Prior Choice and DSGE Model Comparisons

Marco Del Negro
Federal Reserve Bank of Atlanta

Frank Schorfheide*
University of Pennsylvania
CEPR and NBER

Preliminary Version: August 21, 2006

Abstract

It is difficult to form beliefs about the parameters that govern the law of motion of the exogenous processes in DSGE models. We provide a simple way of translating beliefs for the moments of the endogenous variable into a prior for these parameters. We use our approach to investigate the importance of nominal rigidities and show that it affects our assessment of the relative importance of different frictions.

JEL CLASSIFICATION: C32

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

*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. 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.

1 Introduction

What frictions are important in a DSGE model to capture the salient features of the data? In the empirical literature literature this question is often answered by computing the posterior odds of the model with and without the friction of interest (e.g., Smets and Wouters 2003, and Rabanal and Rubio-Ramirez, 2005). The prior distribution for the deep parameters plays a key role in these model comparisons. Priors for DSGE model parameters are typically specified as follows. It is assumed that all the parameters are *a priori* independent. For a subset of the parameters, for instance related to labor supply elasticities, mark-ups, frequencies of price changes, capital adjustment costs there exists micro-econometric evidence on empirically plausible parameter values, which can be used in the specification of a prior. Other parameters, in particular those that determine the law of motion of the exogenous shock processes, we do not have any direct observations that can assist the choice of prior. 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 standard choice of priors in the literature has two shortcomings. First, the independence assumption potentially leads to a prior distribution that assigns non-negligible 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. But identical parameterizations of the exogenous shock processes potentially generate very different dynamics across model specifications. Hence a standard prior chosen for a given model can penalize an alternative specification.

Our approach is motivated as follows. While it is difficult to form beliefs about the parameters that govern the law of motion of the (unobservable) exogenous processes, we do have observations (and beliefs) on the volatility and serial correlation of the various endogenous variables. We provide a simple way of translating these beliefs into a reasonable prior distributions for the parameters of the exogenous shock processes using dummy observations. We argue that our approach can be used to overcome the two shortcomings of the standard prior distribution.

We use our approach to investigate the importance of nominal rigidities and show that it affects our assessment of the relative importance of different frictions. In particular, in contrast with some of the previous literature 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 reproduce the persistence in the labor share, a commonly used measure of marginal costs. We also find that the evidence for indexation in the Phillips Curve becomes rather tenuous once we use a prior that places all models considered on a similar footing.

Section 2 provides a simple example that illustrates that a naive choice of prior distributions can severely distort Bayesian posterior odds for competing models. In Section 3 we are introducing a dummy observations prior based on a quasi-likelihood function from a vector autoregression (VAR), that can be used to induce a prior for parameters that are difficult to quantify. Our approach is subsequently applied to a New Keynesian DSGE model, described in Section 4. Section 5 summarizes our empirical findings and Section 6 concludes.

2 A Simple Example

Consider the following two models, in which y_t is the observed endogenous variable and u_t is an unobserved shock process. In model $\mathcal{M}_1$ the shocks are serially uncorrelated. We we introduce a backward-looking term ϕy_{t-1} on the right-hand-side as is often done in the New Keynesian Phillips Curve literature:

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

In model $\mathcal{M}_2$, on the other hand, the u_t's are serially correlated:

$$\mathcal{M}_2: \quad y_t = \frac{1}{\alpha}\mathbb{E}_t[y_{t+1}] + u_t, \quad u_t = \rho u_{t-1} + \epsilon_t \sim iid(0, \sigma^2). \tag{2}$$

This example is taken from Lubik and Schorfheide (2006). Under the "backward-looking" specification the equilibrium law of motion becomes

$$\mathcal{M}_1: \quad y_t = \frac{1}{2}(\alpha - \sqrt{\alpha^2 - 4\rho\alpha})y_{t-1} + \frac{2\alpha}{\alpha + \sqrt{\alpha^2 - 4\rho\alpha}}\epsilon_t, \tag{3}$$

whereas under the model $\mathcal{M}_2$

$$\mathcal{M}_2: \quad y_t = \rho y_{t-1} + \frac{1}{1 - \rho/\alpha}\epsilon_t. \tag{4}$$

Models $\mathcal{M}_1$ and $\mathcal{M}_2$ are observationally equivalent. The model with the 'backward looking' component is distinguishable from the purely 'forward looking' specification only under a strong *a priori* restriction on the exogenous component, namely $\rho = 0$. 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.

Subsequently, we will illustrate the consequences of seemingly innocuous choices for prior distributions on posterior model odds. In the DSGE model literature, we can, broadly speaking, distinguish two types of parameters: 'deep' taste and technology parameters and 'auxiliary' parameters that determine the law of motion of the exogenous processes. Priors for the deep parameters are often chosen based on micro-econometric evidence, whereas the priors for the auxiliary parameters are either chosen arbitrarily or they are chosen based on some beliefs about the serial correlation and volatility of the endogenous variables. In our example we assume that α is a deep parameter whose prior has been specified based on micro-econometric evidence, whereas ρ and σ are auxiliary parameters. The baseline prior is denoted as Prior 1 and summarized in Table 1.

To start, we use the same prior distribution for models $\mathcal{M}_1$ and $\mathcal{M}_2$. We generate 200 draws from the prior predictive distribution of the sample autocorrelation and standard deviation of y_t. These draws are depicted in Panels (1,1) and (2,1) of Figure 1. Notice that according to model $\mathcal{M}_2$ the prior mean of the autocorrelation of y_t corresponds to the mean of ρ and is approximately 0.5. Under $\mathcal{M}_1$ the reduced-form autocorrelation coefficient is a nonlinear function of both ρ and α. It turns out that the prior mean of the predictive distribution of the autocorrelation is about 0.7. Hence, $\mathcal{M}_1$ and $\mathcal{M}_2$ have seemingly different implications for the observables.

We now generate a sample of $T = 80$ observations from an AR(1) model with autoregressive coefficient 0.9 and shock standard deviation 0.9. We compute the posterior for the two models under Prior 1. Draws from the posterior predictive distribution of the sample moments are plotted in Panels (1,2) and (2,2) of Figure 1. The intersection of the solid lines depict the actual sample moments. Given the fairly tight prior on α and ρ the estimated version of $\mathcal{M}_2$ still under-predicts the sample correlation of the data, whereas $\mathcal{M}_1$ captures it quite well. Log marginal data densities are reported in Table 2. The Bayes factor in favor of $\mathcal{M}_1$ is approximately e^{18}. Whether this value provides a good summary of our post-data model uncertainty depends crucially on how confident we are about the specification of Prior 1. If the prior reflects our intrinsic uncertainty about the parameters then the Bayes factor is appropriate and we are ready to conclude that the 'backward-looking' specification is less desirable than the specification with serially correlated shocks. If, on the other hand, the prior for the auxiliary parameters was fairly arbitrary, then the Bayes factors might be regarded as misleading. After all, the two models are observationally equivalent. Therefore we might regard a Bayes factor of 1 a more reasonable result than a Bayes factor of e^{18}.

We re-estimate model $\mathcal{M}_2$ under an alternative prior, denoted as Prior 2, that puts more weight on large values of ρ. The prior predictive distribution of the sample moments under this prior is depicted in Panel (3,1) of Figure 1. The draws are virtually indistinguishable from those obtained with model $\mathcal{M}_1$ and Prior 1. Indeed, under this modified prior distribution the Bayes factor of $\mathcal{M}_1$ versus $\mathcal{M}_2$ is essentially 1. The example illustrates that a careless choice of prior, in particular the use of identical priors for the parameters of the exogenous processes in different models, can lead to a significant distortion of posterior model odds.

3 Dummy Observation Priors for DSGE Models

In a Bayesian framework, the likelihood function of associated with an econometric model is re-weighted by a prior to obtain a posterior distribution for the model parameters. The prior distribution plays an important role in the estimation of DSGE models. The priors used in actual applications typically down-weigh regions of the parameter space that are at odds with observations not contained in the estimation sample. They might also 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. While, in principle, priors can be gleaned from personal introspection to reflect strongly held beliefs about the validity and quantitative implications of economic theories, in practice most priors are based on some observations.

Priors for DSGE model parameters are typically specified as follows. It is assumed that all the parameters are *a priori* independent. For a subset of the parameters, for instance related to labor supply elasticities, mark-ups, frequencies of price changes, capital adjustment costs there exists micro-econometric evidence on empirically plausible parameter values, which can be used in the specification of a prior. Other parameters, in particular those that determine the law of motion of the exogenous shock processes, we do not have any direct observations that can assist the choice of prior. 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 *standard* choice of priors in the literature 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. Consider a stationary AR(1) model of the form

$$y_t = \theta_1 y_{t-1} + \theta_2 + u_t, \quad u_t \sim iid\mathcal{N}(0, \sigma^2) \tag{5}$$

and $0 \leq \theta < 1$. Since the unconditional mean and variance of y_t are given by

$$\mathbb{E}[y_t] = \frac{\theta_2}{1-\theta_1}, \quad Var[y_t] = \frac{\sigma^2}{1-\theta_1^2}. \tag{6}$$

Hence, if θ_2 and σ are *a priori* independent, the prior assigns a fairly large probability to parameterizations of the model under which the process has a large mean (in absolute values) and a large variance.

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. Unless, the researcher holds strong beliefs about the parameters of the shocks instead of beliefs about reasonable magnitudes of the volatilities and autocorrelations of the observables, the prior distribution should be adjusted for each model specification under consideration. We will subsequently propose a dummy observations prior that can be used to overcome the two shortcomings of the standard prior distribution.

The insight that prior distributions can be represented by dummy observations dates back at least to 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). Dummy observations priors have several desirable features. First, they often lead to priors that are conjugate and allow for a straightforward computation of posterior distributions. Second, they are able to introduce correlation between model parameters without requiring the researcher to specify a complete covariance matrix for all the parameters.

Consider the following VAR:

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + u_t, \quad u_t \sim \mathcal{N}(0, \Sigma), \tag{7}$$

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 in matrix notation as

$$Y = X\Phi + U. \tag{8}$$

Here Y is the $T \times n$ matrix with rows y'_t, X is the $T \times k$ matrix with rows x'_t, U is composed of u'_t and $\Phi = [\Phi_0, \Phi_1, \ldots, \Phi_p]'$. A dummy observations prior for the VAR can be constructed as follows. Collect the T^* dummy observations in the matrices Y^* and X^* and interpret the VAR likelihood function for dummy observations

$$\begin{aligned} &\mathcal{L}(\Phi, \Sigma | Y^*) = \\ &\quad (2\pi)^{-nT^*/2} |\Sigma|^{-T^*/2} \exp\left\{ -\frac{1}{2} tr[\Sigma^{-1}(Y^{*'}Y^* - \Phi' X^{*'} Y^* - Y^{*'} X^* \Phi + \Phi' X^{*'} X^* \Phi)] \right\}. \end{aligned} \tag{9}$$

as prior density for Φ and Σ. Combining (9) with the improper prior $p(\Phi, \Sigma) \propto |\Sigma|^{-(n+1)/2}$ yields

$$\begin{aligned} &p(\Phi, \Sigma | Y^*, X^*) \\ &\quad = c_*^{-1} |\Sigma|^{-\frac{T^*+n+1}{2}} \exp\left\{ -\frac{1}{2} tr[\Sigma^{-1}(Y^{*'}Y^* - \Phi' X^{*'} Y^* - Y^{*'} X^* \Phi + \Phi' X^{*'} X^* \Phi)] \right\}, \end{aligned} \tag{10}$$

which implies that Σ has a marginal inverted Wishart distribution and Φ is multivariate normal conditional on Σ. Since the likelihood function is Gaussian and the VAR is a linear regression model, one only needs to specify the sufficient statistics for the dummy observations: $Y^{*'}Y^*$, $X^{*'}Y^*$, and $X^{*'}X^*$. We will subsequently modify this approach to obtain a prior for (a subset of) the DSGE model parameters.

3.1 General Approach

We denote the DSGE model parameters by the vector θ and decompose it into two components: $\theta = [\theta'_1, \theta_2]'$. Roughly speaking, θ_1 collects the parameters for which we can elicit prior distributions, say, based on micro-econometric and other quantitative evidence not obtained from the estimation sample, and θ_2 is a sub-vector of parameters for which we choose prior distributions such that the model implied autocovariances of the endogenous variables are "realistic" and comparable across models. We assume that the observables have been transformed such that the vector y_t is covariance stationary. We use Γ_{YY}, Γ_{YX} and Γ_{XX} to denote population autocovariances $I\!E[y_t y'_t]$, $I\!E[y_t x'_t]$, and $I\!E[x_t x'_t]$, respectively. If the population autovariances are calculated from a DSGE model conditional on a particular parameterization, we use the notation $\Gamma_{YY}(\theta)$ or $\Gamma_{YY}(\theta_1, \theta_2)$, which serves as a shorthand for $\Gamma_{YY}([\theta'_1, \theta'_2]')$.

We are departing from the traditional approach of constructing prior distributions from dummy observations in that we do not use the actual likelihood function of the DSGE model to obtain a prior density. Instead, we are using a quasi-likelihood function for which we have a low dimensional set of sufficient statistics. More specifically, we are using the likelihood function with a p-th order vector autoregression, given in (9). To relate the DSGE model parameters to the VAR, we define a VAR approximation of the DSGE model through the population least-squares regression:

$$\Phi_*(\theta) = [\Gamma_{XX}(\theta)]^{-1}\Gamma_{XY}(\theta), \quad \Sigma_*(\theta) = \Gamma_{YY}(\theta) - \Gamma_{YX}(\theta)[\Gamma_{XX}(\theta)]^{-1}\Gamma_{XY}(\theta). \tag{11}$$

To simplify notation we collect the autocovariance matrices $\Gamma = \{\Gamma_{YY}, \Gamma_{XY}, \Gamma_{XX}\}$. and will use $\mathcal{L}(\theta_1, \theta_2|\Gamma)$ as shorthand for $\mathcal{L}([\theta_1', \theta_2']'|\Gamma)$ below. Our dummy observations prior for the DSGE model parameters is based on the quasi-likelihood function

$$\begin{aligned} \mathcal{L}(\theta|\Gamma, T^*) &= |\Sigma_*(\theta)|^{-(T^*+n+1)/2} \\ &\times \exp\left\{-\frac{T^*}{2} tr\left[\Sigma_*(\theta)^{-1}(\Gamma_{YY} - 2\Phi_*(\theta)\Gamma_{XY} + \Phi_*'(\theta)\Gamma_{XX}\Phi_*(\theta)\right]\right\}, \end{aligned} \tag{12}$$

where the autocovariance matrices Γ are either constructed from introspection, a pre-sample of actual observations, or an alternative candidate model. The prior (12) places low probability on values of θ for which the DSGE model implied autocovariances strongly differ from the Γ's. The larger T^*, the more concentrated the prior density.

Given our goal of specifying a prior for θ_2 such that the model implied autocovariances are broadly in line with the target autocovariances 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.

$$p_*(\theta_1, \theta_2|\Gamma, T_*) = \underbrace{c_1(\theta_1|\Gamma, T_*)\mathcal{L}(\theta_1, \theta_2|\Gamma, T^*)\pi(\theta_2)}_{p_*(\theta_2|\theta_1, \Gamma, T^*)} p(\theta_1), \tag{13}$$

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 (13) that it depends on a normalization constant that typically cannot be calculated analytically. Hence, (13) would be very difficult to implement in practice.

For the empirical work presented below, we consider the following simplification.

$$p(\theta_1, \theta_2|\Gamma, T_*) = \underbrace{c_1(\underline{\theta}_1|\Gamma, T_*)\mathcal{L}(\underline{\theta}_1, \theta_2|\Gamma, T^*)\pi(\theta_2)}_{p(\theta_2|\Gamma, T_*)} p(\theta_1). \tag{14}$$

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

$$\frac{1}{c_1(\theta_1|\Gamma, T_*)} = \int \mathcal{L}(\underline{\theta}_1, \theta_2|\Gamma, T^*)\pi(\theta_2)d\theta_2.$$

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.

3.2 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 lag length p of the vector autoregressive specification determines how many autocovariances are included in the construction of the prior distribution. The parameter T^* scales the prior distribution. A large value leads to a concentrated prior. Most importantly, one has to choose a suitable Γ matrices. These matrices could corresponds to sample autocovariances calculate from a pre-sample or based on data from a different country, or they could be obtained from a benchmark model. In the empirical application in Section 5 we are using sample autocovariances based on a pre-sample.

Since the functions $\Phi_*(\theta)$ and $\Sigma_*(\theta)$ are highly nonlinear, we need numerical techniques to generate draws from the prior distribution and to compute the normalization constant $c_1(\theta_1|\Gamma, T_*)$. For the application below we use a random-walk Metropolis algorithm, described in detail in An and Schorfheide (2006) to generate draws from the prior distribution. Based on the output of the Metropolis algorithm, Geweke's (1999) modified harmonic mean estimator is used to calculate the normalization constant. While the normalization constant is not needed to generate prior or posterior parameter draws, it is important for the calculation of the marginal likelihood associated with a DSGE model.

3.3 Quasi-Likelihoods versus Parameter Transformations

One of the motivations for our prior distribution based on the quasi-likelihood function was that we wanted to use beliefs about unconditional moments of the endogenous variable to induce a prior for the exogenous shock processes. In principle, one could start from a prior on selected population moments for the endogenous variables and then, by a change of variable argument, deduce an implicit prior for the parameters of the shock processes. While such an approach is not practical for DSGE models that are highly nonlinear in the parameters, we are illustrating the relationship between our dummy observations prior and a prior constructed from a parameter transformation in the context of the AR(1) model given in (5). For simplicity we assume that $\sigma^2 = 1$. Consider the following prior for θ_1 and θ_2:

$$\theta_1 \sim \mathcal{U}[0,1] \quad \text{and} \quad \theta_2 \sim \mathcal{N}(0, \tau^2). \tag{15}$$

While this prior enables a straightforward calculation of the posterior, it is not very convenient if we would like to impose a particular belief about say the unconditional mean and the autocorrelation of y_t.

We will begin by constructing a prior based on a quasi-likelihood function derived from the simple location model $y_t = \phi + u_t$. Suppose we would like to incorporate the belief that the mean of y_t is approximately $\underline{\mu}$. 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

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

Hence, we obtain

$$\mathcal{L}(\theta|\Gamma, T^*) = (2\pi)^{-T^*/2} \exp\left\{ -\frac{T^*}{2}\left(1 + \underline{\mu}^2 - 2\underline{\mu}\frac{\theta_2}{1-\theta_1} + \frac{\theta_2^2}{(1-\theta_1)^2}\right)\right\} \tag{16}$$

Combining the quasi-likelihood function with the baseline prior distribution (15) yields

$$p(\theta_2, \theta_1) \quad \propto \quad \exp\left\{ -\frac{1}{2}\left(-2\theta_2\underline{\mu}(1-\theta_1)\frac{T^*}{(1-\theta_1)^2} + \theta_2^2\left[\frac{T^*}{(1-\theta_1)^2} + \frac{1}{\tau^2}\right]\right)\right\},$$

where $\propto$ denotes proportionality. Therefore,

$$\theta_2|\theta_1 \sim \mathcal{N}\left(\left[1 + \frac{1-\theta_1}{\tau^2 T^*}\right]^{-1} \underline{\mu}(1-\theta_1), \left[\frac{T^*}{(1-\theta_1)^2} + \frac{1}{\tau^2}\right]^{-1}\right) \tag{17}$$

and

$$p(\theta_1) \propto \left[\frac{T^*}{(1-\theta_1)^2} + \frac{1}{\tau^2}\right]^{-1/2} \exp\left\{ -\frac{T^*}{2}\frac{\underline{\mu}^2}{(1-\theta_1)^2}\left[\frac{1}{(1-\theta_1)^2} + \frac{1}{T^*\tau^2}\right]^{-1}\right\} \tag{18}$$

If we let $\tau^2 \longrightarrow \infty$ we obtain

$$p(\theta_1) \propto |1-\theta_1| \quad \text{and} \quad \theta_2|\theta_1 \sim \mathcal{N}\left(\underline{\mu}(1-\theta_1), \frac{1}{T^*}(1-\theta_1)^2\right), \tag{19}$$

which corresponds to (13) in our AR(1) example. The larger T^* the smaller the variance of the conditional distribution of θ_2 given θ_1. 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 (14) we obtain

$$\theta_1 \sim \mathcal{U}[0,1] \quad \text{and} \quad \theta_2 \sim \mathcal{N}\left(\underline{\mu}(1-\underline{\theta}_1), \frac{1}{T^*}(1-\underline{\theta}_1)^2\right). \tag{20}$$

As an alternative, consider a prior constructed from a change of parameters. Suppose that our beliefs about the mean $\mu = I\!E[y_t]$ are represented by $\mu \sim \mathcal{N}(\underline{\mu}, \lambda^2)$. The determinant of the Jacobian of the mapping from $[\theta_2, \theta_1]$ to $[\mu, \theta_1]$ is given by

$$|J| = \begin{vmatrix} \frac{1}{1-\theta_1} & \frac{\theta_2}{(1-\phi)^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{1-\theta_1}$$

Thus,

$$p(\theta_2, \theta_1) \propto \lambda^{-1}(1-\theta_1)^{-1} \exp\left\{ -\frac{1}{2\lambda^2} \left(\frac{\theta_2}{1-\theta_1} - \underline{\mu} \right)^2 \right\} \tag{21}$$

and we obtain the following prior:

$$\theta_1 \sim \mathcal{U}[0,1], \quad \theta_2|\theta_1 \sim \mathcal{N}\Big(\underline{\mu}(1-\theta_1), \lambda^2(1-\theta_1)^2 \Big). \tag{22}$$

Hence the conditional distribution of θ_2 given θ_1 under the change-of-parameter approach is identical to (19) if we set $\lambda = 1/\sqrt{T^*}$. The implied marginal distributions of θ_1 are, however, different. If the context of DSGE model, the change-of-parameter approach is impractical because the autocovariances of y_t are complicated nonlinear functions of θ and it is difficult to calculate the Jacobian associated with the parameter transformation.

3.4 Predictive Likelihoods

Many economists hold the view that models ought to be judged on their out-of-sample predictive performance. If a model generates a predictive density for future observations, then this density provides a fairly natural measure of forecast accuracy. If the actual observation falls into the tails of the predictive density then the score will be low, indicating that the model assigned low probability to an event that occurred. Hence, a sample of T observations could be split into two parts such that we fit the model(s) based on the first τ observations and then assess how well the model predicts the remaining $T - \tau$ observations. Hence, for a parametric model with likelihood function $p(y_1, \ldots, y_T|\theta)$ we can write

$$p(y_{\tau+1}, \ldots, y_T|y_1, \ldots, y_\tau) = \int p(y_{\tau+1}, \ldots, y_T|\theta, y_1, \ldots, y_\tau) p(\theta|y_1, \ldots, y_\tau) d\theta. \tag{23}$$

Here $p(\theta|y_1, \ldots, y_\tau$ 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. It is well known, that the predictive likelihood (23), e.g. Geweke (2005), is closely related to the marginal likelihood:

$$p(y_{\tau+1}, \ldots, y_T|y_1, \ldots, y_\tau) = \frac{p(y_1, \ldots, y_T)}{p(y_1, \ldots, y_\tau)}. \tag{24}$$

An advantage of using predictive likelihoods instead of marginal likelihoods for model comparisons is that the predictive likelihood is less sensitive to the choice of prior distribution for θ and may be well defined even if the prior for θ is improper.

The pre-sample based adjustment proposed in Section 3.1 is similar in spirit to the use of the predictive likelihood. Technically, there are two differences. First, instead of using the actual likelihood function to obtain a distribution of θ conditional on the pre-sample, we are using a quasi-likelihood function, derived from a VAR approximation of the DSGE model. Second, the shape of this quasi-likelihood function allows us to scale the distribution of θ with a single hyperparameter T_*. The hyperparameter adds valuable flexibility to our approach.

3.5 Marginal Likelihood Functions and Predictive Checks

This subsection is very preliminary. We provide a heuristic explanation of why a comparison of prior predictive distributions for sample moments can generate some insights about the outcome of model comparisons based on marginal likelihood functions. Consider two models, $\mathcal{M}_1$ and $\mathcal{M}_2$. We assume that the models share a common likelihood function $p(Y|\theta)$ and can be characterized through different prior distributions $p(\theta|\mathcal{M}_i)$. The odds of model $\mathcal{M}_1$ versus $\mathcal{M}_2$ are updated through the Bayes factor

$$B_T(\mathcal{M}_1, \mathcal{M}_2) = \frac{p(Y|\mathcal{M}_1)}{p(Y|\mathcal{M}_2)} \tag{25}$$

where

$$p(Y|\mathcal{M}_i) = \int p(Y|\theta)p(\theta|\mathcal{M}_i)d\theta$$

is the marginal likelihood associated with model $\mathcal{M}_i$. Suppose that S is a set of sufficient statistics for the two models. Then we can write the likelihood function as

$$p(Y|\theta) = f(Y|S)g(S|\theta).$$

Hence,

$$B_T(\mathcal{M}_1, \mathcal{M}_2) = \frac{\int g(S|\theta)p(\theta|\mathcal{M}_1)d\theta}{\int g(S|\theta)p(\theta|\mathcal{M}_2)d\theta} \tag{26}$$

If the observed value of the sufficient statistic for model $\mathcal{M}_1$ is far in the tails of its (prior) predictive distribution, the odds in favor of $\mathcal{M}_1$ tend to be low.

To the extent that sample moments provide a good approximation of the sufficient statistics associated with a linearized DSGE model, a comparison of their prior predictive distributions across models can give insights of why one model is preferred to another.

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 $i \in [0, 1]$:

$$Y_t = \left[\int_0^1 Y_t(i)^{\frac{1}{1+\lambda_{f,t}}} di \right]^{1+\lambda_{f,t}} \tag{27}$$

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

$$\ln \lambda_{f,t} = (1 - \rho_{\lambda_f}) \ln \lambda_f + \rho_{\lambda_f} \ln \lambda_{f,t-1} + \sigma_{\lambda,f} \epsilon_{\lambda,t}, \tag{28}$$

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 (27). 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:

$$Y_t(i) = \left(\frac{P_t(i)}{P_t} \right)^{-\frac{1+\lambda_{f,t}}{\lambda_{f,t}}} Y_t \quad \text{and} \quad P_t = \left[\int_0^1 P_t(i)^{-\frac{1}{\lambda_{f,t}}} di \right]^{-\lambda_{f,t}}. \tag{29}$$

4.2 Intermediate goods producers

Good i is made using the technology:

$$Y_t(i) = \max \left\{ Z_t^{1-\alpha} K_t(i)^\alpha L_t(i)^{1-\alpha} - Z_t \mathcal{F}, 0 \right\}, \tag{30}$$

where the technology shock Z_t (common across all firms) follows a unit root process, and where $\mathcal{F}$ represent fixed costs faced by the firm. Based on preliminary estimation results we decided to set $\mathcal{F} = 0$ in the empirical analysis. We define technology growth $z_t = \log(Z_t/Z_{t-1})$ and assume that z_t follows the autoregressive process:

$$z_t = (1-\rho_z)\gamma + \rho_z z_{t-1} + \sigma_z \epsilon_{z,t}. \tag{31}$$

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:

$$\frac{K_t(i)}{L_t(i)} = \frac{\alpha}{1-\alpha}\frac{W_t}{R_t^k}, \tag{32}$$

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

$$P_t(i) = (\pi_{t-1})^{\iota_p}(\pi_*)^{1-\iota_p}, \tag{33}$$

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:

$$\begin{aligned} &\max_{\tilde{P}_t(I)} I\!\!E_t\left[\sum_{s=0}^{\infty}\zeta_p^s\beta^s\Xi_{t+s}^p\left(\tilde{P}_t(i)\left(\Pi_{l=1}^s\pi_{t+l-1}^{\iota_p}\pi_*^{1-\iota_p}\right)-MC_{t+s}\right)Y_{t+s}(i)\right] \\ &\text{s.t. } Y_{t+s}(i) = \left(\frac{\tilde{P}_t(i)\left(\Pi_{l=1}^s\pi_{t+l-1}^{\iota_p}\pi_*^{1-\iota_p}\right)}{P_{t+s}}\right)^{-\frac{1+\lambda_{f,t}}{\lambda_{f,t}}} Y_{t+s},\ MC_{t+s} = \frac{\alpha^{-\alpha}W_{t+s}^{1-\alpha}R_{t+s}^{k\ \alpha}}{(1-\alpha)^{(1-\alpha)}Z_{t+s}^{1-\alpha}}, \end{aligned} \tag{34}$$

where $\beta^s\Xi_{t+s}^p$ 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 (29) we obtain the following law of motion for the aggregate price level:

$$P_t = \left[(1-\zeta_p)\tilde{P}_t^{-\frac{1}{\lambda_{f,t}}} + \zeta_p\left(\pi_{t-1}^{\iota_p}\pi_*^{1-\iota_p}P_{t-1}\right)^{-\frac{1}{\lambda_{f,t}}}\right]^{-\lambda_{f,t}}. \tag{35}$$

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:

$$L_t = \left[\int_0^1 L_t(j)^{\frac{1}{1+\lambda_w}}di\right]^{1+\lambda_w}, \tag{36}$$

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:

$$(a) \quad L_t(j) = \left(\frac{W_t(j)}{W_t}\right)^{-\frac{1+\lambda_w}{\lambda_w}} L_t \quad \text{and} \quad (b) \quad W_t = \left[\int_0^1 W_t(j)^{-\frac{1}{\lambda_w}} di\right]^{-\lambda_w}. \tag{37}$$

4.4 Households

The objective function for household j is given by:

$$I\!E_t \sum_{s=0}^{\infty} \beta^s \left[\log(C_{t+s}(j) - hC_{t+s-1}(j)) - \frac{\phi_{t+s}}{1+\nu_l} L_{t+s}(j)^{1+\nu_l} + \frac{\chi}{1-\nu_m}\left(\frac{M_{t+s}(j)}{Z_{t+s}P_{t+s}}\right)^{1-\nu_m}\right] \tag{38}$$

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:

$$\ln \phi_t = (1-\rho_\phi)\ln \phi + \rho_\phi \ln \phi_{t-1} + \sigma_\phi \epsilon_{\phi,t}, \tag{39}$$

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:

$$\begin{aligned} & P_{t+s}C_{t+s}(j) + P_{t+s}I_{t+s}(j) + B_{t+s}(j) + M_{t+s}(j) + T_{t+s}(j) \\ \leq \quad & R_{t+s-1}B_{t+s-1}(j) + M_{t+s-1}(j) + A_{t+s-1}(j) + \Pi_{t+s} + W_{t+s}(j)L_{t+s}(j) \\ & + \left(R^k_{t+s}u_{t+s}(j)\bar{K}_{t+s-1}(j) - P_{t+s}a(u_{t+s}(j))\bar{K}_{t+s-1}(j)\right), \end{aligned} \tag{40}$$

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:

$$K_t(j) = u_t(j)\bar{K}_{t-1}(j), \tag{41}$$

and receive $R^k_t u_t(j)\bar{K}_{t-1}(j)$ 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:

$$\bar{K}_t(j) = (1-\delta)\bar{K}_{t-1}(j) + \mu\left(1 - S\left(\frac{I_t(j)}{I_{t-1}(j)}\right)\right) I_t(j), \tag{42}$$

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:

$$\begin{array}{ll} \max_{\tilde{W}_t(j)} & I\!\!E_t \sum_{s=0}^{\infty} \zeta_w^s \beta^s b_{t+s}\left[-\frac{\phi_{t+s}}{1+\nu_l} L_{t+s}(j)^{1+\nu_l}\right] \\ \text{s.t.} & \text{Eq. (40) for } s = 0, \ldots, \infty, \text{ (37a), and} \\ & W_{t+s}(j) = \left(\Pi_{l=1}^{s}(\pi_* e^\gamma)^{1-\iota_w}(\pi_{t+l-1}e^{z_{t+l-1}})^{\iota_w}\right)\tilde{W}_t(j). \end{array} \tag{43}$$

We again consider only the symmetric equilibrium in which all agents solving (43) will choose the same $\tilde{W}_t(j)$. From (37b) it follows that:

$$W_t = [(1-\zeta_w)\tilde{W}_t^{-\frac{1}{\lambda_w}} + \zeta_w((\pi_* e^\gamma)^{1-\iota_w}(\pi_{t-1}e^{z_{t-1}})^{\iota_w} W_{t-1})^{-\frac{1}{\lambda_w}}]^{-\lambda_w}. \tag{44}$$

Finally, we assume there is a complete set of state contingent securities in nominal terms, which implies that the Lagrange multiplier $\Xi^p_t(j)$ associated with (40) 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:

$$\frac{R_t}{R^*} = \left(\frac{R_{t-1}}{R^*}\right)^{\rho_R}\left[\left(\frac{\pi_t}{\pi_*}\right)^{\psi_1}\left(\frac{Y_t}{Y^*_t}\right)^{\psi_2}\right]^{1-\rho_R} e^{\sigma_R \epsilon_{R,t}}, \tag{45}$$

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 (45) 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

$$P_t G_t + R_{t-1} B_{t-1} + M_{t-1} = T_t + M_t + B_t, \tag{46}$$

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

$$G_t = (1 - 1/g_t) Y_t, \tag{47}$$

where g_t follows the exogenous process:

$$\ln g_t = (1 - \rho_g) \ln g + \rho_g \ln g_{t-1} + \sigma_g \epsilon_{g,t} \tag{48}$$

4.6 Resource Constraint

The aggregate resource constraint:

$$C_t + I_t + a(u_t) \bar{K}_{t-1} = \frac{1}{g_t} Y_t. \tag{49}$$

can be derived by integrating the budget constraint (40) across households, and combining it with the government budget constraint (46) 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 Empirical Application

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: $\delta = 0.025$, $\lambda_w = 0.3$, and $\mathcal{F} = 0$. 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. Loosely speaking, we will refer to θ_1 as the taste-and-technology parameters. and we will use the dummy observations to generate a prior distribution for the parameters that determine the law of motion of the exogenous processes.

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. The Γ matrices are constructed from a pre-sample, and we use $p = 1$ lag of y_t to construct the quasi-likelihood function. In Section 5.2 we describe a *Standard* prior distribution[1] for the DSGE model parameters $\theta = [\theta_1', \theta_2']'$. When constructing our dummy observations prior, we retain the marginal distribution of $p(\theta_1)$ and generate alternative distributions for θ_2. 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, we study Phillips curve dynamics of various model specifications under the standard prior and the dummy observations prior.

[1]The term *standard* does not refer to particular numerical values but rather to the way in which micro-level and pre-sample information is used to justify the marginal prior densities.

5.1 The Data

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*), also in percent.

Our dummy observations prior is based on autocovariance matrices $\hat{\Gamma}_{YY}$, $\hat{\Gamma}_{YX}$, $\hat{\Gamma}_{XX}$ computed from a sample of observations ranging from QIII:1954 to QIV:1980. In addition to various prior statistics, we are also reporting marginal likelihood values for a sample of 100 observations from QI:1981 to QIV:2005.

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 3 and the first four columns of Table 4 and essentially corresponds to the one used in Del Negro, Schorfheide, Smets, and Wouters (2006).

Consider the marginal distributions of the taste-and-technology parameters θ_1. 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.

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 3 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 (14) 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 4. As we consider different DSGE model specifications in Sections 5.4 and 5.5, we keep the standard prior unchanged.

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 4 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 2 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 3 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 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 Assessing the Role of Nominal Rigidities

This section discusses how nominal rigidities, sticky prices and wages, affect the model's ability to describe the data. We compare three 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, and iii) the model without either wage or price stickiness ($\zeta_w = \zeta_p = 0$), which we refer to as the *Flexible Wages and Prices* model. We show how the presence of nominal rigidities changes the models' implications for the 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 likelihoods. We also show 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 likelihoods.

Among others, papers by Rabanal and Rubio-Ramirez (2005), Smets and Wouters (2003) and Christiano, Eichenbaum, and Evans (2005) also address the importance of nominal rigidities in DSGE models. A contribution of our paper is to assess the robustness of the results of the previous literature to changes in the prior distribution for the DSGE model parameters. We find that overall the results of the previous literature are robust: both nominal rigidities – sticky prices and wages – are needed to describe the data.

In addition, we show how the choice of the prior may change the reason *why* the flexible price and wage models are rejected in comparison with the Benchmark. When we use the standard prior on all three models, for instance, we find that both the Flexible Wages and Prices and the Flexible Wages predict that inflation is less autocorrelated than in the Benchmark model. However, when we use the dummy observations prior we find that the prior implications for the autocorrelation of inflation in the Benchmark and the Flexible Wages are roughly the same, while the Flexible Wages and Prices still predicts a much lower degree of autocorrelation. Where the Benchmark and the Flexible Wages still differ, even under the Dummy Observation prior, is in their predictions for the autocorrelation of the labor share, which is much lower for the latter than for the former. Marginal likelihood comparisons, in turn, find that both the Flexible Wages and Prices and the Flexible Wages are less capable of explaining the data than the Benchmark model, even under the dummy observations prior.

Figure 4 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 bottom two panels compare the Benchmark and the Flexible Prices models. The two 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 (O) represent draws from the Benchmark and the alternative model, respectively. The dark and light lines show the median for the two models.

The figure shows that under the dummy observations prior the Benchmark model's predictions are more concentrated than under the standard prior, but the median prediction is largely unchanged: Under both priors the model generates large persistence in both inflation and the labor share. 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. Under the dummy observations prior the predicted inflation autocorrelation rises, but is nowhere as high as that predicted by the Benchmark model. 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. Under the dummy observations prior the differences between the Benchmark and the Flexible Wages predictions for inflation autocorrelation nearly disappear. Differences in the predictions for the autocorrelation in the labor share remain, however. The Flexible Wage model cannot generate the degree of persistence in the labor share afforded by the presence of both nominal rigidities.

An interesting feature of Figure 4 is that the Flexible Wages and Prices model can generate persistence in the labor share, even under the standard prior, while the Flexible Wages model never can. In Figure 5 we therefore compare the impulse response functions for both models, computed under the dummy observations prior. The dashed and solid lines represent the responses for the Flexible Wages and Prices and for the Flexible Wages models, respectively. If both prices and wages are flexible

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

where $\widehat{lsh}_t$ are percentage deviations of the labor share from its steady-state value and $\hat{\lambda}_{f,t}$ is the mark-up shock. Hence, a persistent mark-up shock leads to a persistent labor share. In the Flexible Wages model other shocks affect the labor share as well, and their impulse responses are far less persistent than that of the mark-up shock. For inflation, the persistence of mark-up shocks does not help the Flexible Wages and Prices model. Conversely, the inflation impulse responses for the Flexible Wages model are all quite persistent.

Table 5 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 are different in the three different models, highlighting the idea that these parameters have different meanings depending on the model under consideration. For the Benchmark model the dummy observations prior generates a persistent mark-up shock. For the Flexible Wages model government spending shocks are persistent. For the Flexible Wages and Prices model both government spending and mark-up shocks are very persistent.

Finally, Figure 7 depicts marginal data density differentials relative to the benchmark model. Consistently with the discussion of Figure 4, the use of the Dummy Observation prior narrows the gap between the models. Yet, differences in marginal likelihoods are stark even under the alternative prior.

5.5 Assessing the Phillips Curve

This section focuses on the specification of the New Keynesian Phillips curve relationship, which for the model described in section 4 takes the following log-linear form:

$$\begin{aligned}\widehat{\pi}_t &= \frac{(1-\zeta_p\beta)(1-\zeta_p)}{(1+\iota_p\beta)\zeta_p}\left[\widehat{mc}_t + \frac{\lambda_f}{1+\lambda_f}\widehat{\lambda}_{f,t}\right] \\ &\quad + \frac{\iota_p}{1+\iota_p\beta}\widehat{\pi}_{t-1} + \frac{\beta}{1+\iota_p\beta}I\!E_t[\widehat{\pi}_{t+1}].\end{aligned} \tag{50}$$

where $\widehat{}$ denotes log-deviations from the steady state. Our model implies that (in terms of log deviations) marginal costs equal the labor share:

$$\widehat{mc}_t = \widehat{lsh}_t.$$

A large body of literature (Galí and Gertler 1999, JME 2005 volume, ...) 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, we use full information methods here. Equation (50) shows that in terms of our model the issue boils down to assessing the magnitude of the parameter $\iota_p \in [0,1]$. We therefore compare three models: i) the *Benchmark* model that allows for partial indexation ($\iota_p \in (0,1)$); ii) The very same model with *No Indexation* ($\iota_p = 0$), and iii) the model with *Full Indexation* ($\iota_p = 1$). As shown in the simple example 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 (50) 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 6 shows the prior predictive distributions for the sample autocorrelations of inflation and the correlation between inflation and labor share. The autocorrelation of inflation is clearly a moment of interest in assessing the empirical relevance of the Phillips curve. The correlation between inflation and labor share is also relevant. Galí and Gertler (1999) argue 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 *a priori* implications of the three models considered here for these two moments.

The top two panels compare the predictions for the Benchmark and the No Indexation models. The bottom two panels compare the Benchmark and the Full Indexation models. The two panels on the left use the standard prior while the two panels on the right use the dummy observations prior with 10 dummy observations. In each panel the dark crosses (+) and the lighter circles (O) represent draws from the Benchmark and the alternative model, respectively. The dark and light lines show the median for the two models.

The left two panels show that under the standard prior all three models deliver inflation persistence, although quantitatively the median autocorrelation for the No Indexation model (0.72) is lower than for the Benchmark model (0.86) and for the Full Indexation model (0.93). The right two panels show that under the dummy observations prior the difference between the No Indexation and the Benchmark model in terms of inflation autocorrelation virtually disappears, while the Full Indexation model still predicts slightly higher autocorrelation than the other two models. Interestingly, under the standard prior all three models deliver a negative contemporaneous correlation between inflation and the labor share, in contrast with the prediction of the Phillips curve. Quantitatively the correlation is the more negative the higher the degree of indexation. The dummy observations prior ameliorates this problem. For the Benchmark and the No Indexation models roughly half of the draws in the right hand panel display a positive correlation, while the median is still slightly negative for the Full Indexation model.

The reason why the standard prior delivers a negative contemporaneous correlation between inflation and the labor share is that under this prior mark-up shocks drive most of the co-movement between inflation and the labor share. Recall that in this model the labor share coincides with marginal costs. The mark-up shock $\lambda_{f,t}$ is an important determinant of marginal cost – in fact, the *o*nly determinant in absence of nominal rigidities. From Equation (50) it is also clear that mark-up shocks directly impact inflation. A positive mark-up shock raises inflation and lowers marginal costs, thereby creating a negative correlation between the two.

The standard prior generates by construction reasonable magnitudes for the standard deviations and autocorrelations of the endogenous variables in the Benchmark model, yet generates this counterfactual implication for the cross-sectional moments. The Dummy Observation prior recognizes the fact that this implication is counterfactual also in the pre-sample data, which is used to generate the dummy observations, and modifies the prior for the parameters of the exogenous shocks accordingly. Table 6 summarizes the prior distributions for the shock parameters. The table shows that a key difference between the standard and the dummy observations prior consists in the AR(1) parameters for the $\lambda_{f,t}$ shock. Under the dummy observations prior the standard deviation of $\lambda_{f,t}$ is substantially lower, which implies that these shocks become less important in driving the co-movements between inflation and the labor share under this prior. At the same time however the persistence of the shock increases in all three models. As we have seen in the previous section, mark-up shocks are important in delivering persistence in the labor share. As their standard deviation decreases, their autocorrelation needs to increase to keep the persistence of the labor share roughly unchanged. The autocorrelation of $\lambda_{f,t}$ is somewhat smaller under Full Indexation, which is consistent with the conjecture that the data can be matched with a model without indexation but a fairly persistent mark-up shock.

Figure 7 depicts marginal likelihood differentials relative to the benchmark model. As clear from the discussion in this section, the use of the dummy observations prior narrows the gap between the models. In fact, with $T^* = 10$ dummy observations the No Indexation and the Benchmark model have roughly the same marginal likelihood. The Full Indexation model still delivers a lower marginal likelihood than the other two models even after the correction, although the gap is far less than that obtained for the Flexible Wages and Flexible Wages and Prices models.

6 Conclusion

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.

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.

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

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.

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.

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.

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

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 – Prior Distributions

Name	Domain		Prior 1		Prior 2	
		Density	Para (1)	Para (2)	Para (1)	Para (2)
α	$I\!R^+$	Gamma	2.00	0.10	2.00	0.10
ρ	$[0,1)$	Beta	0.50	0.05	0.73	0.10
σ	$I\!R^+$	InvGamma	1.00	4.00	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: Example – Log Marginal Likelihoods

Specification	$\ln p(Y)$
Model $\mathcal{M}_1$, Prior 1	-105.93
Model $\mathcal{M}_2$, Prior 1	-123.53
Model $\mathcal{M}_2$, Prior 2	-105.70
Model $\mathcal{M}_1$, Prior 3	-108.93
Model $\mathcal{M}_2$, Prior 3	-108.24

Notes: We truncate the prior distribution of α, ρ, σ at the boundary of the indeterminacy region. The marginal likelihoods have been adjusted accordingly.

Table 3: 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 4: 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 autocorrelation (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 5: Prior for Shock Parameters - Benchmark vs. Flex Prices / Wages Models

	Standard Prior		Dummy Obs. Prior Baseline (106)		Dummy Obs. Prior Flex. Wages (105)		Dummy Obs. Prior Flex. Wages, Prices (104)	
	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev
ρ_z	0.400	0.250	0.489	0.129	0.332	0.118	0.326	0.122
ρ_ϕ	0.750	0.250	0.692	0.194	0.769	0.199	0.586	0.338
ρ_{λ_f}	0.750	0.250	0.843	0.120	0.799	0.146	0.884	0.067
ρ_g	0.750	0.250	0.597	0.278	0.840	0.204	0.922	0.141
σ_z	0.376	0.194	1.549	0.388	1.613	0.371	1.667	0.405
σ_ϕ	3.755	1.955	5.392	2.646	1.901	0.783	1.832	0.918
σ_{λ_f}	0.376	0.194	0.191	0.086	0.230	0.084	0.732	0.172
σ_g	0.626	0.323	0.577	0.204	0.789	0.406	0.822	0.320
σ_r	0.250	0.130	0.398	0.115	0.410	0.101	0.414	0.132

Notes: StdDev denotes standard deviation. The support for the distributions of the autocorrelation (standard deviation) parameters is $[0, 1)$ ($\mathbb{R}^+$). See Table 4 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 6: Priors for Shock Parameters – Benchmark vs. Restricted Indexation Models

	Standard Prior		Dummy Obs. Prior Baseline (106)		Dummy Obs. Prior No Indexation (101)		Dummy Obs. Prior Full Indexation (107)	
	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev
ρ_z	0.400	0.250	0.489	0.129	0.490	0.125	0.459	0.127
ρ_ϕ	0.750	0.250	0.692	0.194	0.688	0.204	0.614	0.235
ρ_{λ_f}	0.750	0.250	0.843	0.120	0.872	0.089	0.838	0.130
ρ_g	0.750	0.250	0.597	0.278	0.625	0.287	0.573	0.280
σ_z	0.376	0.194	1.549	0.388	1.628	0.393	1.724	0.499
σ_ϕ	3.755	1.955	5.392	2.646	5.289	2.837	7.252	4.844
σ_{λ_f}	0.376	0.194	0.191	0.086	0.157	0.056	0.209	0.079
σ_g	0.626	0.323	0.577	0.204	0.570	0.241	0.543	0.195
σ_r	0.250	0.130	0.398	0.115	0.398	0.109	0.429	0.109

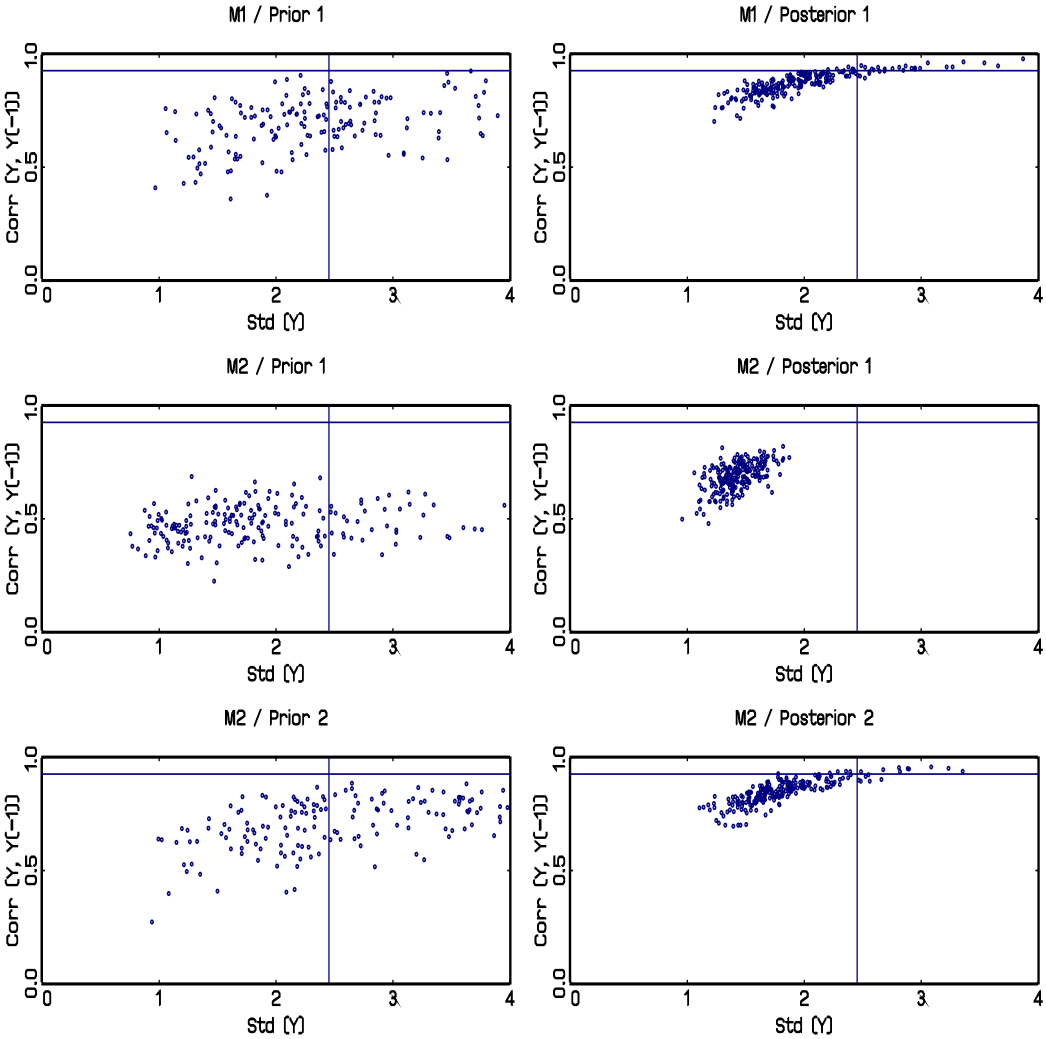
Notes: StdDev denotes standard deviation. The support for the distributions of the autocorrelation (standard deviation) parameters is $[0, 1)$ ($\mathbb{R}^+$). See Table 4 for the marginal densities of the benchmark prior and the (improper) prior that is used to pre-multiply the quasi-likelihood function for the dummy observations. The results are based on $T^* = 10$.

Table 7: Log Marginal Likelihoods $\ln p(Y)$

Specification	Standard	Dummy Obs. Prior	
	Prior	$T^* = 4$	$T^* = 10$
Baseline (106)	-611.95	-611.02	-614.31
Flexible Wages (105)	-635.18	-624.53	-631.05
Flexible Wages and Prices (104)	-677.31	-664.46	-667.07
No Indexation (101)	-612.58	-611.56	-614.53
Full Indexation (107)	-620.34	-616.06	-619.42

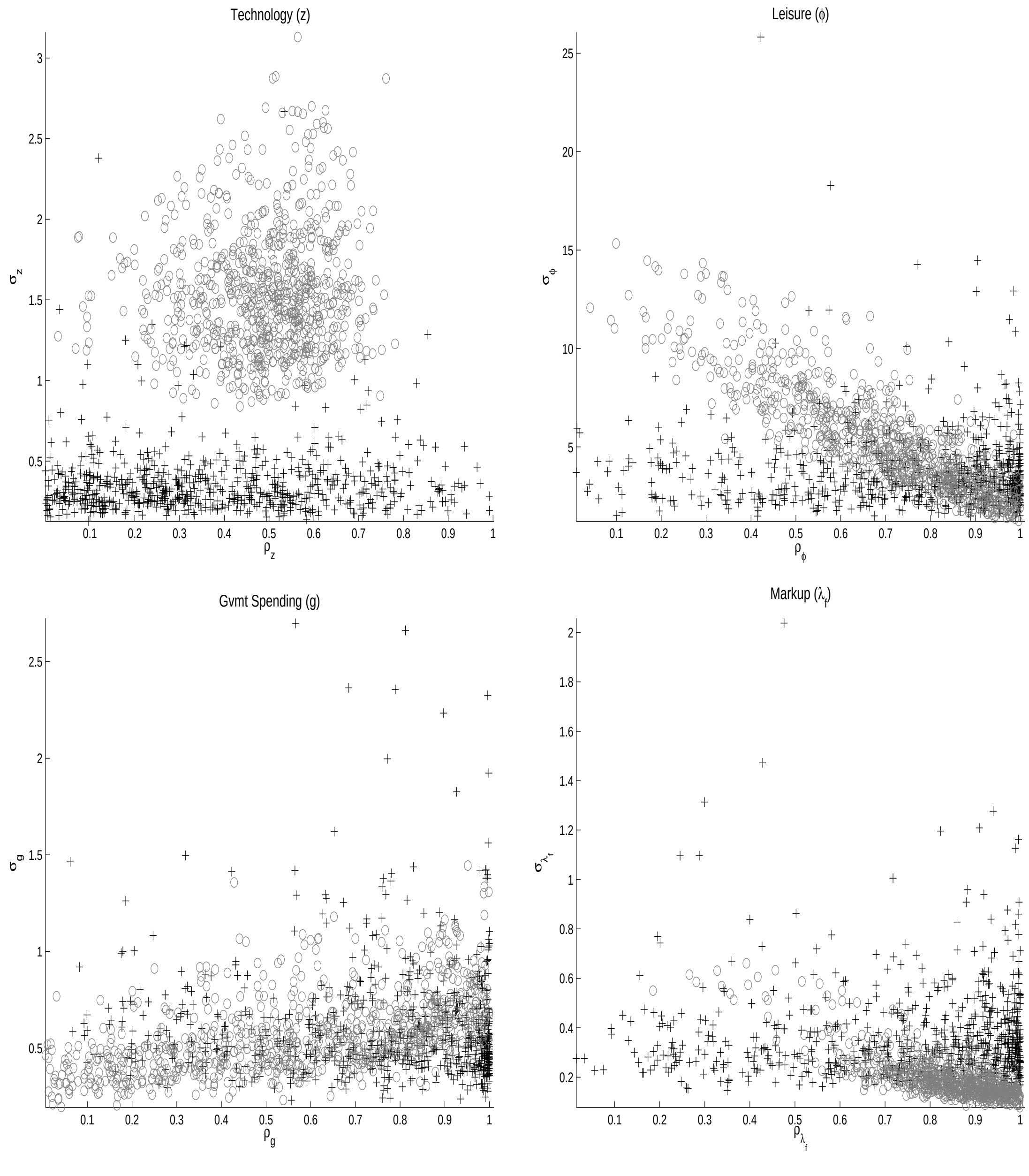
Notes: The marginal data densities are computed based on quarterly U.S. data ranging from QI:1981 to QIV:2005.

Figure 1: Example - Predictive Distributions of Sample Moments



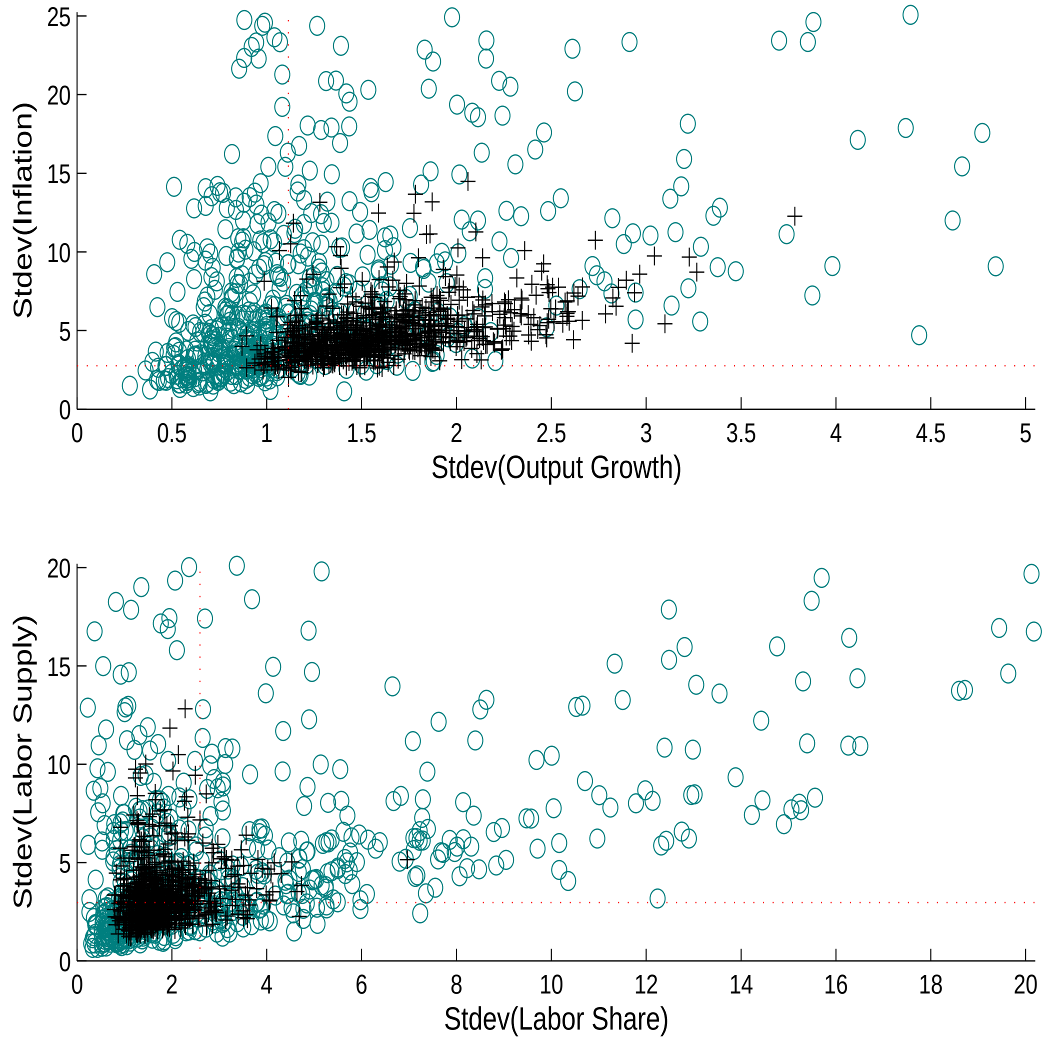
Notes: Each panel depicts 200 draws from predictive distribution for the sample autocorrelation and standard deviation. Intersection of solid lines signifies the actual sample moment.

Figure 2: Priors for Benchmark Model – Shock Parameters



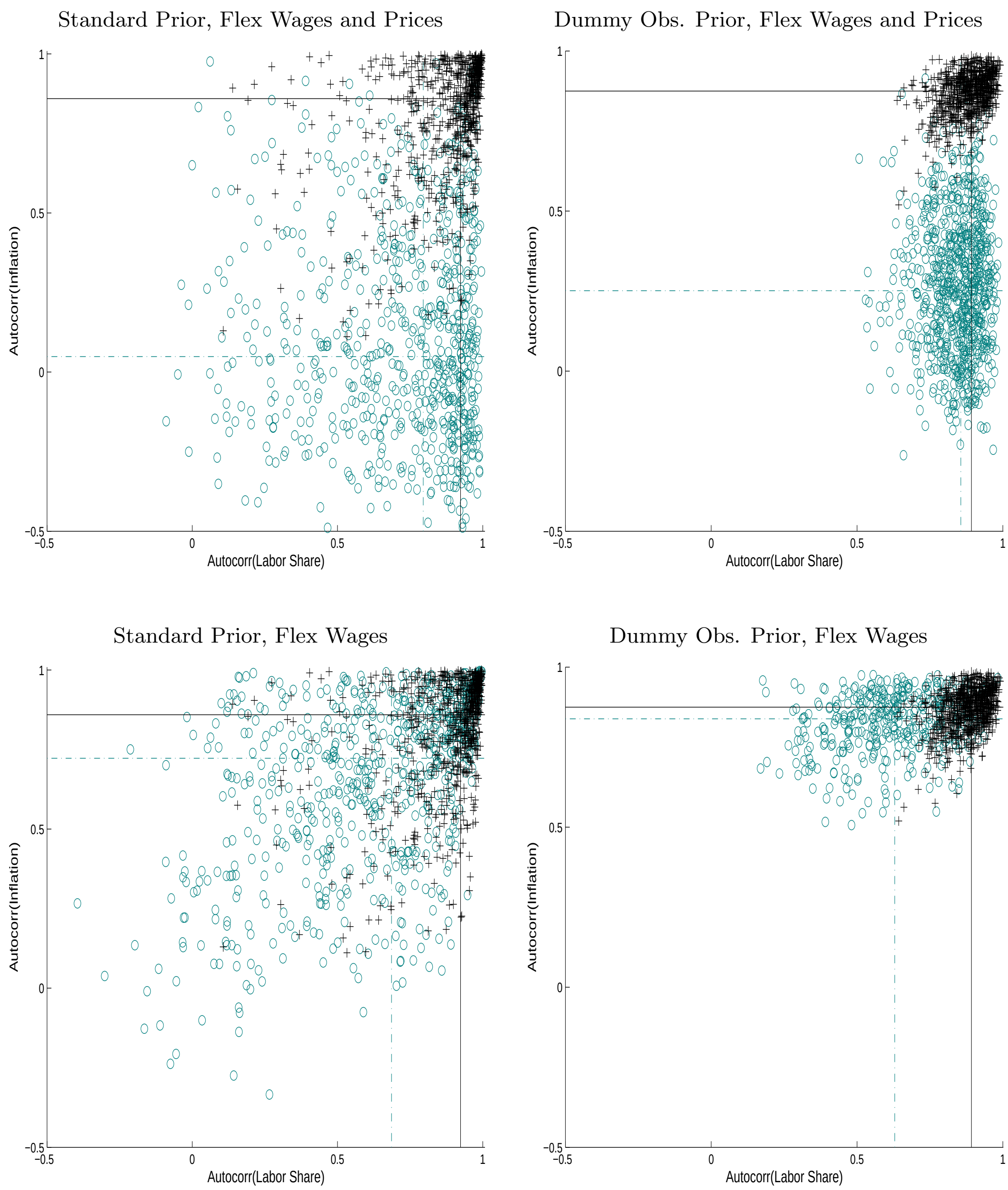
Notes: Each panel depicts draws from the prior distribution of the shock parameters. Black crosses indicate draws from the standard prior, whereas grey circles correspond to draws from the dummy observations prior.

Figure 3: Priors for Benchmark Model - Sample Moments



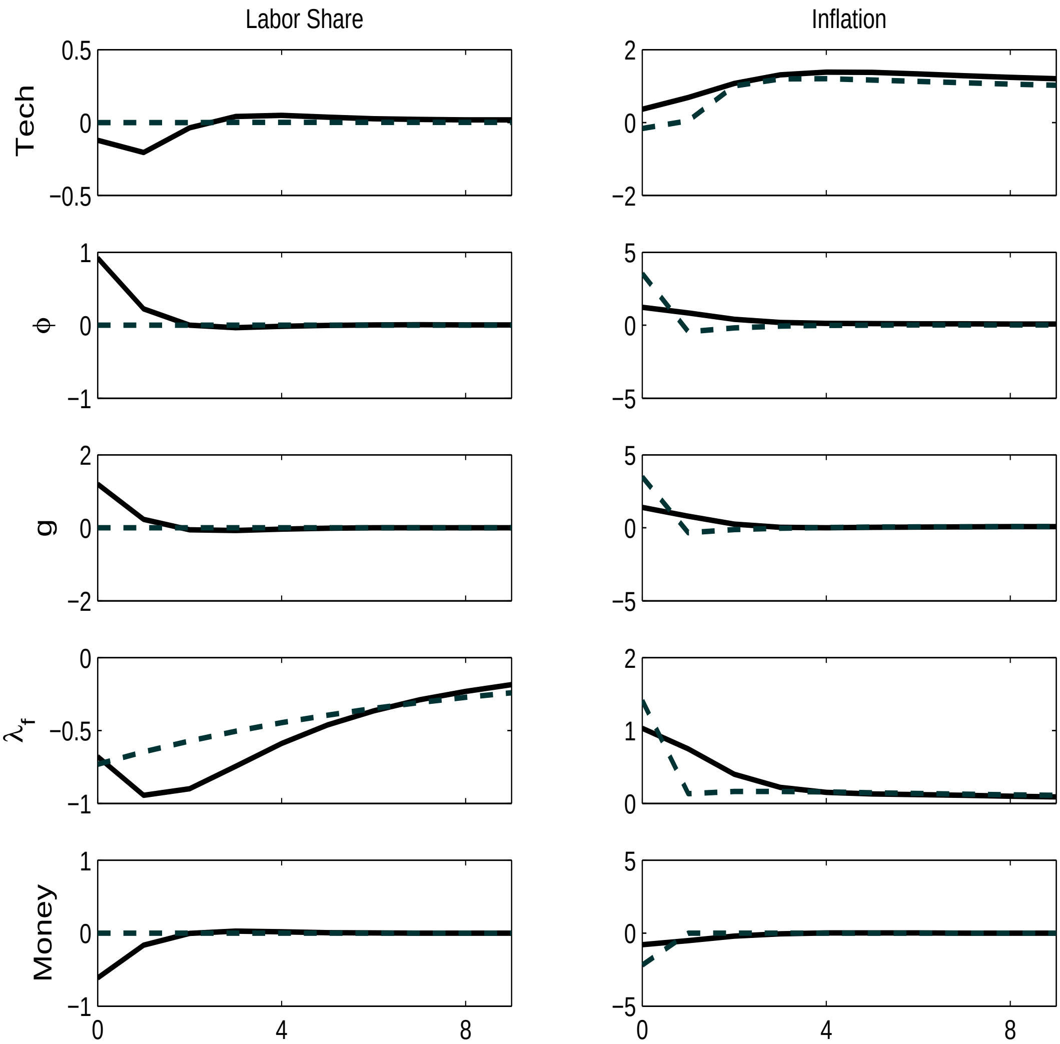
Notes: Each panel depicts draws from the prior predictive distribution of various sample standard deviations, calculated based on 100 artificial observations from the DSGE model. The intersection of the red dotted lines signifies the sample standard deviations computed from the pre-sample that is used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

Figure 4: Nominal Rigidities: Benchmark versus Flex Wages / Prices Model



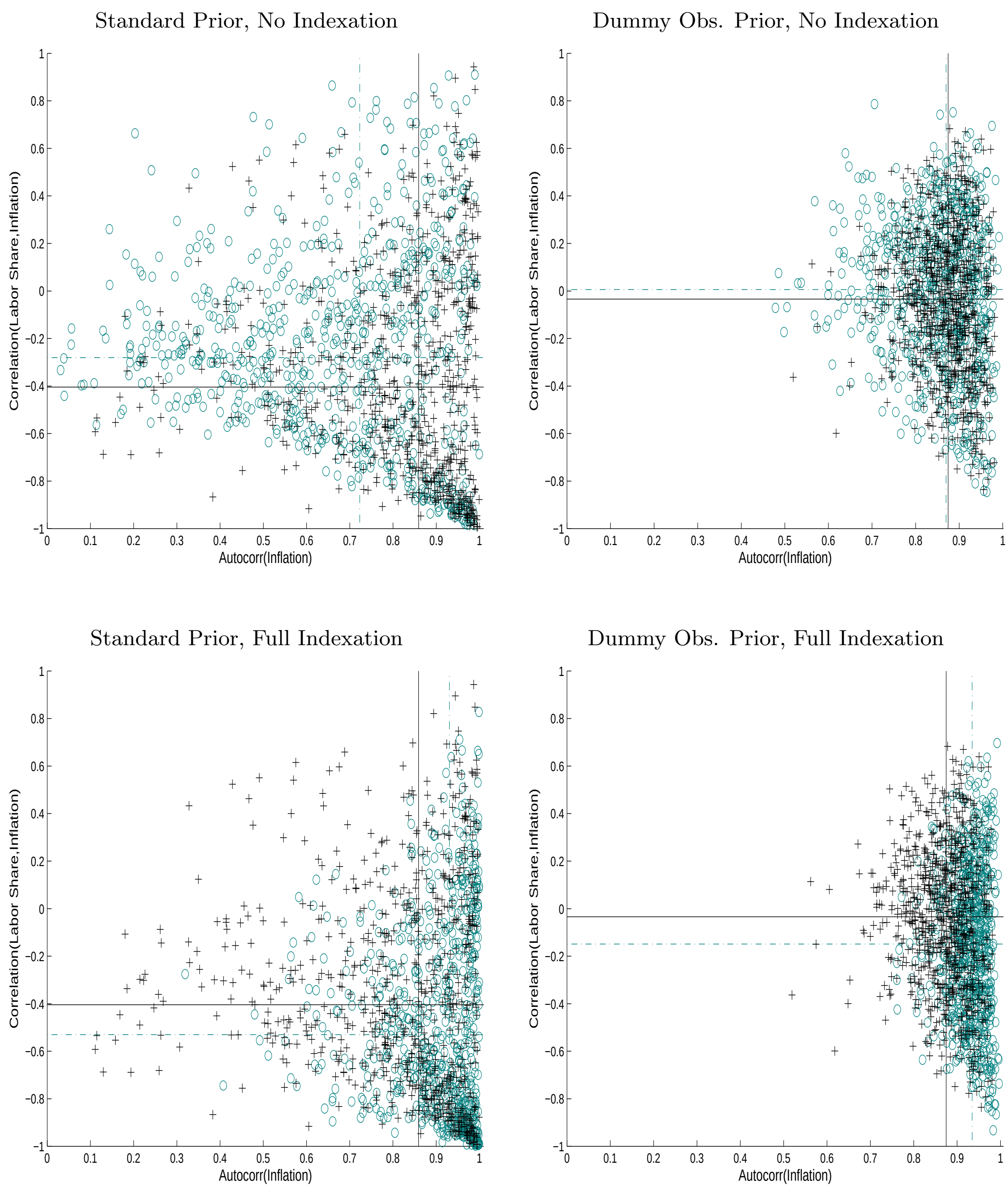
Notes: Each panel depicts draws from the prior predictive distribution of the autocorrelation of inflation and the labor share, calculated based on 100 artificial observations from the DSGE model. Black cross correspond to draws from the Benchmark model, whereas green circles denote draws from the flexible price / wage models. The intersection of the solid black and dashed green lines signifies the median of the prior predictive distributions.

Figure 5: Impulse Response Functions - Flexible Wages vs. Flexible Wages and Prices



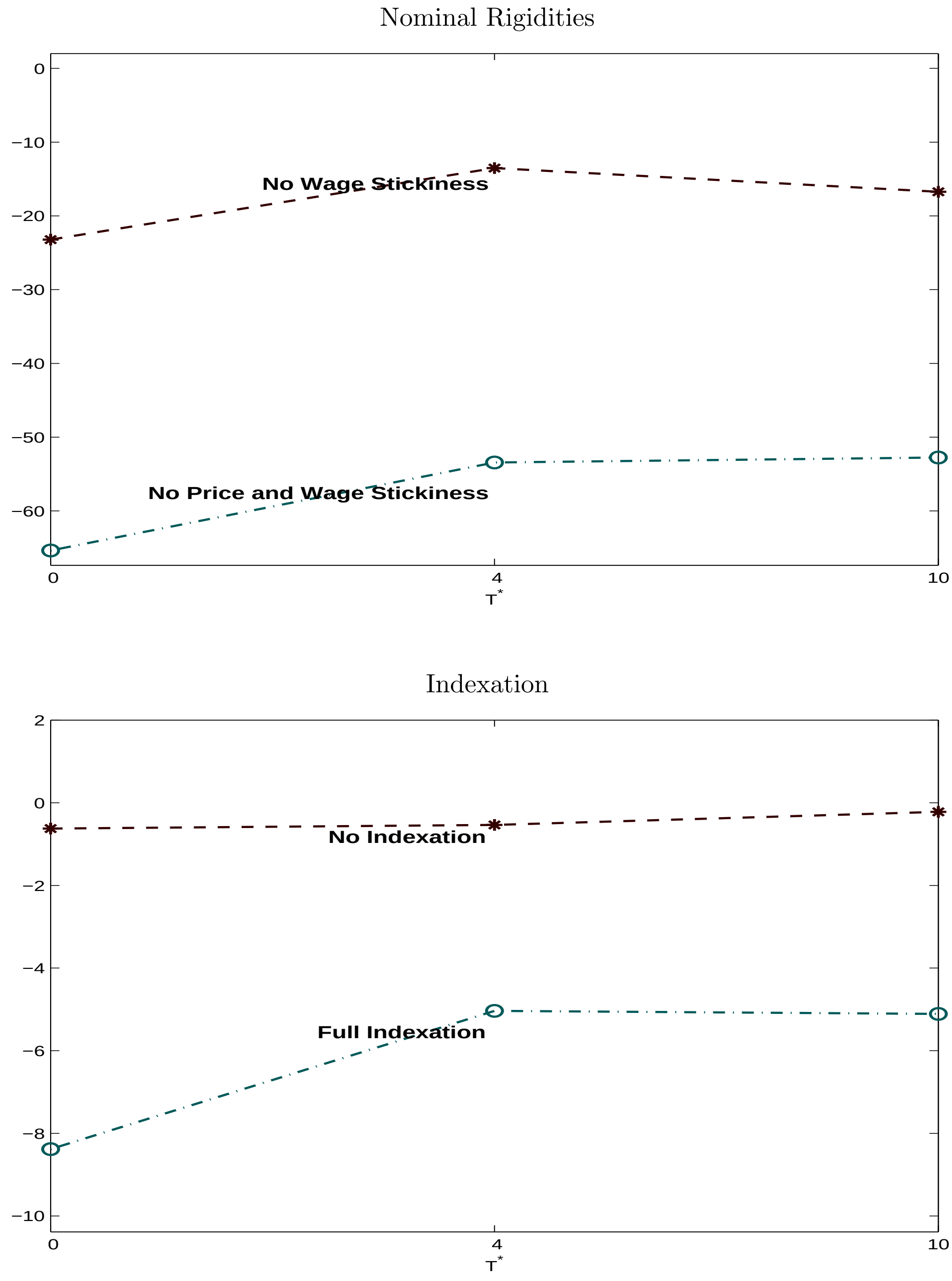
Notes: The left (right) column depicts prior mean responses of the labor share (inflation) to the five structural shocks. Solid lines correspond to flexible wage model, dashed lines signify responses of the flexible wage and price model.

Figure 6: Phillips Curve Relationships: Benchmark versus Restricted Indexation



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 green circles denote draws from the flexible price / wage models. The intersection of the solid black and dashed green lines signifies the median of the prior predictive distributions.

Figure 7: MARGINAL LIKELIHOODS RELATIVE TO BENCHMARK



Notes: Each panel depicts log marginal likelihood differences with respect to the benchmark specification. Negative values indicate that the benchmark model attains a higher marginal likelihood. "0" refers to the standard prior, whereas "4" and "10" refer to the dummy observations prior with $T^* = 4$ and $T^* = 10$, respectively.