Bayesian Estimation of a DSGE Model with Asset Prices∗

Martin Kliem† Harald Uhlig‡
Deutsche Bundesbank University of Chicago

November 4, 2015

Abstract

This paper presents a novel Bayesian method for estimating dynamic stochastic general equilibrium (DSGE) models subject to a constrained posterior distribution for the implied Sharpe ratio. We apply our methodology to a DSGE model with habit formation in consumption and leisure, and show that the constrained estimation produces both reasonable asset-pricing as well as business-cycle implications. Next, we estimate the Smets-Wouters model subject to the same Sharpe ratio constraint. The results move the model closer to reproducing observed risk premia, but at increasing cost to its macroeconomic performance.

JEL classification: C11, E32, E44, G12.
Keywords: Bayesian estimation, stochastic steady-state, prior choice, Sharpe ratio.

∗We would like to thank Fabio Canova, Bong Geun Choi, Lars Hansen, Ivan Jacaard, John Geweke, Holger Müller, B. Ravikumar, Tom Sargent, Frank Schorfheide, Andrea Tamoni, Mu-Chun Wang, Tao Zha as well as three anonymous referees for their helpful comments. Moreover, we have benefited from presentations of earlier drafts at a variety of conferences and talks. This research was previously supported by the Deutsche Forschungsgemeinschaft through the SFB 649 “Economic Risk” and is currently supported by the NSF grant SES-1227280. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Deutsche Bundesbank or its staff.

†Deutsche Bundesbank, Economic Research Center, Wilhelm-Epstein-Str. 14, 60431 Frankfurt am Main, Germany, E-mail: martin.kliem@bundesbank.de, phone: +49 69 9566 4759.
‡The University of Chicago, Department of Economics, 1126 E 59th Street Chicago, IL 60637, E-mail: huhlig@uchicago.edu
1 Introduction

This paper presents a novel Bayesian method for estimating dynamic stochastic general equilibrium (DSGE) models subject to constraints on the posterior distribution for some implied statistic of interest, such as the Sharpe ratio, i.e. the market price for risk. This is akin to imposing a version of the Hansen-Jagannathan bound, see Hansen and Jagannathan (1997). We use the log-linearized law of motion to calculate a version of the Sharpe ratio as minus the ratio of the correlation between the log-stochastic discount factor and the log-return on capital, divided by the standard deviation of the log-return on capital, see equation (18). Starting from an initial, unconstrained prior, we construct a constrained prior so that the resulting implied posterior for some variable of interest (the Sharpe-ratio, for instance) coincides with some a priori chosen distribution, and such that the constrained prior is proportional to the original prior, conditional on that variable. We apply our methodology to a DSGE model with habit formation in consumption and leisure, real wage rigidities, and capital adjustment costs, extending Uhlig (2007). As constraint, we use a sharply peaked density centered at 0.2, a previous estimate of the Sharpe ratio. We show that the estimation subject to this constraint produces a quantitative model with both reasonable asset pricing as well as business cycle implications. Next, we estimate the Smets-Wouters model, see Smets and Wouters (2007), subject to a Sharpe ratio constraint centered around the more modest value of 0.075. The results move the model closer to reproducing observed risk premia, but at increasing cost to its macroeconomic performance.

It can be challenging to specify a dynamic stochastic general equilibrium (DSGE) model with reasonable macroeconomic implications as well as asset pricing implications. Considerable advances have been made here in the literature in recent years. One branch of the literature exploits Epstein-Zin preference specifications developed by Epstein and Zin (1989, 1991), and includes Tallarini (2000), Rudebusch and Swanson (2012), Guvenen (2009) or Piazzesi and Schneider (2007). Another branch of the literature has pursued habit formation specifications, see e.g. Abel (1990), Ljungqvist and Uhlig (2000), Campbell and Cochrane (1999), Boldrin et al. (2001), Uhlig (2007), Ljungqvist and Uhlig (2015) and Campbell et al. (2015). Finally, smartly chosen tail risk distributions and disaster risks have been successfully utilized by Barro (2006), Gabaix (2012) and Gourio (2012).

But even if a model is cleverly crafted for that purpose, the estimation of such a model may not necessarily lead to parameters allowing to deliver on both. In essence, the practical problem appears to boil down to having just a single observation on the size of the risk premium, while there are many observations helping to identify parameters crucial for the macroeconomic dynamics of the model. Thus, rather than calculating the implied posterior for the risk premium among many other statistic, our methodology instead imposes a constraint. The constraint is expressed as a
density and therefore a bit softer than an exact numerical value: this allows the use of standard sampling techniques.

Our procedure, which expands Kliem (2009), adds to the existing literature on endogenous prior choice for Bayesian estimation of DSGE models pioneered by Del Negro and Schorfheide (2008). Similarly to us, these authors propose a methodology to construct prior distributions for DSGE model parameters from beliefs about statistics of interests, which are themselves functions of the parameters. More specifically, they focus on the model-implied $p$-th order vector autoregression. Christiano et al. (2011) pointed out how to modify the approach, if some additional statistics of interest are provided: in the context here, that would be the Sharpe ratio. Like Del Negro and Schorfheide (2008) and Christiano et al. (2011), our approach amounts to multiplying the original prior with a term that involves the statistics of interest as a function of the parameters. The main difference of our approach to Christiano et al. (2011) lies in modifying the prior, so that the marginal posterior for the parameter of interest has a particular shape. This is of advantage if one wants to avoid using the data “twice”: the “prior” view about, say, the Sharpe ratio truly comes from having observed the entire data already, and it may then be odd to use this data again for calculating a posterior for that statistic of interest. The posterior will be informative about the other aspects of the model, and that will be the focus of interest. Furthermore Del Negro and Schorfheide (2008) and Christiano et al. (2011) focus on statistics from a hypothetical pre-sample and thus their sample moments, while we allow for a free implied posterior for the statistic of interest. Our emphasis here is on constraining asset pricing implications, and it may be instructive to compare these approaches, when applying the other two with that focus in mind.

To demonstrate our methodology and to set ourselves a bit of a challenge, we apply our method to two models, which do not easily account for asset pricing observations. The first and main application in section 3 is the habit-formation model in Uhlig (2007), extended to include additional shocks. Unsurprisingly and in line with a large literature, the constrained estimation delivers a high degree of risk aversion with respect to short-term fluctuations in consumption and thereby a low degree of intertemporal substitution. This is simply an unavoidable consequence of employing a log-linearized habit-formation model and seeking to match risk premia on asset markets.

The novel contribution of the methodology and insights here is to respect that consumption fluctuations now arise endogenously in the model, rather than being exogenously given as in much of the asset pricing literature: our estimation methodology then seeks to find the optimal compromise between matching macroeconomic facts and risk premia, taking into account these general equilibrium repercussions. The constrained estimator does so for the model in 3 by increasing the persistence of the response of the investment-to-output ratio, using the capital stock as a long-run buffer to smooth out short-term consumption fluctuations. This increase in the long-run volatility of the investment-output ratio may provide an intriguing connection to
the long-run risk literature for Epstein-Zin preferences, see Bansal and Yaron (2004) and Hansen et al. (2008). The estimates further suggest that labor rigidities rather than external habits in consumption play an important role for producing our results, in line with Uhlig (2007). Remarkably, the volatility of the risk-free return is similar in size to what is observed in the data and the ‘risk-free rate puzzle’ (Weil, 1989) is avoided. The Bayesian posterior odds ratio between the unconstrained and the constrained model is around 6.6, demonstrating that the constrained model is not “obviously false” from the unconstrained perspective.

This is in rather marked contrast to the constrained estimation of the Smets-Wouters model, pursued in section 4. Now, the posterior odds ratio of the unconstrained model over the constrained model is $3.7e^{24}$, even though we set ourselves the more modest target of only matching half of the observed Sharpe ratio. Clear gaps open in the macroeconomic implications of the unconstrained and constrained model. For illustration, we then additionally impose a second constraint in subsection 4.2, asking the estimated model to also deliver an upwardly sloping term structure of interest rates. It turns out that the method now hits a rather hard wall in terms of delivering on both the Sharpe ratio and the term premium, and cannot fully deliver the constraints one is asking for.

This, of course, should not surprise. If one wishes to construct a DSGE model for the purpose of, say, analyzing monetary policy or fiscal policy, and wishes that model to account for asset pricing features such as the Sharpe ratio, it is worth introducing features that it will help to do so. Our procedure is certainly not a “cure-all”. Rather, our methodology is useful for finding a good parameterization, when the model has a chance to succeed in principle, and for finding the least bad parameterization, when it does not.

The paper is organized as follows. Section 2 introduces our methodology. Section 3 applies our method to the habit-formation model of Uhlig (2007). Section 4 examines our procedure when applied to the Smets and Wouters (2007) model, including a multivariate constraint application for illustration. Section 5 concludes the paper. An online supplemental appendix provides auxiliary results.

## 2 Constructing the Constrained Prior

Our methodology can broadly be understood as a Bayesian estimation subject to a constraint: we shall describe it in general terms first, and then provide additional details for our core application to constraining asset pricing implications in a DSGE model.
2.1 Bayesian estimation subject to a posterior density constraint.

Suppose one wishes to impose the constraint (or a priori belief) that a vector of values or statistics \( \omega \in \mathbb{R}^m \), which can be expressed as a function of the parameters \( \omega = \Omega(\theta) \), lies within bounds \( a \leq \omega \leq b \), where \( a, b \in \mathbb{R}^m \). Given data \( X \), a straightforward modification of some unconstrained prior would then be to constrain its domain to this set,

\[
p(\theta|X) \propto \begin{cases} p(\theta)p(X|\theta) & \text{if} \quad \omega(\theta) \in Q \\ 0 & \text{if} \quad \omega(\theta) \not\in Q \end{cases}
\]

where \( Q \equiv \{ \theta : a \leq \omega(\theta) \leq b \} \) (1)

We wish to refine this approach for several reasons, building on insights learned from the specific application to constraining the Sharpe ratio in a DSGE model. First, since such a model will often favor parameters with a low Sharpe ratio, the approach above will lead to a pile-up of the constrained parameters at the lower bound \( a \): the results will then be rather sensitive to specifying that lower bound. Second, our a priori beliefs regarding the Sharpe ratio are more appropriately formulated as a probability distribution, or, more precisely, given by an uncertain estimate of the Sharpe ratio, given observed data, rather than imposing it to lie within some interval. Finally, the procedure becomes both more flexible and more reasonable. We argue that it is more appealing to learn the implications of the model, if the Sharpe ratio is constrained to be high with considerable probability, rather than imposing it to be within some interval. We believe this insight applies more generally to other statistics of interest as well.

Consequently, we use the following approach. The model generates data \( X \): the particular observation at hand is the realization \( \bar{X} \). We start from some “unconstrained prior” \( p(\theta) \) in the parameter vector \( \theta \). We then construct a “constrained prior” \( \tilde{p}_X(\theta) \), such that the following two properties hold:

P1: the implied posterior for the statistic of interest \( \omega = \Omega(\theta) \in \mathbb{R}^m \) coincides with some a priori given distribution \( F(\omega) \), with density \( f(\omega) \) at the given observation \( \bar{X} \).

P2: the constrained prior is proportional to the unconstrained prior, given any \( \omega \).

This approach should be viewed as a modification or refinement of the standard approach of Bayesian estimation subject to a constraint, described above. Rather than imposing an inequality constraint on a parameter or function of the parameters, we impose a probability density. We believe that (P2) is a particularly appealing property and useful benchmark. It says that prior views regarding the parameters remain unchanged, conditional on any particular value for \( \omega \). We only wish to impose that some values of \( \omega \) are more likely than others, compared to the unconstrained
prior. While that necessarily shifts the prior regarding the underlying parameters due to the re-weighting of different \( \omega \)'s, we do not want to additionally impose a change in our prior views regarding various combinations of these parameters, if they all result in the same value of \( \omega \).

More formally, we start from

1. a proper prior \( p(\theta) \), which we call the “unconstrained prior”.
2. a likelihood function \( \ell(X \mid \theta) \).
3. A mapping \( \Omega(\theta) = \omega \in \mathbb{R}^m \).
4. A differentiable probability distribution function \( F(\omega) \) on \( \mathbb{R} \), i.e. an increasing function with \( \lim_{\omega \to -\infty} F(\omega) = 0 \), \( \lim_{\omega \to \infty} F(\omega) = 1 \), and\(^1\) the corresponding probability density function

\[
f(\omega) = \frac{\partial^m F(\omega)}{\partial\omega_1 \partial\omega_2 \ldots \partial\omega_m} \tag{3}
\]

Given any data \( X \), calculate the implied unconstrained posterior for \( \omega \),

\[
G(\omega \mid X) = \int_{\theta \mid \Omega(\theta) \leq \omega} p(\theta \mid X) d\theta \tag{4}
\]

as well as its density \( g(\omega \mid X) \) defined analogously to (3). Define the “transformation function”

\[
h_X(\omega) = \frac{f(\omega)}{g(\omega \mid X)} \tag{5}
\]

Note that \( h \) depends on the particular observation \( X \), indicated by the subscript. We shall use this notation throughout, whenever that dependence needs to be indicated. It is important, that \( h_X \) does not vary with the data, when defining the constrained prior below in equation (6), i.e., \( h(\omega) \) is a function of \( \omega \) only from here onwards.

We claim that the following prior has the two desired properties listed above. Define the constrained prior \( \tilde{p}_X(\theta) \) as

\[
\tilde{p}_X(\theta) = C_X^{-1} p(\theta) h_X(\Omega(\theta)) \tag{6}
\]

where

\[
C_X = \int p(\theta) h_X(\Omega(\theta)) d\theta \tag{7}
\]

\(^1\)We assume throughout that the mathematical conditions for performing the differentiation and integration steps are satisfied.
is the integration constant. Note that (6) implies the second of the desired properties, i.e. that the constrained prior is proportional to the unconstrained prior for any given \( \omega \). We need to verify the first desired property. The implied posterior for \( \omega \), given some observation \( X \) is given by the distribution function

\[
\tilde{G}_X(\omega \mid X) = \int_{\theta \in \Omega(\theta) \leq \omega} \tilde{p}_X(\theta \mid X) d\theta
\]  

(8)

We need to show that \( \tilde{G}_X(\cdot \mid \bar{X}) = F(\cdot) \).

To that end, let \( \tilde{g}_X(\tilde{\omega} \mid X) \) be the density function for the multivariate distribution function \( \tilde{G}_X \), defined analogously to (3). Moreover, let \( \Delta = (\Delta_1, \ldots, \Delta_m) \) be a “volume element” or vector of small positive entries \( \Delta_i > 0 \). Substituting the explicit expression for the posterior, it follows that

\[
\tilde{g}_X(\tilde{\omega} \mid X) \Delta_1 \Delta_2 \cdots \Delta_m \approx \int_{\theta \in \tilde{\omega} \leq \Omega(\theta) \leq \tilde{\omega} + \Delta} \left( \frac{p(\theta) h_X(\Omega(\theta)) \ell(X \mid \theta)}{\int p(\theta) h_X(\Omega(\theta)) \ell(X \mid \theta) d\theta} \right) d\theta
\]  

(9)

Likewise,

\[
g(\tilde{\omega} \mid X) \Delta_1 \Delta_2 \cdots \Delta_m \propto \int_{\theta \in \tilde{\omega} \leq \Omega(\theta) \leq \tilde{\omega} + \Delta} p(\theta) \ell(X \mid \theta) d\theta
\]  

(10)

Therefore, at \( X = \bar{X} \),

\[
\tilde{g}_X(\tilde{\omega} \mid \bar{X}) \Delta_1 \Delta_2 \cdots \Delta_m \propto f(\tilde{\omega}) \Delta_1 \Delta_2 \cdots \Delta_m
\]  

(11)

Integrating and recognizing that both sides of this equation are probability densities delivers

\[
\tilde{G}_X(\tilde{\omega} \mid \bar{X}) = F(\tilde{\omega})
\]

as claimed.

These calculations also show what happens in the limit, as \( F(\omega) \) approaches a pointmass at a particular value for \( \omega = \omega^* \), as it would in the case of estimation subject to an exact constraint. In that case, \( f(\omega) \) and \( h_X(\omega) \) approach a multiple of a Dirac delta function at \( \omega = \omega^* \), and the dependence on the sample \( X \) disappears, except in insisting that the constrained prior \( \tilde{p}_X \) integrates to unity. This should alleviate concerns that our procedure uses the data “twice”; instead, it should be viewed as imposing a soft constraint. Moreover, near the pointmass limit, one nearly does not need to know \( g(\omega \mid X) \), unless one wishes to calculate posterior odds of models and therefore needs \( C_X \); all that one needs is to check that it is not zero, and
then to normalize $\tilde{p}$ to integrate to unity. This is of importance in cases where the unconstrained prior assigns very little mass to the area near $\omega^*$, making it hard to calculate $g(\omega \mid X)$ accurately. Of course, if the constraint is imposed rather softly, the dependence on $\bar{X}$ and the precise calculation of $g(\omega \mid X)$ matters relatively more.

The following algorithm provides a numerical implementation of our methodology:

1. Estimate the model by sampling from the unconstrained posterior $p(\theta \mid X)$
2. Approximate the implied unconstrained probability density function $g(\omega \mid \bar{X})$
3. Calculate the transformation function $h_{\bar{X}}(\omega)$ per (5).
4. Estimate the model with the constrained prior by sampling from $\tilde{p}_{\bar{X}}(\theta \mid X)$.
5. For Bayesian posterior odds ratio calculations and model comparisons, for example, calculate the normalization constant $C_{\bar{X}}$ of equation (7).

The choice of the imposed implied posterior for the variable of interest is entirely up to the researcher employing this methodology and simply a generalization of textbook estimation-subject-to-constraint. From a pure methodological perspective, it should not matter much, whether such constraints are imposed “a priori” in order to express strong beliefs about the properties of some parameter, or “a posteriori” in order to learn more about certain features of a model. From the perspective of econometrics as conversation and rhetoric, see McCloskey (1983), we view this as a practical and appealing way to explore and communicate model properties within a range of specifications of interest, rather than being forced to examine model estimates, that deliver non-credible implications for key variables. For these practical reasons, one may then wish to obtain a data-driven implied posterior for the variable of interest. Indeed, the main difference of our approach to Christiano et al. (2011) lies in modifying the prior, so that the marginal posterior for the parameter of interest has a particular shape. This is of advantage if one wants to avoid using the data “twice”: the “prior” view about, say, the Sharpe ratio truly comes from having observed the entire data already, and it may then be odd to use this data again for calculating a posterior for that statistic of interest. The posterior will be informative about the other aspects of the model, and that will be the focus of interest.

2.2 Constraining the Sharpe Ratio

While we have described the method with the potential of imposing a multivariate constraint, we shall focus on a univariate constraint in the bulk of the application below, except for subsection 4.2. In particular, we shall focus on constraining the Sharpe ratio on the returns to capital investments.

As is well known, many estimated DSGE models typically have poor asset pricing implications, unless designed otherwise or — and that is the perspective in this paper
— unless the estimation pays particular attention to these aspects. From the perspective of model applications, it is interesting to know the properties of the model, if the estimation procedure is constrained to deliver certain asset pricing implications. In our specific example, we seek to constrain our prior to deliver particular implications regarding the Sharpe ratio, i.e. the market price for risk.

We impose the constraint per exploiting the implications for that market price for risk, given the log-linearized dynamics. Following Campbell (1994) and Uhlig (1999), log-linearize the given (and possibly de-trended) model and solve for the recursive law of motion,

$$\hat{y}_t = A\hat{h}_{t-1} + B\epsilon_t ,$$

(12)

where $\hat{y}_t = \log (y_t) - \log (y^{ss})$ is a vector containing all log-linearized model variables and $\hat{h}_t = \log (h_t) - \log (h^{ss})$ is the vector containing all log-linearized state variables of the model, with $y^{ss}$ and $h^{ss}$ as their corresponding steady state values. The entries in the matrices $A$ and $B$ can typically be interpreted as elasticities and are functions of the model parameters $\theta$,

$$A = A(\theta), \ B = B(\theta)$$

Following Lettau and Uhlig (2002) and using the representation (12), we can decompose the log pricing kernel into its conditional expectation and its innovations:

$$\hat{M}_{t+1} = E_t \left[ \hat{M}_{t+1} \right] + b_M \epsilon_{t+1} ,$$

(13)

where $b_M$ indicates the row vector of matrix $B$ with respect to the pricing kernel. Let $\Sigma = E_t \left[ \epsilon_{t+1} \epsilon_{t+1}' \right]$ be the variance-covariance matrix of $\epsilon_t$: we assumed it to be diagonal, but the formulas apply more generally. $\Sigma$ is a function of the underlying parameters $\theta$,

$$\Sigma = \Sigma(\theta)$$

The conditional variance, $\sigma^2_M$ of the pricing kernel is

$$\sigma^2_M = b_M \Sigma b_M' .$$

(14)

Similarly, we can solve for the conditional variances $\sigma^2_{Rk}$ of $\hat{R}_k^t$, the log-linearized return on some asset (where we used the superscript $k$ to refer to the typical case of calculating the return of investing in capital). The conditional covariance of the pricing kernel and the return on the asset, $\sigma_{MRk}$ is

$$\sigma_{MRk} = b_M \Sigma b_{Rk}' ,$$

(15)

with $b_{Rk}$ the row vector of matrix $B$ for the return on the asset $\hat{R}_k^t$.

Given the log-linear law of motion and exploiting the joint normality of the innovations of the logs of the variables, one can now calculate the implied risk premium on any asset, as in Lettau and Uhlig (2002). This is similar to the second-order “update” of the steady state in Schmitt-Grohe and Uribe (2004), in light of the first order
dynamics, and exploits that $E[X] = E[\exp(\log(X))] = \exp(E[\log(X)] + \text{Var}[\log(X)])$, for a normally distributed random variable $\log(X)$. The formulas below would be exact, if the log-linear law of motion and joint normality were exact: the approximation error is in the latter. The implied risk premium on the return of the asset over the risk free return satisfies\footnote{For example, $E_t[R_{t+1}^k] = E_t[\exp(E_t[\log(R_{t+1}^k)])] = \exp(E_t[\log(R_{t+1}^k)] + b_R^k \epsilon_{t+1}) + \sigma_R^2 / 2).$ A few more similar calculations deliver the equations in the text.}

$$\log E_t[R_{t+1}^k] - \log R^f_t = -\sigma_{MR^k}. \tag{16}$$

Likewise, for the Sharpe ratio $\omega$, defined here as in Lettau and Uhlig (2002) as

$$\omega = \frac{\log E_t[R_{t+1}^k] - \log R^f_t}{\sigma_{R^k}} \tag{17}$$

we obtain

$$\omega = -\frac{\sigma_{MR^k}}{\sigma_{R^k}}. \tag{18}$$

This provides the target for our constraint in the constrained estimation procedure. Equation (18) provides us with the mapping $\Omega(\theta) = \omega \in \mathbb{R}$, which is required in the method described above. Put differently, we shall impose constraints on the Sharpe ratio for investing in capital, as this can be measured in the data using stock returns.

As Hansen and Jagannathan (1997) have shown and as was subsequently investigated further in Campbell and Cochrane (2000) and Lettau and Uhlig (2002), the highest possible Sharpe ratio is equal to $\sigma_M$, when the correlation between the pricing kernel and the return of capital is equal to -1. Alternatively, one could therefore use the mapping

$$\Omega(\theta) = \omega = \sigma_M$$

if one wishes to impose a constraint on the maximal market price for risk, i.e. impose a version of the Hansen-Jagannathan bound, see Hansen and Jagannathan (1997).

To implement the procedure practically, one needs the densities $f(\omega)$ and $g(\omega | X)$. We proceed by approximating a sample histogram from the unconstrained posterior with a gamma density, to calculate $g(\cdot)$. With that, one obtains an approximation to $h_X(\omega)$, which is easy to calculate accurately, but one may feel uncomfortable that the approximation is accurate around the desired peak $\omega^* = \text{argmax}_\omega f(\omega)$ of $f(\cdot)$. As an alternative, one could pursue a kernel density estimation procedure for $g(\omega | X)$ in particular around $\omega^*$, but if $\omega^*$ is very unlikely under the unconditional prior, this may result in long computations. In the extreme case, there is no parameter combination resulting in $\omega$ near $\omega^*$, in which case $g(\omega | X)$ is zero in a neighbourhood of $\omega^*$ and the true $h_X(\omega)$ is not defined. With a kernel density approximation procedure for $g(\omega | X)$, this case may be computationally hard to distinguish from the fundamentally rather different case of a very low density. Our (potentially rather mistaken) gamma
approximation to \( g(\omega \mid X) \) avoids these numerical challenges and potentially intense calculations at this stage, and instead produces a numerical challenge in the next step, when sampling from the constrained prior \( \tilde{p}_X(\theta) \). We sample from this prior by first obtaining its peak and calculating the Hessian there, for the purpose of constructing a normal density approximation and starting the MCMC algorithm. Consider first the extreme case that the true \( g(\omega \mid X) \) is zero around \( \omega^* \). The maximization procedure then attempts to essentially climb \( h_X(\omega) \), but then stops at a sharp boundary, encountering a nondifferentiability at the peak of \( \tilde{p}_X(\theta) \) and before reaching \( \omega^* \). For the same reason, it may even encounter difficulties of finding the peak, when using versions of the Newton-Raphson algorithm. While these numerical challenges are solvable in principle, it now is already clear that a parameterization with the desired \( \omega^* \) cannot be obtained. So, rather than doggedly proceeding with the original \( f(\omega) \), we recommend adjusting this function and the target \( \omega^* \) to a more modest value, where these adverse conditions do not hold, and restarting the calculations. Even in the less extreme case, where \( \omega^* \) is reachable in principle, but the true \( g(\omega \mid X) \) falls off rather sharply before reaching that value, the numerical challenges in computing the peak and the Hessian there may be rather similar in practice, and our recommendation of adjusting the target \( \omega^* \) to a more modest value stays the same. One simply needs to recognize that our econometric approach is not a “cure-all” to fundamental deficiencies of the model itself. Some may wish to proceed in this slightly more benign case anyhow, however.

The particular implementation in the context of our first application is discussed in subsection 3.3 and shown in figure 1. Note that, by choosing a Gamma distribution, we implicitly impose that the Sharpe ratio is positive: this seems reasonable on economic grounds. However, for our application the choice of the Gamma distribution itself could also be substituted by, e.g., a log-normal distribution or an other distribution which imposes a positive Sharpe ratio. We approximate \( f(\omega) \) by following our estimates for the Sharpe ratio and assume that \( f(\omega) \) is Gamma distributed with mean 0.2049. For our second application in section 4, we found that \( g(\omega \mid X) \) is numerically ill-behaved around our initial Sharpe ratio target. Given the practical considerations explained above and in order to avoid moving the macroeconomic implications too far, we reduce the mean to a third of that value and set ourselves a more modest goal in terms of asset pricing implications. Moreover, it turns out that we thus need to impose some \( f(\omega) \), which is considerably tighter than would be justified on estimation uncertainty of the Sharpe ratio alone. However, this ensure that the tails of \( h \) die out and that we therefore guarantee a proper prior and, therefore, we choose a standard deviation of 0.001 for \( f(\omega) \). Since our main aim is a flexible implementation of estimation subject to constraint rather than the imposition of a particular data-driven prior, we view this as a feature, not a bug. Moreover, given that the exercise here is to obtain a model estimate in line with the observed Sharpe ratio as well as to illustrate our estimation methodology, we view this as perfectly
adequate: after all, this is part of the prior specification, and not necessarily entirely data-driven.

Finally, we calculate the normalization constant \( C_X = \int p(\theta)h(\Omega(\theta))d\theta \) of equation (7). To that end, we apply an MCMC algorithm and use the Modified Harmonic Mean estimator by Geweke (1999) to compute the integral. If one can show that the Sharpe ratio only depends on a few parameters of the model, and the unconditional prior is a product of densities in these parameters and a density of the other parameters, then one can focus the calculation of the constant entirely on these few parameters, since that is all where the adjustment takes place. This Rao-Blackwellization can then help to achieve accuracy more easily\(^3\). Note that this constant is only used for model comparison, though, since the MCMC-based method for constrained-posterior inference do not require knowledge of the normalizing constant. If such a model comparison is intended as an overall “rough” diagnostic, one may not need to compute the normalizing constant with great precision or other methods, such as a Laplace approximation, may suit the purpose as well.

Mathematically, one can obtain these restrictions on \( \omega \) from the log-linearized dynamics, even if one calculates the log-linearization around a nonstochastic steady state without a risk premium for the various assets. One may feel that doing so builds in a logical inconsistency, however. If the risk premium on investing in, say, capital is imposed to be fairly high, the steady state level of capital should be fairly low, in order to generate this additional risk premium on average. Therefore, we follow Juillard (2011) and Coeurdacier et al. (2011), by evaluating the model around its risky steady state which incorporates information about the stochastic nature of the economic environment. For the numerical solution, we impose that

\[
E \left[ R^k_t \right] = E \left[ \exp \left( -\log \bar{M} - E_t \left[ \hat{M}_{t+1} \right] - \frac{\sigma^2_M}{2} \right) \right],
\]

(19)

as well as

\[
E \left[ R^k_{t+1} \right] = E \left[ \exp \left( -\log \bar{M} - E_t \left[ \hat{M}_{t+1} \right] - \frac{\sigma^2_M}{2} - \sigma_{MR^k} \right) \right].
\]

(20)

Given \( \bar{M} \) as well as variances and covariances, note that the right-hand side can once again be calculated using (12) and the assumption of normally distributed shocks.

Because the conditional second moments depend on the policy function, the stochastic steady state is computationally more demanding than the deterministic steady state. In particular, the calculation requires the steady state and the corresponding dynamics to be jointly determined (see, e.g., de Groot, 2013). Therefore, we get a fixed point problem by solving our model accurately with respect to the stochastic steady state. For this reason, we use an iterative procedure. We start with the deterministic steady state to obtain our policy function. The resulting steady

\(^3\)We are grateful to the editor to pointing this out
state adjustment yields a new set of policy functions, etc.. As discussed in Canton (2002), a few iterations suffice to achieve convergence and to resolve the fixed point problem to a reasonable degree of accuracy.

3 Application 1: Habit Formation and Asset Pricing

We apply our methodology to a dynamic stochastic general equilibrium model with external habit formation or “catching-up with the Joneses” (see Abel, 1990) in both consumption and leisure, building on Uhlig (2007). We extend the model by adding shocks, to permit estimation, as well as solving the model around the stochastic steady state. We briefly describe it here for the sake of completeness.

3.1 The model

Output $y_t$ is produced by a competitive sector of firms. Each firm produces output $y_t$ with capital $k_{t-1}$ in place from the previous period as well as labor $n_t$ per the Cobb-Douglas production function,

$$y_t = k_{t-1}^\theta (e^{z_{P,t} n_t})^{1-\theta}.$$  (21)

where $z_{P,t}$ is a productivity or technology parameter. We assume it to follow a random walk with drift,

$$z_{P,t} = \gamma + z_{P,t-1} + \epsilon_{P,t}, \epsilon_{P,t} \sim N(0,\sigma_P),$$

with $\gamma$ reflecting the trend.

There is a representative household with the utility function

$$U = E \left[ \sum_{t=0}^{\infty} \beta^t \left( (c_t - H_t) (A + (e^{z_{L,t} l_t} - F_t)^\nu) \right)^{1-\eta} - 1 \right].$$  (22)

The discount factor $\beta$ and $A, \nu, \eta$ are parameters, which we assume to satisfy $\nu > 0$ and $\eta > \nu/(\nu + 1)$ in order to assure monotonicity and concavity, see Uhlig (2007). The variables $c_t$ and $l_t$ denote consumption and leisure of the particular household. The utility depends on the economy-wide average level of consumption habit and leisure habit, $H_t$ and $F_t$, evolving according to

$$H_t = e^\gamma \left( (1 - \rho_c) \chi C_{t-1} + \rho_c H_{t-1} \right),$$  (23)

$$F_t = (1 - \rho_l) \psi L_{t-1} + \rho_l F_{t-1},$$  (24)

where $C_t$ and $L_t$ are aggregate average levels of consumption and leisure: in equilibrium, $C_t = c_t$ and $L_t = l_t$. The parameters $\rho_c, \rho_l, \chi$ and $\psi$ determine the persistence and importance of the habit features. The variable $z_{L,t}$ represents a labor supply shock. We assume it to follow an AR(1) process,

$$z_{L,t} = \pi_L z_{L,t-1} + \epsilon_{L,t}, \epsilon_{L,t} \sim N(0,\sigma_L).$$

Total time endowment is normalized to unity, so that total labor supply is $n_t = 1 - L_t$.  


The budget constraint of the agent is
\[ c_t + x_t + T_t = d_t k_{t-1} + w_t n_t. \]  
(25)

Capital accumulation is affected by a depreciation rate \( \delta \) and investment adjustment costs \( g(\cdot) \),
\[ k_t = \left( 1 - \delta + g\left( e^{x_{t-I,t} k_t} \right) \right) k_{t-1}. \]  
(26)

Following Jermann (1998), we assume the adjustment cost function \( g(\cdot) \) to satisfy
\[ g(\tilde{\delta}) = \delta + \exp(\gamma) - 1, \quad g'(\tilde{\delta}) = 1, \quad g''(\tilde{\delta}) = -1/\zeta \quad \forall \zeta > 0, \]
where \( \tilde{\delta} = \exp(\gamma) + \delta - 1 \), to adjust for trend growth. Adjustment costs are affected by
the parameter \( z_{I,t} \), following the AR(1) process:
\[ z_{I,t} = \pi_I z_{I,t-1} + \epsilon_{I,t}, \quad \epsilon_{I,t} \sim N(0, \sigma_I). \]

Given initial capital \( k_{-1} \), the household maximizes its utility by choosing leisure \( l_t \), consumption \( c_t \), and investments \( x_t \) subject to the constraints (25) and (26), taking
as given the exogenous habits \( H_t \) and \( F_t \) and their aggregate evolution, real wages \( w_t \),
dividends \( d_t \) and lump-sum taxes \( T_t \).

The agent’s first-order condition for labor supply yields the frictionless wage or
the marginal rate of substitution,
\[ w_t^f = \frac{U_L}{U_c}. \]  
As motivated in Uhlig (2007), we assume real wage rigidities as postulated by e.g. Hall (2005), Shimer (2005), and
Blanchard and Galí (2007). More precisely, we assume that
\[ w_t = (e^{\gamma} w_{t-1})^\mu \left( e^{\omega + z_{W,t} w_t} \right)^{1-\mu}. \]  
(27)

The parameter \( \omega > 0 \) represents an average wage markup to ensure that \( w > w_t^f \)
locally around the steady state, and that therefore the labor market is (typically) demand constrained. The wage markup \( z_{W,t} \) follows an AR(1) process:
\[ z_{W,t} = \pi_W z_{W,t-1} + \epsilon_{W,t}, \quad \epsilon_{W,t} \sim N(0, \sigma_W) \]
The parameter \( \mu \) reflects the degree of real wage stickiness. In the special case of \( \mu = \omega = z_{W,t} = 0 \), there are no frictions and wages
are fully flexible.

Finally, there is a government, financing an exogenously given stream of expenditures \( G_t \) with lump sum taxes, \( G_t = T_t \). We assume that \( G_t = \bar{g} \exp(z_{P,t-1} + z_{G,t}) \),
where \( \exp(z_{P,t-1}) \) appears to assure a stationary spending-to-output ratio and where \( z_{G,t} \) is assumed to follow the AR(1) process:
\[ z_{G,t} = \pi_G z_{G,t-1} + \epsilon_{G,t}, \quad \epsilon_{G,t} \sim N(0, \sigma_G) \]

For the further analysis, define \( \lambda_t \) as the marginal utility of consumption and define
the stochastic discount factor \( M_t = \beta \lambda_t^{\lambda_{t-1}} \). Risk premia arise from investigating the Lucas asset pricing equation
\[ 1 = E_t [M_{t+1} R_{t+1}] , \]  
(28)

where \( R_{t+1} \) is the one-period return on investing one unit of resources. For investing
in capital, for example, let \( q(\cdot) \) be the shadow price of a unit of capital and \( R_k^t \) be the return for investing in capital.

The five entries of the shock vector \( \epsilon_t = [\epsilon_{P,t}, \epsilon_{L,t}, \epsilon_{f,t}, \epsilon_{W,t}, \epsilon_{G,t}]' \) are assumed to be independent. Equilibrium is defined as usual.

For the numerical analysis, the variables \( k_t, y_t, c_t, H_t, w_t, w_f^t, x_t, \lambda_t, \) and \( G_t \) have to be productivity-detrended to solve the model. This is done by dividing each variable by \( \exp(z_{P,t} - 1) \), except capital \( k_t \), which is detrended with \( \exp(z_{P,t}) \) and \( \lambda_t \) which is detrended by \( \exp(-\eta z_{P,t} - 1) \). Beside this, \( l_t, F_t, n_t, q_t, R_f^t, R_k^t, M_t, \) and \( d_t \) are stationary. We use a logarithmic approximation around the stochastic detrended steady state for our computations and solve for the recursive law of motion. In the following, all detrended variables are marked with \( \sim \) and the log-deviations from the detrended variables are marked with \( \wedge \). For details see the supplemental appendix.

### 3.2 Data

The estimation of the model is based on six time series from 1963:qI to 2008:qII. All data are quarterly and in real terms.\(^4\) For both the unconstrained and constrained estimation, we use the vector of time series \( X_t = [\Delta \bar{y}_t, \Delta \bar{c}_t, \bar{n}_t, \bar{x}_t - \bar{y}_t, \bar{R}_f^t, \bar{R}_q^t] \), where \( \Delta \bar{y}_t \) is the demeaned first difference of log output per capita, \( \Delta \bar{c}_t \) is the demeaned first difference of log consumption per capita, \( \bar{x}_t - \bar{y}_t \) is the demeaned log investment-to-output ratio, \( \bar{n}_t \) is de-meaned hours worked per capita, \( \bar{R}_f^t \) is the demeaned risk free rate, and \( \bar{R}_q^t \) are the excess returns on an aggregate stock market index. Because there is no equivalent variable in our model, we define log excess returns as follows

\[
R_q^t = -\sigma_{MR} + \frac{1}{1 - \Upsilon} \left( \hat{R}_k^t - \hat{R}_f^t \right) + \epsilon_{Q,t},
\]  

(29)

where \( \Upsilon \) is a parameter which can be interpreted as leverage and \( \epsilon_{Q,t} \) is an i.i.d. error term and assumed to be uncorrelated with the stochastic discount factor. One may wish to simply interpret \( \epsilon_{Q,t} \) as a measurement error. Conceptually, however, the term reflects random variations in dividend policies or random fluctuations in demands by heterogeneous investors, which require far more detailed modeling, but which we view as largely immaterial for the aggregate quantitative dynamics of the model.\(^5\) Similar to Section 2.2, we derive the mean excess returns as the negative covariance between the pricing kernel and the excess returns, \( -\sigma_{MR} \). While we can observe the Sharpe

\(^{4}\)See supplemental appendix for details on the source and a description of any data used in this paper.

\(^{5}\)In the corresponding appendix, we present a historical and variance decomposition to investigate the role of \( \epsilon_{Q,t} \) in more detail. Unsurprisingly, this shock explains most of the variation in \( R_q^t \). Per our model assumptions, the pricing kernel does not depend on that shock and therefore the shock does not contribute to the Sharpe ratio. We re-check this model property numerically in the supplemental appendix.
ratio of the total returns on an aggregate stock market index in the data, we cannot observe a Sharpe ratio for the return on economy-wide capital in the data. Hence, we assume that in our economy all assets are priced along the implied market line and therefore excess returns and capital returns share the same Sharpe ratio. Thus, we use $\omega \cdot \sigma_{R}$ as a proxy for the mean excess returns.

### 3.3 Choice of the prior

As in Uhlig (2007), we find it useful to describe our unconstrained prior in terms of an economically meaningful transformation of the model’s parameter. Specifically, consider the Frisch elasticity of labor supply, defined as the elasticity of labor supply to frictionless wages by holding the marginal rate of consumption constant,

$$
\tau = \frac{dn}{dw} \frac{w^f}{n} \frac{1}{\bar{U}_c}. \quad (30)
$$

Given our preference assumptions, this yields

$$
\tau = \frac{U_n}{\bar{n} \left( U_{nn} - \frac{U_{nc}^2}{U_{cc}} \right)} = \frac{1 - \bar{n}}{\bar{n}} \cdot \frac{\eta (1 + \alpha) (1 - \psi)}{\eta (\alpha (1 - \nu) + 1 + \nu) - \nu}, \quad (31)
$$

where $\alpha = A (1 - \psi)^{-\nu} \bar{l}^{-\nu}$. Therefore, rather than specifying a prior for $A$ or $\nu$ we shall specify a prior for $\tau$, and calculate the implied $A$ and $\nu$ from $\tau$ as well as the other parameters and variables in equation (31). We assume that the steady state level of hours worked is $n = 1/3$. From the first-order conditions,

$$
\nu = 1 - (1 - \psi) \frac{1}{1 - \bar{l}} \frac{1}{\tau} - \left( 2 - \frac{1}{\eta} \right) \frac{1}{(1 - \chi) \kappa}, \quad (32)
$$

with

$$
\kappa = \frac{e^\varpi \bar{c} \bar{y}}{1 - \bar{\theta} - \bar{l} \bar{y}}, \quad (33)
$$

where the $\bar{c}/\bar{y}$ is the steady state consumption share of output.\(^6\) Additionally, we can solve for the remaining preference parameters,

$$
\alpha = \frac{\kappa \nu (1 - \chi)}{1 - \psi} - 1 \quad (34)
$$

$$
A = \alpha (1 - \psi)^{\nu} \bar{l}^{\nu} \quad (35)
$$

While the real business cycle literature often assumes a relatively high Frisch elasticity of two or more (King et al., 1988; Prescott, 1986), recent papers of Bayesian

\(^6\)More details regarding steady state calculation can be found in the supplemental appendix.
DSGE model estimation found far smaller values for the Frisch elasticity in a New Keynesian model framework. For example, Justiniano and Primiceri (2008) argue for values between 0.25 and 0.5. These findings are in line with some micro-data based studies, which argue for small values in a range between 0 and 0.7, too (see Pistaferri, 2003, and references therein). To that end, we use a prior for the inverse of the Frisch elasticity, which is Gamma distributed with mean 1.0 and a standard deviation of 0.75. This assumption covers the values used in the different strands of the literature.

For explaining business cycle facts and asset pricing facts simultaneously, we expect that the discount factor $\beta$ as well as the power utility parameter $\eta$ play an important role. The business cycle literature often uses values for the discount factor that are slightly smaller than one to ensure a positive time preference of the representative agent and steady-state risk-free returns comparable to observed returns. However, from an asset pricing perspective, discount factors with much smaller values or values greater than one are postulated. These opposing assumptions are known as the risk-free rate puzzle (see Weil, 1989). However, Kocherlakota (1990) has shown that values for the discount factor above unity can be in line with positive time preference if the economy is growing. For this reason, we use a prior information for the riskless return to ensure positive time preference and solve recursively for the discount factor:

$$\beta = \exp(\eta \gamma - \sigma^2_M/2 - \log(\bar{R}^f))$$

In particular, we assume that the steady state risk-free real quarterly return, $\log(\bar{R}^f)$, is inverted-gamma distributed with mean 0.005 and standard deviation 0.01. This ensures that the mass of the prior is on positive real annual returns which are smaller than 4%. Finally, we assume that the power utility parameter is uniformly distributed between 1 and 200 which implies most of the prior mass on high values. This is in contrast to the common business cycle literature which generally assumes small values and therefore uses quite informative prior, but allows our procedure to consider high degrees of risk aversion when seeking to match asset pricing facts: this is useful when constraining the prior in the next step. The prior for the remaining deep model parameters are chosen in line with the recent literature. An overview of the priors can be found in the supplemental appendix.

In addition, to the steady state labor supply of $1/3$, we also calibrate the growth rate of the economy $\gamma$ and the capital share $\theta$. As mentioned in the previous Subsection, we calibrate the growth rate equal to the observed value of 0.0044 per quarter. The capital share is calibrated to 0.33 as common. Finally, we assume a steady-state government expenditure to output ratio, $\bar{g}/\bar{y}$, of 0.28 and the average wage markup $\varpi$ of 0.05.

To obtain the constraint on $\omega$, we approximate $g(\omega \mid \bar{X})$ using a Gamma distribution with mean 0.02 and standard deviation 0.002. Figure 1 shows the implied unconstrained probability density function $g(\omega \mid \bar{X})$ and its approximation by a Gamma distribution.
3.4 Estimation Results

We estimate the posterior mode of the distribution and employ a random walk Metropolis-Hastings algorithm to approximate the uncertainty distribution of the parameters. We run two chains, each with 300,000 parameter vector draws. The first 75% have been discarded. We provide results for both the unconstrained and the constrained prior. The posterior means for the constrained estimation are $\mu = 0.29, \eta = 108.18, \chi = 0.84, \psi = 0.84, \rho_c = 0.66, \rho_l = 0.07, \gamma = 0.11, \zeta = 7.84, \delta = 0.02, 1/\tau = 7.11, \log(\hat{R}^f) \times 100 = 0.32, \pi_G = 0.93, \pi_I = 0.70, \pi_W = 0.93, \pi_L = 0.65, \sigma_p \times 100 = 0.90, \sigma_I \times 100 = 1.32, \sigma_L \times 100 = 0.29, \sigma_W \times 100 = 2.11, \sigma_G \times 100 = 1.90, \sigma_Q \times 100 = 7.38$. In the supplemental appendix, we show detailed posterior statistics, e.g. the posterior mean and the highest probability density (HPD) interval, using the range between 5% and 95%. The results indicate that the posterior distributions of all structural parameters are well approximated and different from the prior distribution. In the following, we are highlighting the key differences between both estimation results.

By comparing the results of the estimation with unconstrained prior and constrained prior, we find the biggest difference for the power utility parameter $\eta$. This result is expected, because introducing the constrained prior decisively shifts the marginal prior distribution with respect to $\eta$ to high values. Additionally, for this class of preferences, the parameter is directly linked to the agents’ relative risk aversion regarding short-term fluctuations in consumption, which can be calculated as:

$$RRA = \frac{\eta}{1 - \chi},$$  \hspace{1cm} (37)

We calculate the relative risk aversion for every draw from the posterior. This calculation of the relative risk aversion regarding short-term fluctuations in consumption has been criticized by Boldrin et al. (1997), who argue that relative risk aversion
with respect to wealth is more meaningful. Likewise, Swanson (2012) argues that this measure ignores the labor margin. We do not seek to take a stand here on this debate: rather, our results illustrate the intuitive insight that high relative risk aversion for short-term consumption fluctuations is needed to explain stylized asset pricing facts. In particular, the implied posterior mean of the relative risk aversion with unconstrained prior and constrained prior is 39.2 and 719.25, respectively.

For both estimations we identify similar volatilities of the exogenous shocks. This means, both economies face the same “economy-wide” risk. Since high economy-wide risk is therefore not at the heart of matching the high Sharpe ratio according to these estimates, a high relative risk aversion (in the sense described above) is unavoidable, see also Rudebusch and Swanson (2008) or Lettau and Uhlig (2002). For habit-based DSGE models as presented in the present paper, the elasticity of intertemporal substitution is the inverse of the relative risk aversion. This elasticity can be calculated as 0.026 and 0.0013 for the estimation with unconstrained prior and constrained prior, respectively. Compared to the findings by Hall (1988) or Vissing-Jørgensen (2002), these are very small, but unavoidably so due to the need for a high degree of relative risk aversion. Some recent DSGE models likewise postulate high parameter values for external or internal habit and therefore also imply small elasticities. As shown by Uhlig (2007), wage rigidities can be a helpful ingredient to explain a high risk premium in habit-based DSGE models. Our estimation results, however, show that the degree of wage rigidity is small and similar for both estimation. Instead, the constrained estimation prefers a smaller Frisch elasticity $\tau$, compared to the unconstrained estimation: the Frisch elasticity of labor supply $\tau$ decreases from 0.18 for the unconstrained estimation to 0.14 for the constrained estimation. Both values are in line with findings of the microeconomic literature (see Pistaferri, 2003), but are at odds with some of the macroeconomic literature. We feel comfortable with these results, however, since the model is nonetheless capable of matching aggregate labor fluctuations as shown in Subsection 3.6: it is these aggregate observations which are at the heart of the motivation for the large macroeconomic Frisch elasticities used elsewhere. Intuitively, these estimates indicate high labor market rigidities for the estimation with the constrained prior, in line with the insights of Uhlig (2007), but focus on supply elasticities rather than wage elasticities.

Figure 2 shows the unconstrained and constrained prior and posterior distributions for selected key parameters. In particular, Figure 2(a) and 2(b) present the prior and posterior of the unconstrained and constrained estimation respectively. The constrained of the Sharpe ratio affects the prior for $\eta$ strongly. While the prior distribution is still diffuse, the prior is also more concentrated around values of 100. In comparison, as illustrated in Figure 2(c) the prior of the inverse of the Frisch elasticity $1/\tau$ is not affected by the constrained. Finally, Figure 2(d) illustrates the the mechanism behind our approach. The constrained prior and corresponding posterior are close to be identical as intended.
Figure 2: Prior and posterior distribution for selected parameters. The solid grey and solid black lines correspond to the unconstrained prior and posterior distribution, while the dashed grey and dashed black lines correspond to the constrained prior and posterior distribution respectively.

3.5 Implied asset pricing facts

Table 1 shows the implied distribution of the first and conditional second moments for both estimations. In general, the estimation with constrained prior delivers asset price facts similar to those observed in the data. The estimation with unconstrained prior can only explain the moments of the risk-free rate appropriately. Especially, the Sharpe ratio and the risk premium illustrate the well known difficulty of explaining asset pricing facts using standard DSGE models.

The conditional second moments of the risk free rate are similar for both estimations and comparable to the data. In particular, the mean of risk free return for the estimation with constrained prior is slightly smaller but still in line with observations. Additionally, the moments of the return of capital for both estimations are comparable with each other. The high risk aversion parameter $\eta$ delivers the large conditional volatility of the stochastic pricing kernel $\sigma_M$, which in turn is needed to obtain the observed Sharpe ratio: this standard deviation is also the maximal Sharpe ratio for any asset and approximately five times as high as the Sharpe ratio for the return on capital. Put differently, the constrained estimate of our model is in principle compat-
Table 1: Implied quarterly asset pricing facts by the estimated models. All values in percent with the exception of the Sharpe ratio.

<table>
<thead>
<tr>
<th></th>
<th>Unconstrained prior</th>
<th>Constrained prior</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Posterior Mean</td>
<td>5%</td>
<td>95%</td>
</tr>
<tr>
<td>Mean risk-free rate x 100</td>
<td>log($R^f$)</td>
<td>0.47</td>
<td>0.27</td>
</tr>
<tr>
<td>s.d. risk-free return</td>
<td>$\sigma_{R^f}$</td>
<td>0.39</td>
<td>0.35</td>
</tr>
<tr>
<td>s.d. return on capital</td>
<td>$\sigma_{R^k}$</td>
<td>1.19</td>
<td>0.97</td>
</tr>
<tr>
<td>s.d. excess returns</td>
<td>$\sigma_{R^q}$</td>
<td>7.86</td>
<td>7.20</td>
</tr>
<tr>
<td>s.d. pricing kernel</td>
<td>$\sigma_M$</td>
<td>4.16</td>
<td>1.90</td>
</tr>
<tr>
<td>Risk premium ($R^k/R^f$)</td>
<td>$-\sigma_{M R^k}$</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>Mean excess returns</td>
<td>$\omega \cdot \sigma_{R^q}$</td>
<td>0.16</td>
<td>0.11</td>
</tr>
<tr>
<td>Sharpe ratio ($R^k$) x 100</td>
<td>$\omega$</td>
<td>2.03</td>
<td>1.40</td>
</tr>
</tbody>
</table>

Table 1: Implied quarterly asset pricing facts by the estimated models. All values in percent with the exception of the Sharpe ratio.

The improvement regarding the asset pricing implications may worsen the macroeconomic implications and, potentially, dramatically so. Remarkably, the predicted standard deviations and contemporaneous correlations with output for the HP-filtered model variables do not change much, as one moves from the unconstrained to the constrained prior, see table 4 in the supplemental appendix. That table shows that with both, the standard deviation of output is underpredicted a bit and consumption volatility is overpredicted a bit, compared to the data, but overall, that table suggests a decent match to macroeconomic observations either way.

For a more detailed examination, we compare the cross-correlations of the hp-filtered variables from the two model analyses (“unconstrained” and “constrained”) to those generated from a Bayesian vector autoregression (BVAR) with two lags, using HP-filtered data,

7 We fit a BVAR(2) to HP-filtered variables. We assume a weak Normal-Whishart prior for the coefficients and the covariance matrix of the BVAR. For the comparison, we draw 1200 parameter vectors from the posterior of the BVAR as well as 1200 parameter vectors from the posterior distributions of both estimated DSGE models.

To find a difference between the predictions from the constrained and the unconstrained prior, one needs to look elsewhere. We repeat the exercise above, though this

3.6 Implied business cycle facts

The improvement regarding the asset pricing implications may worsen the macroeconomic implications and, potentially, dramatically so. Remarkably, the predicted standard deviations and contemporaneous correlations with output for the HP-filtered model variables do not change much, as one moves from the unconstrained to the constrained prior, see table 4 in the supplemental appendix. That table shows that with both, the standard deviation of output is underpredicted a bit and consumption volatility is overpredicted a bit, compared to the data, but overall, that table suggests a decent match to macroeconomic observations either way.

For a more detailed examination, we compare the cross-correlations of the hp-filtered variables from the two model analyses (“unconstrained” and “constrained”) to those generated from a Bayesian vector autoregression (BVAR) with two lags, using HP-filtered data, see figure 3. Once again, the differences between the unconstrained and constrained predictions are remarkably small to each other, and overall reasonably close to those of the BVAR model, with some obvious differences.

To find a difference between the predictions from the constrained and the unconstrained prior, one needs to look elsewhere. We repeat the exercise above, though this
Figure 3: Predicted Cross-Correlations the DSGEs and the BVAR(2) based on 1200 draws from the corresponding posterior.

Figure 4: Predicted standard deviations of the DSGEs and the BVAR(2) based on 1200 draws from the corresponding posterior.
time fitting a BVAR(2) to the variables used in the estimation of the DSGE model, i.e. the (log) growth rate of output, the (log) growth rate of consumption, the log investment-to-output ratio, log labor and the log of the risk free rate, and comparing the resulting posterior distribution over the unconditional moments to those of the two model analysis. Figure 4 shows the results. While the two model results are fairly similar for the (log) growth rate of output, the (log) growth rate of consumption, log labor and the log of the risk free rate and overall differ from the BVAR(2) results for the (log) growth rate of output and the log of the risk free rate, differences between the two model analyses arise for the log investment-to-output ratio. While the unconstrained estimation predicts values close to those of the BVAR, the constrained estimation predicts a considerably larger standard deviation. One may interpret this as turning short-run shocks into long-run risks. This may point to an interesting connection to the long-run risk literature for Epstein-Zin preferences, see Bansal and Yaron (2004) and Hansen et al. (2008).

3.7 Model comparison

Our procedure allows a formal comparison of the constrained and the unconstrained estimate. We use the Modified Harmonic Mean estimator by Geweke (1999) to calculate the marginal data density of each model. The log marginal data density for the unconstrained model is 3439.74 and 3437.86 for the constrained model. This difference in the marginal data density which implies posterior probabilities of 0.87 vs. 0.13 and a posterior odds ratio of 6.6 in favor of the model estimated with the unconstrained prior. The difference strikes as as remarkably small, and confirms the similarities between both model analyses exhibited in figures 3 and 4. For comparison, we follow Smets and Wouters (2007) and also calculate the log marginal data density for the BVAR(2) model which is 3569.77. The 130 log-point difference indicates that the BVAR(2) fits quite a bit better than the model, though this may be rather unsurprising, given the simplicity of the DSGE model. Indeed, we view these numbers as tolerably close, confirming the impressions from figure 3 that both estimated DSGE models make reasonably similar predictions as the BVAR model about the second moments. In sum, our method has achieved in shifting the model parameters in order to match the Sharpe ratio without giving up too much on the macroeconomic implications, which are tolerable overall to begin with, given the simplicity of the model.

4 Application 2: Smets-Wouters (2007)

The Smets-Wouters model of Smets and Wouters (2007) has become a benchmark medium-scale New Keynesian model, prior to the financial crisis in 2008. Versions of it were in widespread use, in particular at central banks. It still often provides a starting point for post-crisis DSGE models. Given this widespread use of this model,
we think of it as a natural choice to apply our method. We first investigate a single constraint on the Sharpe ratio and next investigate a two-dimensional constraint on both the Sharpe ratio and the slope of the yield curve.

The Smets-Wouters features a representative household with nonseparable utility in consumption and leisure and an external consumption habit. The households own the capital stock, choose the level of capital utilization and pay the investment adjustment costs. Firms produce output with capital and labor. Prices and wages are Calvo sticky. The monetary authority follows a Taylor rule in inflation, output gap and the change in output gap. Seven disturbances drive the model: the technology shock, the risk premium or equity premium shock, the government expenditure shock, the investment adjustment cost shock, the price and wage markup shocks, and the shock to monetary policy. For the details, the reader is referred to Smets and Wouters (2007).

4.1 Constraining the Sharpe Ratio

To apply our method and to highlight characteristics, we try to stay as close as possible with the original code and data files provided by the authors. Nevertheless, it is necessary to extend and to adjust some issues to make the application feasible and also comparable to the former one. In the supplemental appendix, we provide a list of equations for a complete comparison. The set of log-linearized equations is just slightly extended. In particular, we add the asset pricing equations (19) and (20), which were substituted out in the original code. Finally, we add equation (29) which defines log excess returns which goes along with adding the $\Psi$ to the list of parameters and adding $\epsilon_Q$ to the list of exogenous shocks. Next to these changes, we solve the model around a “quasi”-stochastic steady state. More precisely, we correct the deterministic steady state by the conditional second moments from equations (19) and (20). We call this “quasi”-stochastic steady state because we correct just for risk related to the asset pricing equations but not for risk related to, e.g., the price setting equations. This also allows to use a more diffuse initial prior than Smets and Wouters (2007) did. Similarly to our first application, we are not restricting the discount factor $\beta$ to be smaller then one but rather ensure a positive time preference via the mean of the real risk free rate. Hence, we avoid the risk-free rate puzzle by construction and solve recursively for the implied discount factor. In particular, we assume that the expression $100(\bar{R}^f - 1)$ is Gamma distributed with mean 1 and standard deviation 0.5 and calculate the discount factor recursively as

$$\beta = \frac{1}{\bar{R}^f \exp (\sigma_m^2/2) (1 + \gamma/100)^{-\sigma_c}}.$$ 

---

8Data and Dynare code are available on [http://www.aeaweb.org/aer/data/june07/20041254_data.zip](http://www.aeaweb.org/aer/data/june07/20041254_data.zip).
where $\sigma_c$ is the consumption preference parameter in the Smets-Wouters model and $\gamma$ is the quarterly growth rate of the economy. As in our main application, we assume that the prior of $\sigma_c$ is uniformly distributed between 1 and 200. The prior distributions of the remaining parameters are the same as in Smets-Wouters. Finally, we extend the original data set by Smets-Wouters with the time series for excess returns as described in Section 3.2. To keep our estimation results comparable to the original Smets-Wouters paper, we use the same time episode for estimation. The Sharpe ratio in the data for this time is around 0.15, and thus somewhat smaller than the longer time episode used for estimation in section 3. In order to avoid drastic changes to its macroeconomic performance as well as to avoid numerical maximization problems, we soften our Sharpe ratio constraint by imposing a mean of 0.075 instead of 0.15 as in the data. That way, the model will only go half the way of accounting for the observed risk premium: our interest will be in how much the parameters and the macroeconomic implications change.

Figure 5 shows the prior, unconstrained as well as constrained prior, and the posterior distributions for selected key parameters. There are similarities to the findings from our first application. As should be expected, the distribution for the power utility parameter $\sigma_c$ is now centered around a value corresponding to high relative risk aversion. This goes along with a more tight constrained prior distribution for the higher values of $\sigma_c$. The Frisch elasticity of labor supply $1/\sigma_l$ decreases and nominal rigidities increase when imposing the constraint. These effects are similar to those observed in section 3. Interestingly and also similar to the first application, the parameter for habit formation $h$ becomes smaller in both constrained estimations. While this parameter is often mentioned as a key parameter to explain asset pricing facts, our estimation results suggest a smaller importance, especially compared to the curvature parameters, i.e. $\sigma_c$. The estimated standard deviations of the preference shock becomes smaller in the constrained model, because of the higher volatility of the pricing kernel due to, e.g., higher risk aversion. There are some further, but minor differences, concerning the remaining parameters. The technology shock becomes more persistent in the constrained model (see Table in the supplemental appendix), which is again an interesting connection to the long-run risk literature.

Table 2 shows the implied asset pricing facts of the model for both estimation approaches. We are successful in matching the Sharpe ratio constraint: as a reminder, it was set at half of the observed value for this episode. Moreover, the conditional second moments of the risk free rate are similar for both estimations and comparable to the data. Additionally, the moments of the return of capital for both estimations are comparable with each other. As before, the high risk aversion parameter $\sigma_c$ delivers the large conditional volatility of the stochastic pricing kernel $\sigma_M$, which in turn is needed to obtain the Sharpe ratio constraint.

However, the posterior odds ratio is $3.7e^{24}$ to one in favor of the model estimated with the unconstrained prior, compared with 6.6 for the previous application, see

---

9We present detailed estimations results for all parameters in the supplemental appendix.
Figure 5: Prior and posterior distribution for selected parameters of the Smets and Wouters (2007) model. The solid grey and black lines correspond to the unconstrained prior and posterior distribution respectively. The dashed grey and black lines correspond to the contrained prior and posterior distribution respectively.
Figure 6: Predicted Cross-Correlations (HP-filtered) of the DSGEs based on 1200 draws from the corresponding posterior.

![Graph showing cross-correlations](image)

Table 2: Implied quarterly asset pricing facts by the estimated Smets-Wouters models. All values in percent with the exception of the Sharpe ratio.

<table>
<thead>
<tr>
<th></th>
<th>Unconstrained Prior</th>
<th>Constrained Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Posterior</td>
<td>Posterior</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>5%</td>
</tr>
<tr>
<td>s.d. risk-free return $\sigma_{R^f}$</td>
<td>0.24</td>
<td>0.22</td>
</tr>
<tr>
<td>s.d. return on capital $\sigma_{R^k}$</td>
<td>1.76</td>
<td>1.52</td>
</tr>
<tr>
<td>s.d. return on capital $\sigma_{R^q}$</td>
<td>8.12</td>
<td>7.35</td>
</tr>
<tr>
<td>s.d. pricing kernel $\sigma_M$</td>
<td>2.13</td>
<td>1.80</td>
</tr>
<tr>
<td>Risk premium $-\sigma_{MR^k}$</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Sharpe ratio $(R^k) \times 100$ $\omega$</td>
<td>1.77</td>
<td>1.48</td>
</tr>
</tbody>
</table>
subsection 3.7. The Bayesian posterior odds ratio therefore makes the constrained model look non-credible from the unconstrained perspective. In order to illustrate the difficulties of the Smets-Wouters model to jointly explain both asset pricing and macroeconomic facts, we plot selected predictions of second moments for both estimations in Figure 6. In particular, we show the cross-correlation of output, consumption, and investment with inflation respectively based on 1200 parameter vector draws from the posterior distribution. While one prominent result of the original Smets-Wouters model is the success in accounting for the cross-correlations between output and inflation, our results show that this characteristic diminishes by using the constrained prior.

The application to the Smets-Wouters model furthermore shows the limitations of our procedure: while it tries hard to push this model to explain both asset pricing facts and macroeconomic facts, larger gaps between the quantitative model results and the observations emerge. Indeed, trying to push the Smets-Wouters model to fully explain the observed Sharpe ratio rather than just half of it runs into the boundary problem for the constrained prior, as described more generally in subsection 2.2. We therefore followed the general recommendation of subsection 2.2 and pursued the more modest goal of achieving half the observed Sharpe ratio.

4.2 A two-dimensional constraint

As an illustration, we show how to impose a two-dimensional constraint and to calculate its implications. For the Smets-Wouters model just presented, a reasonable extension is to constrain both the bond premium as well as the Sharpe ratio. Like the Sharpe ratio, the upward sloping yield curve is a stylized fact on asset markets which is difficult to account for in standard DSGE models. In this regard, a continuously growing literature has started with the work by Backus et al. (1989) and den Haan (1995) who have postulated this so called bond-premium puzzle.

We calculate the 10-year nominal zero-coupon bond premium over a one-quarter nominal risk-free return. The return of a zero-coupon bond with \( k \) periods maturity left is analogously priced via the Euler equation,

\[
1 = E_t [M_{t,t+k}] R_t^{(k)},
\]

where \( M_{t,t+k} \) is the nominal pricing kernel and \( R_t^{(k)} \) is the safe return after \( k \) periods in \( t+k \). Then, the quarterly bond premium, \( bp_t \), is calculated as \( \sqrt[4]{R_t^{(k)} - R_t^{(1)}} \), which in steady state can be simplified to

\[
bp = \frac{1}{k} \cdot \frac{\sigma_M^{(k)2}}{2} - \frac{\sigma_M^{(1)2}}{2}.
\]

For the unconstrained SW model we find a posterior mean for \( bp \) of approximately –0.85 basis points, which implies a slightly downward sloping yield curve. This find-
Figure 7: Examples for potential distributions regarding application of a multidimensional constraint for 10-year nominal bond premium and Sharpe ratio.
ing is in line with the literature investigating similar class of models. Given our application, let \( \omega = [sr, bp] \) be a vector including the Sharpe ratio \( sr \), and bond premium \( bp \) in per cent. Taking draws from the unconstrained posterior the Figure 7(a) illustrates the corresponding joint histogram. We approximate the joint posterior \( g(\omega | X) \) with a multivariate normal distribution, \( g(\omega) \sim (\bar{\mu}_M, \sum_M) \) as illustrated in Figure 7(b). To approximate the distribution \( f(\omega) \) we use the bond premium data by Gürkaynak et al. (2007) and the excess returns of the S&P 500 as described before.\(^{10}\) In particular, we estimate a VAR(1) model consisting of the normalized excess returns and the demeaned bond premium data to calculate \( f(\omega) \sim (\bar{\mu}_D, \sum_D) \).\(^{11}\) Moreover, to ensure a proper prior, we impose some \( f(\omega) \), which is considerably tighter than would be justified on estimation uncertainty of the joint Sharpe ratio and bond premium process alone, see figure 7(c).

The figure furthermore illustrates the difference between the small negative bond premium implied by the unconstrained SW model and the bond premium of 47 basis points in the data. Our dual constraint estimation procedure only manages to push the bond premium into slightly positive territory, with a posterior mode of 0.62 basis points, and thus rather far from the desired target of 47 basis points. Further experimentation did not result in a full resolution of accounting for both the bond premium and the Sharpe ratio, as desired. Our procedure runs into the boundary problem described more generally in subsection 2.2. This finding illustrates the general remarks in subsection 2.2 that our econometric approach is not a “cure-all”. It may well be that adding features like long-run inflation risk, see Rudebusch and Swanson (2012) and the discussion there, or Epstein-Zin preferences can provide a helpful way out, if one wishes to pursue a modification of the Smets-Wouters model which also accounts for asset market features, but these matters are beyond the scope of the present paper.

5 Conclusion

We have presented a novel Bayesian method for estimating dynamic stochastic general equilibrium (DSGE) models subject to the constraining the posterior distribution of the implied Sharpe ratio. After presenting our methodology generally, we have applied it to two models. The first is a DSGE model with habit formation in consumption and leisure, real wage rigidities and capital adjustment costs, provided by Uhlig (2007) and extended to feature additional shocks. We show that the estimation subject to the constraint on the Sharpe ratio produces a quantitative model with both reasonable asset-pricing as well as business-cycle implications and with tolerable posterior odds compared to the unconstrained model, thus offering more hope than the somewhat

\(^{10}\)Due to limited data availability, we use a data for the 10 year nominal zero-coupon yields from 1971:Q1 onward.

\(^{11}\)See for example Lüttheohl (2007, p. 84) for estimation and asymptotic properties of the sample mean for a mean-adjusted VAR process.
more pessimistic message in Rudebusch and Swanson (2008, 2012) regarding habit-formation models. The particular parameterization delivered by our procedure might have been hard to find otherwise.

The second application is to the Smets-Wouters model in Smets and Wouters (2007), extended with the appropriate asset pricing equations. Now, the constrained estimator yields a parameterization, which is highly unlikely from the perspective of the unconstrained estimator. The discrepancies in macroeconomic implications are now more substantial. The results here are akin to a glass “half-full” or “half-empty”: one can either tolerate the compromise found and celebrate it as progress towards also accounting for asset market features, or point to a more fundamental inability of the model to account for them. It would seem more reasonable to think about useful modifications of that model that could help with the latter.

The broader lesson here is this. If one wishes to construct a DSGE model for the purpose of, say, analyzing monetary policy or fiscal policy, and wishes that model to account for asset pricing features such as the Sharpe ratio, it is worth introducing features that it will help to do so. Our procedure is certainly not a “cure-all”. Rather, our methodology is useful for finding a good parameterization, when the model has a chance to succeed in principle, and for finding the least bad parameterization, when it does not.
References


