Identifying the Sources of Model Misspecification

Atsushi Inoue[☆] Vanderbilt University

Chun-Hung Kuo[☆] International Univ. of Japan

Barbara Rossi[☆] ICREA-Univ. Pompeu Fabra, Barcelona GSE and CREI

Abstract

In this paper we propose an empirical method for detecting and identifying misspecification in structural economic models. Our approach formalizes the common practice of adding "shocks" in the model, and identifies potential misspecification via forecast error variance decomposition and marginal likelihood analyses. The simulation results based on a small-scale DSGE model demonstrate that our method can correctly identify the source of misspecification. Our empirical results show that state-of-the-art medium-scale New Keynesian DSGE models remain misspecified, pointing to asset and labor markets as the sources of the misspecification.

Keywords: DSGE models, marginal likelihood, misspecification

1. Introduction

The recent financial crisis of 2007-2009 uncovered existing structural economic models' difficulties in explaining the data. The limitations of structural models in forecasting are also well known. For example, Edge and Gurkaynak (2010), among others, have shown that the forecast performance of the Smets and Wouters (2007) model is not better than a naïve constant growth-rate model during the Great Moderation period. Moreover, the

[☆]We thank Ricardo Reis, G. Ganics, F. Loria and three anonymous referees for their constructive comments and suggestions. The first and third authors acknowledge financial support from the National Science Foundation through grants 102159 and 1022125, respectively. The authors thank seminar participants at the Applied Time Series Econometrics Workshop at the Federal Reserve Bank of St. Louis, Chung-Hua Institution for Economic Research, International University of Japan, National Taipei University as well as the 2013 EC(2), the Triangle Econometrics and the 2017 ASSA Conferences for helpful comments on earlier versions of the paper. Barbara Rossi gratefully acknowledges support from the European Research Council grant No 615608.

severity and the prolonged duration of the Great Recession have challenged the adequacy of existing predictors and raised the possibility that these models might be misspecified. For example, Ng and Wright (2013) have argued that the features of "financial-crisis-induced" recessions (such as the Great Recession) are distinct from those of the "typical" recessions driven by supply or monetary policy shocks. This distinction may explain why the study of the Great Recession requires alternative models and predictors.

While the above limitations of the existing models are well known, it is unclear whether the recent financial crisis resulted from unexpected shocks or changes in the transmission mechanism. On the one hand, Stock and Watson (2012) argue that the transmission mechanism during the Great Recession was not different than that of any other post-war recession, showing that the larger shocks hitting the economy were the origin of the deep and prolonged recession. On the other hand, other researchers emphasize that existing macroeconomic models do not fully capture the mechanisms behind the Great Recession and argue that substantial modifications are necessary. For example, in an effort to improve the fit of structural economic models during the crisis, Del Negro and Schorfheide (2014) include information from inflation expectations, financial frictions and interest rate spreads: if this is the case, knowing the source of model misspecification will help economists develop alternative models. Therefore, in this paper, we examine the empirical importance of model misspecification in dynamic stochastic general equilibrium (DSGE) models and propose a methodology to shed light on the sources of misspecification responsible for their poor forecasting performance.

To examine misspecification in a structural model, we propose to broaden the treatment of the exogenous processes. Specifically, we consider two types of exogenous processes. The first type of exogenous processes represent structural shocks and are interpreted in the conventional way. Since these shocks are structural, they should be viewed as indispensable ingredients of the model at hand. In contrast, the second type of exogenous processes, which we call "margins", are not structural and their function is to check for model misspecification. In fact, by estimating the margins, we are able to assess: (i) where the misspecification might be located (that is, which parts of the model are most affected by the misspecification); and (ii) how qualitatively important it is.

While the technique that we propose is very general, in this paper we focus on DSGE models given their widespread use in academia and central banks as standard tools for analyzing macroeconomic policies. To provide more intuition and better illustrate our framework, we consider a medium-scale DSGE model, embedded with most New Keynesian features. This model is mildly misspecified in the sense that: (i) the cross-equation and equilibrium restrictions imposed by the model do not exactly hold in every time period; and (ii) the deviations from equilibrium are zero on average. We assume that when solving their optimization problems, the economic agents in the model (both firms and households) take into account the exogenous margin processes, which allow for deviations from equilibrium conditions. We interpret the variances of the margin processes as a measure of the degree of misspecification of the model. To examine where and how large the misspecification is, we conduct forecast error variance decompositions (FEVDs) and marginal likelihood comparisons.

Our empirical application to a state-of-the-art DSGE model highlights two interesting findings. First, our technique points to misspecification in the model's labor demand component. Second, bond markets also show evidence of misspecification, which is persistent in nature. Our findings confirm the existing view that asset and labor markets in the New Keynesian DSGE model are misspecified and suggest that further work in these areas would be beneficial.[4]

Our method is related to several recent contributions. First, our paper is related to Sargent (1989), Ireland (2004), Del Negro and Schorfheide (2009) and Curdia and Reis (2010), among others. Sargent (1989) and Ireland (2004) introduce errors in the measurement equations of the state space version of the model to assess whether the model is misspecified. Del Negro and Schorfheide (2009) develop a framework for Bayesian estima-

[4]Our findings are related to Del Negro, Schorfheide, Smets and Wouters (2007) who, among others, have recently showed that model misspecification cannot be ignored in policy analyses.

tion of possibly misspecified DSGE models by using DSGE-implied parameters as priors for vector autoregressive (VAR) models. Their framework allows for model misspecification and produces the posterior distribution of structural parameters as well as the posterior structural impulse responses based on DSGE priors.[5] Our framework complements these works in that we can identify which parts of the model are misspecified. Curdia and Reis (2010) relax the restriction that exogenous disturbances in structural models are independent processes. They argue that estimating models with correlated disturbances provides a useful check for model misspecification. We consider instead a different way to incorporate misspecification in the models by including independent disturbances in the equilibrium conditions of the model. We also suggest that the analysis of the estimated disturbances provides a way to identify the source of the misspecification and its importance over time. Note also that Watson (1993) proposed a measure of fit for calibrated models; we instead estimate the structural models and provide a different methodology to identify the sources of misspecification. This paper is also related to the literature on the challenges in the estimation of structural macroeconomic models. The challenge we focus on is misspecification, which has been considered by Corradi and Swanson (2007) and Canova and Ferroni (2011). Canova and Ferroni (2011) show that incorrect filtering results in model misspecification but, unlike us, do not search for the source of misspecification. Corradi and Swanson (2007) provide new tools for comparing the empirical joint distribution of historical time series with the empirical distribution of simulated time series based on structural macroeconomic models. Their focus is on detecting whether the whole distribution of a macroeconomic model is correctly specified, whereas our focus is on the first moments of the model and on providing guidance on the sources of model misspecification.

Other challenges that researchers face in the estimation of structural macroeconomic models are related to identification. This paper does not deal with potential lack of identi-

[5] They assume that DSGE models have a VAR representation in their implementation. Many DSGE models have VARMA representations and we do not need to assume a VAR representation although the error due to VAR approximations may be small if the lag is sufficiently large.

fication or weak identification of the models' parameters. That is a different problem and has been analyzed by Dufour, Khalaf and Kichian (2013), Canova and Sala (2009), and Iskrev (2010), among others. If the models' parameters are not identified, then adding margins will not help. Indeed, it is possible that adding margins to an unidentified or weakly identified model may make the estimation more difficult.

Finally, our paper is related to Chari et al. (2007) and Brinca et al. (2016), who introduce time-varying "wedges" to account for deviations between the time series of the observables implied by the model and those actually observed in the data. There is a fundamental difference and two more minor differences between our work and theirs. The fundamental difference is that they focus on business cycle accounting while we focus on detecting model misspecification. In other words, Chari et al. view the model as correctly specified, and the (correlated) wedges are frictions introduced in the model to account for business cycle fluctuations similar to those in the data. Our approach is substantially and philosophically different: we view the margins as a measure of model misspecification, and each margin independently measures the degree of misspecification in a specific part of the model (e.g. an equation). The other two more minor differences are as follows. One is that they consider a neoclassical stochastic growth benchmark model, while we consider a New Keynesian DSGE benchmark model which incorporates several frictions. Thus, our margins (i.e. distortions due to misspecification) reflect model misspecification that is not already accounted for by frictions built into the model. The other difference is that their analysis is based on calibrated parameter values, while ours is based on an estimated model, and thus is robust to incorrectly calibrating the parameter values.

2. An Illustrative Example

In this section, we focus on a simple consumption model to illustrate our method and compare it with those existing in the literature.

Suppose that we want to evaluate the following baseline model:

$$\max_{a_t, c_t} \; E_0[\sum_{t=0}^{\infty} \beta^t (\theta_0 c_t - \frac{\theta_1}{2} c_t^2)] \tag{1}$$

$$s.t. \; a_{t+1} = (1+r)(a_t + y_t - c_t), \tag{2}$$

$$y_t \;=\; y_{t-1} + \varepsilon_t, \tag{3}$$

where $a > 0$, $b > 0$, $\beta(1+r) = 1$, and ε_t is an independent and identically distributed (iid) mean zero random variable with variance σ_ε^2: $\varepsilon_t \overset{iid}{\sim} (0, \sigma_\varepsilon^2)$. We assume that the econometrician observes consumption (c_t), income (y_t) and assets (a_t).

The baseline model, eqs. (1)-(3), is potentially misspecified. We consider separately three different types of misspecification. In the first example, transitory income is present in the true data generating process: $y_t^T \;=\; \rho_{y^T} y_{t-1}^T + \varepsilon_{y^T,t}$, where $|\rho_{y^T}| < 1$ and $\varepsilon_{y^T,t} \overset{iid}{\sim} (0, \sigma_y^2)$. The solution can be described by:

$$c_t \;=\; \frac{r}{r+1} a_t + y_t^P + \frac{r}{1 - \rho_{y^T} + r} y_t^T, \tag{4}$$

$$a_{t+1} \;=\; (1+r)(a_t + y_t - c_t), \tag{5}$$

$$y_t^P \;=\; y_{t-1}^P + \varepsilon_{y^P,t}, \;\; y_t^T \;=\; \rho_{y^T} y_{t-1}^T + \varepsilon_{y^T,t}, \;\; y_t \;=\; y_t^P + y_t^T. \tag{6}$$

In the second example, asset returns are uncertain and stochastic in the true data generating process, so that: $a_{t+1} \;=\; (1 + r_{t+1})(a_t + y_t - c_t)$, where r_t is *iid* with $E(1 + r_{t+1}) = 1/\beta$. Then the solution can be written as:

$$c_t \;=\; \left(1 - \frac{1}{\kappa}\right) a_t + y_t, \tag{7}$$

$$a_{t+1} \;=\; (1 + r_{t+1})(a_t + y_t - c_t), \tag{8}$$

$$y_t \;=\; y_{t-1} + \varepsilon_t, \tag{9}$$

where $\kappa = \beta E[(1 + r_{t+1})^2]$.

In the third example, the asset data available to the econometrician contain measurement error: $\tilde{a}_t = a_t + \xi_t$, where $\xi_t = \rho_\xi \xi_{t-1} + \eta_{\xi,t}$, $|\rho_\xi| < 1$ and $\eta_{\xi,t} \overset{iid}{\sim} (0, \sigma_\eta^2)$. While the solution remains the same as in the baseline case, the econometrician erroneously fits

$$c_t \;=\; \frac{r}{r+1} \tilde{a}_t + y_t \;=\; \frac{r}{r+1} a_t + y_t + \frac{r}{1+r} \xi_t, \tag{10}$$

and

$$\tilde{a}_{t+1} = (1+r)(\tilde{a}_t + y_t - c_t), \tag{11}$$

which is equivalent to

$$a_{t+1} = (1+r)(a_t + y_t - c_t) + (1+r)\xi_t - \xi_{t+1}. \tag{12}$$

2.1. Our Approach

Our approach to detect misspecification introduces "margins" in the optimization problem. We see the flexibility that characterizes our methodology in terms of choosing the location and the number of margins as a strength of our approach; in fact, the researcher can introduce as many margins as he/she wants. As we show in this example, this is not completely arbitrary, as we discipline the way we introduce margins by using the marginal likelihood; in fact, unnecessary margins will be eliminated by the marginal likelihood criterion that we suggest. The marginal likelihood has a built-in term that penalizes over-parameterized models (Fernandez-Villaverde and Rubio-Ramirez, 2004). A model with an unnecessary margin will have a lower marginal likelihood value than a model without it.

To obtain a closed form solution that can be compared with eqs. (4)–(12), we introduce three margins, u_t, v_t and w_t, in the baseline model:

$$\begin{array}{ll} \max_{a_t,c_t} & E_0\{\sum_{t=0}^{\infty} \beta^t[(a+u_t)c_t - \frac{b}{2}c_t^2]\} \\ \text{subject to} & a_{t+1} = (1+r)(1+v_{t+1})(a_t + y_t - c_t + w_t), \\ & y_t \; = \; y_{t-1} + \varepsilon_t, \\ & u_t \; = \; \rho_u u_{t-1} + \eta_{u,t}, \\ & v_t \; = \; \eta_{v,t}, \\ & w_t \; = \; \rho_w w_{t-1} + \eta_{w,t}, \end{array}$$

where $\beta(1+r) = 1$ and

$$\begin{bmatrix} \varepsilon_t \\ \eta_{u,t} \\ \eta_{v,t} \\ \eta_{w,t} \end{bmatrix} \overset{iid}{\sim} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & 0 & 0 & 0 \\ 0 & \sigma_u^2 & 0 & 0 \\ 0 & 0 & \sigma_v^2 & 0 \\ 0 & 0 & 0 & \sigma_w^2 \end{bmatrix} \right).$$

u_t is placed in the linear part of the utility function and v_t is assumed to be iid for analytical tractability. Alternatively, one could add a margin to the utility function, i.e., $\theta_0 c_t - (\theta_1/2)c_t^2 + x_t$. However, this margin would not show up anywhere in the solution and, hence, would be of little help in identifying the sources of misspecification in this model. This observation highlights the importance that a researcher thinks carefully about how to add the margins. We view this as a strength of our approach: margins should be included in a meaningful way.

The solution to this optimization problem is given by

$$\begin{aligned}
c_t &= \left(1 - \frac{1}{\phi}\right) a_t + y_t + \frac{1-\rho_u}{b(\phi-\rho_u)} u_t + \frac{\phi - 1}{\phi - \rho_w} w_t, && (13)\\
a_{t+1} &= (1+r)(1+v_{t+1})(a_t + y_t - c_t + w_t), && (14)\\
y_t &= y_{t-1} + \varepsilon_t, && (15)\\
u_t &= \rho_u u_{t-1} + \eta_{u,t}, && (16)\\
v_t &= \eta_{v,t}, && (17)\\
w_t &= \rho_w w_{t-1} + \eta_{w,t}, && (18)
\end{aligned}$$

where $\phi \equiv (1+r)E[(1+v_{t+1})^2]$.

2.2. How and Why Does Our Approach Work?

Suppose that one fits the above model (eqs. 13-18) when there is transitory income. Because the asset return is constant, v_{t+1} should be (close to) zero. So $\phi = 1 + r$ and $(1-1/\phi) = r/(1+r)$. Furthermore, by comparing eqs. (5) and (14), the term w_t should be close to zero as well. Comparing (13) and (4), in the true model there is a term due to the presence of the transitory income component, $(r/(1-\rho_y^T + r))y_t^T$. In the fitted model, the total income is used as the permanent income and thus we have $y_t^P + y_t^T$. The difference $[(1-\rho_y^T)/(1-\rho_y^T + r)]y_t^T$ should be absorbed by either u_t or w_t on the right-hand side of (13), but w_t is (close to) zero. Thus only u_t will be different from zero, significantly contributing to the FEVDs and successfully capturing the misspecification. Alternatively, comparing the marginal likelihood of models where each margin is removed one-at-a-time will also show that the likelihood decreases the most when u_t is removed while it is virtually unchanged when either w_t or v_t is removed.

In the second case, v_t will be significant because $1+r_{t+1} = (1+r)(1+v_{t+1})$. Once this equation is satisfied, the model is correctly specified and thus u_t and w_t will be insignificant. Hence, again, FEVDs and marginal likelihood analyses will correctly signal the source of the misspecification.

In the last case, the margin in the budget constraint, w_t, will capture the AR(1) measurement error that appears in both the consumption equation and budget constraint; however, given the way the serially correlated measurement error appears in eqs. (10)-(12), the margin in the preference parameter, u_t, might also be significant to pick up any remaining misspecification. v_t will be insignificant since v_t would yield a non-stationary error in the budget constraint that is different from the MA(1) error in equation (12). Either way, the misspecification will be successfully differentiated from the previous two cases.

Therefore, our method successfully distinguishes among the three types of misspecification: (i) misspecification due to erroneous serial correlation structures; (ii) misspecification due to the incorrect assumption of constant asset returns; and (iii) misspecification due to measurement error.

The margins should not be confused with structural shocks. In fact, there is not necessarily a one-to-one mapping between margins and shocks. As this example shows, in the first case the margin included the utility function (u_t) captures the absence of transitory income in the model but is not a preference/utility shock. However, if a researcher were interested in distinguishing between the two, he/she could estimate two models: a model with transitory income and without a preference shock, and a model with a preference shock but without transitory income; the marginal likelihood can distinguish between the two models since they have different implications for the income processes.

2.3. Alternative Approaches

Sargent (1989) and Ireland's (2004) approach: Sargent (1989) and Ireland (2004) introduce (possibly serially correlated) errors in measurement equations of state space models. Although our baseline model is not a state space model, their idea can still be applied as follows:

$$
\begin{aligned}
c_t &= \frac{r}{r+1}a_t + y_t + e_{1t}, && (19)\\
a_{t+1} &= (1+r)(a_t + y_t - c_t) + e_{2t}, && (20)\\
e_{1t} &= \rho_{e_1} e_{1,t-1} + \eta_{1t}, && (21)\\
e_{2t} &= \rho_{e_2} e_{2,t-1} + \eta_{2t}, && (22)
\end{aligned}
$$

where $[\eta_{1t}\ \eta_{2t}]' \overset{iid}{\sim} N(0, \Sigma)$.

In the first example, e_{1t} will be significant, but e_{2t} will not, since only the consumption equation is misspecified. In the last two examples, both e_{1t} and e_{2t} will have non-degenerate distributions because the two equations are both misspecified. Therefore their method cannot tell apart the source of the misspecification in the second and third examples.

Del Negro and Schorfheide's (2009) approach: In their (first) approach, they generalize the dependence structure. The only shock in our benchmark model is the shock to the permanent income process. To make the solution analytically tractable, we increase the dependence in the income process (which is equivalent to including a moving average in the error term) and consider the following generalization of the permanent income process: $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$, where $\phi_1 + \phi_2 = 1$. The solution is:

$$
\begin{aligned}
c_t &= \frac{r}{r+1}a_t + \frac{r(1+r)}{(1-\phi_1+r)(1+r)-\phi_2}y_t + \frac{\phi_2 r}{(1-\phi_1+r)(1+r)-\phi_2}y_{t-1}, && (23)\\
a_{t+1} &= (1+r)(a_t + y_t - c_t), && (24)\\
y_t &= \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t. && (25)
\end{aligned}
$$

Because the consumption equation is misspecified in all the examples, the error in the consumption equation will be captured by the more general structure of y_t, i.e., $\phi_2 \neq 0$. However, a researcher implementing Del Negro and Schorfheide's (2009) approach would be unable to identify the source of misspecification, even though their approach would correctly signal that the model is broadly misspecified.

Curdia and Reis' (2010) approach: Since we need two shocks to apply their method, we introduce transitory income and allow the permanent and the transitory income shocks to be correlated. Suppose that the econometrician observes the permanent and transitory income components, and that the latter follows: $y_t^T = \rho y_{t-1}^T + \varepsilon_{y^T,t}$, where

$$\begin{bmatrix} \varepsilon_{y^P,t} \\ \varepsilon_{y^T,t} \end{bmatrix} \overset{iid}{\sim} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{y^P}^2 & \rho\sigma_{y^P}\sigma_{y^T} \\ \rho\sigma_{y^P}\sigma_{y^T} & \sigma_{y^T}^2 \end{bmatrix} \right) \tag{26}$$

and $\varepsilon_{y^P,t}$ is the shock for the random walk permanent income process. A correlation different from zero, i.e., $\rho \neq 0$, indicates the presence of misspecification. Even if the shocks are correlated, the solution remains the same as eqs. (4)-(5):

$$c_t = \frac{r}{r+1}a_t + y_t^P + \frac{r}{1+r-\rho}y_t^T, \tag{27}$$

$$a_{t+1} = (1+r)(a_t + y_t - c_t), \tag{28}$$

$$y_t = y_t^P + y_t^T, \;\; y_t^P = y_{t-1}^P + \varepsilon_{y^P,t}, \;\; y_t^T = \rho y_{t-1}^T + \varepsilon_{y^T,t}. \tag{29}$$

Because the consumption equation will be misspecified in all the examples, the variance of the transitory income component will be positive. In the first example, because the permanent and transitory incomes are independent in the data generating process, the two shocks will be uncorrelated. When the interest rate is stochastic, that will yield persistent and transitory residuals because the asset equation includes both permanent and transitory income. That will likely make the two errors correlated. In the last example, the serially correlated measurement error in the Euler equation will be detected by the transitory income process. Because the measurement error is independent of the permanent income process, ρ will be zero.[6] One would notice that the model is misspecified, however, because the budget constraint is not satisfied. Thus this method can detect misspecification except for the first example. As with the Del Negro and Schorfheide method, however, it is not clear how to find out exactly which building blocks of the model are misspecified.

Chari, Kehoe and McGrattan's (2007) approach: Our approach is fundamentally and philosophically different from the approach developed by Chari, Kehoe and McGrat-

[6]Since $c_t = \frac{r}{r+1}a_t + \frac{r}{r+1}\xi_t + y_t^P + \frac{r}{1+r-\rho}y_t^T$, then $y_t^T \approx -\frac{1+r-\rho}{r+1}\xi_t$, which is uncorrelated with y_t^P.

tan (2007) and further elaborated in Brinca, Chari, Kehoe and McGrattan (2016). Their method is not designed to detect misspecification; rather, to account for business cycle fluctuations. Their estimated wedges are calculated based on the difference between observables and their counterparts obtained from a model evaluated at given parameter values; thus, their wedges are typically correlated to one another and it is difficult to interpret the marginal contribution of a given wedge. In contrast, the exogenous process that we use in our methodology are crucially different objects from wedges: in fact, our margins are independent of each other and are introduced in the agents' optimization conditions. In our framework, FEVD and Marginal likelihood analyses provide statistically sound ways to detect which margins are significant and responsible for the misspecification. If one nevertheless treats their method as a method to detect misspecification, one may detect misspecification when deviations from equilibrium conditions are large. However, since they are evaluated at pre-specified parameter values rather than estimated, one could be mislead to conclude that the model is misspecified even when it is correctly specified if the parameters are not correctly calibrated.

2.4. Main Features and Limitations of Our Methodology

The above example shows that our method nicely complements existing works by Sargent (1989), Ireland (2004), Del Negro and Schorfheide (2009) and Curdia and Reis (2010) since it is able of not only detecting model misspecification but also identifying which parts of the model are misspecified. The example also shows that our method has the potential to differentiate among the three types of misspecification, and hence detect the sources of misspecification, while existing methods fail to do so. In practice, however, the same margin may capture different types of misspecification. Thus, applying our method only once may not uniquely identify a source of misspecification. We suggest repeating this process until no margin is found to be substantial. For example, if the first margin is substantial in the above example, one may replace the baseline model by a model with transitory income and re-estimate the model with additional margins. If the additional margins are not substantial, we suggest stopping there; otherwise, investigate another baseline model and proceed until no margin is substantial. However, in general, there may be situations where two margins could be far from zero even though there is only one source of misspecification. On the other hand, it is not possible to provide a general theorem proving the precise conditions under which our method works, or does not work. The implementation of the method depends on the specific model and empirical application at hand, and it is up to the researcher to design the margins in an insightful and constructive way. If the researcher is willing to consider all possible margins and compare their marginal likelihoods, the choice of the model is not path dependent.

The margins should not be confused with structural shocks. We introduce margins only after "traditional" structural shocks are included. Our goal is to evaluate the correct specification of the model that only includes the structural shocks. That is, we evaluate the misspecification *relative to a benchmark model*: if the benchmark model has shocks that are interpreted as structural shocks, we maintain those, and include margins on top of them. The specific way we incorporate the time-varying margins in the model is by including them into the agents' optimization problem (such as households' budget constraints): for example, the existence of margins in agents' optimization problems allows for deviations in the prices of relevant goods because the model misspecification will eventually lead to distorted relative prices. As a consequence, these margins can be used to measure both the nature and importance of misspecification.

While it is common in practice to add shocks to a given model to get a better fit, our contribution is to develop this procedure into a formal methodology to evaluate model misspecification. We do not recommend adding shocks where our method identifies a margin: even if a margin is significant, it does not necessarily mean it is a structural shock. Although some margins may have an economic interpretation, others do not, as the examples in our paper demonstrate. While our margins are uncorrelated with each other at all leads and lags, they should not be thought of as shocks, but as measures of misspecification: the uncorrelatedness of the shocks is important to make sure the researcher can identify the source of the misspecification, not because our margins should be thought of as structural shocks. In fact, it is difficult to interpret FEVD and marginal likelihood results if margins are correlated. This is because eliminating a margin may affect other margins, if the margins are correlated. In other words, we view our method as a diagnostic to detect which parts of a model are misspecified. The next step is then to improve the parts of the model that our methodology identifies as misspecified.

We include the margins into the agent's optimization problem, rather than the first order conditions. This is because, in case our method detects misspecification, the researcher can assess how to potentially modify the structural model and what kind of missing dynamics it may need. Technically, the method can be implemented by adding margins to the first order conditions, although the interpretation would be different in the latter case; furthermore, the researcher would be able to add only one margin for each first order condition, and thus he may not be able to locate which part of the model is misspecified. Finally, our approach may be sensitive to the functional forms used. For example, if one misspecifies disutility over labor in the utility function, that may result in a misspecification in the labor market. For our approach to succeed, it is necessary to consider a specification that encompasses all the important features and then add margins to that specification. If a margin is substantial, one needs to investigate the reason why it should be there economically.

In the next section, we examine where and how large the misspecification is by conducting FEVDs and marginal likelihood comparison. Our goal is not to compare models but rather to evaluate the specification of a given model. While margins may have a structural interpretation, our goal is to start with a model of interest without margins and evaluate its specification by measuring the contribution of margins without necessarily giving the margins a structural interpretation. Finally, note that testing for over-identification would not be useful in our context: the J-test for overidentifying restrictions would allow researchers to determine whether a model's moment conditions are correctly specified, but would not shed light on the sources of misspecification if the model fails the test.

3. The Proposed Methodology in Practice and Simulation Analysis

In this section, we provide Monte Carlo simulation evidence based on a simple New Keynesian model: we estimate several models (misspecified and correctly specified ones) and report FEVD analyses and marginal likelihoods to show how to detect the source of model misspecification using our method.

3.1. The Data-Generating Process (DGP)

Our DGP is the linearized simple New Keynesian model:[7]

$$\hat{Y}_t = E_t\left\{\hat{Y}_{t+1}\right\} - \frac{1}{\gamma}\left(\hat{R}_t - E_t\hat{\pi}_{t+1}\right) + \left(\hat{g}_t - E_t\hat{g}_{t+1}\right), \tag{30}$$

$$\hat{\pi}_t = \kappa\left\{(\gamma+\varphi)\hat{Y}_t - (\varphi+1)\hat{z}_t - \gamma\hat{g}_t + \hat{\chi}_t\right\} + \beta E_t\left\{\hat{\pi}_{t+1}\right\}, \tag{31}$$

$$\hat{R}_t = \rho_r\hat{R}_{t-1} + (1-\rho_r)\left\{\gamma_\pi\hat{\pi}_t + \gamma_y\hat{Y}_t\right\} + \hat{\nu}_t, \tag{32}$$

$$\hat{Y}_t = \hat{z}_t + \hat{L}_t, \tag{33}$$

where (30) is the dynamic IS curve, (31) is the New Keynesian Phillips curve (NKPC), (32) is the monetary policy rule, and (33) is the linearized aggregate production function. In the NKPC, $\kappa = \frac{(1-\beta\xi)(1-\xi)}{\xi}$.

The structural shocks follow: $\hat{x}_{t+1} = \rho_x\hat{x}_t + \sigma_x\varepsilon_{x,t+1}$, $\varepsilon_{x,t+1} \overset{iid}{\sim} N(0,1)$, where $x \in \{z, \nu, g, \chi\}$ and $\hat{x}$ denotes the log deviation of x from its steady state value.

3.2. The Proposed Methodology

To demonstrate our methodology, we estimate a correctly specified model, labeled $\mathcal{M}_0$, and three misspecified ones, labeled $\mathcal{M}_l$, $\mathcal{M}_m$ and $\mathcal{M}_g$. It is worth noting that, in our simulation setting, we know exactly whether the models are misspecified and the sources of their misspecification; in reality, however, we do not have such valuable information. In order to mimic realistic circumstances, we assume that we do not know which models are misspecified and introduce several margins into the above four models. There is more than one way to include the margins in the model. As we discuss in the previous section, we

[7]A detailed description of the model can be found in Section A of the Not-for-Publication Appendix.

prefer introducing margins into agents' optimization problems rather than into first order conditions for two reasons. One reason is that there are many ways to write first-order conditions. Another reason is that introducing margins in agents' optimization problems allows us to interpret the margins more clearly. The models we consider are the following:

$\mathcal{M}_0$: The equilibrium conditions of the model $\mathcal{M}_0$ are highly similar to those of the DGP. The only difference is that the dynamic IS curve and the NKPC curve contain three time-varying margins: $\hat{\tau}_{b,t}$ is a bond market margin and $\hat{\tau}_{c,t}$ is a final good margin, both of which enter the household's budget constraint (the first multiplies the lagged bond and the second multiplies consumption in the budget constraint), whereas $\hat{\tau}_{l,t}$ is a time-varying labor margin (it multiplies labor in the cost minimization problem of the intermediate good firms). We assume that all margins follow independent AR(1) processes: $\hat{\tau}_{x,t+1} = \rho_x\hat{\tau}_{x,t} + \sigma_x\varepsilon_{x,t+1}, \quad \varepsilon_{x,t+1} \overset{iid}{\sim} N(0,1)$, where $x \in \{l, c, b\}$. The corresponding equilibrium conditions are:

$$\begin{aligned}
\hat{Y}_t &= \mathrm{E}_t\hat{Y}_{t+1} - \frac{1}{\gamma}\left(\hat{R}_t - \mathrm{E}_t\hat{\pi}_{t+1}\right) + (\hat{g}_t - \mathrm{E}_t\hat{g}_{t+1}) - \frac{1}{\gamma}(\hat{\tau}_{c,t} - \mathrm{E}_t\hat{\tau}_{c,t+1}) - \frac{1}{\gamma}\mathrm{E}_t\tau_{b,t+1}, && (34)\\
\hat{\pi}_t &= \kappa\left\{(\gamma+\varphi)\hat{Y}_t - (\varphi+1)\hat{z}_t - \gamma\hat{g}_t + \hat{\chi}_t\right\} + \beta\mathrm{E}_t\{\hat{\pi}_{t+1}\} + \kappa(\hat{\tau}_{c,t} + \hat{\tau}_{l,t}), && (35)\\
\hat{R}_t &= \rho_r\hat{R}_{t-1} + (1-\rho_r)\left\{\gamma_\pi\hat{\pi}_t + \gamma_y\hat{Y}_t\right\} + \hat{\nu}_t, && (36)\\
\hat{Y}_t &= \hat{z}_t + \hat{L}_t. && (37)
\end{aligned}$$

In sum, model $\mathcal{M}_0$ is characterized by: (i) four equilibrium conditions: (34), (35), (36) and (37); (ii) four exogenous structural shock processes: $\hat{z}_t$, $\hat{\nu}_t$, $\hat{g}_t$ and $\hat{\chi}_t$; (iii) three exogenous margin processes: $\hat{\tau}_l$, $\hat{\tau}_c$ and $\hat{\tau}_b$.

$\mathcal{M}_l$: This model is misspecified since the labor supply shock, $\hat{\chi}_t$, is excluded from the set of structural shocks. In other words, we assume that there are only three structural shocks in the model. Thus, the NKPC becomes:

$$\hat{\pi}_t = \kappa\left\{(\gamma+\varphi)\hat{Y}_t - (\varphi+1)\hat{z}_t - \gamma\hat{g}_t\right\} + \beta\mathrm{E}_t\{\hat{\pi}_{t+1}\} + \kappa(\hat{\tau}_{c,t} + \hat{\tau}_{l,t}). \qquad (38)$$

In sum, model $\mathcal{M}_l$ is characterized by: (i) four equilibrium conditions: (34), (36), (37) and (38); (ii) three exogenous structural shock processes: $\hat{z}_t$, $\hat{\nu}_t$ and $\hat{g}_t$; and (iii) three exogenous margin processes: $\hat{\tau}_l$, $\hat{\tau}_c$ and $\hat{\tau}_b$.

$\mathcal{M}_m$: This model is misspecified because the non-systematic component of the nominal rate in the monetary policy decision, $\hat{\nu}_t$, is assumed to be *iid* rather than the AR(1) process described in the DGP. In other words, the model builder incorrectly assumes that the monetary policy shock follows:

$$\hat{\nu}_{t+1} \;=\; \sigma_\nu \varepsilon_{\nu,t+1}, \quad \varepsilon_{\nu,t+1} \overset{iid}{\sim} N(0,1). \tag{39}$$

In sum, model $\mathcal{M}_m$ is characterized by: (i) four equilibrium conditions: (34), (35), (36) and (37); (ii) four exogenous structural shock processes: $\hat{z}_t$, $\hat{g}_t$, $\hat{\chi}_t$ and $\hat{\nu}_t$, defined in eq. (39); and (iii) three exogenous margin processes: $\hat{\tau}_l$, $\hat{\tau}_c$ and $\hat{\tau}_b$.

$\mathcal{M}_g$: This model is misspecified because the government spending shock $\hat{g}_t$ is excluded from the set of the structural shocks. In this model, the exclusion of the government spending shock affects both the dynamic IS curve and the NKPC:

$$\hat{Y}_t \;=\; \mathrm{E}_t\left\{\hat{Y}_{t+1}\right\} - \frac{1}{\gamma}\left(\hat{R}_t - \mathrm{E}_t\hat{\pi}_{t+1}\right) - \frac{1}{\gamma}\left(\hat{\tau}_{c,t} - \mathrm{E}_t\hat{\tau}_{c,t+1}\right) - \frac{1}{\gamma}\mathrm{E}_t\left\{\tau_{b,t+1}\right\}, \tag{40}$$

$$\hat{\pi}_t \;=\; \kappa\left\{(\gamma+\varphi)\hat{Y}_t - (\varphi+1)\hat{z}_t + \hat{\chi}_t\right\} + \beta\mathrm{E}_t\left\{\hat{\pi}_{t+1}\right\} + \kappa\left(\hat{\tau}_{c,t} + \hat{\tau}_{l,t}\right). \tag{41}$$

In sum, model $\mathcal{M}_g$ is characterized by: (i) four equilibrium conditions: (36), (37), (40) and (41); (ii) four exogenous structural shock processes: $\hat{z}_t$, $\hat{\nu}_t$, $\hat{g}_t$ and $\hat{\chi}_t$; and (iii) three exogenous margin processes: $\hat{\tau}_l$, $\hat{\tau}_c$ and $\hat{\tau}_b$.

Traditionally, FEVDs are used to evaluate the contribution of various structural shocks to the observables (e.g. Ireland, 2001). Usually, researchers assume that the model is correctly specified and all the variation in the observables are fully explained by the structural shocks. In contrast, we consider the possibility that all the models are potentially misspecified. Therefore, the FEVD analysis should reveal the effects of both structural shocks and margins since the model without margins is misspecified. We expect that the margin that plays an important role in the FEVD identifies the source of the misspecification.

The marginal likelihood can be written as the product of one-step-ahead predictive densities that may be highly non-normal because parameters are integrated out in the marginal likelihood. Even if we focus on one-step-ahead predictions, the densities contain more information than the second moments (i.e., FEVD). For this reason, the marginal likelihood complements FEVDs and may provide additional information. We expect that removing the margin that is most related to the type of misspecification has the largest impact on the marginal likelihood. Thus, we propose to remove one margin at a time and compare the marginal likelihoods of the model with all the margins with the model where one of the margins has been removed. Checking which margins reduce the marginal likelihood the most enables researchers to locate the source of model misspecification. Rules of thumb, such as those in Jeffreys (1961, p. 432) and Kass and Raftery (1995, p. 777), can be used to determine whether the results are significant. [8]

3.3. Simulation Results

We simulate data on hours worked $\hat{L}_t$, real output $\hat{Y}_t$, the inflation rate $\hat{\pi}_t$, and the nominal interest rate $\hat{R}_t$ using the DGP described in subsection 3.1 and estimate the four models described in subsection 3.2.[9] The sample size of the simulated data is 100.

The FEVDs in terms of the innovations of the structural shock and margin processes are summarized in Table 1. Panel (a) focuses on model $\mathcal{M}_0$. The structural shocks, especially the monetary policy shock, explain most of the FEVD. In contrast, the contribution of the margins is indeed negligible for any of the four variables, which indicates that the model is correctly specified, as expected: since model $\mathcal{M}_0$ is correctly specified, the three margins are redundant.

Panel (b) focuses on model $\mathcal{M}_l$. Clearly the contribution of the labor margin (ε_l) is substantial for all the observables. Comparing these results with those in panel (a), evidently the role of the labor supply shock is almost replaced by the labor margin. This is because the omission of the labor supply shock distorts the wage in the labor market in model $\mathcal{M}_l$. Thus, the labor margin correctly identifies the source of the misspecification, which is the omission of the labor supply shock.

[8] According to Kass and Raftery (1995), the evidence is *not worth more than a bare mention*, *positive*, *strong* and *very strong* if twice the natural logarithm of the Bayes factor is 0 to 2, 2 to 6, 6 to 10, and is greater than 10, respectively.

[9] The pseudo-true data are log deviations from steady-state values.

Panel (c) reports the FEVD of model $\mathcal{M}_m$. In this case, the bond market margin (ε_b) explains the majority of the forecast errors of all variables, while the labor supply and final good margins explain almost nothing. Again, the bond market margin correctly identifies the source of the misspecification which is related to the bond price, $\hat{R}_t$, and therefore points to misspecification in the monetary policy equation.

Finally, comparing the results of panels (d) and (a), it is clear that the final good margin (ε_c) somehow replaces the role of the government spending shock in panel (d), correctly pointing to the source of the misspecification even though the absolute magnitude of the contribution is not very large compared to the other two cases ($\mathcal{M}_l$ and $\mathcal{M}_m$). The latter may be due to the fact that, in our setting, the role of government spending is relatively small. The lack of the government spending in model $\mathcal{M}_g$ distorts the resource constraint which in turn causes distortion in the final goods market.

Table 2 reports the marginal likelihood of the models with the three margins as well as the likelihood of the models with one margin removed at a time. The first column shows the marginal likelihood of several sub-models based on $\mathcal{M}_0$. Since the marginal likelihood does not change much regardless of which margin is dropped, the marginal likelihood correctly signals that all margins are negligible in the correctly specified model. According to Kass and Raftery (1995, p.777), the magnitude of the differences in the log marginal likelihood is at most *positive*. However, when the labor supply shock is neglected (sub-models based on $\mathcal{M}_l$ in the second column), the model without the labor margin has the lowest marginal likelihood. Similarly, when the process for the monetary policy shock (sub-model based on $\mathcal{M}_m$) is misspecified, the marginal likelihood is the lowest when the bond market margin is removed. When government spending is omitted from the model ($\mathcal{M}_g$), removing the final good margin leads to the lowest marginal likelihood, while the reduction is relatively minor since government spending shocks play a small role in the DGP. The evidence for τ_l and τ_b in $\mathcal{M}_l$ and $\mathcal{M}_m$, respectively, is *very strong* in Kass and Raftery's (1995) terminology while the evidence for τ_c is *positive* in $\mathcal{M}_g$.

Figure 1 shows the population and estimated impulse responses based on posterior mean estimates. As expected, the impulse responses estimated from the correctly specified model with margins ($\mathcal{M}_0$) tend to be close to the population impulse responses (DGP). [10] Omitting the labor supply shock from the model does not have a strong impact on the impulse responses to the other shocks ($\mathcal{M}_l$), and excluding the government spending shock has mild effects ($\mathcal{M}_g$). Misspecifying the monetary policy shock process, instead, has large effects ($\mathcal{M}_m$): The response of the interest rate to the monetary policy shock has the opposite sign.

To provide intuition why our methodology works, let us examine the linearized equilibrium conditions in subsection 3.2. In model $\mathcal{M}_l$, the only equation affected by the misspecification is the NKPC (eq. 38), where the labor supply shock is missing relative to the correctly specified NKPC (eq. 31), and where two of the margins $(\tau_{c,t}, \tau_{l,t})$ show up. The dynamic IS curve, eq. (34), where the margins $\tau_{c,t}$ and $\tau_{b,t}$ show up, is not affected by the misspecification. Thus, the misspecification in the Phillips curve must be captured by $\tau_{l,t}$, which is the only margin that appears in the NKPC but not in the IS curve. This is exactly what we find. Regarding the model $\mathcal{M}_m$, the only equation affected by the misspecification is the Taylor rule. Since the nominal interest rate appears in the IS curve but not in the NKPC, the misspecification will be captured by the margin that appears in the IS but not in the NKPC, that is $\tau_{b,t}$. In model $\mathcal{M}_g$, the misspecification affects both the NKPC and the IS equations; the only margin that enters in both is $(\tau_{c,t})$, which will thus capture the misspecification.

The lesson we learn from the simulation exercise is that introducing margins in the model has the potential to capture the missing channels since the margins correctly reveal which structural shocks are missing from the model. To summarize, FEVD and marginal likelihood analyses may provide useful tools for detecting and identifying model misspecification. When a model is misspecified, the contribution of a margin to the FEVD is substantial and is related to the misspecification, thus correctly signaling the possible cause of

[10] The exception is the impulse response function of the interest rate to the government spending shock. It is due to the biased estimate of ρ_g. See Table C in the Not-for-Publication Appendix.

the misspecification. Moreover, when a margin captures an important aspect of model misspecification, removing it has a large impact on the marginal likelihood. On the other hand, a word of caution: the method may not necessarily identify the source of misspecification: in our Monte Carlo simulation it does, but it depends on the model and its features.

3.4. Suggestions for Practitioners

In practice, one does not know which parts of a model are misspecified and may wonder where the margins should be included. In principle, while one can introduce a margin for each market in the model, such a strategy might yield an over-parameterized model even for Bayesian estimation methods. We suggest two approaches. One is to introduce margins in markets which the researcher suspects are misspecified. Another is to introduce margins everywhere, but impose tight priors on the AR(1) parameters and innovation variances of the margin where misspecification is unlikely. The first approach is suitable when the researcher has a strong view on which parts of the model are correctly specified and which parts are not. The second approach is more agnostic about the nature of the misspecification.

The parsimonious AR(1) specification of margin processes is chosen for estimation although one could consider more general specifications and our method would still be valid. By keeping the specification of the margin processes as simple as AR(1) processes, parameters are estimated more precisely and our procedure is expected to have a better chance in detecting misspecification.

One might worry that adding many margins into the model creates lack-of-identification problems. If that is the case, the researcher can use any of the existing methods for detecting identification problems. For example, if the researcher worries about weak identification, he/she could utilize the methods proposed by Canova and Sala (2009), Iskrev (2010) or Inoue and Rossi (2011). If the researcher worries about lack of local identification, he/she could utilize Komunjer and Ng (2011), while if the researcher worries about lack of global identification, he/she could utilize the method proposed by Qu and Tkachenko (2016). It should be noted that the marginal likelihood criterion can be used to detect misspecification even when some parameters are not identified, however. This is because the value of the marginal likelihood is identified even when parameters are not.

Finally, while in our examples FEVD and marginal likelihood analyses convey the same conclusions, that may not always be the case in practice. In fact, they allow researchers to evaluate different aspects of the model: FEVDs focus on how the margins affect the second moment properties of the observables while the marginal likelihood provides a more general assessment that includes other moments and general characteristics of the overall distribution. On the other hand, the marginal likelihood only assesses the joint performance of the margins on all the observables, while FEVDs provide more detailed information on which observables are affected by which margins. The marginal likelihood has a built-in penalty term for overparameterization. Hence, they both are useful in different ways and we recommend researchers to use both of them.

4. Empirical Application to a Medium-Scale New Keynesian Model

In this section, we consider potential misspecification in a medium-scale New Keynesian model. The model, based on Justiniano et al. (2010), is a stochastic neoclassical growth model with various real and nominal frictions, and is routinely used by researchers and policymakers. The frictions include imperfect competition in the intermediate goods and labor markets, sticky prices and wages, habit formation in consumption, investment adjustment cost and variable capital utilization. We explore whether this model may be misspecified by including several time-varying margins and evaluating their importance. See Table 3 for notation and Sections B and C in the Not-for-Publication Appendix for the description of the model and for its linearized equilibrium conditions, respectively.

Following Justiniano et al. (2010), we estimate the model using seven quarterly aggregate U.S. time series from 1954:III to 2004:IV. We focus on a linearized DSGE model, although our method could be applied to non-linear models as well. We intentionally selected a sample that does not include the zero lower bound, the reason being that our model is linearized and we do not explicitly take into account the zero lower bound constraint. One could potentially apply our methodology to models that include a zero lower bound if one were interested in extending the sample period. The variables in the model are: real output, real consumption, real investment, hours worked, the inflation rate, and the Federal funds rate. Real output, real consumption, real investment and hours worked are per capita. Real output, investment and consumption per capita are obtained by dividing nominal GDP, investment and consumption by the population and the price index. Nominal consumption is defined as the sum of non-durable goods and services expenditures. Nominal investment is defined as the sum of private domestic investment and personal durable goods expenditure. Hours worked per capita are defined as total hours worked in the non-farm business sector divided by the population. Real wages are defined as non-farm business sector hourly compensation divided by the price index. We use the civilian non-institutional population as our population measure. The price index is the GDP deflator, and the quarterly inflation rate is its growth rate. We introduce six margins in the model: labor demand, capital demand, consumption good, bond market, intermediate good and labor market margins. Estimation results are based on 250,000 Markov Chain Monte Carlo (MCMC) draws, and the first 50,000 draws are discarded. Section D in the Not-for-Publication Appendix discusses the priors and posterior estimates of the parameters.

Forecast Error Variance Decomposition. Recall from Sections 2 and 3 that the margins that play the largest role in the FEVDs identify the sources of model misspecification. Table 4 reports the FEVD contribution of the innovations of the structural shock and margin processes (listed in the columns) to the overall variance of the observable variables (listed in the rows) at different forecast horizons ($H = 1, 4, 20$ quarters). Table 4 shows that the labor demand, capital demand and consumption good margins (ε_l, ε_k and ε_c) are extremely small for all the variables of interest. In contrast, the bond market, the intermediate goods demand, and the household labor margins (ε_r, ε_q and ε_h) contribute to explain the variability of several variables, sometimes substantially. For example, the household labor margin explains 58% of one quarter-ahead forecast error variance of wage growth. The effects of the household labor margin are persistent: it explains 38% of the wage growth variation after twenty quarters. This result indicates that the model without household labor margin could be misspecified in the wage growth dynamics. Our results differ from Justiniano et al. (2010), whose wage mark-up shocks explain about 56% of the wage growth variation at the business cycle frequency: according to our results, the wage mark-up shock (ε_w) does not explain more than 10% between 1 and 20 quarters. That is, the model without household labor margin might attribute wage fluctuations to the wage mark-up shock.

At the one-quarter ahead forecast horizon, the intermediate good margin (ε_q) explains 15% of the variation in inflation. Actually, its contribution is even larger than that of technology shocks, which explain about 12% of inflation fluctuations at the same forecast horizon. Note also that its importance decays as the forecast horizon increases. Since the intermediate good margin mainly affects short run fluctuations, our results warn against using the model for forecasting short run inflation if the model misspecification is not properly addressed. It is worth noting that while the intermediate good (ε_q) and household labor margins (ε_h) have crucial effects on nominal variables (that is, wage growth, inflation rates, and interest rates), they have hardly any effects on other variables.

The bond market margin ε_r mainly affects nominal interest rates fluctuations, and its contribution increases with the forecast horizon: at the twenty quarters horizon ($H = 20$), more than 44% of the interest rate variability can be attributed to fluctuations in the bond market margin. The dominant effect of the bond market margin comes at the expense of the wage mark-up and investment shocks, whose total contribution is at most 37%. This result sharply differs from Justiniano et al. (2010), where about 67% of the interest rate variability is explained by wage mark-up and investment shocks. In addition, the bond market margin also contributes to output, investment, and consumption growth as well as hours worked fluctuations at various forecast horizons, although its effects are smaller in magnitude.

Based on the above FEVD analysis, we conclude that margins mainly affect wage growth, the inflation rate, and the interest rate; for other observables, our results are similar to Justiniano et al. (2010): that is, the investment shock is the main driving force behind output growth, investment growth, and hours worked, and the inter-temporal preference shock explains most of the variability in consumption growth.[11]

For comparison, we report the FEVD of the benchmark model without margins in Table 5. To simplify the discussion, we focus on the one quarter-ahead forecast horizon. Regarding the wage growth, the model without margins explains its fluctuations mainly by the wage mark-up shock (62%). However, once margins are taken into account, the contribution of the wage mark-up shock reduces to about 10% only as shown in Table 4. In contrast, the household labor margin explains about 58% of wage growth in the model with margins. Regarding the inflation rate, the contribution of the price markup shock decreases substantially once the margins are introduced.

Overall, our FEVD results in Table 4 indicate that misspecification is mostly captured by the household labor, the bond market and the intermediate goods demand margins, which confirms the existing view that the asset and labor markets in the standard New Keynesian DSGE model are misspecified (see Krause, Lopez-Salido and Lubik, 2008, and Christiano, Motto and Rostagno, 2008, for example).

Marginal Likelihood. We compare the marginal likelihood (calculated by the modified harmonic mean estimator) of: (i) the model with all the margins; (ii) the model with no margins; (iii) the models that remove one margin at a time; and (iv) the model that removes the three margins that, according to the FEVD, are the source of the misspecification. Recall that the logic of this exercise is to investigate the relative importance of margins: if removing a particular margin considerably reduces the marginal likelihood value, it is a signal that the margin is crucial in explaining the dynamics of the observed data. Panel A in Table 6 summarizes the marginal likelihood of the various versions of the benchmark model. First of all, the first two rows display the results for models with and without margins; their difference is *very strong* according to Kass and Raftery (1995).

[11] See Table 3 of the technical appendix in Justiniano et al. (2010).

Interestingly, the model that removes the same three margins that FEVDs identify as the source of misspecification is the one with the lowest likelihood, and the difference is even larger. This finding confirms that the latter are the sources of misspecification, while the small differences in the marginal likelihood suggest that the degree of misspecification may be mild.[12]

Figure 2 reports time series estimates of the margins. By examining the estimate of the bond market margin over time ($\tau_{r,t}$), it is clear that the margin was especially large in the early 1980s, possibly due to the changes in monetary policy around that time.

Serial Correlation in the Shocks. DSGE models rely on a structural interpretation of the exogenous shocks. The exogenous shocks should be invariant to any policy changes (to be robust to the Lucas critique) as well as uncorrelated among each other. If a shock included in a DSGE model is in reality a combination of other shocks, that shock should be interpreted as a reduced-form, rather than structural, shock. Here, we assess the correct specification of the ARMA structure of the wage mark-up shock, as in Chari et al. (2009).

We consider a restricted model which is the same as the benchmark model except that the moving average coefficient of the wage mark-up shock is set to zero. We estimate several variants of the benchmark model, depending on which margins are included or removed, and compare their marginal likelihoods.

Panel B in Table 6 displays the marginal likelihood of the model. Clearly, the marginal likelihood values in Panel B are all lower than the corresponding values in Panel A. One can interpret this finding in two ways. A first interpretation is that the wage mark-up shock indeed follows an ARMA process. Thus, by estimating the model with the wage mark-up shock following an AR process, we are actually estimating a misspecified model. By including the household labor margin (τ_h) into the model, the potential gap due to the labor market misspecification is filled by this margin. This point can be seen clearly by comparing marginal likelihood values of the models with and without the household labor market

[12] We did not investigate all possible margin combinations, as that would result in too many models to consider. The scope of the marginal likelihood analysis was mostly to confirm the results of the FEVDs.

margin. The evidence is *very strong* according to Kass and Raftery (1995). Given that the true model is that with an ARMA-type wage mark-up shock, the marginal likelihood values indeed help us detect the source of model misspecification. A second interpretation is that the wage mark-up shock instead follows an AR process. Since including the household labor margin improves the overall fitting of the model, then some factors related to the labor market must be missing. In other words, the models without the household labor margin are misspecified in the labor market. Either way, our exercise casts doubts on the structural interpretation of the wage mark-up shock.

5. Conclusion

This paper proposes empirical methods for detecting and identifying misspecification in structural economic models. Our approach is based on analyzing FEVDs and marginal likelihoods of DSGE models augmented with margins, where the margins are introduced in the agents' optimization problems to capture potential misspecification. Monte Carlo simulations demonstrate that our method can correctly identify the source of the misspecification. Our empirical results show that a medium-scale New Keynesian DSGE model that incorporates features in the recent empirical macro literature is still severely misspecified, and suggest that asset and labor markets are the sources of the misspecification.

We should note that there are three potential issues with implementing our method: exogeneity of the margins, over-parametrization and non-nesting misspecification. First, because the margin processes are assumed to be exogenous, our method might not correctly identify the location of misspecification if the misspecification was endogenous. One way to address this issue is to let the margins depend on state variables. In our simulation results, however, we are able to successfully identify the misspecification due to the serially correlated omitted frictions in the monetary policy reaction function, which suggests the usefulness of our approach in finding the omitted frictions in the model.

Second, when many markets are included in the model, there may be too many locations for introducing margins and ways for forming priors for these processes. We suggest to either use fewer margins when the prior on the location of the misspecification is strong (e.g. the researcher is confident that there is no misspecification in some parts of the model, but unsure whether there might be misspecification in others, and the researcher has strong opinions on where the misspecification is potentially located), or introduce many margins and impose prior information when the misspecification location is more uncertain (i.e., every part of the model can potentially be misspecified). When neither is possible, another approach is to use a Bayesian model averaging approach to take into account many margins, and let it provide information on the location of the margins.

Lastly, our method of comparing models via their marginal likelihood is clearly appropriate if the true model is one of the models with the margins. Otherwise, one could consider alternative ways to perform the comparison, such as cross-validation and out-of-sample predictions (see Bernardo and Smith, 2000, p.403 and Geweke, 2010). We leave these extensions to future research.

References

Bernardo, J.M., Smith, A.F.M., 2000. Bayesian theory. Wiley Series in Probability and Statistics, Chichester.

Brinca, P., Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2016. Accounting for business cycles. In: Taylor, J.B., Uhlig, H., (Eds.), Handbook of Macroeconomics, volume 2. Elsevier, Amsterdam, pp.1013–1063.

Canova, F., Ferroni, F., 2011. Multiple filtering devices for the estimation of cyclical DSGE models. Quantitative Economics 2, 73-98.

Canova, F., Sala, L., 2009. Back to square one: Identification issues in DSGE models. Journal of Monetary Economics 56, 431-449.

Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2007. Business cycle accounting. Econometrica 75, 781–836.

Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2009. New Keynesian models: Not yet useful for policy analysis. American Economic Journal: Macroeconomics 1, 242–66.

Christiano, L., Motto, R., Rostagno, M., 2008. Financial factors in business cycles. Unpublished manuscript.

Corradi, V., Swanson, N., 2007. Evaluation of dynamic stochastic general equilibrium models based on distributional comparison of simulated and historical data. Journal of Econometrics 136, 699-723.

Curdia, V., Reis, R., 2010. Correlated disturbances and U.S. business cycles. NBER Working Paper 15744.

Del Negro, M., Schorfheide, F., 2009. Monetary policy analysis with potentially misspecified models. American Economic Review 99, 1415–50.

Del Negro, M., Schorfheide, F. 2014. Real-time DSGE model density forecasts during the Great Recession. mimeo.

Del Negro, M., Schorfheide, F., Smets, F., Wouters, R., 2007. On the fit of new Keynesian models. Journal of Business & Economic Statistics 25, 123–143.

Dufour, J.-M., Khalaf, L., Kichian, M., 2013. Identification-robust analysis of DSGE and structural macroeconomic models. Journal of Monetary Economics 60, 340–350.

Edge, R.M., Gurkaynak, R.S., 2010. How useful are estimated DSGE model forecasts for central bankers? Brookings Papers on Economic Activity Fall, 209–259.

Fernandez-Villaverde, J., Rubio-Ramirez, J.F., 2004. Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach. Journal of Econometrics 123, 153–187.

Geweke, J., 2010. Complete and incomplete econometric models. Princeton University Press, Princeton, NJ.

Inoue, A., Rossi, B., 2011. Testing for weak identification in possibly nonlinear models. Journal of Econometrics 161, 246-261.

Ireland, P. N., 2001. Sticky-price models of the business cycle: specification and stability. Journal of Monetary Economics 47, 3–18.

Ireland, P. N., 2004. Technology shocks in the new Keynesian model. The Review of Economics and Statistics 86, 923–936.

Iskrev, N., 2010. Local identification in DSGE models. Journal of Monetary Economics 57, 189–202.

Jeffreys, H., 1961. Theory of Probability, Third Edition. Oxford University Press.

Justiniano, A., Primiceri, G.E., Tambalotti, A., 2010. Investment shocks and business cycles. Journal of Monetary Economics 57, 132–145.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 430, 773–795.

Komunjer, I., Ng, S., 2011. Dynamic identification of dynamic stochastic general equilibrium models. Econometrica 79, 1995–2032.

Krause, M.U., Lopez-Salido, D., Lubik, T.A., 2008. Inflation dynamics with search frictions: A structural econometric analysis. Journal of Monetary Economics 55, 892–916.

Ng, S., Wright, J.H., 2013. Facts and challenges from the Great Recession for forecasting and macroeconomic modeling. Journal of Economic Literature 51, 1120–1154.

Qu, Z., Tkachenko, D., 2016. Global identification in DSGE models allowing for indeterminacy. Review of Economic Studies, forthcoming.

Sargent, T.J., 1989. Two models of measurements and the investment accelerator. Journal of Political Economy 97, 251–287.

Smets, F., Wouters, R., 2007. Shocks and frictions in US business cycles: A Bayesian DSGE approach. American Economic Review 97, 586–606.

Stock, J.H., Watson M.W., 2012. Disentangling the channels of the 2007-2009 recession. Brookings Papers on Economic Activity Spring 2012, 81-135.

Watson, M.W., 1993. Measures of fit for calibrated models. The Journal of Political Economy 101, 1011-1041.

Table 1: Forecast Error Variance Decompositions

Panel (a): Model $\mathcal{M}_0$							
Variable	ε_z	ε_ν	ε_g	ε_χ	ε_l	ε_c	ε_b
Hours	9.36	58.35	15.50	13.17	1.79	1.57	0.26
Output	9.48	58.29	15.47	13.14	1.79	1.57	0.26
Inflation Rate	3.77	88.93	0.68	5.51	1.01	0.04	0.06
Interest Rate	23.97	35.35	0.79	33.74	5.04	0.30	0.82
Panel (b): Model $\mathcal{M}_l$							
Variable	ε_z	ε_ν	ε_g	ε_χ	ε_l	ε_c	ε_b
Hours	9.63	59.19	16.32	-	13.02	1.56	0.28
Output	10.02	58.96	16.24	-	12.96	1.55	0.28
Inflation Rate	3.89	89.05	0.78	-	6.17	0.04	0.08
Interest Rate	25.31	36.85	1.06	-	35.41	0.31	1.06
Panel (c): Model $\mathcal{M}_m$							
Variable	ε_z	ε_ν	ε_g	ε_χ	ε_l	ε_c	ε_b
Hours	11.62	18.14	18.17	14.26	1.31	1.92	34.58
Output	9.47	18.60	18.57	14.61	1.34	1.96	35.45
Inflation Rate	6.65	27.50	1.07	9.60	1.29	0.05	53.84
Interest Rate	3.01	4.60	0.26	5.06	0.41	0.03	86.62
Panel (d): Model $\mathcal{M}_g$							
Variable	ε_z	ε_ν	ε_g	ε_χ	ε_l	ε_c	ε_b
Hours	9.91	61.67	-	15.32	2.13	10.44	0.54
Output	10.17	61.51	-	15.27	2.12	10.39	0.54
Inflation Rate	3.77	88.84	-	5.88	1.09	0.26	0.15
Interest Rate	22.95	34.83	-	34.31	5.25	0.77	1.89

Notes to the table. The table reports the posterior mean of the FEVD (in percentage) for the models we estimated. The forecast horizon is 12 periods. The structural shocks are the technology shock (ε_z), the monetary policy shock (ε_μ), the government spending shock (ε_g), and the labor supply shock (ε_χ). The innovatins to the margin processes are the labor demand margin (ε_l), the final good margin (ε_c), and the bond market margin (ε_r).

Table 2. Log Marginal Likelihood

	$\mathcal{M}_0$	$\mathcal{M}_l$	$\mathcal{M}_m$	$\mathcal{M}_g$
All margins	*480.48*	*477.61*	*364.56*	*481.02*
Remove τ_l	*480.06*	***465.60***	*364.38*	*480.66*
Remove τ_c	*481.74*	*478.79*	*365.36*	***477.96***
Remove τ_b	*481.63*	*478.72*	***354.59***	*481.70*

Notes to the table. The numbers are the log marginal likelihood, calculated via the modified harmonic mean. Lower values of the marginal likelihood indicate that models are more at odds with the data. Thus, the margins whose removal are associated with the lowest likelihood are the margins that are deemed the most necessary to explain the data.

Table 3. Margins and Shocks: Summary Table

Margin	*Description*	*Innovation*	*Reference Equation in the Not-for-Publication Appendix*
		Panel A. Margins	
τ_l	Homogeneous labor market margin	ε_l	Eq. (B.8)
τ_k	Capital market margin	ε_k	Eq. (B.7)
τ_c	Consumption margin	ε_c	Eq. (B.12)
τ_r	Bond market margin	ε_r	Eq. (B.10)
τ_q	Intermediate goods demand margin	ε_q	Eq. (B.3)
τ_h	Household labor margin	ε_h	Eq. (B.16)
		Panel B. Shocks	
η_{mp}	Monetary policy shock	ε_{mp}	Eq. (B.18)
z	Technology shock	ε_z	Eq. (B.6)
g	Government spending shock	ε_g	Eq. (B.17)
μ	Investment shock	ε_μ	Eq. (B.11)
λ_p	Price mark-up shock	ε_p	Eq. (B.2)
λ_w	Wage mark-up shock	ε_w	Eq. (B.15)
b	Intertemporal preference shock	ε_b	Eq. (B.13)

Notes to the table. The table summarizes the shocks and the margins in the model considered in Section 4. For simplicity, we removed time subscripts from the notation.

Table 4. The Forecast Error Variance Decomposition: The Benchmark Model (All Margins Included)

	Shocks							*Margins*					
Series	ε_{mp}	ε_z	ε_g	ε_μ	ε_p	ε_w	ε_b	ε_l	ε_k	ε_c	ε_r	ε_q	ε_h
Forecast Horizon: $H=1$													
Output growth	5.46	11.93	11.00	52.58	1.02	0.29	10.45	0.00	0.00	0.00	7.13	0.05	0.08
Consumption growth	2.16	16.66	1.99	0.39	0.13	4.56	71.07	0.00	0.00	0.02	2.94	0.01	0.06
Investment growth	4.40	4.54	0.03	83.37	1.08	0.38	0.35	0.00	0.00	0.00	5.74	0.04	0.07
Hours	5.03	19.34	10.07	48.13	0.92	0.28	9.55	0.00	0.00	0.00	6.55	0.04	0.09
Wage growth	0.37	10.37	0.00	0.80	15.66	9.61	0.16	0.01	0.00	0.00	0.56	4.30	58.15
Inflation rates	2.68	12.48	0.11	1.87	38.16	21.67	0.35	0.04	0.00	0.00	6.00	15.16	1.48
Interest rates	52.99	6.32	0.64	12.28	1.81	1.62	10.46	0.02	0.00	0.00	11.79	1.67	0.40
Forecast Horizon: H=4													
Output growth	5.69	20.89	8.41	44.98	2.22	1.81	8.17	0.00	0.00	0.00	7.62	0.04	0.16
Consumption growth	2.26	27.73	3.20	0.53	0.32	9.68	52.98	0.00	0.00	0.02	3.19	0.01	0.09
Investment growth	4.88	8.43	0.03	76.65	2.48	0.47	0.39	0.00	0.00	0.00	6.48	0.04	0.13
Hours	9.62	6.65	4.74	53.41	3.90	2.85	5.53	0.00	0.00	0.00	12.93	0.04	0.32
Wage growth	0.42	28.87	0.00	1.21	16.82	6.64	0.10	0.01	0.00	0.00	0.65	3.43	41.84
Inflation rates	4.43	12.20	0.13	2.47	25.01	37.12	0.48	0.02	0.00	0.00	10.27	6.72	1.14
Interest rates	20.31	8.46	0.66	26.71	1.93	5.86	6.52	0.01	0.00	0.00	28.65	0.46	0.42
Forecast Horizon: H=20													
Output growth	5.61	20.79	7.45	43.49	2.49	4.99	7.56	0.00	0.00	0.00	7.39	0.05	0.18
Consumption growth	1.95	29.11	3.20	2.63	0.39	12.90	46.96	0.00	0.00	0.02	2.75	0.01	0.09
Investment growth	4.91	8.10	0.04	75.52	2.82	1.60	0.40	0.00	0.00	0.00	6.43	0.04	0.15
Hours	6.19	5.08	2.19	20.06	7.76	47.04	1.62	0.00	0.00	0.00	9.66	0.02	0.39
Wage growth	0.43	32.53	0.01	1.33	17.00	6.33	0.11	0.01	0.00	0.00	0.67	3.10	38.48
Inflation rates	4.31	7.42	0.12	1.77	14.49	55.26	0.39	0.01	0.00	0.00	11.92	3.62	0.69
Interest rates	8.99	4.91	0.43	20.21	1.21	16.49	3.28	0.01	0.00	0.00	44.05	0.20	0.22

Notes to the table. The table reports the posterior mean of the FEVD (in percentage) for the model with margins. The structural shocks are the monetary policy shock (ε_{mp}), the technology shock (ε_z), the government spending shock (ε_g) , the investment shock (ε_μ), the price mark up shocks (ε_p), the wage mark-up shock (ε_w) and the intertemporal preference shock (ε_b). The innovatins to the margin processes are the labor demand margin (ε_l), the capital demand margin (ε_k), the consumption good margin (ε_c), the bond market margin (ε_r), and the intermediate good margin (ε_q) and the labor market margin (ε_h). The posterior mean of the FEVD is calculated over 200,000 MCMC draws after discarding the first 50,000 draws.

Table 5. The Forecast Error Variance Decomposition: The Benchmark Without Margins

Series	ε_{mp}	ε_z	ε_g	ε_μ	ε_p	ε_w	ε_b
Forecast Horizon: $H = 1$							
Output growth	4.51	14.05	10.67	56.68	2.07	0.97	11.04
Consumption growth	2.03	17.56	2.10	1.37	0.36	3.73	72.84
Investment growth	2.85	4.27	0.02	89.95	1.77	0.06	1.08
Hours	4.32	18.37	10.14	53.79	1.93	0.99	10.46
Wage growth	0.31	9.44	0.00	1.33	26.15	62.41	0.35
Inflation rates	1.74	10.26	0.09	4.07	61.90	20.97	0.97
Interest rates	52.58	7.00	0.69	16.89	5.75	1.69	15.39
Forecast Horizon: $H = 4$							
Output growth	4.28	22.38	7.85	50.39	3.33	3.31	8.45
Consumption growth	1.92	27.39	3.20	1.34	0.59	7.54	58.03
Investment growth	2.91	7.34	0.02	84.81	3.07	0.55	1.29
Hours	6.65	5.29	3.97	64.41	5.66	5.49	8.54
Wage growth	0.36	30.70	0.00	2.24	24.43	42.00	0.28
Inflation rates	3.49	11.95	0.13	7.13	37.72	37.84	1.74
Interest rates	20.94	9.37	0.65	43.19	4.66	5.85	15.34
Forecast Horizon: $H = 20$							
Output growth	4.31	21.98	7.03	49.55	3.53	5.06	8.54
Consumption growth	1.65	26.73	3.04	5.14	0.53	8.53	54.37
Investment growth	2.90	6.88	0.03	84.43	3.26	1.18	1.31
Hours	4.26	4.73	2.05	32.40	8.53	44.80	3.23
Wage growth	0.37	33.64	0.01	2.52	23.85	39.29	0.33
Inflation rates	3.92	9.06	0.13	6.63	26.38	52.12	1.75
Interest rates	12.24	7.11	0.52	49.25	3.24	16.19	11.45

Notes to the table. See the definition of the shocks and margins in the notes to Table 4. The table reports the posterior mean of the FEVD (in percentage) for the model without margins. The FEVD is calculated over 200,000 MCMC draws after discarding the first 50,000 draws.

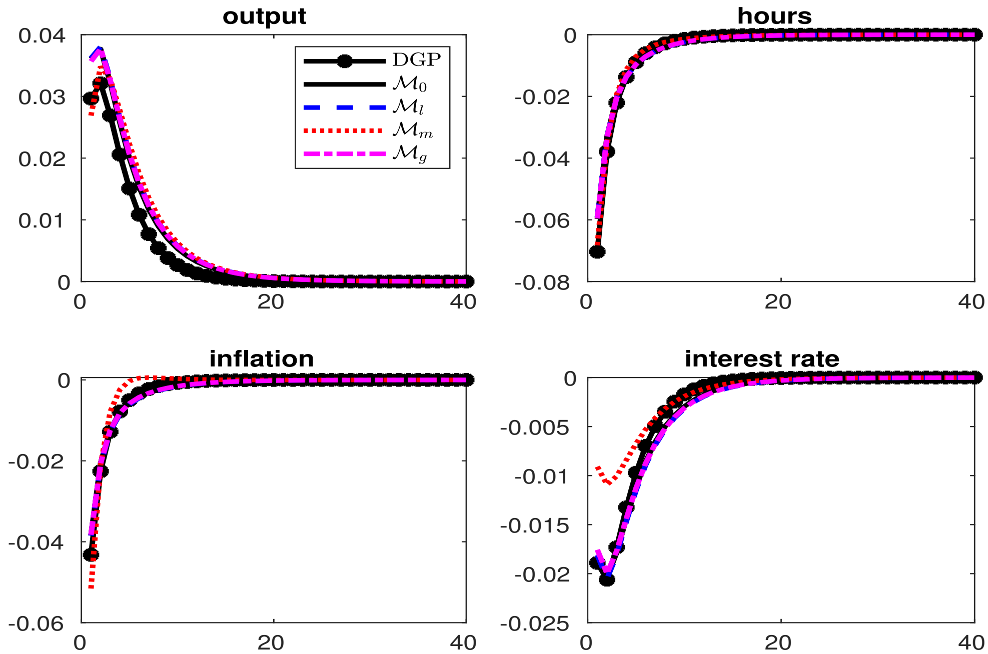
Table 6. Log Marginal Likelihood Values

Panel A. Variants of the Benchmark Model		
Models	*Log Marginal Likelihood*	*Ranking*
All margins	-1158.72	1
Remove all margins	-1166.61	8
Remove τ_l margin	-1161.16	5
Remove τ_k margin	-1159.77	2
Remove τ_c margin	-1161.16	4
Remove τ_r margin	-1163.30	6
Remove τ_q margin	-1160.38	3
Remove τ_h margin	-1163.32	7
Remove $\{\tau_r, \tau_q, \tau_h\}$ margins	-1167.07	9
Panel B. Model with a Simple AR Process of Wage Mark-up Shocks		
Models	*Log Marginal Likelihood*	*Ranking*
All margins	-1165.00	1
Remove all margins	-1192.43	8
Remove τ_l margin	-1166.39	5
Remove τ_k margin	-1165.53	3
Remove τ_c margin	-1165.51	2
Remove τ_r margin	-1168.75	6
Remove τ_q margin	-1166.19	4
Remove τ_h margin	-1190.64	7

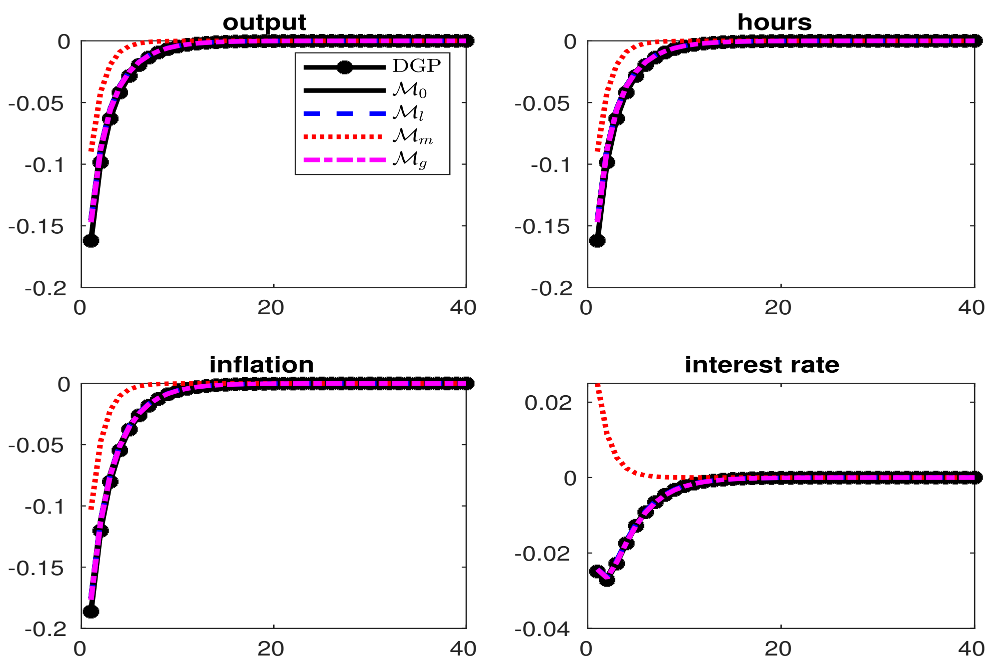
Notes to the table. The table shows values of the log marginal likelihood. Lower values of the likelihood denote models that are the most at odds with the data. Thus, the margins whose removal are associated with the lowest likelihood are the margins that are deemed the most necessary to explain the data.

Figure 1. Impulse Responses

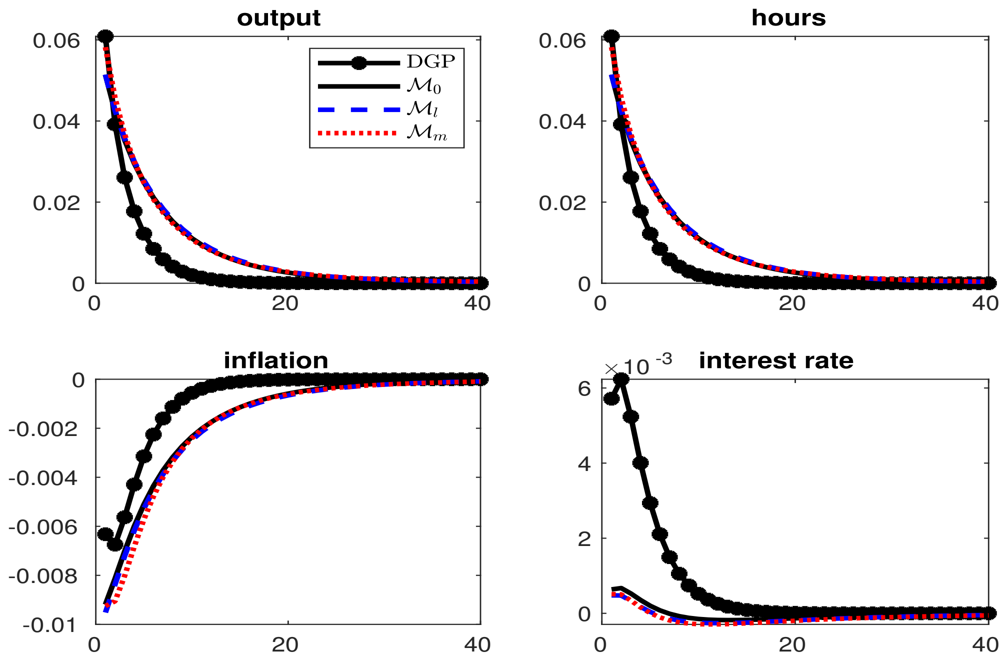
Panel A. Impulse responses to the technology shock



Panel B. Impulse responses to the monetary policy shock



Panel C. Impulse responses to the government spending shock



Panel D. Impulse responses to the labor supply shock

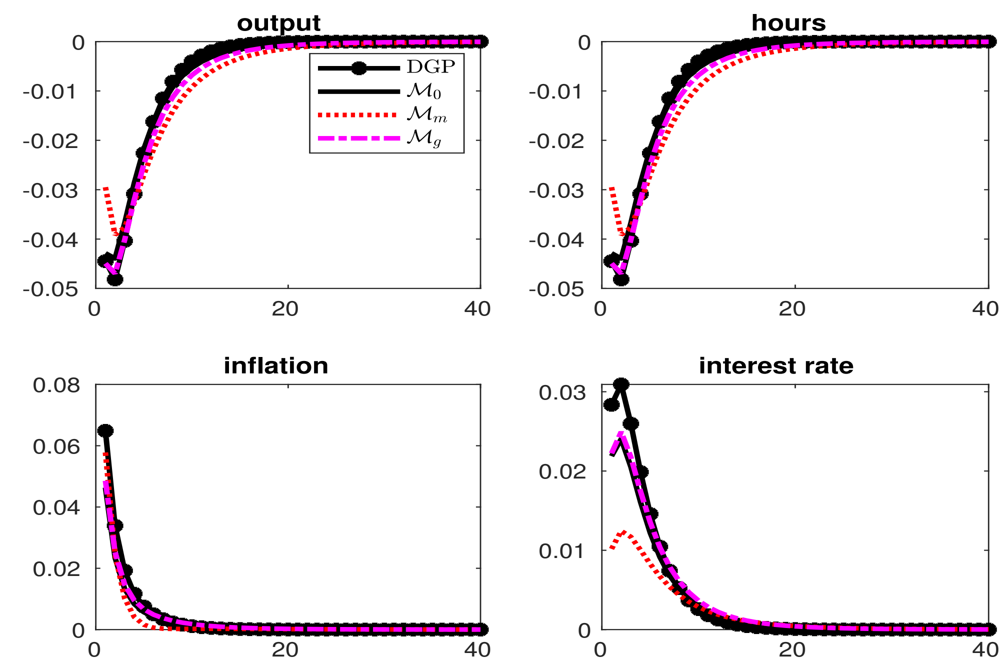
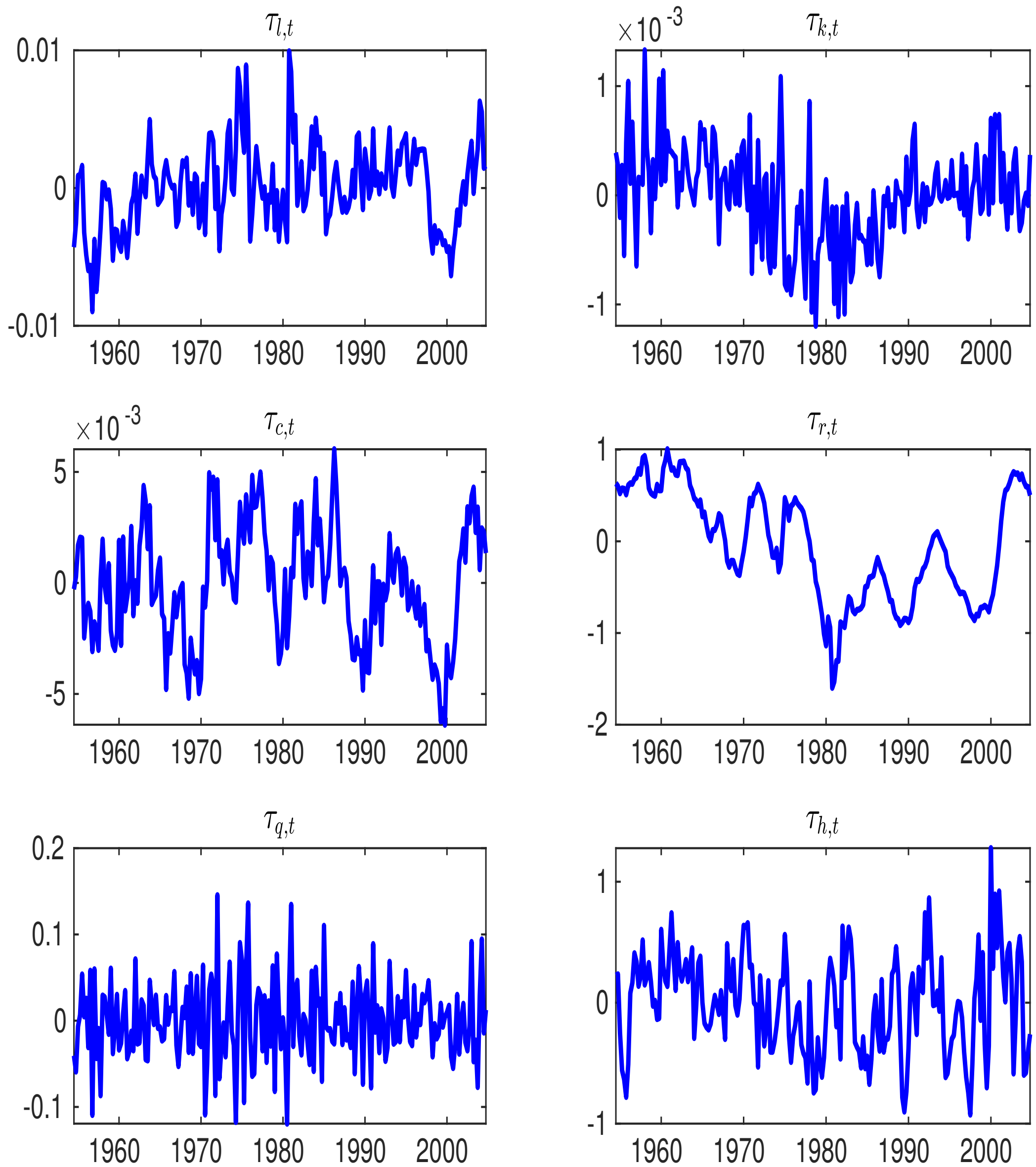


Figure 2. Margins Over Time (Smoothed)



Note to the figure. The figure plots the time series of the estimated margins over time.