

Macroeconometrics - A Discussion

Frank Schorfheide*

University of Pennsylvania
CEPR, NBER, and PIER

March 9, 2016

*Correspondence: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu. Paul Sangrey provided excellent research assistance. A preliminary version of this discussion was presented at the 2015 World Congress of the Econometric Society. Financial support from the National Science Foundation under grant SES 1061725 is gratefully acknowledged.

1 Introduction

The prediction of macroeconomic time series and the effects of monetary and fiscal policy interventions is an exciting and perhaps sometimes mysterious task, associated in equal parts with images of the ancient Oracle at Delphi and folks hunched in front of computer screens and crunching numbers. The popularity of these prediction falls and rises with their perceived accuracy and the 2007-09 recession has certainly generated some disappointments, at least in the eye of the public. Unfortunately, the public underappreciates the fact that economic forecasts tend to be associated with measures of uncertainty. We are at a point at which smartphone weather apps assign probabilities to whether it will rain or not over the next 10 days, while at the same point GDP, unemployment, and inflation forecasts when discussed in major news outlets are always reported as point forecasts, without any probabilities associated with them. Naturally, this opens the door for disappointments.

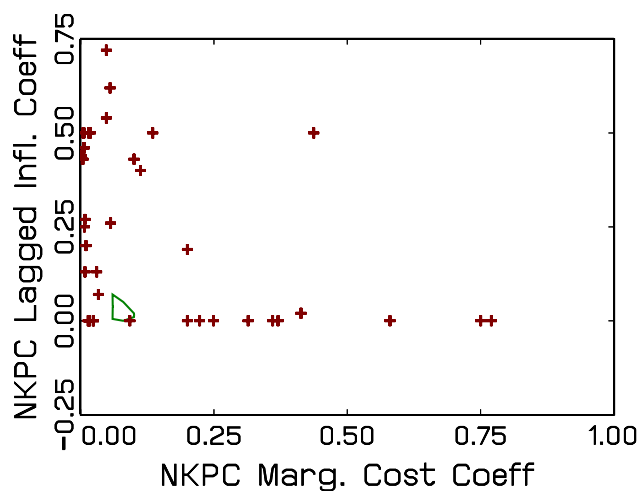
In my view, one of the key tasks and challenges for macroeconometrics is to produce accurate characterizations of uncertainty associated with model parameter estimates, policy effects, and future (or counterfactual) events and developments. The chapters contributed to this volume by Ulrich Müller and Mark Watson, on the one hand, and Harald Uhlig on the other hand, take on the challenge in the context of two different, but equally important settings. Before providing some remarks on the Müller-Watson and Uhlig chapters, let me highlight the difficulty of providing accurate measures of uncertainty by looking back at the 2010 World Congress. The following illustration is taken from Schorfheide (2013).

Figure 1 depicts estimates of the slope κ and the weight γ_b on lagged inflation of the following New Keynesian Phillips curve (NKPC)

$$\tilde{\pi}_t = \gamma_b \tilde{\pi}_{t-1} + \gamma_f \mathbb{E}_t[\tilde{\pi}_{t+1}] + \kappa \widetilde{MC}_t. \quad (1)$$

Here $\tilde{\pi}_t$ is inflation and \widetilde{MC}_t is marginal costs, both in deviations from a long-run mean or trend process. The slope of the NKPC crucially affects the central bank's output-inflation trade-off. Each dot in the figure corresponds to a point estimate of (κ, γ_b) reported in the literature, obtained by estimating a dynamic stochastic general equilibrium (DSGE) model. Model specification, included observations, and sample periods differ across studies. The green circle is a credible interval associated with one of the estimates. The message of the figure is that somehow, the measure of uncertainty is too small. It does not foreshadow the

Figure 1: Estimates of NKPC Parameters



Source: Schorfheide (2013)

variation in point estimates that is obtained by varying details of the model specification and the specifics of the data set.

The common theme of the Müller-Watson and Uhlig chapters is to produce measures of uncertainty that are appropriately sized. The Müller-Watson chapter is about making inference about what happens in the long-run by filtering out short-run fluctuations and noise from the data and focusing on the relevant low-frequency information in the data. It formalizes the notion that if you have 50 years of data and are interested in predicting what happens over a 10 year horizon, then you really just have five non-overlapping observations, which invariably should lead to sizable coverage intervals.

The chapter by Harald Uhlig focuses on uncertainty about the propagation of structural shocks in the context of a vector autoregression (VAR). Here shocks could be exogenous shifts to demand or supply or changes in economic policies. For the sake of concreteness my discussion will use monetary policy shocks as the running example, that is, unanticipated (from the perspective of the public) deviations from some perceived monetary policy rule that sets the nominal interest rate based on the current state of the economy. The difficulty is that one-step-ahead forecast errors of the policy instrument usually do not identify the unanticipated part of the policy change, because some of the forecast errors can be explained

by the systematic reaction to other unanticipated shocks that hit the economy in the current period. The VAR literature of the 1980s and 1990s typically made very strong identifying assumptions about the mapping between forecast errors and shocks. In turn, different assumptions led to different conclusions. With the emergence of the use of sign restrictions in the 2000s, researchers started to make more conservative statements about the propagation of shocks that nicely summarize and encompass results obtained from more dogmatic identification strategies.

2 Low Frequency Econometrics

The Müller-Watson chapter develops a general technique of extracting and processing low frequency information from economic time series. This information can then be used for many different purposes, including inference about the persistent components of time series models, to create heteroskedasticity and autocorrelation robust standard errors, and to generate long-horizon forecasts. It can be applied to univariate as well as multivariate time series and the authors discuss its relationship to spectral analysis in detail. Henceforth, I will refer to this technique as the MW approach. While the chapter outlines a broad research agenda, my remarks will more narrowly focus on an application of their approach to the problem of long-horizon forecasting. The question is the following: suppose the goal is to generate a forecast of average consumption growth over the next five to ten years, should we (i) attempt to model both short-run and long-run dynamics or should we (ii) just write down a model of long-run dynamics?

Approach (i) has the potential advantage that we can exploit possible “cross-equation restrictions.” We can use high-frequency information to estimate the “common” parameters (think of an AR(p) model) and extrapolate from high-frequency behavior to low-frequency behavior, thereby sharpening long-run predictions. Approach (ii) is appealing in situations in which the econometrician has reason to believe that the cross-coefficient restrictions between short-run and long-run dynamics are potentially misspecified.

In the remainder of this section I will compare consumption growth forecasts using the MW approach to forecasts from a parametric local-level model that captures both short-run and

long-run dynamics. To explore the MW approach, let us consider the following specification:

$$\begin{aligned} c_t &= \mu + \frac{g}{T}x_t + \sigma u_t \\ x_t &= x_{t-1} + \sigma\eta_t. \end{aligned} \tag{2}$$

Here c_t is consumption growth and x_t is a local level process, which plays an important role in the asset pricing literature; see Bansal and Yaron (2004).¹ Because x_t is a unit-root process, its variance is growing at rate T . Due to the sample-size dependent loading g/T in the observation equation for consumption, the contribution of the local level process x_t to the variance of consumption growth shrinks at rate $1/T$, which makes it difficult to detect. As we will see below, the sequence of drifting coefficients has been carefully chosen to obtain well-defined limits.

In Schorfheide, Song, and Yaron (2014) we show that specification (2) combined with the assumption that the sequence $\{u_t\}_{t=1}^T$ is serially uncorrelated is unable to capture the negative first-order autocorrelation of monthly consumption growth. Thus, while (2) maybe a good model of long-run consumption growth, it is a poor model of short-run consumption dynamics. A better specification is one that includes MA(1) measurement errors:²

$$\begin{aligned} c_t &= \mu + g/Tx_t + \sigma u_t + \sigma_\epsilon(\epsilon_t - \epsilon_{t-1}) \\ x_t &= x_{t-1} + \sigma\eta_t. \end{aligned} \tag{3}$$

The subsequent estimation of (3) is not based on an asymptotic argument. Thus, the model could be reparameterized in terms of $\varphi = g/T$.

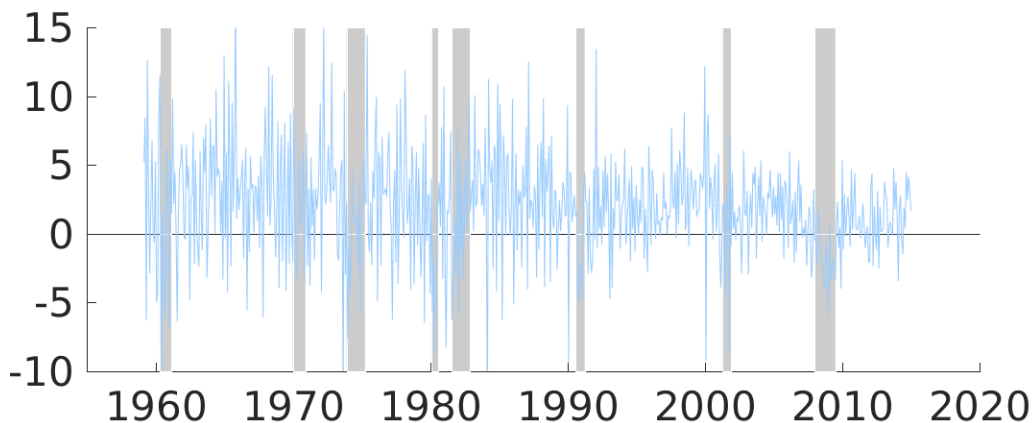
I will subsequently compare long-run forecasts from (2) based on the MW approach to forecasts obtained from Bayesian estimation of (3). Formally, we will focus on the prediction of average consumption growth over H periods:

$$\bar{c}_{T:T+H} = \frac{1}{H} \sum_{h=1}^H c_{T+h}. \tag{4}$$

¹In the long-run risks literature the x_t process is assumed to be stationary but highly persistent, e.g., $\rho_x = 0.99$.

²Schorfheide, Song, and Yaron (2014) provide some justification for the measurement error process based on the construction of monthly consumption by the Bureau of Economic Analysis. Their preferred specification also has the feature that monthly measurement errors average out over annual frequency and it includes stochastic volatility. This discussion focuses on the simpler version in (3).

Figure 2: Monthly U.S. Consumption Growth (Annualized Percent) from 1959-2014



Let me re-iterate that MW emphasize the following potential disadvantages of estimating the parametric model (3): careful modeling of the measurement errors is required; misspecified high-frequency dynamics can contaminate inference about the low-frequency component; a tight parametric specification of the high-frequency dynamics might understate uncertainty about low-frequency implications of the model.

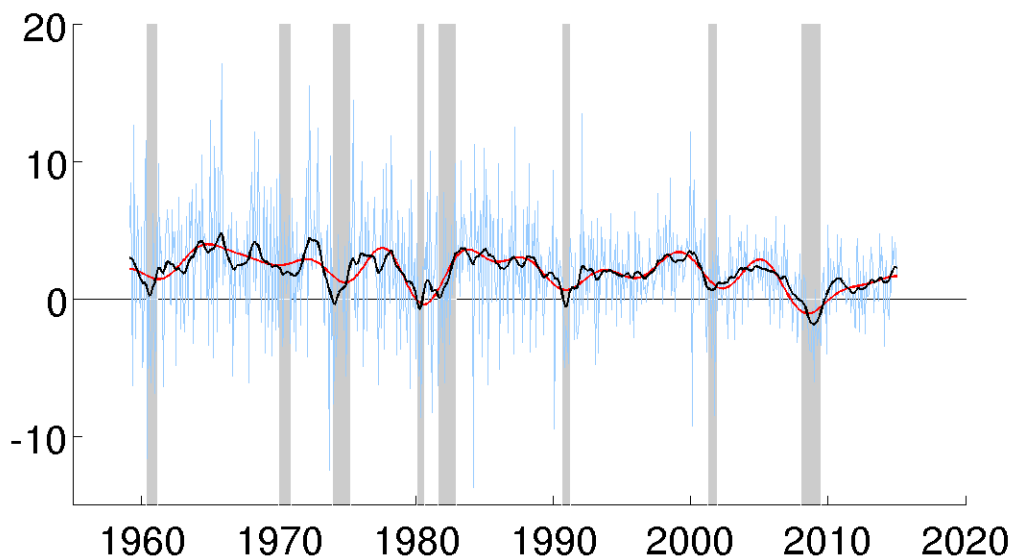
2.1 Data and Low-Frequency Component

Figure 2 plots per capita real consumption expenditure on nondurables and services from the NIPA tables available from the Bureau of Economic Analysis. From a visual inspection of the plot, the local level component x_t is very difficult to detect, because consumption growth data is very noisy.

The first step of the MW approach is to project data $\{c_t\}$ onto cosine functions $\cos(\pi jt/T)$, $j = 1, \dots, q$. Here q is a constant that determines the number of cosine terms considered in the analysis. The standardized regression coefficients are given by

$$C_j = T^{-1/2} \sum_{t=1}^T \sqrt{2} \cos(\pi jt/T) c_t, \quad j = 1, \dots, q. \quad (5)$$

Figure 3: Low Frequency Component of Consumption Growth: Projection onto Cosines versus Smoothed x_t



Notes: The red line depicts \hat{c}_t defined in (7) obtained from the MW approach for $q = 24$. The black line depicts posterior estimates $\mathbb{E}[x_t|c_{1:T}]$ obtained from the Bayesian estimation of the parametric model in (3).

In addition, the sample average is

$$C_0 = T^{-1/2} \sum_{t=1}^T (c_t - \mu). \quad (6)$$

The fitted values \hat{c}_t , defined as

$$\hat{c}_t = C_0 + \sum_{j=1}^q C_j \cos(\pi j t / T), \quad (7)$$

can be interpreted as an estimate (in the time domain) of the low frequency component of consumption growth. They are plotted in Figure 3 together with the raw consumption growth data.

For the parametric model (3) I use a Kalman smoother to extract the hidden local level process x_t . More precisely, I evaluate the likelihood function using the Kalman filter and use a standard random-walk Metropolis-Hastings algorithm (see Herbst and Schorfheide (2015) for a description in the context of DSGE model estimation) to generate parameter draws from the posterior distribution. For each draw, I run the Kalman smoother to compute

$\mathbb{E}[x_t|c_{1:T}, \theta]$, where $\theta = [\mu, \varphi, \sigma, \sigma_\epsilon]'$ and $c_{1:T} = \{c_1, \dots, c_T\}$. Averaging over the θ draws yields an approximation of $\mathbb{E}[x_t|c_{1:T}]$ which is also plotted in Figure 3. Notice that the two extracted low frequency components in Figure 3 look quite similar. I achieved this by experimenting with the choice of q , holding our sample size T fixed. For values of q less (greater) than 24, one obtains a \hat{c}_t that is smoother (more volatile) than $\mathbb{E}[x_t|c_{1:T}]$.

2.2 Inference under the MW Approach

The MW approach replaces the sample of T original observations c_1, \dots, c_T by a sample of $q + 1$ regression coefficients C_0, C_1, \dots, C_q . The distribution of the raw data according to (7) determines the distribution of the regression coefficients. However, due to the averaging in (5) the specification of the short-run dynamics in (2) is not important as $T \rightarrow \infty$. A functional central limit theorem leads to the following convergence result:

$$\begin{aligned} C_j &\implies \sigma \int_0^1 \Psi_j(s) dW_u(s) + \sigma g \int_0^1 \Psi_j(s) W_\eta(s) ds \\ C_0 &\implies \sigma \int_0^1 dW_u(s) + \sigma g \int_0^1 W_\eta(s) ds, \end{aligned} \quad (8)$$

where $\Psi_j(s) = \sqrt{2} \cos(\pi j s)$. One can also approximate the distribution of the long-horizon forecast in (4) by expressing the maximum forecast horizon H as a function of the sample size: $H = \lambda T$.

$$\bar{C} = \frac{1}{\lambda} \sqrt{T} (\bar{c}_{T+1:T+\lfloor \lambda T \rfloor} - \mu) \implies \frac{\sigma}{\lambda} W_u^+(\lambda) + g\sigma W_\eta(1) + \frac{g\sigma}{\lambda} \int_0^\lambda W_\eta^+(s) ds. \quad (9)$$

These calculations imply

$$(C_0, \dots, C_q, \bar{C}) | (\mu, g, \sigma) \implies N(0, \sigma^2 \Sigma(g)). \quad (10)$$

The derivation of the covariance matrix $\Sigma(g)$ is a bit tedious and it is easy to make mistakes in calculating the entries. But once that has been done, the original time series $c_{1:T}$ now has been transformed into realizations of $q + 2$ Gaussian random variables that can be used for inference.

MW have developed sophisticated inference procedures for the parameters based on the approximate small distribution of the $q + 1$ random variables C_0, \dots, C_q . I decided to simply use quasi-Bayesian inference as follows based on the approximate Gaussian likelihood in

(10). Thus, I interpret the right-hand side of (10) as $p(C_0, \dots, C_q | \mu, g, \sigma)$ and specify a prior $p(\mu, g, \sigma)$. According to Bayes Theorem

$$p(\mu, g, \sigma | C_0, \dots, C_q) \propto p(C_0, \dots, C_q | \mu, g, \sigma) p(\mu, g, \sigma), \quad (11)$$

where \propto denotes proportionality. I use the same random walk Metropolis-Hastings algorithm that is used for inference in the parametric local level model (3) to generate draws from the posterior of (μ, g, σ) . Based on the posterior parameter draws, it is straightforward to obtain draws from the posterior predictive distribution by using a Monte Carlo approximation of

$$p(\bar{C} | C_0, \dots, C_q) = \int p(\bar{C} | C_0, \dots, C_q, \mu, g, \sigma) p(\mu, g, \sigma | C_0, \dots, C_q) d(\mu, g, \sigma). \quad (12)$$

2.3 Empirical Results

Table 1 reports posterior means and credible intervals obtained under the MW approach in (11) and from the Bayesian estimation of the parametric state-space model (3). For this illustration I mostly used flat priors so that the posterior mimics the shape of the likelihood function. The priors for μ , σ , and σ are improper, whereas the prior for g is restricted to the bounded interval $[0, 3\sqrt{T}]$. Note that $(g/T)^2 t$ can be interpreted as the signal-to-noise ratio for the local level process x_t . Our prior places an upper bound of 3 on the end-of-sample signal-to-noise ratio.

A few observations stand out. First, the parameter g is very imprecisely estimated. The 90% credible intervals have a width of about 35. Moreover, the estimates differ substantially across the MW approach and the parametric model. For the former the posterior mean is about 10, whereas for the latter the posterior mean is 63. Second, the point estimates of the mean μ are very similar across the two estimation procedures; but the parametric model leads to tighter credible intervals. This is not surprising because it also utilizes information from the entire spectral band. Finally, the estimate of σ is larger under the MW approach, presumably because under the parametric model part of the short-run fluctuations are explained by measurement errors and a larger fraction of the variation in consumption growth is attributed to the local level process.

The resulting forecasts of long-run consumption growth are plotted in Figure 4. Under the MW approach the forecasts are centered around the estimate of μ , whereas the forecasts from the parametric model are centered at $\mu + \mathbb{E}[x_T | c_{1:T}, \theta]$. The MW approach generates

Table 1: Parameter Estimates

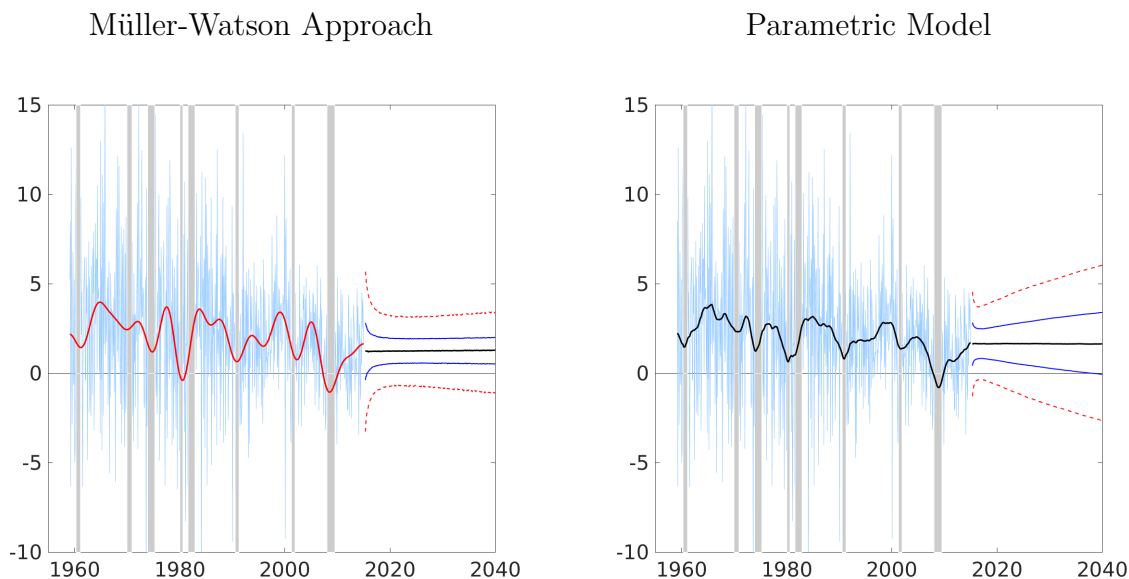
Prior		Posterior	
		Median	90% Interval
Müller-Watson Approach			
μ	$\propto 1$	2.27	(0.42, 4.13)
σ	$\propto \mathbb{I}\{\sigma > 0\}$	5.46	(2.32, 8.63)
g	$\propto \mathbb{I}\{0 \leq g \leq 3\sqrt{T}\}$	10.0	(0.95, 37.7)
Parametric Model			
μ	$\propto 1$	2.23	(0.67, 3.51)
σ	$\propto 1\mathbb{I}\{\sigma > 0\}$	2.71	(2.44, 3.03)
g	$\propto \mathbb{I}\{0 \leq g \leq 3\sqrt{T}\}$	63.2	(40.8, 75.7)
σ_ϵ	$\propto \mathbb{I}\{\sigma_\epsilon > 0\}$	1.94	(1.68, 2.22)

Notes: Sample Size: $T = 671$; $3\sqrt{T} \approx 78$; number of transforms: $q = 24$; number of posterior draws $N = 20,000$. The prior on g is set so the trend explains at most 90% of the variance of consumption growth. $\mathbb{I}\{x > 0\}$ is indicator function that is one if $x > 0$ and zero otherwise.

a lot of uncertainty about short-run forecasts. Mechanically, the predictive intervals diverge as one lets $\lambda \rightarrow 0$ (in the figure, the shortest horizon is $H = 3$). This turns out to be an artifact of the asymptotics which were derived by letting $T \rightarrow \infty$ for fixed λ ; rather than setting $\lambda = 1/T$ before taking the $T \rightarrow \infty$ limit. However, given that the MW approach explicitly removes information about short-run dynamics from the sample by transforming c_1, \dots, c_T into C_1, \dots, C_q , it is fairly intuitive that the intervals for short-horizon predictions are wide.

Under the parametric approach the prediction interval also widens as $H \rightarrow 1$, but it stays bounded. Intuitively, in the short-run, the uncertainty is dominated by the realizations of u_t and ϵ_t . As the forecast horizon increases, these shocks start to average out. In the long-run, the uncertainty is dominated by the unit root process x_t . In my illustration, the parametric model generates more uncertainty about the long-run because of the larger estimate of g , which controls the weight of the local-level process x_t .

Figure 4: Forecasts of Average Consumption Growth



Notes: The figure depicts posterior mean forecasts (solid) as well as 90% prediction intervals (dashed) and 60% prediction intervals (solid). Right panel: the solid line to the left of the forecast origin is $\mathbb{E}[x_t|c_{1:T}]$. Left panel: the solid line to the left of the forecast origin is \hat{c}_t .

2.4 Score Card

The MW approach formalizes the notion that if you have fifty years of data and are interested in making statements about what happens over a decade, you really only have five non-overlapping observations. It does so, by developing a very elegant econometric theory that relies on projecting the original data onto cosine functions of different frequency. In fact, the asymptotics are set up such that as the sample size increases, the frequency band covered by these cosine functions shrinks to zero, so that the number of transformed observations q stays constant. Thus, the resulting inference problem always is a small-sample inference problem, albeit based on approximately normally distributed random variables.

In my view the MW approach is appealing if the goal of the empirical analysis is to ask questions that pertain only to low frequency properties of the data, such as long-horizon forecasting or the estimation of a long-run variance. The implementation of the approach requires the user to select a spectral band, which is defined by q . Unfortunately, there is little guidance on how to do this. In the empirical application, I simply picked q such that

the fitted values from the cosine projection looked like the smoothed values of the local-level process obtained from the estimation of the parametric model. Of course, in practice this is undesirable. An algorithm on how to choose q in view of the question that is being asked and in view of the salient features of the data would be very helpful, in particular for applications in which the substantive conclusion is really sensitive to q .

For a wider adoption, I think it is important to make the procedure as user-friendly as possible. While the formulas look very elegant, the derivation of the likelihood function, that is the elements of $\Sigma(g)$ in (10), is quite tedious because of the various standardizations and coding up the likelihood function can be prone to errors. I am sure that practitioners will appreciate explicit formulas for a broad set of canonical models along with some code that can generate the likelihood functions. I also think that it is important to separate the basic idea of data transformation from the problem of conducting inference based in non-standard small sample settings. Much of the chapter as well as other papers on this research agenda written by the authors focus on the inference problem. While this is certainly important and interesting, it should not turn into an impediment for using the data transformation. All the computations presented in my discussion were based on a fairly basic Metropolis-Hastings algorithm.

3 Structural VARs and Identification

Structural analysis with VARs requires identification assumptions. The chapter by Harald Uhlig provides a critical review of the sign-restriction literature that he pioneered in Uhlig (2005). His key principles are: (i) If you know it, impose it! (ii) If you do not know it, do not impose it! As stated, it is difficult to disagree with these principles. However, in practice, the devil is in the details of the empirical application; in part, because there is a grey area in which there is some uncertainty associated with what we know.

3.1 The Basic Setup

When I teach structural VARs to graduate students, I tend to introduce the identification problem as follows. A structural VAR expresses the vector of one-step-ahead forecast errors

u_t as a function of a vector of structural shock innovations ϵ_t :

$$y_t = \Phi y_{t-1} + u_t, \quad u_t = \Phi_\epsilon \epsilon_t. \quad (13)$$

One can identify Φ and the covariance matrix Σ of u_t from the data. The ϵ_t 's are assumed to be orthogonal to each other and have unit variance. This leads to the restriction

$$\Phi_\epsilon \Phi_\epsilon' = \Sigma. \quad (14)$$

Because Σ is a symmetric matrix, this system of equations leaves Φ_ϵ undetermined. One way of separating the identifiable components of Φ_ϵ from the unidentifiable components is to define Σ_{tr} as the lower triangular Cholesky factor of Σ and to parameterize Φ_ϵ as

$$\Phi_\epsilon = \Sigma_{tr} \Omega, \quad (15)$$

where Ω is an orthogonal matrix.

Because $\Omega \Omega' = I$, it is straightforward to show that Ω does not appear in the Gaussian likelihood function. Noting that up to some trivial normalizations the matrices Σ and Σ_{tr} contain the same information we can write the joint distribution for the data Y and the VAR parameters (Φ, Σ, Ω) as

$$p(Y, \Phi, \Sigma, \Omega) = p(Y|\Phi, \Sigma)p(\Phi, \Sigma)p(\Omega|\Phi, \Sigma), \quad (16)$$

$p(Y|\Phi, \Sigma)$ is the likelihood function, $p(\Phi, \Sigma)$ is the prior for the reduced-form parameters of the VAR, and $p(\Omega|\Phi, \Sigma)$ is the prior for the non-identifiable parameters of the structural VAR model. It can be verified that beliefs about Ω do not get updated, that is, the posterior of Ω conditional on (Φ, Σ) equals the prior

$$p(\Omega|Y, \Phi, \Sigma) = p(\Omega|\Phi, \Sigma). \quad (17)$$

Using this notation, the debates about VAR identification in the empirical macroeconomics literature can be reduced to debates about $p(\Omega|\Phi, \Sigma)$. The sign restriction literature replaced the dogmatic prior distributions of the 1980s and 1990s (zero restrictions and long-run restrictions that can be represented by point mass priors), with more “agnostic” and “less dogmatic” distributions. Thus, the implementation of the above-mentioned principles amounts to the choice of a prior distribution.

Let us represent impulse responses as functions $\theta(\Phi, \Sigma, \Omega)$. Here θ may either be skalar or a vector. In empirical work researchers typically report pointwise coverage sets plotted as “error bands.” Once one subscribes to the notion that we “know” that impulse responses to, say, a contractionary monetary policy shock have to satisfy certain sign restrictions, e.g., interest rates have to rise and monetary aggregates and prices have to fall, then the support of the prior distribution $p(\Omega|\Phi, \Sigma)$ is restricted. In the remainder of this section I focus on the case in which the goal is to identify a single shock so that we can replace Ω by its first column, which I denote by q . In turn, we can replace Ω by q in the densities that appear in (16) and (17). In the absence of sign restrictions q is located on the n -dimensional hypersphere \mathbb{Q} , where n is the number of observables stacked in the vector y_t . The sign restrictions restrict the support of $p(q|\Phi, \Sigma)$ to a subspace $\mathbb{Q}^s(\Phi, \Sigma)$ of \mathbb{Q} .³ The literature on set-identified models has called $\mathbb{Q}^s(\Phi, \Sigma)$ the identified set. Note that its location depends on the reduced-form parameters (Φ, Σ) .

3.2 A Stylized Representation of the Inference Problem

The inference problem can be illustrated through the following simple example. Let $\phi = [\phi_1, \phi_2]'$ be an identifiable reduced form parameter of dimension 2×1 . Here ϕ is the analog of (Φ, Σ) in the VAR. Moreover, let θ be the structural parameter of interest, e.g., an impulse response to a monetary policy shock, in the context of the VAR. Suppose that the unit-length vector $q = [q_1, q_2] \in \mathbb{Q}$ is constrained by the following inequalities

$$q_1 \geq 0 \quad \text{and} \quad q_2 \geq \frac{\phi_1}{\phi_2} q_1, \quad (18)$$

where $\phi_1, \phi_2 > 0$. In this case the identified set is a segment of the unit circle given by

$$\mathbb{Q}^s(\phi) = \left\{ [q_1, q_2] \in \mathbb{Q} \mid 0 \leq q_1 \leq \sqrt{\phi_2^2 / (\phi_1^2 + \phi_2^2)} \right\}. \quad (19)$$

The parameterization of the problem in terms of (ϕ, q) – or, more generally (Φ, Σ, q) – is useful to understand what can be learned from the data and what cannot be learned. The paper by Uhlig (2005) – and the literature that builds on it – also uses this parameterization to specify a prior. In the context of the stylized example, the benchmark prior proposed

³It could be that for some (Φ, Σ) the support is empty, i.e., the reduced-form parameters are inconsistent with the sign restrictions. I abstract from this case to simplify the exposition.

by Uhlig (2005) for q is uniform on $\mathbb{Q}^s(\phi)$, where uniform means invariant under rotations. For $n = 2$ one can easily express q in polar coordinates $[\cos \varphi, \sin \varphi]'$. The benchmark prior assumes that φ is uniformly distributed over the interval that corresponds to the segment $\mathbb{Q}^s(\phi)$ of the unit circle.

While the benchmark prior is uniform on the identified set $\mathbb{Q}^s(\phi)$, it is not uniform on the identified set for the impulse responses. Suppose we consider $\theta = q_1$. Then the identified set for θ is given by the projection of $\mathbb{Q}^s(\phi)$ onto the q_1 ordinate:

$$\Theta(\phi) = \left\{ 0, \sqrt{\phi_2^2 / (\phi_1^2 + \phi_2^2)} \right\}. \quad (20)$$

It is well known that uniform prior distributions are generally not preserved under nonlinear transformations. In our example, a uniformly distributed angle φ , does not translate into a uniform distribution of $\theta = q_1 = \cos \varphi$. The implied prior for θ assigns more probability to sets near the upper bound of the identified set than to sets near the lower bound of the identified set.

3.3 Important Themes in the Literature

The Uhlig chapter presents a critical review of the sign restrictions literature. In the remainder of this discussion I will highlight a few themes in this research agenda that I think are important.

Good Reporting. Reasonable people might disagree on the specification of prior distributions, but everybody should strive to be transparent in their communication. The identified set $\Theta(\Phi, \Sigma)$ is a crucial object for inference in VARs identified with sign restrictions and it should be reported. The reduced-form parameter (Φ, Σ) is unknown and can be replaced by a posterior mean estimate, say $(\hat{\Phi}, \hat{\Sigma})$. We emphasized this in Moon, Schorfheide, Granziera, and Lee (2011): “Since in a Bayesian analysis the prior distribution of the impulse response functions conditional on the reduced form parameters does not get updated, it is useful to report the identified set conditional on some estimate, say, the posterior mean of Φ and Σ so that the audience can judge whether the conditional prior distribution is highly concentrated in a particular area of the identified set.” In addition one could plot the density $p(\theta | \hat{\Phi}, \hat{\Sigma})$ to communicate where the prior mass is located in the identified set conditional on the reduced-form parameter estimates.

Alternative Priors. While the (Φ, Σ, Ω) parameterization of the structural VAR is useful for separating directions in the parameter space in which the sample is informative from directions in which there is no information, it is not clear that it is useful for the elicitation of prior distributions. There is a long history of specifying prior distributions for the reduced-form parameters (Φ, Σ) based on statistical considerations (see, for instance, Doan, Litterman, and Sims (1984), Kadiyala and Karlsson (1997), Sims and Zha (1998)) or based on macroeconomic theory (see, for instance, Ingram and Whiteman (1994) and Del Negro and Schorfheide (2004)). Unfortunately, The elicitation of priors for Ω is more difficult and perhaps a bit unnatural. Del Negro and Schorfheide (2004) derive Ω matrices from DSGE models. A prior distribution for the DSGE model parameters then induces a prior distribution for Ω .

Alternatively, one could elicit prior distributions for (Φ, Φ_ϵ) or write the structural VAR as

$$A_0 y_t = A_1 y_{t-1} + \epsilon_t, \quad (21)$$

where $A_0 = \Phi_\epsilon^{-1}$ and $A_1 = \Phi_\epsilon^{-1} \Phi$. (21) looks more like a dynamic version of a traditional system-of-equations macroeconometric model. For instance, in a three-variable system the equations may correspond to an aggregate supply equation, an aggregate demand equation, and a monetary policy rule. The researcher can then try to specify priors for (A_0, A_1) and truncate this prior such that the desired sign restrictions are satisfied. This approach is pursued in Baumeister and Hamilton (2015) who at great length discuss the elicitation of priors for A_0 in the context of their empirical application. As long as the prior distribution for A_0 and A_1 is proper, the posterior distribution will be proper as well; but updating of the prior takes only place in certain directions of the parameter space.

Inference for the Identified Set. There is a large microeconometrics literature on inference in set-identified models. So far, I focused on inference for an impulse response θ . The microeconometrics literature also considered inference for the identified set $\Theta(\Phi, \Sigma)$. In Moon and Schorfheide (2009) we proposed a naive way of constructing credible sets for $\Theta(\Phi, \Sigma)$: compute a credible set for the identifiable reduced-form parameters (Φ, Σ) . Then take the union of identified sets $\Theta(\Phi, \Sigma)$ for all (Φ, Σ) in the previously-computed credible set. This approach avoids specifying a distribution on $\Theta(\Phi, \Sigma)$. More elaborate implementations of this idea along with a careful formal analysis are provided in Kline and Tamer (2016). The idea has not been applied in the structural VAR setting and personally I find

inference with respect to θ instead of $\Theta(\Phi, \Sigma)$ more compelling in VAR applications.

Multiple Priors and Posterior Bounds. Instead of considering a single prior $p(\Omega|\Phi, \Sigma)$ one could consider a family of prior distributions, say \mathcal{P} . This approach is pursued in Giacomini and Kitagawa (2015). For each $p(\Omega|\Phi, \Sigma) \in \mathcal{P}$ one can compute a posterior distribution $p(\theta|Y)$. Let \mathcal{P}_Y denote the resulting set of posterior distribution. One then can compute upper and lower bounds on, say, the posterior mean of θ , or a credible interval for θ that has coverage probability greater or equal than $1 - \alpha$ for every $p(\theta|Y) \in \mathcal{P}_Y$.

3.4 Score Card

Over the past decade the use of sign restrictions has become very popular in the structural VAR literature and Harald Uhlig's chapter provides a timely critical review of this literature. It is important to understand that sign restrictions define identified sets for impulse responses, which I denoted by $\Theta(\Phi, \Sigma)$. Sign restrictions taken by themselves do not imply prior distributions; they only restrict the domain of prior distributions. Because in set-identified models priors do not get updated in view of the data for some transformations of the model parameters. Thus, a careful elicitation of priors is important and some of the debates in this literature are about how to parameterize the structural VAR to facilitate the elicitation of a prior.

Instead of debating implementation details for structural VARs that are set-identified via sign restrictions, one might ask the broader question of where do sign restrictions come from? DSGE models are often used to motivate sign restrictions. However, once parameterized, they imply much stronger restrictions on structural VAR representations than sign restrictions. In most models, these are restrictions on the contemporaneous movements of the endogenous variables that cannot be represented as zero restrictions. Stepping away from DSGE models, we probably don't believe that many of the popular sign restrictions used in the literature should be literally true: it is not inconceivable that prices fall in response to an expansionary monetary policy shock, because the interest rate drop lowers financing costs for firms, which could temporarily get passed on to consumers in form of lower prices. Thus, we might want to relax the sign restriction and allow for small, temporary price drops after a monetary expansion in our prior distribution.

The structural VAR literature has by now generated hundreds of estimates of impulse response functions to monetary policy, government spending, tax, oil, and technology shocks. The sign-restriction approach has kept the literature more honest by consolidating empirical results from more restrictive identification schemes. Unfortunately, many papers focus on qualitative instead of quantitative aspects of the impulse response functions, i.e., the direction of the response or whether or not error bands cover zero. At this stage, a meta study that aggregates the quantitative results from existing studies would be of great value to the profession.

4 Conclusion

In closing, let me reiterate that a key challenge for macroeconometricians is to deliver tools that provide good characterizations of uncertainty associated with quantitative statements about future developments as well as the effect of policy interventions. The chapters by Ulrich Müller and Mark Watson on low frequency econometrics and by Harald Uhlig on structural VARs successfully confronted this challenge. As a field, macroeconometrics is well and alive and I hope the contributions in this volume will attract talented young scholars to tackle open questions and expand the frontier of knowledge at the interface of macroeconomics and econometrics.

References

- BANSAL, R., AND A. YARON (2004): “Risks For the Long Run: A Potential Resolution of Asset Pricing Puzzles,” *Journal of Finance*, 59, 1481–1509.
- BAUMEISTER, C., AND J. D. HAMILTON (2015): “Sign Restrictions, Structural Vector Autoregressions, and Useful Prior Information,” *Econometrica*, 83(5), 1963–1999.
- DEL NEGRO, M., AND F. SCHORFHEIDE (2004): “Priors from General Equilibrium Models for VARs,” *International Economic Review*, 45(2), 643 – 673.
- DOAN, T., R. LITTERMAN, AND C. A. SIMS (1984): “Forecasting and Conditional Projections Using Realistic Prior Distributions,” *Econometric Reviews*, 3(4), 1–100.

- GIACOMINI, R., AND T. KITAGAWA (2015): “Robust Inference About Partially Identified SVARs,” *Manuscript, UCL*.
- HERBST, E., AND F. SCHORFHEIDE (2015): *Bayesian Estimation of DSGE Models*. Princeton University Press.
- 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.
- 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.
- KLINE, B., AND E. TAMER (2016): “Bayesian Inference in a Class of Partially Identified Models,” *Quantitative Economics*, forthcoming.
- MOON, H. R., AND F. SCHORFHEIDE (2009): “Bayesian and Frequentist Inference in Partially-Identified Models,” *NBER Working Paper*, 14882.
- MOON, H. R., F. SCHORFHEIDE, E. GRANZIERA, AND M. LEE (2011): “Inference for VARs Identified with Sign Restrictions,” *NBER Working Paper*.
- SCHORFHEIDE, F. (2013): “Estimation and Evaluation of DSGE Models: Progress and Challenges,” in *Advances in Economics and Econometrics: Tenth World Congress*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, vol. III, chap. 5, pp. 184–230. Cambridge University Press.
- SCHORFHEIDE, F., D. SONG, AND A. YARON (2014): “Identifying Long-Run Risks: A Bayesian Mixed-Frequency Approach,” *NBER Working Paper*, 20303.
- SIMS, C. A., AND T. ZHA (1998): “Bayesian Methods for Dynamic Multivariate Models,” *International Economic Review*, 39(4), 949–968.
- UHLIG, H. (2005): “What Are the Effects of Monetary Policy on Output? Results From an Agnostic Identification Procedure,” *Journal of Monetary Economics*, 52(2), 381–419.