Forming Priors for DSGE Models

(and How It Affects the Assessment of Nominal Rigidities)

Marco Del Negro
Federal Reserve Bank of Atlanta

Frank Schorfheide*
University of Pennsylvania
CEPR and NBER

September 27, 2006

*Correspondence: Marco Del Negro: Research Department, Federal Reserve Bank of Atlanta, 1000 Peachtree St NE, Atlanta GA 30309-4470. Email: marcodelnegro@frbatlanta.org. Frank Schorfheide: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu. We thank Simon Potter, Frank Smets, as well as participants at the Bank of Finland Conference on "Practical Issues in DSGE Modeling at Central Banks", the 7th EABCN Workshop on "Estimation and Empirical Validation of Structural Models for Business Cycle Analysis", the St. Louis Fed, and the Board of Governors for helpful comments and suggestions. Schorfheide gratefully acknowledges financial support from the Alfred P. Sloan Foundation and the National Science Foundation. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

Abstract

In Bayesian analysis of dynamic stochastic general equilibrium (DSGE) prior distributions for some of the taste-and-technology parameters can be obtained from microeconometric or pre-sample evidence, but it is difficult to elicit priors for the parameters that govern the law of motion of unobservable exogenous processes. Moreover, since it is challenging to formulate beliefs about the correlation of parameters, most researchers assume that all model parameters are independent of each other. We provide a simple method of constructing prior distributions for (a subset of) DSGE model parameters from beliefs about the moments of the endogenous variables. We use our approach to investigate the importance of nominal rigidities and show how the specification of prior distributions affects our assessment of the relative importance of different frictions.

JEL CLASSIFICATION: C32

KEY WORDS: Bayesian Analysis, DSGE Models, Dummy Observations Priors, Model Comparisons, Prior Elicitation

1 Introduction

A growing literature uses Bayesian methods to estimate and evaluate dynamic stochastic general equilibrium (DSGE) models. Moreover, central banks are starting to employ estimated DSGE models for forecasting and policy analysis.[1] Of particular interest is the question what real and nominal frictions have to be included in the DSGE model to capture the salient features of macroeconomic time series. Several approaches are available to answer this questions: a comparison of impulse responses computed from DSGE models and a structural vector autoregression, e.g., Christiano, Eichenbaum, and Evans (2005) or Del Negro, Schorfheide, Smets, and Wouters (2006); an assessment of how far actual sample moments lie in the tails of predictive distributions from DSGE models, e.g., Canova (1994); a comparison of different DSGE model specifications based on their in-sample fit (adjusted for model complexity) or pseudo-out-of-sample fit, e.g. Smets and Wouters (2003), and Rabanal and Rubio-Ramirez (2005). In a Bayesian framework, prior distributions for the DSGE model parameters play an important role for model comparisons. The contribution of this paper is to provide an easily implementable method to elicit prior distributions for DSGE model parameters from beliefs about sample moments of observable variables. This method is then applied to study the role of nominal rigidities in a New Keynesian DSGE model with both nominal and real frictions.

Prior distributions either reflect subjective opinions or summarize information derived from data sets not included in the estimation sample. The latter case is essentially equivalent to simplifying the likelihood function for a larger set of observations that would be too complicated to model directly. For instance, when pre-sample information is used to construct a prior, the tacit assumption is that the structure of the economy could have changed prior to the beginning of the estimation sample, e.g., a drop in volatility of the macroeconomic aggregates and a potentially more active monetary policy since the early 1980s in the U.S or harmonized monetary policy in the Euro Area starting in the late 1990s. Alternatively, data definitions in the pre-sample and the estimation sample period could be different. Priors for parameters that determine labor supply elasticities, mark-ups, frequencies of price changes, and capital adjustment costs are often quantified based on evidence from household or firm-level data sets which makes the specification of a joint likelihood function to cumbersome. As discussed for instance in Chang, Gomes, and Schorfheide (2002)

[1] Some of the literature on Bayesian estimation of DSGE model is reviewed in An and Schorfheide (2006). A December 2005 conference on "DSGE Modeling at Policymaking Institutions" held at the Federal Reserve Board provides a good overview of the state of DSGE Modeling at central banks around the world.

the prior distribution provides a useful device for incorporating micro-level information in the estimation of a aggregate time series model.

There are three aspects of the prior specification that this paper aims to improve upon. First, researchers typically assume that all DSGE model parameters are independent. This assumption is made purely to simplify the analysis and has the drawback that the resulting joint prior distribution assigns non-negligible probability mass to regions of the parameter space where the model is quite unreasonable. Second, since most of the exogenous shock processes are latent, it is difficult to quantify beliefs about their volatilities and autocorrelations. Hence, informally researchers often choose priors that ensure that the model is not inconsistent with the autocovariance patterns observed in the actual sample or a pre-sample. In practice, this amounts to simulating the prior predictive distribution for important sample moments and checking that the prior does not place little or no mass in a neighborhood of important features of the data. The approach of eliciting priors based on beliefs about predictive densities associated with an econometric model dates back at least to Kadane, Dickey, Winkler, Smith, and Peters (1980). Our proposed method will automate the elicitation of priors for the parameters of the exogenous shock processes based on views about reasonable magnitudes for sample moments of observables.

Third, after having specified a prior distribution for the parameters of a benchmark model, researchers often use the same prior distribution for alternative model specifications when assessing the relative importance of various model features. But identical parameterizations of the exogenous shock processes potentially generate very different dynamics across model specifications and hence the use of a common prior for all models can implicitly penalize some specifications and favor others. Starting point of our proposed method are views about sample characteristics of observables, which will be the same across different DSGE model specifications. However, these beliefs will induce model-specific priors for the actual parameters.

Our method for constructing a prior distribution can be summarized as follows. We partition the vector of DSGE model parameters into two components: a sub-vector for which we can elicit prior distributions directly and a sub-vector for which we elicit a prior distribution based on the implied predictive distribution of the DSGE model for the observables. We use a vector autoregression to derive a quasi-likelihood function for the DSGE model, represent the prior views about the sample moments of observables as dummy observations (or sufficient statistics for these dummy observations), and plug these dummy observations into the quasi-likelihood function. The quasi-likelihood function is then interpreted as a prior density for a sub-vector of the DSGE model parameters. We refer to our prior as dummy observations prior.

In the empirical application, we investigate the importance of nominal rigidities both under a standard prior and our proposed dummy observation prior. We document to what extent the assessment of the relative importance of different frictions is sensitive to the choice of prior. We find that models with and without nominal wage rigidities can both explain the persistence of inflation, especially when the latter are endowed with our proposed dummy observations priors. Flexible wage models are rejected, however, because they cannot quite reproduce the persistence in the labor share, a commonly used measure of marginal costs. Overall, sticky prices are much more important than sticky wages in describing the dynamics in the data. We also find that the evidence for dynamic indexation in the Phillips Curve, which generates an additional lagged inflation term, becomes rather tenuous once we use a prior that places all models considered on a similar footing.

The remainder of this paper is organized as follows. Section 2 provides two simple example that illustrate that a naive choice of prior distributions can distort Bayesian posterior odds for competing models. As an alternative, we consider a prior that is derived from beliefs about predictive distributions, using a change-of-variable argument. Unless the naive prior truly summarizes the prior information about model parameters, the change-of-variable prior can help sharpen inference based on model odds and avoid misleading results in the presence of identification problems. It is very difficult to construct the change-of-variable prior for DSGE models. Hence, we are introducing our dummy observations prior in Section 3. While the dummy observations prior inherits some of the desirable properties of the change-of-variable prior, it is much easier to use in practice. We subsequently apply the dummy observations prior to a New Keynesian DSGE model, described in Section 4. Section 5 summarizes our empirical findings and Section 6 concludes.

2 Priors and Model Comparisons in Two Examples

In a Bayesian framework the likelihood function of an econometric model is re-weighted by a prior to obtain a posterior distribution for the model parameters. In the estimation of DSGE models prior distributions play an important role, see for instance the discussions in An and Schorfheide (2006) and Lubik and Schorfheide (2006). The priors used in empirical applications are typically quite informative and down-weigh regions of the parameter space that are at odds with pre-sample information or other observations available to the researcher that are not contained in the estimation sample. The priors often add curvature to a likelihood function that is (nearly) flat in some dimensions of the parameter space and therefore strongly influence the shape of the posterior distribution. In principle, priors can be gleaned from personal introspection and reflect beliefs about the validity and quantitative implications of economic theories, but often they are based on some empirical observations.[2]

As discussed in the Introduction, the *standard* choice of priors in the empirical literature on the estimation and evaluation of DSGE models has two shortcomings. First, the independence assumption potentially leads to a prior distribution that assigns a lot of probability mass to regions of the parameter space where the model is quite unreasonable. Second, after having specified a prior distribution for the parameters of a benchmark model, researchers often use the same prior distribution for alternative model specifications, when assessing the relative importance of various model features. However, identical parameterizations of the exogenous shock processes potentially generate very different dynamics across model specifications. We will illustrate these shortcomings and their consequences for model comparisons in two simple examples.

Example 1: Consider a simple location model, denoted by $\mathcal{M}_1$, of the form

$$\mathcal{M}_1: \quad y_t = \theta + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1) \tag{1}$$

with the following prior distribution for θ:

$$\theta \sim \mathcal{N}(\underline{\mu}, \lambda^2).$$

In addition to $\mathcal{M}_1$ we consider a second model, $\mathcal{M}_2$, that allows for serial correlation in y_t

$$\mathcal{M}_2: \quad y_t = \theta_1 y_{t-1} + \theta_2 + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0,1). \tag{2}$$

Prior beliefs about the autocorrelation coefficient are summarized by $\theta_1 \sim \mathcal{U}[0,1]$. A comparison of the fit of $\mathcal{M}_1$ and $\mathcal{M}_2$ can be used to assess the importance of serial correlation. We will explore two different prior distributions for $\mathcal{M}_2$.

Since both θ in $\mathcal{M}_1$ and θ_2 in $\mathcal{M}_2$ can be interpreted as intercepts of a regression function we could use the same prior distribution for the two coefficients and assume that θ_2 is independent of θ_1:

$$\text{Prior 1}: \quad \theta_1 \sim \mathcal{U}[0,1], \quad \theta_2|\theta_1 \sim \mathcal{N}(\underline{\mu}, \lambda^2). \tag{3}$$

[2] A detailed discussion of prior elicitation techniques can be found in many Bayesian textbooks and the references cited therein, e.g., Bauwens, Lubrano, and Richard (1999, Chapter 4).

Alternatively, we could interpret the prior for $\mathcal{M}_1$ as reflecting the belief that the mean of y_t is normally distributed with mean $\underline{\mu}$ and variance λ^2. According to model $\mathcal{M}_2$ the mean of y_t is given by

$$\mathbb{E}[y_t] = \mu = \frac{\theta_2}{1-\theta_1} \quad \text{if} \quad 0 \leq \theta_1 < 1.$$

Hence, a straightforward change-of-variable argument leads to the following prior:

$$\text{Prior 2}: \quad \theta_1 \sim \mathcal{U}[0,1], \quad \theta_2|\theta_1 \sim \mathcal{N}\Big(\underline{\mu}(1-\theta_1), \lambda^2(1-\theta_1)^2\Big). \tag{4}$$

Thus, under Prior 2 the two parameters of the AR(1) model are not independent anymore. The closer θ_1 is to one, the smaller the mean and variance of θ_2.

The top panels of Figure 1 depict draws from the implicit distribution of the population mean and autocorrelation of y_t for model $\mathcal{M}_2$ under Priors 1 and 2. By construction, the mean of y_t is independent of the autocorrelation under Prior 2, whereas Prior 1 implies that the distribution of μ becomes more and more diffuse as θ_1 approaches 1. Hence, relative to model $\mathcal{M}_1$, Prior 1 for model $\mathcal{M}_2$ places much more mass on parameterizations that imply a very large (in absolute value) mean of y_t.

In the bottom panels of Figure 1 we show draws from the marginal distribution of two observations, y_1 and y_2, under the two priors for $\mathcal{M}_2$. Moreover, we also display draws generated from the marginal distribution of $\mathcal{M}_1$. These marginal distributions are important for model comparisons based on posterior odds. According to Bayes Theorem, model odds are updated as follows:

$$\frac{\pi_{T,1}}{\pi_{T,2}} = \frac{\pi_{0,1}}{\pi_{0,2}} \frac{p(y_1,\ldots,y_T|\mathcal{M}_1)}{p(y_1,\ldots,y_T|\mathcal{M}_2)}, \tag{5}$$

where $p(y_1,\ldots,y_T|\mathcal{M}_i)$ is the marginal likelihood (or data density) for model $\mathcal{M}_i$. Under Prior 1, the marginal data density for $\mathcal{M}_2$ is much more diffuse than under Prior 2, as it assigns considerable mass to very large and very small values of y_t. As a consequence, for small or intermediate values of y_t the posterior odds will tend to favor model $\mathcal{M}_1$, even in presence of a positive correlation between y_1 and y_2. Vice versa, the more concentrated marginal data density under Prior 2, will generate more decisive odds against $\mathcal{M}_1$ if there is a positive correlation between y_1 and y_2. ■

Example 2: Consider the following two rational expectations models adopted from Lubik and Schorfheide (2006). Model $\mathcal{M}_1$ is given by:

$$\mathcal{M}_1: \quad y_t = \frac{1}{\alpha}\mathbb{E}_t[y_{t+1}] + u_t, \quad u_t = \rho_1 u_{t-1} + \epsilon_t, \quad \epsilon_t \sim iid(0,\sigma^2). \tag{6}$$

Thus, the exogenous driving process u_t is serially correlated. Under model $\mathcal{M}_2$ the process u_t is uncorrelated, but the lagged endogenous variable appears on the right-hand-side:

$$\mathcal{M}_2: \quad y_t = \frac{1}{\alpha}I\!\!E_t[y_{t+1}] + \rho_2 y_{t-1} + u_t, \quad u_t = \epsilon_t, \quad \epsilon_t \sim iid(0, \sigma^2). \tag{7}$$

The two models vaguely resemble New Keynesian Phillips curves, in which y_t corresponds to inflation and u_t to marginal costs, which are treated as latent variable in this example. If we restrict the parameters to values for which there exists a unique (stable) rational expectations solution, we obtain the following reduced-form laws of motion. Under the model $\mathcal{M}_1$

$$\mathcal{M}_1: \quad y_t = \rho_1 y_{t-1} + \frac{1}{1 - \rho_1/\alpha}\epsilon_t, \tag{8}$$

whereas $\mathcal{M}_2$ implies that

$$\mathcal{M}_2: \quad y_t = \frac{1}{2}(\alpha - \sqrt{\alpha^2 - 4\rho_2\alpha})y_{t-1} + \frac{2\alpha}{\alpha + \sqrt{\alpha^2 - 4\rho_2\alpha}}\epsilon_t. \tag{9}$$

By setting $\rho_1 = \frac{1}{2}(\alpha - \sqrt{\alpha^2 - 4\rho_2\alpha})$ it is straightforward to verify that there exists a range of parameters for which $\mathcal{M}_1$ and $\mathcal{M}_2$ are observationally equivalent. Although $\mathcal{M}_1$ and $\mathcal{M}_2$ will generate identical reduced form forecasts, the effect of changes in α on the law of motion of y_t is different in the two specifications.

We begin by specifying a prior for the coefficients of $\mathcal{M}_1$. For $i = 1, 2$ let $\theta_{(i)} = [\alpha, \rho_i, \sigma]'$ and $\Theta^D_{(i)}$ be the region of the parameter space for which the rational expectations difference equation has a unique stable solution.[3] In particular, we assume that

$$p(\theta_{(1)}|\mathcal{M}_1) \propto \tilde{p}_\alpha(\alpha)\tilde{p}_\rho(\rho_1)\tilde{p}_\sigma(\sigma)\{\theta \in \Theta^D_{(1)}\}, \tag{10}$$

where $\{\theta \in \Theta^D_{(1)}\}$ is the indicator function that is one if θ lies in the determinacy region of the parameter space and is zero otherwise. The densities $\tilde{p}_{(\cdot)}(\cdot)$ are given in Table 1.

For model $\mathcal{M}_2$ we consider two priors. Prior 1 is taken directly from $\mathcal{M}_1$ without taking into account that the parameters ρ_1 and ρ_2 in the two models have very different effects on the reduced-form dynamics:

$$\text{Prior 1}: \quad p(\theta_{(2)}|\mathcal{M}_2) \propto \tilde{p}_\alpha(\alpha)\tilde{p}_\rho(\rho_2)\tilde{p}_\sigma(\sigma)\{\theta \in \Theta^D_{(2)}\}, \tag{11}$$

where the $\tilde{p}(\cdot)$ densities are the same as for $\mathcal{M}_1$. The top left panel of Figure 2 depicts draws from the implicit prior distribution of the population autocorrelation and standard deviation for y_t under the two models. The bottom left panel shows draws from the marginal

[3] We assume that $\alpha, \sigma \in \mathbb{R}^+$ and $\rho_i \in [0, 1)$. To ensure determinacy in $\mathcal{M}_1$ we require $\alpha > 1$. To guarantee determinacy in $\mathcal{M}_2$ we require $\alpha^2 \geq 4\rho_2\alpha$, $|\alpha - \sqrt{\alpha^2 - 4\rho_2\alpha}| < 2$, and $|\alpha + \sqrt{\alpha^2 - 4\rho_2\alpha}| > 2$.

distribution of two observations y_1 and y_2. It is evident from the plots that using the same prior for the coefficients of the two models generates quite different implications for the observables. Under $\mathcal{M}_2$, y_t is much more persistent than under $\mathcal{M}_1$ and the marginal data density is more spread out. Hence, depending on the realizations of y_t the posterior odds will signal strong evidence in favor of one of the two models.

In model $\mathcal{M}_1$ the prior for ρ_1 essentially captures beliefs about the persistence of y_t. As in Example 1, we will now construct and alternative prior for model $\mathcal{M}_2$ based on beliefs[4] about the autocorrelation of y_t:

$$\frac{1}{2}(\alpha - \sqrt{\alpha^2 - 4\rho_2\alpha}) \sim \text{Beta}(0.5, 0.05).$$

A change-of-variable argument leads to

$$\begin{aligned} \text{Prior 2}: \quad p(\theta_{(2)}|\mathcal{M}_2) \quad &\propto \quad \tilde{p}_\alpha(\alpha)\tilde{p}_\rho\left(\frac{1}{2}(\alpha - \sqrt{\alpha^2 - 4\rho_2\alpha})\right)\tilde{p}_\sigma(\sigma)\{\theta \in \Theta^D_{(2)}\} \\ &\times\left|2\alpha(\alpha^2 - 4\rho_2\alpha)^{-1/2}\right|. \end{aligned} \tag{12}$$

The last term in (12) represents the Jacobian for the parameter transformation and generates *a priori* dependence between α and ρ_2. Since under the transformation $\rho_1 = \frac{1}{2}(\alpha - \sqrt{\alpha^2 - 4\rho_2\alpha})$ the reduced forms associated with models $\mathcal{M}_1$ and $\mathcal{M}_2$ are identical, the two right panels of Figure 2 indicate that the prior distribution of the population moments as well as the marginal distribution of y_1 and y_2 obtained from models $\mathcal{M}_1$ and $\mathcal{M}_2$ are now identical. As a consequence, posterior odds will be equal to prior odds regardless of the realization of the y_t's. In this simple example that can be solved analytically it is straightforward to see that ρ_1 and ρ_2 are different parameters, and have different implications for persistence in the two models. In larger scale DSGE models the persistence of the exogenous processes also has different implications for the reduced form across models, yet that is harder to see since these models cannot be solved analytically. ■

In both examples we compare two types of priors. The first prior is constructed from beliefs about the actual model parameters and assumes that these parameters are independent, as is standard practice in the DSGE model literature. The second prior is derived from beliefs about certain sample characteristics of the endogenous variable y_t. A change-of-variable argument is then used to map a prior formulated on the mean (Example 1) or the autocorrelation (Example 2) of y_t into a prior for the actual model parameters. Under the change-of-variable prior the model parameters are *a priori* dependent. Unless a researcher has direct believes about the model parameters itself, the change-of-variable prior

[4] As in Table 1, the Beta distribution is parameterized in terms of means and standard deviations.

can help sharpen inference based on posterior model odds and avoid misleading results in the presence of identification problems.

Unfortunately, it is difficult to implement change-of-variable priors in the context of DSGE models. The mapping between structural and reduced-form parameters is highly nonlinear and can only be explored numerically, which makes the computation of the Jacobian matrix associated with the parameter transformation impractical. In the next section, we are proposing an alternative approach, based on the use of dummy observations, that allows researchers to incorporate beliefs about reduced-form characteristics of endogenous variables into the construction of a prior distribution, without having to explicitly derive a Jacobian matrix.

3 Dummy Observation Priors for DSGE Models

It is well known that it is possible to interpret the parameters of a natural conjugate prior as the sufficient statistics of a hypothetical sample. A prior constructed from such a hypothetical sample is called a dummy observations prior and the posterior can then be computed based on a mixed sample of actual and hypothetical observations, e.g., Theil and Goldberger (1961). Dummy observations are frequently used to construct prior distributions for vector autoregressions, for instance to represent a version of the so-called Minnesota prior (Doan, Litterman, and Sims, 1984) or to tilt the VAR estimates toward restrictions implied by a DSGE model (Del Negro and Schorfheide, 2004). In case of the Minnesota prior, the researcher typically specifies a sample of dummy observations, whereas for the DSGE model prior in Del Negro and Schorfheide (2004) the structural model is used to generate only the sufficient statistics for the hypothetical sample. We will subsequently propose a dummy observations prior for the DSGE model parameters that inherits some of the desirable properties of the change-of-variable prior discussed in the previous sections and that can be used to overcome the shortcomings of the standard prior distributions used in the DSGE model literature.

3.1 Prior Specification

As in the examples of Section 2, we want to express our beliefs in terms of some simple statistics: means, variances, autocorrelations, *et cetera.* Due to the state-space structure of the DSGE models, low-dimensional sufficient statistics are typically not available. Hence, instead of constructing the dummy observation prior from the actual likelihood function of the DSGE model, we are deriving it based on a quasi-likelihood function for which sufficient statistics are readily available. More specifically, we are using the likelihood function associated with a p-th order vector autoregression:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma), \tag{13}$$

where y_t is an $n \times 1$ vector of observables. Let x_t be the $k \times 1$ vector $[1, y'_{t-1}, \ldots, y'_{t-p}]'$ and re-write the VAR as linear regression model

$$y'_t = x'_t \Phi + u'_t. \tag{14}$$

To relate the DSGE model parameters θ to the VAR parameters Φ, Σ, we assume that the observables have been transformed such that the vector y_t is covariance stationary according to the DSGE model. $\Gamma^D_{YY}(\theta)$, $\Gamma^D_{YX}(\theta)$ and $\Gamma^D_{XX}(\theta)$ denote the population autocovariances $I\!E^D_\theta[y_t y'_t]$, $I\!E^D_\theta[y_t x'_t]$, and $I\!E^D_\theta[x_t x'_t]$, which are calculated from a DSGE model conditional on a particular parameterization θ. We then define a VAR approximation of the DSGE model through the population least-squares regression:

$$\Phi_D(\theta) = [\Gamma^D_{XX}(\theta)]^{-1}\Gamma^D_{XY}(\theta), \quad \Sigma_D(\theta) = \Gamma^D_{YY}(\theta) - \Gamma^D_{YX}(\theta)[\Gamma^D_{XX}(\theta)]^{-1}\Gamma^D_{XY}(\theta). \tag{15}$$

In the multivariate Gaussian linear regression model (14) the sufficient statistics for a set of dummy observations $\{y^*_t, x^*_t\}_{t=1}^{T^*}$ are given by $\sum y^*_t y^{*'}_t$, $\sum y^*_t x^{*'}_t$, and $\sum x^*_t x^{*'}_t$, which we will write as $T^*\Gamma^*_{YY}$, $T^*\Gamma^*_{YX}$, and $T^*\Gamma^*_{XX}$, respectively. Our dummy observations prior for the DSGE model parameters is based on the quasi-likelihood function (premultiplied by $|\Sigma_D(\theta)|^{-(n+1)/2}$)

$$\begin{aligned} \mathcal{L}(\theta|\Gamma, T^*) \quad &= \quad |\Sigma_D(\theta)|^{-(T^*+n+1)/2} \\ &\quad \times \exp\left\{-\frac{T^*}{2} tr\left[\Sigma_D(\theta)^{-1}(\Gamma^*_{YY} - 2\Phi_D(\theta)\Gamma^*_{XY} + \Phi'_D(\theta)\Gamma^*_{XX}\Phi_D(\theta)\right]\right\}, \end{aligned} \tag{16}$$

where the autocovariance matrices $\Gamma^* = \{\Gamma^*_{YY}, \Gamma^*_{XY}, \Gamma^*_{XX}\}$ are either constructed from introspection, a pre-sample of actual observations, or an alternative candidate model. The quasi-likelihood (16) is small at values of θ for which the DSGE model implied autocovariances strongly differ from the Γ^*'s. The parameter T^* captures the precision of our beliefs: The larger T^*, the sharper the peak of $\mathcal{L}(\theta|\Gamma^*, T^*)$.

We proceed by decomposing the vector of DSGE model parameters into two components: $\theta = [\theta'_1, \theta_2]'$. θ_1 collects the parameters for which we can elicit prior distributions directly, say, based on micro-econometric and other quantitative evidence not obtained from the estimation sample. θ_2 is a sub-vector of parameters for which we elicit a prior distribution indirectly by specifying beliefs on autocovariance matrices Γ^*.

A natural approach is to specify a marginal prior distribution for θ_1, denoted by $p(\theta_1)$, and use quasi-likelihood function to generate a conditional prior of θ_2 given θ_1.

$$p_*(\theta_1, \theta_2|\Gamma^*, T^*) = \underbrace{c_1(\theta_1|\Gamma^*, T^*)\mathcal{L}(\theta_1, \theta_2|\Gamma^*, T^*)\pi(\theta_2)}_{p_*(\theta_2|\theta_1, \Gamma^*, T^*)} p(\theta_1), \tag{17}$$

where $c_1(\theta_1|\Gamma^*, T^*)$ is chosen such that

$$\frac{1}{c_1(\theta_1|\Gamma^*, T^*)} = \int \mathcal{L}(\theta_1, \theta_2|\Gamma^*, T^*)\pi(\theta_2)d\theta_2 \quad \text{for all} \quad \theta_1.$$

The disadvantage of the prior defined in (17) that it depends on a normalization constant that typically cannot be calculated analytically. Hence, (17) would be very difficult to implement in practice.

For the empirical work presented in Section 5 we consider the following simplification.

$$p(\theta_1, \theta_2|\Gamma^*, T^*) = \underbrace{c_1(\underline{\theta}_1|\Gamma^*, T^*)\mathcal{L}(\underline{\theta}_1, \theta_2|\Gamma^*, T^*)\pi(\theta_2)}_{p(\theta_2|\Gamma^*, T^*)} p(\theta_1), \tag{18}$$

where $c_1(\underline{\theta}_1|\Gamma^*, T^*)$ is chosen such that

$$\frac{1}{c_1(\underline{\theta}_1|\Gamma^*, T^*)} = \int \mathcal{L}(\underline{\theta}_1, \theta_2|\Gamma^*, T^*)\pi(\theta_2)d\theta_2.$$

This simplification leads to a prior in which θ_1 and θ_2 are independent and the normalization constant does not depend on θ_1. If the prior is used in model comparisons, T^* has to be sufficiently large to ensure that $p(\theta_2|\Gamma^*, T^*)$ (or $p(\theta_2|\theta_1, \Gamma^*, T^*)$) is proper even if $\pi(\theta_2)$ is not.

The rationale for using a VAR approximation of the DSGE model is that we can express our beliefs in the form of the sufficient statistics Γ^*_{YY}, Γ^*_{YX}, and Γ^*_{XX}, that is, in terms of variances and autocorrelations. If we were to use the likelihood of the DSGE model we would have to specify an actual time series for the dummy observations. Furthermore, the following result provides a basis for using the VAR approximation. Let $I\!E^{VAR}_{\Phi,\Sigma}[\cdot]$ denote expectations under the probability distribution generated by the VAR approximation of the DSGE model. Then:

$$I\!E^{VAR}_{\Phi^D(\theta),\Sigma^D(\theta)}[y_t y_t'] = I\!E^D_\theta[y_t y_t'],\ I\!E^{VAR}_{\Phi^D(\theta),\Sigma^D(\theta)}[y_t x_t'] = I\!E^D_\theta[y_t x_t'], I\!E^{VAR}_{\Phi^D(\theta),\Sigma^D(\theta)}[x_t x_t'] = I\!E^D_\theta[x_t x_t']. \tag{19}$$

The result, which can be verified by straightforward matrix manipulations, shows that the VAR(p) approximation and the DSGE model have the same implications for the moments of interest. This means that if we apply our procedure with $T^* = \infty$ and then generate data from the DSGE model, the expectation for the relevant moments is going to be exactly Γ^*_{YY}, Γ^*_{YX}, and Γ^*_{XX}.

3.2 Example 1 – Revisited

Suppose we would like to incorporate the belief that the mean of y_t is approximately $\underline{\mu}$ using the dummy observation approach. Let $\Gamma_{YY} = \underline{\mu}^2 + 1$ and $\Gamma_{YX} = \underline{\mu}$. The restriction function that relates the parameters of the AR(1) model to the location model is given by

$$\phi_D(\theta) = \frac{\theta_2}{1-\theta_1}.$$

Hence, we obtain

$$\mathcal{L}(\theta|\Gamma^*, T^*) = (2\pi)^{-T^*/2} \exp\left\{ -\frac{T^*}{2} \left(1 + \underline{\mu}^2 - 2\underline{\mu}\frac{\theta_2}{1-\theta_1} + \frac{\theta_2^2}{(1-\theta_1)^2} \right) \right\} \tag{20}$$

Combining the quasi-likelihood function with the initial prior distribution

$$p(\theta_1, \theta_2) \propto \{0 \leq \theta_1 < 1\}$$

yields

$$p(\theta_1) \propto |1-\theta_1| \quad \text{and} \quad \theta_2|\theta_1 \sim \mathcal{N}\left(\underline{\mu}(1-\theta_1), \frac{1}{T^*}(1-\theta_1)^2 \right), \tag{21}$$

which corresponds to (17). The larger T^* the smaller the variance of the conditional distribution of θ_2 given θ_1. Hence the conditional distribution of θ_2 given θ_1 under the change-of-parameter approach is identical to (21) if we set $\lambda = 1/\sqrt{T^*}$. Notice, however, that the marginal distribution of θ_1 is not affected by T^*. If we simplify the dummy observations prior by conditioning on a particular value $\underline{\theta}_1$ as in (18) we obtain

$$\theta_1 \sim \mathcal{U}[0,1] \quad \text{and} \quad \theta_2|\theta_1 \sim \mathcal{N}\left(\underline{\mu}(1-\underline{\theta}_1), \frac{1}{T^*}(1-\underline{\theta}_1)^2 \right). \tag{22}$$

3.3 Implementation

In order to implement the proposed dummy observations prior for the sub-vector θ_2 a number of choices have to be made. The parameter T^* scales the prior distribution: the larger T^* the more concentrated the prior. The Γ^* matrices summarize the information contained in the dummy observations. Suppose that $p = 0$. Then Γ^* only contains information about the mean and the covariance matrix of y_t and hence the researcher only uses beliefs about location and scale to construct a prior for θ_2. If $p = 1$ and x_t is composed only of y_{t-1} and the autocovariance matrices in Γ^* are specified in terms of deviations of y_t from its mean, then the prior for θ_2 will indirectly be based on beliefs about the covariance matrix of y_t and first-order autocorrelations. This will be the case considered in the empirical implementation.

The numerical values for the Γ^* matrix could be obtained from introspection, calculated from a pre-sample or based on data from a different country, or they could be obtained from a benchmark model. For instance, suppose the goal is to estimate a DSGE model for the Euro Area. A synthetic Euro-area data set is only available from the mid 1970s onwards. Moreover, the harmonization of monetary policy across the Euro-area countries did not start until the early 1990s. Our method allows the researcher to incorporate sample autocovariances, say computed from a subset of the Euro-area countries, through the dummy observation prior into the estimation of the DSGE model, without having to impose the structure of the likelihood function on this pre-sample. Similarly, it has been well documented that across many countries the volatility of the major macroeconomic time series has dropped substantially in the early 1980s. Without introducing time-varying shock volatilities into the DSGE model as in Justiniano and Primiceri (2006), our method allows us to calculate autocovariance estimates based on a pre-1980 sample, scale them to reflect the reduction in volatility post-1980, and use the scaled autocovariances to obtain the dummy observation prior.

In the empirical application we are going to form the Γ^* on the basis of a pre-sample, without any adjustment, to keep the exercise as simple as possible and avoid any arbitrary choice. In this case, if p were sufficiently large, our approach is similar to the use of a training-sample prior for the estimation of the DSGE model parameters. Moreover, posterior odds comparisons would become essentially predictive likelihood comparisons.[5] An advantage of our approach is that it can be used even in absence of a pre-sample, as long as the researcher

[5]The predictive likelihood associated with a Bayes model is of the form

$$p(y_{\tau+1},\ldots,y_T|y_1,\ldots,y_\tau) = \int p(y_{\tau+1},\ldots,y_T|\theta,y_1,\ldots,y_\tau)p(\theta|y_1,\ldots,y_\tau)d\theta,$$

where $p(\theta|y_1,\ldots,y_\tau)$ is the posterior density of θ given $y_1,\ldots,y_\tau$. and $p(y_{\tau+1},\ldots,y_T|\theta,y_1,\ldots,y_\tau)$ is the predictive density for the "future" observations given the parameter θ. The predictive likelihood is closely related to the marginal likelihood:

$$p(y_{\tau+1},\ldots,y_T|y_1,\ldots,y_\tau) = \frac{p(y_1,\ldots,y_T)}{p(y_1,\ldots,y_\tau)}.$$

has beliefs on the autocovariance matrices Γ^*. Even if she uses a pre-sample, she can easily summarize her confidence in the pre-sample moments via the hyperparameter T^*.

Once Γ^* and T^* have been determined, Markov Chain Monte Carlo techniques can be used to implement Bayesian computations. Due to the nonlinearities of $\Phi_*(\theta)$ and $\Sigma_*(\theta)$ it is not possible to generate draws from the prior distribution of θ directly. In the application in Section 5 we use a random-walk Metropolis algorithm, described in detail for instance in Del Negro and Schorfheide (2004) and An and Schorfheide (2006), to generate draws from the prior distribution. This algorithm only requires us to be able to numerically evaluate the prior density (18) up to the normalization constant. Based on the output of the Metropolis algorithm, Geweke's (1999) modified harmonic mean estimator can be used to calculate the normalization constant $c_1(\underline{\theta}|\Gamma^*, T^*)$. The same algorithms can be used to obtain draws from the posterior distribution. The only modification that is necessary is to replace (18) by the product of prior density and likelihood function of the DSGE model. For a linearized DSGE model the likelihood function can be evaluated with the Kalman filter.

4 The DSGE Model

This section briefly describes the DSGE model, which is taken from Del Negro, Schorfheide, Smets, and Wouters (2006). The model is based on work of Smets and Wouters (2003) and Christiano, Eichenbaum, and Evans (2005) and contains a large number of nominal and real frictions. To make this paper self-contained we subsequently describe the structure of the model economy and the decision problems of the agents in the economy. The exposition closely follows Section 2 of Del Negro, Schorfheide, Smets, and Wouters (2006).

4.1 Final Goods Producers

The final good Y_t is a composite made of a continuum of intermediate goods $Y_t(i)$, indexed by $i \in [0, 1]$:

$$Y_t = \left[\int_0^1 Y_t(i)^{\frac{1}{1+\lambda_{f,t}}} di \right]^{1+\lambda_{f,t}} \tag{23}$$

where $\lambda_{f,t} \in (0, \infty)$ follows the exogenous process:

$$\ln \lambda_{f,t} = (1 - \rho_{\lambda_f}) \ln \lambda_f + \rho_{\lambda_f} \ln \lambda_{f,t-1} + \sigma_{\lambda,f} \epsilon_{\lambda,t}, \tag{24}$$

where $\epsilon_{\lambda,t}$ is an exogenous shock with unit variance that in equilibrium affects the mark-up over marginal costs. The final goods producers are perfectly competitive firms that buy intermediate goods, combine them to the final product Y_t, and resell the final good to consumers. The firms maximize profits

$$P_t Y_t - \int P_t(i) Y_t(i) di$$

subject to (23). Here P_t denotes the price of the final good and $P_t(i)$ is the price of intermediate good i. From their first order conditions and the zero-profit condition we obtain that:

$$Y_t(i) = \left(\frac{P_t(i)}{P_t}\right)^{-\frac{1+\lambda_{f,t}}{\lambda_{f,t}}} Y_t \quad \text{and} \quad P_t = \left[\int_0^1 P_t(i)^{-\frac{1}{\lambda_{f,t}}} di\right]^{-\lambda_{f,t}}. \tag{25}$$

4.2 Intermediate goods producers

Good i is made using the technology:

$$Y_t(i) = Z_t^{1-\alpha} K_t(i)^\alpha L_t(i)^{1-\alpha}, \tag{26}$$

where the technology shock Z_t (common across all firms) follows a unit root process. We define technology growth $z_t = \log(Z_t/Z_{t-1})$ and assume that z_t follows the autoregressive process:

$$z_t = (1-\rho_z)\gamma + \rho_z z_{t-1} + \sigma_z \epsilon_{z,t}. \tag{27}$$

All firms face the same prices for their labor and capital inputs. Hence profit maximization implies that the capital-labor ratio is the same for all firms:

$$\frac{K_t(i)}{L_t(i)} = \frac{\alpha}{1-\alpha} \frac{W_t}{R_t^k}, \tag{28}$$

where W_t is the nominal wage and R_t^k is the rental rate of capital. Following Calvo (1983), we assume that in every period a fraction of firms ζ_p is unable to re-optimize their prices $P_t(i)$. These firms adjust their prices mechanically according to

$$P_t(i) = (\pi_{t-1})^{\iota_p} (\pi_*)^{1-\iota_p}, \tag{29}$$

where $\pi_t = P_t/P_{t-1}$, π_* is the steady state inflation rate of the final good, and $\iota \in [0,1]$. Those firms that are able to re-optimize prices choose the price level $\tilde{P}_t(i)$ that solves:

$$\begin{aligned} &\max_{\tilde{P}_t(i)} \mathbb{E}_t \left[\sum_{s=0}^\infty \zeta_p^s \beta^s \Xi_{t+s}^p \left(\tilde{P}_t(i) \left(\Pi_{l=1}^s \pi_{t+l-1}^{\iota_p} \pi_*^{1-\iota_p}\right) - MC_{t+s}\right) Y_{t+s}(i)\right] \\ &\text{s.t. } Y_{t+s}(i) = \left(\frac{\tilde{P}_t(i)\left(\Pi_{l=1}^s \pi_{t+l-1}^{\iota_p} \pi_*^{1-\iota_p}\right)}{P_{t+s}}\right)^{-\frac{1+\lambda_{f,t}}{\lambda_{f,t}}} Y_{t+s}, \; MC_{t+s} = \frac{\alpha^{-\alpha} W_{t+s}^{1-\alpha} R_{t+s}^{k\ \alpha}}{(1-\alpha)^{(1-\alpha)} Z_{t+s}^{1-\alpha}}, \end{aligned} \tag{30}$$

where $\beta^s \Xi^p_{t+s}$ is today's value of a future dollar for the consumers and MC_t reflects marginal costs. We consider only the symmetric equilibrium where all firms will choose the same $\tilde{P}_t(i)$. Hence from (25) we obtain the following law of motion for the aggregate price level:

$$P_t = \left[(1-\zeta_p)\tilde{P}_t^{-\frac{1}{\lambda_{f,t}}} + \zeta_p \left(\pi_{t-1}^{\iota_p} \pi_*^{1-\iota_p} P_{t-1} \right)^{-\frac{1}{\lambda_{f,t}}} \right]^{-\lambda_{f,t}}. \tag{31}$$

4.3 Labor Packers

There is a continuum of households, indexed by $j \in [0,1]$, each supplying a differentiated form of labor, $L(j)$. The labor packers are perfectly competitive firms that hire labor from the households and combine it into labor services L_t that are offered to the intermediate goods producers:

$$L_t = \left[\int_0^1 L_t(j)^{\frac{1}{1+\lambda_w}} di \right]^{1+\lambda_w}, \tag{32}$$

where $\lambda_w \in (0,\infty)$ is a fixed parameter. From first-order and zero-profit conditions of the labor packers we obtain the labor demand function and an expression for the price of aggregated labor services L_t:

$$(a) \quad L_t(j) = \left(\frac{W_t(j)}{W_t} \right)^{-\frac{1+\lambda_w}{\lambda_w}} L_t \quad \text{and} \quad (b) \quad W_t = \left[\int_0^1 W_t(j)^{-\frac{1}{\lambda_w}} di \right]^{-\lambda_w}. \tag{33}$$

4.4 Households

The objective function for household j is given by:

$$I\!\!E_t \sum_{s=0}^{\infty} \beta^s \left[\log(C_{t+s}(j) - hC_{t+s-1}(j)) - \frac{\phi_{t+s}}{1+\nu_l} L_{t+s}(j)^{1+\nu_l} + \frac{\chi}{1-\nu_m} \left(\frac{M_{t+s}(j)}{Z_{t+s}P_{t+s}} \right)^{1-\nu_m} \right] \tag{34}$$

where $C_t(j)$ is consumption, $L_t(j)$ is labor supply, and $M_t(j)$ is money holdings. Household's preferences display habit-persistence. The exogenous preference shifter ϕ_t, which affects the marginal utility of leisure, is common to all households and evolves as:

$$\ln \phi_t = (1-\rho_\phi) \ln \phi + \rho_\phi \ln \phi_{t-1} + \sigma_\phi \epsilon_{\phi,t}, \tag{35}$$

Real money balances enter the utility function deflated by the (stochastic) trend growth of the economy, so to make real money demand stationary.

The household's budget constraint written in nominal terms is given by:

$$\begin{aligned} & P_{t+s}C_{t+s}(j) + P_{t+s}I_{t+s}(j) + B_{t+s}(j) + M_{t+s}(j) + T_{t+s}(j) \\ \leq \quad & R_{t+s-1}B_{t+s-1}(j) + M_{t+s-1}(j) + A_{t+s-1}(j) + \Pi_{t+s} + W_{t+s}(j)L_{t+s}(j) \\ & + \left(R^k_{t+s} u_{t+s}(j) \bar{K}_{t+s-1}(j) - P_{t+s} a(u_{t+s}(j)) \bar{K}_{t+s-1}(j) \right), \end{aligned} \tag{36}$$

where $I_t(j)$ is investment, $B_t(j)$ are holdings of government bonds, $T_t(j)$ are lump-sum taxes (or subsidies), R_t is the gross nominal interest rate paid on government bonds, $A_t(j)$ is the net cash inflow from participating in state-contingent securities, Π_t is the per-capita profit the household gets from owning firms (households pool their firm shares, and they all receive the same profit), and $W_t(j)$ is the nominal wage earned by household j. The term within parenthesis represents the return to owning $\bar{K}_t(j)$ units of capital. Households choose the utilization rate of their own capital, $u_t(j)$. Households rent to firms in period t an amount of effective capital equal to:

$$K_t(j) = u_t(j)\bar{K}_{t-1}(j), \tag{37}$$

and receive $R_t^k u_t(j)\bar{K}_{t-1}(j)$ in return. They however have to pay a cost of utilization in terms of the consumption good equal to $a(u_t(j))\bar{K}_{t-1}(j)$. Households accumulate capital according to the equation:

$$\bar{K}_t(j) = (1-\delta)\bar{K}_{t-1}(j) + \mu\left(1 - S\left(\frac{I_t(j)}{I_{t-1}(j)}\right)\right) I_t(j), \tag{38}$$

where δ is the rate of depreciation, and $S(\cdot)$ is the cost of adjusting investment, with $S(e^\gamma) = 0$, and $S''(\cdot) > 0$.

The households' wage setting is subject to nominal rigidities á la Calvo (1983). In each period a fraction ζ_w of households is unable to re-adjust wages. For these households, the wage $W_t(j)$ will increase at a geometrically weighted average of the steady state rate increase in wages (equal to steady state inflation π_* times the steady state growth rate of the economy e^γ) and of last period's inflation times last period's productivity $(\pi_{t-1}e^{z_{t-1}})$. The weights are $1-\iota_w$ and ι_w, respectively. Those households that are able to re-optimize their wage solve the problem:

$$\begin{array}{ll} \max_{\tilde{W}_t(j)} & I\!\!E_t \sum_{s=0}^{\infty} \zeta_w^s \beta^s b_{t+s}\left[-\frac{\phi_{t+s}}{1+\nu_l} L_{t+s}(j)^{1+\nu_l}\right] \\ \text{s.t.} & \text{Eq. (36) for } s = 0, \dots, \infty, \text{ (33a), and} \\ & W_{t+s}(j) = \left(\Pi_{l=1}^{s}(\pi_* e^\gamma)^{1-\iota_w}(\pi_{t+l-1}e^{z_{t+l-1}})^{\iota_w}\right)\tilde{W}_t(j). \end{array} \tag{39}$$

We again consider only the symmetric equilibrium in which all agents solving (39) will choose the same $\tilde{W}_t(j)$. From (33b) it follows that:

$$W_t = [(1-\zeta_w)\tilde{W}_t^{-\frac{1}{\lambda_w}} + \zeta_w((\pi_* e^\gamma)^{1-\iota_w}(\pi_{t-1}e^{z_{t-1}})^{\iota_w} W_{t-1})^{-\frac{1}{\lambda_w}}]^{-\lambda_w}. \tag{40}$$

Finally, we assume there is a complete set of state contingent securities in nominal terms, which implies that the Lagrange multiplier $\Xi_t^p(j)$ associated with (36) must be the same for all households in all periods and across all states of nature. This in turn implies that in equilibrium households will make the same choice of consumption, money demand, investment and capital utilization. Since the amount of leisure will differ across households due to the wage rigidity, separability between labor and consumption in the utility function is key for this result.

4.5 Government Policies

The central bank follows a nominal interest rate rule by adjusting its instrument in response to deviations of inflation and output from their respective target levels:

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_R} \left[\left(\frac{\pi_t}{\pi_*}\right)^{\psi_1} \left(\frac{Y_t}{Y_t^*}\right)^{\psi_2}\right]^{1-\rho_R} e^{\sigma_R \epsilon_{R,t}}, \tag{41}$$

where $\epsilon_{R,t}$ is the monetary policy shock, R^* is the steady state nominal rate, Y_t^* is the target level of output, and the parameter ρ_R determines the degree of interest rate smoothing. We set the target level of output Y_t^* in (41) equal to the trend level of output $Y_t^* = Z_t Y^*$, where Y^* is the steady state of the model expressed in terms of detrended variables. The central bank supplies the money demanded by the household to support the desired nominal interest rate.

The government budget constraint is of the form

$$P_t G_t + R_{t-1} B_{t-1} + M_{t-1} = T_t + M_t + B_t, \tag{42}$$

where T_t are total nominal lump-sum taxes (or subsidies), aggregated across all households. Government spending is given by:

$$G_t = (1 - 1/g_t) Y_t, \tag{43}$$

where g_t follows the exogenous process:

$$\ln g_t = (1-\rho_g)\ln g + \rho_g \ln g_{t-1} + \sigma_g \epsilon_{g,t} \tag{44}$$

4.6 Resource Constraint

The aggregate resource constraint:

$$C_t + I_t + a(u_t)\bar{K}_{t-1} = \frac{1}{g_t} Y_t. \tag{45}$$

can be derived by integrating the budget constraint (36) across households, and combining it with the government budget constraint (42) and the zero profit conditions of both labor packers and final good producers.

4.7 Model Solution

As in Altig, Christiano, Eichenbaum, and Lindé (2004) our model economy evolves along stochastic growth path. Output Y_t, consumption C_t, investment I_t, the real wage W_t/P_t, physical capital $\bar{K}_t$ and effective capital K_t all grow at the rate Z_t. Nominal interest rates R_t, inflation π_t, and hours worked L_t are stationary. The model can be rewritten in terms of detrended variables. We find the steady states for the detrended variables and use the method in Sims (2002) to construct a log-linear approximation of the model around the steady state. All subsequent statements about the DSGE model are statements about its log-linear approximation. We collect all the DSGE model parameters in the vector θ, stack the structural shocks in the vector ϵ_t, and derive a state-space representation for:

$$y_t = [\ln(Y_t/Y_{t-1}), \ln L_t, \ln(W_t L_t/Y_t), \pi_t, R_t]'.$$

5 Assessing the Role of Nominal Rigidities

We will now apply the dummy observations prior proposed in Section 3 to the DSGE model outlined in the previous section. Throughout this section we will fix the following parameters: $\delta = 0.025$, $\lambda_w = 0.3$. We choose the mean of the preference shock, ϕ, such that in steady state each household supplies one unit of labor. Hence, ϕ does not appear in the subsequent definition of θ. Using the notation of Section 3 we will partition the DSGE model parameters as follows:

$$\begin{aligned}
\theta_1 &= [\alpha, \zeta_p, \iota_p, s', h, a'', \nu_l, \zeta_w, \iota_w, r^*, \psi_1, \psi_2, \rho_r, \pi^*, \gamma, \lambda_f, g^*, L^{adj}]' \\
\theta_2 &= [\rho_z, \rho_\phi, \rho_{\lambda_f}, \rho_g, \sigma_z, \sigma_\phi, \sigma_{\lambda_f}, \sigma_g, \sigma_r]'
\end{aligned}$$

The parameter L_{adj} captures the units of measured hours worked. The vector θ_1 contains the subset of parameters for which we will specify a prior distribution directly, whereas θ_2 collects the parameters for which we elicit a prior distribution based on the DSGE model implied predictive distribution for the observables. In our empirical application, θ_2 contains the autocorrelations and the standard deviations of the exogenous shock processes.[6]

The remainder of this section is organized as follows. We briefly describe the composition of the vector of observables, y_t, and the data sources in Section 5.1. In Section 5.2

[6]In principle we could also include the steady state parameters r^*, g^*, π^*, L^{adj}, and α into θ_2 to automate the custom of constructing priors for these parameters based on pre-sample averages.

we describe a *standard* prior distribution[7] for the DSGE model parameters $\theta = [\theta_1', \theta_2']'$. Next, we compare this prior with the *dummy observation* prior for θ_2, computed using the approach described in section 3 The Γ^* matrices used to construct the dummy observations are based on sample autocovariance matrices $\hat{\Gamma}_{YY}$, $\hat{\Gamma}_{YX}$, $\hat{\Gamma}_{XX}$ computed from a pre-sample of observations ranging from QIII:1954 to QIV:1980 to specify Γ^*. Section 5.3 compares the implications of the standard prior to those of the dummy observations prior in the benchmark version of our DSGE model. We introduce flexible wages and prices specifications of the DSGE model in Section 5.4 and ask what effect prior distributions have when it comes to the assessment of nominal rigidities. Finally, section 5.5 studies the presence of indexation in the New-Keynesian Phillips curve.

5.1 The Data

We will subsequently use data in two instances. First, we use a pre-sample from QIII:1954 to QIV:1980 to elicit our beliefs on the moments of the endogenous variables, i.e. to construct the Γ^* matrices for the dummy observations prior. Second, we use a sample of 100 observations on output growth, inflation, interest rates, log hours worked, and the log labor share from QI:1981 to QIV:2005 to compute posterior distributions and marginal likelihood values. Our data are obtained from Haver Analytics (Haver mnemonics are in italics). Real output is obtained by dividing the nominal series (*GDP*) by population 16 years and older (*LN16N*), and deflating using the chained-price GDP deflator (*JGDP*). We compute quarter-to-quarter output growth as log difference of real GDP per capita and multiply the growth rates by 100 to convert them into percentages. Our measure of hours worked is computed by taking total hours worked reported in the National Income and Product Accounts (NIPA), which is at annual frequency, and interpolating it using growth rates computed from hours of all persons in the non-farm business sector (*LXNFH*). We divide hours worked by *LN16N* to convert them into per capita terms. We then take the log of the series multiplied by 100 so that all figures can be interpreted as percentage changes in hours worked. The labor share is computed by dividing total compensation of employees (*YCOMP*) obtained from the NIPA by nominal GDP. We then take the log of the labor share multiplied by 100. Inflation rates are defined as log differences of the GDP deflator and converted into annualized percentages. The nominal rate corresponds to the effective Federal Funds Rate (*FFED*),

[7]The term *standard* does not refer to particular numerical values but rather to the approach of specifying a prior directly for the θ_2 i) assuming independence between the different elements of θ_2, and ii) using the same prior for the different specifications considered.

also in percent.

5.2 Prior Distributions

We begin by specifying a standard prior distribution for the entire vector θ of DSGE model parameters. This prior is summarized in Table 2 and the first four columns of Table 3 and essentially corresponds to the one used in Del Negro, Schorfheide, Smets, and Wouters (2006).

We begin with the description of the marginal distributions for the θ_1 parameters, that is, those parameters for which we specify a prior directly. The priors for the degree of price and wage stickiness, ζ_p and ζ_w, are both centered at 0.6, which implies that firms and households re-optimize their prices and wages on average every two and half quarters. The 90% interval is very wide and encompasses findings in micro-level studies of price adjustments such as Bils and Klenow (2004). The priors for the degree of price and wage indexation, ι_p and ι_w, are nearly uniform over the unit interval. The prior for the adjustment cost parameter s' is consistent with the values that Christiano, Eichenbaum, and Evans (2005) use when matching DSGE impulse response functions to consumption and investment, among other variables, to VAR responses.

The prior for the habit persistence parameter h is centered at 0.7, which is the value used by Boldrin, Christiano, and Fisher (2001). These authors find that $h = 0.7$ enhances the ability of a standard DSGE model to account for key asset market statistics. The prior for a'' implies that in response to a 1% increase in the return to capital, utilization rates rise by 0.1 to 0.3%. These numbers are considerably smaller than the one used by Christiano, Eichenbaum, and Evans (2005). The 90% interval for the prior distribution on ν_l implies that the Frisch labor supply elasticity lies between 0.3 and 1.3, reflecting the micro-level estimates at the lower end, and the estimates of Kimball and Shapiro (2003) and Chang and Kim (2006) at the upper end.

We use a pre-sample of observations from QI:1960 to QI:1974 to choose the prior means for the parameters that determine steady states. The prior mean for the technology growth rate is 2% per year. The annualized steady state inflation rate lies between 0.5 and 5.5% and the prior for the inverse of the discount factor r^* implies a growth adjusted real interest rate of 4% on average. The prior means for the capital share α, the substitution parameter λ_f, and the steady state government share $1 - 1/g$ are chosen to capture the labor share of 0.57, the investment-to-output ratio of 0.24, and the government share of 0.21 in the pre-sample. The distribution for ψ_1 and ψ_2 is approximately centered at Taylor's (1993) values, whereas the smoothing parameter lies in the range from 0.17 to 0.83. Finally, the prior for L_{adj} is chosen based on quarterly per capita hours worked in the pre-sample.

The standard priors for the parameters of the shock processes, θ_2, are obtained as follows. Since we model the level of technology Z_t as a unit root root process, the prior for ρ_z, which measures the serial correlation of technology growth z_t, is centered at 0.4. The priors for ρ_ϕ (preference for leisure), ρ_{λ_f} (price markup shocks), ρ_g (government spending) are fairly diffuse and centered around 0.75. Finally, the priors for the standard deviation parameters are chosen to obtain realistic magnitudes for the implied volatility of the endogenous variables. Under the standard prior we assume that the parameters are *a priori* independent. Also, we follow the common approach of keeping the standard prior unchanged as we consider different DSGE model specifications in Sections 5.4 and 5.5.

As an alternative to the standard prior we consider dummy observations priors based on different choices of T^*. We retain the prior for θ_1 described in Table 2 and use the dummy observations to generate a prior for θ_2. Using the notation of Section 3, we combine the quasi-likelihood function in (18) with an initial prior $\pi(\theta_2)$ that is uniform on $[0,1)$ for the autocorrelation parameters and proportional to $1/\sigma$ for the standard deviation parameters, see column 5 of Table 3. In specifying the autocovariance matrices Γ^* that enter the quasi-likelihood function we choose a value of $p = 1$, that is, we use the pre-sample to form beliefs about the contemporaneous covariance matrix and the first-order auto and cross-correlations for the endogenous variables.

5.3 Standard vs. Dummy Observation Prior in Benchmark Model

We use a random-walk Metropolis algorithm to generate parameter draws from the dummy observations prior and directly sample from the standard prior. Table 3 summarizes prior means and standard deviations for the parameters of the exogenous shock processes in the benchmark model. Under dummy observations prior the technology and preference shock are more volatile. Mark-up and technology shock are slightly more persistent, whereas the autocorrelation of the preference and government spending shocks drops.

One of the motivations for the benchmark prior was to be able to generate correlation between the DSGE model parameters and shift probability mass away from parameter combinations that are empirically implausible. The panels of Figure 3 depict bivariate scatter plots of draws generated from the two prior distributions. The dummy observations prior introduces a strong negative correlation between the autocorrelation and standard deviation parameters associated with the preference and mark-up shock.

Figure 4 shows draws from the prior predictive distribution of the sample standard deviations of output growth, hours worked, the labor share, and inflation. These draws are generated as follows. For a subset of our draws from the prior distributions of θ we simulate samples of 100 observations from the DSGE model and compute sample standard deviations. Under the standard prior the predictive distribution of these sample standard deviations has fat tails. The figure shows many draws in which the standard deviation of inflation exceeds 15, which is extreme given the U.S. post-war experience. Under the dummy observations prior, the probability mass is shifted away from these extreme values and the predictive distribution concentrates in a more plausible range.

5.4 Sticky Prices vs. Sticky Wages

This section discusses how nominal rigidities, sticky prices and wages, affect the model's ability to describe the data. We compare four specifications: i) the *Benchmark* model described in Section 4, ii) the very same model without wage stickiness ($\zeta_w = 0$), which we refer to as the *Flexible Wages* model, iii) the *Flexible Prices* model, which has no price stickiness ($\zeta_p = 0$), and iv) the model without either wage or price stickiness ($\zeta_w = \zeta_p = 0$), called *Flexible Wages and Prices* model. We show how the presence of nominal rigidities changes the models' implications for some important sample moments of the endogenous variables using prior predictive distributions. In turn, we use these prior predictive checks to help explain the model rankings obtained from Bayesian marginal likelihood comparisons. Moreover, we document how the use of the dummy observation prior in place of the standard prior changes the *a priori* model's implications and, as a consequence, the marginal likelihood values. Among others, papers by Rabanal and Rubio-Ramirez (2005), Smets and Wouters (2003) and Christiano, Eichenbaum, and Evans (2005) have addressed the importance of nominal rigidities in DSGE models. Our paper contributes to this literature by assessing the robustness of previous findings to changes in the prior distribution of the DSGE model parameters.

Figure 5 shows the prior predictive distributions for the sample autocorrelations of inflation and the labor share. The top two panels compare the predictions for the Benchmark and the Flexible Wages and Prices models. The middle two panels compare the Benchmark and the Flexible Prices models, while the bottom two panels compare the Benchmark and the Flexible Wages models. The panels on the left use the standard prior while the two panels on the right use the dummy observations prior with $T^* = 10$ dummy observations. In each panel the dark crosses (+) and the lighter circles (◯) represent draws from the Benchmark and the alternative model, respectively. The dark and light lines show the medians of the marginal prior predictive distributions for the two models. The thick gray cross indicates the corresponding sample moments for the actual observations, which are used to compute the marginal likelihood values reported in Table 5.

The comparison of the distribution of the dark crosses between the left-hand-side and right-hand-side panels in Figure 5 shows that under the dummy observations prior the Benchmark model's predictions are more concentrated than under the standard prior. Under both priors the Benchmark model generates large persistence in both inflation and the labor share. Under the standard prior the median inflation and labor share autocorrelation are about 0.86 and 0.92 respectively. On the contrary, for the Flexible Wages and Prices model the autocorrelation of inflation is negative roughly fifty percent of the times under the standard prior (top left panel). Under the dummy observations prior the predicted inflation autocorrelation rises, but is not as high as that predicted by the Benchmark model (top right panel).

Even under the standard prior the Flexible Prices model (left middle panel) is much closer to the Benchmark in terms of its *a priori* predictions than the Flexible Wages and Prices model. The median inflation and labor share autocorrelations are about 0.7 and 0.8, respectively. Under the dummy observations prior (right middle panel) the difference between the Flexible Prices and the Benchmark models' predictions narrows. The marginal distribution for the autocorrelation of the labor share is roughly the same for the two models, but the Flexible Prices model still predicts a slightly lower autocorrelation of inflation than the Benchmark.

For the Flexible Wages model the standard prior implies that the predicted autocorrelation of inflation, while higher than for the Flexible Wages and Prices model, is still lower than for the Benchmark model (bottom left panel). Under the dummy observations prior the differences between the Benchmark and the Flexible Wages predictions for inflation autocorrelation nearly disappear (bottom right panel). Differences in the predictions for the autocorrelation in the labor share remain, however. Under the dummy observations prior the median autocorrelation of the labor share is 0.63 for the Flexible Wages model. The Flexible Wages model cannot generate the degree of persistence in the labor share afforded by the presence of both nominal rigidities.

An interesting feature of Figure 5 is that the two specification with flexible prices can generate persistence in the labor share, even under the standard prior, while the Flexible Wages is not able to do so under either prior. In order to explain this result, in Figure 6 we compare the impulse response functions for both models, computed for parameter values corresponding to the mean of the dummy observations prior. The dashed, dash-and-dotted, and solid lines represent the responses for the Flexible Wages and Prices, the Flexible Prices, and the Flexible Wages models, respectively. If prices are flexible, then movements in the labor share are solely due to the mark-up shock:

$$\widehat{lsh}_t = -\hat{\lambda}_{f,t},$$

where $\widehat{\ }$ denotes log deviations from the steady state. Hence, the persistence of the labor share is directly determined by the autocorrelation of the mark-up shock. Figure 6 shows that the mark-up shock generates a negative correlation between inflation and the labor share. As pointed out by by Galí and Gertler (1999), such a negative correlation is counterfactual. In our sample, the correlation is about 0.4. Hence, both models with flexible prices face a trade-off when trying to simultaneously match the persistence in the labor share and the positive correlation between the labor share and inflation, which has a negative impact on its overall fit. Figure 6 also documents that the inflation responses in the Flexible Wages and Prices model are generally short-lived, which is consistent with the low inflation persistence documented in the top panels of Figure 5.

For the model with flexible prices and sticky wages, inflation responses to the leisure preference shock ϕ_t and to technology shocks are quite persistent, resulting in an autocorrelation of inflation close to the Benchmark's. Yet this models suffers the same trade-off as the Flexible Wages and Prices model in matching cross-correlations of inflation and the labor share.

If wages are flexible and prices are sticky, the four other shocks affect the labor share as well and their impulse responses are far less persistent than that of the mark-up shock. As a consequence the autocorrelation of the labor share decreases relative to the flexible price models, while inflation essentially remains as persistent as in the Benchmark model.

Table 4 summarizes the prior distributions for the shock parameters under the standard and the dummy observation prior. The table shows that the dummy observations prior generates changes in the persistence of the exogenous processes that differ in the three model specifications, highlighting that each model provides a distinct propagation mechanism for the exogenous shocks. The marginal distribution of the shock parameters shifts as follows relative to the standard prior. Across all models the mean of the standard deviation of the technology shock, σ_z, increases by at least a factor of 5. In the Benchmark model the autocorrelation of the mark-up shock, ρ_{λ_f} increases while its volatility drops from 0.38 to 0.19. In the two specifications without price stickiness both the autocorrelation and the volatility of $\lambda_{f,t}$ increase as the shock process determines the law of motion for the labor share. If wages are assumed to be flexible, the persistence and volatility of the government spending shock rises.

Finally, Table 5 summarizes log marginal likelihood ratios relative to the Benchmark model. In the last row of Table 5 we report the log marginal likelihood values for the Benchmark model. The use of the dummy observations prior narrows the gap between the Benchmark and the alternative models, which is consistent with the findings presented in Figure 5. However, log marginal likelihood ratios are large even under the alternative prior. The Flexible Wages model is in part penalized for its inability to reproduce the persistence in the labor share series. The Flexible Prices model overall performs worse than the Flexible Wages model. The reason for this ranking possibly lies in its inability to generate both persistent labor share dynamics as well as a strong positive correlation between inflation and the labor share. The specification in which both prices and wages are flexible performs worst.

Direct comparison of our results with existing literature on estimated DSGE models on U.S. data is difficult either because other studies use somewhat different data (or detrend the data as in Rabanal and Rubio-Ramirez 2005), or a different model validation approach (impulse responses to a monetary shock as in Christiano, Eichenbaum, and Evans 2005). Still, it is important to put our findings in perspective. Christiano, Eichenbaum, and Evans 2005 find that wage rigidities are more important than price rigidities – indeed, that price rigidities seem to matter very little. Our results on the importance of price rigidities, both in absolute terms and relative to wage stickiness, are quite different. Partly that reflects the inclusion of the labor share among our observables: as shown in Figure 6 the labor share does not move at all after a monetary shock in absence of price stickiness, and this implication may be counterfactual. In general, in our analysis the fit of flexible price models hinges on whether mark-up shocks, the only ones that can move the labor share, can reproduce the dynamic correlations between the labor share and the other endogenous variables. These shocks are not considered in their analysis.[8] Rabanal and Rubio-Ramirez (2005) do not

[8]There is an ongoing debate on the merits of limited information approaches, such as the one pursued by Christiano, Eichenbaum, and Evans, and the full information approach used here. Valid arguments can

assess the fit of a model without price stickiness, hence we cannot make any inference from their study on the importance of sticky prices. Also, they use a model without capital accumulation. They do find that wage stickiness is very important – the difference in log-marginal likelihoods for the models with and without sticky wages is about 147, much larger than what we find. An interesting question is to what extent the lack of real side frictions in their model enhances the importance of wage stickiness. Finally, Smets and Wouters (2003b) find that both sticky prices and sticky wages are important, with sticky prices being more important than sticky wages. Relative to their benchmark model, which shares many features with ours, the marginal likelihood drops by 226 and 26 log points when the ζ_p and ζ_w parameters are set equal to .42, respectively. That is, these authors find very large drop in fit by constraining the degree of price and wage stickiness, let alone eliminating it.

5.5 Assessing the Phillips Curve

This section focuses on the specification of the New Keynesian Phillips curve relationship, which for our Benchmark model takes the following log-linear form:

$$\begin{aligned}\widehat{\pi}_t \;&= \frac{(1-\zeta_p\beta)(1-\zeta_p)}{(1+\iota_p\beta)\zeta_p}\left[\widehat{lsh}_t + \frac{\lambda_f}{1+\lambda_f}\widehat{\lambda}_{f,t}\right] \\ &+ \frac{\iota_p}{1+\iota_p\beta}\widehat{\pi}_{t-1} + \frac{\beta}{1+\iota_p\beta}I\!E_t[\widehat{\pi}_{t+1}].\end{aligned} \tag{46}$$

In terms of log deviations from the steady state the labor share $\widehat{lsh}_t$ is identical to the marginal costs. A large body of literature (Eichenbaum and Fisher 2003, Galí and Gertler 1999, Sbordone 2002, Galí, Gertler and Lopez-Salido 2005, Rudd and Whelan 2005, among several others) has investigated whether the lagged inflation term $\widehat{\pi}_{t-1}$ needs to be incorporated in order for the Phillips curve to adequately describe the dynamics of inflation. While much of the literature studies the issue using single equation methods (Lindé, 2005, and Rabanal and Rubio-Ramirez, 2005, are exceptions), we use full information methods. The importance of lagged inflation is determined by the parameter ι_p, which in this model captures what Eichenbaum and Fisher (2003) call dynamic indexation, that is, the extent to which prices for those firms that are not able to re-optimize are indexed by past inflation rather than steady state inflation.

We therefore compare the *Benchmark* model, which allows for partial dynamic indexation for both firms and workers ($\iota_p \in (0,1)$, $\iota_w \in (0,1)$), to the same model with *No Dynamic Indexation* for either firms or workers ($\iota_p = 0$ $\iota_w = 0$). As shown in the Example 2

be made in favor of either approach. Here we simply emphasize how the full information implications of the model drive our results on the importance of price rigidities.

in Section 2, the choice of prior for a comparison of the three specifications is not innocuous: a model that assigns a large coefficient to the lagged inflation term in (46) and imposes a small autocorrelation in the mark-up shock, might generate similar dynamics as a model without indexation and a persistent mark-up shock.

Figure 7 shows the *a priori* implications of the two specifications for two moments that are important for the empirical assessment of the Phillips curve: the persistence of inflation and the correlation between inflation and marginal costs. Galí and Gertler (1999) have argued that the positive correlation found in the data between inflation and the labor share is *prima facie* evidence in support of the Phillips curve. We therefore investigate the prior predictive distribution for these two moments generated by the Benchmark (dark +) and the No-Dynamic-Indexation (light ◯) model. As in Section 5.4, the left panel of Figure 7 is based on the standard prior while the draws depicted in the right panel are obtained from the dummy observations prior with $T^* = 10$. The dark and light lines show medians of the predictive distributions for the two models. The thick gray cross signifies the sample moments computed from the actual U.S. data.

The left panel shows that both model specifications are able to generate inflation persistence under the standard prior, although quantitatively the median autocorrelation for the No-Dynamic-Indexation model (0.72) is lower than for the Benchmark model (0.86). The right two panel shows that under the dummy observations prior the difference between the No-Dynamic-Indexation and the Benchmark model in terms of inflation autocorrelation virtually disappears. The predictive distribution for the correlation of the labor share and inflation is in general fairly diffuse for all models and priors. Draws range approximately from -0.8 to 0.8, indicating that the DSGE model does not generate any sharp predictions with respect to this correlation. The median correlation for the two models is slightly negative (less than -0.2) under the standard prior and about zero for the dummy observations prior. As could be seen from the impulse-response functions in Figure 6, the mark-up shock generates a negative correlation between inflation and the labor share. According to Table 4 the volatility of the mark-up shock drops under the dummy observations prior for both the Benchmark and the No-Dynamic-Indexation model and hence the mark-up shock becomes less important for the co-movement of inflation and the labor share. Log marginal likelihood ratios are for the No-Dynamic-Indexation model, reported in Table 5, show that under the dummy observation prior with $T^* = 10$ the two models are essentially equivalent and the evidence in favor of dynamic indexation has vanished. The results here are in line with the findings of Del Negro, Schorfheide, Smets and Wouters (2006), who show that the evidence from the impulse responses functions comparison between VAR and DSGE models in favor of dynamic indexation is tenuous.

Our finding that the Benchmark model and the No-Dynamic-Indexation model are essentially observationally equivalent seems to be at odds with the single-equation literature on the New-Keynesian Phillips curve, which tends to emphasize the importance of lagged inflation. The seemingly conflicting results can be reconciled by taking a closer look at the role of the mark-up shock. As mentioned previously, in our model marginal costs are identical to the labor share (in terms of log deviations). The mark-up shock in (46) has in general two interpretations. On the one hand, it might capture changes in the degree of monopolistic competition over the business cycle. On the other hand, in reality the labor share might be an imperfect measure of marginal costs and the mark-up shock picks up this misspecification. In both cases an autocorrelated mark-up shock serves as a substitute to dynamic indexation in generating inflation persistence, as it captures the autocorrelation in the residual of the purely forward-looking Phillips curve.

Figure 8 shows draws from the prior ($\bigcirc$) and posterior (+) for the indexation parameter ι_p and the autocorrelation of the markup shock ρ_{λ_f} under the standard and the dummy observations prior. The prior draws of ι_p and ρ_{λ_f} are independent by construction. Under the standard prior the posterior draws have a strong negative correlation. The marginal posterior for ι_p has most of its mass between 0.2 and 0.7. The dummy observation prior, on the other hand, places more mass on higher values of the autocorrelation parameter, as noted above. As a consequence, under this prior the marginal posterior on ι_p concentrates between 0 and 0.3. Since in general equilibrium the mark-up shock determines a large fraction of the labor share dynamics, its estimated autocorrelation tends to be high. Yet, once we have fairly persistent markup shocks, dynamic indexation is no longer needed in the DSGE model. Single-equation generalized method of moments (GMM) estimates ignore the $\lambda_{f,t}$ term in (46), in part because latent variables are difficult to handle in a GMM framework, and therefore find that lagged inflation is important to explain the inflation data.

6 Conclusion

The careful specification of prior distributions is an important task in the Bayesian analysis of DSGE models. Since the priors used in this literature tend to be rather informative, priors do affect posterior parameter estimates as well as posterior model odds. For some

DSGE model parameters we can elicit prior distributions directly, often based on micro-level evidence as it has been done in the literature on calibrated equilibrium models for about two decades now. For other parameters, including those that determine the law of motion of the exogenous shocks, direct elicitation of prior distributions is very difficult. We find it advantageous to elicit priors for these parameters based on beliefs about predictive distributions. As one considers different model specifications, it seems reasonable to hold the beliefs about the predictive distributions constant and implicitly construct a new prior for the parameters of each model specification.

The contribution of this paper is to provide a procedure based on dummy observations and a quasi-likelihood function for the DSGE model that automates the elicitation. We apply our so-called dummy observation prior to a New Keynesian DSGE model and assess the role of various features of the model. We compare a Benchmark specification to versions of the model with flexible prices, flexible wages, and both. While the the use of the dummy observation prior narrows the gap between the model, the Benchmark specification in which both prices and wages remains to be preferred. We also show that once the dummy observation prior is used the small evidence in favor of dynamic indexation that we find under the standard prior completely vanishes.

References

Altig, David, Lawrence Christiano, Martin Eichenbaum, and Jesper Lindé (2004): "Firm-Specific Capital, Nominal Rigidities, and the Business Cycle," *Manuscript*, Northwestern University.

An, Sungbae, and Frank Schorfheide (2006): "Bayesian Analysis of DSGE Models," *Working Paper*, **06-5**, Federal Reserve Bank of Philadelphia.

Bauwens, Luc, Michel Lubrano, and Jean-Francois Richard (1999): *"Bayesian Inference in Dynamic Econometric Models,"* Oxford University Press, Oxford.

Bils, Mark, and Peter Klenow (2004): "Some Evidence on the Importance of Sticky Prices," *Journal of Political Economy*, **112**, 947-985.

Boldrin, Michele, Lawrence Christiano, and Jonas Fisher (2001): "Habit Persistence, Asset Returns, and the Business Cycle," *American Economic Review*, **91**, 149-166.

Calvo, Guillermo (1983): " Staggered Prices in a Utility-Maximizing Framework," *Journal of Monetary Economics*, **12**, 383-398.

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

Chang, Yongsung, and Sun-Bin Kim (2006): "From Individual to Aggregate Labor Supply: A Quantitative Analysis based on Heterogeneous Agent Macroeconomy," *International Economic Review*, **47**, 1-27.

Christiano, Lawrence, Martin Eichenbaum, and Charles Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, **113**, 1-45.

Del Negro, Marco, and Frank Schorfheide (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, **45**, 643-673.

Del Negro, Marco, Frank Schorfheide, Frank Smets, and Rafael Wouters (2006): "On the Fit of New-Keynesian Models," *Journal of Business & Economic Statistics*, with discussions and rejoinder, forthcoming.

Doan, Thomas, Robert Litterman, and Christopher Sims (1984): "Forecasting and Conditional Projections Using Realistic Prior Distributions". *Econometric Reviews*, **3**, 1-100.

Eichenbaum, Martin and Jonas D.M. Fisher (2003): "Evaluating the Calvo Model of Sticky Prices," *Federal Reserve Bank of Chicago Working Paper* 2003-23.

Galí, Jordi, and Mark Gertler (1999): " Inflation Dynamics: A Structural Econometric Analysis," *Journal of Monetary Economics*, **44**, 195-222.

Galí, Jordi, Gertler, Mark, and J. David López-Salido (2005): "Robustness of the Estimates of the Hybrid New Keynesian Phillips Curve," *Journal of Monetary Economics*, **52**, 1107-1118.

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

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

Greenwood, Jeremy, Zvi Hercovitz, and Per Krusell (1998): "Long-Run Implications of Investment-Specific Technological Change," American Economic Review, **87**(3), 342-362.

Justiniano, Alejandro and Giorgio E. Primiceri (2006) "The Time Varying Volatility of Macroeconomic Fluctuations." *NBER Working Paper No. 12022*.

Kadane, J.B., J.M. Dickey, R.L. Winkler, W.S. Smith, and S.C. Peters (1980): "Interactive Elicitation of Opinion for a Normal Linear Model," *Journal of the American Statistical Association*, **75**, 845-854.

Kimball, Miles, and Matthew Shapiro (2003): "Labor Supply: Are the Income and Substitution Effects Both Large or Both Small?" *Manuscript*, University of Michigan.

Kydland Finn, and Edward Prescott (1982): "Time to Build and Aggregate Fluctuations," *Econometrica*, **50**, 1345-70.

Lindé, Jesper (2005): "Estimating New Keynesian Phillips Curves: A Full Information Maximum Likelihood Approach," *Journal of Monetary Economics*, **52**, 1135-1149.

Lubik, Thomas and Frank Schorfheide (2006): "A Bayesian Look at New Open Economy Macroeconomics" in Mark Gertler and Kenneth Rogoff (eds.) *NBER Macroeconomics Annual 2005*, MIT Press, Cambridge, MA.

Rotemberg, Julio, and Michael Woodford (1997): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," *NBER Macroeconomics Annual*, **12**, 297-246.

Rabanal, Pau and Juan Francisco Rubio-Ramirez, (2005): "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach," *Journal of Monetary Economics*, **52**, 1151-1166.

Rudd, Jeremy and Karl Whelan (2005): "New Tests of the New Keynesian Phillips Curve," *Journal of Monetary Economics*, **52**, 1167-1181.

Sbordone, Argia M. (2002): "Prices and Unit Labor Costs: A New Test of Price Stickiness," *Journal of Monetary Economics*, **49**, 265-292.

Sbordone, Argia M. (2005): "Do Expected Future Marginal Costs Drive Inflation Dynamics?," *Journal of Monetary Economics*, **52**, 1183-1197.

Schorfheide, Frank (2000): "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, **15**, S645-670.

Sims, Christopher (2002): "Solving Rational Expectations Models," *Computational Economics*, **20**(1-2), 1-20.

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

Smets, Frank, and Raf Wouters (2003a): "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, **1**, 1123-75.

Smets, Frank, and Raf Wouters (2003b): "Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach," *mimeo*, European Central Bank.

Taylor, John (1993): "Discretion versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, **39**, 195-214.

Theil, Henry and Arthur S. Goldberger (1961): "On Pure and Mixed Estimation in Economics". *International Economic Review*, **2**, 65-78.

Table 1: EXAMPLE 2 – PRIOR DISTRIBUTIONS

Name	Domain	Density	Para (1)	Para (2)
$p_\alpha(\cdot)$	$I\!R^+$	Gamma	2.00	0.10
$p_\rho(\cdot)$	$[0, 1)$	Beta	0.50	0.05
$p_\sigma(\cdot)$	$I\!R^+$	InvGamma	1.00	4.00

Notes: Para (1) and Para (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; s and ν for the Inverse Gamma distribution, where $p_{\mathcal{IG}}(\sigma|\nu, s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. The effective prior is truncated at the boundary of the determinacy region.

Table 2: Prior Distribution for Taste-and-Technology Parameters

	Support	Density	Mean	StdDev	90% LB	90% UB
α	[0,1)	Beta	0.400	0.100	0.234	0.562
ζ_p	[0,1)	Beta	0.600	0.200	0.292	0.935
ι_p	[0,1)	Beta	0.500	0.280	0.061	0.942
s'	$\mathbb{R}^+$	Gamma	4.000	1.500	1.561	6.248
h	[0,1)	Beta	0.700	0.050	0.620	0.782
a''	$\mathbb{R}^+$	Gamma	0.200	0.100	0.049	0.349
ν_l	$\mathbb{R}^+$	Gamma	2.000	0.750	0.784	3.138
ζ_w	[0,1)	Beta	0.600	0.200	0.290	0.937
ι_w	[0,1)	Beta	0.500	0.280	0.057	0.936
r^*	$\mathbb{R}^+$	Gamma	2.000	1.000	0.457	3.473
ψ_1	$\mathbb{R}^+$	Gamma	1.550	0.370	0.990	2.089
ψ_2	$\mathbb{R}^+$	Gamma	0.200	0.100	0.048	0.349
ρ_r	[0,1)	Beta	0.500	0.200	0.168	0.825
π^*	$\mathbb{R}$	Normal	3.000	1.500	0.556	5.435
γ	$\mathbb{R}^+$	Gamma	2.000	1.000	0.475	3.469
λ_f	$\mathbb{R}^+$	Gamma	0.150	0.100	0.010	0.288
g^*	$\mathbb{R}^+$	Gamma	0.300	0.100	0.141	0.457
L^{adj}	$\mathbb{R}$	Normal	252.0	10.00	235.7	268.6

Notes: The prior distributions for the taste-and-technology parameters are identical for both the standard and the dummy observations prior. *StdDev* denotes standard deviation, *LB* and *UB* refer to lower and upper bounds of a 90% credible interval. The following parameters are fixed: $\delta = 0.025$, $\lambda_w = 0.3$, $\mathcal{F} = 0$. We assume that the taste-and-technology parameters are *a priori* independent.

Table 3: Prior for Shock Parameters – Benchmark Model

	Standard Prior			Dummy Obs. Prior		
	Density	Mean	StdDev	Initial	Mean	StdDev
ρ_z	Beta	0.400	0.250	Uniform	0.489	0.129
ρ_ϕ	Beta	0.750	0.250	Uniform	0.692	0.194
ρ_{λ_f}	Beta	0.750	0.250	Uniform	0.843	0.120
ρ_g	Beta	0.750	0.250	Uniform	0.597	0.278
σ_z	InvGamma	0.376	0.194	$1/\sigma_z$	1.549	0.388
σ_ϕ	InvGamma	3.755	1.955	$1/\sigma_\phi$	5.392	2.646
σ_{λ_f}	InvGamma	0.376	0.194	$1/\sigma_{\lambda_f}$	0.191	0.086
σ_g	InvGamma	0.626	0.323	$1/\sigma_g$	0.577	0.204
σ_r	InvGamma	0.250	0.130	$1/\sigma_r$	0.398	0.115

Notes: StdDev denotes standard deviation. The support for the distributions of the auto-correlation (standard deviation) parameters is $[0, 1)$ $(\mathbb{R}^+)$. The column *Initial* refers to the (improper) prior that is used to pre-multiply the quasi-likelihood function for the dummy observations. The results are based on $T^* = 10$.

Table 4: Prior for Shock Parameters - Standard vs Dummy Observation

Prior:	Standard Baseline	Dummy Obs. Baseline	Dummy Obs. Flex Wages & Prices	Dummy Obs. Flex Prices	Dummy Obs. Flex Wages	Dummy Obs. No Dynamic Indexation
ρ_z	0.400 (0.250)	0.489 (0.129)	0.326 (0.122)	0.520 (0.100)	0.332 (0.118)	0.490 (0.125)
ρ_ϕ	0.750 (0.250)	0.692 (0.194)	0.586 (0.338)	0.722 (0.166)	0.769 (0.199)	0.688 (0.204)
ρ_{λ_f}	0.750 (0.250)	0.843 (0.120)	0.884 (0.067)	0.896 (0.066)	0.799 (0.146)	0.872 (0.089)
ρ_g	0.750 (0.250)	0.597 (0.278)	0.922 (0.141)	0.512 (0.286)	0.840 (0.204)	0.625 (0.287)
σ_z	0.376 (0.194)	1.549 (0.388)	1.667 (0.405)	1.703 (0.412)	1.613 (0.371)	1.628 (0.393)
σ_ϕ	3.755 (1.955)	5.392 (2.646)	1.832 (0.918)	5.705 (2.327)	1.901 (0.783)	5.289 (2.837)
σ_{λ_f}	0.376 (0.194)	0.191 (0.086)	0.732 (0.172)	0.720 (0.171)	0.230 (0.084)	0.157 (0.056)
σ_g	0.626 (0.323)	0.577 (0.204)	0.822 (0.320)	0.376 (0.130)	0.789 (0.406)	0.570 (0.241)
σ_r	0.250 (0.130)	0.398 (0.115)	0.414 (0.132)	0.410 (0.098)	0.410 (0.101)	0.398 (0.109)

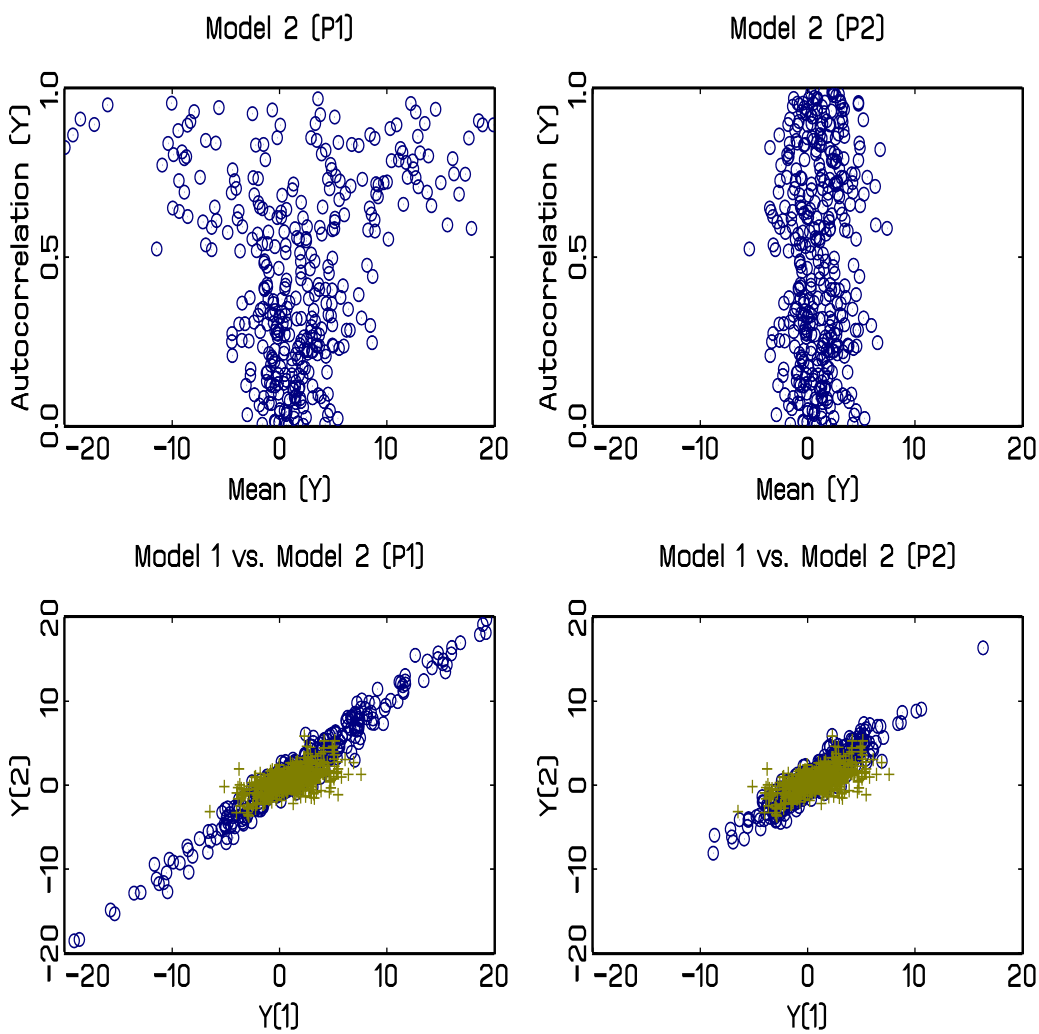
Notes: StdDev denotes standard deviation. The support for the distributions of the autocorrelation (standard deviation) parameters is $[0, 1)$ ($\mathbb{R}^+$). See Table 3 for the marginal densities of the benchmark prior and the (improper) prior that is used to pre-multiply the quasi-likelihood function for the dummy observations. The results are based on $T^* = 10$.

Table 5: Log Marginal Likelihoods $\ln p(Y)$ Relative to Benchmark

Specification	Standard	Dummy Obs. Prior	
	Prior	$T^* = 4$	$T^* = 10$
Flexible Wages and Prices	-65.36	-53.44	-52.76
Flexible Prices	-44.92	-42.77	-38.76
Flexible Wages	-23.23	-13.51	-16.74
No Indexation	-0.63	-0.54	-0.22
Benchmark	-611.95	-611.02	-614.31

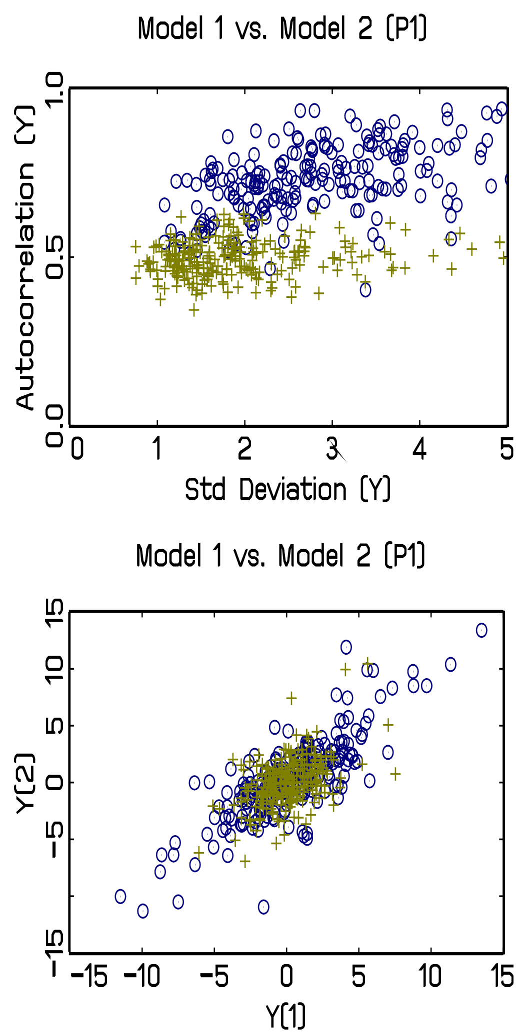
Notes: The marginal likelihoods are computed based on quarterly U.S. data ranging from QI:1981 to QIV:2005. We report $\ln p(Y|\mathcal{M}_0)$ for the Benchmark specification and log marginal likelihood ratios for all other models. Negative entries indicate a deterioration of fit relative to the Benchmark specification.

Figure 1: Example 1 - Moments and Predictive Densities



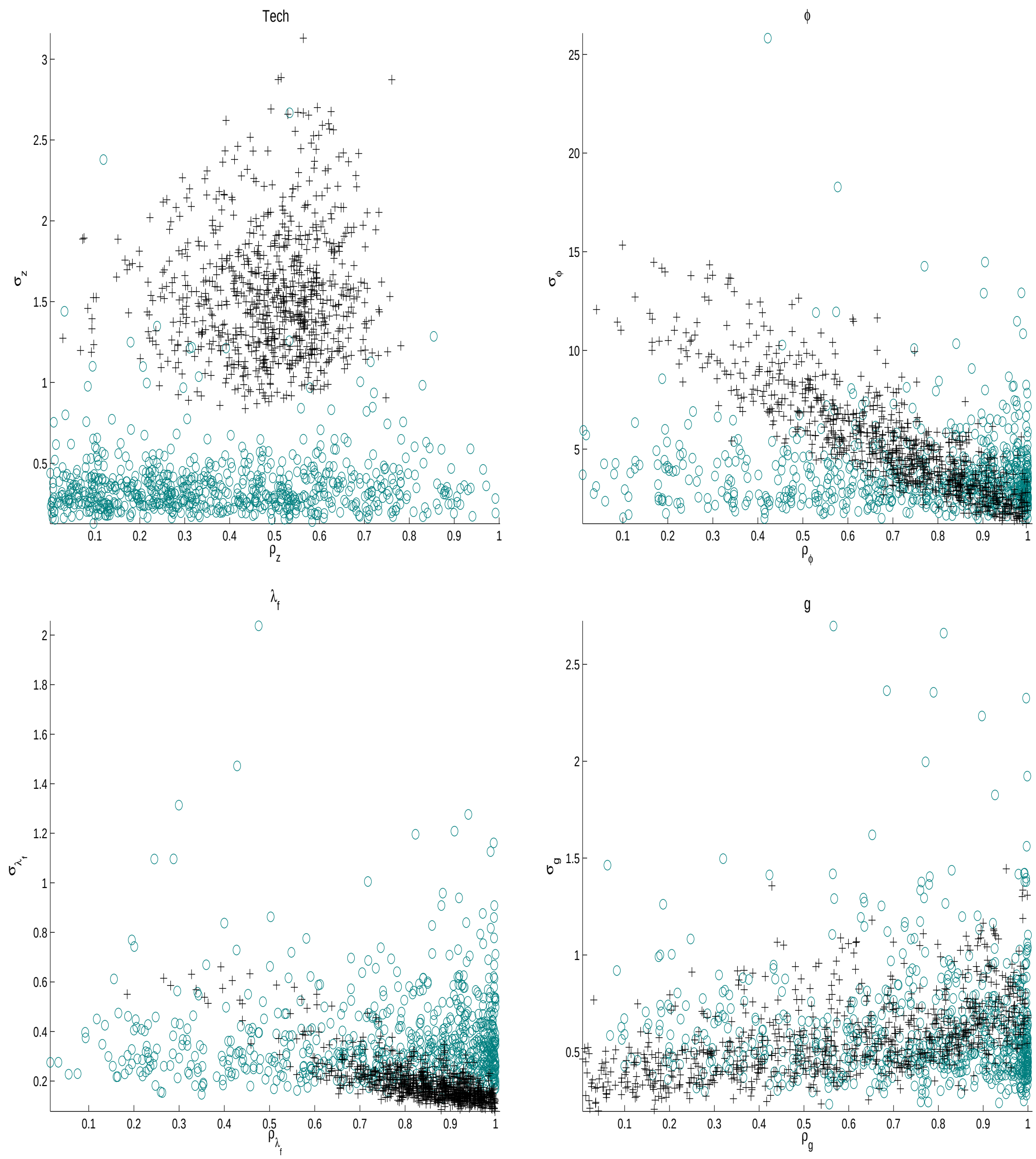
Notes: Top panels depict 400 draws from the implicit prior distribution for mean and autocorrelation of y for Model 2. Bottom panels depict draws from predictive distribution for two observations, y_1 and y_2. Blue circles correspond to Model 2, green crosses to Model 1.

Figure 2: Example 2 - Moments and Predictive Densities



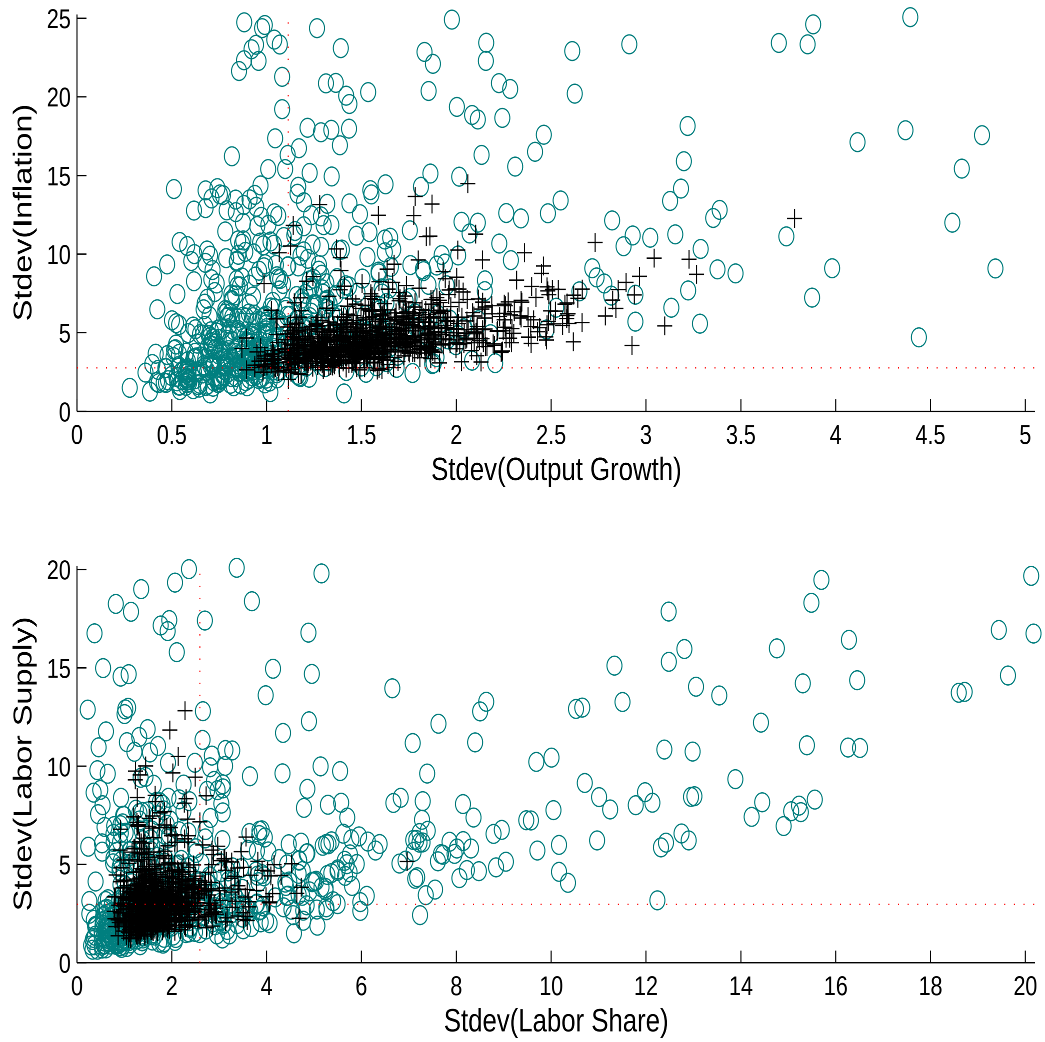
Notes: Top panels depict 400 draws from the implicit prior distribution for mean and autocorrelation of y for Model 2. Bottom panels depict draws from predictive distribution for two observations, y_1 and y_2. Blue circles correspond to Model 2, green crosses to Model 1.

Figure 3: Priors for Benchmark Model – Shock Parameters



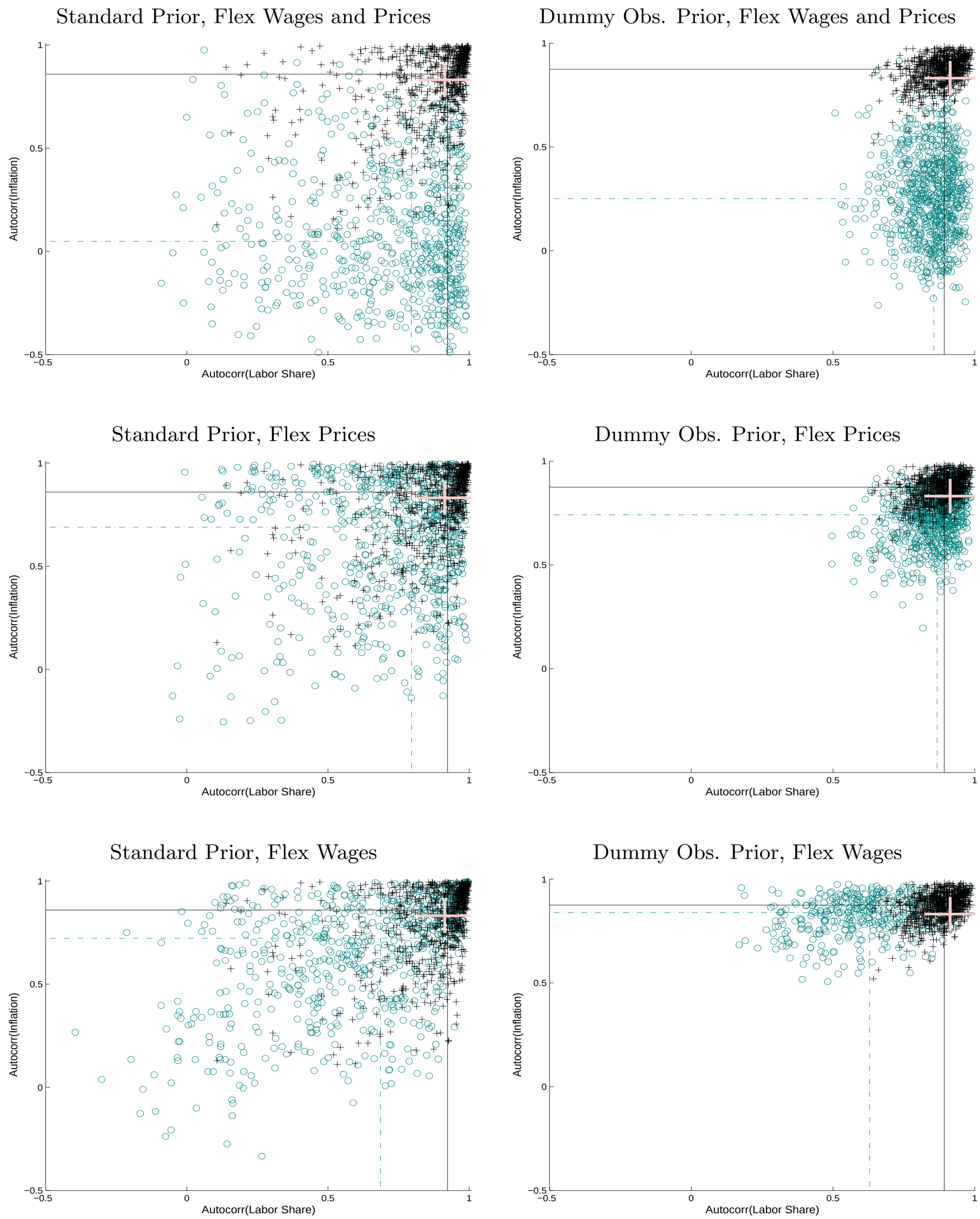
Notes: Each panel depicts draws from the prior distribution of the shock parameters. Grey circles indicate draws from the standard prior, whereas black crosses correspond to draws from the dummy observations prior.

Figure 4: Priors for Benchmark Model - Sample Moments



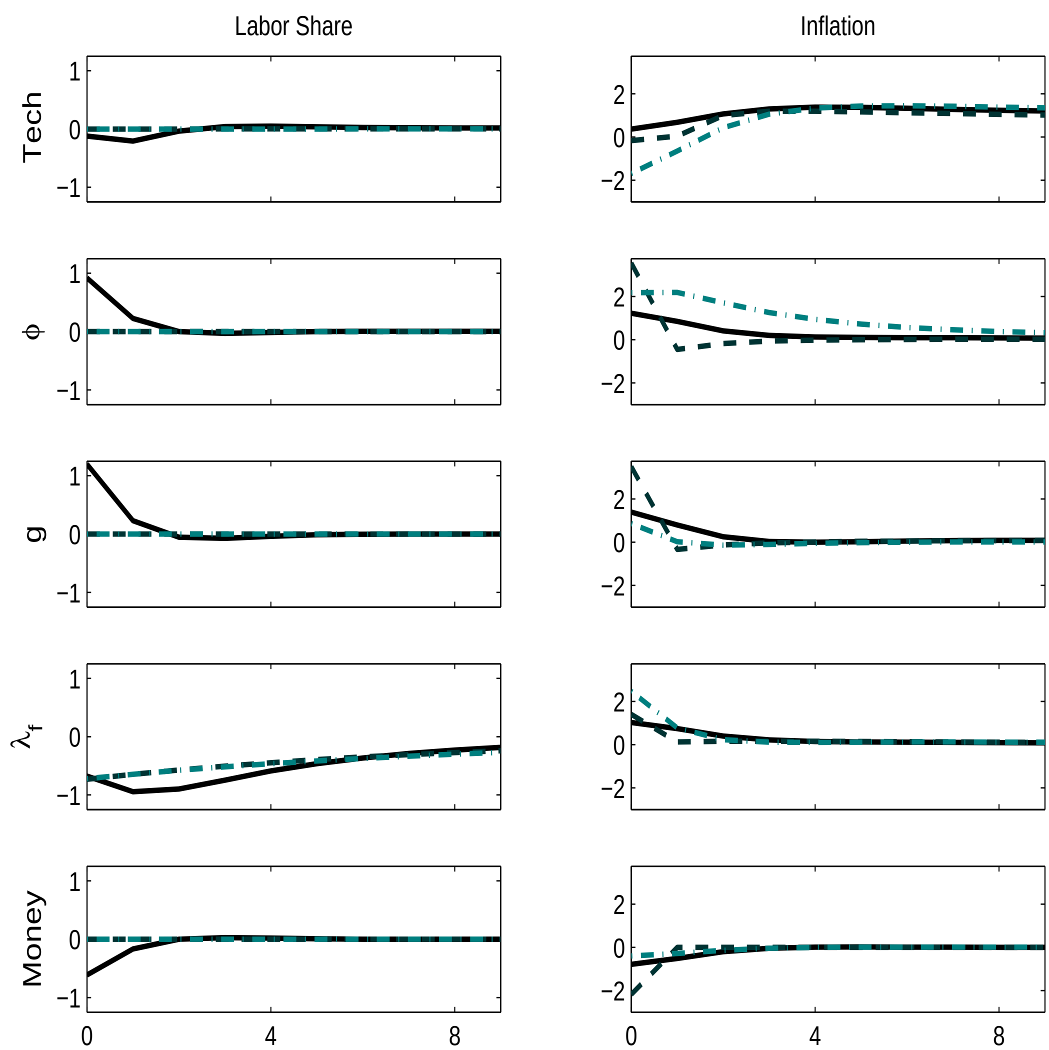
Notes: Each panel depicts draws from the prior predictive distribution of various sample standard deviations, calculated based on 100 artificial observations from the DSGE model. Grey circles indicate draws from the standard prior, whereas black crosses correspond to draws from the dummy observations prior. The intersection of the red dotted lines signifies the sample standard deviations computed from the pre-sample that is used to generate the Γ matrices for the dummy observations prior.

Figure 5: Nominal Rigidities: Benchmark versus Flex Wages / Prices Model



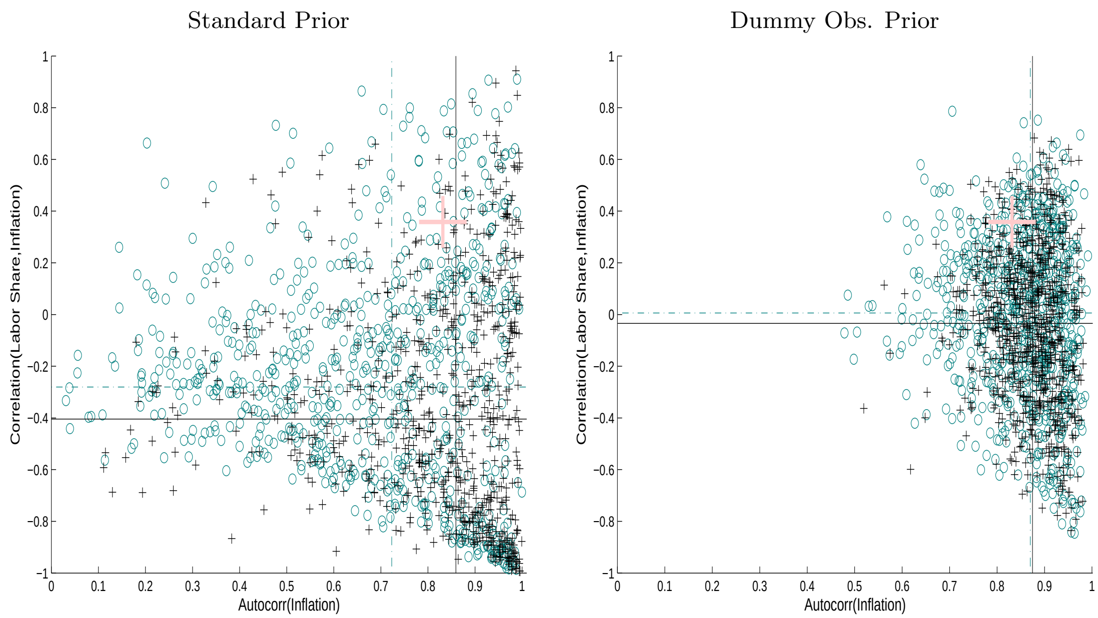
Notes: Each panel depicts draws from the prior predictive distribution of the autocorrelation of inflation and the labor share, calculated based on 100 artificial observations from the DSGE model. Black cross correspond to draws from the Benchmark model, whereas gray circles denote draws from the flexible price / wage models. The intersection of the solid black and dashed gray lines signifies the median of the prior predictive distributions. The thick gray cross indicates the corresponding moments for the data, i.e. the sample used in the estimation.

Figure 6: Impulse Response Functions - Flex Wages / Prices Models



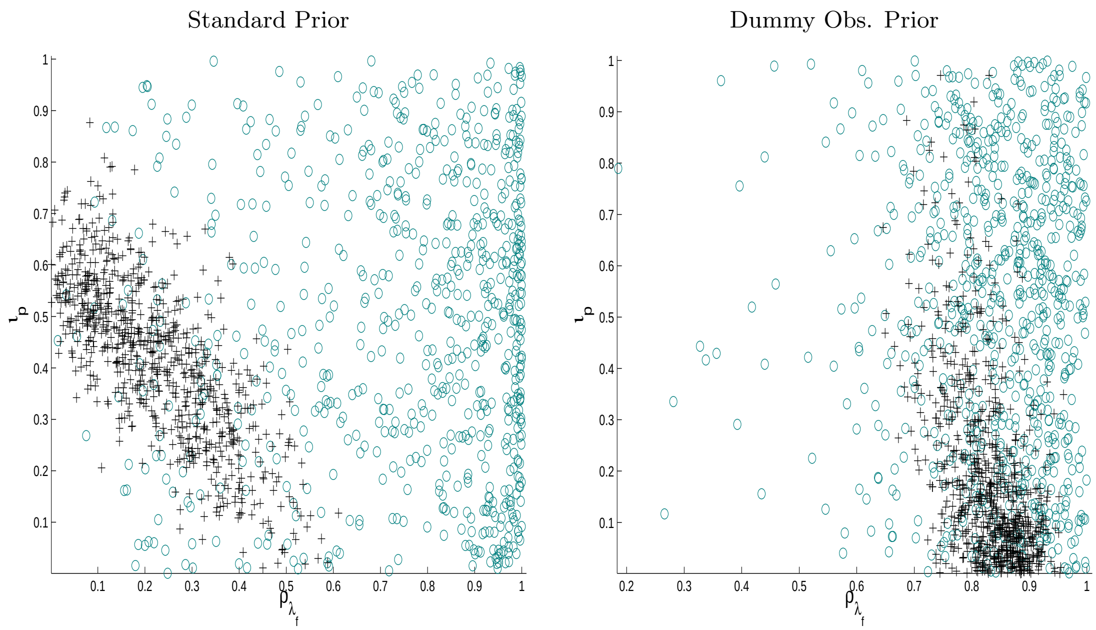
Notes: The left and right columns depict prior mean responses of the labor share and inflation, respectively, to the five structural shocks for the Flexible Wages and Prices model (dashed lines), the Flexible Wages/sticky prices model (black solid lines), and the Flexible Prices/sticky wage model (gray dash-and-dotted lines). The impulse responses are computed under the dummy observations prior.

Figure 7: Assessing the Phillips Curve: Benchmark versus No Dynamic Indexation



Notes: The top panels depict draws from the prior predictive distribution of the autocorrelation of inflation and the labor share, while the bottom panels depict draws from the prior predictive distribution of the contemporaneous correlation between labor share and inflation and labor share and output growth, respectively. The moments are calculated based on 100 artificial observations from the DSGE model. Black cross correspond to draws from the Benchmark model, whereas gray circles denote draws from the No Dynamic Indexation model. The intersection of the solid black and dashed gray lines signifies the median of the prior predictive distributions. The thick gray cross indicates the corresponding moments for the data, i.e. the sample used in the estimation.

Figure 8: Prior and Posterior Distribution of Dynamic Indexation Parameters and Mark-up Shocks Autocorrelation



Notes: Each panel depicts draws from the prior (gray circles) and the posterior (dark crosses) distribution of the parameters ι_p (dynamic indexation for prices), and ρ_{λ_f} (autocorrelation of the mark-up shock).