Bayesian Macroeconometrics

Marco Del Negro
Federal Reserve Bank of New York

Frank Schorfheide*
University of Pennsylvania
CEPR and NBER

April 18, 2010

Prepared for
Handbook of Bayesian Econometrics

*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. Ed Herbst and Maxym Kryshko provided excellent research assistant. We are thankful for the feedback received from the editors of the Handbook John Geweke, Gary Koop, and Herman van Dijk as well as comments by Giorgio Primiceri, Dan Waggoner, and Tao Zha.

Contents

1 Introduction

One of the goals of macroeconometric analysis is to provide quantitative answers to substantive macroeconomic questions. Answers to some questions, such as whether gross domestic product (GDP) will decline over the next two quarters, can be obtained with univariate time-series models by simply exploiting serial correlations. Other questions, such as what are the main driving forces of business cycles, require at least a minimal set of restrictions, obtained from theoretical considerations, that allow the identification of structural disturbances in a multivariate time-series model. Finally, macroeconometricians might be confronted with questions demanding a sophisticated theoretical model that is able to predict how agents adjust their behavior in response to new economic policies, such as changes in monetary or fiscal policy.

1.1 Challenges for Inference and Decision Making

Unfortunately, macroeconometricians often face a shortage of observations necessary for providing precise answers. Some questions require high-dimensional empirical models. For instance, the analysis of 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 highly parameterized multicountry vector autoregressive models. High-dimensional models are also necessary in applications in which it is reasonable to believe that parameters evolve over time, for instance, because of changes in economic policies. Thus, sample information alone is often insufficient to enable sharp inference about model parameters and implications. Other questions do not necessarily require a very densely parameterized empirical model, but they do demand identification restrictions that are not self-evident and that are highly contested in the empirical literature. For instance, an unambiguous measurement of the quantitative response of output and inflation to an unexpected reduction in the federal funds rate remains elusive. Thus, documenting the uncertainty associated with empirical findings or predictions is of first-order importance for scientific reporting.

Many macroeconomists have a strong preference for models with a high degree of theoretical coherence such as dynamic stochastic general equilibrium (DSGE) models. In these models, decision rules of economic agents are derived from assumptions

about agents' preferences and production technologies and some fundamental principles such as intertemporal optimization, rational expectations, and competitive equilibrium. In practice, this means that the functional forms and parameters of equations that describe the behavior of economic agents are tightly restricted by optimality and equilibrium conditions. Thus, likelihood functions for empirical models with a strong degree of theoretical coherence tend to be more restrictive than likelihood functions associated with atheoretical models. A challenge arises if the data favor the atheoretical model and the atheoretical model generates more accurate forecasts, but a theoretically coherent model is required for the analysis of a particular economic policy.

1.2 How Can Bayesian Analysis Help?

In Bayesian inference, a prior distribution is updated by sample information contained in the likelihood function to form a posterior distribution. Thus, to the extent that the prior is based on *nonsample* information, it provides the ideal framework for combining different sources of information and thereby sharpening inference in macroeconometric analysis. This combination of information sets is prominently used in the context of DSGE model inference in Section 4. Through informative prior distributions, Bayesian DSGE model inference can draw from a wide range of data sources that are (at least approximately) independent of the sample information. These sources might include microeconometric panel studies that are informative about aggregate elasticities or long-run averages of macroeconomic variables that are not included in the likelihood function because the DSGE model under consideration is too stylized to be able to explain their cyclical fluctuations.

Many macroeconometric models are richly parameterized. Examples include the vector autoregressions (VARs) with time-varying coefficients in Section 5 and the multicountry VARs considered in Section 6. In any sample of realistic size, there will be a shortage of information for determining the model coefficients, leading to very imprecise inference and diffuse predictive distributions. In the context of time-varying coefficient models, it is often appealing to conduct inference under the assumption that either coefficient change only infrequently, but by a potentially large amount, or that they change frequently, but only gradually. Such assumptions can be conveniently imposed by treating the sequence of model parameters as a

stochastic process, which is of course nothing but a prior distribution that can be updated with the likelihood function.

To reduce the number of parameters in a high-dimensional VAR, one could of course set many coefficients equal to zero or impose the condition that the same coefficient interacts with multiple regressors. Unfortunately, such *hard* restrictions rule out the existence of certain spillover effects, which might be undesirable. Conceptually more appealing is the use of *soft* restrictions, which can be easily incorporated through probability distributions for those coefficients that are "centered" at the desired restrictions but that have a small, yet nonzero, variance. An important and empirically successful example of such a prior is the Minnesota prior discussed in Section 2.

An extreme version of lack of sample information arises in the context of structural VARs, which are studied in Section 2. Structural VARs can be parameterized in terms of reduced-form parameters, which enter the likelihood function, and an orthogonal matrix Ω, which does not enter the likelihood function. Thus, Ω is not identifiable based on the sample information. In this case, the conditional distribution of Ω given the reduced-form parameters will not be updated, and its conditional posterior is identical to the conditional prior. Identification issues also arise in the context of DSGE models. In general, as long as the joint prior distribution of reduced-form and nonidentifiable parameters is proper, meaning that the total probability mass is one, so is the joint posterior distribution. In this sense, the lack of identification poses no conceptual problem in a Bayesian framework. However, it does pose a challenge: it becomes more important to document which aspects of the prior distribution are not updated by the likelihood function and to recognize the extreme sensitivity of those aspects to the specification of the prior distribution.

Predictive distributions of future observations such as aggregate output, inflation, and interest rates are important for macroeconomic forecasts and policy decisions. These distributions need 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 simultaneously is conceptually straightforward. To the extent that the substantive analysis requires a researcher to consider multiple theoretical and empirical frameworks, Bayesian analysis allows the researcher to assign probabilities to competing model specifications

and update these probabilities in view of the data. Throughout this chapter, we will encounter a large number of variants of VARs (sections 2 and 3) and DSGE models (section 4) that potentially differ in their economic implications. With posterior model probabilities in hand, inference and decisions can be based on model averages (section 7).

Predictions of how economic agents would behave under counterfactual economic policies never previously observed require empirical models with a large degree of theoretical coherence. The DSGE models discussed in Section 4 provide an example. As mentioned earlier, in practice posterior model probabilities often favor more flexible, nonstructural time-series models such as VARs. Nonetheless, Bayesian methods offer a rich tool kit for linking structural econometric models to more densely parameterized reference models. For instance, one could use the restrictions associated with the theoretically coherent DSGE model only loosely, to center a prior distribution on a more flexible reference model. This idea is explored in more detail in Section 4.

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. 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 unanticipated by the public. After discussing various identification schemes and their implementation, we devote the remainder of Section 2 is devoted to a discussion of advanced topics such as inference in restricted or overidentified VARs. As an empirical illustration, we measure the effects of an unanticipated change in monetary policy using a four-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

models, these stochastic trends are often common to several time series. For example, in many countries the ratio (or log difference) of aggregate consumption and investment is stationary. This observation is consistent with a widely used version of the neoclassical growth model (King, Plosser, and Rebelo (1988)), in which the exogenous technology process follows a random walk. One 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) popularized the concept of cointegration. While frequentist analysis of nonstationary 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, as pointed out by Sims and Uhlig (1991). Nonetheless, the Bayesian literature has experienced a lively debate about how to best analyze VECMs. Most of the controversies are related to the specification of prior distributions. We will focus on the use of informative priors in the context of an empirical model for U.S. output and investment data. Our prior is based on the balanced-growth-path implications of a neoclassical growth model. However, we also discuss an important strand of the literature that, instead of using priors as a tool to incorporate additional information, uses them to regularize or smooth the likelihood function of a cointegration model in areas of the parameter space in which it is very nonelliptical.

Modern dynamic macroeconomic theory implies fairly tight cross-equation restrictions for vector autoregressive processes, and in Section 4 we turn to Bayesian inference with DSGE models. The term *DSGE model* is typically used to refer to a broad class 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 the solution of intertemporal optimization problems determines the decision rules, given the specification of preferences and technology. Moreover, agents potentially face uncertainty with respect to total factor productivity, for instance, 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 the specified distribution of 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 empirical work with DSGE models employs Bayesian methods. Section 4 discusses inference with linearized as well as nonlinear DSGE models and reviews various approaches for evaluating the empirical fit of DSGE models. As an illustration, we conduct inference with a simple stochastic growth model based on U.S. output 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 parameters (TVP), discussed in Section 5. 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 nonstationary autoregressive law of motion and models in which parameters evolve according to a finite-state Markov-switching (MS) 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.

Because of the rapid advances in information technologies, macroeconomists now have 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 modeling is to avoid a 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 dynamic factor models (DFMs). 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, such as 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, or 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, a word on notation. We use $Y_{t_0:t_1}$ to denote the sequence of observations or random variables $\{y_{t_0}, \ldots, y_{t_1}\}$. If no ambiguity arises, we sometimes drop the time subscripts and abbreviate $Y_{1:T}$ by Y. θ often 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 follow the Appendix of this Handbook. We use *iid* to abbreviate independently and identically distributed. If $X|\Sigma \sim MN_{p\times q}(M, \Sigma \otimes P)$ is matricvariate Normal and $\Sigma \sim IW_q(S, \nu)$ has an Inverted Wishart distribution, we say that $(X, \Sigma) \sim MNIW(M, P, S, \nu)$. Here $\otimes$ is the Kronecker product. We use I to denote the identity matrix and use a subscript indicating the dimension if necessary. $tr[A]$ is the trace of the square matrix A, $|A|$ is its determinant, and $vec(A)$ stacks the columns of A. Moreover, we let $\|A\| = \sqrt{tr[A'A]}$. If A is a vector, then $\|A\| = \sqrt{A'A}$ is its length. We use $A_{(.j)}$ ($A_{(j.)}$) to denote the j'th column (row) of a matrix A. Finally, $\mathcal{I}\{x \geq a\}$ is the indicator function equal to one if $x \geq a$ and equal to zero otherwise.

2 Vector Autoregressions

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 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 discusses how to use dummy observations to construct prior distributions and reviews the widely

used Minnesota prior. In Section 2.3, we consider a reduced-form VAR that is expressed in terms of deviations from a deterministic trend. Section 2.4 is devoted to structural VARs in which innovations are expressed as functions of structural shocks with a particular economic interpretation, for example, an unanticipated change in monetary policy. Finally, Section 2.5 provides some suggestions for further reading.

INSERT FIGURE 1 HERE

2.1 A Reduced-Form VAR

Vector autoregressions are linear time-series models, designed to capture the joint dynamics of multiple time series. 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 real GDP from a linear time trend, annualized inflation rates computed from the GDP deflator, and the effective federal funds rate. These series are obtained from the FRED database maintained by the Federal Reserve Bank of St. Louis. We will subsequently illustrate the VAR analysis using the three series plotted in Figure 1. Let y_t be an $n \times 1$ 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 the reduced-form representation of a VAR(p), because the u_t's are simply one-step-ahead forecast errors and do not have a specific economic interpretation.

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 the conditional distribution of y_t is Normal:

$$u_t \sim iidN(0, \Sigma). \tag{2}$$

We are now in a position to characterize the joint distribution of a sequence of observations $y_1, \ldots, y_T$. 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 Φ and

Σ, is called (conditional) likelihood function when it is viewed as function of the parameters. It can be factorized as

$$p(Y_{1:T}|\Phi,\Sigma,Y_{1-p:0}) = \prod_{t=1}^{T} p(y_t|\Phi,\Sigma,Y_{1-p:t-1}). \tag{3}$$

The conditional likelihood function can be conveniently expressed if the VAR is written as a multivariate linear regression model in matrix notation:

$$Y = X\Phi + U. \tag{4}$$

Here, the $T \times n$ matrices Y and U and the $T \times k$ matrix X are defined as

$$Y = \begin{bmatrix} y_1' \\ \vdots \\ y_T' \end{bmatrix}, \quad X = \begin{bmatrix} x_1' \\ \vdots \\ x_T' \end{bmatrix}, \quad x_t' = [y_{t-1}', \ldots, y_{t-p}', 1], \quad U = \begin{bmatrix} u_1' \\ \vdots \\ u_T' \end{bmatrix}. \tag{5}$$

In a slight abuse of notation, we abbreviate $p(Y_{1:T}|\Phi,\Sigma,Y_{1-p:0})$ by $p(Y|\Phi,\Sigma)$:

$$\begin{aligned} p(Y|\Phi,\Sigma) \quad \propto \quad & |\Sigma|^{-T/2} \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}\hat{S}]\right\} \\ & \times \exp\left\{-\frac{1}{2} tr[\Sigma^{-1}(\Phi-\hat{\Phi})'X'X(\Phi-\hat{\Phi})]\right\}, \end{aligned} \tag{6}$$

where

$$\hat{\Phi} = (X'X)^{-1}X'Y, \quad \hat{S} = (Y - X\hat{\Phi})'(Y - X\hat{\Phi}). \tag{7}$$

$\hat{\Phi}$ is the maximum-likelihood estimator (MLE) of Φ, 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 the posterior distribution is of the form

$$(\Phi,\Sigma)|Y \sim MNIW\left(\hat{\Phi}, (X'X)^{-1}, \hat{S}, T-k\right). \tag{8}$$

Detailed derivations for the multivariate Gaussian linear regression model can be found in Zellner (1971). Draws from this posterior can be easily obtained by direct Monte Carlo sampling.

Algorithm 2.1: Direct Monte Carlo Sampling from Posterior of VAR Parameters

For $s = 1, \ldots, n_{sim}$:

1. Draw $\Sigma^{(s)}$ from an $IW(\hat{S}, T-k)$ distribution.

2. Draw $\Phi^{(s)}$ from the conditional distribution $MN(\hat{\Phi}, \Sigma^{(s)} \otimes (X'X)^{-1})$. $\square$

An important challenge in practice is to cope with the dimensionality of the parameter matrix Φ. 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. If the sample is restricted to the post-1982 period, after the disinflation under Fed Chairman Paul Volcker, the sample size shrinks to 96 observations. Now imagine estimating a two-country VAR for the U.S. and the Euro Area on post-1982 data, which doubles the number of parameters. Informative prior distributions can compensate for lack of sample information, and we will subsequently discuss alternatives to the improper prior used so far.

2.2 Dummy Observations and the Minnesota Prior

Prior distributions can be conveniently represented by dummy observations. This insight dates back at least to Theil and Goldberger (1961). 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 Φ and Σ. Using the same arguments that lead to (8), we deduce that up to a constant the product $p(Y^*|\Phi, \Sigma) \cdot |\Sigma|^{-(n+1)/2}$ can be interpreted as a $MNIW(\underline{\Phi}, (X^{*'}X^*)^{-1}, \underline{S}, T^* - k)$ prior for Φ and Σ, where $\underline{\Phi}$ and $\underline{S}$ are obtained from $\hat{\Phi}$ and $\hat{S}$ in (7) by replacing Y and X with Y^* and X^*. Provided that $T^* > k+n$ and $X^{*'}X^*$ is invertible, the prior distribution is proper. Now let $\bar{T} = T + T^*$, $\bar{Y} = [Y^{*'}, Y']'$, $\bar{X} = [X^{*'}, X']'$, and let $\bar{\Phi}$ and $\bar{S}$ be the analogue of $\hat{\Phi}$ and $\hat{S}$ in (7); then we deduce that the posterior of (Φ, Σ) 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. Prior and likelihood are conjugate if the posterior belongs to the same distributional family as the prior distribution.

A widely used prior in the VAR literature is the so-called Minnesota prior, which dates back to Litterman (1980) 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 on 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, possibly 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 Φ at a value that implies a random-walk behavior for each of the components of y_t. The random-walk approximation is taken for convenience and could be replaced by other representations. For instance, if some series have very little serial correlation because they have been transformed to induce stationarity – for example log output has been converted into output growth – then an *iid* approximation might be preferable. In Section 4, we will discuss how DSGE model restrictions could be used to construct a prior.

The Minnesota prior can be implemented either by directly specifying a distribution for Φ 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 Φ, it tends to be difficult to elicit beliefs about the correlation between elements of the Φ 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 is typically specified conditional on several hyperparameters. Let $Y_{-\tau:0}$ be a presample, and let $\underline{y}$ and $\underline{s}$ be $n \times 1$ vectors of means and standard deviations. The remaining hyperparameters are stacked in the 5×1 vector λ with elements λ_i. In turn, we will specify the rows of the matrices Y^* and X^*. To simplify the exposition, suppose that $n = 2$ and $p = 2$. The dummy observations are interpreted as observations from the regression model (4). We begin with dummy observations that generate a prior distribution for Φ_1. For illustrative purposes, the dummy observations are plugged into (4):

$$\begin{bmatrix} \lambda_1\underline{s}_1 & 0 \\ 0 & \lambda_1\underline{s}_2 \end{bmatrix} = \begin{bmatrix} \lambda_1\underline{s}_1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1\underline{s}_2 & 0 & 0 & 0 \end{bmatrix} \Phi + \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}. \tag{9}$$

According to the distributional assumption in (2), the rows of U are normally distributed. Thus, we can rewrite the first row of (9) as

$$\lambda_1\underline{s}_1 = \lambda_1\underline{s}_1\phi_{11} + u_{11}, \quad 0 = \lambda_1\underline{s}_1\phi_{21} + u_{12}$$

and interpret it as

$$\phi_{11} \sim \mathcal{N}(1, \Sigma_{11}/(\lambda_1^2 \underline{s}_1^2)), \quad \phi_{21} \sim \mathcal{N}(0, \Sigma_{22}/(\lambda_1^2 \underline{s}_1^2)).$$

ϕ_{ij} denotes the element i, j of the matrix Φ, and Σ_{ij} corresponds to element i, j of Σ. The hyperparameter λ_1 controls the tightness of the prior.[1]

The prior for Φ_2 is implemented with the dummy observations

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \lambda_1 \underline{s}_1 2^{\lambda_2} & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 \underline{s}_2 2^{\lambda_2} & 0 \end{bmatrix} \Phi + U, \tag{10}$$

where the hyperparameter λ_2 is used to scale the prior standard deviations for coefficients associated with y_{t-l} according to $l^{-\lambda_2}$. A prior for the covariance matrix Σ, centered at a matrix that is diagonal with elements equal to the presample variance of y_t, can be obtained by stacking the observations

$$\begin{bmatrix} \underline{s}_1 & 0 \\ 0 & \underline{s}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Phi + U \tag{11}$$

λ_3 times.

The remaining sets of dummy observations provide a prior for the intercept Φ_c 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 lagged values of a variable $y_{i,t}$ are at the level $\underline{y}_i$, the same value $\underline{y}_i$ is likely to be a good forecast of $y_{i,t}$, regardless of the value of other variables:

$$\begin{bmatrix} \lambda_4 \underline{y}_1 & 0 \\ 0 & \lambda_4 \underline{y}_2 \end{bmatrix} = \begin{bmatrix} \lambda_4 \underline{y}_1 & 0 & \lambda_4 \underline{y}_1 & 0 & 0 \\ 0 & \lambda_4 \underline{y}_2 & 0 & \lambda_4 \underline{y}_2 & 0 \end{bmatrix} \Phi + U. \tag{12}$$

The *co-persistence dummy observations*, proposed by Sims (1993) reflect the belief that when all lagged y_t's are at the level $\underline{y}$, y_t tends to persist at that level:

$$\begin{bmatrix} \lambda_5 \underline{y}_1 & \lambda_5 \underline{y}_2 \end{bmatrix} = \begin{bmatrix} \lambda_5 \underline{y}_1 & \lambda_5 \underline{y}_2 & \lambda_5 \underline{y}_1 & \lambda_5 \underline{y}_2 & \lambda_5 \end{bmatrix} \Phi + U. \tag{13}$$

[1] Consider the regression $y_t = \phi_1 x_{1,t} + \phi_2 x_{2,t} + u_t$, $u_t \sim iidN(0,1)$, and suppose that the standard deviation of $x_{j,t}$ is s_j. If we define $\tilde{\phi}_j = \phi_j s_j$ 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 $\tilde{\phi}_j \sim \mathcal{N}(0, \lambda^2)$, then $\phi_j \sim \mathcal{N}(0, \lambda^2/s_j^2)$. The s_j terms that appear in the definition of the dummy observations achieve this scale adjustment.

The strength of these beliefs is controlled by λ_4 and λ_5. These two sets of dummy observations introduce correlations in prior beliefs about all coefficients, including the intercept, in a given equation.

VAR inference tends to be sensitive to the choice of hyperparameters. If $\lambda = 0$, then all the dummy observations are zero, and the VAR is estimated under an improper prior. The larger the elements of λ, the more weight is placed on various components of the Minnesota prior vis-á-vis the likelihood function. From a practitioner's view, an empirical Bayes approach of choosing λ based on the marginal likelihood function

$$p_\lambda(Y) = \int p(Y|\Phi,\Sigma)p(\Phi,\Sigma|\lambda)d(\Phi,\Sigma) \tag{14}$$

tends to work well for inference as well as for forecasting purposes. If the prior distribution is constructed based on T^* dummy observations, then an analytical expression for the marginal likelihood can be obtained by using the normalization constants for the MNIW distribution (see Zellner (1971)):

$$p_\lambda(Y) = (2\pi)^{-nT/2}\frac{|\bar{X}'\bar{X}|^{-\frac{n}{2}}|\bar{S}|^{-\frac{\bar{T}-k}{2}}}{|X^{*\prime}X^*|^{-\frac{n}{2}}|S^*|^{-\frac{T^*-k}{2}}}\frac{2^{\frac{n(\bar{T}-k)}{2}}\prod_{i=1}^n\Gamma[(\bar{T}-k+1-i)/2]}{2^{\frac{n(T^*-k)}{2}}\prod_{i=1}^n\Gamma[(T^*-k+1-i)/2]}. \tag{15}$$

As before, we let $\bar{T} = T^* + T$, $\bar{Y} = [Y^{*\prime}, Y']'$, and $\bar{X} = [X^{*\prime}, X']'$. The hyperparameters $(\bar{y}, \bar{s}, \lambda)$ enter through the dummy observations X^* and Y^*. S^* ($\bar{S}$) is obtained from $\hat{S}$ in (7) by replacing Y and X with Y^* and X^* ($\bar{Y}$ and $\bar{X}$). We will provide an empirical illustration of this hyperparameter selection approach in Section 2.4. Instead of conditioning on the value of λ that maximizes the marginal likelihood function $p_\lambda(Y)$, one could specify a prior distribution for λ and integrate out the hyperparameter, which is commonly done in hierarchical Bayes models. A more detailed discussion of selection versus averaging is provided in Section 7.

A potential drawback of the dummy-observation prior is that one is forced to treat all equations symmetrically when specifying a prior. In other words, the prior covariance matrix for the coefficients in all equations has to be proportional to $(X^{*\prime}X^*)^{-1}$. For instance, if the prior variance for the lagged inflation terms in the output equation is 10 times larger than the prior variance for the coefficients on lagged interest rate terms, then it also has to be 10 times larger in the inflation equation and the interest rate equation. Methods for relaxing this restriction and alternative approaches of implementing the Minnesota prior (as well as other VAR priors) are discussed in Kadiyala and Karlsson (1997).

2.3 A Second Reduced-Form VAR

The reduced-form VAR in (1) is specified with an intercept term that determines the unconditional mean of y_t if the VAR is stationary. However, this unconditional mean also depends on the autoregressive coefficients $\Phi_1, \ldots, \Phi_p$. Alternatively, one can use the following representation, studied, for instance, in Villani (2009):

$$y_t = \Gamma_0 + \Gamma_1 t + \widetilde{y}_t, \quad \widetilde{y}_t = \Phi_1 \widetilde{y}_{t-1} + \ldots + \Phi_p \widetilde{y}_{t-p} + u_t, \quad u_t \sim iidN(0, \Sigma). \tag{16}$$

Here Γ_0 and Γ_1 are $n \times 1$ vectors. 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 nonstationary. This alternative specification makes it straightforward to separate beliefs about the deterministic trend component from beliefs about the persistence of fluctuations around this trend.

Suppose we define $\Phi = [\Phi_1, \ldots, \Phi_p]'$ and $\Gamma = [\Gamma_1', \Gamma_2']'$. Moreover, let $\widetilde{Y}(\Gamma)$ be the $T \times n$ matrix with rows $(y_t - \Gamma_0 - \Gamma_1 t)'$ and $\widetilde{X}(\Gamma)$ be the $T \times (pn)$ matrix with rows $[(y_{t-1} - \Gamma_0 - \Gamma_1(t-1))', \ldots, (y_{t-p} - \Gamma_0 - \Gamma_1(t-p))']$; then the conditional likelihood function associated with (16) is

$$\begin{aligned} & p(Y_{1:T}|\Phi, \Sigma, \Gamma, Y_{1-p:0}) \\ & \quad \propto \quad |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} tr\left[\Sigma^{-1} (\widetilde{Y}(\Gamma) - \widetilde{X}(\Gamma)\Phi)' (\widetilde{Y}(\Gamma) - \widetilde{X}(\Gamma)\Phi) \right] \right\}. \end{aligned} \tag{17}$$

Thus, as long as the prior for Φ and Σ conditional on Γ is $MNIW$, the posterior of $(\Phi, \Sigma)|\Gamma$ is of the $MNIW$ form.

Let L denote the temporal lag operator such that $L^j y_t = y_{t-j}$. Using this operator, one can rewrite (16) as

$$\left(I - \sum_{j=1}^{p} \Phi_j L^j \right) (y_t - \Gamma_0 - \Gamma_1 t) = u_t.$$

Now define

$$z_t(\Phi) = \left(I - \sum_{j=1}^{p} \Phi_j L^j \right) y_t, \quad W_t(\Phi) = \left[\left(I - \sum_{j=1}^{p} \Phi_j \right), \left(I - \sum_{j=1}^{p} \Phi_j L^j \right) t \right]$$

with the understanding that $L^j t = t - j$. Thus, $z_t(\Phi) = W_t(\Phi)\Gamma + u_t$ and the likelihood function can be rewritten as

$$\begin{aligned} & p(Y_{1:T}|\Phi, \Sigma, \Gamma, Y_{1-p:0}) \\ & \quad \propto \quad \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} (z_t(\Phi) - W_t(\Phi)\Gamma)' \Sigma^{-1} (z_t(\Phi) - W_t(\Phi)\Gamma) \right\}. \end{aligned} \tag{18}$$

Thus, it is straightforward to verify that as long as the prior distribution of Γ conditional on Φ and Σ is matricvariate Normal, the (conditional) posterior distribution of Γ is also Normal. Posterior inference can then be implemented via Gibbs sampling, which is an example of a so-called *Markov chain Monte Carlo* (MCMC) algorithm discussed in detail in Chib (This Volume):

Algorithm 2.2: Gibbs Sampling from Posterior of VAR Parameters

For $s = 1, \ldots, n_{sim}$:

1. Draw $(\Phi^{(s)}, \Sigma^{(s)})$ from the MNIW distribution of $(\Phi, \Sigma)|(\Gamma^{(s-1)}, Y)$.
2. Draw $\Gamma^{(s)}$ from the Normal distribution of $\Gamma|(\Phi^{(s)}, \Sigma^{(s)}, Y)$. $\square$

To illustrate the subtle difference between the VAR in (1) and the VAR in (16), we consider the special case of two univariate AR(1) processes:

$$y_t = \phi_1 y_{t-1} + \phi_c + u_t, \quad u_t \sim iidN(0,1), \tag{19}$$

$$y_t = \gamma_0 + \gamma_1 t + \widetilde{y}_t, \quad \widetilde{y}_t = \phi_1 \widetilde{y}_{t-1} + u_t, \quad u_t \sim iidN(0,1). \tag{20}$$

If $|\phi_1| < 1$ both AR(1) processes are stationary. The second process, characterized by (20), allows for stationary fluctuations around a linear time trend, whereas the first allows only for fluctuations around a constant mean. If $\phi_1 = 1$, the interpretation of ϕ_c in model (19) changes drastically, as the parameter is now capturing the drift in a unit-root process instead of determining the long-run mean of y_t. Schotman and van Dijk (1991) make the case that the representation (20) is more appealing, if the goal of the empirical analysis is to determine the evidence in favor of the hypothesis that $\phi_1 = 1$.[2] Since the initial level of the latent process $\tilde{y}_0$ is unobserved, γ_0 in (20) is nonidentifiable if $\phi_1 = 1$. Thus, in practice it is advisable to specify a proper prior for γ_0 in (20).

In empirical work researchers often treat parameters as independent and might combine (19) with a prior distribution that implies $\phi_1 \sim U[0, 1-\xi]$ and $\phi_c \sim N(\underline{\phi}_c, \lambda^2)$. For the subsequent argument, it is assumed that $\xi > 0$ to impose stationarity. Since the expected value of $I\!E[y_t] = \phi_c/(1-\phi_1)$, this prior for ϕ_1 and ϕ_c has the following implication. Conditional on ϕ_c, the prior mean and variance of the population mean $I\!E[y_t]$ increases (in absolute value) as $\phi_1 \longrightarrow 1-\xi$. In turn,

[2] Giordani, Pitt, and Kohn (This Volume) discuss evidence that in many instances the so-called centered parameterization of (20) can increase the efficiency of MCMC algorithms.

this prior generates a fairly diffuse distribution of y_t that might place little mass on values of y_t that appear *a priori* plausible.

Treating the parameters of Model (20) as independent – for example, $\phi_1 \sim U[0, 1-\xi]$, $\gamma_0 \sim N(\underline{\gamma}_0, \lambda^2)$, and $\gamma_1 = 0$ – avoids the problem of an overly diffuse data distribution. In this case $\mathbb{E}[y_t]$ has *a priori* mean $\underline{\gamma}_0$ and variance λ^2 for every value of ϕ_1. For researchers who do prefer to work with Model (19) but are concerned about *a priori* implausible data distributions, the co-persistence dummy observations discussed in Section 2.2 are useful. With these dummy observations, the implied prior distribution of the population mean of y_t conditional on ϕ_1 takes the form $\mathbb{E}[y_t]|\phi_1 \sim N(\underline{y}, (\lambda_5(1-\phi_1))^{-2})$. While the scale of the distribution of $\mathbb{E}[y_t]$ is still dependent on the autoregressive coefficient, at least the location remains centered at $\underline{y}$ regardless of ϕ_1.

2.4 Structural VARs

Reduced-form VARs summarize the autocovariance properties of the data and provide a useful forecasting tool, but they lack economic interpretability. We will consider two ways of adding economic content to the VAR specified in (1). First, one can turn (1) into a dynamic simultaneous equations model by premultiplying it with a matrix A_0, such that the equations could be interpreted as, for instance, monetary policy rule, money demand equation, aggregate supply equation, and aggregate demand equation. Shocks to these equations can in turn be interpreted as monetary policy shocks or as innovations to aggregate supply and demand. To the extent that the monetary policy rule captures the central bank's systematic reaction to the state of the economy, it is natural to assume that the monetary policy shocks are orthogonal to the other innovations. More generally, researchers often assume that shocks to the aggregate supply and demand equations are independent of each other.

A second way of adding economic content to VARs exploits the close connection between VARs and modern dynamic stochastic general equilibrium models. In the context of a DSGE model, a monetary policy rule might be well defined, but the notion of an aggregate demand or supply function is obscure. As we will see in Section 4, these models are specified in terms of preferences of economic agents and production technologies. The optimal solution of agents' decision problems combined with an equilibrium concept leads to an autoregressive law of motion for

the endogenous model variables. Economic fluctuations are generated by shocks to technology, preferences, monetary policy, or fiscal policy. These shocks are typically assumed to be independent of each other. One reason for this independence assumption is that many researchers view the purpose of DSGE models as that of generating the observed comovements between macroeconomic variables through well-specified economic propagation mechanisms, rather than from correlated exogenous shocks. Thus, these kinds of dynamic macroeconomic theories suggest that the one-step-ahead forecast errors u_t in (1) are functions of orthogonal fundamental innovations in technology, preferences, or policies.

To summarize, one can think of a structural VAR either as a dynamic simultaneous equations model, in which each equation has a particular structural interpretation, or as an autoregressive model, in which the forecast errors are explicitly linked to such fundamental innovations. We adopt the latter view in Section 2.4.1 and consider the former interpretation in Section 2.4.2.

2.4.1 Reduced-Form Innovations and Structural Shocks

A straightforward calculation shows that we need to impose additional restrictions to identify a structural VAR. Let ϵ_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

$$u_t = \Phi_\epsilon \epsilon_t = \Sigma_{tr} \Omega \epsilon_t. \tag{21}$$

Here, Σ_{tr} refers to the unique lower-triangular Cholesky factor of Σ with nonnegative diagonal elements, and Ω is an $n \times n$ orthogonal matrix. The second equality ensures that the covariance matrix of u_t is preserved; that is, Φ_ϵ has to satisfy the restriction $\Sigma = \Phi_\epsilon \Phi_\epsilon'$. Thus, our structural VAR is parameterized in terms of the reduced-form parameters Φ and Σ (or its Cholesky factor Σ_{tr}) and the orthogonal matrix Ω. The joint distribution of data and parameters is given by

$$p(Y, \Phi, \Sigma, \Omega) = p(Y|\Phi, \Sigma) p(\Phi, \Sigma) p(\Omega|\Phi, \Sigma). \tag{22}$$

Since the distribution of Y depends only on the covariance matrix Σ and not on its factorization $\Sigma_{tr} \Omega \Omega' \Sigma_{tr}'$, the likelihood function here is the same as the likelihood function of the reduced-form VAR in (6), denoted by $p(Y|\Phi, \Sigma)$. The identification problem arises precisely from the absence of Ω in this likelihood function.

We proceed by examining the effect of the identification problem on the calculation of posterior distributions. Integrating the joint density with respect to Ω yields

$$p(Y, \Phi, \Sigma) = p(Y|\Phi, \Sigma)p(\Phi, \Sigma). \tag{23}$$

Thus, the calculation of the posterior distribution of the reduced-form parameters is not affected by the presence of the nonidentifiable matrix Ω. The conditional posterior density of Ω can be calculated as follows:

$$p(\Omega|Y, \Phi, \Sigma) = \frac{p(Y, \Phi, \Sigma)p(\Omega|\Phi, \Sigma)}{\int p(Y, \Phi, \Sigma)p(\Omega|\Phi, \Sigma)d\Omega} = p(\Omega|\Phi, \Sigma). \tag{24}$$

The conditional distribution of the nonidentifiable parameter Ω 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), Poirier (1998), and Moon and Schorfheide (2009). We can deduce immediately that draws from the joint posterior distribution $p(\Phi, \Sigma, \Omega|Y)$ can in principle be obtained in two steps.

Algorithm 2.3: Posterior Sampler for Structural VARs

For $s = 1, \ldots, n_{sim}$:

1. Draw $(\Phi^{(s)}, \Sigma^{(s)})$ from the posterior $p(\Phi, \Sigma|Y)$.
2. Draw $\Omega^{(s)}$ from the conditional prior distribution $p(\Omega|\Phi^{(s)}, \Sigma^{(s)})$. □

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 Ω such that the conditional distribution of Ω, given the reduced-form parameters, reduces to a point mass. Priors for Ω are typically referred to as identification schemes because, conditional on Ω, the relationship between the forecast errors u_t and the structural shocks ϵ_t is uniquely determined. Cochrane (1994), Christiano, Eichenbaum, and Evans (1999), and Stock and Watson (2001) provide detailed surveys.

To present various identification schemes that have been employed in the literature, we consider a simple bivariate VAR(1) without intercept; that is, we set $n = 2$, $p = 1$, and $\Phi_c = 0$. For the remainder of this subsection, it is assumed that the eigenvalues of Φ_1 are all less than one in absolute value. This eigenvalue restriction guarantees that the VAR can be written as infinite-order moving average (MA(∞)):

$$y_t = \sum_{j=0}^{\infty} \Phi_1^j \Sigma_{tr} \Omega \epsilon_t. \tag{25}$$

We will refer to the sequence of partial derivatives

$$\frac{\partial y_{t+j}}{\partial \epsilon_t} = \Phi_1^j \Sigma_{tr} \Omega, \quad j = 0, 1, \ldots \tag{26}$$

as the impulse-response function. In addition, macroeconomists are often interested in so-called variance decompositions. A variance decomposition measures the fraction that each of the structural shocks contributes to the overall variance of a particular element of y_t. In the stationary bivariate VAR(1), the (unconditional) 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 the 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

$$\Gamma_{yy}^{(i)} = \sum_{j=0}^{\infty} \Phi_1^j \Sigma_{tr} \Omega \mathcal{I}^{(i)} \Omega' \Sigma_{tr}' (\Phi^j)'. \tag{27}$$

Thus, the fraction of the variance of $y_{j,t}$ explained by shock i is $[\Gamma_{yy,0}^{(i)}]_{(jj)} / [\Gamma_{yy,0}]_{(jj)}$. Variance decompositions based on h-step-ahead forecast error covariance matrices $\sum_{j=0}^{h} \Phi_1^j \Sigma (\Phi^j)'$ can be constructed in the same manner. Handling these nonlinear transformations of the VAR parameters in a Bayesian framework is straightforward, because one can simply postprocess the output of the posterior sampler (Algorithm 2.3). Using (26) or (27), each triplet $(\Phi^{(s)}, \Sigma^{(s)}, \Omega^{(s)})$, $s = 1, \ldots, n_{sim}$, can be converted into a draw from the posterior distribution of impulse responses or variance decompositions. Based on these draws, it is straightforward to compute posterior moments and credible sets.

For $n = 2$, the set of orthogonal matrices Ω can be conveniently characterized by an angle φ and a parameter $\xi \in \{-1, 1\}$:

$$\Omega(\varphi, \xi) = \begin{bmatrix} \cos\varphi & -\xi \sin\varphi \\ \sin\varphi & \xi \cos\varphi \end{bmatrix} \tag{28}$$

where $\varphi \in (-\pi, \pi]$. Each column represents a vector of unit length in $\mathbb{R}^2$, and the two vectors are orthogonal. The determinant of Ω equals ξ. Notice that $\Omega(\varphi) = -\Omega(\varphi + \pi)$. Thus, rotating the two vectors by 180 degrees simply changes the sign of the impulse responses to both shocks. Switching from $\xi = 1$ to $\xi = -1$ changes

the sign of the impulse responses to the second shock. We will now consider three different identification schemes that restrict Ω conditional on Φ and Σ.

Example 2.1 (Short-Run Identification): Suppose that y_t is composed of output deviations from trend, $\tilde{y}_t$, and that the federal funds rate, R_t, and the vector ϵ_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, following an earlier literature, Boivin and Giannoni (2006b) assume in a slightly richer setting that the private sector does not respond to monetary policy shocks contemporaneously. This assumption can be formalized by considering the following choices of φ and ξ in (28): (i) $\varphi = 0$ and $\xi = 1$; (ii) $\varphi = 0$ and $\xi = -1$; (iii) $\varphi = \pi$ and $\xi = 1$; and (iv) $\varphi = \pi$ and $\xi = -1$. It is common in the literature to normalize the direction of the impulse response by, for instance, considering responses to *expansionary* monetary policy and technology shocks. The former could be defined as shocks that lower interest rates upon impact. Since by construction $\Sigma_{22}^{tr} \geq 0$, interest rates fall in response to a monetary policy shock in cases (ii) and (iii). Likewise, since $\Sigma_{11}^{tr} \geq 0$, output increases in response to $\epsilon_{z,t}$ in cases (i) and (ii). Thus, after imposing the identification and normalization restrictions, the prior $p(\Omega|\Phi, \Sigma)$ assigns probability one to the matrix Ω that is diagonal with elements 1 and -1. Such a restriction on Ω is typically referred to as a short-run identification scheme. A short-run identification scheme was used in the seminal work by Sims (1980). $\square$

Example 2.2 (Long-Run Identification): Now suppose y_t is composed of inflation, π_t, and output growth: $y_t = [\pi_t, \Delta \ln \tilde{y}_t]'$. As in the previous example, we maintain the assumption that business-cycle fluctuations are generated by monetary policy and technology shocks, but now reverse the ordering: $\epsilon_t = [\epsilon_{R,t}, \epsilon_{z,t}]'$. We now use the following identification restriction: unanticipated changes in monetary policy shocks do not raise output in the long run. The long-run response of the log-level of output to a monetary policy shock can be obtained from the infinite sum of growth-rate responses $\sum_{j=0}^{\infty} \partial \Delta \ln \tilde{y}_{t+j} / \partial \epsilon_{R,t}$. Since the stationarity assumption implies that $\sum_{j=0}^{\infty} \Phi_1^j = (I - \Phi_1)^{-1}$, the desired long-run response is given by

$$[(I - \Phi_1)^{-1} \Sigma_{tr}]_{(2.)} \Omega_{(.1)}(\varphi, \xi), \tag{29}$$

where $A_{(.j)}$ ($A_{(j.)}$) is the j'th column (row) of a matrix A. This identification scheme has been used, for instance, by Nason and Cogley (1994) and Schorfheide (2000). To obtain the orthogonal matrix Ω, we need to determine the φ and ξ

such that the expression in (29) equals zero. Since the columns of $\Omega(\varphi, \xi)$ are composed of orthonormal vectors, we need to find a unit length vector $\Omega_{(.1)}(\varphi, \xi)$ that is perpendicular to $[(I - \Phi_1)^{-1}\Sigma_{tr}]'_{(2.)}$. Notice that ξ does not affect the first column of Ω; it only changes the sign of the response to the second shock. Suppose that (29) equals zero for $\tilde{\varphi}$. By rotating the vector $\Omega_{(.1)}(\tilde{\varphi}, \xi)$ by 180 degrees, we can find a second angle φ such that the long-run response in (29) equals zero. Thus, similar to Example 2.1, we can find four pairs (φ, ξ) such that the long-run effect (29) of a monetary policy shock on output is zero. While the shapes of the response functions are the same for each of these pairs, the sign will be different.

We could use the same normalization as in Example 2.1 by considering the effects of expansionary technology shocks (the level of output rises in the long run) and expansionary monetary policy shocks (interest rates fall in the short run). To implement this normalization, one has to choose one of the four (φ, ξ) pairs. Unlike in Example 2.1, where we used $\varphi = 0$ and $\xi = -1$ regardless of Φ and Σ, here the choice depends on Φ and Σ. However, once the normalization has been imposed, $p(\Omega|\Phi, \Sigma)$ remains a point mass. 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 estimates of the impulse response function and to inference that is very sensitive to lag length choice and prefiltering of the observations. This point dates back to Sims (1972) and a detailed discussion in the structural VAR context can be found in Leeper and Faust (1997). More recently, the usefulness of long-run restrictions has been debated in the papers by Christiano, Eichenbaum, and Vigfusson (2007) and Chari, Kehoe, and McGrattan (2008).

Example 2.3 (Sign-Restrictions): As before, let $y_t = [\pi_t, \Delta \ln \tilde{y}_t]'$ and $\epsilon_t = [\epsilon_{R,t}, \epsilon_{z,t}]'$. The priors for $\Omega|(\Phi, \Sigma)$ in the two preceding examples were degenerate. Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005) propose to be more agnostic in the choice of Ω. Suppose we restrict only the direction of impulse responses by assuming that monetary policy shocks move inflation and output in the same direction upon impact. In addition, we normalize the monetary policy shock to be expansionary; that is, output rises. Formally, this implies that $\Sigma_{tr}\Omega_{(.1)}(\varphi, \xi) \geq 0$ and is referred to as a sign-restriction identification scheme. It will become clear subsequently that sign restrictions only partially identify impulse responses in the sense that they deliver (nonsingleton) sets. Since by construction $\Sigma_{11}^{tr} \geq 0$, we can deduce

from (28) and the sign restriction on the inflation response that $\varphi \in (-\pi/2, \pi/2]$. Since $\Sigma_{22}^{tr} \geq 0$ as well, the inequality restriction for the output response can be used to sharpen the lower bound:

$$\Sigma_{21}^{tr} \cos\varphi + \Sigma_{22} \sin\varphi \geq 0 \quad \text{implies} \quad \varphi \geq \underline{\varphi}(\Sigma) = \arctan\big(-\Sigma_{21}/\Sigma_{22}\big).$$

The parameter ξ can be determined conditional on Σ and φ by normalizing the technology shock to be expansionary. To implement Bayesian inference, a researcher now has to specify a prior distribution for $\varphi|\Sigma$ with support on the interval $[\underline{\varphi}(\Sigma), \pi/2]$ and a prior for $\xi|(\varphi, \Sigma)$. In practice, researchers have often chosen a uniform distribution for $\varphi|\Sigma$ as we will discuss in more detail below. $\square$

For short- and long-run identification schemes, it is straightforward to implement Bayesian inference. One can use a simplified version of Algorithm 2.3, in which $\Omega^{(s)}$ is calculated directly as function of $(\Phi^{(s)}, \Sigma^{(s)})$. For each triplet (Φ, Σ, Ω), suitable generalizations of (26) and (27) can be used to convert parameter draws into draws of impulse responses or variance decompositions. With these draws in hand, one can approximate features of marginal posterior distributions such as means, medians, standard deviations, or credible sets. In many applications, including the empirical illustration provided below, researchers are interested only in the response of an n-dimensional vector y_t to one particular shock, say a monetary policy shock. In this case, one can simply replace Ω in the previous expressions by its first column $\Omega_{(.1)}$, which is a unit-length vector.

Credible sets for impulse responses are typically plotted as error bands around mean or median responses. It is important to keep in mind that impulse-response functions are multidimensional objects. However, the error bands typically reported in the literature have to be interpreted point-wise, that is, they delimit the credible set for the response of a particular variable at a particular horizon to a particular shock. In an effort to account for the correlation between responses at different horizons, Sims and Zha (1999) propose a method for computing credible bands that relies on the first few principal components of the covariance matrix of the responses.

Bayesian inference in sign-restricted structural VARs is more complicated because one has to sample from the conditional distribution of $p(\Omega|\Phi, \Sigma)$. Some authors, like Uhlig (2005), restrict their attention to one particular shock and parameterize only one column of the matrix Ω. Other authors, like Peersman (2005), construct responses for the full set of n shocks. In practice, sign restrictions are imposed not

just on impact but also over longer horizons $j > 0$. Most authors use a conditional prior distribution of $\Omega|(\Phi, \Sigma)$ that is *uniform.* Any r columns of Ω can be interpreted as an orthonormal basis for an r-dimensional subspace of $\mathbb{R}^n$. The set of these subspaces is called Grassmann manifold and denoted by $\mathcal{G}_{r,n-r}$. Thus, specifying a prior distribution for (the columns of) Ω can be viewed as placing probabilities on a Grassmann manifold. A similar problem arises when placing prior probabilities on cointegration spaces, and we will provide a more extensive discussion in Section 3.3. A uniform distribution can be defined as the unique distribution that is invariant to transformations induced by orthonormal transformations of $\mathbb{R}^n$ (James (1954)). For $n = 2$, this uniform distribution is obtained by letting $\varphi \sim U(-\pi, \pi]$ in (28) and, in case of Example 2.3, restricting it to the interval $[-\underline{\varphi}(\Sigma), \pi/2]$. Detailed descriptions of algorithms for Bayesian inference in sign-restricted structural VARs for $n > 2$ can be found, for instance, in Uhlig (2005) and Rubio-Ramírez, Waggoner, and Zha (2010).

Illustration 2.1: We consider a VAR(4) based on output, inflation, interest rates, and real money balances. The data are obtained from the FRED database of the Federal Reserve Bank of St. Louis. Database identifiers are provided in parentheses. Per capita output is defined as real GDP (GDPC96) divided by the civilian noninstitutionalized population (CNP16OV). We take the natural log of per capita output and extract a deterministic trend by OLS regression over the period 1959:I to 2006:IV.[3] The deviations from the linear trend are scaled by 100 to convert them into percentages. Inflation is defined as the log difference of the GDP deflator (GDPDEF), scaled by 400 to obtain annualized percentage rates. Our measure of nominal interest rates corresponds to the average federal funds rate (FEDFUNDS) within a quarter. We divide sweep-adjusted M2 money balances by quarterly nominal GDP to obtain inverse velocity. We then remove a linear trend from log inverse velocity and scale the deviations from trend by 100. Finally, we add our measure of detrended per capita real GDP to obtain real money balances. The sample used for posterior inference is restricted to the period from 1965:I to 2005:I.

We use the dummy-observation version of the Minnesota prior described in Section 2.2 with the hyperparameters $\lambda_2 = 4$, $\lambda_3 = 1$, $\lambda_4 = 1$, and $\lambda_5 = 1$. We consider

[3] This deterministic trend could also be incorporated into the specification of the VAR. However, in this illustration we wanted (i) to only remove a deterministic trend from output and not from the other variables and (ii) to use Algorithm 2.1 and the marginal likelihood formula (15) which do not allow for equation-specific parameter restrictions.

Table 1: Hyperparameter Choice for Minnesota Prior

λ_1	0.01	0.10	0.50	1.00	2.00
$\pi_{i,0}$	0.20	0.20	0.20	0.20	0.20
$\ln p_\lambda(Y)$	-914.35	-868.71	-888.32	-898.18	-902.43
$\pi_{i,T}$	0.00	1.00	0.00	0.00	0.00

five possible values for λ_1, which controls the overall variance of the prior. We assign equal prior probability to each of these values and use (15) to compute the marginal likelihoods $p_\lambda(Y)$. Results are reported in Table 1. The posterior probabilites of the hyperparameter values are essentially degenerate, with a weight of approximately one on $\lambda_1 = 0.1$. The subsequent analysis is conducted conditional on this hyperparameter setting.

Draws from the posterior distribution of the reduced-form parameters Φ and Σ can be generated with Algorithm 2.1, using the appropriate modification of $\hat{S}$, $\hat{\Phi}$ and X, described at the beginning of Section 2.2. To identify the dynamic response to a monetary policy shock, we use the sign-restriction approach described in Example 2.3. In particular, we assume that a contractionary monetary policy shock raises the nominal interest rate upon impact and for one period after the impact. During these two periods, the shock also lowers inflation and real money balances. Since we are identifying only one shock, we focus on the first column of the orthogonal matrix Ω. We specify a prior for $\Omega_{(.1)}$ that implies that the space spanned by this vector is uniformly distributed on the relevant Grassman manifold. This uniform distribution is truncated to enforce the sign restrictions given (Φ, Σ). Thus, the second step of Algorithm 2.3 is implemented with an acceptance sampler that rejects proposed draws of Ω for which the sign restrictions are not satisfied. Proposal draws $\tilde{\Omega}$ are obtained by sampling $Z \sim N(0, I)$ and letting $\tilde{\Omega} = Z/\|Z\|$.

Posterior means and credible sets for the impulse responses are plotted in Figure 2. According to the posterior mean estimates, a one-standard deviation shock raises interest rates by 40 basis points upon impact. In response, the (annualized) inflation rate drops by 30 basis points, and real money balances fall by 0.4 percent. The posterior mean of the output response is slightly positive, but the 90% credible set ranges from -50 to about 60 basis points, indicating substantial uncertainty about the sign and magnitude of the real effect of unanticipated changes in monetary policy

under our fairly agnostic prior for the vector $\Omega_{(.1)}$. □

INSERT FIGURE 2 HERE

2.4.2 An Alternative Structural VAR Parameterization

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 premultiply 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

$$A_0' y_t = A_1 y_{t-1} + \ldots A_p y_{t-p} + A_c + \epsilon_t, \quad \epsilon_t \sim iidN(0, I). \tag{30}$$

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 (30) is that the coefficients have direct behaviorial interpretations. For instance, one could impose identifying restrictions on A_0 such that the first equation in (30) 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 (30) is provided in Sims and Zha (1998). As in (5), let $x_t' = [y_{t-1}', \ldots, y_{t-p}', 1]$ and Y and X be matrices with rows y_t', x_t', respectively. Moreover, we use E to denote the $T \times n$ matrix with rows ϵ_t'. Finally, define $A = [A_1, \ldots, A_p, A_c]'$ such that (30) can be expressed as a multivariate regression of the form

$$YA_0 = XA + E \tag{31}$$

with likelihood function

$$p(Y|A_0, A) \propto |A_0|^T \exp\left\{-\frac{1}{2} tr[(YA_0 - XA)'(YA_0 - XA)]\right\}. \tag{32}$$

The term $|A_0|^T$ is the determinant of the Jacobian associated with the transformation of E into Y. 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 matricvariate Normal.

Sims and Zha (1998) propose prior distributions that share the Kronecker structure of the likelihood function and hence lead to posterior distributions that can

be evaluated with a high degree of numerical efficiency, that is, without having to invert matrices of the dimension $nk \times nk$. Specifically, 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

$$A|A_0 \sim MN\Big(\underline{A}(A_0), \lambda^{-1} I \otimes \underline{V}(A_0)\Big), \tag{33}$$

where the matrix of means $\underline{A}(A_0)$ and the covariance matrix $\underline{V}(A_0)$ are potentially functions of A_0 and λ is a hyperparameter that scales the prior covariance matrix. 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:

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

Combining the likelihood function (32) with the prior (33) leads to a posterior for A that is conditionally matricvariate Normal:

$$A|A_0, Y \sim MN\Big(\bar{A}(A_0), I \otimes \bar{V}(A_0)\Big), \tag{34}$$

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. An example of such restrictions, based on a structural VAR analyzed by Robertson and Tallman (2001), is provided next.

Example 2.4: Suppose 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 Robertson and Tallman (2001) are summarized in Table 2.4.2. Each row in the table corresponds to a behavioral equation labeled on the left-hand side of the row. 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. The entries in the table imply that the

Table 2: Identification Restrictions for A_0'

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

Notes: Each row in the table represents a behavioral equation labeled on the left-hand side of the row: information market (Inform), monetary policy rule (MP), money demand (MD), and three equations that characterize the production sector of the economy (Prod). The column labels reflect the observables: commodity prices (Pcom), monetary aggregate (M2), federal funds rate (R), real GDP (Y), consumer price index (CPI), and unemployment (U). A 0 entry denotes a coefficient set to zero. □

only variables that enter contemporaneously into the monetary policy rule (MP) are the federal funds rate (R) and M2. The structural VAR here is overidentified, because the covariance matrix of the one-step-ahead forecast errors of a VAR with $n = 6$ has in principle 21 free elements, whereas the matrix A_0 has only 18 free elements. Despite the fact that overidentifying restrictions were imposed, the system requires a further normalization. One can 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 require that the diagonal elements of A_0 all be nonnegative. In practice, this normalization can be imposed by postprocessing the output of the posterior sampler: for all draws $(A_0', A_1, \ldots, A_p, A_c)$ multiply the i'th row of each matrix by -1 if $A_{0,ii} < 0$. This normalization works well if the posterior support of each diagonal element of A_0 is well away from zero. Otherwise, this normalization may induce bimodality in distributions of other parameters. □

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 (33), 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

$$p(Y|A_0) = \int p(Y|A_0, A)p(A|A_0)dA \propto |A_0|^T \exp\left\{-\frac{1}{2}tr[A_0'\bar{S}A_0]\right\}, \qquad (35)$$

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, q_i is the number of unrestricted elements of $A_{0(.i)}$, and U_i is an $n \times q_i$ matrix, composed of orthonormal column vectors. Under the assumption that $b_i \sim N(\underline{b}_i, \underline{\Omega}_i)$, independently across i, we obtain

$$p(b_1, \ldots, b_n|Y) \propto |[U_1 b_1, \ldots, U_n b_n]|^T \exp\left\{-\frac{T}{2}\sum_{i=1}^{n} b_i' S_i b_i\right\}, \qquad (36)$$

where $S_i = U_i'(\bar{S} + \underline{\Omega}_i^{-1})U_i$ and A_0 can be recovered from the b_i's. Now consider the conditional density of $b_i|(b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n)$:

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

Since b_i also appears in the determinant, its distribution is not Normal. Characterizing the distribution of b_i requires a few additional steps. 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 introduce the parameters $\beta_1, \ldots, \beta_{q_i}$ and reparameterize the vector b_i as a linear combination of the w_j's:

$$b_i = V_i \sum_{j=1}^{q_i} \beta_j w_j. \qquad (37)$$

By the orthonormal property of the w_j's, we can verify that the conditional posterior of the β_j's is given by

$$\begin{aligned} & p(\beta_1, \ldots, \beta_{q_i}|Y, b_1, \ldots, b_{i-1}, b_{i+1}, \ldots, b_n) \qquad (38) \\ & \propto \left(\sum_{j=1}^{q_i} |[U_1 b_1, \ldots, \beta_j V_i w_j, \ldots, U_n b_n]|\right)^T \exp\left\{-\frac{T}{2}\sum_{j=1}^{q_i}\beta_j^2\right\} \\ & \propto |\beta_1|^T \exp\left\{-\frac{T}{2}\sum_{j=1}^{q_i}\beta_j^2\right\}. \end{aligned}$$

The last line follows because $w_2, \ldots, w_{q_i}$ by construction falls in the space spanned by $U_j b_j$, $j \neq i$. Thus, all β_j's are independent of each other, β_1 has a Gamma

distribution, and β_j, $2 \leq j \leq q_i$, are normally distributed. Draws from the posterior of A_0 can be obtained by Gibbs sampling.

Algorithm 2.4: Gibbs Sampler for Structural VARs

For $s = 1, \ldots, n_{sim}$:

1. Draw $A_0^{(s)}$ conditional on $(A^{(s-1)}, Y)$ as follows. For $i = 1, \ldots, n$ generate $\beta_1, \ldots, \beta_{q_i}$ from (38) conditional on $(b_1^{(s)}, \ldots, b_{i-1}^{(s)}, b_{i+1}^{(s-1)}, \ldots, b_n^{(s-1)})$, define $b_i^{(s)}$ according to (37), and let $A_{0(.i)}^{(s)} = U_i b_i^{(s)}$.

2. Draw $A^{(s)}$ conditional on $(A_0^{(s)}, Y)$ from the matricvariate Normal distribution in (34). $\square$

2.5 Further VAR Topics

The literature on Bayesian analysis of VARs is by now extensive, and our presentation is by no means exhaustive. A complementary survey of Bayesian analysis of VARs including VARs with time-varying coefficients and factor-augmented VARs can be found in Koop and Korobilis (2010). Readers who are interested in using VARs for forecasting purposes can find algorithms to compute such predictions efficiently, possibly conditional on the future path of a subset of variables, in Waggoner and Zha (1999). Rubio-Ramírez, Waggoner, and Zha (2010) provide conditions for the global identification of VARs of the form (30). Our exposition was based on the assumption that the VAR innovations are homoskedastic. Extensions to GARCH-type heteroskedasticity can be found, for instance, in Pelloni and Polasek (2003). Uhlig (1997) proposes a Bayesian approach to VARs with stochastic volatility. We will discuss VAR models with stochastic volatility in Section 5.

3 VARs with Reduced-Rank Restrictions

It is well documented that many economic time series such as aggregate output, consumption, and investment exhibit clear trends and tend to be very persistent. 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. Examples are the so-called *Great Ratios*, such as the consumption-output or investment-output ratio (see Klein and Kosobud (1961)). The left panel of Figure 3

depicts log nominal GDP and nominal aggregate investment for the United States over the period 1965-2006 (obtained from the FRED database of the Federal Reserve Bank of St. Louis) and the right panel 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 mean-reverting. The observation that particular linear combinations of nonstationary economic time series appear to be stationary has triggered a large literature on cointegration starting in the mid 1980's; see, for example, Engle and Granger (1987), Johansen (1988), Johansen (1991), and Phillips (1991).

Insert Figure 3 Here

More formally, the dynamic behavior of a univariate 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 nonstationary. Unit-root processes are often called integrated of order one, I(1), because stationarity can be induced by taking first differences $\Delta y_t = (1-L)y_t$. If a linear combination of univariate 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 in Section 3.1 that one can impose cotrending restrictions in a VAR by restricting some of the eigenvalues of its characteristic polynomial to unity. This leads to the so-called vector error correction model, which takes the form of a reduced-rank regression. Such restricted VARs have become a useful and empirically successful tool in applied macroeconomics. In Section 3.2, we discuss Bayesian inference in cointegration systems under various types of prior distributions.

3.1 Cointegration Restrictions

Consider the reduced-form VAR specified in (1). Subtracting y_{t-1} from both sides of the equality leads to

$$\Delta y_t = (\Phi_1 - I)y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \Phi_c + u_t, \quad u_t \sim iidN(0, \Sigma). \quad (39)$$

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 (39) as

$$\Delta y_t = \Pi_* y_{t-1} + \Pi_1 \Delta y_{t-1} + \ldots + \Pi_{p-1} \Delta y_{t-p+1} + \Pi_c + u_t, \tag{40}$$

where

$$\Pi_* = -\Phi(1) \quad \text{and} \quad \Phi(z) = I - \sum_{j=1}^{p} \Phi_j z^j.$$

$\Phi(z)$ is the characteristic polynomial of the VAR. If the VAR has unit roots, – that is, $|\Phi(1)| = 0$ – then the matrix Π_* is of reduced rank. If the rank of Π_* equals $r < n$, we can reparameterize the matrix as $\Pi_* = \alpha\beta'$, where α and β are $n \times r$ matrices of full column rank. This reparameterization leads to the so-called vector error correction or vector equilibrium correction (VECM) representation:

$$\Delta y_t = \alpha\beta' y_{t-1} + \Pi_1 \Delta y_{t-1} + \ldots + \Pi_{p-1} \Delta y_{t-p+1} + \Pi_c + u_t, \tag{41}$$

studied by Engle and Granger (1987).

A few remarks are in order. It can be easily verified that the parameterization of Π_* in terms of α and β is not unique: for any nonsingular $r \times r$ 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 matrices α and β, it is useful to define a matrix $\alpha_\perp$ and $\beta_\perp$ of full column rank and dimension $n \times (n-r)$ such that $\alpha'\alpha_\perp = 0$ and $\beta'\beta_\perp = 0$. If no root of $\Phi(z) = 0$ lies inside the unit circle and $\alpha_\perp'\beta_\perp$ has full rank, then (41) implies that y_t can be expressed as (Granger's Representation Theorem):

$$y_t = \beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp' \sum_{\tau=1}^{t}(u_t + \Pi_c) + \Psi(L)(u_t + \Pi_c) + P_{\beta_\perp} y_0. \tag{42}$$

$\Gamma = I - \sum_{j=1}^{p-1} \Pi_j$, $P_{\beta_\perp}$ 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. The columns of β are called cointegration vectors. Moreover, y_t has $n - r$ common stochastic trends given by $(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp' \sum_{\tau=1}^{t}(u_t + \Pi_c)$. A detailed exposition can be found, for instance, in the monograph by Johansen (1995).

If y_t is composed of log GDP and investment, a visual inspection of Figure 3 suggests that the cointegration vector β is close to $[1, -1]'$. Thus, according to (41) the growth rates of output and investment should be modeled as functions of lagged growth rates as well as the log investment-output ratio. Since in this example $\beta_\perp$ is

2×1 and the term $(\alpha_\perp' \Gamma \beta_\perp)^{-1} \alpha_\perp' \sum_{\tau=1}^{t} (u_t + \Pi_c)$ is scalar, Equation (42) highlights the fact 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 to specifying a prior distribution for Π_* 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 with Gaussian Prior for β

Define $\Pi = [\Pi_1, \ldots, \Pi_{p-1}, \Pi_c]'$ and let $u_t \sim N(0, \Sigma)$. Inspection of (41) suggests that conditional on α and β, 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. A Gibbs sampler to generate draws from the posterior distribution of the VECM typically has the following structure:

Algorithm 3.1: Gibbs Sampler for VECM

For $s = 1, \ldots, n_{sim}$:

1. Draw $(\Pi^{(s)}, \Sigma^{(s)})$ from the posterior $p(\Pi, \Sigma | \Pi_*^{(s-1)}, Y)$.
2. Draw $\Pi_*^{(s)}$ from the posterior $p(\Pi_* | \Pi^{(s)}, \Sigma^{(s)}, Y)$. $\square$

To simplify the subsequent exposition, we will focus on inference for $\Pi_* = \alpha\beta'$ conditional on Π and Σ for the remainder of this section (Step 2 of Algorithm 3.1). To do so, we study the simplified model

$$\Delta y_t = \Pi_* y_{t-1} + u_t, \quad \Pi_* = \alpha\beta', \quad u_t \sim iidN(0, \Sigma), \tag{43}$$

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

In this section, we consider independent priors $p(\alpha)$ and $p(\beta)$ that are either flat or Gaussian. Geweke (1996) used such priors to study inference in the reduced-rank regression model. Throughout this subsection we normalize $\beta' = [I_{r \times r}, B_{r \times (n-r)}']$

The prior distribution for β is induced by a prior distribution for B. This normalization requires that the elements of y_t be ordered such that each of these variables appears in at least one cointegration relationship. We will discuss the consequences of this normalization later on.

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 presample 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. For brevity, we refer to this class of priors as balanced-growth-path priors. An informative prior for α 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 Σ (both subsequently omitted from our notation), the likelihood function is of the form

$$p(Y|\alpha,\beta) \propto |\Sigma|^{-T/2} \exp\left\{ -\frac{1}{2} tr[\Sigma^{-1}(\Delta Y - X\beta\alpha')'(\Delta Y - X\beta\alpha')] \right\}. \tag{44}$$

In turn, we will derive conditional posterior distributions for α and β based on the likelihood (44). We begin with the posterior of α. Define $\tilde{X} = X\beta$. Then

$$p(\alpha|Y,\beta) \propto p(\alpha) \exp\left\{ -\frac{1}{2} tr[\Sigma^{-1}(\alpha\tilde{X}'\tilde{X}\alpha' - 2\alpha\tilde{X}'\Delta Y)] \right\}. \tag{45}$$

Thus, as long as the prior of $vec(\alpha')$ is Gaussian, the posterior of $vec(\alpha')$ is multivariate Normal. If the prior has the same Kronecker structure as the likelihood function, then the posterior is matricvariate Normal.

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

$$\Delta Y = X_1\alpha' + X_2 B\alpha' + U.$$

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

$$Z = X_2 B\alpha' + U. \tag{46}$$

The fact that B is right-multiplied by α' complicates the analysis. The following steps are designed to eliminate the α' term. Post-multiplying (46) by the matrix

$C = [\alpha(\alpha'\alpha)^{-1}, \alpha_\perp]$ yields the seemingly unrelated regression

$$\big[\tilde{Z}_1,\ \tilde{Z}_2\big] = X_2\big[B,\ 0\big] + \big[\tilde{U}_1,\ \tilde{U}_2\big], \tag{47}$$

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 about $\tilde{U}_1$, which sharpens the inference for B. Formally, let $\tilde{\Sigma} = C'\Sigma C$ and partition $\tilde{\Sigma}$ conforming with $\tilde{U} = [\tilde{U}_1, \tilde{U}_2]$. The mean and variance of $\tilde{Z}_1$ conditional on $\tilde{Z}_2$ are given by $(\tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{Z}_2 + X_2B)$ and $\tilde{\Sigma}_{1|2} = \tilde{\Sigma}_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21}$, respectively. Define $\tilde{Z}_{1|2} = \tilde{Z}_1 - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{Z}_2$. Then we can deduce

$$p(B|Y,\alpha) \propto p(\beta(B)) \exp\left\{ -\frac{1}{2} tr\left[\tilde{\Sigma}_{1|2}^{-1}(\tilde{Z}_{1|2} - X_2B)'(\tilde{Z}_{1|2} - X_2B)\right]\right\}. \tag{48}$$

Thus, if the prior distribution for B is either flat or Normal, then the conditional posterior of B given α is Normal.

Algorithm 3.2: Gibbs Sampler for Simple VECM with Gaussian Priors

For $s = 1, \ldots, n_{sim}$:

1. Draw $\alpha^{(s)}$ from $p(\alpha|\beta^{(s-1)}, Y)$ given in (45).
2. Draw $B^{(s)}$ from $p(B|\alpha^{(s)}, Y)$ given in (48) and let $\beta^{(s)} = [I, B^{(s)'}]'$. □

Illustration 3.1: We use the VECM in (41) with $p = 4$ and the associated moving-average representation (42) to extract a common trend from the U.S. investment and GDP data depicted in Figure 3. We use an improper prior of the form

$$p(\Pi, \Sigma, \alpha, B) \propto |\Sigma|^{-(n+1)/2} \exp\left\{ -\frac{1}{2\lambda}(B - (-1))^2 \right\},$$

where $\lambda \in \{0.01, 0.1, 1\}$. The prior distribution for the cointegration vector $\beta = [1, B]'$ is centered at the balanced-growth-path values $[1, -1]'$. Draws from the posterior distribution are generated through a Gibbs sampler in which Step 2 of Algorithm 3.1 is replaced by the two steps described in Algorithm 3.2. The posterior density for B is plotted in Figure 4 for the three parameterizations of the prior variance λ. The posterior is similar for all three choices of λ, indicating that the data

are quite informative about the cointegration relationship. For each prior, the posterior mean of B is about -1.07, with most of the mass of the distributions placed on values less than -1, indicating a slight violation of the balanced-growth-path restriction. Using posterior draws based on $\lambda = 0.10$, Figure 5 plots the decompositions of log nominal aggregate investment and log nominal GDP into common trends and stationary fluctuations around those trends. The plots in the left column of the Figure display the common trend $\beta_\perp(\alpha_\perp'\Gamma\beta_\perp)^{-1}\alpha_\perp'\sum_{\tau=1}^{t}(u_t + \Pi_c)$ for each series, while the plots in the right column show the demeaned stationary component $\Psi(L)u_t$. National Bureau of Economic Research (NBER) recession dates are overlayed in gray. □

Insert Figure 4 Here

Insert Figure 5 Here

3.3 Further Research on Bayesian Cointegration Models

The Bayesian analysis of cointegration systems has been an active area of research, and a detailed survey is provided by Koop, Strachan, van Dijk, and Villani (2006). Subsequently, we consider two strands of this literature. The first strand points out that the columns of β in (41) should be interpreted as a characterization of a subspace of $\mathbb{R}^n$ and that priors for β are priors over subspaces. The second strand uses prior distributions to regularize or smooth the likelihood function of a cointegration model in areas of the parameter space in which it is very nonelliptical.

We begin by reviewing the first strand. Strachan and Inder (2004) and Villani (2005) emphasize that specifying a prior distribution for β amounts to placing a prior probability on the set of r-dimensional subspaces of $\mathbb{R}^n$ (Grassmann manifold $\mathcal{G}_{r,n-r}$), which we previously encountered in the context of structural VARs in Section 2.4.1. Our discussion focuses on the output-investment example with $n = 2$ and $r = 1$. In this case the Grassmann manifold consists of all the lines in $\mathbb{R}^2$ that pass through the origin. Rather than normalizing one of the ordinates of the cointegration vector β to one, we can alternatively normalize its length to one and express it in terms of polar coordinates. For reasons that will become apparent subsequently, we let

$$\beta(\varphi) = [\cos(-\pi/4 + \pi(\varphi - 1/2)), \sin(-\pi/4 + \pi(\varphi - 1/2))]', \quad \varphi \in (0, 1].$$

The one-dimensional subspace associated with $\beta(\varphi)$ is given by $\lambda\beta(\varphi)$, where $\lambda \in \mathbb{R}$. In our empirical illustration, we used a balanced-growth-path prior that was centered at the cointegration vector $[1, -1]'$. This vector lies in the space spanned by $\beta(1/2)$. Thus, to generate prior distributions that are centered at the balanced-growth-path restriction, we can choose a Beta distribution for φ and let $\varphi \sim B(\gamma, \gamma)$. If $\gamma >> 1$, then the prior is fairly dogmatic.

As γ approaches 1 from above it becomes more diffuse. In fact, if $\gamma = 1$, then $\varphi \sim U(0, 1]$, and it turns out that the subspaces associated with $\beta(\varphi)$ are uniformly distributed on the Grassmann manifold (see James (1954)). This uniform distribution is defined to be the unique distribution that is invariant under the group of orthonormal transformations of $\mathbb{R}^n$. For $n = 2$, this group is given by the set of orthogonal matrices specified in (28), which rotate the subspace spanned by $\beta(\varphi)$ around the origin. Villani (2005) proposes to use the uniform distribution on the Grassman manifold as a reference prior for the analysis of cointegration systems and, for general n and r, derives the posterior distribution for α and β using the ordinal normalization $\beta' = [I, B']$.

Strachan and Inder (2004) are very critical of the ordinal normalization, because a flat and apparently noninformative prior on B in $\beta' = [I, B']$ favors the cointegration spaces near the region where the linear normalization is invalid, meaning that some of the first r variables do not appear in any cointegration vector. Instead, these authors propose to normalize β according to $\beta'\beta = I$ and develop methods of constructing informative and diffuse priors on the Grassmann manifold associated with β.

We now turn to the literature on regularization. Kleibergen and van Dijk (1994) and Kleibergen and Paap (2002) use prior distributions to correct irregularities in the likelihood function of the VECM, caused by local nonidentifiability of α and B under the ordinal normalization $\beta' = [I, B']$. As the loadings α for the cointegration relationships $\beta' y_{t-1}$ approach zero, B becomes nonidentifiable. If the highly informative balanced-growth-path prior discussed previously were replaced by a flat prior for B – that is $p(B) \propto \textit{constant}$ – to express diffuse prior beliefs about cointegration relationships, then the conditional posterior of B given $\alpha = 0$ is improper, and its density integrates to infinity. Under this prior, the marginal posterior density of α 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 α tends to favor near-zero values for which the cointegration relationships are poorly identified.

Kleibergen and Paap (2002) propose the following alternative. The starting point is a singular-value decomposition of a (for now) unrestricted $n\times n$ matrix Π_*', which takes the form:

$$\Pi_*' = VDW' = \left[\begin{array}{cc} V_{11} & V_{12} \\ V_{21} & V_{22} \end{array}\right]\left[\begin{array}{cc} D_{11} & 0 \\ 0 & D_{22} \end{array}\right]\left[\begin{array}{cc} W_{11}' & W_{21}' \\ W_{12}' & W_{22}' \end{array}\right]. \tag{49}$$

V and W are orthogonal $n\times n$ matrices, and D is a diagonal $n\times n$ matrix. The partitions V_{11}, D_{11}, and W_{11} are of dimension $r\times r$, and all other partitions conform. Regardless of the rank of Π_*', it can be verified that the matrix can be decomposed as follows:

$$\begin{aligned}\Pi_*' &= \left[\begin{array}{c} V_{11} \\ V_{21} \end{array}\right] D_{11} \left[\begin{array}{cc} W_{11}' & W_{21}' \end{array}\right] + \left[\begin{array}{c} V_{12} \\ V_{22} \end{array}\right] D_{22} \left[\begin{array}{cc} W_{12}' & W_{22}' \end{array}\right] \\ &= \beta\alpha' + \beta_\perp \Lambda \alpha_\perp', \end{aligned} \tag{50}$$

where

$$\beta = \left[\begin{array}{c} I \\ B \end{array}\right], \quad B = V_{21}V_{11}^{-1}, \quad \text{and} \quad \alpha' = V_{11}D_{11}[W_{11}', W_{21}'].$$

The matrix Λ is chosen to obtain a convenient functional form for the prior density below:

$$\Lambda = (V_{22}'V_{22})^{-1/2}V_{22}D_{22}W_{22}'(W_{22}W_{22}')^{-1/2}.$$

Finally, the matrices $\beta_\perp'$ and $\alpha_\perp'$ take the form $\beta_\perp' = M_\beta'[V_{12}'\ V_{22}']$ and $\alpha_\perp' = M_\alpha'[W_{12}'\ W_{22}']$, respectively. Here M_α and M_β are chosen such that the second equality in (50) holds. For $\Lambda = 0$ the rank of the unrestricted Π_*' in (50) reduces to r and we obtain the familiar expression $\Pi_*' = \beta\alpha'$.

The authors start from a flat prior on Π_*: that is, $p(\Pi_*) \propto \textit{constant}$, ignoring the rank reduction generated by the r cointegration relationships. They proceed by deriving a conditional distribution for Π_* given $\Lambda = 0$, and finally use a change of variables to obtain a distribution for the parameters of interest, α and B. Thus,

$$p(\alpha, B) \propto |J_{\Lambda=0}(\Pi_*(\alpha, B, \Lambda))| \propto |\beta'\beta|^{(n-r)/2}|\alpha\alpha'|^{(n-r)/2}. \tag{51}$$

Here, $J_{\Lambda=0}(\Pi_*(\alpha, B, \Lambda))$ is the Jacobian associated with the mapping between Π_* and (α, B, Λ). This prior has the property that as $\alpha \longrightarrow 0$ its 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).

4 Dynamic Stochastic General Equilibrium Models

The term *DSGE model* is typically 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 total factor productivity, for instance, or the nominal interest rate set by a central bank. This uncertainty is generated by exogenous stochastic processes that shift technology, for example, 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. In a Bayesian framework, this likelihood function can be used to transform a prior distribution for the structural parameters of the DSGE model into a posterior distribution. This posterior is the basis for substantive inference and decision making. DSGE models can be used for numerous tasks, such as studying the sources of business-cycle fluctuations and the propagation of shocks to the macroeconomy, generating predictive distributions for key macroeconomic variables, and analyzing the welfare effects of economic policies, taking both parameter and model uncertainty into account.

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 are discussed in Section 4.2. Bayesian inference on the parameters of a linearized DSGE model is discussed in Section 4.3. Extensions to models with indeterminacies or stochastic volatility, and to models solved with nonlinear techniques are discussed in Sections 4.4, 4.5, and 4.6, respectively. Section 4.7 discusses numerous methods of documenting the performance of DSGE models and comparing them to less restrictive models such as vector autoregressions. Finally, we provide a brief discussion of some empirical applications in Section 4.8. A detailed survey of Bayesian techniques for the estimation and evaluation of DSGE models is provided in An and Schorfheide (2007a).

INSERT FIGURE 6 HERE

4.1 A Prototypical DSGE Model

Figure 6 depicts postwar aggregate log output, hours worked, and log labor productivity for the US. Precise data definitions are provided in Ríos-Rull, Schorfheide, Fuentes-Albero, Kryshko, and Santaeulalia-Llopis (2009). 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 is exogenous changes in total factor productivity. We will illustrate the techniques discussed in this section with the estimation of a stochastic growth model based on observations on aggregate output and hours worked.

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

$$\mathbb{E}_t\left[\sum_{s=0}^{\infty}\beta^{t+s}\left(\ln C_{t+s}-\frac{(H_{t+s}/B_{t+s})^{1+1/\nu}}{1+1/\nu}\right)\right] \tag{52}$$

subject to a sequence of budget constraints

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

The household receives the labor income $W_t H_t$, where W_t is the hourly wage. It owns the capital stock K_t and rents it to the firms at the rate R_t. Capital accumulates according to

$$K_{t+1} = (1-\delta)K_t + I_t, \tag{53}$$

where I_t is investment and δ is the depreciation rate. The household uses the discount rate β, and B_t is an exogenous preference shifter that can be interpreted as a labor supply shock. If B_t increases, then the disutility associated with hours worked falls. Finally, ν is the aggregate 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:

$$\frac{1}{C_t} = \beta\mathbb{E}\left[\frac{1}{C_{t+1}}(R_{t+1}+(1-\delta))\right] \quad \text{and} \quad \frac{1}{C_t}W_t = \frac{1}{B_t}\left(\frac{H_t}{B_t}\right)^{1/\nu}. \tag{54}$$

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

$$Y_t = (A_t H_t)^{\alpha} K_t^{1-\alpha}. \tag{55}$$

The stochastic process A_t represents the exogenous labor augmenting technological progress. Firms solve a static profit maximization problem and choose labor and capital to equate marginal products of labor and capital with the wage and rental rate of capital, respectively:

$$W_t = \alpha \frac{Y_t}{H_t}, \quad R_t = (1-\alpha)\frac{Y_t}{K_t}. \tag{56}$$

An equilibrium is a sequence of prices and quantities such that (i) the representative household maximizes utility and firms maximize profits taking the prices as given, and (ii) markets clear, implying that

$$Y_t = C_t + I_t. \tag{57}$$

To close the model, we specify a law of motion for the two exogenous processes. Log technology evolves according to

$$\ln A_t = \ln A_0 + (\ln \gamma)t + \ln \widetilde{A}_t, \quad \ln \widetilde{A}_t = \rho_a \ln \widetilde{A}_{t-1} + \sigma_a \epsilon_{a,t}, \quad \epsilon_{a,t} \sim iidN(0,1), \tag{58}$$

where $\rho_a \in [0,1]$. If $0 \leq \rho_a < 1$, the technology process is trend stationary. If $\rho_a = 1$, then $\ln A_t$ is a random-walk process with drift. Exogenous labor supply shifts are assumed to follow a stationary AR(1) process:

$$\ln B_t = (1-\rho_b)\ln B_* + \rho_b \ln B_{t-1} + \sigma_b \epsilon_{b,t}, \quad \epsilon_{b,t} \sim iidN(0,1), \tag{59}$$

and $0 \leq \rho_b < 1$. To initialize the exogenous processes, we assume

$$\ln \widetilde{A}_{-\tau} = 0 \quad \text{and} \quad \ln B_{-\tau} = 0.$$

The solution to the rational expectations difference equations (53) to (59) 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. Since we will subsequently solve the model by constructing a local approximation of its dynamics near a steady state, it is useful to detrend the model variables as follows:

$$\widetilde{Y}_t = \frac{Y_t}{A_t}, \ \widetilde{C}_t = \frac{C_t}{A_t}, \ \widetilde{I}_t = \frac{I_t}{A_t}, \ \widetilde{K}_{t+1} = \frac{K_{t+1}}{A_t}, \ \widetilde{W}_t = \frac{W_t}{A_t}. \tag{60}$$

The detrended variables are mean reverting. This bounds the probability of experiencing large deviations from the log-linearization point for which the approximate solution becomes inaccurate. According to our timing convention, K_{t+1} refers to capital at the 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 (53) to (57) in terms of the detrended variables:

$$\begin{aligned}
\frac{1}{\widetilde{C}_t} &= \beta \mathbb{E}\left[\frac{1}{\widetilde{C}_{t+1}} e^{-a_{t+1}}(R_{t+1}+(1-\delta))\right], \quad \frac{1}{\widetilde{C}_t}\widetilde{W}_t = \frac{1}{B_t}\left(\frac{H_t}{B_t}\right)^{1/\nu} \\
\widetilde{W}_t &= \alpha\frac{\widetilde{Y}_t}{H_t}, \quad R_t = (1-\alpha)\frac{\widetilde{Y}_t}{\widetilde{K}_t}e^{a_t} \\
\widetilde{Y}_t &= H_t^{\alpha}\left(\widetilde{K}_t e^{-a_t}\right)^{1-\alpha}, \quad \widetilde{Y}_t = \widetilde{C}_t + \widetilde{I}_t, \quad \widetilde{K}_{t+1} = (1-\delta)\widetilde{K}_t e^{-a_t} + \widetilde{I}_t.
\end{aligned} \tag{61}$$

The process a_t is defined as

$$a_t = \ln\frac{A_t}{A_{t-1}} = \ln\gamma + (\rho_a - 1)\ln\widetilde{A}_{t-1} + \sigma_a\epsilon_{a,t}. \tag{62}$$

This log ratio is always stationary, because if $\rho_a = 1$ the $\ln\widetilde{A}_{t-1}$ term drops out. Finally, we stack the parameters of the DSGE model in the vector θ:

$$\theta = [\alpha, \beta, \gamma, \delta, \nu, \ln A_0, \rho_a, \sigma_a, \ln B_*, \rho_b, \sigma_b]'. \tag{63}$$

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 θ. For instance, the rental rate of capital, the capital-output, and the investment-output ratios are given by

$$R_* = \frac{\gamma}{\beta} - (1-\delta), \quad \frac{\widetilde{K}_*}{\widetilde{Y}_*} = \frac{(1-\alpha)\gamma}{R_*}, \quad \frac{\widetilde{I}_*}{\widetilde{Y}_*} = \left(1 - \frac{1-\delta}{\gamma}\right)\frac{\widetilde{K}_*}{\widetilde{Y}_*}. \tag{64}$$

In a stochastic environment, the detrended variables follow a stationary law of motion, even if the underlying technology shock is nonstationary. Moreover, if $\rho_a = 1$, the model generates a number of cointegration relationships, which according to (60) are obtained by taking pairwise 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

The solution to the equilibrium conditions (59), (61), and (62) leads to a probability distribution for the endogenous model variables, indexed by the vector of structural

parameters θ. 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 in terms of both the state and the control variables, and the solution of the optimization problems requires numerical techniques. The solution of the DSGE model can be written as

$$s_t = \Phi(s_{t-1}, \epsilon_t; \theta), \tag{65}$$

where s_t is a vector of suitably defined state variables and ϵ_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_* ($\widetilde{X}_*$), 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 (61) are given by the following system of linear expectational difference equations:

$$\begin{aligned}
\widehat{C}_t &= I\!E_t\left[\widehat{C}_{t+1} + \widehat{a}_{t+1} - \frac{R_*}{R_* + (1-\delta)}\widehat{R}_{t+1}\right] \\
\widehat{H}_t &= \nu\widehat{W}_t - \nu\widehat{C}_t + (1+\nu)\widehat{B}_t, \quad \widehat{W}_t = \widehat{Y}_t - \widehat{H}_t, \\
\widehat{R}_t &= \widehat{Y}_t - \widehat{K}_t + \widehat{a}_t, \quad \widehat{K}_{t+1} = \frac{1-\delta}{\gamma}\widehat{K}_t + \frac{\widetilde{I}_*}{\widetilde{K}_*}\widehat{I}_t - \frac{1-\delta}{\gamma}\widehat{a}_t, \\
\widehat{Y}_t &= \alpha\widehat{H}_t + (1-\alpha)\widehat{K}_t - (1-\alpha)\widehat{a}_t, \quad \widehat{Y}_t = \frac{\widetilde{C}_*}{\widetilde{Y}_*}\widehat{C}_t + \frac{\widetilde{I}_*}{\widetilde{Y}_*}\widehat{I}_t, \\
\widehat{A}_t &= \rho_a\widehat{A}_{t-1} + \sigma_a\epsilon_{a,t}, \quad \widehat{a}_t = \widehat{A}_t - \widehat{A}_{t-1}, \quad \widehat{B}_t = \rho_b\widehat{B}_{t-1} + \sigma_b\epsilon_{b,t}.
\end{aligned} \tag{66}$$

A multitude of techniques are available for solving linear rational expectations models (see, for instance, Sims (2002b)). Economists focus on solutions that guarantee a nonexplosive law of motion for the endogenous variables that appear in (66), with the loose justification that any explosive solution would violate the transversality conditions associated with the underlying dynamic optimization problems. For the neoclassical growth model, the solution takes the form

$$s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t. \tag{67}$$

The system matrices Φ_1 and Φ_ϵ are functions of the DSGE model parameters θ, 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.

Like all DSGE models, the linearized neoclassical growth model has some apparent counterfactual implications. Since fluctuations are generated by two exogenous disturbances, $\widehat{A}_t$ and $\widehat{B}_t$, the likelihood function for more than two variables is degenerate. The model predicts that certain linear combinations of variables, such as the labor share $\widehat{lsh} = \widehat{H}_t + \widehat{W}_t - \widehat{Y}_t$, are constant, which is clearly at odds with the data. To cope with this problem authors have added either so-called measurement errors, Sargent (1989), Altug (1989), and Ireland (2004), or additional shocks as in Leeper and Sims (1995) and more recently Smets and Wouters (2003). In the subsequent illustration, we restrict the dimension of the vector of observables y_t to $n = 2$, so that it matches the number of exogenous shocks. Our measurement equation takes the form

$$y_t = \Psi_0(\theta) + \Psi_1(\theta)t + \Psi_2(\theta)s_t. \tag{68}$$

Equations (67) and (68) provide a state-space representation for the linearized DSGE model. If the innovations ϵ_t are Gaussian, then the likelihood function can be obtained from the Kalman filter, which is described in detail in Giordani, Pitt, and Kohn (This Volume).

In the subsequent empirical illustration, we let y_t consist of log GDP and log hours worked. In this case, Equation (68) becomes

$$\begin{bmatrix} \ln GDP_t \\ \ln H_t \end{bmatrix} = \begin{bmatrix} \ln Y_0 \\ \ln H_* \end{bmatrix} + \begin{bmatrix} \ln \gamma \\ 0 \end{bmatrix} t + \begin{bmatrix} \widehat{Y}_t + \widehat{A}_t \\ \widehat{H}_t \end{bmatrix},$$

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. Notice that even though the DSGE model was solved in terms of the detrended model variable $\widehat{Y}_t$, the trend generated by technology $(\ln \gamma)t + \widehat{A}_t$ is added in the measurement equation. Thus, we are able to use nondetrended log real GDP as an observable and to learn about the technology growth rate γ and its persistence ρ_a from the available information about the level of output.

Although we focus on the dynamics of output and hours in this section, it is instructive to examine the measurement equations that the model yields for output

and investment. Suppose we use the GDP deflator to convert the two series depicted in Figure 3 from nominal into real terms. Then, we can write

$$\begin{bmatrix} \ln GDP_t \\ \ln I_t \end{bmatrix} = \begin{bmatrix} \ln Y_0 \\ \ln Y_0 + (\ln \widetilde{I}_* - \ln \widetilde{Y}_*) \end{bmatrix} + \begin{bmatrix} \ln \gamma \\ \ln \gamma \end{bmatrix} t + \begin{bmatrix} \widehat{A}_t + \widehat{Y}_t \\ \widehat{A}_t + \widehat{I}_t \end{bmatrix}.$$

This representation highlights the common trend in output and investment generated by the technology process $\widehat{A}_t$. If $\rho_a = 1$ then the last line of (66) implies that $\widehat{A}_t$ follows a random-walk process and hence induces nonstationary dynamics. In this case, the model implies the following cointegration relationship:

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

Recall that both $\widehat{Y}_t$ and $\widehat{I}_t$ are stationary, even if $\rho_a = 1$. We used this model implication in Section 3.2 as justification of our informative prior for the cointegration vector. In contrast, the posterior estimates of the cointegration vector reported in Illustration 3.1 suggest that the balanced-growth-path implication of the DSGE model is overly restrictive. In practice, such a model deficiency may lead to posterior distributions of the autoregressive coefficients associated with shocks other than technology that concentrate near unity.

4.3 Bayesian Inference

Although most of the literature on Bayesian estimation of DSGE models uses fairly informative prior distributions, 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 use other sources of information that do 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 important, the choice of prior should be properly documented.

For concreteness, suppose the neoclassical growth model is estimated based on aggregate output and hours data over the period 1955 to 2006. There are three important sources of information that are approximately independent of the data that enter the likelihood function and therefore 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; and (iii) macroeconomic data, including observations on output and hours worked, prior to 1955. Consider source (i). It is apparent from (64) that long-run averages of real interest rates, capital-output ratios, and investment-output ratios are informative about α, β, and δ. Moreover, the parameter α equals the labor share of income in our model. Since none of these variables directly enters the likelihood function, it is sensible to incorporate this information through the prior distribution. The parameters ρ_a, ρ_b, σ_a, and σ_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 presample evidence, that is, information from source (iii). Del Negro and Schorfheide (2008) provide an approach for automating this type of prior elicitation. Finally, microeconometric estimates of labor supply elasticities – an example of source (ii) – could be used to specify a prior for the Frisch elasticity ν, accounting for the fact that most of the variation in hours worked at the aggregate level is due to the extensive margin, that is, to individuals moving in and out of unemployment.

Because of the nonlinear relationship between the DSGE model parameters θ and the system matrices Ψ_0, Ψ_1, Ψ_2, Φ_1 and Φ_ϵ in (67) and (68), the marginal and conditional distributions of the elements of θ do not fall into the well-known families of probability distributions. Up to now, the most commonly used procedures for generating draws from the posterior distribution of θ are the Random-Walk Metropolis (RWM) Algorithm described in Schorfheide (2000) and Otrok (2001) or the Importance Sampler proposed in DeJong, Ingram, and Whiteman (2000). The basic RWM Algorithm takes the following form

Algorithm 4.1: Random-Walk Metropolis (RWM) Algorithm for DSGE Model

1. Use a numerical optimization routine to maximize the log posterior, which up to a constant is given by $\ln p(Y|\theta) + \ln p(\theta)$. Denote the posterior mode by $\tilde{\theta}$.

2. Let $\tilde{\Sigma}$ be the inverse of the (negative) Hessian computed at the posterior mode $\tilde{\theta}$, which can be computed numerically.

3. Draw $\theta^{(0)}$ from $N(\tilde{\theta}, c_0^2\tilde{\Sigma})$ or directly specify a starting value.

4. For $s = 1, \ldots, n_{sim}$: draw ϑ from the proposal distribution $N(\theta^{(s-1)}, c^2\tilde{\Sigma})$. 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,

$$r(\theta^{(s-1)}, \vartheta|Y) = \frac{p(Y|\vartheta)p(\vartheta)}{p(Y|\theta^{(s-1)})p(\theta^{(s-1)})}. \quad \square$$

If the likelihood can be evaluated with a high degree of precision, then the maximization in Step 1 can be implemented with a gradient-based numerical optimization routine. The optimization is often not straightforward as the posterior density is typically not globally concave. Thus, it is advisable to start the optimization routine from multiple starting values, which could be drawn from the prior distribution, and then set $\tilde{\theta}$ to the value that attains the highest posterior density across optimization runs.

The evaluation of the likelihood typically involves three steps: (i) the computation of the steady state; (ii) the solution of the linear rational expectations system; and (iii) the evaluation of the likelihood function of a linear state-space model with the Kalman filter. While the computation of the steady states is trivial in our neoclassical stochastic growth model, it might require the use of numerical equation solvers for more complicated DSGE models. Any inaccuracy in the computation of the steady states will translate into an inaccurate evaluation of the likelihood function that makes use of gradient-based optimization methods impractical. Chib and Ramamurthy (2010) recommend using a simulated annealing algorithm for Step 1. In some applications we found it useful to skip Steps 1 to 3 by choosing a reasonable starting value, such as the mean of the prior distribution, and replacing $\tilde{\Sigma}$ in Step 4 with a matrix whose diagonal elements are equal to the prior variances of the DSGE model parameters and whose off-diagonal elements are zero.

Based on practitioners' experience, Algorithm 4.1 tends to work well if the posterior density is unimodal. The scale factor c_0 controls the expected distance between the mode and the starting point of the Markov chain. The tuning parameter c is typically chosen to obtain a rejection rate of about 50%. In this case, reasonable perturbations of the starting points lead to chains that after 100,000 to 1,000,000 iterations provide very similar approximations of the objects of interest, for example posterior means, medians, standard deviations, and credible sets. An and Schorfheide (2007b) describe a hybrid MCMC algorithm with transition mixture to deal with a bimodal posterior distribution. Most recently, Chib and Ramamurthy

(2010) have developed a multiblock 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 Chib (This Volume).

Illustration 4.1: The prior distribution for our empirical illustration is summarized in the first five columns of Table 3. Based on National Income and Product Account (NIPA) data, published by the Bureau of Economic Analysis, we choose the prior means for α, β, and δ to be consistent with a labor share of 0.66, an investment-to-output ratio of about 25%, and an annual interest rate of 4%. These choices yield values of $\underline{\alpha} = 0.66$, $\underline{\beta} = 0.99$, and $\underline{\delta} = 0.025$ in quarterly terms. As is quite common in the literature, we decided to use dogmatic priors for β and δ. Fixing these parameters is typically justified as follows. Conditional on the adoption of a particular data definition, the relevant long-run averages computed from NIPA data appear to deliver fairly precise measurements of steady-state relationships that can be used to extract information about parameters such as β and δ, resulting in small prior variances. The use of a dogmatic prior can then be viewed as a (fairly good) approximation of a low-variance prior. For illustrative purpose, we use such a low-variance prior for α. We assume that α has a Beta distribution with a standard deviation of 0.02.

An important parameter for the behavior of the model is the labor supply elasticity. As discussed in Ríos-Rull, Schorfheide, Fuentes-Albero, Kryshko, and Santaeulalia-Llopis (2009), *a priori* plausible values 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 (1988) 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 that 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 3 are marginal distributions. A joint prior is typically obtained by taking the product of the marginals for all elements of θ, 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) propose to multiply an initial prior $\tilde{p}(\theta)$ constructed from marginal distributions for the individual elements of θ by a function $f(\theta)$ that reflects beliefs about steady-state relationships and autocovariances. This function is generated by interpreting long-run averages of variables that do not appear in the model and presample autocovariances of y_t as noisy measures of steady states and population autocovariances. For example, let $lsh_*(\theta)$ be the model-implied labor share as a function of θ and $\widehat{lsh}$ a sample average of postwar U.S. labor shares. Then $\ln f(\theta)$ could be defined as $-(lsh_*(\theta) - \widehat{lsh})^2/(2\lambda)$, where λ reflects the strength of the belief about the labor share. The overall prior then takes the form $p(\theta) \propto \tilde{p}(\theta)f(\theta)$.

The prior distribution is updated based on quarterly data on aggregate output and hours worked ranging from 1955 to 2006. Unlike in Figure 6, we do not remove a deterministic trend from the output series. We apply the RWM Algorithm to generate 100,000 draws from the posterior distribution of the parameters of the stochastic growth model. The scale parameter in the proposal density is chosen to be $c = 0.5$, which leads to a rejection rate of about 50%. Posterior means and 90% credible intervals, computed from the output of the posterior simulator, are summarized in the last four columns of Table 3. 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 lie 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 α is essentially not updated in view of the data. The posterior means of the labor supply elasticity are 0.42 and 0.70, respectively, which is in line with the range of estimates reported in Ríos-Rull, Schorfheide, Fuentes-Albero, Kryshko, and Santaeulalia-Llopis (2009). These relatively small values of ν 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 1.1% for the technology shock and 0.7% for the preference shock. We used a logarithmic transformation of γ, which can be interpreted as the average quarterly growth rate of the economy and is estimated

to be 0.3% to 0.4%. The estimates of $\ln H_*$ and $\ln Y_0$ capture the level of the two series. Once draws from the posterior distribution have been generated, they can be converted into other objects of interest such as responses to structural shocks. □

4.4 Extensions I: Indeterminacy

Linear rational expectations systems can have multiple stable solutions, and this 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) estimated interest rate feedback rules based on U.S. postwar 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

$$y_t = \frac{1}{\theta}\mathbb{E}_t[y_{t+1}] + \epsilon_t, \quad \epsilon_t \sim iidN(0,1), \quad \theta \in (0,2]. \tag{69}$$

Here, θ should be interpreted as the structural parameter, which is scalar. It can be verified that if, on the one hand, $\theta > 1$, the unique stable equilibrium law of motion of the endogenous variable y_t is given by

$$y_t = \epsilon_t. \tag{70}$$

If, on the other hand, $\theta \leq 1$, one obtains a much larger class of solutions that can be characterized by the ARMA(1,1) process

$$y_t = \theta y_{t-1} + (1+M)\epsilon_t - \theta\epsilon_{t-1}. \tag{71}$$

Here, the scalar parameter $M \in \mathbb{R}$ is used to characterize all stationary solutions of (69). M is completely unrelated to the agents' tastes and technologies characterized by θ, but it does affect the law of motion of y_t if $\theta \leq 1$. From a macroeconomist's perspective, M captures an indeterminacy: based on θ alone, the law of motion of y_t is not uniquely determined.

Table 3: 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.
α	$[0,1)$	Beta	0.66	0.02	0.65	[0.62, 0.68]	0.65	[0.63, 0.69]
ν	$\mathbb{R}^+$	Gamma	2.00	1.00	0.42	[0.16, 0.67]	0.70	[0.22, 1.23]
$4\ln\gamma$	$\mathbb{R}$	Normal	0.00	0.10	.003	[.002, .004]	.004	[.002, .005]
ρ_a	$\mathbb{R}^+$	Beta	0.95	0.02	0.97	[0.95, 0.98]	1.00	
σ_a	$\mathbb{R}^+$	InvGamma	0.01	4.00	.011	[.010, .012]	.011	[.010, .012]
ρ_b	$\mathbb{R}^+$	Beta	0.80	0.10	0.98	[0.96, 0.99]	0.98	[0.96, 0.99]
σ_b	$\mathbb{R}^+$	InvGamma	0.01	4.00	.008	[.007, .008]	.007	[.006, .008]
$\ln H_*$	$\mathbb{R}$	Normal	0.00	10.0	-0.04	[-0.08, 0.01]	-0.03	[-0.07, 0.02]
$\ln Y_0$	$\mathbb{R}$	Normal	0.00	100	8.77	[8.61, 8.93]	8.39	[7.93, 8.86]

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 ν for the Inverted Gamma distribution, where $p_{IG}(\sigma|\nu,s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. To estimate the stochastic growth version of the model we set $\rho_a = 1$. The parameters $\beta = 0.99$ and $\delta = 0.025$ are fixed. □

From an econometrician's perspective, one needs to introduce this auxiliary parameter M to construct the likelihood function. The likelihood function has the following features. According to (70), the likelihood function is completely flat (does not vary with θ and M) for $\theta > 1$ because all parameters drop from the equilibrium law of motion. If $\theta \leq 1$ and $M = 0$ the likelihood function does not vary with θ because the roots of the autoregressive and the moving-average polynomial in the ARMA(1,1) process (71) cancel. If $\theta \leq 1$ and $M \neq 0$, then the likelihood function exhibits curvature. In a Bayesian framework, this irregular shape of the likelihood function does not pose any conceptual challenge. In principle, one can combine proper priors for θ and M and obtain a posterior distribution. However, in more realistic applications the implementation of posterior simulation procedures require extra care. Lubik and Schorfheide (2004) divided the parameter space into Θ_D and Θ_I (for model (69) $\Theta_D = (1, 2]$ and $\Theta_D = [0, 1]$) along the lines of the determinacy-indeterminacy boundary, treated the subspaces as separate models, generated posterior draws for each subspace separately, and used marginal likelihoods to obtain posterior probabilities for Θ_D and Θ_I.

4.5 Extensions II: Stochastic Volatility

One of the most striking features of postwar U.S. GDP data is the reduction in the volatility of output growth around 1984. This phenomenon has been 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 ϵ_t in (67) to vary stochastically over time. The authors adopt a specification in which log standard deviations evolve according to an autoregressive process. An alternative approach would be to capture the Great Moderation with Markov-switching shock standard deviations (see Section 5).

In the context of the stochastic growth model, consider for instance the technology shock $\epsilon_{a,t}$. We previously assumed in (58) that $\epsilon_{a,t} \sim N(0, 1)$. Alternatively, suppose that

$$\epsilon_{a,t} \sim N(0, v_t^2), \quad \ln v_t = \rho_v \ln v_{t-1} + \eta_t, \quad \eta_t \sim iidN(0, \omega^2). \tag{72}$$

Justiniano and Primiceri (2008) solved the linear rational expectational system obtained from the log-linearized equilibrium conditions of their DSGE model and then augmented the linear solution by equations that characterize the stochastic volatility of the exogenous structural shocks. Their approach amounts to using (67) and

assuming that the element $\epsilon_{a,t}$ in the shock vector ϵ_t evolves according to (72). The following Gibbs sampler can be used to generate draws from the posterior distribution.

Algorithm 4.2: Metropolis-within-Gibbs Sampler for DSGE Model with Stochastic Volatility

For $s = 1, \ldots, n_{sim}$:

1. Draw $\theta^{(s)}$ conditional on $(\theta|v_{1:T}^{(s-1)}, Y)$. Given the sequence $v_{1:T}^{(s-1)}$ the likelihood function of the state-space model can be evaluated with the Kalman filter. Consequently, the RWM step described in Algorithm 4.1 can be used to generate a draw $\theta^{(s)}$.

2. Draw $\epsilon_{a,1:T}^{(s)}$ conditional on $(\theta^{(s)}, v_{1:T}^{(s-1)}, Y)$ using the simulation smoother of Carter and Kohn (1994), described in Giordani, Pitt, and Kohn (This Volume).

3. Draw $(\rho_v^{(s)}, \omega^{(s)})$ conditional on $(v_{1:T}^{(s-1)}, Y)$ from the Normal-Inverse Gamma posterior obtained from the AR(1) law of motion for $\ln v_t$ in (72).

4. Draw $v_{1:T}^{(s)}$ conditional on $(\epsilon_{a,1:T}^{(s)}, \rho_v^{(s)}, \omega^{(s)}, Y)$. Notice that (72) can be interpreted as a 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) and are discussed in Jacquier and Polson (This Volume) and Giordani, Pitt, and Kohn (This Volume). □

The empirical model of Justiniano and Primiceri (2008) ignores any higher-order dynamics generated from the nonlinearities of the DSGE model itself on grounds of computational ease. As we will see in the next subsection, Bayesian inference is more difficult to implement for DSGE models solved with nonlinear techniques.

4.6 Extension III: General Nonlinear DSGE Models

DSGE models are inherently nonlinear, as can be seen from the equilibrium conditions (61) associated with our stochastic growth model. Nonetheless, given the magnitude of the business-cycle fluctuations of a country like the United States or the Euro area, many researchers take the stand that the equilibrium dynamics are

well approximated by a linear state-space system. 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 a gross return $R_{j,t}$, the linearized consumption Euler equation takes the form

$$\widehat{C}_t = I\!E_t\left[\widehat{C}_{t+1} + \widehat{a}_{t+1} - \widehat{R}_{j,t+1}\right], \tag{73}$$

implying that all assets yield the same expected return. Thus, log-linear approximations have the undesirable feature (for asset-pricing applications) that risk premiums disappear.

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 models is to use second-order perturbation methods. A comparison of solution methods for DSGE models can be found in Aruoba, Fernández-Villaverde, and Rubio-Ramírez (2004). Second, the evaluation of the likelihood function becomes more costly because both the state transition equation and the measurement equation of the state-space model are nonlinear. Thus, (67) and (68) are replaced by (65) and

$$y_t = \Psi(s_t; \theta). \tag{74}$$

Fernández-Villaverde and Rubio-Ramírez (2007) and Fernández-Villaverde and Rubio-Ramírez (2008) show how a particle filter can be used to evaluate the likelihood function associated with a DSGE model. A detailed description of the particle filter is provided in Giordani, Pitt, and Kohn (This Volume).

Bayesian analysis of nonlinear DSGE models is currently an active area of research and faces a number of difficulties that have not yet been fully resolved. For the particle filter to work in the context of the stochastic growth model described above, the researcher has to introduce measurement errors in (74). Suppose that $\{s_{t-1}^{(i)}\}_{i=1}^N$ is a collection of particles whose empirical distribution approximates $p(s_{t-1}|Y_{1:t-1}, \theta)$. Without errors in the measurement equation, a proposed particle $\tilde{s}_t^{(i)}$ has to satisfy the following two equations:

$$y_t = \Psi(\tilde{s}_t^{(i)}; \theta) \tag{75}$$

$$\tilde{s}_t^{(i)} = \Phi(s_{t-1}^{(i)}, \epsilon_t^{(i)}; \theta). \tag{76}$$

If $\tilde{s}_t^{(i)}$ is sampled from a continuous distribution, the probability that (75) is satisfied is zero. Thus, in the absence of measurement errors, $\tilde{s}_t^{(i)}$ needs to be sampled from a discrete distribution. One can plug (76) into (75), eliminating $\tilde{s}_t^{(i)}$, and then find all real solutions $\tilde{\epsilon}$ of ϵ for the equation $y_t = \Psi(\Phi(s_{t-1}^{(i)}, \epsilon; \theta); \theta)$. Based on the $\tilde{\epsilon}'s$, one can obtain the support points for the distribution of $\tilde{s}_t^{(i)}$ as $\Phi(s_{t-1}^{(i)}, \tilde{\epsilon}; \theta)$. In practice, this calculation is difficult if not infeasible to implement, because the nonlinear equation might have multiple solutions.

If errors $\eta_t \sim N(0, \Sigma_\eta)$ are added to the measurement equation (74), which in the context of our stochastic growth model amounts to a modification of the DSGE model, then (75) turns into

$$y_t = \Psi(\tilde{s}_t^{(i)}; \theta) + \eta_t. \tag{77}$$

This equation 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 η_t that are small relative to Σ_η. Some authors – referring to earlier work by Sargent (1989), Altug (1989), or Ireland (2004) – make measurement errors part of the specification of their empirical model. In this case, it is important to realize that one needs to bound the magnitude of the measurement error standard deviations from below to avoid a deterioration of the particle filter performance as these standard deviations approach zero.

4.7 DSGE Model Evaluation

An important aspect of empirical work with DSGE models is the evaluation of fit. We will distinguish three approaches. First, a researcher might be interested in assessing whether the fit of a stochastic growth model improves if one allows for convex investment adjustment costs. Posterior odds of a model with adjustment costs versus a model without are useful for such an assessment. Second, one could examine to what extent a 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. Finally, a researcher might want to compare one or more DSGE models to a more flexible reference model such as a VAR. We consider three methods of doing so. Such comparisons can be used to examine whether a particular

DSGE model captures certain important features of the data. Alternatively, they can be used to rank different DSGE model specifications.

4.7.1 Posterior Odds

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

$$\frac{\pi_{i,T}}{\pi_{j,T}} = \frac{\pi_{i,0}}{\pi_{j,0}} \times \frac{p(Y|\mathcal{M}_i)}{p(Y|\mathcal{M}_j)}. \tag{78}$$

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

$$p(Y|\mathcal{M}_i) = \int p(Y|\theta_{(i)}, \mathcal{M}_i) p(\theta_{(i)}) d\theta_{(i)} \tag{79}$$

is the marginal likelihood function. The key challenge in posterior odds comparisons is the computation of the marginal likelihood that 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) and Chib and Jeliazkov (2001) 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-Ramírez (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. Section 7 provides a more detailed discussion of model selection and model averaging based on posterior probabilities.

Illustration 4.2: We previously estimated two versions of the neoclassical stochastic growth model: a version with a trend-stationary technology process and a version with a difference-stationary exogenous productivity process. The log-marginal data densities $\ln p(Y|\mathcal{M}_i)$ are 1392.8 and 1395.2, respectively. If the prior probabilities for the two specifications are identical, these marginal data densities imply that the posterior probability of the difference-stationary specification is approximately 90%. □

4.7.2 Predictive Checks

A general discussion of the role of predictive checks in Bayesian analysis can be found in Lancaster (2004), Geweke (2005), and Geweke (2007). Predictive checks

can be implemented based on either the prior or the posterior distribution of the DSGE model parameters θ. Let $Y^*_{1:T}$ be a hypothetical sample of length T. The predictive distribution for $Y^*_{1:T}$ based on the time t information set $\mathcal{F}_t$ is

$$p(Y^*_{1:T}|\mathcal{F}_t) = \int p(Y^*_{1:T}|\theta)p(\theta|\mathcal{F}_t)d\theta. \tag{80}$$

We can then use $\mathcal{F}_0$ to denote the prior information and $\mathcal{F}_T$ to denote the posterior information set that includes the sample $Y_{1:T}$. Draws from the predictive distribution can be obtained in two steps. First, generate a parameter draw $\tilde{\theta}$ from $p(\theta|\mathcal{F}_t)$. Second, simulate a trajectory of observations $Y^*_{1:T}$ from the DSGE model conditional on $\tilde{\theta}$. The simulated trajectories can be converted into sample statistics of interest, $\mathcal{S}(Y^*_{1:T})$, 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 value of the statistic $\mathcal{S}(Y_{1:T})$ based on the actual data and assess how far it lies in the tails of its predictive distribution. If $\mathcal{S}(Y_{1:T})$ 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. Because the prior predictive distribution conveys the implications of models without having to develop methods for formal posterior inference, prior predictive checks can be very useful at an early stage of model development. Canova (1994) was the first author to use prior predictive checks to assess implications of a stochastic growth model driven solely by a technology shock. Prior predictive distributions are closely related to marginal likelihoods. A comparison of (79) and (80) for $t = 0$ indicates that the two expressions are identical. In its implementation, the prior predictive check replaces $Y^*_{1:T}$ in (80) with $Y_{1:T}$ and tries to measure whether the density that the Bayesian model assigns *a priori* to the observed data is high or low. One can make the procedure more easily interpretable by replacing the high-dimensional data matrix Y with a low-dimensional statistic $\mathcal{S}(Y)$.

In posterior predictive checks, the distribution of the parameters, $p(\theta|\mathcal{F}_T)$, is conditioned on the observed data $Y_{1:T}$. In its core, the posterior predictive check works like a frequentist specification test. If $\mathcal{S}(Y_{1:T})$ falls into the tails (or low-density region) of the predictive distribution derived from the estimated model, then the model is discredited. Chang, Doh, and Schorfheide (2007) use posterior predictive checks to determine whether a stochastic growth model, similar to the

one analyzed in this section, is able to capture the observed persistence of hours worked.

4.7.3 VARs as Reference Models

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

Models of Moments: Geweke (2010) points out that many DSGE models are too stylized to deliver a realistic distribution for the data Y that is usable 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 two variables. Suppose we collect these population moments in the vector φ, which in turn is a function of the DSGE model parameters θ. Thus, a prior distribution for θ 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 φ. Let $p(\varphi|Y,\mathcal{M}_0)$ denote the posterior distribution of population characteristics as obtained from the VAR. Geweke (2010) shows that

$$\frac{\pi_{1,0}\int p(\varphi|\mathcal{M}_1)p(\varphi|Y,\mathcal{M}_0)d\varphi}{\pi_{2,0}\int p(\varphi|\mathcal{M}_2)p(\varphi|Y,\mathcal{M}_0)d\varphi} \tag{81}$$

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 (81) is large, if there is a strong overlap between the predictive densities for φ 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) and has been used, for instance, to examine asset-pricing implications of DSGE models. In practice, the densities $p(\varphi|\mathcal{M}_i)$ and $p(\varphi|Y,\mathcal{M}_0)$ can be approximated by Kernel density estimates based on draws of φ. Draws of φ can be obtained by transforming draws of the DSGE model and VAR parameters, respectively.

Loss-Function-Based Evaluation: Schorfheide (2000) proposes a Bayesian framework for a loss function-based evaluation of DSGE models. As in Geweke (2010)'s framework, the researcher is interested in the relative ability of two DSGE models to capture a certain set of population moments φ, which are transformations of model

parameters θ. Unlike in Geweke (2010), the DSGE models are assumed to deliver a probability distribution for the data Y. Suppose there are two DSGE models, $\mathcal{M}_1$ and $\mathcal{M}_2$, and a VAR that serves as a 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 $\pi_{i,T}$, $i=0,1,2$. Second, one can form a predictive density for φ by averaging across the three models

$$p(\varphi|Y) = \sum_{i=0,1,2} \pi_{i,T} p(\varphi|Y,\mathcal{M}_i). \tag{82}$$

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, however, none of the DSGE models fits well, then the predictive density is dominated by the VAR. Third, one specifies a loss function $L(\hat{\varphi},\varphi)$, for example $L(\hat{\varphi},\varphi) = \|\hat{\varphi}-\varphi\|^2$, under which a point prediction $\hat{\varphi}$ of φ is to be evaluated. For each DSGE model, the prediction $\hat{\varphi}_{(i)}$ is computed by minimizing the expected loss under the DSGE model-specific posterior:

$$\hat{\varphi}_{(i)} = \text{argmin}_{\tilde{\varphi}} \int L(\tilde{\varphi},\varphi) p(\varphi|Y,\mathcal{M}_i) d\varphi, \quad i=1,2.$$

Finally one can compare DSGE models $\mathcal{M}_1$ and $\mathcal{M}_2$ based on the posterior expected loss $\int L(\hat{\varphi}_{(i)},\varphi)p(\varphi|Y)d\varphi$, computed under the overall posterior distribution (82) that averages the predictions of the reference model and all DSGE models. In this procedure, 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. The starting point is the VAR specified in Equation (1). Assuming that the data have been transformed such that y_t is stationary, let $\mathbb{E}_\theta^D[\cdot]$ be the expectation under the DSGE model conditional on parameterization θ 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:

$$\Phi^*(\theta) = \Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta), \quad \Sigma^*(\theta) = \Gamma_{YY}(\theta) - \Gamma_{YX}(\theta)\Gamma_{XX}^{-1}(\theta)\Gamma_{XY}(\theta). \tag{83}$$

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 Φ and Σ from the restriction functions:

$$\Phi, \Sigma|\theta \sim MNIW\Big(\Phi^*(\theta), [\lambda T \Gamma_{XX}(\theta)]^{-1}, \lambda T \Sigma^*(\theta), \lambda T - k\Big). \tag{84}$$

This prior distribution can be interpreted as a posterior calculated from a sample of $T^* = \lambda T$ artificial observations generated from the DSGE model with parameters θ. Here, λ is a hyperparameter, and T denotes the actual sample size.

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 ϵ_t, that is $u_t = \Sigma_{tr}\Omega\epsilon_t$, see (21). Let $A_0(\theta)$ be the contemporaneous impact of ϵ_t on y_t according to the DSGE model. With a QR factorization, the initial response of y_t to the structural shocks can be uniquely decomposed into

$$\left(\frac{\partial y_t}{\partial \epsilon_t'}\right)_{DSGE} = A_0(\theta) = \Sigma_{tr}^*(\theta)\Omega^*(\theta), \tag{85}$$

where $\Sigma_{tr}^*(\theta)$ is lower-triangular and $\Omega^*(\theta)$ is an orthogonal matrix. The initial impact of ϵ_t on y_t in the VAR, in contrast, is given by

$$\left(\frac{\partial y_t}{\partial \epsilon_t'}\right)_{VAR} = \Sigma_{tr}\Omega. \tag{86}$$

To identify the DSGE-VAR, we maintain the triangularization of its covariance matrix Σ and replace the rotation Ω in (86) with the function $\Omega^*(\theta)$ that appears in (85). 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 posterior short-run responses of the VAR with those from the DSGE model.

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

$$p_\lambda(Y, \Phi, \Sigma, \theta) = p(Y|\Phi, \Sigma)p_\lambda(\Phi, \Sigma|\theta)p(\Omega|\theta)p(\theta), \tag{87}$$

with the understanding that the distribution of $\Omega|\theta$ is a point mass at $\Omega^*(\theta)$. Since Φ and Σ can be conveniently integrated out, we can first draw from the marginal

posterior of θ and then from the conditional distribution of (Φ, Σ) given θ. This leads to the following algorithm.

Algorithm 4.3: Posterior Draws for DSGE-VAR

1. Use Algorithm 4.1 to generate a sequence of draws $\theta^{(s)}$, $s = 1, \ldots, n_{sim}$, from the posterior distribution of θ, given by $p_\lambda(\theta|Y) \propto p_\lambda(Y|\theta)p(\theta)$. The marginal likelihood $p_\lambda(Y|\theta)$ is obtained by straightforward modification of (15). Moreover, compute $\Omega^{(s)} = \Omega^*(\theta^{(s)})$.

2. For $s = 1, \ldots, n_{sim}$: draw a pair $(\Phi^{(s)}, \Sigma^{(s)})$ from its conditional MNIW posterior distribution given $\theta^{(s)}$. The MNIW distribution can be obtained by the modification of (8) described in Section 2.2. $\square$

Since the empirical performance of the DSGE-VAR procedure crucially depends on the weight placed on the DSGE model restrictions, it is useful to consider a data-driven procedure to select λ. As in the context of the Minnesota prior, a natural criterion for the choice of λ is the marginal data density

$$p_\lambda(Y) = \int p_\lambda(Y|\theta)p(\theta)d\theta. \tag{88}$$

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

$$\hat{\lambda} = \text{argmax}_{\lambda \in \Lambda}\, p_\lambda(Y). \tag{89}$$

If $p_\lambda(Y)$ peaks at an intermediate value of λ, 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 also been used as a tool for model evaluation and comparison in Del Negro, Schorfheide, Smets, and Wouters (2007) and for policy analysis with potentially misspecified DSGE models in Del Negro and Schorfheide (2009).

4.8 DSGE Models in Applied Work

Much of the empirical analysis with DSGE models is conducted with Bayesian methods. Since the literature is fairly extensive and rapidly growing, we do not attempt to provide a survey of the empirical work. Instead, we will highlight a few important contributions and discuss how Bayesian analysis has contributed to the proliferation of estimated DSGE models. The first published papers that conduct Bayesian inference in DSGE models are DeJong, Ingram, and Whiteman (2000), Schorfheide (2000), and Otrok (2001). Smets and Wouters (2003) document that a DSGE model that is built around the neoclassical growth model presented previously and enriched by habit formation in consumption, capital adjustment costs, variable factor utilization, nominal price and wage stickiness, behavioral rules for government spending and monetary policy, and numerous exogenous shocks could deliver a time-series fit and forecasting performance for a vector of key macroeconomic variables that is comparable to a VAR. Even though posterior odds comparison, literally taken, often favor VARs, the theoretical coherence and the ease with which model implications can be interpreted make DSGE models an attractive competitor.

One reason for the rapid adoption of Bayesian methods is the ability to incorporate nonsample information, meaning data that do not enter the likelihood function, through the use of prior distributions. Many of the priors used by Smets and Wouters (2003) as well as in subsequent work are fairly informative, and over the past five years the literature has become more careful about systematically documenting the specification of prior distributions in view of the available nonsample information. From a purely computational perspective, this kind of prior information often tends to smooth out the shape of the posterior density, which improves the performance of posterior simulators. Once parameter draws have been obtained, they can be easily converted into objects of interest. For instance, Justiniano, Primiceri, and Tambalotti (2009) study the relative importance of investment-specific technology shocks and thereby provide posterior distributions of the fraction of the business-cycle variation of key macroeconomic variables explained by these shocks.

A large part of the literature tries to assess the importance of various propagation mechanisms that are useful for explaining observed business-cycle fluctuations. Bayesian posterior model probabilities are widely employed to compare competing model specifications. For instance, Rabanal and Rubio-Ramírez (2005) compare the relative importance of wage and price rigidities. Unlike standard frequentist likeli-

hood ratio tests, posterior odds remain applicable, even if the model specifications under consideration are nonnested, for example, a DSGE model with sticky wages versus a DSGE model with sticky prices.

DSGE models with nominal rigidities are widely used to analyze monetary policy. This analysis might consist of determining the range of policy rule coefficients that guarantees a unique stable rational expectations solution and suppresses self-fulfilling expectations, of choosing interest-rate feedback rule parameters that maximize the welfare of a representative agent or minimizes a convex combination of inflation and output-gap volatility, or in finding a welfare-maximizing mapping between the underlying state variables of the economy and the policy instruments. The solution of these optimal policy problems always depends on the unknown taste and technology parameters. The Bayesian framework enables researchers and policy makers to take this parameter uncertainty into account by maximizing posterior expected welfare. A good example of this line of work is the paper by Levin, Onatski, Williams, and Williams (2006). Several central banks have adopted DSGE models as tools for macroeconomic forecasting, for example, Adolfson, Lindé, and Villani (2007) and Edge, Kiley, and Laforte (2009). An important advantage of the Bayesian methods described in this section is that they deliver predictive distributions for the future path of macroeconomic variables that reflect both parameter uncertainty and uncertainty about the realization of future exogenous shocks.

5 Time-Varying Parameters Models

The parameters of the models presented in the preceding sections were assumed to be time-invariant, implying that economic relationships are stable. In Figure 7, we plot quarterly U.S. GDP-deflator inflation from 1960 to 2006. Suppose one adopts the view that the inflation rate can be decomposed into 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 and 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 trade-off between unemployment and inflation and gradually revised its target upward. In the early 1980s, however, it realized that the long-run

Phillips curve is essentially vertical and that the high inflation had led to a significant distortion of the economy. Under the chairmanship of Paul Volcker, the Fed decided to disinflate, that is, to reduce the target inflation rate. This time-variation in the target rate could be captured either by a slowly-varying autoregressive process or through a regime-switching process that shifts from a 2.5% target to a 7% target and back.

This section considers models that can capture structural changes in the economy. Model parameters 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 discussed subsequently can be written in state-space form, and much of the technical apparatus needed for Bayesian inference can be found in Giordani, Pitt, and Kohn (This Volume). We focus on placing the TVP models in the context of the empirical macroeconomics literature and discuss specific applications in Section 5.3. There are other important classes of nonlinear time-series models such as threshold vector autoregressive models, Geweke and Terui (1993) and Koop and Potter (1999), for instance, in which the parameter change is linked directly to observables rather than to latent state variables. Due to space constraints, we are unable to discuss these models in this chapter.

5.1 Models with Autoregressive Coefficients

Most of the subsequent discussion is devoted to VARs with parameters that follow an autoregressive law of motion (section 5.1.1). Whenever time-varying parameters 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 parameters 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

While VARs with time-varying coefficients were estimated with Bayesian methods almost two decades ago, see, for instance, Sims (1993), their current popularity in

empirical macroeconomics is largely due to Cogley and Sargent (2002), who took advantage of the MCMC innovations in the 1990s. They estimated 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 time-varying features of business cycles in the United States and other countries.

Cogley and Sargent (2002) set out to investigate time-variation in US inflation persistence using a three-variable VAR with inflation, unemployment, and interest rates. The rationale for their reduced-form specification is provided by models in which the policy maker and/or agents in the private sector gradually learn about the dynamics of the economy and consequently adapt their behavior (see Sargent (1999)). 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 captured by models in which the coefficients are allowed to vary in each period. Cogley and Sargent (2002)'s work was criticized by Sims (2002a), 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 (2005b) address this criticism of their earlier work by adding time-varying volatility to their model. Our subsequent exposition of a TVP VAR allows for drifts in both the conditional mean and the variance parameters.

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

$$y_t = X'_t \phi_t + u_t, \tag{90}$$

where we replaced the vector of constant coefficients, ϕ with a vector of time-varying coefficients, ϕ_t. We let the parameters evolve according to the random-walk process:

$$\phi_t = \phi_{t-1} + \nu_t, \quad \nu_t \sim iidN(0, Q). \tag{91}$$

We restrict the covariance matrix Q to be diagonal and the parameter innovations ν_t to be uncorrelated with the VAR innovations u_t. The u_t innovations are also

normally distributed, but unlike in Section 2, their variance now evolves over time:

$$u_t \sim N(0, \Sigma_t), \quad \Sigma_t = B^{-1} H_t (B^{-1})'. \tag{92}$$

In the decomposition of Σ_t, the matrix B is a lower-triangular matrix with ones on the diagonal, and H_t is a diagonal matrix with elements $h_{i,t}^2$ following a geometric random walk:

$$\ln h_{i,t} = \ln h_{i,t-1} + \eta_{i,t}, \quad \eta_{i,t} \sim iidN(0, \sigma_i^2). \tag{93}$$

Notice that this form of stochastic volatility was also used in Section 4.5 to make the innovation variances for shocks in DSGE models time varying.

The prior distributions for Q and the σ_i's can be used to express beliefs about the magnitude of the period-to-period drift in the VAR coefficients and the changes in the volatility of the VAR innovations. In practice these priors are chosen to ensure that the shocks to (91) and (93) are small enough that the short- and medium-run dynamics of y_t are not swamped by the random-walk behavior of ϕ_t and H_t. If the prior distributions for ϕ_0, Q, B, and the σ_i's are conjugate, then one can use the following Gibbs sampler for posterior inference.

Algorithm 5.1: Gibbs Sampler for TVP VAR

For $s = 1, \ldots, n_{sim}$:

1. Draw $\phi_{1:T}^{(s)}$ conditional on $(B^{(s-1)}, H_{1:T}^{(s-1)}, Q^{(s-1)}, \sigma_1^{(s-1)} \ldots \sigma_n^{(s-1)}, Y)$. (90) and (91) provide a state-space representation for y_t. Thus, $\phi_{1:T}$ can be sampled using the algorithm developed by Carter and Kohn (1994), described in Giordani, Pitt, and Kohn (This Volume).

2. Draw $B^{(s)}$ conditional on $(\phi_{1:T}^{(s)}, H_{1:T}^{(s-1)}, Q^{(s-1)}, \sigma_1^{(s-1)} \ldots \sigma_n^{(s-1)}, Y)$. Conditional on the VAR parameters ϕ_t, the innovations to equation (90) are known. According to (92), Bu_t is normally distributed with variance H_t:

$$Bu_t = H_t^{\frac{1}{2}} \epsilon_t, \tag{94}$$

   where ϵ_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 (30) described in detail in Section 2.4.2.

3. Draw $H_{1:T}^{(s)}$ conditional on $(\phi_{1:T}^{(s)}, B^{(s)}, Q^{(s-1)}, \sigma_1^{(s-1)} \dots \sigma_n^{(s-1)}, Y)$. Conditional on ϕ_t and B, we can write the i'th equation of (94) as $z_{i,t} = B_{(i.)}u_t \sim N(0, h_{i,t}^2)$, which is identical to (72). Thus, as in Section 4.5, one can use the algorithms of Jacquier, Polson, and Rossi (1994) or Kim, Shephard, and Chib (1998) to draw the sequences $h_{i,t:T}$.

4. Draw $Q^{(s)}$ conditional on $(\phi_{1:T}^{(s)}, B^{(s)}, H_{1:T}^{(s)}, \sigma_1^{(s-1)} \dots \sigma_n^{(s-1)}, Y)$ from the appropriate Inverted Wishart distribution derived from (91).

5. Draw $\sigma_1^{(s)} \dots \sigma_n^s$ conditional on $(\phi_{1:T}^{(s)}, B^{(s)}, H_{1:T}^{(s)}, Q^{(s)}, Y)$ from the appropriate Inverted Gamma distributions derived from (93). □

For the initial vector of VAR coefficients, ϕ_0, Cogley and Sargent (2002) and Cogley and Sargent (2005b) 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 VAR with a flat prior on a presample. Del Negro (2003) advocates the use of a shrinkage prior with tighter variance than Cogley and Sargent's to partly overcome the problem of overfitting. Imposing the restriction that for each t all roots of the characteristic polynomial associated with the VAR coefficients ϕ_t lie outside the unit circle introduces a complication that we do not explore here. Koop and Potter (2008) discuss how to impose such a restriction efficiently.

Primiceri (2005) extends the above TVP VAR by also allowing the nonzero off-diagonal elements of the contemporaneous covariance matrix B to evolve as random-walk processes. If one is willing to assume that the lower-triangular B_t's identify structural shocks, then this model generalizes the constant-coefficient structural SVAR discussed in Section 2.4 with $\Omega = I$ to a TVP environment. Primiceri (2005) uses a structural TVP VAR for interest rates, inflation, and unemployment to estimate a time-varying monetary policy rule for the postwar United States. 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 (30). Finally, no cointegration restrictions are imposed on the VAR specified in (90). A Bayesian analysis of a TVP cointegration model can be found in Koop, Leon-Gonzalez, and Strachan (2008).

5.1.2 DSGE Models with Drifting Parameters

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

$$\mathbb{E}_t\left[\sum_{s=0}^{\infty}\beta^{t+s}\left(\ln C_{t+s}-\frac{(H_{t+s}/B)^{1+1/\nu}}{1+1/\nu}\right)\right]. \tag{95}$$

We can interpret our original objective function (52) as a generalization of (95), in which we have replaced the constant parameter B,which affects the disutility associated with working, 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 (66), 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

$$\widehat{H}_t=\nu\widehat{W}_t-\nu\widehat{C}_t+(1+\nu)\widehat{B}_t. \tag{96}$$

Now imagine replacing the constant Frisch elasticity ν in (52) and (95) by a time-varying process ν_t. In a log-linear approximation of the equilibrium conditions, the time-varying elasticity will appear as an additional additive shock in (96) and therefore be indistinguishable in its dynamic effects from B_t; provided that the steady-state ratio $H_*/B_*\neq 1$. If $H_*/B_*=1$, then ν_t has no effects on the first-order dynamics. 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. Fernández-Villaverde and Rubio-Ramírez (2008) 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 the different effects of a typical monetary policy shock and a shock that changes the central bank's reaction to deviations from the inflation target, for instance, the authors use a second-order perturbation method to solve the model and the particle filter to approximate its likelihood function. Thus, the topic of DSGE models with time-varying autoregressive parameters has essentially been covered in Section 4.6.

5.2 Models with Markov-Switching Parameters

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 7. The piecewise constant path of the target can be generated by a MS model but not by the drifting-parameter model of the previous subsection. We will begin with a discussion of MS coefficients in the context of a VAR (section 5.2.1) and then consider the estimation of DSGE models with MS parameters (section 5.2.2).

5.2.1 Markov-Switching VARs

MS models have been popularized in economics by the work of Hamilton (1989), who used them to allow for different GDP-growth-rate dynamics 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

$$y_t' = x_t'\Phi(K_t) + u_t', \quad u_t \sim iidN(0, \Sigma(K_t)) \tag{97}$$

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

$$\pi_{lm} = P[K_t = l \mid K_{t-1} = m], \quad l, m \in \{1, \ldots, M\}.$$

For simplicity, suppose that $M = 2$ and all elements of $\Phi(K_t)$ and $\Sigma(K_t)$ switch simultaneously, without any restrictions. We denote the values of the VAR parameter matrices in state $K_t = l$ by $\Phi(l)$ and $\Sigma(l)$, $l = 1, 2$, respectively. If the prior distributions of $(\Phi(l), \Sigma(l))$ are MNIW and the priors for the regime-switching probabilities π_{11} and π_{22} are independent Beta distributions, then posterior inference in this simple MS VAR model can be implemented with the following Gibbs sampler

Algorithm 5.2: Gibbs Sampler for Unrestricted MS VARs

For $s = 1, \ldots, n_{sim}$:

1. Draw $(\Phi^{(s)}(l), \Sigma^{(s)}(l))$ conditional on $(K_{1:T}^{(i-1)}, \pi_{11}^{(i-1)}, \pi_{22}^{(i-1)}, Y)$. Let $\mathcal{T}_l$ be a set that contains the time periods when $K_t = l$, $l = 1, 2$. Under a conjugate prior, the posterior of $\Phi(l)$ and $\Sigma(l)$ is MNIW, obtained from the regression $y_t' = x_t'\Phi(l) + u_t$, $u_t \sim N(0, \Sigma(l))$, $t \in \mathcal{T}_l$.

2. Draw $K_{1:T}^{(s)}$ conditional on $(\Phi^{(s)}(l), \Sigma^{(s)}(l), \pi_{11}^{(i-1)}, \pi_{22}^{(i-1)}, Y)$ using a variant of the Carter and Kohn (1994) approach, described in detail in Giordani, Pitt, and Kohn (This Volume).

3. Draw $\pi_{11}^{(s)}$ and $\pi_{22}^{(s)}$ conditional on $(\Phi^{(s)}(s), \Sigma^{(s)}(s), K_{1:T}^{(s)}, Y)$. If one ignores the relationship between the transition probabilities and the distribution of K_1, then the posteriors of $\pi_{11}^{(s)}$ and $\pi_{22}^{(s)}$ take the form of Beta distributions. If K_1 is distributed according to the stationary distribution of the Markov chain, then the Beta distributions can be used as proposal distributions in a Metropolis step. □

If one imposes the condition that $\pi_{22} = 1$ and $\pi_{12} = 0$, then model (97) becomes a change-point model in which state 2 is the final state.[4] 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 has been a structural break in postwar U.S. GDP growth toward 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) consider 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 (2009) explore posterior inference in change-point models under various types of prior distributions. Koop, Leon-Gonzalez, and Strachan (2009) consider a modification of Primiceri (2005)'s framework where parameters evolve according to a change-point model and study the evolution over time of the monetary policy transmission mechanism in the United States.

In a multivariate setting, the unrestricted MS VAR in (97) with coefficient matrices that are *a priori* independent across states may involve a large number of

[4] More generally, for a process with M states one would impose the restrictions $\pi_{MM} = 1$ and $\pi_{j+1,j} + \pi_{jj} = 1$.

coefficients, and parameter restrictions can compensate for lack of sample information. For instance, Paap and van Dijk (2003) start from the VAR specification used in Section 2.3 that expresses y_t as a deterministic trend and autoregressive deviations from this trend. The authors impose the restriction that only the trend is affected by the MS process:

$$y_t = y_t^* + \Gamma_0(K_t) + \widetilde{y}_t, \quad \widetilde{y}_t = \Phi_1 \widetilde{y}_{t-1} + \ldots + \Phi_p \widetilde{y}_{t-p} + u_t, \quad u_t \sim iidN(0, \Sigma), \tag{98}$$

where

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

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

Thus far, we have focused on reduced-form VARs with MS parameters. Sims and Zha (2006) extend the structural VAR given in (30) to a MS setting:

$$y_t' A_0(K_t) = x_t' A(K_t) + \epsilon_t', \quad \epsilon_t \sim iidN(0, I) \tag{99}$$

where ϵ_t is a vector of orthogonal structural shocks and x_t is defined as in Section 2.1. The authors reparameterize the $k \times n$ matrix $A(K_t)$ as $D(K_t) + GA_0(K_t)$, where $\bar{S}$ is a $k \times n$ with the $n \times n$ identity matrix in the first n rows and zeros elsewhere. Thus,

$$y_t' A_0(K_t) = x_t'(D(K_t) + GA_0(K_t)) + \epsilon_t'. \tag{100}$$

If $D(K_t) = 0$, then the reduced-form VAR coefficients are given by $\Phi = A(K_t)[A_0(K_t)]^{-1} = G$ and the elements of y_t follow random-walk processes, as implied by the mean of the Minnesota prior (see Section 2.2). Loosely speaking, if the prior for $D(K_t)$ is centered at zero, the prior for the reduced-form VAR is centered at a random-walk representation.

To avoid a proliferation of parameters, Sims and Zha (2006) impose constraints on the evolution of $D(K_t)$ across states. Let $d_{i,j,l}$ correspond to the coefficient associated with lag l of variable i in equation j. The authors impose that $d_{i,j,l}(K_t) = \delta_{i,j,l}\lambda_{i,j}(K_t)$. This specification allows for shifts in $D(K_t)$ to be equation or variable dependent but rules out lag dependency. The authors use their setup to estimate MS VAR 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 (100) is obtained

by merging and generalizing Algorithms 2.4 and 5.2. Details are provided in Sims, Waggoner, and Zha (2008).

5.2.2 DSGE Models with Markov-Switching Coefficients

A growing number of papers incorporates Markov-switching effects in DSGE models. Consider the nonlinear equilibrium conditions of our stochastic growth model in (61). The most rigorous and general treatment of Markov-switching coefficients would involve replacing the vector θ with a function of the latent state K_t, $\theta(K_t)$, and solving the nonlinear model while accounting for the time variation in θ. Since the implementation of the solution and the subsequent computation of the likelihood function are very challenging, the literature has focused on various short-cuts, which introduce Markov-switching in the coefficients of the linearized model given by (66).

Following Sims (2002b), we write the linearized equilibrium conditions of the DSGE model in the following canonical form:

$$\Gamma_0(\theta)x_t = C(\theta) + \Gamma_1(\theta)x_{t-1} + \Psi(\theta)\epsilon_t + \Pi(\theta)\eta_t. \tag{101}$$

For the stochastic growth model presented in Section 4, θ is defined in (63), and the vector x_t can be defined as follows:

$$x_t = \left[\widehat{C}_t, \widehat{H}_t, \widehat{W}_t, \widehat{Y}_t, \widehat{R}_t, \widehat{I}_t, \widehat{K}_{t+1}, \widehat{A}_t, \widehat{a}_t, \widehat{B}_t, I\!E_t[\widehat{C}_{t+1}], I\!E_t[\widehat{a}_{t+1}], I\!E_t[\widehat{R}_{t+1}]\right]'.$$

The vector η_t comprises the following one-step-ahead rational expectations forecast errors:

$$\eta_t = \left[(\widehat{C}_t - I\!E_{t-1}[\widehat{C}_t]), (\widehat{a}_t - I\!E_{t-1}[\widehat{a}_t]), (\widehat{R}_t - I\!E_{t-1}[\widehat{R}_t])\right]'$$

and ϵ_t stacks the innovations of the exogenous shocks: $\epsilon_t = [\epsilon_{a,t}, \epsilon_{b,t}]'$. With these definitions, it is straightforward, albeit slightly tedious, to rewrite (66) in terms of the canonical form (101). In most applications, including our stochastic growth model, one can define the vector x_t such that the observables y_t can, as in Section 4.2, be expressed simply as a linear function of x_t; that is:

$$y_t = \Psi_0(\theta) + \Psi_1(\theta)t + \Psi_2(\theta)x_t. \tag{102}$$

Markov-switching can be introduced into the linearized DSGE model by expressing the DSGE model parameters θ as a function of a hidden Markov process K_t, which we denote by $\theta(K_t)$.

Schorfheide (2005) considers a special case of this Markov-switching linear rational expectations framework, because in his analysis the process K_t affects only the target inflation rate of the central bank, which can be low or high. Using the same notation as in Section 5.2.1, the number of states is $M = 2$, and the state transition probabilities are denoted by π_{lm}. If we partition the parameter vector $\theta(K_t)$ into a component θ_1 that is unaffected by the hidden Markov process K_t and a component $\theta_2(K_t)$ that varies with K_t and takes the values $\theta_2(l)$, $l = 1, 2$, the resulting rational expectations system can be written as

$$\Gamma_0(\theta_1)x_t = C(\theta_1, \theta_2(K_t)) + \Gamma_1(\theta_1)x_{t-1} + \Psi(\theta_1)\epsilon_t + \Pi(\theta_1)\eta_t \tag{103}$$

and is solvable with the algorithm provided in Sims (2002b). The solution takes the special form

$$y_t = \Psi_0 + \Psi_1 t + \Psi_2 x_t, \quad x_t = \Phi_1 x_{t-1} + \Phi_\epsilon[\mu(K_t) + \epsilon_t] + \Phi_0(K_t), \tag{104}$$

where only Φ_0 and μ depend on the Markov process K_t (indirectly through $\theta_2(K_t)$), but not the matrices Ψ_0, Ψ_1, Ψ_2, Φ_1, and Φ_ϵ. Equation (104) defines a (linear) Markov-switching state-space model, with the understanding that the system matrices are functions of the DSGE model parameters θ_1 and $\theta_2(K_t)$. Following a filtering approach that simultaneously integrates over x_t and K_t, discussed in Kim and Nelson (1999b), Schorfheide (2005) constructs an approximate likelihood that depends only on θ_1, $\theta_2(1)$, $\theta_2(2)$ and the transition probabilities π_{11} and π_{22}. This likelihood function is then used in Algorithm 4.1 to implement posterior inference.

The analysis in Schorfheide (2005) is clearly restrictive. For instance, there is a large debate in the literature about whether the central bank's reaction to inflation and output deviations from target changed around 1980. A candidate explanation for the reduction of macroeconomic volatility in the 1980s is a more forceful reaction of central banks to inflation deviations. To capture this explanation in a Markov-switching rational expectations model, it is necessary that not just the intercept in (101) but also the slope coefficients be affected by the regime shifts. Thus, subsequent work by Davig and Leeper (2007) and Farmer, Waggoner, and Zha (2009) is more ambitious in that it allows for switches in all the matrices of the canonical rational expectations model:

$$\Gamma_0(\theta(K_t))x_t = C(\theta(K_t)) + \Gamma_1(\theta(K_t))x_{t-1} + \Psi(\theta(K_t))\epsilon_t + \Pi(\theta(K_t))\eta_t.$$

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

5.3 Applications of Bayesian TVP Models

Bayesian TVP models have been applied to several issues of interest, including macroeconomic forecasting, for example, Sims (1993) and Cogley, Morozov, and Sargent (2005). Here, we shall focus on one specific issue, namely, the debate over whether the dynamics of U.S. inflation changed over the last quarter of the 20th century and, to the extent that they have, whether monetary policy played a major role in affecting inflation dynamics. Naturally, this debate evolved in parallel to the debate over the magnitude and causes of the Great Moderation, that is, the decline in the volatility of business cycles around 1984 initially documented by Kim and Nelson (1999a) and McConnell and Perez-Quiros (2000). Whatever the causes of the changes in output dynamics were – shocks, monetary policy, or other structural changes – it is likely that these same causes affected the dynamics of inflation.

Bayesian inference in a TVP VAR yields posterior estimates of the reduced-form coefficients ϕ_t in (90). Conditioning on estimates of ϕ_t for various periods between 1960 and 2000, Cogley and Sargent (2002) compute the spectrum of inflation based on their VAR and use it as evidence that both inflation volatility and persistence have changed dramatically in the United States. Cogley and Sargent (2005b) find that their earlier empirical results are robust to time-variation in the volatility of shocks and argue that changes in the monetary policy rule are partly responsible for the changes in inflation dynamics. Based on an estimated structural TVP VAR, Primiceri (2005) argues that monetary policy has indeed changed since the 1980s but that the impact of these changes on the rest of the economy has been small. 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. Sims and Zha (2006) conduct inference with a MS VAR and find no support for the hypothesis that the parameters of the monetary policy rule differed pre- and post-1980. To the contrary, they provide evidence that it was the behavior of the private sector that changed and that shock heteroskedasticity is important. Similarly, using an AR time-varying coefficients VAR identified with sign restrictions Canova and Gambetti (2009) find little evidence that monetary policy has become more aggressive in

responding to inflation since the early 1980s. Cogley and Sbordone (2008) use a TVP VAR to assess the stability of the New Keynesian Phillips curve during the past four decades.

Given the numerical difficulties of estimating nonlinear DSGE models, there currently exists less published empirical work based on DSGE models with time-varying coefficients. Two notable exceptions are the papers by Justiniano and Primiceri (2008) discussed in Section (4.5) and Fernández-Villaverde and Rubio-Ramírez (2008). The latter paper provides evidence that after 1980 the U.S. central bank has changed interest rates more aggressively in response to deviations of inflation from the target rate. The authors also find that the estimated frequency of price changes has decreased over time. This frequency is taken as exogenous within the Calvo framework they adopt.

6 Models for Data-Rich Environments

We now turn to inference with models for data sets that have a large cross-sectional and time-series dimension. Consider the VAR(p) from Section 2:

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

where y_t is an $n \times 1$ 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. For instance, in Illustration 2.1 the time-series dimension was approximately $T = 160$ and the cross-sectional dimension was $n = 4$. This section focuses on applications in which the ratio T/n is relatively small, possibly less than 5.

High-dimensional VARs are useful for applications that involve large cross sections of macroeconomic indicators for a particular country – for example, GDP and its components, industrial production, measures of employment and compensation, housing starts and new orders of capital goods, price indices, interest rates, consumer confidence measures, *et cetera*. Examples of such data sets can be found in Stock and Watson (1999) and Stock and Watson (2002). Large-scale VARs are also frequently employed in the context of multicountry econometric modeling. For instance, to study international business cycles among OECD countries, y_t might

be composed of aggregate output, consumption, investment, and employment for a group of 20 to 30 countries, which leads to $n > 80$.

In general, for the models considered in this section there will be a shortage of sample information to determine parameters, leading to imprecise inference and diffuse predictive distributions. Priors can be used to impose either hard or soft parameter restrictions and thereby to sharpen inference. Hard restrictions involve setting combinations of VAR coefficients equal to zero. 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 for GDP growth in a given country enter only the trade-weighted averages of the other countries' GDP growth rates. Second, one could use very informative, yet nondegenerate, prior distributions for the many VAR coefficients, which is what is meant by soft restrictions. Both types of restrictions are discussed in Section 6.1. Finally, one could express y_t as a function of a lower-dimensional vector of variables called factors, possibly latent, that drive all the comovement among the elements of y_t, plus a vector ζ_t of so-called idiosyncratic components, which evolve independently from one another. In such a setting, one needs only to parameterize the evolution of the factors, the impact of these on the observables y_t, and the evolution of the univariate idiosyncratic components, rather than the dynamic interrelationships among all the elements of the y_t vector. Factor models are explored in Section 6.2.

6.1 Restricted High-Dimensional VARs

We begin by directly imposing hard restrictions on the coefficients of the VAR. As before, define the $k \times 1$ vector $x_t = [y'_{t-1}, \ldots, y'_{t-p}, 1]'$ and the $k \times n$ matrix $\Phi = [\Phi_1, \ldots, \Phi_p, \Phi_c]'$, where $k = np + 1$. Moreover, let $X_t = I_n \otimes x_t$ and $\phi = vec(\Phi)$ with dimensions $kn \times n$ and $kn \times 1$, respectively. Then we can write the VAR as

$$y_t = X'_t\phi + u_t, \quad u_t \sim iidN(0, \Sigma). \tag{105}$$

To incorporate the restrictions on ϕ, we reparameterize the VAR as follows:

$$\phi = M\theta. \tag{106}$$

θ is a vector of size $\kappa << nk$, and the $nk \times \kappa$ matrix M induces the restrictions by linking the VAR coefficients ϕ to the lower-dimensional parameter vector θ. The elements of M are known. For instance, M could be specified such that the

coefficient in Equation i, $i = 1, .., n$, associated with the l'th lag of variable j is the sum of an equation-specific, a variable-specific parameter, and a lag-specific parameter. Here, θ would comprise the set of all $n + n + p$ equation/variable/lag-specific parameters, and M would be an indicator matrix of zeros and ones that selects the elements of θ associated with each element of ϕ. The matrix M could also be specified to set certain elements of ϕ equal to zero and thereby exclude regressors from each of the n equations of the VAR. Since the relationship between ϕ and θ is linear, Bayesian inference in this restricted VAR under a Gaussian prior for θ and an Inverted Wishart prior for Σ is straightforward.

To turn the hard restrictions (106) into soft restrictions, one can construct a hierarchical model, in which the prior distribution for ϕ conditional on θ has a nonzero variance:

$$\phi = M\theta + \nu, \quad \nu \sim N(0, V), \tag{107}$$

where ν is an $nk \times 1$ vector with $nk \times nk$ covariance matrix V. The joint distribution of parameters and data can be factorized as

$$p(Y, \phi, \theta) = p(Y|\phi)p(\phi|\theta)p(\theta). \tag{108}$$

A few remarks are in order. First, (108) has the same form as the DSGE-VAR discussed in Section 4.7.3, except that the conditional distribution of ϕ given θ is centered at the simple linear restriction $M\theta$ rather than the rather complicated VAR approximation of a DSGE model. Second, (108) also nests the Minnesota prior discussed in Section 2.2, which can be obtained by using a degenerate distribution for θ concentrated at $\underline{\theta}$ with a suitable choice of M, $\underline{\theta}$, and V. Third, in practice the choice of the prior covariance matrix V is crucial for inference. In the context of the Minnesota prior and the DSGE-VAR, we expressed this covariance matrix in terms of a low-dimensional vector λ of hyperparameters such that $\|V(\lambda)\| \longrightarrow 0$ ($\|V(\lambda)\| \longrightarrow \infty$) as $\|\lambda\| \longrightarrow \infty$ ($\|\lambda\| \longrightarrow 0$) and recommended conditioning on a value of λ that maximizes the marginal likelihood function $p_\lambda(Y)$ over a suitably chosen grid.

Finally, since the discrepancy between the posterior mean estimate of ϕ and the restriction $M\theta$ can be reduced by increasing the hyperparameter λ, the resulting Bayes estimator of ϕ is often called a *shrinkage* estimator. De Mol, Giannone, and Reichlin (2008) consider a covariance matrix V that in our notation takes the form $V = \Sigma \otimes (I_k/\lambda^2)$ and show that there is a tight connection between these shrinkage estimators and estimators of conditional mean functions obtained from factor

models, which we will discuss below. They document empirically that with a suitably chosen shrinkage parameter the forecast performance of their Bayes predictor constructed from a large number of regressors is similar to the performance of a predictor obtained by regressing y_t on the first few principal components of the regressors x_t, as is often done in the factor model literature.

Canova and Ciccarelli (2009) allow the deviations of ϕ from the restricted subspace characterized by $M\theta$ to differ in each period t. Formally, they allow for time-variation in ϕ and let

$$\phi_t = M\theta + \nu_t, \quad \nu_t \sim iidN(0, V). \tag{109}$$

The deviations ν_t from the restriction $M\theta$ are assumed to be independent over time, which simplifies inference. In fact, the random deviations ν_t can be merged with the VAR innovations u_t, resulting in a model for which Bayesian inference is fairly straightforward to implement. Inserting (109) into (105), we obtain the system

$$y_t = (X_t'M)\theta + \zeta_t. \tag{110}$$

The $n \times \kappa$ matrix of regressors $X_t'M$ essentially contains weighted averages of the regressors, where the weights are given by the columns of M. The random vector ζ_t is given by $\zeta_t = X_t'\nu_t + u_t$ and, since x_t contains lagged values of y_t, forms a Martingale difference sequence with conditional covariance matrix $X_t'VX_t + \Sigma$. If one chooses a prior covariance matrix of the form $V = \Sigma \otimes (I_k/\lambda^2)$, then the covariance matrix of ζ_t reduces to $(1 + (x_t'x_t)/\lambda^2)\Sigma$. The likelihood function (conditional on the initial observations $Y_{-p+1:0}$) takes the convenient form

$$\begin{aligned} p(Y_{1:T}|\theta, \lambda) \quad \propto \quad & \left|(1 + (x_t'x_t)/\lambda^2)\Sigma\right|^{-1/2} \\ & \times \prod_{t=1}^{T} \exp\left\{-\frac{1}{2(1 + (x_t'x_t)/\lambda^2)}(y_t - X_t'M\theta)'\Sigma^{-1}(y_t - X_t'M\theta)\right\}, \end{aligned} \tag{111}$$

and Bayesian inference under a conjugate prior for θ and Σ is straightforward.

Canova and Ciccarelli (2009) further generalize expression (109) by assuming that the vector θ is time-varying and follows a simple autoregressive law of motion. They discuss in detail how to implement Bayesian inference in this more general environment. The authors interpret the time-varying θ_t as a vector of latent factors. Their setting is therefore related to that of the factor models described in the next subsection. In multicountry VAR applications, M could be chosen such that y_t is a

function of lagged country-specific variables and, say, average lagged output growth and unemployment across countries. If most of the variation in the elements of y_t is due to the cross-sectional averages, then the business cycles in the various countries are highly synchronized. Canova and Ciccarelli (2009) use their framework to study the convergence in business cycles among G7 countries.

6.2 Dynamic Factor Models

Factor models describe the dynamic behavior of a possibly large cross section of observations as the sum of a few common components, which explain comovements, and of series-specific components, which capture idiosyncratic dynamics of each series. While factor models have been part of the econometricians' toolbox for a long time – the *unobservable index* models by Sargent and Sims (1977) and Geweke (1977)), for example – the contribution of Stock and Watson (1989) generated renewed interest in this class of models among macroeconomists. These authors use a factor model to exploit information from a large cross section of macroeconomic time series for forecasting. While Stock and Watson (1989) employ maximum likelihood methods, Geweke and Zhou (1996) and Otrok and Whiteman (1998) conduct Bayesian inference with dynamic factor models. Our baseline version of the DFM is introduced in Section 6.2.1, and posterior inference is described in Section 6.2.2. Some applications are discussed in Section 6.2.3. Finally, Section 6.2.4 surveys various extensions of the basic DFM.

6.2.1 Baseline Specification

A DFM decomposes the dynamics of n observables $y_{i,t}$, $i = 1, \ldots, n$, into the sum of two unobservable components:

$$y_{i,t} = a_i + \lambda_i f_t + \xi_{i,t}, \quad t = 1, \ldots, T. \tag{112}$$

Here, f_t is a $\kappa \times 1$ vector of factors that are common to all observables, and $\xi_{i,t}$ is an idiosyncratic process that is specific to each i. Moreover, a_i is a constant, and λ_i is a $1 \times \kappa$ vector of loadings that links $y_{i,t}$ to the factor f_t. The factors follow a vector autoregressive processes of order q:

$$f_t = \Phi_{0,1} f_{t-1} + \ldots + \Phi_{0,q} f_{t-q} + u_{0,t}, \quad u_{0,t} \sim iidN(0, \Sigma_0), \tag{113}$$

where Σ_0 and the $\Phi_{0,j}$ matrices are of dimension $\kappa \times \kappa$ and $u_{0,t}$ is a $\kappa \times 1$ vector of innovations. We used 0-subscripts to denote parameter matrices that describe the law of motion of the factors. The idiosyncratic components follow autoregressive processes of order p_i:

$$\xi_{i,t} = \phi_{i,1}\xi_{i,t-1} + \ldots + \phi_{i,p_i}\xi_{i,t-p_i} + u_{i,t}, \quad u_{i,t} \sim iidN(0, \sigma_i^2). \tag{114}$$

At all leads and lags, the $u_{i,t}$ innovations are independent across i and independent of the innovations to the law of motion of the factors $u_{0,t}$. These orthogonality assumptions are important to identifying the factor model, as they imply that all comovements in the data arise from the factors.

Without further restrictions, the latent factors and the coefficient matrices of the DFM are not identifiable. One can premultiply f_t and its lags in (112) and (113) as well as $u_{0,t}$ by a $\kappa \times \kappa$ invertible matrix H and postmultiply the vectors λ_i and the matrices $\Phi_{0,j}$ by H^{-1}, without changing the distribution of the observables. There are several approaches to restricting the parameters of the DFM to normalize the factors and achieve identification. We will provide three specific examples in which we impose restrictions on Σ_0 and the first κ loading vectors stacked in the matrix

$$\Lambda_{1,\kappa} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_\kappa \end{bmatrix}.$$

The loadings λ_i for $i > \kappa$ are always left unrestricted.

Example 6.1: Geweke and Zhou (1996) restrict $\Lambda_{1,\kappa}$ to be lower-triangular:

$$\Lambda_{1,\kappa} = \Lambda_{1,\kappa}^{tr} = \begin{bmatrix} X & 0 \cdots 0 & 0 \\ \vdots & \ddots & \vdots \\ X & X \cdots X & X \end{bmatrix}. \tag{115}$$

Here, X denotes an unrestricted element, and 0 denotes a zero restriction. The restrictions can be interpreted as follows. According to (115), factor $f_{2,t}$ does not affect $y_{1,t}$, factor $f_{3,t}$ does not affect $y_{1,t}$ and $y_{2,t}$, and so forth. However, these zero restrictions alone are not sufficient for identification because the factors and hence the matrices $\Phi_{0,j}$ and Σ_0 could still be transformed by pre- and postmultiplication of an arbitrary invertible lower-triangular $\kappa \times \kappa$ matrix H_{tr} without changing the distribution of the observables. Under this transformation, the factor innovations

become $H_{tr}u_{0,t}$. Since Σ_0 can be expressed as the product of the unique lower-triangular Choleski factor $\Sigma_{0,tr}$ and its transpose, one can choose $H_{tr} = \Sigma_{0,tr}^{-1}$ such that the factor innovations reduce to a vector of independent standard Normals. To implement this normalization, we simply let

$$\Sigma_0 = I_\kappa. \tag{116}$$

Finally, the signs of the factors need to be normalized. Let $\lambda_{i,i}$, $i = 1, \ldots, \kappa$, be the diagonal elements of $\Lambda_{1,\kappa}$. The sign normalization can be achieved with a set of restrictions of the form

$$\lambda_{i,i} \geq 0, \quad i = 1, \ldots, \kappa. \tag{117}$$

Thus, (115), (116), and (117) provide a set of identifying restrictions. □

Example 6.2: Suppose we start from the normalization in the previous example and proceed with premultiplying the factors by the diagonal matrix H that is composed of the diagonal elements of $\Lambda_{1,\kappa}^{tr}$ in (115) and postmultiplying the loadings by H^{-1}. This transformation leads to a normalization in which $\Lambda_{1,\kappa}$ is restricted to be lower-triangular with ones on the diagonal and Σ_0 is a diagonal matrix with nonnegative elements. The one-entries on the diagonal of $\Lambda_{1,\kappa}$ also take care of the sign normalization. Since under the normalization $\lambda_{i,i} = 1$, $i = 1, \ldots, \kappa$, factor $f_{i,t}$ is forced to have a unit impact on $y_{i,t}$, there exists a potential pitfall. For instance, imagine that there is only one factor and that $y_{1,t}$ is uncorrelated with all other observables. Imposing $\lambda_{1,1} = 1$ may result in a misleading inference for the factor as well as for the other loadings. □

Example 6.3: Suppose we start from the normalization in Example 6.1 and proceed with premultiplying the factors by the matrix $H = \Lambda_{1,\kappa}^{tr}$ in (115) and postmultiplying the loadings by H^{-1}. This transformation leads to a normalization in which $\Lambda_{1,\kappa}$ is restricted to be the identity matrix and Σ_0 is an unrestricted covariance matrix. As in Example 6.2, the one-entries on the diagonal of $\Lambda_{1,\kappa}$ take care of the sign normalization. □

Finally, one might find it attractive to impose overidentifying restrictions. For concreteness, imagine that the factor model is used to study comovements in output across U.S. states, and let $y_{i,t}$ correspond to output in state i in period t. Moreover, suppose that the number of factors is $\kappa = 3$, where $f_{1,t}$ is interpreted as a national business cycle and $f_{2,t}$ and $f_{3,t}$ are factors that affect the Eastern and Western regions, respectively. In this case, one could impose the condition that $\lambda_{i,j} = 0$ if state i does not belong to region $j = 2, 3$.

6.2.2 Priors and Posteriors

We now describe Bayesian inference for the DFM. To simplify the notation, we will discuss the case in which the lag length in (114) is the same for all i ($p_i = p$) and $q \leq p+1$. As we did previously in this chapter, we adopt the convention that $Y_{t_0:t_1}$ and $F_{t_0:t_1}$ denote the sequences $\{y_{t_0}, \ldots, y_{t_1}\}$ and $\{f_{t_0}, \ldots, f_{t_1}\}$, respectively. Premultiply (112) by $1 - \phi_{i,1}L \cdots - \phi_{i,p}L^p$, where L here denotes the lag operator. The quasi-differenced measurement equation takes the form

$$\begin{aligned} y_{i,t} &= a_i + \lambda_i f_t + \phi_{i,1}(y_{i,t-1} - a_i - \lambda_i f_{t-1}) + \ldots \\ &\quad + \phi_{i,p}(y_{i,t-p} - a_i - \lambda_i f_{t-p}) + u_{i,t}, \quad \text{for} \quad t = p+1, .., T. \end{aligned} \tag{118}$$

Let $\theta_i = [a_i, \lambda_i, \sigma_i, \phi_{i,1}, .., \phi_{i,p}]'$ be the parameters entering (118) and θ_0 be the parameters pertaining to the law of motion of the factors (113). The joint distribution of data, parameters, and latent factors can be written as

$$\begin{aligned} &p(Y_{1:T}, F_{0:T}, \{\theta_i\}_{i=1}^n, \theta_0) \\ &= \left[\prod_{t=p+1}^{T}\left(\prod_{i=1}^{n} p(y_{i,t}|Y_{i,t-p:t-1}, F_{t-p:t}, \theta_i)\right) p(f_t|F_{t-q:t-1}, \theta_0)\right] \\ &\quad \times \left(\prod_{i=1}^{n} p(Y_{i,1:p}|F_{0:p}, \theta_i)\right) p(F_{0:p}|\theta_0) \left(\prod_{i=1}^{n} p(\theta_i)\right) p(\theta_0). \end{aligned} \tag{119}$$

To obtain the factorization on the right-hand side of (119), we exploited the fact that the conditional distribution of $y_{i,t}$ given $(Y_{1:t-1}, F_{0:t}, \theta_i)$ depends on lagged observables only through $Y_{i,t-p:t-1}$ and on the factors only through $F_{t-p:t}$. Moreover, the distribution of f_t conditional on $(Y_{1:t-1}, F_{0:t-1}, \theta_0)$ is a function only of $F_{t-q:t-1}$. The distributions $p(y_{i,t}|Y_{i,t-p:t-1}, F_{t-p:t}, \theta_i)$ and $p(f_t|F_{t-q:t-1}, \theta_0)$ can easily be derived from expressions (118) and (113), respectively.

The term $p(Y_{i,1:p}|F_{0:p}, \theta_i)$ in (119) represents the distribution of the first p observations conditional on the factors, which is given by

$$\left[\begin{array}{c} y_{i,1} \\ \vdots \\ y_{i,p} \end{array}\right] \Bigg| (F_{0:p}, \theta_i) \sim N\left(\left[\begin{array}{c} a_i + f_1 \\ \vdots \\ a_i + f_p \end{array}\right], \Sigma_{i,1:p}(\theta_i)\right). \tag{120}$$

The matrix $\Sigma_{i,1:p}(\theta_i)$ is the covariance matrix of $[\xi_{i,1}, \ldots, \xi_{i,p}]'$, which can be derived from the autoregressive law of motion (114) by assuming that $\xi_{i,-(\tau+1)} = \ldots = \xi_{i,-(\tau+p)} = 0$ for some $\tau > 0$. If the law of motion of $\xi_{i,t}$ is stationary for all θ_i in the

support of the prior, one can set $\tau = \infty$, and $\Sigma_{i,1:p}$ becomes the covariance matrix associated with the unconditional distribution of the idiosyncratic shocks. Detailed derivations can be found in Otrok and Whiteman (1998). The initial distribution of the factors $p(F_{0:p}|\theta_0)$ can be obtained in a similar manner using (113).

The remaining terms, $p(\theta_i)$ and $p(\theta_0)$, represent the priors for θ_i and θ_0, which are typically chosen to be conjugate (see, for example, Otrok and Whiteman (1998)). Specifically, the priors on the constant term a_i and the loadings λ_i are normal, namely, $N(\underline{a}_i, \underline{V}_{a_i})$ and $N(\underline{\lambda}_i, \underline{V}_{\lambda_i})$. If the $\lambda_{i,i}$, $i = 1, \ldots, \kappa$ elements are restricted to be nonnegative to resolve the sign-indeterminacy of the factors as in Example 6.1, then the density associated with the prior for λ_i needs to be multiplied by the indicator function $\mathcal{I}\{\lambda_{i,i} \geq 0\}$ to impose the constraint (117). The autoregressive coefficients for the factors and the idiosyncratic shocks have a Normal prior. Define $\phi_0 = [vec(\Phi_{0,1})', .., vec(\Phi_{0,q})']'$ and assume that Σ_0 is normalized to be equal to the identity matrix. The prior for ϕ_0 is $N(\underline{\phi}_0, \underline{V}_{\phi_0})$. Likewise, the prior for $\phi_i = [\phi_{i,1}, .., \phi_{i,p}]'$ is $N(\underline{\phi}_i, \underline{V}_{\phi_i})$. In some applications, it may be desirable to truncate the prior for ϕ_0 (ϕ_i) to rule out parameters for which not all of the roots of the characteristic polynomial associated with the autoregressive laws of motion of f_t and $\xi_{i,t}$ lie outside the unit circle. Finally, the prior for the idiosyncratic volatilities σ_i can be chosen to be of the Inverted Gamma form.

A Gibbs sampler can be used to generate draws from the posterior distribution. The basic structure of the sampler is fairly straightforward though some of the details are tedious and can be found, for instance, in Otrok and Whiteman (1998). Conditional on the factors, Equation (112) is a linear Gaussian regression with AR(p) errors. The posterior density takes the form

$$p(\theta_i|F_{0:T}, \theta_0, Y_{1:T}) \propto p(\theta_i) \left(\prod_{t=p+1}^{T} p(y_{i,t}|Y_{i,t-p:t-1}, F_{t-p:t}, \theta_i) \right) p(Y_{i,1:p}|F_{0:p}, \theta_i). \tag{121}$$

Under a conjugate prior, the first two terms on the right-hand side correspond to the density of a Normal-Inverted Gamma distribution. The last term reflects the effect of the initialization of the AR(p) error process, and its log is not a quadratic function of θ_i. Draws from the distribution associated with (121) can be obtained with the procedure of Chib and Greenberg (1994).

If the prior for $\lambda_{i,i}$, $i = 1, \ldots, \kappa$ includes the indicator function $\mathcal{I}\{\lambda_{i,i} \geq 0\}$, one can use an acceptance sampler that discards all draws of θ_i for which $\lambda_{i,i} < 0$. If

the prior of the loadings does not restrict $\lambda_{i,i} \geq 0$, $i = 1, \ldots, \kappa$, but is symmetric around zero, then one can resolve the sign indeterminacy by postprocessing the output of the (unrestricted) Gibbs sampler: for each set of draws $(\{\theta_i\}_{i=1}^n, \theta_0, F_{0:T})$ such that $\lambda_{i,i} < 0$, flip the sign of the i'th factor and the sign of the loadings of all n observables on the ith factor. Hamilton, Waggoner, and Zha (2007) discuss the sign normalization and related normalization issues in other models at length. Since the errors $\xi_{i,t}$ in equation (112) are independent across i, the sampling can be implemented one i at a time, which implies that computational cost is linear in the size of the cross section.

Conditional on the factors, the posterior for the coefficients θ_0 in (113) is obtained from a multivariate generalization of the preceding steps. Its density can be written as

$$p(\theta_0|F_{0:T}, \{\theta_i\}_{i=1}^n, Y_{1:T}) \propto \left(\prod_{t=p+1}^{T} p(f_t|F_{t-p:t-1}, \theta_0) \right) p(\theta_0) p(F_{0:p}|\theta_0). \tag{122}$$

The first term on the right-hand side corresponds to the conditional likelihood function of a VAR(q) and has been extensively analyzed in Section 2. If the prior for θ_0 is conjugate, the first two terms are proportional to the density of a MNIW distribution if Σ_0 is unrestricted and corresponds to a multivariate normal density if the DFM is normalized such that $\Sigma_0 = I$. The last terms capture the probability density function of the initial factors $f_0, \ldots, f_p$. Thus, θ_0 cannot be directly sampled from, say, a MNIW distribution. As in the case of θ_i, one can use a variant of the procedure proposed by Chib and Greenberg (1994).

In the third block of the Gibbs sampler, one draws the factors $F_{0:T}$ conditional on $(\{\theta_i\}_{i=1}^n, \theta_0, Y_{1:T})$. Two approaches exist in the Bayesian DFM literature. Otrok and Whiteman (1998) explicitly write out the joint Normal distribution of the observations $Y_{1:T}$ and the factors $F_{0:T}$, $p(Y_{1:T}, F_{0:T}|\{\theta_i\}_{i=1,n}, \theta_0)$ and derive the posterior distribution $p(F_{0:T}|\{\theta_i\}_{i=1,n}, \theta_0, Y_{1:T})$ using the formula for conditional means and covariance matrices of a multivariate normal distribution.[5] Their approach involves inverting matrices of size T and hence becomes computationally expensive for data sets with a large time-series dimension. An alternative is to cast the DFM into a linear state-space form and apply the algorithm of Carter and Kohn (1994) for sampling from the distribution of the latent states, described in Giordani, Pitt, and

[5] If $X = [X_1', X_2']$ is distributed $N(\mu, \Sigma)$ then $X_1|X_2$ is distributed $N\big(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2),\ \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\big)$, where the partitions of μ and Σ conform with the partitions of X.

Kohn (This Volume). To avoid the increase in the dimension of the state vector with the cross-sectional dimension n, it is convenient to exclude the AR(p) processes $\xi_{i,t}$ from the state vector and to use the quasi-differenced measurement equation (118) instead of (112).

We will now provide some more details on how to cast the DFM into state-space form with *iid* measurement errors and a VAR(1) state-transition equation. For ease of notation, we shall subsequently assume that the factor f_t is scalar ($\kappa = 1$). Stacking (118) for all i, one obtains the measurement equation

$$(I_n - \sum_{j=1}^{p} \tilde{\Phi}_j L^j)\tilde{y}_t = (I_n - \sum_{j=1}^{p} \tilde{\Phi}_j)\tilde{a} + \Lambda^* \tilde{f}_t + \tilde{u}_t, \quad t = p+1, \ldots, T, \tag{123}$$

where L is the temporal lag operator, $\tilde{y}_t = [y_{1,t}, \ldots, y_{n,t}]'$, $\tilde{a} = [a_1, \ldots, a_n]'$, $\tilde{u}_t = [u_{1,t}, \ldots, u_{n,t}]'$, the $\tilde{\Phi}_j$'s are diagonal $n \times n$ matrices with elements $\phi_{1,j}, \ldots, \phi_{n,j}$, and

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

Due to the quasi-differencing, the random variables $\tilde{u}_t$ in the measurement equation (123) are *iid*. The $(p+1) \times 1$ vector $\tilde{f}_t$ collects the latent states and is defined as $\tilde{f}_t = [f_t, .., f_{t-p}]'$. The state-transition equation is obtained by expressing the law of motion of the factor (113) in companion form

$$\tilde{f}_t = \tilde{\Phi}_0 \tilde{f}_{t-1} + \tilde{u}_{0,t}, \tag{124}$$

where $\tilde{u}_{0,t} = [u_{0,t}, 0, .., 0]'$ is an *iid* $(p+1) \times 1$ random vector and $\tilde{\Phi}_0$ is the $(p+1) \times (p+1)$ companion form matrix

$$\tilde{\Phi}_0 = \begin{bmatrix} [\Phi_{0,1}, \ \ldots, \ \Phi_{0,q}, \ 0_{1\times(p+1-q)}] \\ I_p \qquad\qquad 0_{p\times 1} \end{bmatrix}. \tag{125}$$

Since (123) starts from $t = p+1$ as opposed to $t = 1$, one needs to initialize the filtering step in the Carter and Kohn (1994) algorithm with the conditional distribution of $p(F_{0:p}|Y_{1:p}, \{\theta_i\}_{i=1}^n, \theta_0)$. As mentioned above, this conditional distribution can be obtained from the joint distribution $p(F_{0:p}, Y_{1:p}|\{\theta_i\}_{i=1}^n, \theta_0)$ by using the formula for conditional means and covariance matrices of a multivariate normal distribution. Del Negro and Otrok (2008) provide formulas for the initialization. The Gibbs sampler can be summarized as follows

Algorithm 6.1: Sampling from the Posterior of the DFM

For $s = 1, \ldots, n_{sim}$:

1. Draw $\theta_i^{(s)}$ conditional on $(F_{0:T}^{(s-1)}, \theta_0^{(s-1)}, Y_{1:T})$ from (121). This can be done independently for each $i = 1, \ldots, n$.
2. Draw $\theta_0^{(s)}$ conditional on $(F_{0:T}^{(s-1)}, \{\theta_i^{(s)}\}_{i=1}^n, Y_{1:T})$ from (122).
3. Draw $F_{0:T}^{(s)}$, conditional on $(\{\theta_i^{(s)}\}_{i=1}^n, \theta_0^{(s)}, Y_{1:T})$.

We have omitted the details of the conditional posterior distributions. The exact distributions can be found in the references given in this section. Last, we have not discussed the issue of determining the number of factors κ. In principle, one can regard DFMs with different κ's as individual models and treat the determination of the number of factors as a model selection or a model averaging problem, which will be discussed in more detail in Section 7. In practice, the computation of marginal likelihoods for DFMs, which are needed for the evaluation of posterior model probabilities, is numerically challenging. Lopes and West (2004) discuss the computation of marginal likelihoods for a static factor model in which the factors are *iid*. The authors also consider a MCMC approach where the number of factors is treated as an unknown parameter and is drawn jointly with all the other parameters.

6.2.3 Applications of Dynamic Factor Models

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 United States)? 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. The authors estimate a DFM 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 country-specific cycles. These factors are assumed to evolve independently from one another. The authors find that international business-cycle comovement is significant. In terms of the variance decomposition of output in the G7 countries, for instance, world cycles are on average as important as country-specific cycles, in the

sense that world and country-specific cycles explain a similar share of the variance of output growth. For the entire world, country-specific cycles are, not surprisingly, much more important than world cycles. Regional cycles are not particularly important at all, suggesting that integration is no higher within regions than across regions.

The study of house prices is another interesting application of factor models. House prices have both an important national and regional component, where the former is associated with nationwide conditions (for example, stance of monetary policy and the national business cycle), while the latter is associated with regional business cycles and other region-specific conditions (for example, migration and demographics). Del Negro and Otrok (2007) apply dynamic factor models to study regional house prices in the US.

In a Bayesian framework estimating models where regional or country-specific factors are identified by imposing the restriction that the respective factors have zero loadings on series that do not belong to that region or country is quite straightforward. Models with such restrictions are harder to estimate using nonparametric methods such as principal components. Moreover, using Bayesian methods, we can conduct inference on the country factors even if the number of series per country is small, as is the case in Kose, Otrok, and Whiteman (2003), while nonparametric methods have a harder time characterizing the uncertainty that results from having a small cross section.

6.2.4 Extensions and Alternative Approaches

We briefly discuss four extensions of the basic DFM presented above. These extensions include Factor Augmented VARs, DFMs with time-varying parameters, hierarchical DFMs, and hybrid models that combine a DSGE model and a DFM.

Factor Augmented VARs: Bernanke, Boivin, and Eliasz (2005) introduce Factor augmented VARs (or FAVARs). The FAVAR approach introduces two changes to the standard factor model. First, the FAVAR allows for additional observables $y_{0,t}$, for example, the federal funds rate, to enter the measurement equation, which becomes

$$y_{i,t} = a_i + \gamma_i y_{0,t} + \lambda_i f_t + \xi_{i,t}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{126}$$

where $y_{0,t}$ and γ_i are $m \times 1$ and $1 \times m$ vectors, respectively. Second, the observable vector $y_{0,t}$ and the unobservable factor f_t are assumed to jointly follow a vector autoregressive process of order q:

$$\begin{bmatrix} f_t \\ y_{0,t} \end{bmatrix} = \Phi_{0,1} \begin{bmatrix} f_{t-1} \\ y_{0,t-1} \end{bmatrix} + \ldots + \Phi_{0,q} \begin{bmatrix} f_{t-q} \\ y_{0,t-q} \end{bmatrix} + u_{0,t}, \quad u_{0,t} \sim iidN(0, \Sigma_0), \tag{127}$$

which is the reason for the term *factor augmented* VAR. The $\Phi_{0,j}$ matrices are now of size $(\kappa + m) \times (\kappa + m)$. The innovation vector $u_{0,t}$ is still assumed to be normally distributed with mean 0 and variance Σ_0, with the difference that the variance-covariance matrix Σ_0 is no longer restricted to be diagonal. The idiosyncratic components $\xi_{i,t}$ evolve according to (114), and the innovations to their law of motion $u_{i,t}$ are subject to the distributional assumptions $u_{i,t} \sim N(0, \sigma_i^2)$. Moreover, we maintain the assumption that the innovations $u_{i,t}$ are independent across i and independent of $u_{0,t}$ at all leads and lags. In order to achieve identification, Bernanke, Boivin, and Eliasz (2005) assume that (i) the $\kappa \times \kappa$ matrix obtained by stacking the first κ λ_i's equals the identity I_κ (as in Example 6.3) and (ii) the $\kappa \times m$ matrix obtained by stacking the first κ γ_i's is composed of zeros.

The appeal of the FAVAR is that it affords a combination of factor analysis with the structural VAR analysis described in Section 2.4. In particular, one can assume that the vector of reduced-form shocks $u_{0,t}$ relates to a vector of structural shocks $\epsilon_{0,t}$ as in (21):

$$u_{0,t} = \Sigma_{0,tr} \Omega_0 \epsilon_{0,t}, \tag{128}$$

where Σ_0^{tr} is the unique lower-triangular Cholesky factor of Σ_0 with nonnegative diagonal elements, and Ω_0 is an arbitrary orthogonal matrix. Bernanke, Boivin, and Eliasz (2005) apply their model to study the effects of monetary policy shocks in the United States. They identify monetary policy shocks by imposing a short-run identification scheme where Ω_0 is diagonal as in Example 2.1. This identification implies that the central bank responds contemporaneously to the information contained in the factors. In contrast, unanticipated changes in monetary policy only affect the factors with a one-period lag.

At least in principle, conducting inference in a FAVAR is a straightforward application of the tools described in Section 6.2.2. For given factors, obtaining the posterior distribution for the parameters of (126) and (127) is straightforward. Likewise, the factors can be drawn using expressions (126) and the first κ equations of the VAR

in (127), as the measurement and transition equations, respectively, in a state-space representation.

Time-Varying Parameters: For the same reasons that it may be useful to allow parameter variation in a VAR as we saw in Section 5, we may want to allow for time-variation in the parameters of a factor model. For instance, comovements 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, and so forth). Del Negro and Otrok (2008) accomplish that by modifying the standard factor model in two ways. First, they make the loadings vary over time. This feature allows for changes in the sensitivity of individual series to common factors. The second innovation amounts to introducing stochastic volatility in the law of motion of the factors and the idiosyncratic shocks. This feature accounts for changes in the relative importance of common factors and of idiosyncratic shocks. Both loadings and volatilities evolve according to a random walk without drift as in Cogley and Sargent (2005b). Del Negro and Otrok (2008) apply this model to study the time-varying nature of international business cycles, in the attempt to determine whether the Great Moderation has country-specific or international roots. Mumtaz and Surico (2008) introduce time-variation in the law of motion of the factors (but not in any of the other parameters) and use their model to study cross-country inflation data.

Hierarchical factors: Ng, Moench, and Potter (2008) pursue a modeling strategy different from the one outlined in Section 6.2.1. Their approach entails building a hierarchical set of factor models, where the hierarchy is determined by the level of aggregation. For concreteness, in the study of international business cycles – the application discussed in the previous section – the three levels of aggregation are country, regional, and world. Only the most disaggregated factors – the country-level factors – would appear in the measurement equation (112). In turn, the country factors evolve according to a factor model in which the common components are the factors at the next level of aggregation (the regional factors). Similarly, the regional factors evolve according to a factor model in which the common components are the the world factors. This approach is more parsimonious than the one used by Kose, Otrok, and Whiteman (2003).

Combining DSGE Models and Factor Models: Boivin and Giannoni (2006a) estimate a DSGE-DFM that equates the latent factors with the state variables

of a DSGE model. Accordingly, the factor dynamics are therefore subject to the restrictions implied by the DSGE model and take the form

$$f_t = \Phi_1(\theta_{DSGE}) f_{t-1} + \Phi_\epsilon(\theta_{DSGE})\epsilon_t, \tag{129}$$

where the vector f_t now comprises the minimal set of state variables associated with the DSGE model and θ_{DSGE} is the vector of structural DSGE model parameters. In the context of the simple stochastic growth model analyzed in Section 4, this vector would contain the capital stock as well as the two exogenous processes. Equation (129) is then combined with measurement equations of the form (112). Since in the DSGE-DFM the latent factors have a clear economic interpretation, it is in principle much easier to elicit prior distributions for the loadings λ_i. For instance, suppose $y_{i,t}$ corresponds to log GDP. The solution of the stochastic growth model delivers a functional relationship between log GDP and the state variables of the DSGE model. This relationship can be used to center a prior distribution for λ_i. Details of how to specify such a prior can be found in Kryshko (2010).

As before, define $\theta_i = [a_i, \lambda_i, \sigma_i, \phi_{i,1}, \ldots, \phi_{i,p}]'$, $i = 1, \ldots, n$. Inference in a DSGE-DFM can be implemented with a Metropolis-within-Gibbs sampler that iterates over (i) the conditional posterior distributions of $\{\theta_i\}_{i=1}^n$ given $(F_{1:T}, \theta_{DSGE}, Y_{1:T})$; (ii) the conditional distribution of $F_{1:T}$ given $(\{\theta_i\}_{i=1}^n, \theta_{DSGE}, Y_{1:T})$; and (iii) the distribution of θ_{DSGE} given $(\{\theta_i\}_{i=1}^n, Y_{1:T})$. Steps (i) and (ii) resemble Steps 1 and 3 in Algorithm 6.1, whereas Step (iii) can be implemented with a modified version of the Random-Walk-Metropolis step described in Algorithm 4.1. Details are provided in Boivin and Giannoni (2006a) and Kryshko (2010).

Boivin and Giannoni (2006a) use their DSGE-DFM to relate DSGE model variables such as aggregate output, consumption, investment, hours worked, wages, inflation, and interest rates to multiple observables, that is, multiple measures of employment and labor usage, wage rates, price inflation, and so forth. Using multiple (noisy) measures implicitly allows a researcher to obtain a more precise measure of DSGE model variables – provided the measurement errors are approximately independent – and thus sharpens inference about the DSGE model parameters and the economic state variables, as well as the shocks that drive the economy. Kryshko (2010) documents that the space spanned by the factors of a DSGE-DFM is very similar to the space spanned by factors extracted from an unrestricted DFM. He then uses the DSGE-DFM to study the effect of unanticipated changes in technology

and monetary policy, which are elements of the vector ϵ_t in (129), on a large cross section of macroeconomic variables.

7 Model Uncertainty

The large number of vector autoregressive and dynamic stochastic general equilibrium models encountered thus far, combined with great variation in the implications for policy across models, makes the problem of model uncertainty a compelling one in macroeconometrics. More specifically, in the context of VARs there is uncertainty about the number of lags and cointegration relationships as well as appropriate restrictions for identifying policy rules or structural shocks. In the context of a 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, a model is formally defined as a joint distribution of data and parameters. Thus, both the likelihood function $p(Y|\theta_{(i)}, \mathcal{M}_i)$ and the prior density $p(\theta_{(i)}|\mathcal{M}_i)$ are part of the specification of a model $\mathcal{M}_i$. Model uncertainty is conceptually not different from parameter uncertainty, which is illustrated in the following example.

Example 7.1: Consider the two (nested) models:

$$\begin{aligned}\mathcal{M}_1: &\quad y_t = u_t, \quad u_t \sim iidN(0,1),\\ \mathcal{M}_2: &\quad y_t = \theta_{(2)}x_t + u_t, \quad u_t \sim iidN(0,1), \quad \theta_{(2)} \sim N(0,1).\end{aligned}$$

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 $\pi_{i,0}$. Suppose we assign prior probability $\pi_{1,0} = \lambda$ to $\mathcal{M}_1$. Then the mixture of $\mathcal{M}_1$ and $\mathcal{M}_2$ is equivalent to a model $\mathcal{M}_0$

$$\mathcal{M}_0: \quad y_t = \theta_{(0)}x_t + u_t, \quad u_t \sim iidN(0,1), \quad \theta_{(0)} \sim \begin{cases} 0 & \text{with prob. } \lambda \\ N(0,1) & \text{with prob. } 1-\lambda \end{cases}. \quad \square$$

In principle, one could try to construct a prior distribution on a sufficiently large parameter space such that model uncertainty can be represented as parameter uncertainty. However, as evident from the example, this prior distribution would have to assign nonzero probability to certain lower-dimensional subspaces, which complicates the computation of the posterior distribution. Thus, in most of the applications considered in this chapter such an approach is impractical, and it is useful to regard restricted versions of a large encompassing model as models themselves, for example VARs of lag length $p = 1, \ldots, p_{max}$ and cointegration rank $r = 1, \ldots, n$ or a collection of linearized DSGE models, which can all be nested in an unrestricted state-space model.

The remainder of this section is organized as follows. Section 7.1 discusses the computation of posterior model probabilities and their use in selecting among a collection of models. Rather than first selecting a model and then conditioning on the selected model in the subsequent analysis, it may be more desirable to average across models and to take model uncertainty explicitly into account when making decisions. We use a stylized optimal monetary policy example to highlight this point in Section 7.2. In many macroeconomic applications, in particular those that are based on DSGE models, posterior model probabilities are often overly decisive, in that one specification essentially attains posterior probability one and all other specifications receive probability zero. These decisive probabilities found in individual studies are difficult to reconcile with the variation in results and model rankings found across different studies and therefore are in some sense *implausible*. In view of potentially implausible posterior model probabilities, a decision maker might be inclined to robustify her decisions. These issues are discussed in Section 7.3.

7.1 Posterior Model Probabilities and Model Selection

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

$$\pi_{i,T} = \frac{\pi_{i,0} p(Y_{1:T}|\mathcal{M}_i)}{\sum_{j=1}^{M} \pi_{j,0} p(Y_{1:T}|\mathcal{M}_j)}, \quad p(Y_{1:T}|\mathcal{M}_i) = \int p(Y_{1:T}|\theta_{(i)}, \mathcal{M}_i) p(\theta_{(i)}|\mathcal{M}_i) d\theta_{(i)}, \tag{130}$$

where $p(Y_{1:T}|\mathcal{M}_i)$ is the marginal likelihood or data density associated with model $\mathcal{M}_i$. As long as the likelihood functions $p(Y_{1:T}|\theta_{(i)}, \mathcal{M}_i)$ and prior densities $p(\theta_{(i)}|\mathcal{M}_i)$ are properly normalized for all models, the posterior model probabilities are well defined. Since for any model $\mathcal{M}_i$

$$\ln p(Y_{1:T}|\mathcal{M}_i) = \sum_{t=1}^{T} \ln \int p(y_t|\theta_{(i)}, Y_{1,t-1}, \mathcal{M}_i) p(\theta_{(i)}|Y_{1,t-1}, \mathcal{M}_i) d\theta_{(i)}, \tag{131}$$

log marginal likelihoods can be interpreted as the sum of one-step-ahead predictive scores. The terms on the right-hand side of (131) provide a decomposition of the one-step-ahead predictive densities $p(y_t|Y_{1,t-1}, \mathcal{M}_i)$. This decomposition highlights the fact that inference about the parameter $\theta_{(i)}$ is based on time $t-1$ information, when making the prediction for y_t. The predictive score is small whenever the predictive distribution assigns a low density to the observed y_t. It is beyond the scope of this chapter to provide a general discussion of the use of posterior model probabilities or odds ratios for model comparison. A survey is provided by Kass and Raftery (1995). In turn, we shall highlight a few issues that are important in the context of macroeconometric applications.

We briefly mentioned in Sections 2.2 (hyperparameter choice for Minnesota prior) and 4.3 (prior elicitation for DSGE models) that in practice priors are often based on presample (or training sample) information. Since in time-series models observations have a natural ordering, we could regard observations $Y_{1:T^*}$ as presample and $p(\theta|Y_{1:T^*})$ as a prior for θ that incorporates this presample information. Conditional on $Y_{1:T^*}$, the marginal likelihood function for subsequent observations $Y_{T^*+1:T}$ is given by

$$p(Y_{T^*+1:T}|Y_{1:T^*}) = \frac{p(Y_{1:T})}{p(Y_{1:T^*})} = \int p(Y_{T^*+1:T}|Y_{1:T^*}, \theta) p(\theta|Y_{1:T^*}) d\theta. \tag{132}$$

The density $p(Y_{T^*+1:T}|Y_{1:T^*})$ is often called predictive (marginal) likelihood and can replace the marginal likelihood in (130) in the construction of posterior model probabilities, provided the prior model probabilities are also adjusted to reflect the presample information $Y_{1:T^*}$. As before, it is important that $p(\theta|Y_{1:T^*})$ be a proper density. In the context of a VAR, a proper prior could be obtained by replacing the dummy observations Y^* and X^* with presample observations. Two examples of papers that use predictive marginal likelihoods to construct posterior model probabilities are Schorfheide (2000), who computes posterior odds for a collection of VARs and DSGE models, and Villani (2001), who uses them to evaluate lag length

and cointegration rank restrictions in vector autoregressive models. A more detailed discussion of predictive likelihoods can be found in Geweke (2005). An application of predictive likelihoods to forecast combination and model averaging is provided by Eklund and Karlsson (2007).

While the calculation of posterior probabilities is conceptually straightforward, it can be computationally challenging. There are only a few instances, such as the VAR model in (1) with conjugate MNIW prior, in which the marginal likelihood $p(Y) = \int p(Y|\theta)p(\theta)d\theta$ can be computed analytically. In fact, for priors represented through dummy observations the formula is given in (15). We also mentioned in Section 4.7.1 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)'s modified harmonic mean estimator or the method proposed by Chib and Jeliazkov (2001). A more detailed discussion of numerical approximation techniques for marginal likelihoods is provided in Chib (This Volume). 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 (1978), which is known as Schwarz Criterion or Bayesian Information Criterion (BIC). Phillips (1996) and Chao and Phillips (1999) provide extensions to nonstationary time-series models and reduced-rank VARs.

Schorfheide (2000) compares Laplace approximations of marginal likelihoods for two small-scale DSGE models and bivariate VARs with 2-4 lags to numerical approximations based on a modified harmonic mean estimator. The VARs were specified such that the marginal likelihood could be computed exactly. The approximation error of the numerical procedure was at most 0.02 for log densities, whereas the error of the Laplace approximation was around 0.5. While the exact marginal likelihood was not available for the DSGE models, the discrepancy between the modified harmonic mean estimator and the Laplace approximation was around 0.1 on a log scale. While the results reported in Schorfheide (2000) are model and data specific, the use of numerical procedures to approximate marginal likelihood functions is generally preferable for two reasons. First, posterior inference is typically based on simulation-based methods, and the marginal likelihood approximation can often be constructed from the output of the posterior simulator with very little additional effort. Second, the approximation error can be reduced to a desired level by increasing

the number of parameter draws upon which the approximation is based.

Posterior model probabilities are often used to select a model specification upon which any subsequent inference is conditioned. While it is generally preferable to average across all model specifications with nonzero posterior probability, a model selection approach might provide a good approximation if the posterior probability of one model is very close to one, the probabilities associated with all other specifications are very small, and the loss of making inference or decisions based on the highest posterior probability model is not too large if one of the low probability models is in fact correct. We shall elaborate on this point in Example 7.2 in Section 7.2. A rule for selecting one out of M models can be formally derived from the following decision problem. Suppose that a researcher faces a loss of zero if she chooses the "correct" model and a loss of $\alpha_{ij} > 0$ if she chooses model $\mathcal{M}_i$ although $\mathcal{M}_j$ is correct. If the loss function is symmetric in the sense that $\alpha_{ij} = \alpha$ for all $i \neq j$, then it is straightforward to verify that the posterior expected loss is minimized by selecting the model with the highest posterior probability. A treatment of model selection problems under more general loss functions can be found, for instance, in Bernardo and Smith (1994).

If one among the M models $\mathcal{M}_1, \ldots, \mathcal{M}_M$ is randomly selected to generate a sequence of observations $Y_{1:T}$, then under fairly general conditions the posterior probability assigned to that model will converge to one as $T \longrightarrow \infty$. In this sense, Bayesian model selection procedures are consistent from a frequentist perspective. An early version of this result for general linear regression models was proved by Halpern (1974). The consistency result remains valid if the marginal likelihoods that are used to compute posterior model probabilities are replaced by Laplace approximations (see, for example, Schwarz (1978) and Phillips and Ploberger (1996)). These Laplace approximations highlight the fact that log marginal likelihoods can be decomposed into a goodness-of-fit term, comprising the maximized log likelihood function $\max_{\theta_{(i)} \in \Theta_{(i)}} \ln p(Y_{1:T}|\theta_{(i)}, \mathcal{M}_i)$ and a term that penalizes the dimensionality, which in case of Schwarz's approximation takes the form of $-(k_i/2)\ln T$, where k_i is the dimension of the parameter vector $\theta_{(i)}$. Moreover, the consistency is preserved in nonstationary time-series models. Chao and Phillips (1999), for instance, prove that the use of posterior probabilities leads to a consistent selection of cointegration rank and lag length in vector autoregressive models.

7.2 Decision Making and Inference with Multiple Models

Economic policy makers are often confronted with choosing policies under model uncertainty.[6] Moreover, policy decisions are often made under a fairly specific loss function that is based on some measure of welfare. This welfare loss function might either be fairly ad-hoc – for example, the variability of aggregate output and inflation – or micro-founded albeit model-specific – for instance, the utility of a representative agent in a DSGE model. The optimal decision from a Bayesian perspective is obtained by minimizing the expected loss under a mixture of models. Conditioning on the highest posterior probability model can lead to suboptimal decisions. At a minimum, the decision maker 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 following example provides an illustration.

Example 7.2: Suppose that output y_t and inflation π_t are related to each other according to one of the two Phillips curve relationships

$$\mathcal{M}_i: \; y_t = \theta(\mathcal{M}_i)\pi_t + \epsilon_{s,t}, \quad \epsilon_{s,t} \sim iidN(0,1), \quad i = 1,2, \tag{133}$$

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

$$\pi_t = m_t + \epsilon_{d,t}, \quad \epsilon_{d,t} \sim iidN(0,1), \tag{134}$$

where $\epsilon_{d,t}$ is a demand shock. Finally, assume that up until period T monetary policy was $m_t = 0$. All variables in this model are meant to be in log deviations from some steady state.

In period T, the central bank is considering a class of new monetary policies, indexed by δ:

$$m_t = -\epsilon_{d,t} + \delta\epsilon_{s,t}. \tag{135}$$

δ controls the strength of the central bank's reaction to supply shocks. This class of policies is evaluated under the loss function

$$\widetilde{L}_t = (\pi_t^2 + y_t^2). \tag{136}$$

[6] Chamberlain (This Volume) studies the decision problem of an individual who chooses between two treatments from a Bayesian perspective.

If one averages with respect to the distribution of the supply shocks, the expected period loss associated with a particular policy δ under model $\mathcal{M}_i$ is

$$L(\mathcal{M}_i, \delta) = (\delta\theta(\mathcal{M}_i) + 1)^2 + \delta^2. \tag{137}$$

To provide a numerical illustration, we let

$$\theta(\mathcal{M}_1) = 1/10, \quad \theta(\mathcal{M}_2) = 1, \quad \pi_{1,T} = 0.61, \quad \pi_{2,T} = 0.39.$$

Here, $\pi_{i,T}$ denotes the posterior probability of model $\mathcal{M}_i$ at the end of period T. We will derive the optimal decision and compare it with two suboptimal procedures that are based on a selection step.

First, from a Bayesian perspective it is optimal to minimize the posterior risk (expected loss), which in this example is given by

$$\mathcal{R}(\delta) = \pi_{1,T} L(\mathcal{M}_1, \delta) + \pi_{2,T} L(\mathcal{M}_2, \delta). \tag{138}$$

A straightforward calculation leads to $\delta^* = \text{argmin}_\delta \, \mathcal{R}(\delta) = -0.32$ and the posterior risk associated with this decision is $\mathcal{R}(\delta^*) = 0.85$. Second, suppose that the policy maker had proceeded in two steps: (i) select the highest posterior probability model; and (ii) conditional on this model, determine the optimal choice of δ. The highest posterior probability model is $\mathcal{M}_1$, and, conditional on $\mathcal{M}_1$, it is optimal to set $\delta^*(\mathcal{M}_1) = -0.10$. The risk associated with this decision is $\mathcal{R}(\delta^*(\mathcal{M}_1)) = 0.92$, which is larger than $\mathcal{R}(\delta^*)$ and shows that it is suboptimal to condition the decision on the highest posterior probability model. In particular, this model-selection-based procedure completely ignores the loss that occurs if in fact $\mathcal{M}_2$ is the correct model.

Third, suppose that the policy maker relies on two advisors $\mathcal{A}_1$ and $\mathcal{A}_2$. Advisor $\mathcal{A}_i$ recommends that the policy maker implement the decision $\delta^*(\mathcal{M}_i)$, which minimizes the posterior risk if only model $\mathcal{M}_i$ is considered. If the policy maker implements the recommendation of advisor $\mathcal{A}_i$, taking into account the posterior model probabilities $\pi_{i,T}$, then Table 4 provides the matrix of relevant expected losses. Notice that there is a large loss associated with $\delta^*(\mathcal{M}_2)$ if in fact $\mathcal{M}_1$ is the correct model. Thus, even though the posterior odds favor the model entertained by $\mathcal{A}_1$, it is preferable to implement the recommendation of advisor $\mathcal{A}_2$ because $\mathcal{R}(\delta^*(\mathcal{M}_2)) < \mathcal{R}(\delta^*(\mathcal{M}_1))$. However, while choosing between $\delta^*(\mathcal{M}_1)$ and $\delta^*(\mathcal{M}_2)$ is preferable to conditioning on the highest posterior probability model, the best among the two decisions, $\delta^*(\mathcal{M}_2)$, is inferior to the optimal decision δ^*, obtained by minimizing

Table 4: Expected Losses

Decision	$\mathcal{M}_1$	$\mathcal{M}_2$	Risk $\mathcal{R}(\delta)$
$\delta^* = -0.32$	1.04	0.56	0.85
$\delta^*(\mathcal{M}_1) = -0.1$	0.99	0.82	0.92
$\delta^*(\mathcal{M}_2) = -0.5$	1.15	0.50	0.90

the overall posterior expected loss. In fact, in this numerical illustration the gain from averaging over models is larger than the difference between $\mathcal{R}(\delta^*(\mathcal{M}_1))$ and $\mathcal{R}(\delta^*(\mathcal{M}_2))$. □

In more realistic applications, the two simple models would be replaced by more sophisticated DSGE models. These models would themselves involve unknown parameters. Cogley and Sargent (2005a) provide a nice macroeconomic illustration of the notion that one should not implement the decision of the highest posterior probability model if it has disastrous consequences in case one of the other models is correct. The authors consider a traditional Keynesian model with a strong output and inflation trade-off versus a model in which the Phillips curve is vertical in the long run. According to Cogley and Sargent's analysis, the posterior probability of the Keynesian model was already very small by the mid-1970s, and the natural rate model suggested implementing a disinflation policy. However, the costs associated with this disinflation were initially very high if, in fact, the Keynesian model provided a better description of the U.S. economy. The authors conjecture that this consideration may have 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 is given by the mixture

$$p(y_{T+h}|Y_{1:T}) = \sum_{i=1}^{M} \pi_{i,T} p(y_{T+h}|Y_{1:T}, \mathcal{M}_i). \tag{139}$$

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. There exists an extensive literature on applications of Bayesian model averaging. For instance,

Min and Zellner (1993) use posterior model probabilities to combine forecasts, and Wright (2008) uses Bayesian model averaging to construct exchange rate forecasts. If the goal is to generate point predictions under a quadratic loss function, then it is optimal to average posterior mean forecasts from the M models, using the posterior model probabilities as weights. This is a special case of Bayesian forecast combination, which is discussed in more general terms in Geweke and Whiteman (2006). Strachan and van Dijk (2006) average across VARs with different lag lengths and cointegration restrictions to study the dynamics of the Great Ratios.

If the model space is very large, then the implementation of model averaging can be challenging. Consider the empirical Illustration 2.1, which involved a 4-variable VAR with 4 lags, leading to a coefficient matrix Φ with 68 elements. Suppose one constructs submodels by restricting VAR coefficients to zero. Based on the exclusion of parameters, one can in principle generate $2^{68} \approx 3 \cdot 10^{20}$ submodels. Even if one restricts the set of submodels by requiring that a subset of the VAR coefficients are never restricted to be zero and one specifies a conjugate prior that leads to an analytical formula for the marginal likelihoods of the submodels, the computation of posterior probabilities for all submodels can be a daunting task. As an alternative, George, Ni, and Sun (2008) develop a stochastic search variable selection algorithm for a VAR that automatically averages over high posterior probability submodels. The authors also provide detailed references to the large literature on Bayesian variable selection in problems with large sets of potential regressors. In a nutshell, George, Ni, and Sun (2008) introduce binary indicators that determine whether a coefficient is restricted to be zero. An MCMC algorithm then iterates over the conditional posterior distribution of model parameters and variable selection indicators. However, as is typical of stochastic search applications, the number of restrictions actually visited by the MCMC simulation is only a small portion of all possible restrictions.

Bayesian model averaging has also become popular in growth regressions following the work of Fernandez, Ley, and Steel (2001), Sala-i Martin, Doppelhofer, and Miller (2004), and Masanjala and Papageorgiou (2008). The recent empirical growth literature has identified a substantial number of variables that potentially explain the rate of economic growth in a cross section or panel of countries. Since there is uncertainty about exactly which explanatory variables to include in a growth regression, Bayesian model averaging is an appealing procedure. The paper by Sala-i Martin, Doppelhofer, and Miller (2004) uses a simplified version of Bayesian model averag-

ing, in which marginal likelihoods are approximated by Schwarz (1978)'s Laplace approximation and posterior means and covariances are replaced by maxima and inverse Hessian matrices obtained from a Gaussian likelihood function.

7.3 Difficulties in Decision-Making with Multiple Models

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. 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. Moreover, suppose that – as is the case in New Keynesian DSGE models – 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 nullify 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, then under relevant loss functions we should completely disregard the possibility that wages are sticky. In a related study, Del Negro, Schorfheide, Smets, and Wouters (2007) compare versions of DSGE models with nominal rigidities in which those households (firms) that are unable to reoptimize their wages (prices) are indexing their past price either 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) 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 in this sense implausible from a meta perspective.

The problem of implausible odds has essentially two dimensions. First, each DSGE model corresponds to a stylized representation of a particular economic mechanism, such as wage or price stickiness, augmented by auxiliary mechanisms that are designed to capture the salient features of the data. By looking across studies, one encounters several representations of essentially the same basic economic mechanism, but each representation attains a different time-series fit and makes posterior probabilities appear fragile across studies. Second, in practice macroeconometricians often work with incomplete model spaces. That is, in addition to the models that are being formally analyzed, researchers have in mind a more sophisticated structural model, which may be too complicated to formalize or too costly (in terms of intellectual and computational resources) to estimate. In some instances, a richly parameterized vector autoregression that is only loosely connected to economic theory serves as a stand-in. In view of these reference models, the simpler specifications are potentially misspecified. For illustrative purpose, we provide two stylized examples in which we explicitly specify the sophisticated reference model that in practice is often not spelled out.

Example 7.3: Suppose that a macroeconomist assigns equal prior probabilities to two stylized models $\mathcal{M}_i$: $y_t \sim iidN(\mu_i, \sigma_i^2)$, $i = 1, 2$, where μ_i and σ_i^2 are fixed. In addition, there is a third model $\mathcal{M}_0$ in the background, given by $y_t \sim iidN(0, 1)$. For the sake of argument, suppose it is too costly to analyze $\mathcal{M}_0$ formally. If a sequence of T observations were generated from $\mathcal{M}_0$, the expected log posterior odds of $\mathcal{M}_1$ versus $\mathcal{M}_2$ would be

$$\begin{aligned} I\!E_0\left[\ln \frac{\pi_{1,T}}{\pi_{2,T}}\right] &= I\!E_0\left[-\frac{T}{2}\ln\sigma_1^2 - \frac{1}{2\sigma_1^2}\sum_{t=1}^{T}(y_t-\mu_1)^2 \right. \\ &\quad \left. -\left(-\frac{T}{2}\ln\sigma_2^2 - \frac{1}{2\sigma_2^2}\sum_{t=1}^{T}(y_t-\mu_2)^2\right)\right] \\ &= -\frac{T}{2}\left[\ln\sigma_1^2 + \frac{1}{\sigma_1^2}(1+\mu_1^2)\right] + \frac{T}{2}\left[\ln\sigma_2^2 + \frac{1}{\sigma_2^2}(1+\mu_2^2)\right], \end{aligned}$$

where the expectation is taken with respect to $y_1, \ldots, y_T$ under $\mathcal{M}_0$. Suppose that the location parameters μ_1 and μ_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 model. If the two models are based on similar auxiliary assumptions, that is, $\sigma_1^2 \approx \sigma_2^2$, then the posterior odds are clearly driven by the key economic contents of the two models. If, however, the auxiliary assumptions made in the two models are very different, it is possible that the posterior odds and hence the ranking of models $\mathcal{M}_1$ and $\mathcal{M}_2$ are dominated by the auxiliary assumptions, σ_1^2 and σ_2^2, rather than by the economic contents, μ_1 and μ_2, of the models. □

Example 7.4: This example is adapted from Sims (2003). Suppose that a researcher considers the following two models. $\mathcal{M}_1$ implies $y_t \sim iidN(-0.5, 0.01)$ and model $\mathcal{M}_2$ implies $y_t \sim iidN(0.5, 0.01)$. There is a third model, $\mathcal{M}_0$, given by $y_t \sim iidN(0, 1)$, that is too costly to be analyzed formally. The sample size is $T = 1$. Based on equal prior probabilities, the posterior odds in favor of model $\mathcal{M}_1$ are

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

Thus, for values of y_1 less than -0.05 or greater than 0.05 the posterior odds are greater than $e^5 \approx 150$ in favor of one of the models, which we shall term *decisive*. The models $\mathcal{M}_1$ ($\mathcal{M}_2$) assign a probability of less than 10^{-6} outside the range $[-0.55, -0.45]$ ($[0.45, 0.55]$). Using the terminology of the prior predictive checks described in Section 4.7.2, for observations outside these ranges one would conclude that the models have severe difficulties explaining the data. For any observation falling into the intervals $(-\infty, -0.55]$, $[-0.45, -0.05]$, $[0.05, 0.45]$, and $[0.55, \infty)$, one would obtain decisive posterior odds and at the same time have to conclude that the empirical observation is difficult to reconcile with the models $\mathcal{M}_1$ and $\mathcal{M}_2$. At the same time, the reference model $\mathcal{M}_0$ assigns a probability of almost 0.9 to these intervals. □

As illustrated through these two stylized examples, the problems in the use of posterior probabilities in the context of DSGE models are essentially twofold. First, DSGE models tend to capture one of many possible representations of a particular economic mechanism. Thus, one might be able to find versions of these models that preserve the basic mechanisms but deliver very different odds. Second, the models often suffer from misspecification, which manifests itself through low posterior probabilities in view of more richly parameterized vector autoregressive models that are less tightly linked to economic theory. Posterior odds exceeding e^{50} in a sample of

120 observations are suspicious (to us) and often indicate that we should compare different models or consider a larger model space.

Sims (2003) recommends introducing 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 multimodal 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 a challenge. Geweke (2010) proposes to deal with incomplete model spaces by pooling models. This pooling amounts essentially to creating a convex combination of one-step-ahead predictive distributions, which are derived from individual models. The time-invariant weights of this mixture of models is then estimated by maximizing the log predictive score for this mixture (see Expression (131)).

In view of these practical limitations associated with posterior model probabilities, a policy maker might find it attractive to robustify her decision. In fact, there is a growing literature in economics that studies the robustness of decision rules to model misspecification (see Hansen and Sargent (2008)). Underlying this robustness is often a static or dynamic two-person zero-sum game, which we illustrate in the context of Example 7.2.

Example 7.2, Continued: Recall the monetary policy problem described at the beginning of this section. Suppose scepticism about the posterior probabilities $\pi_{1,T}$ and $\pi_{2,T}$ generates some concern about the robustness of the policy decision to perturbations of these model probabilities. This concern can be represented through the following game between the policy maker and a fictitious adversary, called nature:

$$\min_{\delta} \max_{q\in[0,1/\pi_{1,T}]} \quad q\pi_{1,T}L(\mathcal{M}_1,\delta) + (1-q\pi_{1,T})L(\mathcal{M}_2,\delta) + \frac{1}{\tau}\Big[\pi_{1,T}\ln(q\pi_{1,T}) + (1-\pi_{1,T})\ln(1-q\pi_{1,T})\Big]. \tag{140}$$

Here, nature uses q to distort the posterior model probability of model $\mathcal{M}_1$. 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 (140) penalizes the distortion as a function of the Kullback-Leibler divergence between the undistorted and distorted probabilities. If τ is equal to zero, then the penalty is infinite and nature will not distort $\pi_{1,T}$. If, however, $\tau = \infty$, then conditional on a particular δ nature will set

Table 5: Nash Equilibrium as a Function of Risk Sensitivity τ

τ	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

$q = 1/\pi_{1,T}$ if $L(\mathcal{M}_1, \delta) > L(\mathcal{M}_2, \delta)$ and $q = 0$ otherwise. For selected values of τ, the Nash equilibrium is summarized in Table 5. In our numerical illustration, $L(\mathcal{M}_1, \delta) > L(\mathcal{M}_2, \delta)$ in the relevant region for δ. 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 δ to a supply shock. □

The particular implementation of robust decision making in Example 7.2 is very stylized. While it is our impression that in actual decision making a central bank is taking the output of formal Bayesian analysis more and more seriously, the final decision about economic policies is influenced by concerns about robustness and involves adjustments of model outputs in several dimensions. These adjustments may reflect some scepticism about the correct formalization of the relevant economic mechanisms as well as the availability of information that is difficult to process in macroeconometric models such as VARs and DSGE models.

References

Adolfson, M., J. Lindé, and M. Villani (2007): "Forecasting Performance of an Open Economy Dynamic Stochastic General Equilibrium Model," *Econometric Reviews*, 26(2-4), 289–328.

Altug, S. (1989): "Time-to-Build and Aggregate Fluctuations: Some New Evidence," *International Economic Review*, 30(4), 889–920.

An, S., and F. Schorfheide (2007a): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26(2-4), 113–172.

——— (2007b): "Bayesian Analysis of DSGE Models–Rejoinder," *Econometric Reviews*, 26(2-4), 211–219.

Aruoba, S. B., J. Fernández-Villaverde, and J. F. Rubio-Ramírez (2004): "Comparing Solution Methods for Dynamic Equilibrium Economies," *Journal of Economic Dynamics and Control*, 30(12), 2477–2508.

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

Bernardo, J. E., and A. F. Smith (1994): *Bayesian Theory*. John Wiley & Sons, Hoboken.

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

Boivin, J., and M. P. Giannoni (2006a): "DSGE Models in a Data Rich Enviroment," *NBER Working Paper*, 12772.

——— (2006b): "Has Monetary Policy Become More Effective," *Review of Economics and Statistics*, 88(3), 445–462.

Canova, F. (1994): "Statistical Inference in Calibrated Models," *Journal of Applied Econometrics*, 9, S123–144.

Canova, F., and M. Ciccarelli (2009): "Estimating Multi-country VAR Models," *International Economic Review*, 50(3), 929–959.

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

Canova, F., and L. Gambetti (2009): "Structural Changes in the US Economy: Is There a Role for Monetary Policy?," *Journal of Economic Dynamics and Control*, 33(2), 477–490.

Carter, C., and R. Kohn (1994): "On Gibbs Sampling for State Space Models," *Biometrika*, 81(3), 541–553.

Chamberlain, G. (This Volume): "Bayesian Aspects of Treatment Choice," in *Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. K. van Dijk. Oxford University Press.

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.

Chao, J., and P. C. Phillips (1999): "Model Selection in Partially-Nonstationary Vector Autoregressive Processes with Reduced Rank Structure," *Journal of Econometrics*, 91(2), 227–271.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2008): "Are Structural VARs with Long-Run Restrictions Useful in Developing Business Cycle Theory?," *Journal of Monetary Economics*, 55(8), 1337–1352.

Chib, S. (This Volume): "Introduction to Simulation and MCMC Methods," in *Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. K. van Dijk. Oxford University Press.

Chib, S., and E. Greenberg (1994): "Bayes Inference in Regression Models with ARMA(p,q) Errors," *Journal of Econometrics*, 64(1-2), 183–206.

Chib, S., and I. Jeliazkov (2001): "Marginal Likelihoods from the Metropolis Hastings Output," *Journal of the American Statistical Association*, 96(453), 270–281.

Chib, S., and S. Ramamurthy (2010): "Tailored Randomized Block MCMC Methods with Application to DSGE Models," *Journal of Econometrics*, 155(1), 19–38.

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.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary Policy Shocks: What Have We Learned and to What End," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1a, chap. 2, pp. 65–148. North Holland, Amsterdam.

——— (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1), 1–45.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2007): "Assessing Structural VARs," in *NBER Macroeconomics Annual 2006*, ed. by D. Acemoglu, K. Rogoff, and M. Woodford, vol. 21, pp. 1–72. MIT Press, Cambridge.

Clarida, R., J. Gali, and M. Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115(1), 147–180.

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

Cogley, T., S. Morozov, and T. J. 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., and T. J. Sargent (2002): "Evolving Post-World War II U.S. Inflation Dynamics," in *NBER Macroeconomics Annual 2001*, ed. by B. S. Bernanke, and K. Rogoff, vol. 16, pp. 331–88. MIT Press, Cambridge.

——— (2005a): "The Conquest of US Inflation: Learning and Robustness to Model Uncertainty," *Review of Economic Dynamics*, 8(2), 528–563.

——— (2005b): "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 (2008): "Trend Inflation, Indexation, and Inflation Persistence in the New Keynesian Phillips Curve," *American Economic Review*, 98(5), 2101–2126.

Davig, T., and E. M. Leeper (2007): "Generalizing the Taylor Principle," *American Economic Review*, 97(3), 607–635.

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.

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

——— (2000): "A Bayesian Approach to Dynamic Macroeconomics," *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 US'," *Federal Reserve Bank of 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.," *Federal Reserve Bank of New York Staff Report*, 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.

——— (2009): "Monetary Policy with Potentially Misspecified Models," *American Economic Review*, 99(4), 1415–1450.

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.

Edge, R., M. Kiley, and J.-P. Laforte (2009): "A Comparison of Forecast Performance Between Federal Reserve Staff Forecasts, Simple Reduced-Form Models, and a DSGE Model," *Federal Reserve Board of Governors Finance and Economics Discussion Paper Series*, 2009-10.

Eklund, J., and S. Karlsson (2007): "Forecast Combination and Model Averaging Using Predictive Measures," *Econometric Reviews*, 26(2-4), 329–363.

Engle, R. F., and C. W. Granger (1987): "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55(2), 251–276.

Farmer, R., D. Waggoner, and T. Zha (2009): "Understanding Markov Switching Rational Expectations Models," *Journal of Economic Theory*, 144(5), 1849–1867.

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.

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74(4), 1059–1087.

——— (2008): "How Structural are Structural Parameters?," in *NBER Macroeconomics Annual 2007*, ed. by D. Acemoglu, K. Rogoff, and M. Woodford, vol. 22. University of Chicago Press, Chicago, University of Chicago Press.

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

Geweke, J. (1977): "The Dynamic Factor Analysis of Economic Time Series," in *Latent Variables in Socio-Economic Models*, ed. by D. J. Aigner, and A. S. Goldberger, chap. 19. North Holland, Amsterdam.

——— (1996): "Bayesian Reduced Rank Regression in Econometrics," *Journal of Econometrics*, 75(1), 121–146.

——— (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18(1), 1–126.

——— (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Hoboken.

——— (2007): "Bayesian Model Comparison and Validation," *American Economic Review Papers and Proceedings*, 97, 60–64.

——— (2010): *Complete and Incomplete Econometric Models*. Princeton University Press, Princeton.

Geweke, J., and N. Terui (1993): "Bayesian Threshold Autoregressive Models for Nonlinear Time Series," *Journal of Time Series Analysis*, 14(5), 441–454.

Geweke, J., and C. H. Whiteman (2006): "Bayesian Forecasting," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. W. Granger, and A. Timmermann, vol. 1, pp. 3–80. North Holland, Amsterdam.

Geweke, J., and G. Zhou (1996): "Measuring the Pricing Error of the Arbitrage Pricing Theory," *Review of Financial Studies*, 9(2), 557–587.

Giordani, P., M. K. Pitt, and R. Kohn (This Volume): "Bayesian Inference for Time Series State Space Models," in *Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. K. van Dijk. Oxford University Press.

Halpern, E. F. (1974): "Posterior Consistency for Coefficient Estimation and Model Selection in the General Linear Hypothesis," *Annals of Statistics*, 2(4), 703–712.

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

Hansen, L. P., and T. J. Sargent (2008): *Robustness*. Princeton University Press, Princeton.

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.

Ireland, P. N. (2004): "A Method for Taking Models to the Data," *Journal of Economic Dynamics and Control*, 28(6), 1205–1226.

Jacquier, E., and N. G. Polson (This Volume): "Bayesian Econometrics in Finance," in *Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. K. van Dijk. Oxford University Press.

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

James, A. T. (1954): "Normal Multivariate Analysis and the Orthogonal Group," *Annals of Mathematical Statistics*, 25(1), 40–75.

Johansen, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12(2-3), 231–254.

——— (1991): "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models," *Econometrica*, 59(6), 1551–1580.

——— (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, New York.

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

Justiniano, A., G. E. Primiceri, and A. Tambalotti (2009): "Investment Shocks and Business Cycles," *NBER Working Paper*, 15570.

Kadane, J. B. (1974): "The Role of Identification in Bayesian Theory," in *Studies in Bayesian Econometrics and Statistics*, ed. by S. E. Fienberg, and A. Zellner, pp. 175–191. North Holland, Amsterdam.

Kadiyala, K. R., and S. Karlsson (1997): "Numerical Methods for Estimation and Inference in Bayesian VAR-Models," *Journal of Applied Econometrics*, 12(2), 99–132.

Kass, R. E., and A. E. Raftery (1995): "Bayes Factors," *Journal of the American Statistical Association*, 90(430), 773–795.

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, Cambridge.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65(3), 361–393.

King, R. G., C. I. Plosser, and S. Rebelo (1988): "Production, Growth, and Business Cycles: I The Basic Neoclassical Model," *Journal of Monetary Economics*, 21(2-3), 195–232.

Kleibergen, F., and R. Paap (2002): "Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration," *Journal of Econometrics*, 111(2), 223–249.

Kleibergen, F., and H. K. van Dijk (1994): "On the Shape of the Likelihood / Posterior in Cointegration Models," *Econometric Theory*, 10(3-4), 514–551.

Klein, L. R., and R. F. Kosobud (1961): "Some Econometrics of Growth: Great Ratios of Economics," *Quarterly Journal of Economics*, 75(2), 173–198.

Koop, G., and D. Korobilis (2010): "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics," in *Foundations and Trends in Econometrics*. Now Publisher, forthcoming.

Koop, G., R. Leon-Gonzalez, and R. Strachan (2008): "Bayesian Inference in the Time Varying Cointegration Model," *Rimini Center for Economic Analysis Working Paper*, 23-08.

Koop, G., R. Leon-Gonzalez, and R. W. Strachan (2009): "On the Evolution of the Monetary Policy Transmission Mechanism," *Journal of Economic Dynamics and Control*, 33(4), 997–1017.

Koop, G., and S. M. Potter (1999): "Bayes Factors and Nonlinearity: Evidence from Economic Time Series," *Journal of Econometrics*, 88(2), 251–281.

——— (2007): "Estimation and Forecasting in Models with Multiple Breaks," *Review of Economic Studies*, 74(3), 763–789.

——— (2008): "Time-Varying VARs with Inequality Restrictions," *Manuscript, University of Strathclyde and FRB New York*.

——— (2009): "Prior Elicitation in Multiple Change-point Models," *International Economic Review*, 50(3), 751–772.

Koop, G., R. Strachan, H. K. van Dijk, and M. Villani (2006): "Bayesian Approaches to Cointegration," in *Palgrave Handbook of Econometrics*, ed. by T. C. Mills, and K. P. Patterson, vol. 1, pp. 871–898. Palgrave Macmillan, Basingstoke United Kingdom.

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

Kryshko, M. (2010): "Data-Rich DSGE and Dynamic Factor Models," *Manuscript, University of Pennsylvania*.

Lancaster, T. (2004): *An Introduction to Modern Bayesian Econometrics*. Blackwell Publishing.

Leeper, E. M., and J. Faust (1997): "When Do Long-Run Identifiying Restrictions Give Reliable Results?," *Journal of Business & Economic Statistics*, 15(3), 345–353.

Leeper, E. M., and C. A. Sims (1995): "Toward a Modern Macroeconomic Model Usable for Policy Analysis," in *NBER Macroeconomics Annual 1994*, ed. by S. Fischer, and J. J. Rotemberg, pp. 81–118. MIT Press, Cambridge.

Levin, A., A. Onatski, J. C. Williams, and N. Williams (2006): "Monetary Policy Under Uncertainty in Micro-founded Macroeconometric Models," in *NBER Macroeconomics Annual 2005*, ed. by M. Gertler, and K. Rogoff, vol. 20, pp. 229–287. MIT Press, Cambridge.

Litterman, R. B. (1980): "Techniques for Forecasting with Vector Autoregressions," Ph.D. thesis, University of Minnesota.

Lopes, H. F., and M. West (2004): "Bayesian Model Assessment in Factor Analysis," *Statistica Sinica*, 14(1), 41–67.

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

Masanjala, W. H., and C. Papageorgiou (2008): "Rough and Lonely Road to Prosperity: A Reexamination of the Sources of Growth in Africa Using Bayesian Model Averaging," *Journal of Applied Econometrics*, 23(5), 671–682.

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(5), 1464–76.

Min, C.-K., and A. Zellner (1993): "Bayesian and Non-Bayesian Methods for Combining Models and Forecasts with Applications to Forecasting International Growth Rates," *Journal of Econometrics*, 56(1-2), 89–118.

Moon, H. R., and F. Schorfheide (2009): "Bayesian and Frequentist Inference in Partially-Identified Models," *NBER Working Paper*, 14882.

Mumtaz, H., and P. Surico (2008): "Evolving International Inflation Dynamics: World and Country Specific Factors," *CEPR Discussion Paper*, 6767.

Nason, J. M., and T. Cogley (1994): "Testing the Implications of Long-Run Neutrality for Monetary Business Cycle Models," *Journal of Applied Econometrics*, 9, S37–70.

Ng, S., E. Moench, and S. M. Potter (2008): "Dynamic Hierarchical Factor Models," *Manuscript, Columbia University and FRB New York*.

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), 997–1014.

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 & Statistics*, 21(4), 547–563.

Peersman, G. (2005): "What Caused the Millenium Slowdown? Evidence Based on Vector Autoregressions," *Journal of Applied Econometrics*, 20(2), 185–207.

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. (1991): "Optimal Inference in Cointegrated Systems," *Econometrica*, 59(2), 283–306.

——— (1996): "Econometric Model Determination," *Econometrica*, 64(4), 763–812.

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

Poirier, D. (1998): "Revising Beliefs in Nonidentified Models," *Econometric Theory*, 14(4), 483–509.

Primiceri, G. E. (2005): "Time Varying VARs and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852.

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

Ríos-Rull, J.-V., F. Schorfheide, C. Fuentes-Albero, M. Kryshko, and R. Santaeulalia-Llopis (2009): "Methods versus Substance: Measuring the Effects of Technology Shocks," *NBER Working Paper*, 15375.

Robertson, J. C., and E. W. Tallman (2001): "Improving Federal Funds Rate Forecasts in VAR Models Used for Policy Analysis," *Journal of Business & Economic Statistics*, 19(3), 324–330.

Rogerson, R. (1988): "Indivisible Labor Lotteries and Equilibrium," *Journal of Monetary Economics*, 21(1), 3–16.

Rubio-Ramírez, J. F., D. Waggoner, and T. Zha (2010): "Structural Vector Autoregressions: Theory of Identification and Algorithms for Inference," *Review of Economic Studies*, forthcoming.

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. (1989): "Two Models of Measurements and the Investment Accelerator," *Journal of Political Economy*, 97(2), 251–287.

——— (1999): *The Conquest of American Inflation*. Princeton University Press, Princeton.

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

——— (2008): "DSGE Model-Based Estimation of the New Keynesian Phillips Curve," *FRB Richmond Economic Quarterly*, Fall Issue, 397–433.

Schotman, P. C., and H. K. van Dijk (1991): "On Bayesian Routes to Unit Roots," *Journal of Applied Econometrics*, 6(4), 387–401.

Schwarz, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6(2), 461–464.

Sims, C. A. (1972): "The Role of Approximate Prior Restrictions in Distributed Lag Estimation," *Journal of the American Statistical Association*, 67(337), 169–175.

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

——— (1993): "A 9 Variable Probabilistic Macroeconomic Forecasting Model," in *Business Cycles, Indicators, and Forecasting*, ed. by J. H. Stock, and M. W. Watson, vol. 28 of *NBER Studies in Business Cycles*, pp. 179–214. University of Chicago Press, Chicago.

——— (2002a): "Comment on Cogley and Sargent's 'Evolving post World War II U.S. Inflation Dynamics' ," in *NBER Macroeconomics Annual 2001*, ed. by B. S. Bernanke, and K. Rogoff, vol. 16, pp. 373–379. MIT Press, Cambridge.

——— (2002b): "Solving Linear Rational Expectations Models," *Computational Economics*, 20(1-2), 1–20.

——— (2003): "Probability Models for Monetary Policy Decisions," *Manuscript, Princeton University*.

Sims, C. A., and H. Uhlig (1991): "Understanding Unit Rooters: A Helicopter Tour," *Econometrica*, 59(6), 1591–1599.

Sims, C. A., D. Waggoner, and T. Zha (2008): "Methods for Inference in Large Multiple-Equation Markov-Switching Models," *Journal of Econometrics*, 146(2), 255–274.

Sims, C. A., 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.

——— (2006): "Were There Regime Switches in U.S. Monetary Policy?," *American Economic 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.

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

Stock, J. H., and M. W. Watson (1989): "New Indices of Coincident and Leading Economic Indicators," in *NBER Macroeconomics Annual 1989*, ed. by O. J. Blanchard, and S. Fischer, vol. 4, pp. 351–394. MIT Press, Cambridge.

——— (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., and B. Inder (2004): "Bayesian Analysis of the Error Correction Model," *Journal of Econometrics*, 123(2), 307–325.

Strachan, R., and H. K. van Dijk (2006): "Model Uncertainty and Bayesian Model Averaging in Vector Autoregressive Processes," *Manuscript, Tinbergen Institute*, 06/5.

Theil, H., and A. S. Goldberger (1961): "On Pure and Mixed Estimation in Economics," *International Economic Review*, 2(3), 65–78.

Uhlig, H. (1997): "Bayesian Vector Autoregressions with Stochastic Volatility," *Econometrica*, 65(1), 59–73.

——— (2005): "What Are the Effects of Monetary Policy on Output? Results From an Agnostic Identification Procedure," *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(2), 326–357.

——— (2009): "Steady State Priors for Vector Autoregressions," *Journal of Applied Econometrics*, 24(4), 630–650.

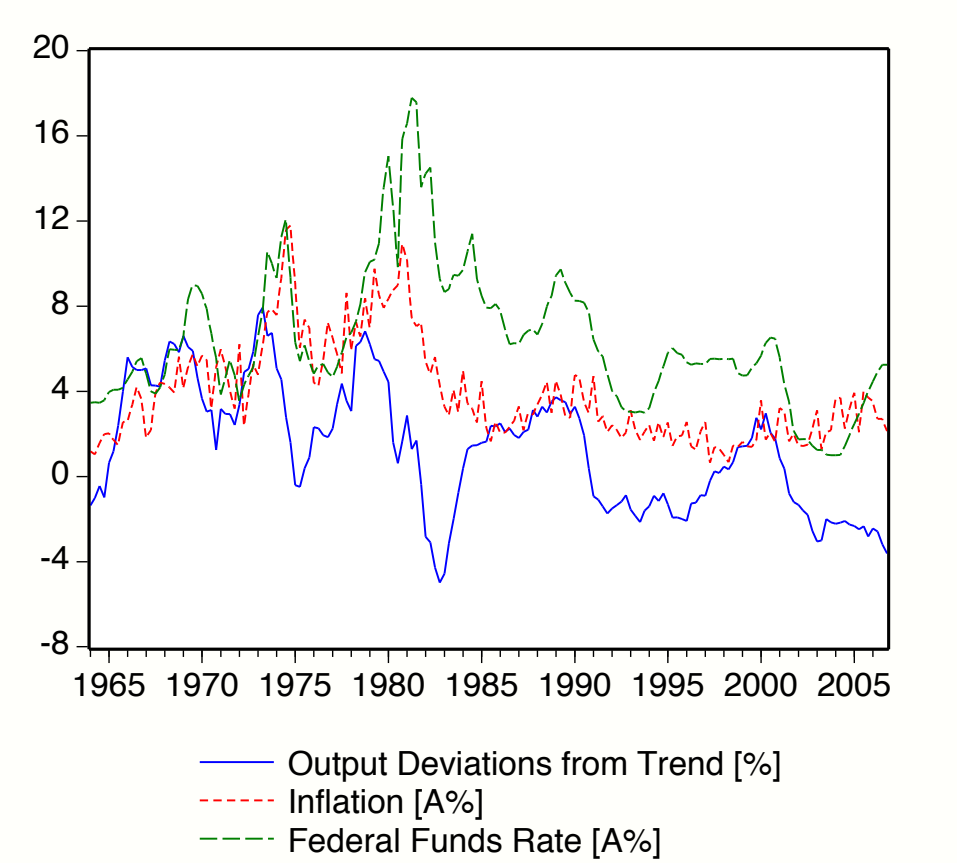
Waggoner, D., and T. Zha (1999): "Conditional Forecasts In Dynamic Multivariate Models," *Review of Economics and Statistics*, 81(4), 639–651.

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

Wright, J. (2008): "Bayesian Model Averaging and Exchange Rate Forecasting," *Journal of Econometrics*, 146, 329–341.

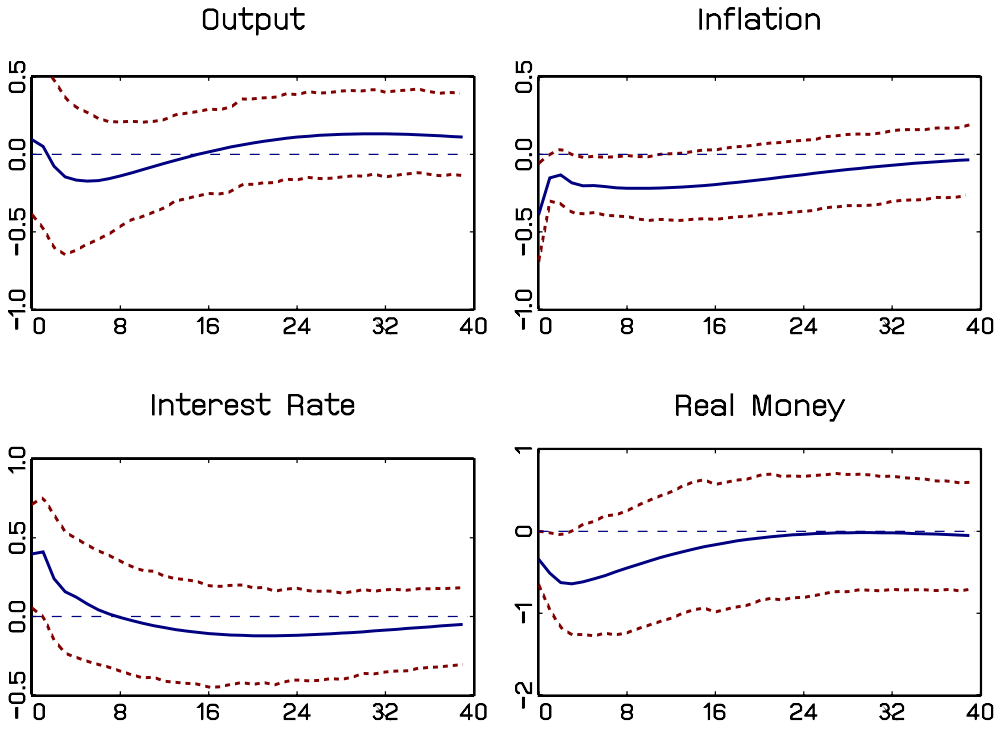
Zellner, A. (1971): *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons, Hoboken.

Figure 1: Output, Inflation, and Interest Rates



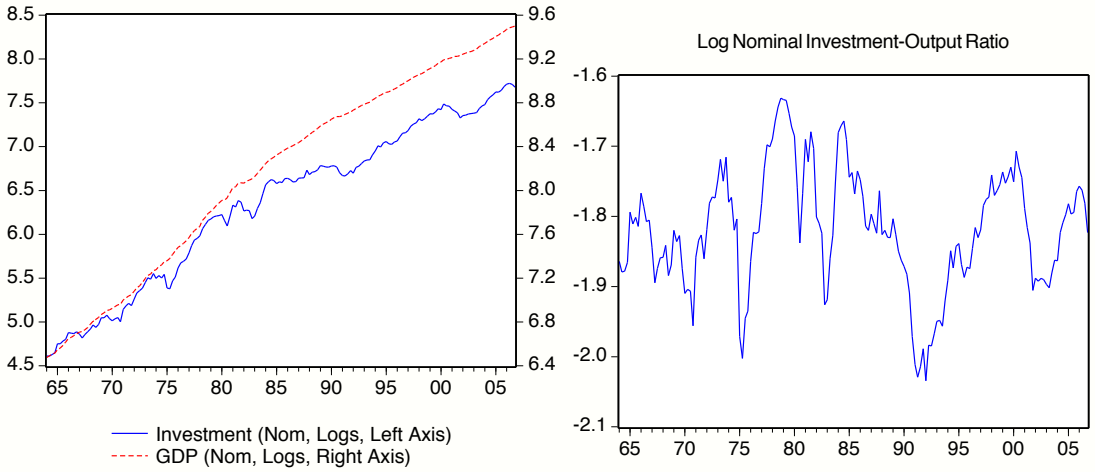
Notes: The figure depicts U.S. data from 1964:Q1 to 2006:Q4. Output is depicted in percentage deviations from a linear deterministic trend. Inflation and interest rates are annualized ($A\%$).

Figure 2: Response to a Monetary Policy Shock



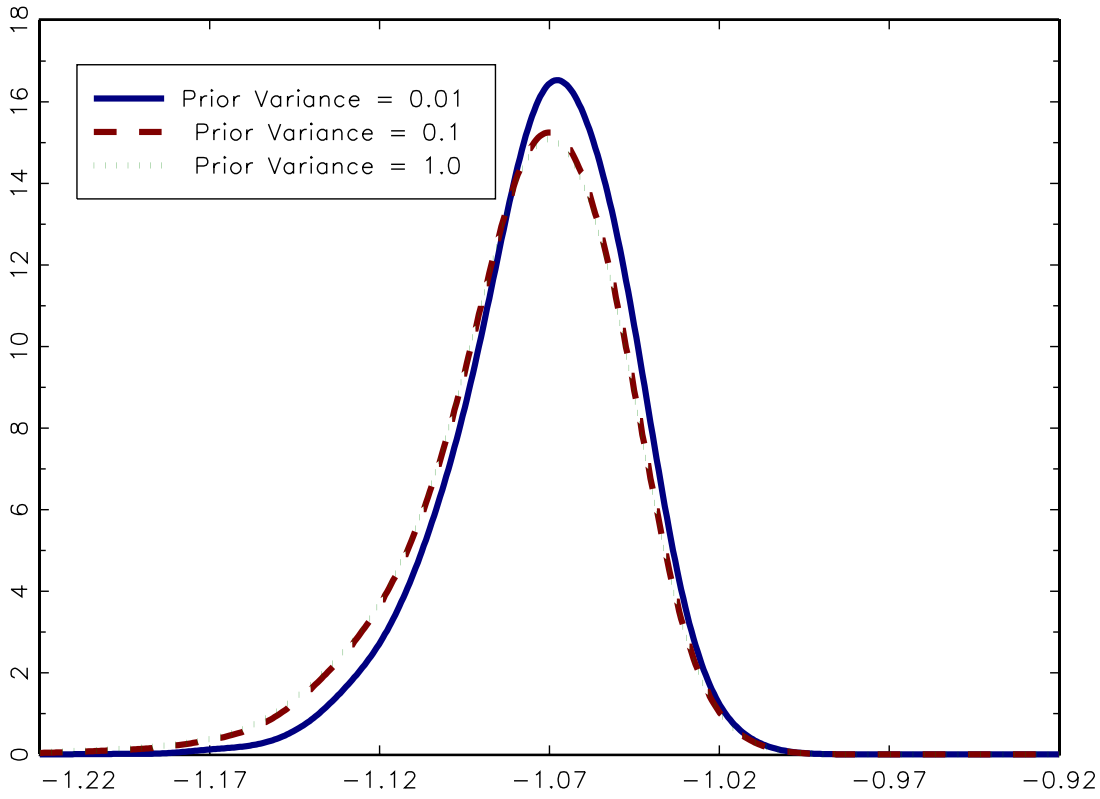
Notes: The figure depicts 90% credible bands and posterior mean responses for a VAR(4) to a one-standard deviation monetary policy shock.

Figure 3: Nominal Output and Investment



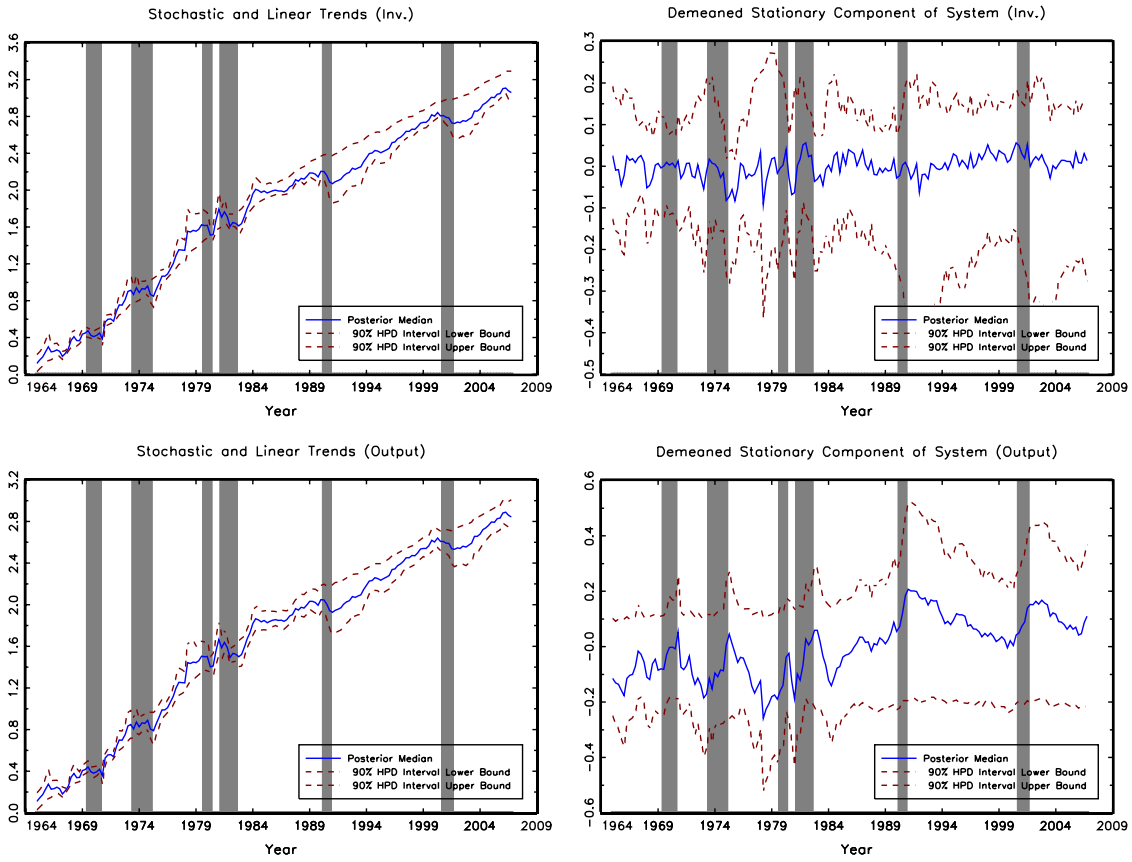
Notes: The figure depicts U.S. data from 1964:Q1 to 2006:Q4.

Figure 4: Posterior Density of Cointegration Parameter



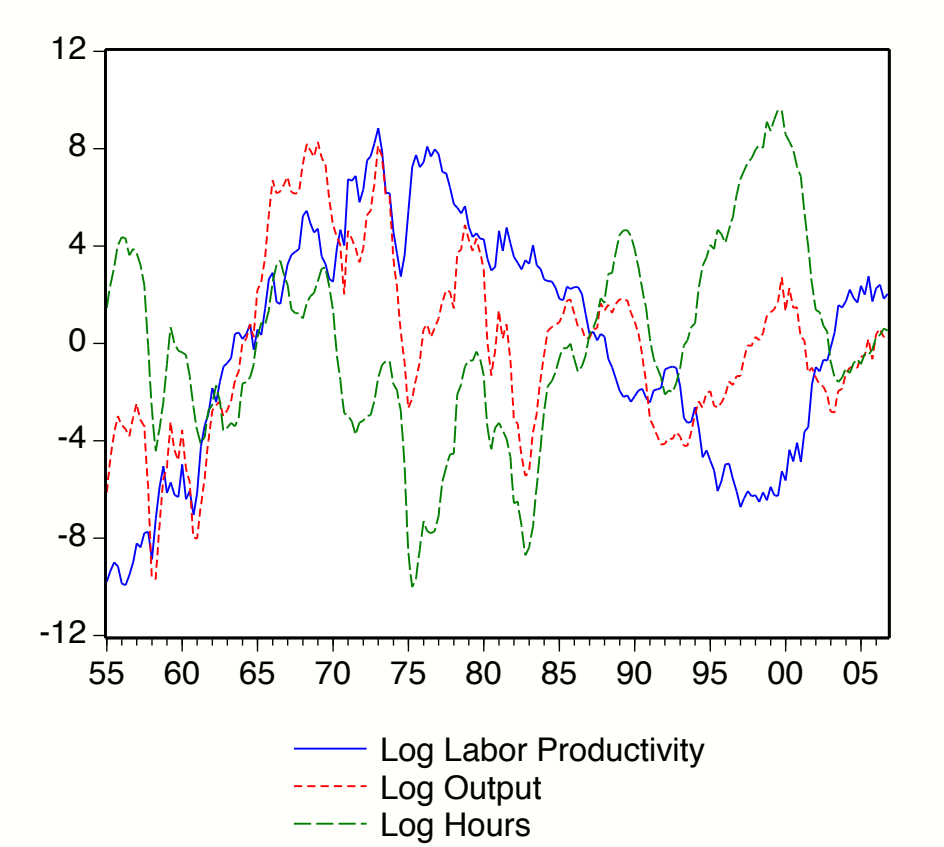
Notes: The figure depicts Kernel density approximations of the posterior density for B in $\beta = [1, B]'$ based on three different priors: $B \sim N(-1, 0.01)$, $B \sim N(-1, 0.1)$, and $B \sim N(-1, 1)$.

Figure 5: Trends and Fluctuations



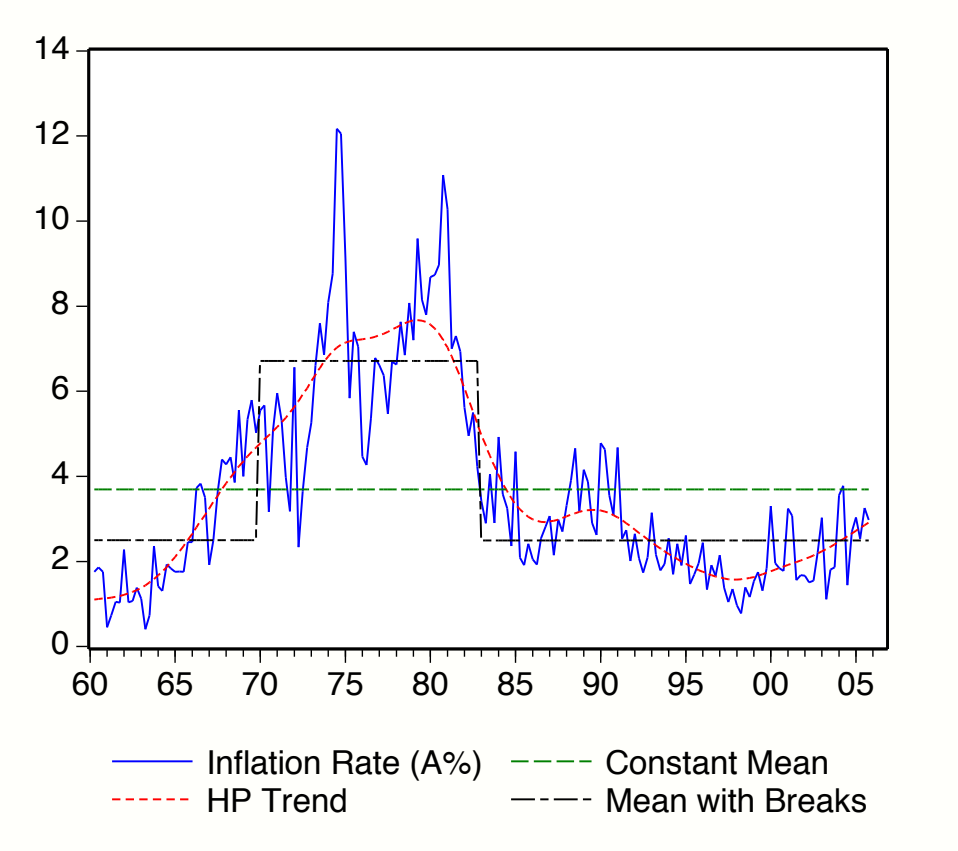
Notes: The figure depicts posterior medians and 90% credible intervals for the common trends in log investment and output as well as deviations around these trends. The gray shaded bands indicate NBER recessions.

Figure 6: 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.

Figure 7: 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 (A%). The sample ranges from 1960:Q1 to 2005:Q4.