or:

$$Bu_t = H_t^{\frac{1}{2}}\epsilon_t, \tag{87}$$

   where $\epsilon_t$ is a vector of standard normals. Thus, the problem of sampling from the posterior distribution of $B$ under a conjugate prior is identical to the problem of sampling from the posterior distribution of $A_0$ in the structural VAR specification (18) described in detail in Section 2.3.2.[c1]

3. Conditional on $\phi_t$ and $B$ we can write the $i$'th equation of (87) as $z_{i,t} = B_{i\cdot}u_t \sim N(0, h_{i,t}^2)$, which is identical to (62). Thus, as in Section 4.4 we can use the algorithms of Jacquier, Polson, and Rossi (1994) or Kim, Shephard, and Chib (1998, ADD) to draw the sequence $h_{i,t,T}$.

4. Conditional on the sequence $h_{i,1,T}$ the posterior of $\sigma_i$ in (86) is of the Inverse-Gamma form.

5. Conditional on the sequence $\phi_{1,T}$ the posterior of $Q$ in (84) the Inverted Wishart form.

For the initial vector of VAR coefficients, $\phi_0$, Cogley and Sargent (2001) and Cogley and Sargent (2005) use a prior of the form $\phi_0 \sim N(\underline{\phi}_0, \underline{V}_0)$, where $\underline{\phi}_0$ and $\underline{V}_0$ are obtained by estimating a fixed-coefficient coefficient VAR with flat prior based on a pre-sample. Del Negro (2003) advocates the use of a shrinkage prior with tighter variance than Cogley and Sargent's to partly overcome the problem of over-fitting. The prior for $Q$ controls the magnitude of the period-to-period drift in the VAR coefficients and the priors for $\sigma_i$ determines the magnitude of changes in the volatility of the VAR innovations. These priors have to be chosen very carefully. Imposing the restriction that for each $t$ the parameter vector $\phi_t$ is such that the VAR (83) is stationary introduces an additional complication that we

[c1] fs: I believe our previous description was incorrect because it ignored the Jacobian term.

do not explore here. Koop and Potter (2008) discuss how to impose stationarity restrictions efficiently.

Primiceri (2005) extends the above framework by allowing the parameter of the contemporaneous covariance matrix ($B$) to evolve as well following a random walk. The matrix $B_t$ still has a lower triangular structure. This extension turns the reduced form VAR (83) into a structural model – at least to the extent that one is comfortable with the lower triangular identification scheme – in that the model now allows for the contemporaneous relationships among the variables to be time-varying. Hence, for instance, one can estimate a time-varying monetary policy reaction function, as in Primiceri's application.[14] Del Negro (2003) suggests an alternative approach where time-variation is directly imposed on the parameters of the structural model – that is, the parameters of the VAR in equation (18). Finally, no cointegration restrictions are imposed in the VAR (83). A Bayesian analysis of a time-varying coefficient cointegration model can be found in Koop, Leon-Gonzalez, and Strachan (2008).

### 5.1.2 DSGE Models with Drifting Coefficients

dsgearcoef

Recall the stochastic growth model introduced in Section 4.1. Suppose that we change the objective function of the household to

$$\mathbb{E}_t\left[\sum_{s=0}^{\infty}\beta^{t+s}\left(\ln C_{t+s}-\frac{(H_{t+s}/B)^{1+1/\nu}}{1+1/\nu}\right)\right]. \tag{88}$$

Under this specification $B$ is a parameter that affects the disutility associated with working. Thus, we can interpret our original objective function (88) as a generalization of (88), in which we have replaced the constant parameter $B$ by a time-varying parameter $B_t$. But in our discussion of the DSGE model in Section 4.1 we never mentioned time-varying parameters, we simply referred to $B_t$ as a labor supply or preference shock. Thus, a time-varying parameter is essentially just another shock.

If the DSGE model is log-linearized, as in (58), then all structural shocks (or time-varying coefficients) appear additively in the equilibrium conditions. For instance, the preference shock appears in the labor supply function

$$\widehat{H}_t=\nu\widehat{W}_t-\widehat{C}_t+(1+\nu)\widehat{B}_t. \tag{89}$$

[14] Primiceri (2005) estimates a tri-variate VAR with interest rates, unemployment, and inflation on post-war U.S. data.

Now imagine replacing the constant Frisch elasticity $\nu$ in (45) and (88) by a time-varying process $\nu_t$. After log-linearizing the equilibrium conditions the time-varying elasticity would show up as an additional additive shock in (89) and hence, be indistinguishable in its dynamic effects from $B_t$. Thus, for additional shocks or time-varying parameters to be identifiable, it is important that the log-linear approximation be replaced by a nonlinear solution technique. Fernandez-Villaverde and Rubio-Ramirez (2007b) take a version of the constant-coefficient DSGE model estimated by Smets and Wouters (2003) and allow for time variation in the coefficients that determine the interest-rate policy of the central bank and the degree of price and wage stickiness in the economy. To capture, for instance, the different effects of a typical monetary policy shock and a shock that changes central bank's reaction to deviations from the inflation target the authors use a second-order perturbation method to solve the model and the particle filter to approximate its likelihood functions. Thus, the topic of DSGE model with time-varying autoregressive parameters has essentially been covered in Section 4.5.

## 5.2 Models with Markov-Switching Coefficients

msmodels

Markov-Switching (MS) models represent an alternative to drifting autoregressive coefficients in time series models with time-varying parameters. MS models are able to capture sudden changes in time series dynamics. Recall the two different representations of a time-varying target inflation rate in Figure 6. The piecewise constant path of the target can be generated by a MS model, but not by the drifting-coefficients model. We will begin with a discussion of MS coefficients in the context a VAR (Section 5.2.1) and then consider the estimation of DSGE models with MS coefficients (Section 5.2.2).

### 5.2.1 Markov-Switching VARs

msvar

Markov-Switching models have been popularized in economics by the work of Hamilton (1989), who used them to model GDP growth rates in recession and expansion states. We will begin by adding regime-switching to the coefficients of the reduced form VAR specified in (1), which we write in terms of a multivariate linear regression model as

$$y_t' = x_t'\Phi(s_t) + u_t', \quad u_t \sim N(0, \Sigma(s_t)) \tag{90}$$

using the same definitions of $\Phi$ and $x_t$ as in Section 2.1. Unlike before, the coefficient vector $\Phi$ is now a function of $s_t$. Here $s_t$ is a discrete $M$-state Markov process with time-invariant transition probabilities

$$\pi_{lm} = P[\{s_t = l\}|\{s_{t-1} = m\}], \quad l, m \in \{1, \ldots, M\}.$$

For simplicity, suppose that $M = 2$ and all elements of $\Phi(s_t)$ and $\Sigma(s_t)$ switch simultaneously, without any restrictions. Posterior inference in this simplified MS VAR model can be implemented with a Gibbs sampler that iterates over the following conditional distributions:

1. Posterior distribution of $\Phi(s)$ and $\Sigma(s)$, $s = 1, 2$, conditional on $s_{1,T}$. Let $\mathcal{T}_1$ be a set that contains the time periods when $s_t = 1$ and $\mathcal{T}_2$ be the set that contains the times when $s_t = 2$. Under a conjugate prior, the posterior of $\Phi(s)$ and $\Sigma(s)$ is MNIW, obtained from the regression $y_t' = x_t'\Phi(s) + u_t$, $u_t \sim N(0, \Sigma(s))$, $t \in \mathcal{T}_s$.

2. Draws from the posterior of $s_{1,T}$ conditional on $\Phi(s)$ and $\Sigma(s)$, $s = 1, 2$, can be obtained with a variant of the Carter and Kohn (1994) approach, described in detail in Chapter [Time Series].

If one imposes that $\pi_{22} = 1$ and $\pi_{12} = 0$ then model (90) becomes a change-point model in which state 2 is the final state. Alternatively, such a model can be viewed as a structural break model in which at most one break can occur, but the time of the break is unknown. Kim and Nelson (1999a) use a change-point model to study whether there been a structural break in post-war GDP growth towards stabilization. By increasing the number of states and imposing the appropriate restrictions on the transition probabilities, one can generalize the change-point model to allow for several breaks. Chopin and Pelgrin (2004) considers a setup that allows the joint estimation of the parameters and the number of regimes that have actually occurred in the sample period. Koop and Potter (2007) and Koop and Potter (FC) explore posterior inference in change-point models under various types of prior distributions.

In a multivariate setting the unrestricted MS VAR in (90) with coefficient matrices that are *a priori* independent across states is typically not empirically not viable, because it involves too many coefficients that need to be estimated on too few observations. Introducing dependence of coefficients across Markov states complicates Step 1. of our posterior simulator.

Sims and Zha (2006) extend the structural VAR given in (18) to a Markov-Switching setting:

$$y_t' A_0(s_t) = x_t' A(s_t) + \epsilon_t', \tag{91}$$

where $\epsilon_t$ is a vector of orthogonal structural shocks with unit variance. To cope with the dimensionality problem the authors re-parameterize the $k \times n$ matrix $A(s_t)$ as $D(s_t) + \bar{S}A_0(s_t)$, where $\bar{S}$ is a $k \times n$ with the $n \times n$ identity matrix in the first $n$ rows, and zeros elsewhere. Thus,

$$y_t' A_0(s_t) = x_t'(D(s_t) + \bar{S}A_0(s_t)) + \epsilon_t'. \tag{92}$$

Notice that if $D(s_t) = 0$, then the reduced-form VAR coefficients are given by $\Phi = \bar{S}$ and the elements of $y_t$ follow random walk processes, as implied by the Minnesota prior. If the prior for $D(s_t)$ is centered at zero, one can impose restriction, for instance constrain some elements of $D(s_t)$ to be the same across states, and at the same time maintain the feature that conditional on the state $s_t$ the prior for the reduced form VAR is centered at a random walk representation.[c1] Sims and Zha (2006) use their setup to estimate specifications in which (i) only the coefficients of the monetary policy rule change across Markov states, (ii) only the coefficients of the private-sector equations switch, and (iii) only coefficients that implicitly control innovation variances (heteroskedasticity) change. The Gibbs sampler for the parameters of (92) is derived by extending the approach in Waggoner and Zha (2003), as discussed in detail by Sims, Waggoner, and Zha (2006).

Paap and Van Dijk (2003) consider a restricted MS VAR in which the Markov process affects the trend of a vector time series $y_t$ but leaves the fluctuations around this trend unchanged. Let

$$y_t = y_t^* + \Gamma_0(s_t) + \widetilde{y}_t, \quad \widetilde{y}_t = \Phi_1 \widetilde{y}_{t-1} + \ldots + \Phi_p \widetilde{y}_{t-p} + u_t, \tag{93}$$

where

$$y_t^* = y_{t-1}^* + \Gamma_1(s_t).$$

The model captures growth rate differentials between recessions and expansions and is used by the authors to capture the joint dynamics of U.S. aggregate output and consumption.

### 5.2.2 DSGE Models with Markov-Switching Coefficients

msdsge

There is a growing number of papers that incorporates Markov-switching effects in DSGE models. Consider the nonlinear equilibrium conditions of our stochastic growth model in (55). The most rigorous and general treatment of Markov-Switching coefficients would involve replacing the vector $\theta$ by a function of the latent state $s_t$, $\theta(s_t)$, and solving the nonlinear model while accounting for the time variation in $\theta$. Since the implementation of the solution and the subsequent computation of the likelihood function is very challenging the literature has focused on various short-cuts, which introduce Markov switching in the coefficients of the linearized model given by (58).

Following Sims (2002, ADD) we write the linearized equilibrium conditions in the following canonical form

$$\Gamma_0 x_t = C + \Gamma_1 x_{t_1} + \Psi \epsilon_t + \Pi \eta_t. \tag{94}$$

Here the vector $\eta_t$ is comprised of one-step ahead rational expectations forecast errors. In the context of our stochastic growth model it would for instance contain $\eta_t^C = \widehat{C}_t - I\!E_{t-1}[\widehat{C}_t]$, which enable us to re-cast (58) in the canonical form. In most applications one can define the vector $x_t$ such that the observables $y_t$ can, as in Section 4.2 be simply expressed as a linear function of $x_t$, that is,

$$y_t = \Psi_0 + \Psi_1 t + \Psi_2 x_t. \tag{95}$$

We could now introduce Markov-switching directly into the system matrices $\Gamma_0$, $\Gamma_1$, $\Psi$, and $\Pi$ – and potentially into the coefficient matrices of the measurement equations as well.

Schorfheide (2005) considers a special case of this Markov-switching linear rational expectations framework, because in his analysis the switching only affects the target inflation rate of the central bank. The resulting system can be written as

$$\Gamma_0 x_t = \Gamma_1 x_{t_1} + (\Psi \epsilon_t + C(s_t)) + \Pi \eta_t$$

and is solvable with the algorithm provided in Sims (2002, ADD). However, there is a large debate in the literature about whether the central bank's reaction to inflation and output deviations from target has changed around 1980. A candidate

[c1] fs: The exposition is still a bit cryptic

explanation for the reduction of macroeconomic volatility in the 1980s is a more forceful reaction of central banks to inflation deviations. This can be captured by a monetary policy rule in which the slope coefficients (rather than the intercept) are subject to regime-switches. Thus, subsequent work by Davig and Leeper (2007, ADD), Farmer, Waggoner, and Zha (2007, ADD), and Bianchi (2008, ADD) is more ambitious in that it allows for switches in the other system matrices as well:

$$\Gamma_0(s_t)x_t = \Gamma_1(s_t)x_{t_1} + (\Psi(s_t)\epsilon_t + C(s_t)) + \Pi(s_t)\eta_t.$$

Characterizing the full set of solutions for this general MS linear rational expectations model and conditions under which there exists a unique stable solution is the subject of ongoing research.

In the simplified model of Schorfheide (2005) the solution to the rational expectations model takes the form

$$x_t = \Phi_1 x_{t-1} + \Phi_\epsilon(s_t) \cdot (\mu(s_t) + \epsilon_t) + \Phi_0(s_t) \tag{96}$$

Equations (95) and (96) define a (linear) Markov-switching state-space model, with the understanding that the system matrices are functions of a time-invariant parameter vector $\theta$. Posterior inference can be implemented in two different ways. Following a filtering approach that simultaneously integrates over $x_t$ and $s_t$, discussed in Kim and Nelson (1999, ADD), Schorfheide (2005) constructs an approximate likelihood that only depends on $\theta$. This likelihood function is then used in the RWM algorithm described in Section 4.3. Alternatively, one could construct a Gibbs sampler that iterates over the blocks $x_{1,T}$, $s_{1,T}$, and $\theta$. Drawing from the conditional distribution of $\theta$ however, requires a Metropolis step due to the nonlinear relationship between the deep structural parameters and the system matrices. One would probably need to tailor this Metropolis step as a function of the latent states at each step of the Gibbs sampler.

## 5.3 Applications of Bayesian Time-Varying Coefficients Models: The 'Great Moderation' and 'Inflation Persistence' Debates

tvpapps

The work by Cogley and Sargent (2001) started a robust debate on whether the dynamics of U.S. inflation, and in particular its persistence, had changed over the last quarter of the 20th century and, to the extent that it had, whether monetary policy played a major role in affecting inflation's dynamics. Naturally, this debate

evolved in parallel to that on the magnitude and causes of the 'great moderation', that is, the decline in the volatility of business cycles during the same period, started by the contributions of McConnell and Perez-Quiros (2000) and Kim and Nelson (1999a). Whatever the causes of the changes in the dynamics of output – whether shocks, policies, or other structural changes – it is likely that these same causes affected the dynamics of inflation. Some of the paper discussed in this section discuss both, while others focus on only one of the two phenomena.

Cogley and Sargent (2001) compute the spectrum of inflation obtained from their time-varying coefficient VAR and use it as evidence that both inflation volatility and persistence have changed dramatically from the 70s to the more recent period. Cogley and Sargent (2005) finds that this result is robust to considering time-variation in the volatility of shocks, and argue that changes in the monetary policy rule are partly responsible for it.[15] Using a structural VAR, Primiceri (2005) finds that monetary policy indeed changed since the 70s, but that the impact of these changes on the rest of the economy was fairly negligible. He claims that variation in the volatility of the shocks is the main cause for the lower volatility of both inflation and business cycles in the post-Volcker period.[16] Sims and Zha (2006) arrives to similar conclusions using a structural regime-switching VAR. Their findings differ from Primiceri's however, to the extent that they find no support to the hypothesis that the parameters of the monetary policy equation have changed. Similarly, using a AR time-varying coefficients VAR identified with sign restrictions Canova and Gambetti (forthcoming) find little evidence that monetary policy has become more aggressive in responding to inflation in the recent period.

Benati and Surico (forthcoming) question the ability of structural VAR methods to correctly identify the sources of the great moderation. They estimate a DSGE model with Bayesian methods, and they find that changing the reaction to inflation in the policy rule can deliver shifts in the volatility of inflation and the output gap similar

[15]Pivetta and Reis (2007) challenge the findings of Cogley and Sargent (2001) and Cogley and Sargent (2005) on the variation of inflation persistence on the ground that variations in inflation persistence are not statistically significant. Cogley, Primiceri, and Sarget (2008) address this criticism by: i) focusing on a different measure of inflation persistence, which the time-varying coefficient model estimates with more precision than that in Cogley and Sargent (2001), and ii) distinguishing between the persistence of inflation itself and that of the 'inflation gap', the difference between inflation and the time-varying target of the central bank. They find that changes in persistence of the inflation gap are statistically significant.

[16]Benati (2008a) reaches similar conclusions for the UK.

to those observed in the data. They claim that the VAR representation implied by such a model under the two regimes would incorrectly identify the sources of the great moderation as a change in the variance of the shocks, and that the change in the VAR impulse response functions to a policy shock would be small.[17]

We conclude by mentioning a few other applications of VARs with time variation in the coefficients. These include forecasting, as in the seminal work by Sims (1993) and the exercise conducted by Cogley, Morozov, and Sargent (2005) using U.K. data, and estimation of a time-varying Phillips Curve, as in Cogley and Sbordone (forthcoming).

[17]Benati and Surico (2008) draw a connection between inflation predictability and the reaction to inflation in the policy rule.

# 6 Models for Data-Rich Environments

panel

This section addresses the estimation of models designed to deal with data sets that have a large cross-sectional and time-series dimension. Consider the VAR(p) from Section 2:

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_c + u_t, \quad t = 1, \ldots, T \tag{97}$$

where $y_t$ is a $n \times 1$ vector. Without mentioning it explicitly, our previous analysis was tailored to situations in which the time series dimension $T$ of the data set is much larger than the cross-sectional dimension $n$. This section focuses on situations in which $T$ and $n$ are of similar size, such that the ratio of observations to parameters for each equation of the vector autoregressive system is too small to enable a meaningful estimation of the coefficients without further restrictions. For instance, in the United States and several other countries statistical agencies release at the monthly or quarterly frequency a number of indicators that are potentially informative on the state of the business cycle. These include well-known measures such as GDP and its components, industrial production or employment, but also indices of consumer confidence, capacity utilization, new orders of non-defense capital goods, *et cetera*. Examples of this data set are those in Stock and Watson (1999) and Stock and Watson (2002). If one wants to use all available information for forecasting, the size of the cross section renders a VAR with typical lag-length (four and thirteen for quarterly and monthly data, respectively) heavily over-parameterized. Similarly, if we want to study international business cycle synchronization the sheer number of countries for which measures of GDP, consumption, and investment are available (say, OECD countries) implies that each VAR equation would have a very large number of regressors.

We will consider the following approaches to cope with the curse of dimensionality. First, one could impose "hard" restrictions, such as setting many of the coefficients in the VAR to zero, or alternatively re-parameterize the model so to reduce the number of free parameters (Section 6.1).[18] Second, one could use very informative, yet non-dogmatic prior distributions for the many coefficients in the VAR (Section 6.2). Finally one could express $y_t$ as a function of a lower dimensional vector of "factors"

[18] For instance, Stock and Watson (2005), who study international business cycles using output data for the G7 countries, impose the restriction that in the equation describing output evolution for a given country enter only the trade-weighted averages of the other countries GDP growth rates.

which by itself follows a vector autoregressive process (Section 6.3). We will in turn explore all three possibilities.

## 6.1 Restricted High-Dimensional VARs

varrestricted

Canova, Ciccarelli, and Ortega (2007) and Canova and Ciccarelli (forthcoming) introduce large-scale VARs where priors reduce the dimensionality of the parameter space.[19] Specifically, let $y_t$, $x_t$, and $\Phi$ be defined as in section 2 (recall that $x_t$ includes the constant and $p$ lags of $y_t$) and let $\phi$ denote $vec(\Phi)$. The VAR can then be represented as:

$$y_t = \left(I_n \otimes x_t'\right)\phi + u_t, \tag{98}$$

where $I_n$ is an $n \times n$ identity matrix, and $u_t$ is a vector of reduced form residuals with variance $\Sigma$. The key idea in Canova and Ciccarelli (forthcoming) consists in reducing the number of parameters that need to be estimated by making $\phi$ time-varying and assuming the following relationship:

$$\phi_t = \Xi\theta + \nu_t, \tag{99}$$

where $\theta$ is a vector of size $\kappa << nk$, $\nu_t$ is a $nk \times 1$ vector of i.i.d. variables (independent from $u_t$ for all leads and lags) with variance $\mathcal{V}$, and the $nk \times \kappa$ matrix incorporates pre-specified weights (or loadings). Equation (99) states that the VAR coefficients $\phi_t$ are known functions of a lower dimensional set of parameters $\theta$ via a set of restrictions $\Xi\theta$, plus a number of deviations $\nu_t$ from such restrictions. An example of these restrictions would be to impose that a coefficient in equation $i$ $(i = 1, .., n)$ associated with variable $j$ $(j = 1, .., n)$ that enters lagged $l$ times $(l = 1, .., p)$ is the sum of an equation-specific parameter, a variable-specific parameter, and a lag-specific parameter. Here $\theta$ would be comprised of the set of all $n + n + p$ equation/variable/lag-specific parameters, and $\Xi$ would be an indicator matrix of zeros and ones that picks the subset of $\theta$ associated with each element of $\phi$.

The fact that the deviations from the restrictions $\nu_t$ are i.i.d. both over time (and over the elements of $\phi$) makes estimation very convenient. In fact, the random

[19] Canova (2005) and Canova, Ciccarelli, and Ortega (2007) use the approach to study the transmission of US shocks to Latin America and convergence in business cycles among G7 countries, respectively.

deviations $\nu_t$ can be merged with the VAR innovations $u_t$, resulting in a model that is straightforward to estimate. Inserting (99) into (98) we obtain the SUR system:

$$y_t = \mathcal{X}_t\theta + \zeta_t. \tag{100}$$

The matrix of regressors $\mathcal{X}_t = (I_n \otimes x_t')\,\Xi$ essentially contains weighted averages of the regressors, where the weights are given by the columns of $\Xi$, and where the variance of $\zeta_t = (I_n \otimes x_t')\,\nu_t + u_t$ is given by[20]

$$\left(I_n \otimes x_t'\right)\mathcal{V}\left(I_n \otimes x_t'\right)' + \Sigma.$$

The appeal of Canova and Ciccarelli's framework is that inference about $\theta$ can be conducted using the tools discussed in section 2.

Canova and Ciccarelli (forthcoming) actually generalize expression (99) by assuming that the vector $\theta_t$ is time-varying and follows a simple autoregressive law of motion. The vector $\theta_t$ can therefore be interpreted as a vector of factors. The structure of the estimation problem is therefore very similar to that of the factor models described later in section 6.3, and estimation can be conducted using similar methods (see Canova and Ciccarelli (forthcoming) for details).

### 6.2 Shrinkage Estimation of High-Dimensional VARs

varshrink

If one solution to the curse of dimensionality is to reparameterize the model so that we have fewer free parameters, another is to have a (possibly strong) a-priori belief that the parameters in a regression are zero. Such approach is known as ridge regressions (or Stein estimator in a classical setup) and amounts to estimating the VAR using a prior of the the type $N(0, \sigma^2 I)$. What is the outcome of ridge regressions when both the cross-sectional and the time series dimensions are large? De Mol, Giannone, and Reichlin (2008) show that there is a tight connection between ridge regressions and factor models. De Mol, Giannone, and Reichlin (2008) compare different solutions to the curse of dimensionality in univariate forecasting problems, e.g. in situations where we want to forecast a specific variable using a potentially large number of regressors. A Classical approach to the curse of dimensionality is to compute the covariance matrix of the regressors, obtain its principal components, and use those associated with the largest eigenvalues as regressors.[21] De Mol, Gi-

[20] Note that if $\mathcal{V}$ and $\Sigma$ have a kronecker structure (as it may be the case for country VARs, for instance) the computation of the variance simplifies considerably.

[21] The Bayesian analog to this solution amounts to extracting a few factors via the methods described later in this section, and then forecast using these factors.

annone, and Reichlin (2008) consider ridge regressions as an alternative, and show that the difference between the two solutions amounts to the following: While under principal components forecasts we use only the first few principal components, under ridge regressions we use all principal components multiplied by a weight that declines with the size of the associated eigenvalues. De Mol, Giannone, and Reichlin (2008) show how to set the shrinkage parameter $\sigma^2$ to obtain consistency of the ridge regression forecasts as both the time-series and cross-sectional size increase to infinity. They also show that with a suitably chosen shrinkage parameter the empirical forecast performance for ridge regression forecasts with a large number of regressors is similar to that obtained with principal components. Banbura, Giannone, and Reichlin (2008) use the theoretical insights from De Mol, Giannone, and Reichlin (2008) to choose the shrinkage parameters in estimating VARs with Minnesota priors in large cross-sections. They argue that these large scale VARs outperform smaller scale VARs in terms if forecasting accuracy.

## 6.3 Dynamic Factor Models

dfm6

Factor models also represent an approach for dealing with the curse of dimensionality. Factor models describe the dynamic behavior of a possibly large cross section of observations as the sum of a few common components, which capture comovements, and of series-specific components, which describe the idiosyncratic dynamics of each individual series. Factor models have been part of the econometricians' toolbox for a long time (see for instance the "unobservable index" models by Sargent and Sims (1977) and Geweke (1977)), but the contribution of Stock and Watson (1989) generated renewed interest in this class of models. Stock and Watson (1989) use a factor model to exploit information from a large cross-section of macroeconomic time-series for forecasting. While Stock and Watson (1989) use maximum likelihood, Bayesian estimation of this class of models follows the work by Otrok and Whiteman (1998). The remainder of this subsection is organized as follows. The baseline version of the dynamic factor model is introduced in Section 6.3.1 and posterior inference is described in Section 6.3.2. Some applications are discussed in Section 6.3.3. Finally, we consider various extensions. We introduce time-varying parameters (Section 6.3.4), consider factor-augmented VARs (Section 6.3.5), and combine DSGE models and factor models (Section 6.3.6).

### 6.3.1 Baseline Specification

dfm6baseline

A dynamic factor model decomposes the dynamics of observables $y_{i,t}$, $i = 1, \ldots, n$, $t = 1, \ldots, T$ into the sum of two unobservable components, one that affects all $y_i$s, namely the factors $f_t$ (a $\kappa \times 1$ vector, if we have $\kappa$ factors), and one that is idiosyncratic, e.g. specific to each $i$:

$$y_{i,t} = a_i + \lambda_i f_t + \xi_{i,t}, \tag{101}$$

where $a_i$ is the constant and $\lambda_i$ is the $1 \times \kappa$ vector of loadings of series $i$ to the common factors. The factors follow a vector autoregressive processes of order $q$:

$$f_t = \Phi_{0,1} f_{t-1} + .. + \Phi_{0,q} f_{t-q} + u_{0,t}, \text{ and} \tag{102}$$

where the $\Phi$ matrices are of size $\kappa \times \kappa$ (in the literature these matrices are often assumed to be diagonal). The idiosyncratic components follow autoregressive processes of order $p_i$:

$$\xi_{i,t} = \phi_{i,1} \xi_{i,t-1} + .. + \phi_{i,p_i} \xi_{i,t-p_i} + \sigma_i u_{i,t}, \tag{103}$$

where $\sigma_i$ is the standard deviation of the idiosyncratic component, $u_{i,t} \sim \mathcal{N}(0, I_\kappa)$, and $u_{i,t} \sim \mathcal{N}(0,1)$ for $i = 1, .., n$ are the innovations to the law of motions (102) and (103), respectively.[22] These innovations are i.i.d. over time and across $i$ (including $i = 0$, e.g. the vector $u_{0,t}$ is orthogonal to all of the $u_{i,t}$ innovations). The latter is the key identifying assumption in the model, as it postulates that all comovements in the data arises from the factors.[23] [24] Another important identifying assumption involves the fact that the factors are not rotation-invariant: we can premultiply the factors in (101) and (102) by a $\kappa \times \kappa$ orthonormal matrix $H$ (such that $H'H = I_\kappa$), and post-multiply the vector $\lambda_i$ and the $\Phi_{0,j}$ matrices by $H'$, and obtain

[22] Note that the matrix $\Sigma_0$ is set equal to the $\kappa \times \kappa$ identity matrix $I_\kappa$, which is a standard normalization assumption given that the scale of the loadings $\lambda_i$s and $\sigma_0$ cannot be separately identified.

[23] Stock and Watson (1989) include lags of the factors in (101), which we omit for simplicity.

[24] Following the work of Chamberlain and Rothschild (1983) the literature has also considered "approximate" factor models, that is, models where there is limited cross-sectionally correlation among the idiosyncratic components. Doz, Giannone, and Reichlin (2008) show that even in this situation maximum likelihood estimation of a model that ignores the weak correlation among the idiosyncratic components delivers consistent estimates of the factors. Their result is important from a Bayesian perspective, as it shows that Bayesian estimates of the factors can be consistent even when the factor structure is not exact.

an observationally equivalent model. This problem is often addressed by imposing constraints on the loading vector $\lambda_i$: for instance, if we had two factors we would impose that for at least one $i$ we have that the second element of $\lambda_i$ equals zero. Another problem is that the factors are sign-indeterminate: if we multiply factors and loadings by $-1$ we obtain an observationally equivalent model. Hamilton and Waggoner (2007) provide a solution to this normalization issue. The problem can also be sidestepped of the prior on the loadings $\lambda_i$ is not centered at zero. While the likelihood would still be sign-invariant, the posterior would no longer be so.[25]

### 6.3.2 Priors and Posteriors

dfm6baselinepp

It is convenient to use conjugate priors for the parameters. Specifically, the priors on both the constant terms $a_i$ and the loadings $\lambda_i$ are normal, namely $N(\underline{a}_i, \underline{A}_i^{-1})$ and $N(\underline{\beta}_i, \underline{B}_i^{-1})$, respectively. The autoregressive coefficients for the factors and the idiosyncratic shocks have a truncated normal prior. Define $\phi_0$ as $\phi_0 = (vec(\Phi_{0,1})', .., vec(\Phi_{0,q})')'$. The prior for $\phi_0$ is $N(\underline{\phi}_0, \underline{V}_0^{-1})I_{S\phi_0}$, where $I_{S\phi_0}$ is an indicator function that places zero prior mass on the region of the parameter space where we have non stationarity. Likewise, the prior for $\phi_i = (\phi_{i,1}, .., \phi_{i,p_i})'$ is $N(\underline{\phi}_i, \underline{V}_i^{-1})I_{S\phi_i}$, where $I_{S\phi_i}$ is similarly defined. The prior for the idiosyncratic volatility $\sigma_i$ is also given by an inverse gamma distribution: $IG(\nu_i, s_i^2)$.

The intuition behind the Gibbs sampler is straightforward: (i) for given factors, equations (101) are simple regression models with AR($p_i$) errors, each independent from one another (since the $\xi_{i,t}$ are independent across $i$); (ii) for given parameters, the factors can be drawn using expressions (101) and (102), as the measurement and transition equations, respectively, in a state-space representation. We now further elaborate on the details of the Gibbs sampler.

(i) The key insight from Otrok and Whiteman (1998) is that, conditional on the factors, drawing from the conditional posterior distribution of $\{a_i,\ \phi_i, \sigma_i^2\}$ amounts to straightforward application of the procedure in Chib and Greenberg (1994) since equation (101) is simply a regression where the errors follow an AR($p_i$) process given

[25] This normalization issue implies that using a time series of zeros as starting values for the factors may delay convergence. Initializing the factors at some reasonable value different than zero can help. For instance, the cross-sectional average of the series can be a reasonable starting point for a common factor.

by equation (103) (see Otrok and Whiteman (1998) for a detailed description). [26] Moreover, since the errors in equation (101) are independent across $i$, the procedure can be implemented one $i$ at the time, which implies that computational cost is linear in the size of the cross-section. Finally, conditional on the factors, the VAR coefficients in (102) can be drawn using the methodology discussed in section 2.

(ii) In the second block of the Gibbs sampler we draw the factors conditional on all other parameters. Two approaches to drawing the factors exist in the Bayesian factor model literature. One is that followed Otrok and Whiteman (1998) approach, who explicitly write the likelihood of the observations $y_{i,t}$, the prior for the factors (given by equation 102), and then combining the two to derive the posterior distribution of $\{f_t\}_{t=1}^T$. The alternative is to use the state-space representation given by equations (101) and (102), and apply the algorithm developed by Carter and Kohn (1994)to draw from the posterior distribution of the factors. The first approach involves inverting size-$T$ matrices, and hence becomes computationally expensive for large $T$. In the second approach it becomes essential to eliminate the idiosyncratic terms from the state vector if one wants to estimate data sets where $n$ is large (for otherwise the dimension of the state vector would increase with $n$, since the idiosyncratic terms in (103) are not i.i.d.). This is accomplished by quasi-differencing expression (101) as in Kim and Nelson (1999c) and Quah and Sargent (1992).

We describe the second approach, assuming notational simplicity that we have only one factor. We first write the measurement equation (101) and the law of motion of the idiosyncratic components in stacked form:

$$\tilde{y}_t = \tilde{a} + \tilde{b} f_t + \tilde{\xi}_t \text{ for } t = 1, .., T \tag{104}$$

$$\tilde{\xi}_t = \tilde{\Phi}_1 \tilde{\xi}_{t-1} + .. + \Phi_p \tilde{\xi}_{t-p} + \tilde{u}_t, \tag{105}$$

where $\tilde{y}_t = (y_{1,t}, \ldots, y_{n,t})'$, $\tilde{a} = (a_1, \ldots, a_n)'$, $\tilde{b} = (\lambda_1, \ldots, \lambda_n)'$, $\tilde{\xi}_t = (\xi_{1,t}, \ldots, \xi_{n,t})'$, $p = max_{i=1,..,n}(p_i)$, $\tilde{u}_t = (\sigma_1 u_{1,t}, \ldots, \sigma_n u_{n,t})'$ and the $\tilde{\Phi}_j$s are diagonal $n \times n$ matrices with elements $\phi_{1,j}, \ldots, \phi_{n,j}$ on the diagonal. Next we express the laws of motion of the factor (equation 102) in companion form:

$$\tilde{f}_t = \tilde{\Phi}_0 \tilde{f}_{t-1} + \tilde{u}_{0,t}, \tag{106}$$

[26] Importantly, Chib and Greenberg (1994) include the first $p_i$ observation in computing the likelihood.

where $\tilde{f}_t = (f_t, .., f_{t-q+1})'$ and $\tilde{u}_{0,t} = (u_{0,t}, 0, .., 0)'$ are $q \times 1$ vector and $\tilde{\Phi}_0$ is the companion matrix:

$$\tilde{\Phi}_0 = \begin{bmatrix} \Phi_{0,1}.. & \Phi_{0,q} \\ I_{q-1} & 0 \end{bmatrix}. \tag{107}$$

We consider the case $q-1=p$ (again for ease of notation) and pre-multiply expression (104) by the matrix $I_n - \sum_{j=1}^{p} \tilde{\Phi}_j L^j$ and obtain the system:

$$\tilde{y}_t^* = \tilde{a}^* + \Lambda^* \tilde{f}_t + \tilde{u}_t \text{ for } t = p+1, .., T \tag{108}$$

where $\tilde{y}_t^* = (I_n - \sum_{j=1}^{p} \tilde{\Phi}_j L^j)\tilde{y}_t$, $\tilde{a}^* = (I_n - \sum_{j=1}^{p} \tilde{\Phi}_j L^j)\tilde{a}$, and:

$$\Lambda^* = \begin{bmatrix} \lambda_1 & -\lambda_1\phi_{1,1} & \dots & -\lambda_1\phi_{1,p} \\ \vdots & & \ddots & \vdots \\ \lambda_n & -\lambda_n\phi_{n,1} & \dots & -\lambda_n\phi_{n,p} \end{bmatrix}.$$

The errors in the measurement equation (108) are now i.i.d.. We are then ready to draw the factors using the Carter and Kohn (1994) procedure, where the transition equation is given by equation (106).[27]

Note that equation (108) starts from $t = p+1$ since we are using the first $p$ observations for pre-whitening. If we were to condition on these first $p$ observations, would use the mean and variance of the unconditional ergodic distribution of the state vector $\tilde{f}_t$ to initialize the Kalman filter. There may be cases where the initial conditions matter, either because $T$ is not too large (relative to $p$) or because the factor is very persistent. Del Negro and Otrok (2008) provide formulas for the mean and variance of the state $\tilde{f}_p$ conditional on the first $p$ observations.

### 6.3.3 Applications of Dynamic Factor Models

dfmapps

How integrated are international business cycles? Are countries more integrated in terms of business cycle synchronization within a region (say, within Europe) than across regions (say, France and the US)? Has the degree of comovement changed significantly over time as trade and financial links have increased? These are all natural questions to address using a dynamic factor model, which is precisely what Kose, Otrok, and Whiteman (2003) do. For instance, Kose, Otrok, and Whiteman

[27] Note that in the the cases $q-1 > p$ one simply needs to add $q-1-p$ columns of zeros to the matrix $\Lambda^*$. Viceversa, in the the case $q-1 < p$ one adds $p-q+1$ columns of zeros to the matrix $\tilde{\Phi}_0$.

(2003) estimate a dynamic factor model on a panel of annual data on output, investment, and consumption for 60 countries and about 30 years. The model includes a world factor that captures the world business cycle, regional factors that capture region-specific cycles (say, Latin America), and from country-specific cycles.[28] [29]

Kose, Otrok, and Whiteman (2003) find that international business cycle comovement is significant. For output in the G7 countries for instance, world cycles are on average as important as country-specific cycles. For the entire world country specific cycles are not surprisingly much more important than world cycles. Regional cycles are not that important at all, suggesting that integration is no higher within regions than across regions. Kose, Otrok, and Whiteman (2008) assess to what extent the importance of world cycles has changed from Bretton Woods period to the early 70s/mid80s, and finally to the current (called 'globalization') period. They argue that for output comovement was highest in the early 70s/mid80s period, which they call 'common shock' period, but that the importance of the world business cycle has also increased from the Bretton Woods to the 'globalization' period. In Kose, Otrok, and Eswar (2008) the authors provide evidence of 'decoupling' during the 'globalization' period. That is, the authors show that industrialized and emerging economies have become more integrated within each group, but less integrated across groups (note that here the regions are defined very differently than in Kose, Otrok, and Whiteman (2003)). Finally, Crucini, Kose, and Otrok (2008) try to explain the unobserved factors, and in particular the world factors, in terms of observables such as productivity, oil prices, *et cetera*. They find that productivity is the main driving force of international business cycles.[30]

[28] Note that in maximum likelihood/Bayesian methods it is quite natural to construct regional or country-specific factors by simply imposing the restriction that that the factor has zero loading on series that do not belong to that region/country. Moreover, the county factor can be estimated even if the number of series per country is small, as is the case in Kose, Otrok, and Whiteman (2003). Non-parametric methods such as principal components have a harder dealing with the first issue (imposing zero restriction on loadings) and with characterizing the uncertainty that results from estimating country factors with a small cross-section per country.

[29] Ng, Moench, and Potter (2008) take a different modeling strategy, and use hierarchical factors. In this case they would estimate factors for each country , and then impose that the country factors themselves are driven by a factor model, where the factors would be the regional factors. Next, they would postulate that regional factors also evolve according to a factor model, where the common factors are the world factors. It turns out that this approach is more parsimonious than Kose, Otrok, and Whiteman (2003)'s.

[30] The study of house prices is another interesting application of factor models, given that these have both an important national (due to interest rates) and a regional (due to migration, demo-

### 6.3.4 Dynamic Factor Models with Time-Varying Parameters

dfmtvp

For the same reasons that it may be useful to allow parameter variation in a VAR as we saw in chapter 5, it may be useful to allow for time-variation among the parameters of a factor model. For instance, comovement across countries may have changed as a result of increased financial or trade integration, or because of monetary arrangements (monetary unions, switches from fixed to flexible exchange rates, *et cetera*). Del Negro and Otrok (2008) accomplish that by modifying the standard factor model in three ways. First, they make the loadings time-varying. This feature allows for changes in the sensitivity of individual series to common factors. The measurement equation (101) becomes:

$$y_{i,t} = a_i + \lambda_{i,t} f_t + \xi_{i,t} \tag{109}$$

where they control the evolution of the factor loadings by requiring that they follow a random walk without drift as in Cogley and Sargent (2005):

$$\lambda_{i,t} = \lambda_{i,t-1} + \sigma_{\eta_i} \eta_{i,t}. \tag{110}$$

Importantly for identification purposes, they assume that $\eta_{i,t} \sim \mathcal{N}(0,1)$ and is independent across $i$.

The second innovation amounts to introducing stochastic volatility in the law of motion of the factors and the idiosyncratic shocks. The transition equations, (102) and (103) become:

$$f_t = \phi_{0,1} f_{t-1} + .. + \phi_{0,q} f_{t-q} + e^{h_{0,t}} u_{0,t} \quad \text{,and} \tag{111}$$

$$\xi_{i,t} = \phi_{i,1} \xi_{i,t-1} + .. + \phi_{i,p_i} \xi_{i,t-p_i} + \sigma_i e^{h_{i,t}} u_{i,t}, \tag{112}$$

respectively, where to keep notation simple we assume we have only one factor. The terms $e^{h_{i,t}}$ represents the stochastic volatility components, where $h_{i,t}$ follows a random walk process:

$$h_{i,t} = h_{i,t-1} + \sigma_{\zeta_i} \zeta_{i,t}, \ i = 0, \ 1, .., n \tag{113}$$

with $\zeta_{i,t} \sim \mathcal{N}(0,1)$ and is independent across $i$ (note that $h_{0,t}$ denotes the factor's stochastic volatility term). We assume that $h_{i,t} = 0$ for $t \leq 0, \ i = 0, 1, ., n$, that

graphics, ...) components. Del Negro and Otrok (2007) apply dynamic factor models to study regional house prices in the US.

is, before the beginning of the sample period there is no stochastic volatility: This assumption allows us to derive an ergodic distribution for the initial conditions.[31] Del Negro and Otrok (2008) assume that the priors for the standard deviations of the innovations to the law of motions of the loadings ($\{\sigma_{\eta_i}\}_{i=1}^n$) and stochastic volatilities ($\{\sigma_{\zeta_i}\}_{i=0}^n$ (where $i = 0$ denotes the stochastic volatility for the factors' law of motion) follow an inverse gamma distributions $IG(\nu_{\eta_i}, s_{\eta_i}^2)$ and $IG(\nu_{\zeta_i}, s_{\zeta_i}^2)$, respectively.

The Gibbs sampling procedure reduces to four main blocks. (i) The first block conditions on the factors, time-varying loadings, and stochastic volatilities to sample from the posterior of the constant term $a_i$, the autoregressive parameters $\{\phi_{i,1}, \ldots, \phi_{i,p_i}\}$, and the non time-varying component of the variance $\sigma_i^2$ as described in Section 6.3.1 (the procedure is modified to take the heteroskedasticity introduced by the stochastic volatilities into account). (i) In the next block, which is again similar to the one described in Section 6.3.2, the factors $f_t$ are drawn conditional on all other parameters using the state space representation of the model, as in Carter and Kohn (1994). (iii) The third block draws the time-varying loadings $\lambda_{i,t}$, again using Carter and Kohn's algorithm. In this block the factors are treated as known quantities. (iv) The last block samples the stochastic volatilities using the procedure of Kim, Shephard, and Chib (1998).

Del Negro and Otrok (2008) apply this model to study the time-varying nature of international business cycles. Mumtaz and Surico (2008), in a related development, introduces time-variation in the law of motion of the factors (but not in any of the other parameters) and use their model to cross-country inflation data.

### 6.3.5 Factor Augmented VARs

fvar

Bernanke, Boivin, and Eliasz (2005) introduce Factor augmented VARs (or FAVARs). The FAVAR approach introduces two changes to the standard factor model. First, FAVAR allows for additional observables (the Fed Funds rate, for instance) to enter the measurement equation, which becomes:

$$y_{i,t} = a_i + \lambda_i f_t + \gamma_i x_t + \xi_{i,t} \tag{114}$$

[31] The assumption that $h_{0,0} = 0$ also makes it possible to identify the scale of the loadings $\lambda_{i,t}$ while the assumption that $h_{i,0} = 0$ makes it possible to identify the non-time varying component of the idiosyncratic shocks standard deviation $\sigma_i$. See the discussion in Del Negro and Otrok (2008).

where $x_t$ and $\gamma_i$ are $m \times 1$ and $1 \times m$ vectors, respectively. Second, the unobservable factors and the observable $x_t$ are assumed to jointly follow a vector autoregressive processes of order $q$:

$$\begin{bmatrix} f_t \\ x_t \end{bmatrix} = \Phi_{0,1} \begin{bmatrix} f_{t-1} \\ x_{t-1} \end{bmatrix} + .. + \Phi_{0,q} \begin{bmatrix} f_{t-q} \\ x_{t-q} \end{bmatrix} + u_{0,t}, \text{ and} \tag{115}$$

where the $\Phi_{0,j}$ matrices are now of size $(\kappa+m)\times(\kappa+m)$. While in early applications of FAVAR (Bernanke, Boivin, and Eliasz (2005)) estimation involved a two-step procedure where first the factors were extracted from a large cross-section and then the VAR parameters were estimated, later approaches involved joint estimation of the factors and the VAR parameters. At least in principle, estimating a FAVAR is a straightforward application of the tools described in Section 6.3.2. For given factors, obtaining the posterior distribution for the parameters of (114) and (115) is straightforward. Likewise, the factors can be drawn using expressions (114) and the first $\kappa$ equations of the VAR in (115), as the measurement and transition equations, respectively, in a state-space representation. The appeal of the FAVAR is that it afford a combination of factor analysis with the structural VAR analysis described in Section 2.3. As a consequence, one can directly model the impact of variations in some particular observables (e.g., the Fed Funds rate) on the evolution of the factors. For instance, Bernanke, Boivin, and Eliasz (2005) apply their model to study the effects of monetary policy shocks in the U.S. Belviso and Milani (2006) also apply a FAVAR to study the effects of monetary policy shocks, and in addition attempt to provide the factors with a structural interpretation. Ludvigson and Ng (2008) apply the methodology to study bond risk premia.

### 6.3.6 Combining DSGE Models and Factor Models

dfmdsge

Boivin and Giannoni (2006) estimate a factor model where the factors are the states from a DSGE model and the factor dynamics are therefore subject to the restrictions implied by the model. As discussed in Section 4, the solution of a log-linearized DSGE models can be written as:

$$s_t = \Phi_1(\theta) s_{t-1} + \Phi_\xi(\theta)\epsilon_t. \tag{116}$$

where the state vector $s_t$ includes all endogenous and exogenous variables relevant to the model. Define a vector $f_t$ of all the economic concepts for which we have

measurement – e.g., output, inflation, *et cetera* – and a matrix $F$ that selects $f_t$ out of $s_t$:[32]

$$f_t = F s_t. \tag{117}$$

Standard practice in the estimation of DSGE models is to select only one empirical counterpart for each economic concept, so that the measurement equation can be written in vectorized forms as:

$$y_t = f_t. \tag{118}$$

where for convenience we ignore the constants. Boivin and Giannoni (2006) argue that in fact there are several empirical counterparts for each economic concept – for instance, inflation can be measured using the GDP deflator, the Consumer Price Index, the Personal Consumption Expenditures deflator, and so on – and that the measurement equations should be generalized accordingly as:

$$y_t = \Lambda^Y f_t + \xi_t, \tag{119}$$

The vector $\xi_t$ stacks all measurement errors, which eveolve independently from one another according to the law of motion (103). Each equation of (119) can be written like the measurement equation (101) described in Section 6.3, with the additional restriction that each observable loads only onto one factor – the economic concept $f_t^j$ for which $y_{i,t}$ is one of the empirical counterparts:

$$y_{i,t} = \lambda_{i,j}^Y f_t^j + \xi_{i,t}, \tag{120}$$

We can now substitute $f_t$ for $s_t$ and the measurement equations become:

$$y_{i,t} = \lambda_i s_t + \xi_{i,t}, \tag{121}$$

where $\lambda_i$ obeys the restriction

$$\lambda_i = \lambda_{i,j}^Y F.$$

Boivin and Giannoni (2006) also consider variables (for instance, consumer confidence) that do not have a clear counterpart in the economic model. For these variables the vector $\lambda_i$ is unrestricted.

Estimation of these model by and large follows the Gibbs sampler described in Section 6.3.2. Given the states $s_t$, the $\lambda_i$ and the other parameters of (103) can be

[32] We can always suitably expand the definition of $s_t$ to incorporate whatever variable of interest the model produces.

recovered equation by equation using the approach in Chib and Greenberg (1994). For given vector $\theta$ of DSGE model parameters, the states can be drawn using the procedure in Carter and Kohn (1994, ADD). The main difference arises when it comes to draw the parameters of the law of motion of the states (116). These parameters are now function of the vector $\theta$ of DSGE model parameters. Conditional on $\lambda_i$ and the other parameters of (103), $\theta$ can be drawn using a Metropolis-Hastings step, where the measurement and transition equations are given by (121) and (116), respectively.

Boivin and Giannoni (2006) argue that their approach to estimating DSGE model in a "data rich environment" provides several advantages relative to the standard practice. First of all, the additional information coming from the large cross section can provide better estimates of the states $s_t$ and of the structural shocks driving the economy.[33] In addition, they provide evidence that their procedure can deliver different estimates of the DSGE model parameters relative to the standard approach. Boivin and Giannoni (2008) discuss optimal monetary policy in a in a "data rich environment".

Kryshko, Schorfheide, and Sill (2008) provide an approach for forecasting variables that do not explicitly appear in the DSGE model that is to some extent in the spirit of that in Boivin and Giannoni (2006), but is computationally much simpler.[34] Specifically, Kryshko, Schorfheide, and Sill (2008) estimate the DSGE using the standard approach. Conditional on the posterior mean of those estimates, they use the Kalman filter to obtain the filtered states $s_t$. In the next step, they use these states as regressors in a measurement equation like (121), where the errors $\xi_{i,t}$ follow an autoregressive law of motion like (103). Unlike Boivin and Giannoni (2006), the $\lambda_i$ in Kryshko, Schorfheide, and Sill (2008) are generally unrestricted. However, these authors use the DSGE model to form a prior for $\lambda_i$. The approach is straightforward to apply as it does not involve a feedback between the $y_{i,t}$ and the states, as in Boivin and Giannoni (2006). Precisely for this reason, however, their approach differs from Boivin and Giannoni (2006) as it does not exploit the data-rich environment to extract estimates of the states.

[33] Bernanke, Boivin, and Eliasz (2005) make a similar point in the context of a structural VAR.

[34] Monti, Giannone, and Reichlin (2008) consider a similar approach where information at different frequency (monthly, quarterly) is used for forecasting with a DSGE model. They postulate a so-called bridge equation that links the observables in the DSGE model to the monthly variables that are not explicitly included in the DSGE model.

# 7 Model Uncertainty

An important aspect of empirical work in macroeconomics is model uncertainty. We have encountered "model" uncertainty in various places throughout this chapter. For instance, in the context of VARs there is uncertainty about the appropriate number of lags and cointegration restrictions. In the context of DSGE model, a researcher might be uncertain whether price stickiness, wage stickiness, informational frictions, or monetary frictions are quantitatively important for the understanding of business cycle fluctuations and should be accounted for when designing monetary and fiscal policies. In view of the proliferation of hard-to-measure coefficients in time-varying parameter models, there is uncertainty about the importance of such features in empirical models. Researchers working with dynamic factor models are typically uncertain about the number of factors necessary to capture the comovements in a cross-section of macroeconomic or financial variables.

In a Bayesian framework, model uncertainty is conceptually not different from parameter uncertainty. Consider the following simple example. We could distinguish the following two (nested) models:

$$\mathcal{M}_1: \quad y_t = u_t, \quad \mathcal{M}_2: \quad y_t = \theta_{(2)} x_t + u_t.$$

Here $\mathcal{M}_1$ restricts the regression coefficient $\theta_{(2)}$ in $\mathcal{M}_2$ to be equal to zero. Bayesian analysis allows us to place probabilities on the two models, denoted by $\pi_{i,0}$, and specify a prior distribution for $\theta_{(2)}$ in model $\mathcal{M}_2$. Suppose that $\pi_{1,0} = \lambda$ and $\theta_{(2)} \sim N(0,1)$. Then the mixture of $\mathcal{M}_1$ and $\mathcal{M}_2$ is equivalent to a model $\mathcal{M}_0$

$$\mathcal{M}_0: \quad y_t = \theta_{(0)} x_t + u_t, \quad u_t \sim iidN(0,1), \quad \theta \in \mathbb{R}.$$

with a prior distribution for $\theta_{(0)}$ that is of the form

$$\theta_{(0)} \sim \begin{cases} 0 & \text{with prob. } \lambda \\ N(0,1) & \text{with prob. } 1-\lambda \end{cases}$$

More generally, we can always construct a prior on a sufficiently large parameter space such that model uncertainty can be represented as parameter uncertainty. However, in many applications it is convenient to refer restricted versions of a large encompassing model, e.g. a VAR whose lag length is restricted to $p$, as models themselves.

The remainder of this section is organized as follows. Section 7.1 discusses the computation of posterior model probabilities and their use to select among a collection of models. However, rather than first selecting a model and then conditioning on the selected model in the subsequent analysis, it is more desirable to average across models and to explicitly take model uncertainty into account when making decisions. We use a stylized optimal monetary policy example to highlight this point in Section 7.3. In many macroeconomic applications, in particular those that are based on DSGE models, posterior model probabilities often seem to be implausible, in that one specification essentially attains posterior probability one and all other specification receive probability zero. These probabilities seem *implausible* because they do not capture the variation of results found across different studies. In view of potentially *implausible* posterior model probabilities a decision maker might be inclined to robustify her decisions and we discuss robustness considerations in Section 7.4.

## 7.1 Posterior Model Probabilities and Selection

modelsel

Suppose we have a collection of $M$ models denoted by $\mathcal{M}_1$ through $\mathcal{M}_M$. Each model has a parameter vector $\theta_{(i)}$ and prior probability $\pi_{i,0}$. The posterior probabilities are given by

$$\pi_{i,T} = \frac{\pi_{i,0} p(Y|\mathcal{M}_i)}{\sum_{j=1}^{M} \pi_{j,0} p(Y|\mathcal{M}_j)}, \quad p(Y|\mathcal{M}_i) = \int p(Y|\theta_{(i)}, \mathcal{M}_i) p(\theta_{(i)}|\mathcal{M}_i) d\theta_{(i)}, \quad (122)$$

where $p(Y|\mathcal{M}_i)$ is the marginal likelihood or data density associated with model $\mathcal{M}_i$. As long as the likelihood functions $p(Y|\theta_{(i)}, \mathcal{M}_i)$ and prior densities $p(\theta_{(i)}|\mathcal{M}_i)$ are properly normalized for all models the posterior model probabilities are well defined. Improper prior densities[35], however, pose a problem. We could pre-multiply an improper prior by an arbitrary factor $c$, which would result in an (arbitrary) scaling of the marginal likelihood and leave the posterior probabilities indeterminate. One can obtain meaningful posterior model probabilities an improper prior $p(\theta)$ (we omitted the $i$ subscript and the model indicator from the notation) by replacing the marginal likelihood $p(Y)$ with a predictive likelihood of the form

$$p(Y_{k+1,T}|Y_{1,k}) = \int p(Y_{k+1,T}|Y_{1,k}, \theta) p(\theta|Y_{1,k}) d\theta,$$

[35] Prior densities for which $\int p(\theta) d\theta = \infty$.

provided the $k$-observation posterior $p(\theta|Y_{1,k})$ is proper.[36] Since for any model with a parameter vector $\theta$

$$\ln p(Y_{1,T}) = \sum_{t=1}^{T} \ln p(y_t|Y_{1,t-1}) = \sum_{t=1}^{T} \ln \int p(y_t|\theta, Y_{1,t-1}) p(\theta|Y_{1,t-1}) d\theta$$

log marginal likelihoods can be interpreted as pseudo-out-of-sample one-step-ahead predictive scores.

While the calculation of posterior probabilities is conceptually straightforward, it can be computationally challenging. First, there are only a few instance, such as the VAR model in (1) with conjugate MNIW prior, in which the marginal likelihood $p(Y)$ can be computed analytically. If fact, for priors represented through dummy observations the formula is given in (80). We also mentioned in Section 4.6.2 that for a DSGE model, or other models for which posterior draws have been obtained using the RWM Algorithm, numerical approximations to marginal likelihoods can be obtained using Geweke (1999, ADD)'s approach or the method proposed by Chib and Jeliazkov (2001, ADD). A more detailed discussion of numerical approximation techniques for marginal likelihoods is provided in Chapter [Bayesian Computation]. Finally, marginal likelihoods can be approximated analytically using a so-called Laplace approximation, which approximates $\ln p(Y|\theta) + \ln p(\theta)$ by a quadratic function centered at the posterior mode or the maximum of the likelihood function. The most widely-used Laplace approximation is the one due to Schwarz (1979, ADD). Phillips (1996, ADD) and Chao and Phillips (1999, ADD) provide extensions to non-stationary time series models and reduced rank VARs.

Second, the model space might be quite large and require the calculation of marginal likelihoods for many specifications. Consider the empirical illustration in Section 2, which involved a 4-variable VAR with 4 lags, which leads to a coefficient matrix $\Phi$ with 68 elements. Suppose we construct submodels by restricting VAR coefficients to zero. Based on the exclusion of parameters we can easily generate $2^{68} \approx 3 \cdot 10^{20}$ submodels. Even if we use a conjugate prior that leads to an analytical formula for the marginal likelihoods of the submodels, the computation of the posterior probabilities is a daunting task. As an alternative to computing the marginal likelihood for each of the $2^{68}$ specifications, George, Ni, and Sun (2008) develop a stochastic search algorithm for VAR restrictions.

[36] For instance, Villani (2001) discusses lag length selection using predictive marginal likelihoods based on improper priors in the context of VARs with cointegration restrictions.

The literature distinguishes between model selection and model averaging. We will provide a few remarks on model selection and defer a discussion of model averaging, which is a more natural approach in a Bayesian framework, to Section 7.2. Suppose that a researcher faces a loss of zero if she chooses the "correct" model and a loss of one if an incorrect model is chosen. Under this loss function, it is straightforward to verify that the posterior expected loss is minimized by selecting the model with the highest posterior probability. Thus, Bayesian model selection simply amounts to choosing the highest posterior probability models.

¿From a frequentist perspective Bayesian model selection procedures are consistent in the sense that if the data generating process is contained in one of the models under consideration then the posterior probability of this model will converge to one as the sample size tends to infinity. If the models $\mathcal{M}_i$ are nested then the smallest model that contains the data generating process will be chosen. This consistency result tends to remain valid if the marginal likelihoods are replaced by Laplace approximations, e.g. Schwarz (1979, to be added) and Phillips and Ploberger (1996). Moreover, the consistency is preserved even in non-stationary time series models. For instance, Chao and Phillips (1999, to be added) show formally that a Bayesian model selection criterion leads to a consistent selection of cointegration rank and lag length in vector autoregressive models. Fernandez-Villaverde and Rubio-Ramirez (2004) emphasize that if none of the models in the model space is "true" model selection based on posterior odds leads asymptotically to the model that is closest to the data-generating process in terms of the Kullback-Leibler divergence.

## 7.2 Decision Making with Multiple Models

decisionmaking

For many applications the zero-one loss function underlying the model selection problem is not particularly attractive. Economic policy makers are often confronted with choosing policies under model uncertainty. We will illustrate a prototypical decision problem. Suppose that output $y_t$ and inflation $\pi_t$ are related to each other according to one of the two Phillips curve relationships

$$\mathcal{M}_1 : \; y_t = \frac{1}{10}\pi_t + \epsilon_{s,t}, \quad \mathcal{M}_2 : \; y_t = \pi_t + \epsilon_{s,t}, \tag{123}$$

where $\epsilon_{s,t}$ is a cost (supply) shock. The demand side of the economy leads to the following relationship between inflation and money $m_t$:

$$\pi_t = m_t + \epsilon_{d,t}, \tag{124}$$

where $\epsilon_{d,t}$ is a demand shock. Finally, up until period $T$ monetary policy was $m_t = 0$. However, in period $T$ the central back is considering a policy of the form

$$m_t = -\epsilon_{d,t} + \delta\epsilon_{s,t} \tag{125}$$

for $t > T$ under the loss function

$$\widetilde{L}_t = (\pi_t^2 + y_t^2). \tag{126}$$

All variables in this model are meant to be in log deviations from some steady state and the shocks are assumed to be iid $N(0,1)$.

Now define $\theta(\mathcal{M}_i)$ such that $\theta(\mathcal{M}_1) = 1/10$ and $\theta(\mathcal{M}_2) = 1$. If one averages with respect to the distribution of the supply shocks the expected loss associated with a policy $\delta$ under model $\mathcal{M}_i$ is

$$L(\mathcal{M}_i, \delta) = [(\delta\theta(\mathcal{M}_i) + 1)^2 + \delta^2]. \tag{127}$$

Suppose that after observing the data the posterior probabilities are $\pi_{1,T} = 0.61$ and $\pi_{2,T} = 0.39$. There are essentially three ways of arriving at a decision $\delta$.

First, one could condition on the highest posterior probability model which is $\mathcal{M}_1$. For $\theta = 1/10$ the loss function (127) is minimized at $\delta^*(\mathcal{M}_1) = -0.10$. However, this decision completely ignores the loss that occurs if in fact $\mathcal{M}_2$ is the correct model. Second, one can derive decisions $\delta^*(\mathcal{M}_i)$ that are optimal for model $\mathcal{M}_i$, $i = 1, 2$, and examine the loss of $\delta^*(\mathcal{M}_i)$ if model $\mathcal{M}_j$ is correct. The matrix of expected losses is summarized in Table 3. Notice that there is a large loss associated with $\delta^*(\mathcal{M}_2)$ if in fact $\mathcal{M}_1$ is the correct model. The last column provides the posterior risks associated with the two decision, which is an average of the expected loss reported in columns (2) and (3), weighted by the posterior probabilities of the two models:

$$\mathcal{R}(\delta) = \pi_{1,T}L(\mathcal{M}_1, \delta) + \pi_{2,T}L(\mathcal{M}_2, \delta). \tag{128}$$

It turns out to be optimal to implement the decision associated with model $\mathcal{M}_2$ despite $\mathcal{M}_1$ being the highest posterior probability model. Finally, one can choose $\delta$ to directly minimize the posterior risk in (128). This leads to $\delta^* = -0.32$ and the posterior risk associated with this decision is $\mathcal{R}(\delta^*) = 0.85$.

Insert Table Here

This simple example illustrates that conditioning on the highest posterior probability model can clearly lead to sub-optimal decisions. At a minimum one should account for the loss of a decision that is optimal under $\mathcal{M}_i$ if in fact one of the other models $\mathcal{M}_j$, $j \neq i$, is correct. The decision that is optimal from a Bayesian perspective is obtained by minimizing the expected loss under a mixture of models.

In more realistic applications the two simple models would be replaced by more sophisticated DSGE models. These models would itself depend on unknown parameters. Cogley and Sargent (2005, ADD) provide a nice illustration of the notion that one should not implement the decision of the highest posterior probability model if it has disastrous consequences if one of the other models is correct. They consider a traditional Keynesian model with a strong output and inflation trade-off versus a model that incorporates the natural rate hypothesis. While according to their analysis the posterior probability on the Keynesian model was very small by the mid 1970s and the natural rate model suggested to implement a disinflation policy, the costs associated with this disinflation if in fact the Keynesian model is correct were very high. This consideration delayed the disinflation until about 1980.

Often, loss depends on future realizations of $y_T$. In this case predictive distributions are important. Consider for example a prediction problem. The $h$-step ahead predictive density can be expressed as the mixture

$$p(y_{T+h}|Y_{1,T}) = \sum_{i=1}^{M} \pi_{i,T} p(y_{T+h}|Y_{1,T}, \mathcal{M}_i). \tag{129}$$

Thus, $p(y_{T+h}|Y_{1,T})$ is the result of the Bayesian averaging of model-specific predictive densities $p(y_{T+h}|Y_{1,T})$. Notice that only if the posterior probability of one of the models is essentially equal to one, conditioning on the highest posterior probability leads to approximately the same predictive density as model averaging. If the goal is to generate point predictions under a quadratic loss function then it would be optimal to average point forecast from the $M$ models. This is often referred to as Bayesian forecast combination. There exists an extensive literature on applications of Bayesian model averaging. For instance, Min and Zellner (1990, ADD) use posterior model probabilities to combine forecasts, Wright (2008) uses Bayesian model averaging to construct exchange rate forecasts, and Strachan and van Dijk (2006) average across VARs with different lag lengths and cointegration restrictions to study the dynamics of the "Great Ratios."

Bayesian model averaging has become popular in growth regressions following the work of Fernandez, Ley, and Steel (2001), Sala-i Martin, Doppelhofer, and Miller (2004), Masanjala and Papageorgiou (Forthcoming). In particular, Sala-i Martin, Doppelhofer, and Miller (2004) use a combination of Bayesian model averaging and uninformative priors on individual regression coefficients called BACE (Bayesian Averaging of Classical Estimates). The BACE approach computes posterior odds for different specifications in growth regressions, and then assesses the magnitude and statistical significance of regressors by taking weighted averages of the posterior means and variances across specifications using the posterior odds as weights. Ciccone and Jarociski (2007) show that the BACE approach may not be robust to differences in the datasets because such differences can result in large disparities in the posterior odds across models. They argue that more informative priors on the regression coefficients robustify the Bayesian model averaging approach to cross-country.

## 7.3 Implausible Posterior Probabilities

implausibleodds

While Bayesian model averaging is conceptually very attractive, it very much relies on the notion that the posterior model probabilities provide a "plausible" characterization of model uncertainty, rather than merely one that is formally correct. Consider a central bank deciding on its monetary policy. Suppose that *a priori* the policy makers entertain the possibility that either wages or prices of intermediate goods producers are subject to nominal rigidities. These rigidities have the effect that wage (or price) setters are not able to adjust their nominal wages (prices) optimally, which distorts relative wages (prices), and ultimately, leads to the use of an inefficient mix of labor (intermediate goods). The central bank could use its monetary policy instrument to avoid the necessity of wage (price) adjustments and thereby nullifying the effect of the nominal rigidity.

Based on the tools and techniques in the preceding Sections, one could now proceed by estimating two models, one in which prices are sticky and wages are flexible, and one in which prices are flexible and wages are sticky. Results for such an estimation, based on a variant of the Smets and Wouters (2007) models, have been reported, for instance, in Table 5 of Del Negro and Schorfheide (2008). According to their estimation, conducted under various prior distributions, U.S. data favor the sticky price version of the DSGE model with odds that are greater than $e^{40}$. Such

odds are not uncommon in the DSGE model literature. If these odds are taken literally, they would imply that we should completely disregard the possibility that wages are sticky. Related, Del Negro, Schorfheide, Smets, and Wouters (2007) compare versions of DSGE models with nominal rigidities in which those households (firms) that are unable to re-optimize their wages (prices) are either indexing their past price by the long-run inflation rate or by last period's inflation rate (dynamic indexation). According to their Figure 4, the odds in favor of the dynamic indexation are greater than $e^{20}$, which again seems very decisive.

Schorfheide (2008, ADD) surveys a large number of DSGE model-based estimates of price and wage stickiness and the degree of dynamic indexation. While the papers included in this survey build on the same theoretical framework, variations in some details of the model specification as well as in the choice of observables lead to a significant variation in parameter estimates and model rankings. Thus, posterior model odds from any individual study, even though formally correct, appear to be overly decisive and thereby implausible.

Part of the problem is that each model represents a stylized representation of a particular economic concept, such as wage or price stickiness, augmented by auxiliary mechanisms that are designed to make the model fit the data. By looking across studies one encounters several representations of what is essentially the same basic concept, but each representation leads to a different time series fit and makes posterior probabilities appear fragile across studies. This problem is exacerbated by model misspecification. We shall provide two stylized example. The second example is taken from Sims (2003, ADD).

**Example 1:** Suppose that the world works according to $y_t \sim iidN(0,1)$ and a macroeconomist considers two stylized representations $\mathcal{M}_i$: $y_t \sim iid(\mu_i, \sigma_i^2)$, $i = 1, 2$, where $\mu_i$ and $\sigma_i^2$ are fixed in both models. Assuming that the two models have the same prior probability, the expected log posterior odds ratio is

$$\begin{aligned} I\!E\left[\ln \frac{\pi_{1,T}}{\pi_{2,T}}\right] &= -\frac{T}{2}\left[\ln \sigma_1^2 + \frac{1}{\sigma_1^2}(1+\mu_1^2)\right] + \frac{T}{2}\left[\ln \sigma_2^2 + \frac{1}{\sigma_2^2}(1+\mu_2^2)\right] \\ &\approx -\frac{T}{2}\left[3(\sigma_1^2-1)^2 + \mu_1^2 - 2(\sigma_2^2-1)^2 - \mu_2^2\right] \end{aligned}$$

The approximation is obtained by a second-order Taylor expansion around the "true" values $\mu = 0$ and $\sigma^2 = 1$. Suppose that the location parameters $\mu_1$ and $\mu_2$ capture the key economic concept, such as wage or price stickiness, and the scale

parameters are generated through the various auxiliary assumptions that are made to obtain a fully-specified DSGE models. If one of the models is well specified, say $\mathcal{M}_1$, in the sense that $\mu_1 \approx \mu_0$, then posterior probabilities for $\mathcal{M}_1$ and $\mathcal{M}_2$ would favor $\mathcal{M}_1$, possibly in a decisive manner, even if auxiliary assumptions lead to some distortions $\sigma_1^2 \neq \sigma_0^2$. These odds would be fairly robust to changes in the auxiliary assumptions (from $\sigma_i^2$ to $\tilde{\sigma}_i^2$). However, if it turns out that both $\mu_1$ and $\mu_2$ are quite different from $\mu_0$, then changes in the auxiliary assumptions underlying the model have a much greater effect on the relative odds of the two misspecified models. For instance, if $\mu_1^2 \approx \mu_2^2$ then the expected log posterior odds depend solely on the $\sigma_i^2$'s, that is, the auxiliary assumptions.

**Example 2,** (Sims, 2003): Suppose that according to model $\mathcal{M}_1$ $y_t \sim iidN(0, 0.01)$ and according to model $\mathcal{M}_2$ $y_t \sim iidN(1, 0.01)$ and that the sample size is $T = 1$. Based on equal prior probabilities, the log posterior odds in favor of model $\mathcal{M}_1$ are

$$\ln \frac{\pi_{1,T}}{\pi_{2,T}} = -\frac{1}{2 \cdot 0.01}[y_1^2 - (y_1 - 1)^2] = 100(y_1 - 1/2)^2.$$

Thus, for values of $y_1$ less than 0.45 or greater than 0.55 the posterior odds are greater than 100 in favor of one of the models. But suppose that $y_1 = 0.6$. While the posterior odds in favor of $\mathcal{M}_2$ are decisive, the likelihood to observe a value in the vicinity of 0.6 under either model is very small.

Thus, the key problems in the use of posterior probabilities in the context of DSGE models are that the models often suffer from misspecification, they tend to capture one of many possible representations of a particular economic mechanism which means that one might be able to find versions of these models that are much closer together and deliver less decisive odds. Posterior odds in the magnitude of $e^{20}$ or $e^{40}$ are suspicious and often indicate that we should compare a different models. Sims (2003) recommends to introduce continuous parameters such that different sub-model specifications can be nested in a larger encompassing model. The downside of creating these encompassing models is that it is potentially difficult to properly characterize multi-modal posterior distributions in high-dimensional parameter spaces. Hence, a proper characterization of posterior uncertainty about the strength of various competing decision-relevant economic mechanisms remains elusive.

## 7.4 Robustness

robustness

There is a growing literature in economics that studies the robustness of decision rules to model misspecification, see Hansen and Sargent (Book, ADD). Underlying this robustness is typically a static or dynamic two-person zero-sum game. The decision maker, in our case the central bank choosing a monetary policy indexed by $\delta$, is minimizing a loss function while a malevolent fictitious other, "nature", chooses the misspecification to harm the decision maker. We consider a simple illustration in which "nature" distorts the model probabilities subject to a penalty function that is increasing in the size of the distortion. The central bank's decision is robust, if it corresponds to a Nash equilibrium in the two person game.

In our Bayesian framework the risk-sensitivity that is inherent in a robust control approach can be introduced as follows by the following fictitious game between "nature" and the policy maker:

$$\min_{\delta} \max_{q\in[0,1/\pi_{1,T}]} \quad q\pi_{1,T}L(\mathcal{M}_1,\delta) + (1-q\pi_{1,T})L(\mathcal{M}_2,\delta) + \frac{1}{\tau}\Big[\pi_{1,T}\ln(q\pi_{1,T}) + (1-\pi_{1,T})\ln(1-q\pi_{1,T})\Big]. \tag{130}$$

Here "nature" uses $q$ to distort the posterior model probability of model $\mathcal{M}_1$. This distortion may reflect concerns about the plausible updating of model odds in light of the data (see Section 7.3). To ensure that the distorted probability of $\mathcal{M}_1$ lies in the unit interval the domain of $q$ is restricted to $[0, 1/\pi_{1,T}]$. The second term in (130) is a penalizes the distortion as a function of the Kullback-Leibler divergence between the undistorted and distorted probabilities. If $\tau$ is equal to zero, then the penalty is infinite and "nature" will not distort $\pi_{1,T}$. If on the other hand $\tau = \infty$, then conditional on a particular $\delta$ "nature" will set $q = 1/\pi_{1,T}$ if $L(\mathcal{M}_1,\delta) > L(\mathcal{M}_2,\delta)$ and $q = 0$ otherwise. For selected values of $\tau$ the Nash equilibrium is summarized in Table 4. In our example $L(\mathcal{M}_1,\delta) > L(\mathcal{M}_2,\delta)$ in the relevant region for $\delta$. Thus, nature has an incentive to increase the probability of $\mathcal{M}_1$ and in response the policy maker reduces (in absolute terms) her response $\delta$ to a supply shock.

Insert Table Here

# 8 Concluding Remarks

conclusion

The Bayesian paradigm provides a rich framework for inference and decision making with modern macroeconometric models such as DSGE models and VARs.[c1] The econometric methods can be tailored to cope with the challenges in this literature: potential model misspecification and a trade-off between theoretical coherence and empirical fit, identification problems, and estimation of models with many parameters based on relatively few observations. Advances in Bayesian computations let the researcher efficiently deal with numerical complications that arise in models with latent state variables, such as regime-switching models, or nonlinear state-space models.

[c1] fs: To be edited.

## References

Adolfson, M., M. K. Andersson, J. Lind, M. Villani, and A. Vredin (2007): "Modern Forecasting Models in Action: Improving Macroeconomic Analyses at Central Banks," *International Journal of Central Banking*, 3(4), 111–144.

Adolfson, M., S. Laseen, J. Linde, and M. Villani (2007): "Bayesian estimation of an open economy DSGE model with incomplete pass-through," *Journal of International Economics*, 72(2), 481 – 511.

Adolfson, M., S. Lasen, J. Lind, and M. Villani (2008): "Empirical Properties Of Closed- And Open-Economy Dsge Models Of The Euro Area," *Macroeconomic Dynamics*, 12, 2–19.

Adolfson, M., J. Lind, and M. Villani (2007): "Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model," *Econometric Reviews*, 26, 289–328.

Adolfson, M., J. Linde, S. Laseen, and M. Villiani (2005): "The Role of Sticky Prices in an Open Economy DSGE Model," *Journal of the European Economic Association*, 3(2-3), 444 – 457.

Aldofson, M., S. Laseen, J. Linde, and M. Villani (2008): "Evaluating an estimated new Keynesian small open economy model," *Journal of Economic Dynamics and Control*, 32(4), 2690 – 2721.

An, S., and F. Schorfheide (2007): "Bayesian Analysis of DSGE Models–Rejoinder," *Econometric Reviews*, 26, 211 – 219.

Banbura, M., D. Giannone, and L. Reichlin (2008): "Large Bayesian VARs," *ECB Working Paper*.

Belviso, F., and F. Milani (2006): "Structural Factor-Augmented VARs (SFAVARs) and the Effects of Monetary Policy," *Topics in Macroeconomics*, 6(3).

Benati, L. (2008a): "The Great Moderation in the United Kingdom," *Journal of Money, Credit, and Banking*, 40(1), 121 – 147.

——— (2008b): "Investigating Inflation Persistence Across Monetary Regimes," *Quarterly Journal of Economics*, 123(3), 1005 – 1060.

Benati, L., and P. Surico (2008): "Evolving US Monetary Policy and the Decline of Inflation Predictability," *Journal of the European Economic Association*, 6(2-3), 634 – 646.

——— (forthcoming): "VAR Analysis and the Great Moderation," *American Economic Review*.

Bernanke, B. S., J. Boivin, and P. Eliasz (2005): "Measuring the Effects of Monetary Policy," *Quarterly Journal of Economics*, 120(1), 387 – 422.

Blanchard, O. J., and D. Quah (1989): "The Dynamic Effects of Aggregate Demand and Supply Disturbances," *American Econometic Review*, 79(4), 655–673.

Boivin, J., and M. P. Giannoni (2006): "DSGE Models in a Data Rich Enviroment," *Working Paper*.

——— (2008): "Optimal Monetary Policy in a Data-rich Enviroment," *Working Paper*.

Canova, F. (2005): "The Transmission of US Shocks to Latin American," *Journal of Applied Econometrics*, 20, 229–251.

Canova, F., and M. Ciccarelli (forthcoming): "Estimating Multi-country VAR Models," *International Economic Review*.

Canova, F., M. Ciccarelli, and E. Ortega (2007): "Similarities and Convergence in G-7 Cycles," *Journal of Monetary Economics*, 54(3), 850 – 878.

Canova, F., and L. Gambetti (forthcoming): "Structural changes in the US economy: is there a role for monetary policy?," *Journal of Economic Dynamics and Control*.

Canova, F., and G. D. Nicolo (2002): "Monetary Disturbances Matter for Business Fluctuations in the G-7," *Journal of Monetary Economics*, 49(4), 1131–1159.

Carter, C., and R. Kohn (1994): "On Gibbs sampling for state space models," *Biometrika*, 81(3), 541 – 553.

Chamberlain, G., and M. Rothschild (1983): "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51 (5), 1281–1304.

Chang, Y., T. Doh, and F. Schorfheide (2007): "Non-stationary Hours in a DSGE Model," *Journal of Money, Credit, and Banking*, 39(6), 1357 – 1373.

Chib, S., and E. Greenberg (1994): "Bayes Inference in Regression Models with ARMA ( p, q) Errors," *Journal of Econometrics*, 64, 183–206.

Chopin, N., and F. Pelgrin (2004): "Bayesian inference and state number determination for hidden Markov models: an application to the information content of the yield curve about inflation," *Journal of Econometrics*, 123(2), 327 – 244.

Ciccone, A., and M. Jarociski (2007): "Determinants of Economic Growth: Will Data Tell?," *CEPR Discussion Paper*, 6544.

Cochrane, J. H. (1994): "Shocks," *Carnegie Rochester Conference Series on Public Policy*, 41(4), 295–364.

Cogley, T., S. Morozov, and T. Sargent (2005): "Bayesian Fan Charts for U.K. Inflation: Forecasting Sources of Uncertainty in an Evolving Monetary System," *Journal of Economic Dynamics and Control*, 29(11), 1893 – 1925.

Cogley, T., G. E. Primiceri, and T. J. Sarget (2008): "Inflation-gap Persistence in the U.S.," *Working Paper*.

Cogley, T., and T. J. Sargent (2001): "Evolving Post-World War II U.S. Inflation Dynamics," *NBER Macroeconomics Annual*, 16, 331–73, Cambridge and London: MIT Press.

——— (2005): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post-WWII US," *Review of Economic Dynamics*, 8(2), 262 – 302.

Cogley, T., and A. M. Sbordone (forthcoming): "Trend inflation and inflation persistence in the New Keynesian Phillips Curve," *American Econometic Review*.

Cristadoro, R., A. Gerali, S. Neri, and M. Pisani (2008): "Real Exchange-Rate Volatility and Disconnect: An Empirical Investigation," *Bank of Italy Working Paper*, (660).

Crucini, M. J., M. A. Kose, and C. Otrok (2008): "What are the Driving Forces of International Business Cycles?," *Working Paper*.

De Mol, C., D. Giannone, and L. Reichlin (2008): "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?," *Journal of Econometrics*, 146(2), 318 – 328, Honoring the research contributions of Charles R. Nelson.

DeJong, D., B. ingram, and C. Whiteman (1996): "A Bayesian Approach to Calibration," *Journal of Business Economics and Statistics*, 14(4), 1–9.

DeJong, D. N., B. F. Ingram, and C. H. Whiteman (2000): "A Bayesian Approach to Dynamic Macroeconomic," *Journal of Econometrics*, 98(2), 203 – 223.

Del Negro, M. (2003): "Discussion of Cogley and Sargent's "Drifts and Volatilities: Monetary Policy and Outcomes in the Post WWII U.S."," *FRB Atlanta Working Paper*, 2003-06.

Del Negro, M., and C. Otrok (2007): "99 Luftballoons: Monetary Policy and the House Price Boom Across the United States," *Journal of Monetary Economics*, 54(7), 1962 – 1985.

——— (2008): "Dynamic Factor Models with Time-Varying Parameters. Measuring Changes in International Business Cycles.," *FRBNY Staff Report Number 325*.

Del Negro, M., and F. Schorfheide (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45(2), 643 – 673.

——— (2008): "Forming priors for DSGE models (and how it affects the assessment of nominal rigidities)," *Journal of Monetary Economics*, 55(7), 1191 – 1208.

——— (forthcoming): "Monetary Policy with Potentially Misspecified Models," *American Economic Review*.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007): "On the Fit of New Keynesian Models," *Journal of Business and Economic Statistics*, 25(2), 123 – 162.

Doan, T., R. Litterman, and C. A. Sims (1984): "Forecasting and Conditional Projections Using Realistic Prior Distributions," *Econometric Reviews*, 3(4), 1–100.

Doh, T. (2007): "What does the Yield Curve Tell Us About the Federal Reserve's Implicit Inflation Target?," *Federal Reserve Bank of Kansas Research Working Paper*, (07-10).

Doz, C., D. Giannone, and L. Reichlin (2008): "A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models," *CEPR Discussion Papers 5724*.

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

Faust, J. (1998): "The Robustness of identified VAR Conclusions about Money," *Carnegie Rochester Conference Series on Public Policy*, 49(4), 207–244.

Fernandez, C., E. Ley, and M. F. J. Steel (2001): "Model uncertainty in cross-country growth regressions," *Journal of Applied Econometrics*, 16(5), 563 – 576.

Fernandez-Villaverde, J., and J. F. Rubio-Ramirez (2004): "Comparing dynamic equilibrium models to data: a Bayesian approach," *Journal of Econometrics*, 123(1), 153 – 187.

——— (2007a): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74, 1059 – 1087.

——— (2007b): "How Structural are Structural Parameters?," *NBER Macroeconomics Annual*.

George, E. I., S. Ni, and D. Sun (2008): "Bayesian Stochastic Search for VAR Model Restrictions," *Journal of Econometrics*, 142(1), 553 – 580.

Gertler, M., L. Sala, and A. Trigari (forthcoming): "An Estimated Monetary DSGE Model with Unemployment and Staggered Nominal Wage Bargaining," *Journal of Money, Credit, and Banking*.

Geweke, J. (1977): "The Dynamic Factor Analysis of Economic Time Series," *in D.J. Aigner and A.S. Goldberger eds. "Latent variables in socio-economic models", Amsterdam, North-Holland Publishing, Ch. 19*.

——— (2007): "Models, Computational Expirements, and Reality," *Working Paper*.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econemetrica*, 57(2), 357 – 384.

Hamilton, J. D., and T. Waggoner, Daniel F.and Zha (2007): "Normalization in Econometrics," *Econemetric Reviews*, 26(2-4), 221 – 252.

Iacoviello, M., and S. Neri (2008): "Housing Market Spillovers: Evidence from an Estimated DSGE Model," *Working Paper*.

Ingram, B., and C. Whiteman (1994): "Supplanting the Minnesota Prior- Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors," *Journal of Monetary Economics*, 49(4), 1131–1159.

Jacquier, E., N. G. Polson, and P. E. Rossi (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business andconomic Statistics*, 12(4), 371–389.

Justiniano, A., and B. Preston (2006): "Can Structural Small Open Economy Models Account for Foreign Disturbances?," *Working Paper*.

——— (forthcoming): "Monetary Policy and Uncertainty in an Empirical Small Open Economy Model," *Journal of Applied Econometrics*.

Justiniano, A., and G. Primiceri (2008): "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98(3), 604 – 641.

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2008): "Investment Shocks and Business Cycles," *Working Paper*, (6739).

Kim, C., and C. R. Nelson (1999a): "Has the U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov-Switching Model of the Business Cycle," *Review of Economics and Statistics*, 81(4), 608 – 618.

Kim, C.-J., and C. Nelson (1999b): *State-Space Models with Regime Switching*. MIT Press.

Kim, C.-J., and C. R. Nelson (1999c): *State-Space Models with Regime Switching*. MIT Press.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic volatility: Likelihood inference and comparison with ARCH models," *Review of Economic Studies*, 6, 361–393.

Kleibergen, F., and R. Paap (2002): "Priors, posteriors and bayes factors for a Bayesian analysis of cointegration," *Journal of Econometrics*, 111(2), 223 – 249.

Koop, G., R. Leon-Gonzalez, and R. Strachan (2008): "Bayesian Inference in the Time Varying Cointegration Model," *Working Paper*.

——— (FC): "Efficient Posterior Simulation for Cointegrated Models with Priors On the Cointegration Space," *Econometric Reviews*.

Koop, G., and S. Potter (2008): "Time-varying VARs with Inequality Restrictions," *Working Paper*.

——— (FC): "Prior Elicitation in Multiple Change-point Models," *International Economic Review*.

Koop, G., and S. M. Potter (2007): "Estimation and Forecasting in Models with Multiple Breaks," *Review of Economic Studies*, 74, 763 – 789.

Koop, G., R. Strachan, H. van Dijk, and M. Villani (2004): "Bayesian Approaches to Cointegration," *Working Paper*.

Kose, M. A., C. Otrok, and P. Eswar (2008): "Global Business Cycles: Convergence or Decoupling," *Working Paper*.

Kose, M. A., C. Otrok, and C. H. Whiteman (2003): "International Business Cycles: World, Region, and Country-Specific Factors," *American Econometic Review*, 93(4), 1216–1239.

——— (2008): "Understanding the evolution of world business cycles," *Journal of International Economics*, 75(1), 110 – 130.

Krause, M. U., D. Lopez-Salido, and T. Lubik (forthcoming): "Inflation Dynamics with Search Frictions: A Structural Econometric Approach," *Journal of Monetary Economics*.

Krause, M. U., and T. A. Lubik (2007): "Aggregate Hours Adjustment in Frictional Labor Markets," *Working Paper*.

Kryshko, M., F. Schorfheide, and K. Sill (2008): "DSGE Model-based Forecasts of Non-modelled Variables," *Philadelphia Fed. Working Paper*, 08-17.

Kydland, F. E., and E. C. Prescott (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, 50(6), 1345–70.

Laforte, J.-P. (2007): "Pricing Models: A Bayesian DSGE Approach for the US Economy," *Journal of Money, Credit, and Banking*, 39(1), 127 – 154.

Levin, A., A. Onatski, J. C. Williams, and N. Williams (2005): "Monetary Policy Under Uncertainty in Micro-founded Macroeconometric Models," *NBER Macroeconomics Annual*, pp. 229 – 289.

Lubik, T. A., and F. Schorfheide (2004): "Testing for Indeterminancy: An Application to U.S. Monetary Policy," *American Economic Review*, 94(1), 190 – 217.

——— (2007): "Do Central Banks Respond to Exchange Rate Movements? A Structural Investigation," *Journal of Monetary Economics*, 54(4), 1069 – 1087.

Ludvigson, S. C., and S. Ng (2008): "A Factor Analysis of Bond Premia," *Working Paper*.

Mankiw, N. G., and R. Reis (2007): "Sticky Information in General Equilibrium," *Journal of the European Economic Association*, 5(2-3), 603 – 613.

Masanjala, W. H., and C. Papageorgiou (Forthcoming): "Rough and Lonely Road to Prosperity: A Reexamination of the Sources of Growth in Africa Using Bayesian Model Averaging," *Journal of Applied Econometrics*.

McConnell, M. M., and G. Perez-Quiros (2000): "Output Fluctuations in the United States: What Has Changed since the Early 1980's?," *American Economic Review*, 90, 1464–76.

Milani, F. (2006): "A Bayesian DSGE Model with Infinite-Horizon Learning: Do "Mechanical" Sources of Persistence Become Superfluous?," *International Journal of Central Banking*, 6.

——— (2007): "Expectations, Learning, and Macroeconomic Persistance," *Journal of Monetary Economics*, 54(7), 2065 – 2082.

——— (2008): "Learning, Monetary Policy Rules, and Macroeconomic Stability," *Journal of Economic Dynamics and Control*, 32(10), 3148 – 3165.

Monti, F., D. Giannone, and L. Reichlin (2008): "Incorporating Conjunctural Analysis into Structural Models," *Working Paper*.

Mumtaz, H., and P. Surico (2008): "Evolving International Inflation Dynamics: World and Country Specific Factors," *Working Paper*.

Ng, S., E. Moench, and S. Potter (2008): "Dynamic Hierarchical Factor Models," *Working Paper*.

Ni, S., and D. Sun (2003): "Noninformative priors and frequentist risks of bayesian estimators of vector-autoregressive models," *Journal of Econometrics*, 115(1), 159 – 197.

Otrok, C. (2001): "On Measuring the Welfare Costs of Business Cycles," *Journal of Monetary Economics*, 45(1), 61 – 92.

Otrok, C., and C. H. Whiteman (1998): "Bayesian Leading Indicators: Measuring and Predicting Economic Conditions in Iowa," *International Economic Review*, 39(4).

Paap, R., and H. K. Van Dijk (2003): "Bayes Estimates of Markov Trends in Possibly Cointegrated Series: An Application to U.S. Consumption and Income," *Journal of Business Economics and Statistics*, 21(4), 547 – 563.

Pelloni, G., and W. Polasek (2003): "Macroeconomic Effects of Sectoral Shocks in Germany, The U.K. and, The U.S.: A VAR-GARCH-M Approach," *Computational Economics*, 21(1), 65 – 85.

Phillips, P. C. B., and W. Ploberger (1996): "An Asymptotic Theory of Bayesian Inference for Time Series," *Econometrica*, 64(2), 318 – 412.

Pivetta, F., and R. Reis (2007): "The Persistence of Inflation in the United States," *Journal of Economic Dynamics and Control*, 31(4), 1326 – 1358.

Primiceri, G. E. (2005): "Time Varying VARs and Monetary Policy," *Review of Economic Studies*, 72, 821 – 852.

Quah, D., and T. J. Sargent (1992): "A Dynamic Index Model for Large Cross Sections," *Discussion Paper, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis*, 77.

Rabanal, P. (2007): "Does Inflation Increase After a Monetary Policy Tightening? Answers Based on an Estimated DSGE," *Journal of Economic Dynamics and Control*, 31(3), 903 – 937.

Rabanal, P., and J. F. Rubio-Ramirez (2005): "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach," *Journal of Monetary Economics*, 52(6), 1151 – 1166.

Reis, R. (2008a): "Optimal Monetary Policy Rules in an Estimated Sticky-Information Model," *Working Paper*.

——— (2008b): "A Sticky-Information General-Equilibrium Model for Policy Analysis," *Working Paper*.

Rubio-Ramrez, J. F., D. Waggoner, and T. Zha (2008): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Atlanta Fed Working Paper*, 2008-18.

Sala, L., U. Sderstrm, and A. Trigari (2008): "Monetary policy under uncertainty in an estimated model with labor market frictions," *Journal of Monetary Economics*, 55(5), 983– – 1006, Carnegie-Rochester Conference Series on Public Policy: Labor Markets, Macroeconomic Fluctuations, and Monetary Policy November 9-10, 2007.

Sala-i Martin, X., G. Doppelhofer, and R. I. Miller (2004): "Determinants of Long-term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*, 94(4), 813 – 835.

Sargent, T. J. (1999): *The Conquest of American Inflation*. Princeton University Press, Princeton, New Jersey.

Sargent, T. J., and C. A. Sims (1977): "Business Cycle Modeling Without Pretending To Have Too Much A Priori Economic Theory," *in Christopher A. Sims et al., "New Methods in Business Cycle Research", Minneapolis, Federal Reserve Bank of Minneapolis*.

Schorfheide, F. (2000): "Loss Function-based Evaluation of DSGE Model," *Journal of Applied Econometrics*, 15(6), 645 – 670.

——— (2005): "Learning and Monetary Policy Shifts," *Review of Economic Dynamics*, 8(2), 392 – 419.

Sims, C. (1980): "Macroeconomics and Reality," *Econometrica*, 48(4), 1–48.

Sims, C., and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), 949 – 968.

——— (1999): "Error Bands for Impulse Responses," *Econometrica*, 67(5), 1113 – 1155.

Sims, C. A. (1993): "A 9 Variable Probabilistic Macroeconomic Forecasting Model," *Business Cycles, Indicators and Forecasting, NBER Studies in Business Cycles*, 28, 179–214, J.H. Stock and M.W. Watson (eds).

——— (2001): "Comment on Cogley and Sargent's 'Evolving post World War II U.S. inflation dynamics.' ," *NBER macroeconomics Annual*, 16.

Sims, C. A., D. F. Waggoner, and T. Zha (2006): "Methods for Inference in Large Multiple-Equation Markov-Switching Models," *Atlanta Fed Working Paper*, 2006-22.

Sims, C. A., and T. Zha (2006): "Were There Regime Switches in U.S. Monetary Policy?," *American Economeic Review*, 96(1), 54 – 81.

Smets, F., and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123 – 1175.

——— (2005a): "Bayesian New Neoclassical Synthesis (NNS) Models: Modern Tools for Central Banks," *Journal of the European Economic Association*, 3(2-3), 422 – 433.

Smets, F., and R. Wouters (2005b): "Comparing Shocks and Frictions in US and Euro Area Business Cycles," *Journal of Applied Econometrics*, 20(2), 161 – 183.

Smets, F., and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97(3), 586 – 606.

Stock, J. J., and M. W. Watson (1989): "New Indices of Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual*, 4, 351–394.

——— (1999): "Forecasting Inflation," *Journal of Monetary Economics*, 44(2), 293 – 335.

——— (2001): "Vector Autoregressions," *Journal of Economic Perspectives*, 15(4), 101–115.

——— (2002): "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20(2), 147 – 162.

——— (2005): "Understanding Changes In International Business Cycle Dynamics," *Journal of the European Economic Association*, 3(5), 968 – 1006.

Strachan, R. a. (2003): "Valid Bayesian Estimation of the Cointegrating Error Correction Model," *Journal of Business and Economic Statistics*, 21(1), 185 – 195.

Strachan, R. W., and H. van Dijk (2006): "Model Uncertainty and Bayesian Model Averaging in Vector Autoregressive Processes," *Working Paper*.

Uhlig, H. (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedurestar, open," *Journal of Monetary Economics*, 52(2), 381 – 419.

Villani, M. (2001): "Fractional Bayesian Lag Length Inference in Multivariate Autoregressive Processes," *Journal of Time Series Analysis*, 22(1), 67 – 86.

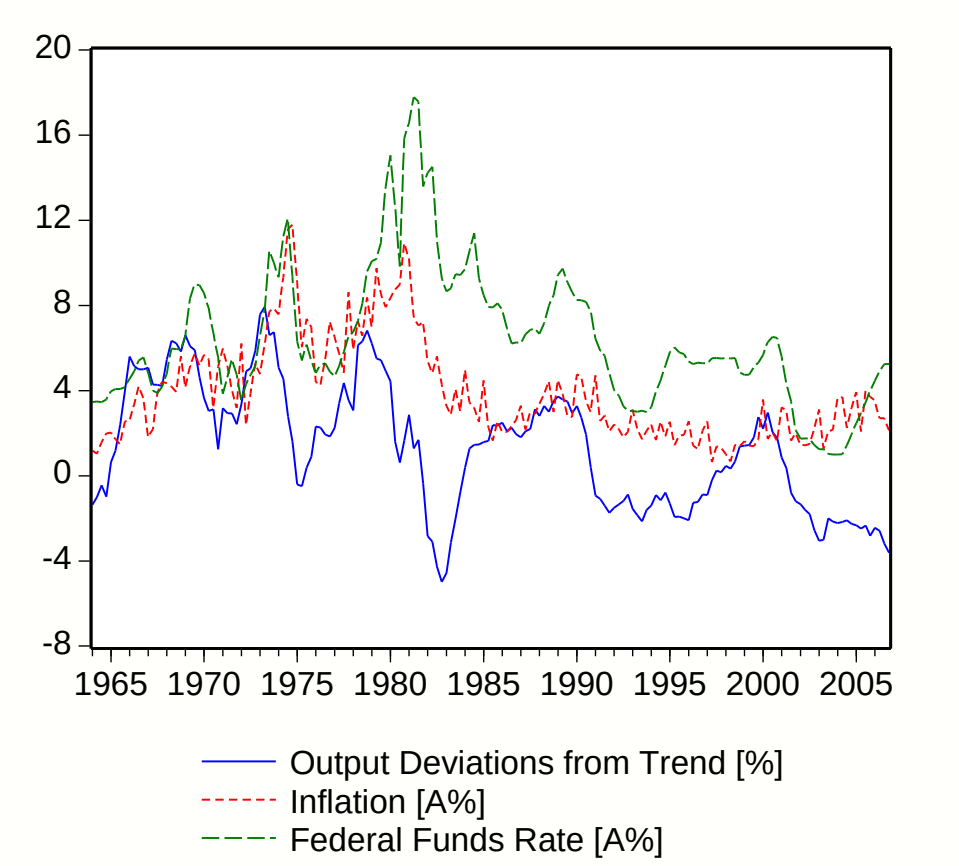
——— (2005): "Bayesian Reference Analysis of Cointegration," *Econometric Theory*, 21, 326 – 357.

——— (2008): "Steady State Priors for Vector Autoregressions," *Journal of Applied Econometrics*, forthcoming.

Waggoner, D. F., and T. Zha (1999): "Conditional Forecasts In Dynamic Multivariate Models," *Review of Economics and Statistics*, 81(4), 639 – 651.

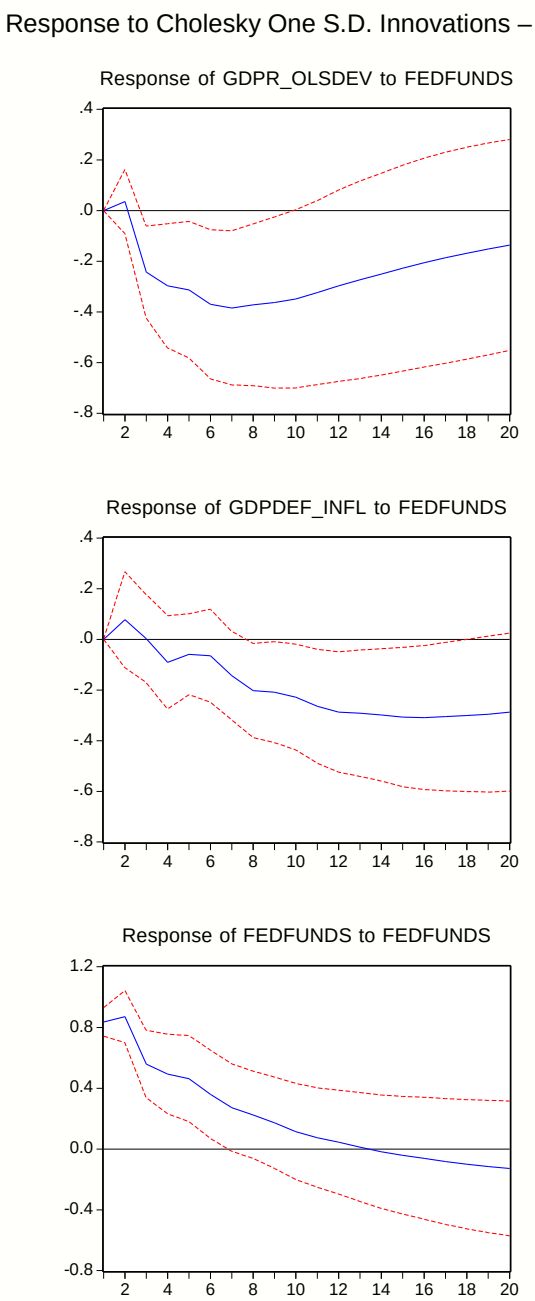
——— (2003): "A Gibb's Sampler for Structural VARs," *Journal of Economic Dynamics and Control*, 28(2), 349 – 366.

Figure 1: Output, Inflation, and Interest Rates




*Notes:* The figure depicts U.S. data from 1964:Q1 to 2006:Q4. Output is depicted in percentage deviations from a linear deterministic trend.

Figure 2: Response to a Monetary Policy Shock



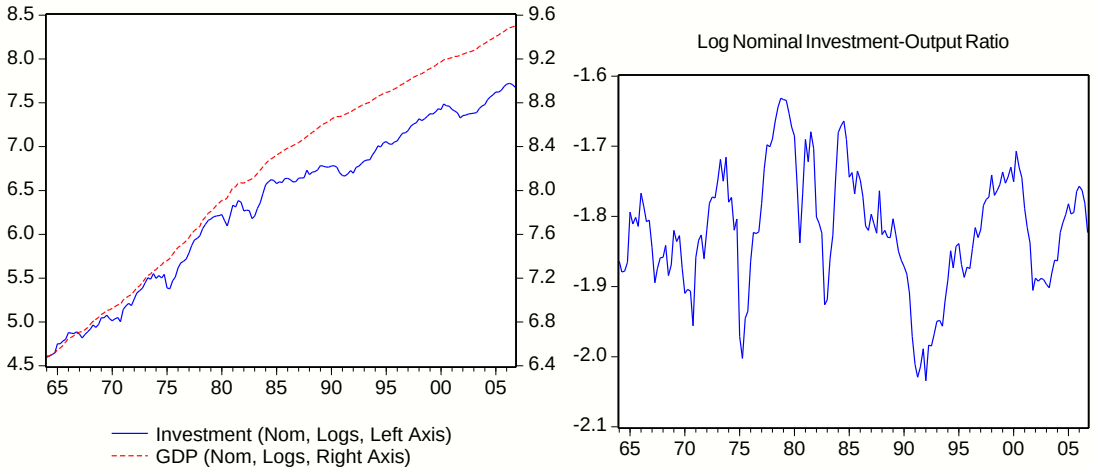

*Notes:* The figure depicts 90% credible bands and posterior mean responses for a VAR(4) to a one-standard deviation monetary policy shock.

Table 1: Identification Restrictions for $A_0$

| | Inform | MP | MD | Prod | Prod | Prod |
|---|---|---|---|---|---|---|
| Pcom | X | 0 | 0 | 0 | 0 | 0 |
| M2 | X | X | X | 0 | 0 | 0 |
| R | X | X | X | 0 | 0 | 0 |
| Y | X | 0 | X | X | X | X |
| CPI | X | 0 | X | 0 | X | X |
| U | X | 0 | 0 | 0 | 0 | X |

*Notes:* A zero entry denotes a coefficient restriction.

Figure 3: Nominal Output and Investment

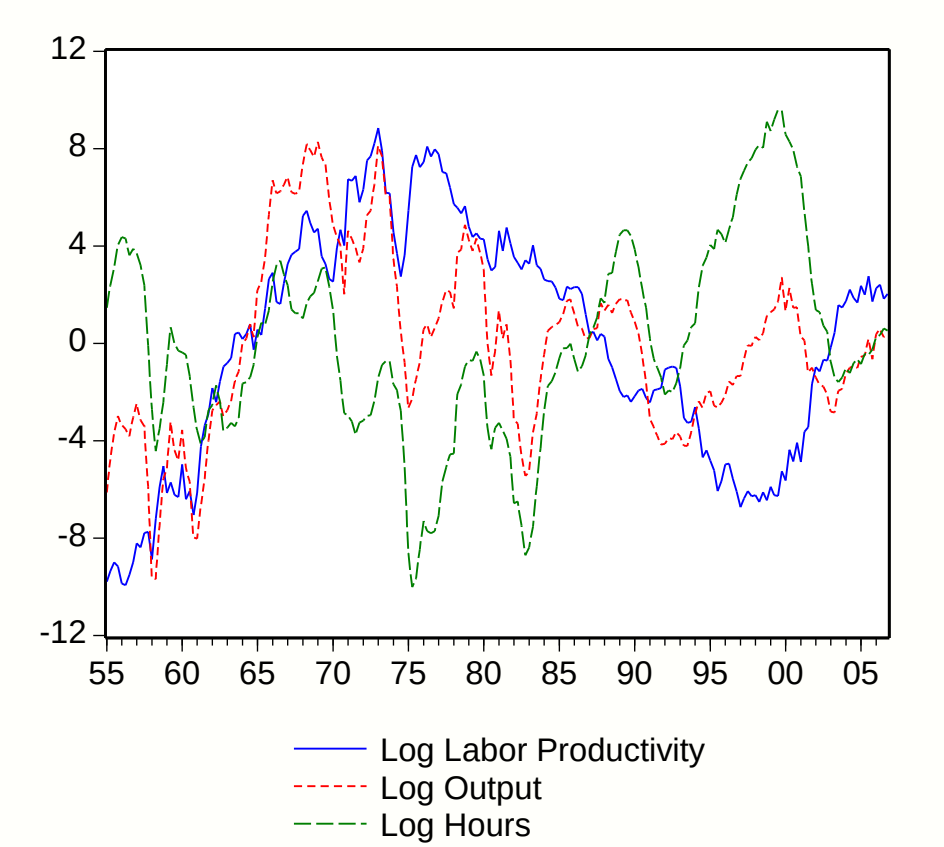



*Notes:* The figure depicts U.S. data from 1964:Q1 to 2006:Q4.

Figure 4: Prior and Posterior Cointegration Parameter

*Notes:* TBA

Figure 5: Aggregate Output, Hours, and Labor Productivity



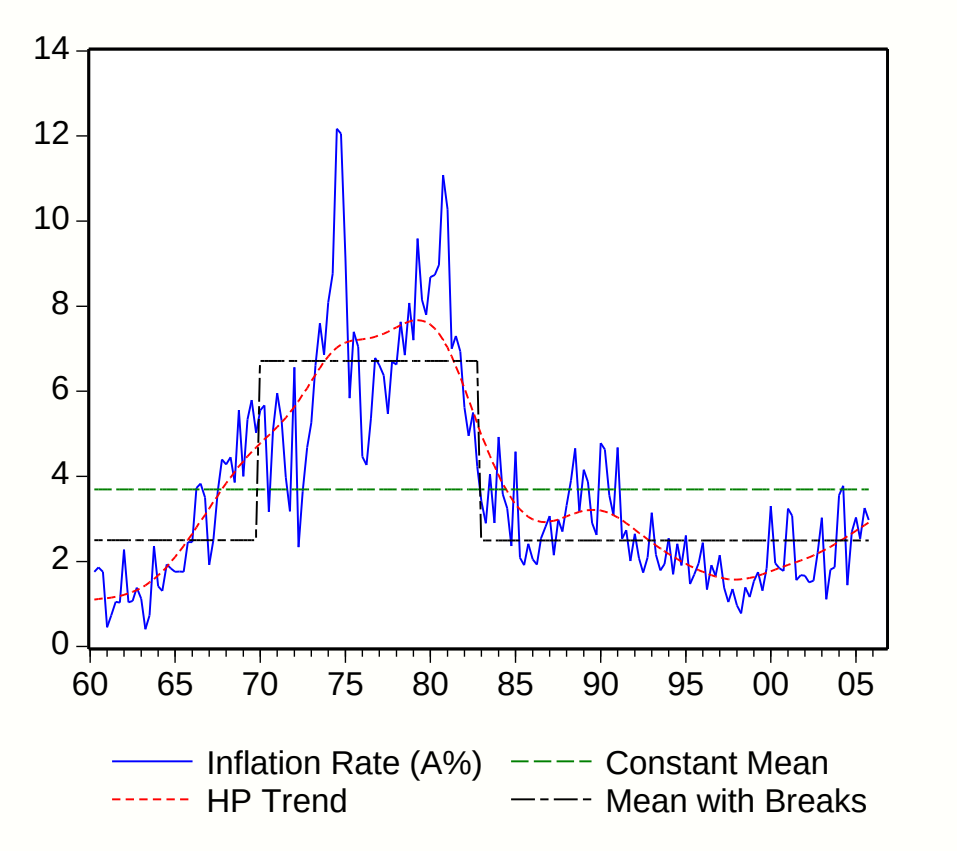

*Notes:* Output and labor productivity are depicted in percentage deviations from a deterministic trend, and hours are depicted in deviations from its mean. Sample period is 1955:Q1 to 2006:Q4.

Table 2: Prior and Posterior Distribution for DSGE Model Parameters

| | | | Prior | | Posterior | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Det. Trend | | Stoch. Trend | |
| Name | Domain | Density | Para (1) | Para (2) | Mean | 90% Intv. | Mean | 90% Intv. |
| $\alpha$ | $[0,1)$ | Beta | 0.34 | 0.02 | 0.35 | [0.31, 0.38] | 0.34 | [0.31, 0.37] |
| $\nu$ | $\mathbb{R}^+$ | Gamma | 2.00 | 1.00 | 0.44 | [0.20, 0.70] | 0.69 | [0.27, 1.08] |
| $\gamma_a$ | $\mathbb{R}$ | Normal | 0.00 | 0.10 | .002 | [.001, .003] | .002 | [.001, .003] |
| $\rho_a$ | $\mathbb{R}^+$ | Beta | 0.95 | 0.02 | 0.97 | [0.95, 0.99] | 1.00 | |
| $\sigma_a$ | $\mathbb{R}^+$ | InvGamma | 0.01 | 4.00 | .007 | [.006, .008] | .007 | [.006, .008] |
| $\rho_b$ | $\mathbb{R}^+$ | Beta | 0.80 | 0.10 | 0.97 | [0.96, 0.99] | 0.98 | [0.96, 0.99] |
| $\sigma_b$ | $\mathbb{R}^+$ | InvGamma | 0.01 | 4.00 | .011 | [.009, .012] | .012 | [.010, .014] |
| $\ln H_*$ | $\mathbb{R}$ | Normal | 0.00 | 10.0 | -0.04 | [-0.08, 0.00] | -0.03 | [-0.07, 0.02] |
| $\ln Y_0$ | $\mathbb{R}$ | Normal | 0.00 | 100 | 8.73 | [8.55, 8.92] | 8.24 | [7.80, 8.74] |

*Notes:* Para (1) and Para (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; $s$ and $\nu$ for the Inverse Gamma distribution, where $p_{\mathcal{IG}}(\sigma|\nu,s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. To estimate the stochastic growth version of the model we set $\rho_a = 1$. The parameters $\beta = 0.99$ and $\delta = 0.013$ are fixed.

Figure 6: Inflation and Measures of Trend Inflation




*Notes:* Inflation is measured as quarter-to-quarter changes in the log GDP deflator, scaled by 400 to convert it into annualized percentages. The sample ranges from 1960:Q1 to 2005:Q4.

Table 3: Expected Losses

| Decision | $\mathcal{M}_1$ | $\mathcal{M}_2$ | Risk $\mathcal{R}(\delta)$ |
|---|---|---|---|
| $\delta^*(\mathcal{M}_1) = -0.1$ | 0.99 | 0.82 | 0.92 |
| $\delta^*(\mathcal{M}_2) = -0.5$ | 1.15 | 0.50 | 0.90 |

Table 4: Nash Equilibrium as Function of Risk Sensitivity $\tau$

| $\tau$ | 0.00 | 1.00 | 10.0 | 100 |
|---|---|---|---|---|
| $q^*(\tau)$ | 1.00 | 1.10 | 1.43 | 1.60 |
| $\delta^*(\tau)$ | -0.32 | -0.30 | -0.19 | -0.12 |