Correlated Disturbances and U.S. Business Cycles*

Vasco Cúrdia
FRB San Francisco

Ricardo Reis
LSE

September 2020

Abstract

The dynamic stochastic general equilibrium (DSGE) models that are used to study business cycles typically assume that exogenous disturbances are independent autoregressions of order one. This paper relaxes this tight and arbitrary restriction, by allowing for disturbances that have a rich contemporaneous and dynamic correlation structure. Our first contribution is a new Bayesian econometric method that uses conjugate conditionals and Gibbs sampling to estimate DSGE models. It is considerably more efficient than conventional algorithms, in draws and time taken, so it makes estimation with correlated disturbances feasible. Our second contribution is to show that allowing for correlated disturbances can provide a specification check on models and robustify inferences about key parameters, pointing to directions over which the model is failing to endogenously match the correlation structure of the data.

JEL codes: E30, E10.

*Contacts: vasco.curdia@sf.frb.org and r.a.reis@lse.ac.uk. This is a significantly revised version of a paper that we circulated in 2010 and 2011. We are especially grateful to Marco Del Negro, Juan Rubio-Ramirez and Frank Schorfheide for helping us improve our method and to Craig Burnside, Fabio Canova, Lawrence Christiano, Martin Eichenbaum, Alejandro Justiniano, Jonathan Parker, Giorgio Primiceri, Chris Sims, Bent Sorensen, and Mike Woodford for useful comments and discussions. The views expressed in this paper are those of the authors and do not necessarily reflect positions of the Federal Reserve Bank of San Francisco or the Federal Reserve System. This project has received funding from the European Union's Horizon 2020 research and innovation programme, INFL, under grant number No. GA: 682288.

1 Introduction

A typical macroeconomic model proposes a theory for the behavior of economic agents that links a set of exogenous disturbances to predictions on a different set of endogenous variables. Because the disturbances are taken as given by the theory, by definition they are unexplained. The common practice in dynamic stochastic general-equilibrium (DSGE) models is to impose very strict assumptions on the processes driving disturbances, usually that they follow independent AR(1)s. This paper develops new estimation techniques for models with a rich correlation structure for the disturbance vector and applies them to robustify inferences on U.S. business cycles.

Our first contribution is a new Bayesian econometric technique to estimate dynamic macroeconomic models with potentially rich processes for the disturbances. We show that the economic structure of the models implies that key conditional posterior distributions belong either exactly or approximately to a family of conjugate distributions with known analytical form. We propose a new *conjugate-conditionals algorithm* that exploits this knowledge to efficiently characterize the posterior. The method relies on two well-known Bayesian techniques: Gibbs sampling and data augmentation. When applied to DSGE models that assume that disturbances are independent first-order autoregressions, AR(1)s, our method significantly speeds up estimation. Because the parameters associated with the disturbances are part of the conjugate conditional distributions, the efficiency gains are potentially larger with correlated disturbances because the algorithm takes advantage of this knowledge. Therefore, the method makes feasible the estimation of DSGE models that were previously prohibitively numerically costly by breaking a curse of dimensionality that plagues existing algorithms.

Our second contribution is methodological. In the simultaneous-equation reduced-form macroeconomic model tradition, there has long been a careful treatment of disturbances. Researchers routinely allow for rich dynamic cross and auto-correlations across disturbances, sometimes estimated non-parametrically. This literature has convincingly established that arbitrary restrictions on the disturbances can lead to inconsistent estimates of key parameters and impulse responses and can lead researchers astray in attempts to endogenize incorrectly-identified disturbances.[1] However, DSGE macroeconometric models arbitrarily assume that disturbances are independent AR(1)s, which is potentially dangerous for inference. From a Bayesian perspective, since researchers are typ-

[1] See Cochrane and Orcutt (1949), Zellner (1962), and Newey and West (1987) for the evolution on dealing with disturbances, and Fair (2004) for a recent careful application.

ically uncertain about the source and nature of the disturbances, generalizing the disturbance process ensures that this uncertainty is reflected in the posterior distribution.

We envision three possible uses for correlated disturbances, and illustrate these with applications to a real business cycle model, a new Keynesian model, and a medium-scale DSGE. First, allowing for correlated disturbances lets the data speak freely on the dimensions along which the model is inadequate. Therefore, it provides a check or test for model misspecification, as previously argued by Del Negro and Schorfheide (2009), and by Herbst and Schorfheide (2015) in their textbook treatment of correlated disturbances. We suggest that, after estimating a model with independent disturbances, a researcher should check whether allowing for correlated disturbances significantly improves fit and whether the estimated correlations are large and statistically significant. If so, the pattern of the estimated correlations should provide useful information regarding which part(s) of the model fail to endogenously match the data.

Second, allowing for more flexible specifications than the independent AR(1) should robustify inferences in DSGE models as reflected in the posterior for the economic parameters. The argument is similar to the practice of adjusting standard errors in linear regressions to allow for unknown heteroskedasticity and autocorrelation in the disturbances (Stock, 2010). It is even more important to be careful with the disturbances in DSGE models than in linear regressions, because correlations will lead to not just inefficient but also inconsistent estimates in these non-linear models. After allowing for correlated disturbances, researchers can compute revised posterior distributions and see in which direction they change.

Third, a researcher would like to be confident that the model captures the relevant co-movements among macroeconomic variables endogenously, without needing to rely on exogenous correlated disturbances. Examining the key propagation mechanisms of the model and its predictions for co-movement of variables when there are correlated disturbances should suggest whether this propagation happens endogenously or not. If not, the pattern of correlation of the disturbances can suggest the path to improving the model by including some new endogenous mechanism, towards the ultimate aim of a model with truly "structural" disturbances that are uncorrelated with each other and over time.

At the same time, researchers must be aware that this ultimate aim is a chimera. Decades of estimating DSGE models and of adding internal propagation mechanisms has not led the serial correlation of the assumed AR(1)s to be driven to the ideal of zero.

We are not proposing the use of models with correlated disturbances as an end in itself. Rather, we see our method as (i) providing a useful specification check, (ii) allowing for more robust inferences against the possibility that disturbances are correlated, and (iii) highlighting directions for improving the endogenous properties of the model. These are all that researchers can hope for in the process of building better models, pursuing an ultimate aim of a truly structural model that is unachievable. Allowing for correlated disturbances is an admission that the model is not correctly specified but, at the same time, models are never perfectly specified, so it is important to have techniques to estimate models with a rich correlation structure. If nothing else, this allows for a test of how far the shocks are from independence and it points to the directions in which to improve the model.

After a brief literature review and discussion of some issues, the paper is organized as follow. Section 2 presents the conjugate-conditionals estimation method for a broad class of equilibrium macroeconomic models. Section 3 presents three concrete models to which we will apply the method: a canonical real business-cycle model, a compact new Keynesian model, and a larger medium-scale DSGE. Section 4 compares the efficiency of our algorithm estimating these models relative to a standard Metropolis-Hastings approach. Section 5 discusses how to use the method to look for misspecification and robustify inferences. Section 6 concludes.

1.1 Literature review

The closest paper to this one is Ireland (2004). He adds measurement errors to the reduced-form equations of a DSGE model and allows them to follow a VAR(1), proceeding to estimate the model by maximum likelihood and to statistically test for structural stability. We differ in several respects. First, our focus is on the exogenous disturbances of the model, not on measurement error (which we will even abstract from). A key distinction between disturbances and measurement errors is that the properties of the disturbance process affect the behavioral responses of the agents in the model, whereas the properties of the measurement error only affect the job of the econometrician. For instance, if productivity disturbances are more persistent, agents in the model will engage in less intertemporal substitution in consumption and hours worked, altering the response of all endogenous variables. Instead, more persistent measurement errors only mechanically drive a difference between the endogenous variables and the observations. Second, from an econometric perspective, while both Ireland's and our approaches exploit the state-space representation of the model, Ireland's focus is on dealing with the measurement equation, while ours is on the state equation. Third, a few other differences are that we take a Bayesian approach, we allow for VARs of higher order than one, and we focus on implications for business cycles.

Del Negro and Schorfheide (2009) also emphasize the need for robustifying inferences from DSGEs. Their preferred approach is to merge the versatility of a VAR with the tight restrictions of a DSGE in an innovative method that uses the DSGE to provide priors for the VAR. They also contrast their approach with the alternative of allowing for flexible processes for the disturbances as we do. As they note, our approach fits into their general framework for dealing with misspecification in policy analysis. Their empirical analysis is constrained to independent AR(2) processes though, and part of their criticisms focus on researchers judiciously picking which correlations to model. We instead allow for a more flexible and more general correlation structure for the disturbances. More recently, Inoue et al (2020) and Den Haan and Drechsler (2020) pursue different methods to add disturbances to models in order to assess their misspecification. They follow our cue in imposing as little structure as possible on the disturbances.

Another related paper is Chib and Ramamurthy (2010). Like us, they use the insight of Gibbs sampling to propose an alternative to Metropolis estimation. However, while our blocks are suggested by the structure of the model, in their work it is the statistical properties of the data that guide the blocking of parameters. More concretely, at each step they cluster the parameters into arbitrary blocks to reduce the number of draws that are necessary to characterize the posterior distribution. Instead, our algorithm clusters the parameters into two groups, the economic parameters and those related to the disturbance processes. While our clustering is not efficient in the sense of minimizing the number of draws, as there may correlation between the two groups of parameters, its virtue is that one of the groups has exact or approximate conjugate distributions. Therefore, while we may need more draws than Chib and Ramamurthy (2010), our algorithm is easier to implement and faster to execute. More generally, one can see our algorithm as a variant of theirs, where our key contribution is to point out that DSGE models lend themselves to data augmentation and conjugate distributions that significantly speed up the estimation.

Finally, Chib and Greenberg (1994, 1995) develop Bayesian algorithms to estimate linear regression models with ARMA errors. Some of their conditional distribution results are similar to ours. We share the focus on estimating models allowing for general processes for the disturbances, and in the use of Gibbs sampling to make estimation feasible in spite of the large number of nuisance parameters. Differently, and crucially, we focus on DSGE models. As the booming literature of the last decade has shown, they pose significant new challenges to estimation relative to regression models.

A few papers have moved beyond the assumption of independent AR(1) disturbances, but typically in only special ways. Within closed-economy models, Chari, Kehoe, and McGrattan (2007) allow for a restricted VAR(1) where the productivity disturbance is special in that it Granger-causes all others, and Smets and Wouters (2007) allow two of their seven disturbances to follow an ARMA(1,1) and two others to be contemporaneously correlated. Schmitt-Grohe and Uribe (2010) find that a common shock to total factor productivity and investment-specific productivity explain an important share of the business cycle. More directly, when researchers have measured supposedly structural disturbances directly, they usually find them to be strongly correlated with each other (e.g., Evans, 1992, with military spending and Solow residuals).

In the open-economy literature, it is more common to assume that disturbances are correlated across countries, starting with the work of Backus, Kehoe and Kydland (1992). Justiniano and Preston (2010) estimate an open-economy DSGE model and find that correlated cross-country disturbances can partially account for the exchange rate disconnect puzzle. Rabanal, Rubio-Ramirez and Tuesta (2011) allow for cointegration among technological disturbances and find they can explain the volatility of real exchange rates.

As these papers on closed-economy and open-economy business cycles show, as models grow larger, with more disturbances and more emphasis on accounting for the data beyond just a few moments, there is a natural tendency to allow for correlated disturbances. We take a step further than this literature and allow for a richer and more general correlation between disturbances.

1.2 Three issues: simplicity, identification and orthogonalization

A natural objection to allowing for correlated disturbances is that it is harder to give them a structural interpretation than, say, AR(1) disturbances. While we are sympathetic with this objection, we are uncomfortable with its implications. Even though the estimates from independent AR(1)s for a vector of variables are simpler to interpret than those from a VAR, few (if any) researchers would argue in favor of the former instead of the latter. This apparent simplicity comes with estimation biases and incorrect inferences. Moreover, as the applications in this paper show, it is possible to interpret estimates with correlated disturbances. Once this is done, what becomes hard to understand is what was captured by estimates that assumed, for instance, that government spending was exogenous. Looking forward, we would expect that once researchers become used to models with correlated disturbances, this objection will become mute as it did just a few years after VARs became popular. In any case, the contribution of this paper is to argue that even when researchers prefer to assume independent disturbances, they should apply our specification check on whether their inferences are robust to allowing for correlated disturbances.

A second, more difficult, issue is identification. As noted by Sargent (1978) in estimating dynamic labor demands, it will often be difficult to empirically distinguish between endogenous sluggishness mechanisms, and exogenous persistent disturbances.[2] More generally, the issue is similar to the old argument (Griliches, 1967) that it is difficult to separately identify a linear regression with both a lagged dependent variable and an autocorrelated disturbance. Komunjer and Ng (2011) have provided a set of conditions for identification of DSGE models involving the rank of the information matrix, and which includes the case of correlated disturbances. In all of the applications of this paper, we exhaustively checked that their condition was satisfied, and did not find problems, but they will surely appear in other models.[3] Looking forward, we find compelling the argument that when there is an identification problem, the disturbance parameter responsible for it is set to zero so that the endogenous mechanisms have primacy in explaining the data.

Third, and related, whenever disturbances are contemporaneously correlated, one must orthogonalize them to produce impulse responses and variance decompositions. This is not a major objection since the arguments used in the VAR literature to argue for particular orthogonalizations can be directly applied to DSGEs with correlated disturbances. Of course, note that orthogonalizations are not restrictions, so this has no effect on the estimates or fit of the model, and that here it is the model disturbances not some VAR residuals that are being orthogonalized. An in-between alternative arises when disturbances are dynamically correlated, but contemporaneously uncorrelated with each other, and so orthogonalized this way. We explore this as well. More generally though, it is a virtue rather than a vice to bring attention to the need for thinking hard about iden-

[2] It is important to note that even if the exogenous disturbances could follow an arbitrary process, e.g. an infinite order VARMA, in many DSGE models, the economic and statistical parameters would still be identified. As noted by Sargent (1978) and many others, rational expectations models impose cross-equation restrictions that both identify the models as well as give them testable predictions.

[3] Reicher (2015) investigates identification more thoroughly within our model of correlated disturbances.

tification and orthogonalization in estimating DSGE models. These are central issues in all empirical work and should not be assumed away as the assumption of independent disturbances implicitly does.

2 The conjugate-conditionals method

A dynamic stochastic model links a vector of exogenous disturbances, s_t, to a vector of endogenous variables, y_t, of size n_s and n_y, respectively. The disturbances are hit by exogenous mean-zero innovations, e_t, and the researcher observes only a vector of observables, x_t, of size n_e and n_x, respectively. The econometrician observes a sample of these observables in $t = 1, ..., T$ with the convention that a variable dated t is determined at that date. The sample realization from $t = 1$ to date j is denoted by $x^j \equiv \{x_t\}_{t=1}^{j}$.

The model comes with two sets of parameters. First is the vector ε, which are the *economic parameters* determining preferences, technologies, and other constraints, of size n_ε. Second is the n_σ vector σ of *statistical parameters*, that determine the correlation structure and the volatility of the disturbances s_t. The distinction between the two will become clearer as we present the model. The inference problem is to characterize the distribution of these parameters given the data, that is the posterior: $p(\varepsilon, \sigma | x^T)$. Section 2.1 presents the economic model, section 2.2 derives the conjugate-conditionals algorithm, and section 2.3 discusses extensions.

2.1 The model

A broad class of dynamic macroeconomic models has the following structure:

$$y_t = \Lambda_1(\varepsilon, \sigma) y_{t-1} + \Lambda_2(\varepsilon, \sigma) s_t + \Lambda_3(\varepsilon, \sigma)(L) s_{t-1}, \tag{1}$$

$$s_t = \Phi(\sigma)(L) s_{t-1} + e_t, \text{ with } e_t \text{ i.i.d. and } Var(e_t) = \Omega(\sigma), \tag{2}$$

$$x_t = H_1(\varepsilon) + H_2(L) y_t + H_3(L) s_t. \tag{3}$$

The $\Phi(\sigma)(L) = \sum_{i=1}^{k} \Phi_i(\sigma) L^{i-1}$, a matrix lag polynomial of order k and similarly for $H_2(L)$, $H_3(L)$ and $\Lambda_3(\varepsilon, \sigma)(L)$. All the matrices are conformable, and their elements are functions of the sub-set of parameters of (ε, σ) that are indicated in brackets. We explain each of these relations in turn in the next three sub-sections.

2.1.1 The economic dynamics

Equation (1) in the model describes the economic (behavioral and accounting) relations between the endogenous variables and the exogenous disturbances. It nests most linear (or linearized) dynamic economic models that are described by a system of equations:

$$\Psi_0(\varepsilon)y_t = \Psi_1(\varepsilon)y_{t-1} + \Psi_2(\varepsilon)(L)s_t + \Psi_3(\varepsilon)w_t, \tag{4}$$

where the vector of endogenous disturbances w_t has the property that $\mathbb{E}_{t-1} w_t = 0$ and can capture terms involving $\mathbb{E}_t(y_{t+1})$. The Ψ_i matrices typically have many zero elements and have more elements than n_ε, embodying the cross-equation restrictions that come from optimal behavior, technologies and other constraints and which are affected by the economic parameters ε. As Blanchard and Khan (1982) and Sims (2002) among many others have shown, equation (1) is the solution, or reduced-form, of these models.

The matrices $\Lambda_i(\varepsilon, \sigma)$ in this solution are typically complicated non-linear functions of all the parameters. Therefore, while the model has a state-space representation, estimating it requires moving well beyond the standard techniques in state-space estimation (Durbin and Koopman, 2001).[4] While each particular DSGE imposes a tight link between parameters and the matrices in $\Lambda_j(\varepsilon, \sigma)$, across models, little that can be said in general. There is an important exception to this statement, and one that is crucial to the conjugate-conditionals method. By the principle of certainty equivalence, the parameters in the reduced-form solution of the model do not depend on the covariances in Ω. That is the $\Lambda_j(\varepsilon, \sigma)$ matrices depend on the parameters in $\Phi(\sigma)$ but not on the parameters in $\Omega(\sigma)$.

2.1.2 The exogenous law of motion for the disturbances

Equation (3) requires that disturbances are linear processes that are well approximated by a vector autoregression of finite order k. The contribution of this paper is to allow for $k \geq 1$, unrestricted Φ_i square matrices, and an unrestricted positive definite matrix Ω. Because s_t is exogenous, its correlations cannot be explained but must be assumed. It is then desirable to assume as little as possible on these measures of our ignorance and focus instead on the tight restrictions imposed by the model on the endogenous variables.

This leads to a much larger vector of statistical disturbances σ. Note that the param-

[4]There is another difference relative to state-space models that one should not get confused about. In our model and notation, y_t are not the state variables. Rather, y_t includes *all* of the variables in the economic model, including states, controls, or any other variable.

eters ε do not appear in equation (2), and it is this exclusion that defines the statistical parameters, σ. They are the collection of the parameters in $\Phi(L)$ and Ω that describe the dynamics of the disturbances. In many studies, the object of interest is the economic parameters, so that the elements of σ are viewed as nuisance parameters that must also be estimated.

An inevitable feature of this general approach is that the elements of e_t no longer have a structural interpretation without explicit orthogonalization assumptions as in the VAR literature. (However, inferences on the economic parameters ε are invariant to these orthogonalizations.) There is one way that some researchers may find useful of preserving a structural interpretation to the elements of e_t as innovations to each of the components of s_t: imposing the assumption that Ω is diagonal. Disturbances are then still dynamically correlated, but now assumed to be contemporaneously uncorrelated. This is similar to the seemingly unrelated regressions approach to dynamic models in time-series econometrics.

The literature has instead often assumed that $k = 1$, and both Φ_1 and Ω are diagonal, so each of the elements of s_t is an independent AR(1). One argument for this assumption is that it reduces the number of nuisance statistical parameters from $n_\sigma = kn_s + n_s(n_s+1)/2$ to $n_\sigma = 2n_s$. There is a curse of dimensionality as k increases, since the computational complexity of most estimation algorithms explodes even for modest values of k. However, as we will show, this is not a limitation of the theory, but rather of the particular algorithms being used.

2.1.3 Measurement and inference

Finally, equation (2) allows for a general linear link between model variables and data observables. In many applications, $x_t = y_t$ so the endogenous variables are observed.[5] Other times, $H_1(\varepsilon)$ includes the steady-state of the model, which depends on the economic parameters, while $H_2(L)$ and $H_3(L)$ are typically simple data transformations that adjust units. We abstract from measurement error in these observations to avoid confusion with the economic disturbances specified in the model. Including measurement error does not change our conclusions significantly, although it requires a clear distinction between them and the disturbances.

We now have all the ingredients to define the inference problem. Following the re-

[5] We will treat y_t as deviations from a steady-state, so we omit constants from (1)-(3), but it is straightforward to include these.

cent literature, we take a Bayesian perspective.[6] Starting with a prior distribution for the parameters, $q(\varepsilon,\sigma)$, the model in (1)-(3) provides a sampling distribution (or likelihood function) $f\left(x^T \middle| \varepsilon,\sigma\right)$ for the sample x^T, which defines, by Bayes rule, the posterior distribution for the parameters:

$$p(\varepsilon,\sigma| x^T) = f\left(x^T \middle| \varepsilon,\sigma\right) q(\varepsilon,\sigma)/p(x^T). \tag{5}$$

2.2 The conjugate-conditionals method to characterize the posterior

There is rarely an analytical form for the posterior distribution, so it must be characterized numerically. This is usually done with Markov Chain Monte Carlo (MCMC) algorithms, that draw a new (ε,σ) pair from an approximate distribution conditional on the last draw, in a way that ensures convergence of the J draws to the posterior distribution.[7] As the number of parameters grows, this suffers from a curse of dimensionality. We propose a new algorithm that breaks the curse by exploiting the economic structure of the model. Because its central observation is to use knowledge that some conditional posterior distributions are exactly or approximately conjugate, we label it the *conjugate-conditionals* algorithm. This section presents the algorithm in three steps. First, it introduces a few technical assumptions for it to hold, next it presents the results in which it lies, and finally it presents the algorithm.

2.2.1 Technical assumptions

We make the following technical assumptions:

Assumption. *The likelihood and prior functions satisfy the following restrictions:*
a) The distributions $p(\varepsilon| x^T,\sigma)$, $p(\sigma| x^T,\varepsilon)$ and $p(s^T| x^T,\varepsilon,\sigma)$ are not point masses, that is they are not degenerate in the sense of the random variables being almost surely constant.
b) $f(e_t| \varepsilon,\sigma)$ and $f(s^k| \varepsilon,\sigma)$ are normal distributions.
c) $q(\varepsilon)$ is a non-degenerate distribution, that is ε is not almost surely constant.
d) $q(\Omega)$ is an inverse-Wishart distribution and *$q(\Phi| \Omega)$is a normal distribution.*

Assumption a) requires that, given the observed data, there is more than one set of parameters that could have generated it with non-zero probability, so there is a legitimate

[6]Throughout the paper, we use $p(.)$ to denote a general posterior distribution, $f(.)$ to denote a sampling distribution, and $q(.)$ to denote a prior distribution.

[7] Readers that are very familiar with the literature on Bayesian estimation may find the next few pages tiresome, and are welcome to skip to the algorithm in section 2.3.

estimation problem. The assumption strengthens this basic requirement in two ways. First, it requires that conditioning on each of the two sub-sets of parameters, ε or σ, again we still have a non-trivial statistical estimation problem. If this was not the case, some of the steps in the algorithm would be trivial or redundant. Second, it requires that the observables are not enough to recover the disturbances. Otherwise the statistical problem would boil down to estimating the VAR in equation (2).

Assumption b) is standard in the literature: innovations are independent and identically normally distributed, and the initial unobserved states in the k lags of the VAR are also normal, so that the observations x^T are normally distributed.

Assumption c) puts only the weakest restriction on the prior for the economic parameters for our method to work. Finally, assumption d) sets the priors for the statistical parameters to the standard in the VAR literature, although not as common in the DSGE literature.

2.2.2 Two results on which the method rests

Our algorithm relies on two observations. First, by the principle of Gibbs sampling, we can break the sampling from the joint posterior at step j into drawing $\sigma^{(j)}$ from the conditional $p(\sigma^{(j)} \big| x^T, \varepsilon^{(j-1)})$ followed by drawing $\varepsilon^{(j)}$ from the conditional $p(\varepsilon^{(j)} \big| x^T, \sigma^{(j)})$. This well-known alternative to the random-walk Metropolis has here a natural application in separating statistical and economic parameters.

Moreover, while we are interested in the parameters, there is also uncertainty on the realization of the innovations e^T and thus the disturbances s^T. Focusing on the first Gibbs step, note that by the definition of a marginal distribution $p(\sigma| x^T, \varepsilon) = \int p(\sigma, s^T | x^T, \varepsilon) ds^T$, so drawing from the conditional for the statistical parameters is equivalent to drawing from the joint distribution for σ and s^T, retaining only the σ draws. This is often referred to in the statistics literature as data augmentation.

Finally, note that drawing from $p(\sigma, s^T | x^T, \varepsilon)$ can be split by Gibbs sampling again into drawing from $p(s^T | x^T, \varepsilon, \sigma)$ and $p(\sigma| x^T, \varepsilon, s^T)$ in succession. Combining all of these in a formal statement, we have:

Proposition 1. *Starting at step j with $(\varepsilon^{(j-1)}, \sigma^{(j-1)})$, then:*
a) drawing $\sigma^{(j)}$ from the conditional $p(\sigma^{(j)} \big| x^T, \varepsilon^{(j-1)})$ and then drawing $\varepsilon^{(j)}$ from the conditional $p(\varepsilon^{(j)} \big| x^T, \sigma^{(j)})$ converges in distribution to a set of draws from $p(\varepsilon, \sigma| x^T)$.
b) drawing $\sigma^{(j)}$ and $s^{T(j)}$ from the joint distribution $p(\sigma^{(j)}, s^{T(j)} \big| x^T, \varepsilon^{(j-1)})$, and storing only the

$\sigma^{(j)}$ draws gives a set of draws from $p(\sigma^{(j)} \big| x^T, \varepsilon^{(j-1)})$
c) drawing $s^{T(j)}$ from the conditional $p(s^{T(j)} \big| x^T, \varepsilon^{(j-1)}, \sigma^{(j-1)})$ and then drawing $\sigma^{(j)}$ from the conditional $p(\sigma^{(j)} \big| x^T, \varepsilon^{(j-1)}, s^{T(j)})$ converges in distribution to a set of draws from $p(\sigma^{(j)}, s^{T(j)} \big| x^T, \varepsilon^{(j-1)})$.

Proof: Result b) is the definition of a marginal distribution in relation to the joint distribution. Results a) and c) are applications of the convergence of the Gibbs sampler. The proof follows the same steps as Tierney (1994), where the crucial assumptions are a) and c) ensuring that the Markov chain defined by the Gibbs sampler is irreducible.⧫

The next observation focuses on result c) of the previous proposition. Conditional on the parameters, the model in equations (1)-(3) is a state-space system and the uncertainty on the disturbances s^T fits into a standard signal extraction problem. Therefore, the conditional distribution $p(s^T | x^T, \varepsilon, \sigma)$ is normal with mean and variance given by variants of the Kalman smoother (the disturbance smoother to be concrete). Moreover, conditional on the disturbances s^T, equation (2) is a standard vector autoregression. If the prior distribution Ω is an inverse-Wishart, then the posterior distribution is also an inverse-Wishart. In turn, if the variability in the innovations e_t is much smaller than the variability in the initial disturbances, then approximately all of the information about Φ in the system (1)-(3) is contained only in the second equation, and a normal prior for Φ leads to a normal posterior distribution.[8] The parameters of these distributions are known analytically: the mean and covariance of the normal distribution are the output of the disturbance smoother and the parameter of the inverse Wishart can be found in most Bayesian statistics textbooks (e.g., Geweke, 2005).

More formally:

Proposition 2. *The following two distributions belong to known families, with analytical means and variances:*
a) the posterior distribution for the disturbances, conditional on the data and the parameters, $p(s^T | x^T, \varepsilon, \sigma)$ is normal.
b) the posterior distribution for the variance of the innovations, conditional on the data, the other parameters, and the disturbances $p(\Omega | x^T, \varepsilon, \Phi, s^T) = p(\Omega | s^T)$, and it is an inverse-Wishart.

Proof: Equations (2)-(3) define a linear state-space system. Assumption b) states that innovations and initial conditions are normal. Therefore, the disturbances are normal,

[8]It is common practice to set the prior variance of the initial conditions equal to the unconditional variance predicted by the system. If the economic system has significant propagation and magnification, then this variance should be considerably larger than the variance of the innovations.

proving result a). To prove the second result, note that by certainty equivalence, only equation (2) involves the covariance matrix Ω. Moreover, no x^T or ε appear in that equation. It therefore follows that $p(\Omega|\, x^T, \varepsilon, s^T, \Phi) = p(\Omega|\, s^T)$. But then, using assumption d), it is a standard result from linear regression that, since the prior is an inverse-Wishart, so is the posterior.♦

2.2.3 The conjugate-conditionals method

Based on the two results, the output of the following hybrid, Metropolis-within-Gibbs (or block-Metropolis) algorithm will converge to a set of draws from the posterior distribution of the parameters:

Algorithm *At draw j:*
Step 1) draw $s^{T(j)}$ *from* $p(s^{T(j)} \big|\, x^T, \varepsilon^{(j-1)}, \sigma^{(j-1)})$*, the known distribution in proposition 2;*
Step 2) draw $\Omega^{(j)}$ *from* $p(\Omega|\, s^{T(j)})$*, the known distribution in proposition 2;*
Step 3) draw $\Phi^{(j)}$ *from a proposal distribution that approximates* $p(\Phi|\, x^T, \varepsilon^{(j-1)}, s^{T(j)}, \Omega^{(j)})$ *and accept or reject this draw with a Metropolis-Hastings probability;*
Step 4) draw $\varepsilon^{(j)}$ *from a proposal distribution that approximates* $p(\varepsilon|\, x^T, \sigma^{(j)})$ *and accept or reject this draw with a Metropolis-Hastings probability.*

The first two steps are easy even for a very large number of disturbances n_s, number of lags, k, and number of observations T. Most software programs can take draws from the multivariate normal quickly and, while the Kalman filter recursions can take some time, they were required anyway in order to calculate the likelihood function of the problem. The Kalman smoother provides the posterior means and variances recursively.[9]

As for the third step, we do not have the exact distribution, but we have a good guess. The autocorrelation parameters Φ enter both the reduced-form solution of the model in equation (1), as well as the VAR in equation (2). But, if the variance of the innovations e^T is much smaller than the variance of the prior for the initial states and endogenous variables, then this filtering problem has an approximate solution where only the information in the VAR is relevant. That is, in this limit case, $p(\Phi|\, x^T, \varepsilon, s^T, \Omega) \approx p(\Phi|\, s^T, \Omega)$. But then, we have another conjugate conditional, since the posterior for Φ is also normal and the formulae for the mean and variance are the standard linear regression formulae.

We have found that a particular implementation of this approximate proposal works

[9]Carter and Kohn (1994) show that sampling from the joint distribution is considerably more efficient. We use their approach as described in Chib's (2001) algorithm 14.

remarkably well, converging quickly. Following Geweke (1989), we use an independence-Metropolis step sampling from a t-distribution instead of the normal in the previous paragraph, to allow for fatter tails. To be clear, recall that this is the proposal density, so there are many other alternatives that would lead to consistent estimates. What the argument in the previous paragraph strongly suggests, and our experience confirms, is that a t-distribution for $p(\Phi | s^T, \Omega)$, using the mean and variance from normal conjugate formulae, provides a good approximation to the target distribution $p(\Phi | x^T, \varepsilon, s^T, \Omega)$, as judged by how quickly the draws converge and the very high acceptance rate that we obtain for the draws.[10]

Finally, for step 4, our algorithm does not make any significant improvement over the literature.[11] We neither know $p(\varepsilon | x^T, \sigma)$, nor is there any hope of having even an approximate result beyond very specific models, since the parameters ε usually enter the system in a highly non-linear way. In practice, we used a random-walk Metropolis for this step, drawing $\varepsilon^{(j)}$ from a normal with mean $\varepsilon^{(j-1)}$ and covariance matrix equal to the inverse-Hessian at the mode of the posterior, scaled to reach an acceptance rate around one quarter. We have tried several alternatives: independent Metropolis, rejection sampling, and modifying the random-walk Metropolis to have the new draws depend on $\sigma^{(j)}$. None of these clearly dominated the more conventional random-walk Metropolis.[12]

To conclude, the conjugate-conditionals algorithm draws from the expanded parameter vector $(\varepsilon, \sigma, s^T)$ in turn, exploiting the knowledge that the conditional distribution for s^T is known, while we have a good guess for the conditional distribution for σ. Allowing for correlated disturbances may dramatically increase the number of parameters in σ, but because the conditional posterior distribution for the covariance matrix is known analytically, and because we have a good approximating distribution for the conditional posterior distribution for the correlation coefficients, the curse of dimensionality is broken. Estimating a DSGE with correlated disturbances is not significantly harder than one with independent AR(1) disturbances, because it is not harder to draw from normals and inverse-Wishart distributions of higher dimension. Because it uses our knowledge of particular slices of the posterior distribution that we are trying to characterize, this algorithm

[10] Some readers, accustomed to the Metropolis-Hastings random walk sampling often used in economics, may find it puzzling that we refer to high acceptance rates as a measure of efficiency. This is the case because we are using *independence* Metropolis.

[11] Note, that since steps 1-3 and 4 are the two blocks of one Gibbs sampler, we could move step 4 before the other three.

[12] Meyer-Gohde and Neuhoff (2015) extend our method by replacing the Metropolis with a recursive jump Markov Chain Monte Carlo.

should be more efficient than the standard Metropolis algorithm.[13]

2.3 Relaxing the assumptions

Some features of the setup are central to our method. They are also central to the definition of a DSGE model. First, our argument applies only to models with a state-space linear set of equations (1)-(3) where the matrices $\Lambda_i(\varepsilon, \sigma)$ are non-linear functions of the parameters, with many restrictions imposed by the theory. Second, the assumption that the economic parameters ε do not affect the law of motion for the disturbances in equation (2) is crucial for the ability to deal separately with the two types of parameters, but it is as much an assumption as it the definition of what σ and ε are. Third, the principle of certainty equivalence is important to draw separately from these two classes of parameters, but it applies to all linearized DSGE models.

The other assumptions can be relaxed in many ways. Starting with equation (3), we could allow for $H_2(\varepsilon)(L)$ and $H_3(\varepsilon)(L)$, so the measurement equation linking the endogenous variables to the observable can depend on the economic parameters. This requires no change in the algorithm, although we have trouble finding economic models to which this extension would be useful.

Turning to the assumptions on prior distributions, they can be somewhat relaxed. There are alternative conjugate priors to the normal-inverse-Wishart family. Kadiyala and Karlsson (1997) discuss combinations of diffuse, normal, Wishart and Minnesota prior distributions that deliver conjugate families for VARs. Sims and Zha (1998) propose an alternative, with a normal conjugate family for the distribution of Φ conditional on Ω, which puts fewer restrictions on the prior variance than the one in our assumption and has some computational advantages, although the posterior for the covariance matrix Ω stops being conjugate.

The assumption that is the focus of this paper is the unrestricted VAR for the disturbances. There are two alternatives, partly discussed in section 2.2. The first is to have disturbances follow independent AR(k)s, so the Φ_k and Ω matrices are all diagonal. Adapting the priors in the assumption to $i = 1, ..., n_s$ independent normals for $[\Phi_j(i)]_{j=1}^k$, and $i = 1, ...n_s$ independent inverse-gammas squared for each of $\Omega(i)$, our results follow. Our algorithm can then be used in models that make the dominant assumption in literature

[13] The statement has to be qualified, because it is possible that the co-dependence between ε and σ is so strong that the Metropolis algorithm ends up dominating the Gibbs-sampler. In our experience, this is not the typical case.

that disturbances are independent AR(1)s, as well as higher-order autoregressions.

The second case is to have dynamic but not contemporaneous correlation, so the Φ_j are unrestricted but the Ω must be diagonal. In this case, using the normal priors for Φ from the assumption, and the independent inverse gamma priors for $\Omega(i)$ just described, again our results follow.

More generally, we may wish to impose that some of the elements of Φ_j and Ω are either zero, or appear more than once in the matrices. In this case, equation (2) is a system of seemingly unrelated regressions (SUR). Collecting the disturbances into the vector $\bar{s}$ of size $n_s(T-k)$, it is written as $\bar{s} = Z\beta + \varepsilon$, with $\varepsilon \sim N(0, \Omega \otimes I_{t-k})$, where Z contains the lagged states as well as blocks of zeros allowing for a rich set of restrictions on the VAR. The coefficients β include the elements of Φ. As long as the prior for $\beta | \Omega$ is normal and the prior for Ω^{-1} is the Wishart distribution as described in assumption 2, then our results on conjugate distributions still hold (Zellner, 1962).

Yet another possibility is to allow the disturbances to have a factor structure. Caldara et al (2014) extend our method to allow for this case.

Finally, in some applications, the researcher may want to impose the constraint that the VAR in equation (2) is stationary. This affects the distribution for Φ in step 3, which is now truncated to the stationarity region. However, our experience is that still using as proposal the t-distribution based on the approximate-normal result, but truncating it to only accept stationary draws, has almost no effect on the performance of the algorithm. This is not entirely surprising; the truncation does not affect the relative density of different draws in the stationary region, so it has little effect on the importance sampling algorithm.

3 Three DSGE models and the data

This section presents three familiar DSGE models in which we now allow for correlated disturbances. The log-linearized solution to each of them takes the form in equation (1), and for each we will allow for general VAR disturbances as in equation (2).

3.1 A real business cycle model

The best-known and simplest DSGE model is due to Prescott (1986), extended to include government spending following Baxter and King (1993) and Christiano and Eichenbaum

(1992). This model has three merits for our purposes. First, it is sufficiently simple that the effect of correlated disturbances can be grasped intuitively. Second, it only has two structural parameters to estimate, making it easy to transparently evaluate the efficiency of the method. And third, it is a canonical model that many macroeconomists carry in their mind, so we can re-examine some classic puzzles that it has generated from the perspective of correlated shocks.

A social planner chooses sequences of consumption and hours, $\{C_t, N_t\}_{t=0}^{\infty}$, to maximize:

$$\mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t \left\{ \frac{\left[C_t \left(1 - N_t\right)^{\theta} \right]^{1-1/\gamma} - 1}{1 - 1/\gamma} + V(G_t) \right\} \right], \tag{6}$$

subject to

$$Y_t = C_t + K_t - (1-\delta)K_{t-1} + G_t, \tag{7}$$

$$Y_t = \left(A_t N_t\right)^{1-\alpha} K_{t-1}^{\alpha}. \tag{8}$$

The notation is standard.[14] Utility increases with consumption and leisure and the benefits of government spending enter additively through the function $V(.)$, so they have no effect on the positive predictions of the model. Equation (7) states that output equals consumption plus investment plus government spending, and equation (8) is a neoclassical production function.

The log-linearized solution to this model takes the form of equation (1). Some of the parameters are easily pinned down by steady-state relations[15] Two of the parameters are not, and they are crucial to the model's business-cycle predictions. First, the elasticity of intertemporal substitution, γ, determines the willingness of households to shift resources over time. It is a key determinant of how strongly savings and labor supply respond to persistent productivity changes, and thus of the model's ability to generate sizable output fluctuations. Second, the parameter θ pins down the steady-state elasticity of labor supply

[14]In particular: C_t is private consumption, G_t is government consumption, N_t is the fraction of hours in a quarter spent at work, K_t is capital, Y_t is output, A_t is total factor productivity, β is the discount factor, γ is the intertemporal elasticity of substitution, θ determines the relative utility from leisure and consumption, δ is the geometric depreciation rate, and α is the labor share.

[15]Namely: the discount factor, $\beta = 0.995$, to generate a steady-state risk-free annual real interest rate of 2%, the production parameter, $\alpha = 0.33$, to match the capital income share, the depreciation rate $\delta = 0.015$ to roughly match econometric estimates and the average U.S. capital-output ratio, the average level of productivity is normalized $\bar{A} = 1$, and the average government spending $\bar{G} = 0.2$, its historical average share of GDP.

with respect to wages. It is the key determinant of the size of the fluctuations in hours worked. These are the economic parameters in the vector $\varepsilon = (\gamma, \theta)$. Their priors have a gamma distribution and follow conventions in this literature in their modes: $\gamma = 2/3$ and $\theta = 4.85$ (to generate a steady-state value of 0.2 for N).

Turning to the data, and so the specification of equation (3), this simple model has been used to account for movement in two variables: output and hours worked. So, in log-deviations from the steady-state $x_t = (\hat{Y}_t, \hat{L}_t)$. Following the RBC literature, we use U.S. data for non-farm business sector hours and output per capita that is quarterly, HP-filtered, and goes from 1948:1 to 2008:2, although we use the data before 1960:1 only to calibrate the priors.

Finally, the disturbances are $s_t = (\ln(A_t), \ln(G_t/\bar{G})) \equiv (\hat{A}_t, \hat{G}_t)$. In all advanced economies, government spending is certainly not an independent process in the data, and via the payment of unemployment benefits or countercyclical fiscal policy, G_t typically responds to A_t at least with a lag, and persists over time through slow fiscal adjustments. In the other direction, perhaps private productivity responds with a lag to some forms of government spending like infrastructures or the enforcement of contracts. There is a strong prior case for allowing these disturbances to be correlated, especially in such a simple model.

3.2 A new Keynesian model

The second model we consider is the canonical 3-equation NK model used to explain the co-movement of inflation, output, and nominal interest rates. We use the version of this model in Cúrdia et al (2015), which adds three sources of inertia to the model: interest-rate smoothing, internal habit formation, and indexation to the past of sticky prices. With these, this otherwise textbook model fits the data reasonably well.

The three equations, mapping into the general model of equation (1) are:

$$\tilde{x}_t = \mathbb{E}_t(\tilde{x}_{t+1}) - \varphi_\gamma^{-1}\left[i_t - \mathbb{E}_t(\pi_{t+1}) - r_t^e\right] \tag{9}$$

$$\pi_t - \zeta\pi_{t-1} = \xi\left(\omega x_t^e + \varphi_\gamma \tilde{x}_t\right) + \beta\,\mathbb{E}_t(\pi_{t+1} - \zeta\pi_t) + u_t, \tag{10}$$

$$i_t = \rho i_{t-1} + (1-\rho)\left[r_t^e + \pi_t^* + \phi_\pi(\pi_t - \pi_t^*) + \phi_x x_t^e\right] + \varepsilon_t^i, \tag{11}$$

where: $\tilde{x}_t$ is the marginal utility of consumption; x_t^e is the efficient output gap, i_t is the nominal interest rate, π_t is inflation, and π_t^* is a time-varying inflation target, all in log-deviations from their steady states. The first equation above is a standard Euler equation, or intertemporal IS, whereby the marginal utility of consumption today is equal to its expectation next period minus the real interest rate. The second equation is a standard Phillips curve relating the gap between actual inflation and its indexed component to marginal costs that depend on the output gap, and its expectations next period. The third equation is a Taylor rule with smoothing of interest rates.

The model has three more auxiliary equations for the dynamics of the exogenous efficient levels. Letting $x_t^e = \tilde{y}_t - \tilde{y}_t^e$, then $\tilde{y}_t$ is output and y_t^e is its efficient level, the equilibrium value in a counterfactual economy with no price rigidities and no monopoly distortions. There is likewise an efficient real interest rate r_t^e. These evolve according to:

$$\tilde{x}_t = \left[x_t^e - \eta_\gamma \left(\tilde{y}_{t-1} - \tilde{y}_{t-1}^e\right)\right] - \beta\eta_\gamma \, \mathbb{E}_t \left(x_{t+1}^e - \eta_\gamma x_t^e\right), \tag{12}$$

$$\tilde{y}_t^e = -\frac{\varphi_\gamma}{\omega} \left\{\tilde{y}_t^e - \eta_\gamma \left(\tilde{y}_{t-1}^e - \gamma_t\right) - \beta\eta_\gamma \, \mathbb{E}_t \left[\tilde{y}_{t+1}^e + \gamma_{t+1}^y - \eta_\gamma \tilde{y}_t^e\right]\right\} + \frac{\beta\eta_\gamma\omega^{-1}}{1-\beta\eta_\gamma} \, \mathbb{E}_t \, \delta_{t+1}^y, \tag{13}$$

$$r_t^e = \mathbb{E}_t \, \gamma_{t+1}^y + \mathbb{E}_t \, \delta_{t+1}^y - \omega \left(\mathbb{E}_t \, \tilde{y}_{t+1}^e - \tilde{y}_t^e\right). \tag{14}$$

The first equation defines the marginal utility as a function of output, and the next two equations define the efficient levels of output and the real interest rate, respectively, relative to which the gaps in the model are defined. These 6 equations constitute the 6 equations captured in our general specification in equation (1).

Turning to the economic parameters, they are listed, together with their priors in table 1. All the values mimic those in Cúrdia et al (2015), and are common in the NK literature.

The model is used to explain data on output, inflation, and nominal interest rates: $x_t = (y_t, \pi_t, i_t)$, defining equation (3) in the general setup. To measure them, we use quarterly data on unfiltered series for the log-change in real GDP, the log-change in the GDP deflator, and the federal funds rate from 1987:3 to 2009:3.

Finally, the focus of this paper are the five disturbances $s_t = (\varepsilon_t^i, \pi_t^*, u_t, \gamma_t, \delta_t)$, to the monetary policy rule, the time-varying inflation target, markups, productivity, and preferences, respectively. Usually they are each assumed to follow an independent AR(1), sometimes further restricted so that the autoregressive coefficient is zero. But again, these assumptions are easily rejected in the macro series. To give but one example, with such a simple policy rule, the shocks to monetary policy are surely partly a response, and so correlated with, changes in the other three shocks. Or, for another example, any attempt at measuring productivity and markups directly has found the two to be strongly dynamically correlated.

3.3 A medium-scale DSGE

Our final model is the medium-scale DSGE model estimated by Smets and Wouters (2003, 2007). It provides a good fit to the data, and it has become influential in the study of business cycles and monetary policy.

The model involves much new notation, and it is reasonably well known, so we relegate its presentation to an appendix. Like the other models, it fits into the general setup for our method in equation (1). The economic parameters, and their respective priors, are likewise stated in the appendix, closely following the choices of Smets and Wouters (2003, 2007). Likewise, for the data, we mimic their choices, with only a few minor exceptions detailed in the appendix.

This model has seven exogenous disturbances: total factor productivity, investment-specific productivity, risk premium, government spending, monetary policy, price markups, and wage markups. Following the DSGE tradition, Smets and Wouters assume that they all follow independent AR(1)s, but already in order to fit the data, they allowed for two exceptions. First, they included two first-order moving average terms for the price and wage markup disturbances to fit high-frequency movements in the data. Second, they allowed for contemporaneous correlation between government spending and total factor productivity. So, the case for correlated disturbances was already in this original DSGE model, but done so by highly restrictive prior restrictions taken ex post to improve model fit.

3.4 Correlated disturbances

The focus of this paper is on allowing the disturbances to follow a vector autoregression of order k as in equation (2) where the Φ_i and Ω are unrestricted matrices and σ is the vector of statistical parameters. Table 2 lists the prior distributions for these parameters.

In the RBC model, there are $4k+3$ elements in σ. We estimated VARs of orders 1 to 6 with very similar results. While the marginal likelihood is higher for order 6, we focus on the $k=1$ case because the results are easier to interpret and the difference in marginal likelihood is less than 3 log points. The prior mode for the four AR(1) parameters (the diagonal terms of Φ_0 and Ω_0) is set to match four moments in the data before 1960: the variances and serial correlations of output and hours. For the three remainder statistical parameters, the non-diagonal elements in Φ_0 and Ω, the prior mode is zero, but they have a precision according to the extension of the Minnesota prior discussed in Kadiyala and

Karlsson (1997), tighter around zero the further away we move from the diagonal.

In the NK model, we also concentrate attention on the VAR(1), and likewise set the mode of the non-diagonal elements of Φ_0 and Ω at zero with a Minnesota prior. For the ten diagonal elements we choose priors comparable to those used in Cúrdia et al (2015). The (self) autocorrelation coefficients are centered at 0.5 for the non-monetary policy disturbances, 0.2 for the policy rule shock, and 0.95 for the (slow moving) time-varying inflation target. The standard deviations of the innovations are centered at one percentage point annualized, with the exception of the inflation target which is centered at half a percentage point.

For the Smets-Wouters DSGE model, we take a different approach. We impose the restriction, already discussed in section 2 that the disturbances are dynamically, but not contemporaneously correlated. This allows us to discuss this approach to orthogonalization and slightly reduces the number of parameters in this large model that takes time to estimate. Because there are so many parameters, we list the priors for this model in the appendix.

For the study of efficiency, we also consider versions of the models with independent AR(1)s. In those cases, the prior variances of the off-diagonal elements of Φ_0 and Ω_0 are set to zero.

4 The computational efficiency of the method

This section evaluates the efficiency of our estimation method against the common alternative in the literature: the random-walk Metropolis. At step j, it draws a proposal $(\varepsilon, \sigma)^{(j)}$ from a normal density with mean $(\varepsilon, \sigma)^{(j-1)}$ and some pre-defined covariance matrix, accepting this draw with a probability that depends on the ratio $p(\varepsilon, \sigma)^{(j)} / p(\varepsilon, \sigma)^{(j-1)}$, keeping $(\varepsilon, \sigma)^{(j-1)}$ in case of rejection. This algorithm is robust in the sense that it usually explores well the posterior distribution with minimal input from the researcher. The other side to this robustness is that, because it uses almost no knowledge of the shape of the posterior, the algorithm can take many draws to converge. Experience with DSGE models has found that it can take millions of draws to converge if there are more than ten parameters to estimate, a bound that is quickly crossed with correlated disturbances.[16]

We compare the methods both on actual data, as well as on average across ten Monte Carlo samples of simulated data. Each was generated by first taking a random draw for

[16] For an alternative sequential Monte Carlo algorithm, see Herbst and Schforheide (2014).

the parameters from their posterior distribution, and then taking random draws to the disturbances in the model. We do this for the RBC and the NK models. The larger-scale DSGE takes a long time to estimate for even one sample, and often fails to converge for the usual algorithm making comparisons difficult.

The comparison is along two metrics. The first is how many draws it took to have at least 300 effective draws, as defined by Geweke (1992). Measuring effective draws involves adjusting for the serial correlation across draws to provide a measure of the number of independent-looking draws. We have tried cut-offs other than 300, and four alternative measures of convergence of the algorithm, all with very similar results. The second is how long it took on a standard computer to achieve this convergence. Computational times are more unreliable because they are too dependent on hardware, software, and coding specifics. We still report them because an individual draw can take longer to take in one method relative to another.[17]

The results are in table 3, broken by panels according to the model that is estimated. Starting with the RBC model in the top panel, assuming independent AR(1) disturbances, the conjugate-conditionals method takes approximately one-eighth of the draws to converge on its estimates in the data as the conventional algorithm. The conjugate distributions take longer to evaluate in each draw, however, so the time savings are smaller but still substantial, with a reduction of 57% in the time to deliver estimates. The Monte Carlo samples reinforce the efficiency gains of the conjugate-conditionals algorithm: it lowers the number of draws on average by 92% and the time taken by 73%.

With VAR(1) disturbances, there are now 9 instead of 6 parameters to estimate. In this case, the method provides no consistent gains relative to Metropolis-Hastings in terms of number of draws, and it takes about 3 times longer to run.

The second panel estimates the NK model. With independent AR(1) disturbances, the efficiency gains in draws are 71% with the actual data and 94% with simulated data. On time, the conjugate-conditional method is on average 76% faster, mostly because sometimes the two methods are just as fast, but sometimes, the traditional algorithm performs very poorly. In this case, where there are more parameters at play, the VAR(1) case makes the efficiency of the algorithm more visible. The traditional method most often fails to

[17] In terms of implementation details: we run four parallel chains, and discard the first 10,000 simulations in each to remove the influence of initial conditions. The proposal density for ε in the conjugate-conditionals algorithm is a random-walk Metropolis. The covariance matrix for the Metropolis algorithm is the Hessian at the mode of the posterior (found by numerical maximization), multiplied by a scale factor to obtain approximately a 25% acceptance rate. We fine-tune this proposal density after the number of effective independent draws reaches 50.

converge after 40 million draws, at which point we stop it, so that in practice it is infeasible to estimate a NK model with correlated disturbances using traditional methods. This barrier is removed with the conjugate-conditionals method.

Overall, the evidence leads to three conclusions. First that the benefits of the conjugate-conditionals method are visible even with independent AR(1) disturbances across all models. Second, that the method leads to large efficiency gains in terms of number of draws (with the exception of the RBC-VAR1 model) and in the case of the NK-VAR model, it makes estimation feasible where the standard algorithm fails. Third, that the efficiency gains in terms of time are smaller, but still often sizeable.

5 Inference with correlated disturbances

In the previous section, we showed that estimating models with correlated disturbances is feasible. Now, we argue that this is desirable from three perspectives. First, because the macroeconomic data strongly prefers models without arbitrary AR(1) restrictions, and rejects specifications with these restrictions. Second, because the posterior estimates of the economic parameters are significantly affected by imposing these restrictions on the nuisance statistical parameters. And third, because an assessment of the internal propagation properties of these dynamic models changes when we allow for correlated disturbances.

5.1 The impact of correlated disturbances in the RBC model

Table 4 reports the posterior distributions for the RBC model with either AR(1) or VAR disturbances.

Misspecification. The log marginal predictive density of the model is 26 points higher with correlated disturbances than with independent AR(1)s. Therefore, the posterior odds ratio is an overwhelming e^{26} in favor of the former.[18] This is in spite of the prior with correlated disturbances being centered around the independent-disturbances model, with shrinking variances as one moves towards the cross-correlations. Therefore, the marginal likelihood would, all else equal, favor the more parsimonious AR(1) specification.

The data estimates of the cross-correlation terms also overwhelmingly reject the zero restrictions. The three non-diagonal terms of the Φ and Ω matrices do not include zero in

[18]Here and everywhere, we calculated the marginal likelihood using a harmonic mean with the truncated multivariate normal distribution as the weighting function, as described in Geweke (2005).

their 90% credible sets. The largest correlated-disturbance term is the lagged productivity term in the law of motion for government spending. According to these estimates, when productivity falls, there is a lagged increase in government spending. This matches what we would expect from the automatic and discretionary stabilizers in U.S. fiscal policy.

Posteriors for economic parameters. Turning to the economic parameters, with independent disturbances, figure 2 plots a view from above of the bivariate histogram for the two economic parameters in the posterior distribution, under independent or correlated disturbances. There is remarkably little overlap between the two posterior distributions, and the posterior with independent disturbances is very far from the prior.

Focusing on the intertemporal elasticity of substitution, the mean of the posterior with independent disturbances is 1.4. This is not just above the prior, but it is especially substantially higher than the usual value of 0.2 that comes from Euler-equation estimates (Hall, 1988, Yogo, 2004). With correlated disturbances instead, the elasticity of intertemporal substitution is much lower, with a mean of 0.43 and a 5% bound of 0.29, bringing the DSGE estimates in line with the single-equation Euler equation estimates. Figure 2 strikingly shows how distorted inferences would be by imposing the independence restriction.

Inferences on internal propagation. With only two variables and two disturbances, only one orthogonalization condition is needed and it is easy to check alternatives and their implications for impulse responses and variance decompositions. Following Evans (1992), we use a Choleski decomposition with the innovations to government spending ordered first. This is an orthogonalization, not a restriction, and it applies to the disturbances, not the variables.[19] Figure 3 plots impulse response responses to one standard-deviation innovations to the two disturbances, with the legend showing the median unconditional variance decomposition between parentheses. With independent disturbances, the RBC model suffer from three well-known problems in its internal propagation when compared with reduced-form estimates from VARs.

1) The output persistence puzzle. In response to an improvement in productivity, output increases both because of the higher productivity, and also because the representative household chooses to work longer today when the returns to working are higher. How-

[19]We also tried ordering productivity first, as well as estimating a model with dynamic but not contemporaneous correlation between the disturbances. The solution of the three puzzles was robust to these alternatives. The results are also robust to the order of the VAR.

ever, as Cogley and Nason (1995) noted, the persistence of the output response closely mirrors the persistence of the productivity disturbance, whereas most reduced-form estimates of these responses are more gradual. Figure 3 shows that with correlated disturbances, the response of output to a productivity disturbance is significantly more delayed. An increase in productivity now leads to a subsequent fall in government spending. While this initially makes the impact on output smaller, after a few periods, it boosts output up partially solving the output persistence puzzle.

2) The hours-productivity puzzle. Gali (1999) first, and Dedola and Neri (2007) more recently, estimated that hours fall after improvements in productivity. In figure 3 though, with independent disturbances, hours increase strongly after a productivity improvement. With correlated disturbances, instead, an improvement in productivity has a delayed and persistent contractionary effect on hours. While the improvement in productivity increases hours, the subsequent fall in government spending lowers them and the net impact is close to zero, matching the results from the VAR literature.

3) The sources-of-business-cycles puzzle. According to the variance decompositions with independent disturbances, government spending disturbances account for half of the variance of output and most of the variance of hours. The findings in typical VAR studies (e.g., Shapiro and Watson, 1986, Fisher 2006) instead which attribute a larger role to productivity.[20] With correlated disturbances, productivity accounts for a much larger fraction of the business cycle. Much of the earlier predominance of government spending was due to its response to productivity. In line with the VAR evidence, productivity now accounts for three quarters of the variance of output and 64% of the variance of hours.[21].

Conclusion on the RBC model. Introducing correlated disturbances improves the fit of the model to the data by allowing the model to account for countercyclical fiscal policy that is a strong feature of the data. The estimates of the intertemporal elasticity of substitution are significantly larger, and impulse responses and variance decompositions from DSGE full-information methods become consistent with the estimates from reduced-form VARs and limited-information methods. Treating government spending as exogenous in spite of its clear counter-cyclicality in the data is the main source of misspecification.

[20] The 90% credible sets for the variance decompositions output are (17, 79) and (21, 83) and for hours (3, 12) and (89, 97), for A_t and G_t respectively.

[21] The 90% credible sets for the variance decompositions output are (58, 83) and (17, 42) and for hours (44, 75) and (25, 56), for A_t and G_t respectively.

5.2 The impact of correlated disturbances in the NK model

The posterior distributions for the NK model are in table 5.

Misspecification. Moving from independent to correlated disturbances raises the log marginal predictive density of the model by 31 log points, a fairly significant amount. Interestingly, the individual marginal posterior 90% credible sets for each of the cross-correlation of the disturbances includes zero. The data strongly disfavors assuming that all disturbances are independent, as is commonly done, but does not point to one particular correlation alone. This provides a word of caution to the sometimes-used procedure of allowing for a few selected cross-correlations between disturbances in the model.

Posteriors for economic parameters. There are two noticeable differences between the posterior estimates with independent and correlated disturbances. First, the parameter measuring the intrinsic persistence of inflation, ζ, here driven by the indexation of unadjusted prices to past inflation, falls from a median of 0.33 to 0.20. This parameter is a notorious weak spot of NK models: the micro evidence strongly rejects it, its microfoundations are weak, and it surely changes with policy. Including correlated disturbances correctly points to the reduced-form nature of this parameter, and just as in the RBC model, including correlated disturbances brings the model closer to other sources of evidence by shrinking it. The data prefers that persistence comes from exogenous disturbances then from a misspecified microfoundation, and therefore suggests changing the model of price rigidity.

Second, the relative weight of each shock changes significantly, once these shocks are allowed to co-vary. The variance of preference and markup shocks increase from $(1.16, 0.20)$ to $(1.65, 0.26)$. In the other direction, the variance of the shocks to productivity and to the inflation target fall from from $(0.03, 0.49)$ to $(0.01, 0.18)$. This indicates that decompositions of the sources of business-cycle variability across shocks in this model are fragile. This is particularly worrisome for optimal policy exercises in this model, which depend strongly on whether shocks arise from changes in natural rates as opposed to markups.

Inferences on the power of monetary policy. The main use of the NK model is to understand the co-movement in the economy after a monetary policy shock. Figure 3 shows the impulse responses of inflation, interest rates, output growth and the output gap following a 25bp shocks to monetary policy. The monetary structural shocks are identified through the standard Choleski assumption that shocks to it only affect the statistical disturbances to interest rates on impact; other disturbances only move with at least a one-quarter lag.

Comparing the two panels, the median point of the impulse responses is not very different. But, the credible sets are much wider. Allowing for correlated disturbances leads to an acknowledgment of higher uncertainty in the statistical estimates, much as what usually happens in linear regressions when adjusting standard errors for heteroskedascitity and autocorrelation. A researcher that assumes independent disturbances would infer with confidence that tighter monetary policy lowers inflation for at least 6 quarters, and that it has non-zero impact on output growth and gap for 2 quarters. A researcher that allows for correlated disturbances is much less confident on these predictions. In fact, if she is skeptical of the effectiveness of monetary policy, she may well conclude that monetary policy has little to no effect on output and an effect on inflation that is gone within one year. Consistent with the estimate of price stickiness, this also points to looking for alternative models of nominal rigidities.

Conclusion on the NK model. Independent, as opposed to correlated, disturbances are again strongly rejected in the data. The main effect of allowing for correlated disturbances is to cast doubt on the apparent certainty in the standard estimates of impulse response and variance decompositions. These inferences appear fragile, as the estimated uncertainty is higher than previously appreciated. Moreover, correlated disturbances shine a light on the model of nominal rigidities, and especially price indexation, as the weak spot of the model, supporting debates in the theoretical literature.

5.3 The impact of correlated disturbances in the Smets-Wouters model

Finally, we turn to the Smets-Wouters model with dynamically but not contemporaneously correlated disturbances.[22]

Misspecification. The two models with independent or correlated disturbances do roughly well at explaining the data: the log marginal predictive densities of the models with independent and dynamically-correlated disturbances are within 5 log points of each other. The criteria heavily penalizes (as it should) the large increase in the nuisance parameters of the model.

[22] All the estimates are based on 3 million draws, preceded by another 6 million draws used to burn in and to calibrate the proposal densities.

Table 6 displays the estimated dynamic correlations across disturbances. Quite a few are significantly different from zero. Starting with the correlation between total factor productivity and government spending, it is large and goes in both directions (Φ_{AG} and Φ_{GA}). While the DSGE model is more involved than the simple RBC model, it is still missing an important role for fiscal policy rules that respond to the business cycle.[23] This result provides a justification for why Smets and Wouters (2007) assumed a correlation between these two variables. However, it is not enough, as there are several other significant cross-correlations. All but one of them involve either the risk-premium disturbance or investment-specific productivity. These estimates suggest that the absence of financial frictions are the main weakness of this model that researchers should focus on.

Posteriors for economic parameters. The Smets-Wouters model is the canonical model to study the impact on the economy of different monetary policy rules. Figure 4 plots the bivariate histogram of the posterior distribution for two of the parameters in the estimated policy function for nominal interest rates. One of them is the implied steady-state target inflation rate, while the other is the coefficient on inflation in the nominal interest rate rule. Both of them capture how averse to inflation the Federal Reserve has been, one in terms of average inflation, and the other in terms of how shocks to inflation trigger policy responses.

With correlated disturbances, there is little change in the posterior for trend inflation. This is perhaps not too surprising insofar as this parameter is being pinned down by average inflation in the sample. However, with correlated rather than independent disturbances, the policy response to inflation is much lower. Unlike the original Smets-Wouters estimates, the correlated estimates portray monetary policy has having been not so aggressive in response to inflation between 1966 and 2004.

Inferences on what drives the business cycle. Table 7 shows the median variance decompositions for output, hours, real wages and inflation in the short run (1 quarter ahead), the long run (unconditionally), and at business cycle frequencies (2 years and 8 years ahead).

With independent disturbances, the fluctuations in output and hours are accounted mostly by government spending, risk premium and investment-specific productivity at

[23] Alternatively, both models assume a closed economy, so G may be capturing net exports. Since both theory and casual empirics suggest that the trade balance is sensitive to the business cycle, this reinforces our point that it is inappropriate to assume an exogenous G.

the shorter horizon. At longer horizons, as Smets and Wouters (2007) emphasized, it is the wage-markup disturbance that dominates.

With correlated disturbances, the conclusions are similar at the short-run frequencies but very different at the business-cycle and long-run frequencies. Focusing on the variance of output, wage markups go from accounting for 47% and 49% at the 8-year and infinite horizon with independent disturbances, to only 25% and 9% with correlated disturbances. The two productivity disturbances and government spending now explain 80% of the variance of output in the long run, and as much as 35% at the 2-year horizon. Looking instead at the variance of inflation, again the role of wage markup declines significantly when we allow for correlated disturbances, and the difference from the independent-disturbances case increases with the horizon. Across all series, there is an increase in the role of productivity and government spending in accounting for the business cycle. Therefore, we find that the much-debated finding that markup disturbances are important is not robust.[24]

Conclusion on the DSGE model. The DSGE model has the virtue of fitting the data quite well, and being routinely estimated within central banks. Allowing for correlated disturbances does not significantly improve the fit of the model. Monetary policy is estimated to have been less aggressive in response to inflation. There are two economic take-aways for modelers. First, the pattern of cross correlations points to the absence of financial frictions as the main flaw of the model. Second, it suggests that trying to endogenize fiscal policy is more important than trying to explain markups, as much of the literature has done.

6 Conclusion

DSGE modeling has made great strides in the last two decades, in particular in the area of estimation and statistical inference. At the same time, there are still some holes in our knowledge that must be filled. This paper identified one of these holes: the strong and incredible restrictions that models typically place on the exogenous disturbances. Using well-known points in simultaneous-equation econometrics, we argued that these restrictions could severely hamper the model's ability to fit the data and severely bias

[24]See Chari, Kehoe, McGrattan (2009) for some of the debate, and Justiniano and Primiceri (2013) for an alternative estimation approach that converges with our results that wage markups are not as important as previously thought.

inferences on key parameters and model predictions. We proposed the alternative of allowing for correlated disturbances, in the tradition of Zellner (1962).

The main obstacle to allowing for correlated disturbances is that it introduces a large number of nuisance parameters. We proposed a new method for estimating DSGE models, based on using conjugate families for some conditional posterior distributions. The algorithm is also valid and useful with uncorrelated disturbances, and with correlated disturbances it makes previously infeasible estimation now possible.

We applied the method to three models, and found that in all of them there was evidence of misspecification due to assuming independent AR(1) disturbances, that some parameter estimates were robust to correlated disturbances, but some key ones were not, and that the answers that the models give to important economic questions change significantly when allowing for correlated disturbances, pointing to fragility in the model's endogenous propagation mechanisms.

References

Gelman, Andrew, John B. Carlin, Hal S. Stern and Donald B. Rubin (1998). *Bayesian Data Analysis*. Chapman & Hall / CRC Press.

Backus David K., Patrick J. Kehoe and Finn E. Kydland (1992). "International Real Business Cycles." *Journal of Political Economy*, 100 (4), 745-775.

Baxter, Marianne and Robert G. King (1993). "Fiscal Policy in General Equilibrium." *American Economic Review*, 83 (3), 315-334.

Blanchard, Olivier J. and Charles M. Kahn (1980). "The Solution of Linear Difference Models under Rational Expectations." *Econometrica*, 48 (5), 1305-1311.

Caldara, Dario, Richard Harrison and Anna Lipinska (2014). "Practical Tools for Policy Analysis in DSGE Models with Missing Shocks." *Journal of Applied Econometrics*, 29(7), 1145-1163.

Carter, C. K. and R. Kohn (1994). "On Gibbs Sampling for State Space Models." *Biometrika*, 81, 541-553.

Chari, V. V., Patrick J. Kehoe and Ellen R. McGrattan (2007). "Business Cycle Accounting." *Econometrica*, 75 (3), 781-836.

Chari, V. V., Patrick J. Kehoe and Ellen R. McGrattan (2009). "New Keynesian Models: Not Yet Useful for Monetary Policy." *American Economic Journal: Macroeconomics*, 1 (1), 242-266.

Chib, Siddartha (2001). "Markov Chain Monte Carlo Methods: Computation and Inference." In *Handbook of Econometrics*, edited by J. J. Heckman and E. E. Leamer, volume 5, chapter 57, pp. 3569-3649.

Chib, Siddartha and Edward Greenberg (1994). "Bayes Inference in Regression Models with ARMA(p,q) Errors." *Journal of Econometrics*, 64, 183-206.

Chib, Siddartha and Edward Greenberg (1995). "Hierarchical Analysis of SUR Models with Extensions to Correlated Serial Errors and Time-Varying Parameter Models. " *Journal of Econometrics*, 68, 339-360.

Chib, Siddartha and Srikanth Ramamurthy (2010). "Tailored Randomized-block MCMC Methods with Application to DSGE Models." *Journal of Econometrics*, 155, 19-38.

Christiano, Lawrence J. and Martin Eichenbaum (1992). "Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations." *American Economic Review*, 82 (3), 430-450.

Cochrane, D. and G. Orcutt (1949). "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms." *Journal of the American Statistical*

Association, 44, 32-61.

Cogley, Timothy and James M. Nason (1995). "Output Dynamics in Real-Business-Cycle Models." *American Economic Review*, 85 (3), 492-511.

Cúrdia, Vasco, Andrea Ferrero, Ging Cee Ng, and Andrea Tambalotti (2014). "Has U.S. Monetary Policy Tracked the Efficient Interest Rate?" *Journal of Monetary Economics*, 70, 72-83.

Dedola, Luca and Stefano Neri (2007). "What does a technology shock do? A VAR analysis with model-based sign restrictions." *Journal of Monetary Economics*, 54 (2), 512-549.

Del Negro, Marco and Frank Schorfheide (2009). "Monetary Policy Analysis with Potentially Misspecified Models. " *American Economic Review*, 99 (4), 1415-1450.

Den Haan, Wouter and Thomas Drechsel (2020). "Agnostic Structural Disturbances (ASDs): Detecting and reducing misspecification in empirical macroeconomic models" *Journal of Monetary Economics*, forthcoming.

Durbin, James and Siem J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press: Oxford, UK.

Evans, Charles L. (1992). "Productivity Shocks and Real Business Cycles." *Journal of Monetary Economics*, 29 (2), 191-208.

Fair, Ray S. (2004). *Estimating How the Macroeconomy Works*. Harvard University Press.

Fisher, Jonas M. (2006). "The Dynamic Effects of Neutral and Investment-Specific Technology Shocks" *Journal of Political Economy*, 114 (3), 413-451.

Gali, Jordi (1999). "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 89 (1), 249-271.

Geweke, John (1989). "Bayesian Inference in Econometric Models Using Monte Carlo Integration." *Econometrica*, 57 (6), 1317-1339

Geweke, John (1992). "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments." In J. M. Bernardo et al, eds., *Bayesian Statistics vol. 4*, Clarendon Press: Oxford UK.

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

Griliches, Zvi (1967). "Distributed Lags: A Survey." *Econometrica*, 35, 16-49.

Hall, Robert E. (1988). "Intertemporal Substitution in Consumption." *Journal of Political Economy*, 96 (2), 339-57.

Herbst, Edward and Frank Schorfheide (2014). "Sequential Monet Carlo Sampling for

DSGE Models." *Journal of Applied Econometrics*, 29 (7), 1073-1098.

Herbst, Edward and Frank Schorfheide (2015). *Bayesian Estimation of DSGE Models*. Princeton University Press, Princeton NJ.

Inoue, Atsushi, Chun-Hung Kuo, and Barbara Rossi (2020). "Identifying the Sources of Model Misspecification." *Journal of Monetary Economics*, 110, 1-18.

Ireland, Peter N. (2004). "A method for taking models to the data." *Journal of Economic Dynamics and Control*, 28 (6), 1205-26.

Justiniano, Alejandro and Bruce Preston (2010). "Can Structural Small Open Economy Models Account for the Influence of Foreign Shocks?" *Journal of International Economics*, 81 (1), 61-74.

Justiniano, Alejandro and Giorgio E. Primiceri (2013). "Is There a Trade-off between Inflation and Output Stablization." *American Economic Journal: Macroeconomics*, 5 (2), 1-31.

Kadiyala, K. Rao and Sune Karlsson (1997). "Numerical Methods for Estimation and Inference in Bayesian VAR-Models." *Journal of Applied Econometrics*, 12 (2), 99-132.

Komunjer, Ivana and Serena Ng (2011). "Dynamic Identification of DSGE Models." *Econometrica*, 79 (6), 1995-2032.

Meyer-Gohde, Alexander and Daniel Neuhoff (2015). "Generalized Exogenous Processes in DSGE: A Bayesian Approach." Humboldt University manuscript.

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica*, 55 (3), 703-708.

Prescott, Edward C. (1986). "Theory Ahead of Business Cycle Measurement." *FRB Minneapolis Quarterly Reiview,* 9-22.

Rabanal, Pau, Juan F. Rubio-Ramirez and Vicente Tuesta (2011). "Cointegrated TFP Shocks and International Business Cycles." *Journal of Monetary Economics*, 58, 156-171.

Reicher, Claire (2015). "A Note on the Identification of Dynamic Economic Models with Generalized Shock Processes." *Oxford Bulletin of Economics and Statistics*, 78(3), 412-423.

Sargent, Thomas J. (1978). "Estimation of Dynamic Labor Demand Schedules under Rational Expectations." *Journal of Political Economy*, 86(6), 1009-1044.

Schmitt-Grohe, Stephanie and Martin Uribe (2010). "Business Cycles with a Common Trend in Neutral and Investment-Specific Productivity." *Review of Economic Dynamics*, forthcoming.

Shapiro, Matthew D. and Mark W. Watson (1986). "Sources of Business Cycle Fluctuations." *NBER Macroeconomics Annual*, 3, 111-148.

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

Sims, Christopher A. and Tao Zha. (1998). "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review*, 39 (4), 949-968.

Smets, Frank and Rafael Wouters (2003). "An Estimated Stochastic Dynamic General Equilibrium Model of the Euro Area." *Journal of the European Economic Association*, 1 (5), 1123-1175.

Smets, Frank and Rafael Wouters (2007). "Shocks and Frictions in U.S. Business Cycles: A Bayesian DSGE Approach." *American Economic Review*, 97 (3), 586-606.

Stock, James H. (2010). "The Other Transformation in Econometric Practice: Robust Tools for Inference." *Journal of Economic Perspectives*, 24 (2), 83-94.

Tierney, L. (1994). "Markov Chains for Exploring Posterior Distributions." *Annals of Statistics*, 22, 1701-62.

Yogo, Motohiro (2004). "Estimating the Elasticity of Intertemporal Substitution When Instruments Are Weak." *Review of Economics and Statistics*, 86 (3), 797-810.

Zellner, Arnold (1962). "An Efficient Method of Estimating Seemingly Unrelated Regressions and Test for Aggregation Bias" *Journal of the American Statistical Association*, 57, 348-368.

Appendix – For Online Publication

This appendix contains extra results to accompany our manuscript "Correlated Disturbances and U.S. Business Cycles." Sections A to C we map the three models in section 3 to the general structure in section 2. Section D discusses other convergence diagnostics to complement the discussion in section 4. Section E present the parameter priors and posteriors for the Smets-Wouters model used in section 5.

A The Real Business Cycle Model

The first model is described in detail, including the exact matrices in the state-space representation. For the other models, we only show the key pieces in terms of our notation, without much discussion.

The endogenous variables in this model are:

$$y_t \equiv \left(\hat{Y}_t, \hat{K}_t, \hat{N}_t\right),$$

corresponding to output, capital and labor, in log-deviations from steady state. The exogenous disturbances are:

$$s_t \equiv \left(\hat{A}_t, \hat{G}_t\right),$$

corresponding to productivity and government expenditure shocks.

The log-linear representation of the model is:

$$\begin{aligned}
0 &= \mathbb{E}_t\left[\frac{\alpha\beta Y}{K}\left(\hat{Y}_{t+1}-\hat{K}_t\right)+\gamma^{-1}\left(\hat{Y}_t-\hat{Y}_{t+1}\right)+\left(\theta(1-\gamma^{-1})\frac{N}{1-N}-\frac{\gamma^{-1}}{1-N}\right)\left(\hat{N}_t-\hat{N}_{t+1}\right)\right],\\
0 &= \left(C-Y\right)\hat{Y}_t+G\hat{G}_t-\frac{C}{1-N}\hat{N}_t+K\hat{K}_t-(1-\delta)K\hat{K}_{t-1},\\
0 &= \hat{Y}_t-(1-\alpha)(\hat{A}_t+\hat{N}_t)-\alpha\hat{K}_{t-1}.
\end{aligned}$$

In capitals, (Y, K, N) correspond to the steady state values. The parameters (α, β, δ) are calibrated, while parameters (γ, θ) are estimated together with the disturbance parameters characterizing the law of motion for the disturbances.

The log-linear equations satisfy the canonical form proposed in Sims (2002) and presented as equation (4) in the manuscript:

$$\Psi_0(\varepsilon) y_t = \Psi_1(\varepsilon) y_{t-1} + \Psi_2(\varepsilon) s_t + \Psi_3(\varepsilon) w_t, \tag{A1}$$

where the matrices are:

$$\Psi_0(\varepsilon) \equiv \begin{bmatrix} \gamma^{-1} - \frac{\alpha\beta Y}{K} & \theta(1-\gamma^{-1})\frac{N}{1-N} - \frac{\gamma^{-1}}{1-N} & 0 \\ C - Y & -\frac{C}{1-N} & K \\ 1 & -(1-\alpha) & 0 \end{bmatrix},$$

$$\Psi_1(\varepsilon) \equiv \begin{bmatrix} \gamma^{-1} & \theta(1-\gamma^{-1})\frac{N}{1-N} - \frac{\gamma^{-1}}{1-N} & -\frac{\alpha\beta Y}{K} \\ 0 & 0 & (1-\delta)K \\ 0 & 0 & \alpha \end{bmatrix},$$

$$\Psi_2(\varepsilon) \equiv \begin{bmatrix} 0 & 0 \\ 0 & -G \\ 1-\alpha & 0 \end{bmatrix}, \quad \Psi_3(\varepsilon) \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

and w_t is a single endogenous expectations shock, defined as

$$w_t \equiv \left(\frac{\alpha\beta Y}{K} - \gamma^{-1}\right)\left(\hat{Y}_t - \mathbb{E}_{t-1}\hat{Y}_t\right) - \left(\theta(1-\gamma^{-1})\frac{N}{1-N} - \frac{\gamma^{-1}}{1-N}\right)\left(\hat{N}_t - \mathbb{E}_{t-1}\hat{N}_t\right).$$

As discussed in Sims (2002), the solution to this forward-backward looking problem is given by

$$y_t = \Lambda_1(\varepsilon,\sigma) y_{t-1} + \Lambda_2(\varepsilon,\sigma) s_t \tag{A2}$$

where $\Lambda_1(\varepsilon,\sigma)$ are nonlinear functions of (ε,σ).[25] This fits equation (1) in the manuscript.

Turning to equation (2) in the manuscript, the shock structure takes the form of:

$$\begin{aligned} s_t &= \Phi(L) s_{t-1} + e_t, \\ Var(e_t) &= \Omega, \end{aligned}$$

just as mentioned in the paper, with $e_t \equiv \left(e_t^A, e_t^G\right)$ the vector of innovations. The vector of statistical parameters to be estimated depends on the specific case considered. For the

[25]Notice that this representation is true for solutions to the rational expectations equilibrium (REE) that exist and are unique. In the implementation it is usual for authors to discard parameter values that do not satisfy existence and uniqueness, which is equivalent to considering a joint prior that is truncated so as to give zero probability to such outcomes.

AR(1) case, we have:

$$\Phi_{AR1} \equiv \begin{bmatrix} \Phi_A & 0 \\ 0 & \Phi_G \end{bmatrix}, \qquad \Omega_{AR1} \equiv \begin{bmatrix} \Omega_A & 0 \\ 0 & \Omega_G \end{bmatrix},$$

and the statistical parameters vector is

$$\sigma_{AR1} \equiv (\Phi_A, \Phi_G, \Omega_A, \Omega_G).$$

For the VAR(1) case then we get

$$\Phi_{VAR1} \equiv \begin{bmatrix} \Phi_{AA} & \Phi_{AG} \\ \Phi_{GA} & \Phi_{GG} \end{bmatrix}, \qquad \Omega_{VAR1} \equiv \begin{bmatrix} \Omega_{AA} & \Omega_{AG} \\ \Omega_{GA} & \Omega_G \end{bmatrix},$$

and the statistical parameters vector is

$$\sigma_{VAR1} \equiv (\Phi_{AA}, \Phi_{AG}, \Phi_{GG}, \Phi_{GA}, \Omega_{AA}, \Omega_{AG}, \Omega_{GG}),$$

and notice that because Ω is a covariance matrix then it must be the case that $\Omega_{GA} = \Omega_{AG}$.

Finally, the observation equations are:

$$\begin{aligned} Y_t^{obs} &= \hat{Y}_t, \\ N_t^{obs} &= \hat{N}_t, \end{aligned}$$

where $x_t \equiv \left(Y_t^{obs}, N_t^{obs}\right)$ is the vector of observables. Relative to equation (3) in the manuscript, $H_1 = H_3(L) = 0$ and H_2 is a matrix with ones and zeros mapping the endogenous variables y_t to the observable variables x_t,

$$H_2 \equiv \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Turning to the assumption in section 2.1, assumption a) holds by inspection of the state-space system and assumption b) holds as long as the disturbances to the exogenous processes are assumed to be drawn from a normal distribution.

B The New Keynesian Model

All variables and parameters are explained in Cúrdia et al (2010). Here we just show the representation in terms of our assumptions, without any further interpretation.

The log-linear equations are:

$$\tilde{x}_t = \mathbb{E}_t\,\tilde{x}_{t+1} - \varphi_\gamma^{-1}\left(i_t - E_t\pi_{t+1} - r_t^e\right), \tag{A3}$$

$$\tilde{x}_t = \left[x_t^e - \eta_\gamma\left(\tilde{y}_t^L - \tilde{y}_t^{e,L}\right)\right] - \beta\eta_\gamma\,\mathbb{E}_t\left(x_{t+1}^* - \eta_\gamma x_t^*\right), \tag{A4}$$

$$x_t^e = \tilde{y}_t - \tilde{y}_t^e \tag{A5}$$

$$\tilde{y}_t^e = -\frac{\varphi_\gamma}{\omega}\left\{\tilde{y}_t^e - \eta_\gamma\left(\tilde{y}_t^{e,L} - \gamma_t\right) - \beta\eta_\gamma\,\mathbb{E}_t\left[\tilde{y}_{t+1}^e + \gamma_{t+1}^y - \eta_\gamma\tilde{y}_t^e\right]\right\} + \frac{\beta\eta_\gamma\omega^{-1}}{1-\beta\eta_\gamma}E_t\delta_{t+1}^y, \tag{A6}$$

$$r_t^e = \mathbb{E}_t\,\gamma_{t+1}^y + \mathbb{E}_t\,\delta_{t+1}^y - \omega\left(\mathbb{E}_t\,\tilde{y}_{t+1}^e - \tilde{y}_t^e\right), \tag{A7}$$

$$\tilde{\pi}_t = \xi\left(\omega x_t^e + \varphi\tilde{x}_t\right) + \beta\,\mathbb{E}_t\,\tilde{\pi}_{t+1} + u_t, \tag{A8}$$

$$\tilde{\pi}_t = \pi_t - \zeta\pi_{t-1}, \tag{A9}$$

$$i_t = \rho i_{t-1} + (1-\rho)\left(\phi_\pi\pi_t + \phi_x x_t^e\right) + r_t, \tag{A10}$$

$$\tilde{y}_t^{e,L} = \tilde{y}_{t-1}^e, \tag{A11}$$

$$\tilde{y}_t^L = \tilde{y}_{t-1}, \tag{A12}$$

$$\Delta y_t = \tilde{y}_t - \tilde{y}_{t-1} + \gamma_t, \tag{A13}$$

$$\gamma_t^y = \gamma_t, \tag{A14}$$

$$\delta_t^y = \delta_t. \tag{A15}$$

The observation equations are:

$$\begin{aligned}\Delta Y_t^a &= \gamma^a + 400\Delta y_t\\ \pi_t^a &= \pi^* + 400\pi_t\\ i_t^a &= (r^a + \pi^*) + 400 i_t,\end{aligned}$$

where ΔY_t^a is the growth rate of real GDP, π_t^a is the inflation rate, i_t^a is the Federal Funds target rate.

The observable variables are:

$$x_t \equiv \left(\Delta Y_t^a, \pi_t^a, i_t^a\right).$$

The endogenous variables are:

$$y_t \equiv \left(\tilde{y}_t, \tilde{\pi}_t, i_t, \tilde{x}_t, x_t^e, x_t^*, \tilde{y}_t^e, r_t^e, \Delta y_t, \tilde{y}_t^L, \tilde{y}_t^{e,L}, \gamma_t^y, \delta_t^y\right).$$

Finally, the exogenous disturbances are:

$$s_t \equiv (\gamma_t, \delta_t, u_t, r_t),$$

with innovations:

$$e_t \equiv \left(e_t^\gamma, e_t^\delta, e_t^u, e_t^r\right).$$

Turning to the parameters, the economic parameters:

$$\varepsilon \equiv (\omega, \xi, \eta, \zeta, \rho, \phi_\pi, \phi_x, \pi^*, r^a, \gamma^a).$$

Other parameters showing up in equations are either calibrated or are a combination of estimated parameters. The statistical parameters depend on the assumption:

$$\begin{aligned}
\sigma_{AR1} &\equiv (\Phi_\gamma, \Phi_\delta, \Phi_u, \Phi_r, \Omega_\gamma, \Omega_\delta, \Omega_u, \Omega_r), \text{ or} \\
\sigma_{VAR1} &\equiv (\Phi_{\gamma\gamma}, \Phi_{\gamma\delta}, \Phi_{\gamma u}, \Phi_{\gamma r}, \Phi_{\delta\gamma}, \Phi_{\delta\delta}, \Phi_{\delta u}, \Phi_{\delta r}, \Phi_{u\gamma}, \Phi_{u\delta}, \Phi_{uu}, \Phi_{ur}, \Phi_{r\gamma}, \Phi_{r\delta}, \Phi_{ru}, \Phi_{rr}, \\
&\quad \Omega_{\gamma\gamma}, \Omega_{\gamma\delta}, \Omega_{\gamma u}, \Omega_{\gamma r}, \Omega_{\delta\delta}, \Omega_{\delta u}, \Omega_{\delta r}, \Omega_{uu}, \Omega_{ur}, \Omega_{rr}).
\end{aligned}$$

Equations (A3)-(A15) satisfy the canonical form in equation (A1), hence the solution to this model can be represented as in equation (A2), consistent with our paper's framework. Notice that, as before, in this model we defined several auxiliary variables that allow us to map the model exactly to the framework needed in our paper. The number of observables, n_x, is smaller than the number of disturbances, n_s, hence again it is possible to apply the Kalman filter and generate a well behaved likelihood function.

C The Smets-Wouters DSGE Model

We follow Smets and Wouters (2007) closely, including keeping their notation in this appendix as much as we can. The only change is for the statistical parameters to fit our general setup in section 2.

The notation refers to: y_t is output, c_t is consumption, i_t is investment, q_t is the value of capital, l_t is hours worked, z_t is capital utilization, r_t is the nominal interest rate, π_t is inflation, w_t is the real wage, k_t is capital installed, μ_t^p is the price mark-up, and μ_t^w is the wage mark-up. The disturbances are all denoted by s_t with the superscript denoting the type of shock.

The endogenous variables are:

$$\begin{aligned} y_t \equiv \; & (\tilde{y}_t, c_t, i_t, z_t, l_t, r_t, q_t, k_t, w_t, \pi_t, \mu_t^p, \mu_t^w, \tilde{y}_t^p, c_t^p, i_t^p, z_t^p, l_t^p, r_t^{rp}, i_t^p, q_t^p, z_t^p, k_t^p, w_t^p, \\ & c_t^L, i_t^L, \pi_t^L, w_t^L, c_t^{pL}, i_t^{pL}, \Delta y_t, \Delta y_t^p, \Delta i_t, \Delta c_t, \Delta w_t \Big) . \end{aligned}$$

The observable variables are:

$$x_t \equiv \left(\Delta Y_t^a, \Delta c_t^{obs}, \Delta i_t^{obs}, \Delta w_t^{obs}, l_t^{obs}, \pi_t^{obs}, r_t^{obs} \right) .$$

Finally, the exogenous disturbances are:

$$s_t \equiv \left(s_t^g, s_t^b, s_t^i, s_t^a, s_t^p, s_t^w, s_t^r \right)$$

with innovations:

$$e_t \equiv \left(e_t^g, e_t^b, e_t^i, e_t^a, e_t^p, e_t^w, e_t^r \right) .$$

Turning to parameters, the list of economic parameters is:

$$\varepsilon \equiv \left(\bar{\gamma}, \bar{l}, \bar{\pi}, 100 \left(\beta^{-1} - 1 \right), \varphi, \sigma_c, \lambda, \xi_w, \sigma_l, \xi_p, \iota_w, \iota_p, \psi, \Phi_{sw}, r_\pi, \rho_{sw}, r_y, r_{\Delta y}, \alpha \right) .$$

The structural parameters are: $\gamma^* = 100(\gamma - 1)$ is the steady-state growth rate, l^* is the steady-state hours worked, π^* is the steady-state inflation rate, β is the discount factor, ϕ is one plus the share of fixed costs in production, σ_c is the elasticity of intertemporal substitution keeping labor fixed, λ is the degree of habit formation, ξ_w is the degree of wage stickiness, σ_l is the wage elasticity of labor supply, ξ_p is the degree of price stickiness, ι_w is the degree of wage indexation, ι_p is the degree of price indexation, ψ is a positive function of the steady-state elasticity of the capital utilization adjustment cost function that is φ, Φ_{SW} is the gross steady-state labor markup, ρ_{SW}, r_π, r_y and $r_{\Delta y}$ are the monetary policy-rule parameters, and α is the capital share.

The reduced-form parameters are linked to structural parameters according to: $i_y = (\gamma - 0.975)k_y$, $c_1 = (\lambda/\gamma)(1 + \lambda/\gamma)$, $c_2 = [(\sigma_c - 1)(W_*^h L_*/C_*)/[\sigma_c(1 + \lambda/\gamma)]$, and $c_3 = (1 - \lambda/\gamma)/[(1 + \lambda/\gamma)\sigma_c]$, $i_1 = 1/(1 + \beta\gamma^{(1-\sigma_c)})$, $i_2 = i_1/\gamma^2\varphi$, $q_1 = 0.975\beta\gamma^{-\sigma_c}$, $k_1 = 0.975/\gamma$, $k_2 = (1 - k_1)(1 + \beta\gamma^{1-\sigma_c})\gamma^2\varphi$, $\pi_1 = \iota_p/(1 + \beta\gamma^{1-\sigma_c}\iota_p)$, $\pi_2 = \pi_1\beta\gamma^{1-\sigma_c}/\iota_p$, $\pi_3 =$

$(\pi_1/\iota_p)\left\{(1-\beta\gamma^{1-\sigma_c}\xi_p)(1-\xi_p)/\left\{\xi_p\left[10(\phi-1)+1\right]\right\}\right\}$, $w_1 = i_1$, $w_2 = w_1(1+\beta\gamma^{1-\sigma_c}\iota_w)$, $w_3 = \iota_w w_1$, $w_4 = w_1\left\{(1-\beta\gamma^{1-\sigma_c}\xi_w)(1-\xi_w)/\left\{\xi_w\left[10(\Phi_{SW}-1)+1\right]\right\}\right\}$, $\gamma^* = 100(\gamma-1)$, and k_y is the steady-state capital-output ratio and R_*^k is the steady-state rental rate of capital. (Other parameters showing up in equations are either calibrated or some combination of estimated parameters.)

The statistical parameters depend on the two cases we considered:

$$\begin{aligned}
\sigma_{AR1} &\equiv \left(\Phi_g,\Phi_b,\Phi_i,\Phi_a,\Phi_p,\Phi_w,\Phi_r,\Omega_g,\Omega_b,\Omega_i,\Omega_a,\Omega_p,\Omega_w,\Omega_r\right)\text{ , or}\\
\sigma_{VAR1^*} &\equiv \left(\Phi_{gg},\Phi_{gb},\Phi_{gi},\Phi_{ga},\Phi_{gp},\Phi_{gw},\Phi_{gr},...\Phi_{rg},\Phi_{rb},\Phi_{ri},\Phi_{ra},\Phi_{rp},\Phi_{rw},\Phi_{rr},\right.\\
&\qquad \left.\Omega_g,\Omega_b,\Omega_i,\Omega_a,\Omega_p,\Omega_w,\Omega_r\right),
\end{aligned}$$

where σ_{VAR1^*} stands for the dynamic VAR(1) — dynamic correlated shocks, but contemporaneously independent — which is the case discussed in our paper (the full list of parameters is shown in Table A.2 below).

Finally, turning to the model, the log-linear set of equations is (referring to equation (4) in the manuscript):

$$\begin{aligned}
\tilde{y}_t &= \left(0.82-i_y\right)c_t + i_y i_t + R_*^k k_y z_t + s_t^g,\\
c_t &= c_1 c_t^L + (1-c_1)\,\mathbb{E}_t\,c_{t+1} + c_2(l_t - \mathbb{E}_t\,l_{t+1}) - c_3(r_t - \mathbb{E}_t\,\pi_{t+1} + s_t^b),\\
i_t &= i_1 i_t^L + (1-i_1)E_t i_{t+1} + i_2 q_t + s_t^i,\\
q_t &= q_1\,\mathbb{E}_t\,q_{t+1} + (1-q_1)\,\mathbb{E}_t\left(l_{t+1} - k_{t+1} + w_{t+1}\right) - (r_t - \mathbb{E}_t\,\pi_{t+1} + s_t^b),\\
\tilde{y}_t &= \phi\left[\alpha k_{t-1} + \alpha z_t + (1-\alpha)l_t + s_t^a\right],\\
z_t &= \left[(1-\psi)/\psi\right](l_t - k_t + w_t),\\
k_t &= k_1 k_{t-1} + (1-k_1)i_t + k_2 s_t^i,\\
\pi_t &= \pi_1\pi_t^L + \pi_2\,\mathbb{E}_t\,\pi_{t+1} - \pi_3\mu_t^p + s_t^p,\\
w_t &= w_1 w_t^L + (1-w_1)\left(\mathbb{E}_t\,w_{t+1} + \mathbb{E}_t\,\pi_{t+1}\right) - w_2\pi_t + w_3\pi_t^L - w_4\mu_t^w + s_t^w,\\
\mu_t^p &= \alpha(k_{t-1} + z_t - l_t) - w_t + s_t^a,\\
\mu_t^w &= w_t - \left[\sigma_l l_t + (c_t - c_{t-1}\lambda/\gamma)/(1-\lambda/\gamma)\right],\\
r_t &= \rho r_{t-1} + (1-\rho)[r_\pi\pi_t + r_y(\tilde{y}_t - \tilde{y}_t^p)] + r_{\Delta y}\left(\Delta y_t - \Delta y_t^p\right) + s_t^r,\\
\tilde{y}_t^p &= \left(0.82-i_y\right)c_t^p + i_y i_t^p + R_*^k k_y z_t^p + s_t^g,\\
c_t^p &= c_1 c_t^{pL} + (1-c_1)\,\mathbb{E}_t\,c_{t+1}^p + c_2(l_t^p - \mathbb{E}_t\,l_{t+1}^p) - c_3(r_t^{rp} + s_t^b),
\end{aligned}$$

$$
\begin{aligned}
i_t^p &= i_1 i_t^{pL} + (1-i_1)\,\mathbb{E}_t\, i_{t+1}^p + i_2 q_t^p + s_t^i, \\
q_t^p &= q_1 E_t q_{t+1}^p + (1-q_1)\,\mathbb{E}_t \left(l_{t+1}^p - k_{t+1}^p + w_{t+1}^p\right) - (r_t^{rp} + s_t^b), \\
\tilde{y}_t^p &= \phi\left[\alpha k_{t-1}^p + \alpha z_t^p + (1-\alpha) l_t^p + s_t^a\right], \\
z_t^p &= \left[(1-\psi)/\psi\right]\left(l_t^p - k_t^p + w_t^p\right), \\
k_t^p &= k_1 k_{t-1}^p + (1-k_1) i_t^p + k_2 s_t^i, \\
w_t^p &= \alpha(k_{t-1}^p + z_t^p - l_t^p) + s_t^a, \\
w_t^p &= \left[\sigma_l l_t^p + \left(c_t^p - c_{t-1}^p \lambda/\gamma\right)/\left(1-\lambda/\gamma\right)\right], \\
c_t^L &= c_{t-1}, \\
i_t^L &= i_{t-1}, \\
\pi_t^L &= \pi_{t-1}, \\
w_t^L &= w_{t-1}, \\
c_t^{pL} &= c_{t-1}^p, \\
i_t^{pL} &= i_{t-1}^p, \\
\Delta y_t &= \tilde{y}_t - \tilde{y}_{t-1}, \\
\Delta y_t^p &= \tilde{y}_t^p - \tilde{y}_{t-1}^p, \\
\Delta c_t &= c_t - c_{t-1}, \\
\Delta i_t &= i_t - i_{t-1}, \\
\Delta w_t &= w_t - w_{t-1}.
\end{aligned}
$$

The observation equations, referring to equation (3) in the manuscript are:

$$
\begin{aligned}
\Delta y_t^{obs} &= \bar{\gamma} + \Delta y_t, \\
\Delta c_t^{obs} &= \bar{\gamma} + \Delta c_t, \\
\Delta i_t^{obs} &= \bar{\gamma} + \Delta i_t, \\
\Delta w_t^{obs} &= \bar{\gamma} + \Delta w_t, \\
l_t^{obs} &= \bar{l} + l_t, \\
\pi_t^{obs} &= \bar{\pi} + \pi_t, \\
r_t^{obs} &= \bar{r} + r_t,
\end{aligned}
$$

where $\left(\Delta Y_t^a, \Delta c_t^{obs}, \Delta i_t^{obs}, \Delta w_t^{obs}\right)$ represent the real growth rates of GDP, consumption, investment and wages, l_t^{obs} hours worked, π_t^{obs} inflation rate, and r_t^{obs} the federal funds rate.

D Other convergence diagnostics

We used four metrics to assess convergence and relative efficiency. First, the R statistic, which compares the variance of each parameter estimate between and within chains, to estimate the factor by which these could be reduced by continuing to take draws. This statistic is always larger or equal than one, and a cut-off of 1.001 is often used. We report the maximum of these statistics across all the parameters. Second, the number of effective draws, *neff*, in each chain for each parameter, which corrects for the serial correlation across draws following Geweke (1992). The larger this is, the more efficient the algorithm, and we again report the minimum of these statistics across parameters and chains. This was the statistic used in the main manuscript. Third, the number of effective draws in total, *mneff*, which combines the previous two corrections applied to the mixed simulations from the four chains (Gelman et al, 1998: 298), where again we report the minimum across parameters. Finally, the number of rejections at the 5% level of the z-test that the mean of the parameter draws in two separated parts of the chain is the same. This is the separated partial means test, *SPM*, of Geweke (1992) and fewer rejections implies being closer to convergence.

E More detailed estimates of the Smets-Wouters model

The model and notation were defined in section A.3 of this appendix.

The posterior distributions, with independent AR(1) disturbances and dynamic correlated VAR(1) disturbances, are described in tables A.2 and A.3, respectively. The impulse responses at the median of the posterior are in figure A.4 for both independent and correlated disturbances. Finally, the credible sets for the variance decompositions are in table A.5.

Tables and Figures

Table 1: **Prior distribution for economic parameters across the models**

			Percentile		
Parameter	Density	Mode	5	50	95
Panel A. RBC model					
γ	Gamma	0.6667	0.2347	1.1559	3.3311
θ	Gamma	4.8480	2.7326	5.3290	9.2117
Panel B. NK model					
ω	Gamma	0.9600	0.6953	0.9867	1.3501
ξ	Gamma	0.0750	0.0342	0.0918	0.1938
η	Beta	0.6667	0.2486	0.6143	0.9024
ζ	Beta	0.6667	0.2486	0.6143	0.9024
ρ	Beta	0.7632	0.4266	0.7166	0.9162
ϕ_π	Normal	1.5000	1.0888	1.5000	1.9112
ϕ_x	Normal	0.5000	0.1710	0.5000	0.8290
π^*	Normal	2.0000	0.3551	2.0000	3.6449
r^a	Normal	2.0000	0.3551	2.0000	3.6449
γ^a	Normal	3.0000	2.4243	3.0000	3.5757

Table 2: **Prior distribution for statistical parameters across the models**

Parameter	Density	Mode	5	50	95
Panel A. RBC model with independent AR(1)s					
Φ_A	Normal	0.7525	0.5125	0.7456	0.9361
Φ_G	Normal	0.4255	0.1979	0.4248	0.6487
$\Omega_A(\times 10^4)$	Inv-Gamma^2	1.0348	0.6579	1.5392	4.9752
Ω_G	Inv-Gamma^2	0.2297	0.1452	0.4002	1.7917
Panel B. RBC model with unrestricted VAR(1)					
$\Phi_{A,A}$	Normal	0.7525	0.5062	0.7442	0.9348
$\Phi_{G,G}$	Normal	0.4255	0.1856	0.4218	0.6482
$\Phi_{A,G}$	Normal	0.0000	-0.4183	0.0012	0.4127
$\Phi_{G,A}$	Normal	0.0000	-0.1302	-0.0002	0.1324
$\Omega_{A,A}(\times 10^4)$	Inv-Wishart	0.8184	0.6552	1.4796	4.6492
$\Omega_{G,G}$	Inv-Wishart	0.2583	0.2046	0.4803	1.5041
$\Omega_{A,G}(\times 10^2)$	Inv-Wishart	0.0000	-0.8163	-0.0017	0.8130
Panel C. NK model witth independent AR(1)s					
$\Phi_\delta, \Phi_\gamma, \Phi_u$	Normal	0.5000	0.1961	0.4993	0.7854
Φ_{mp}	Normal	0.2000	0.0491	0.2004	0.3559
Φ_{π^*}	Normal	0.9500	0.9115	0.9491	0.9820
$\Omega_\delta, \Omega_\gamma, \Omega_u, \Omega_{mp}, \Omega_{\pi^*}$	Inv-Gamma^2	0.3846	0.2430	0.6487	2.7170
Panel D. RBC model with unrestricted VAR(1)					
$\Phi_{\delta,\delta}, \Phi_{\gamma,\gamma}, \Phi_{u,u}$	Normal	0.5000	0.3006	0.4981	0.6856
$\Phi_{mp,mp}$	Normal	0.2000	-0.0018	0.1985	0.3870
Φ_{π^*,π^*}	Normal	0.9500	0.7160	0.9138	1.0035
$\Phi_{\delta,\gamma}, \Phi_{\delta,u}, \Phi_{\delta,mp}, \Phi_{\delta,\pi^*}$	Normal	0.0000	-0.1952	0.0003	0.1863
$\Phi_{\gamma,\delta}, \Phi_{\gamma,u}, \Phi_{\gamma,mp}, \Phi_{\gamma,\pi^*}$	Normal	0.0000	-0.1919	0.0003	0.1871
$\Phi_{u,\delta}, \Phi_{u,\gamma}, \Phi_{u,mp}, \Phi_{u,\pi^*}$	Normal	0.0000	-0.1875	0.0002	0.1872
$\Phi_{mp,\delta}, \Phi_{mp,\gamma}, \Phi_{mp,u}, \Phi_{mp,\pi^*}$	Normal	0.0000	-0.1857	0.0006	0.1855
$\Phi_{\pi^*,\delta}, \Phi_{\pi^*,\gamma}, \Phi_{\pi^*,u}, \Phi_{\pi^*,mp}$	Normal	0.0000	-0.1884	-0.0002	0.1919
$\Omega_{\delta,\delta}, \Omega_{\gamma,\gamma}, \Omega_{u,u}, \Omega_{mp,mp}, \Omega_{\pi^*,\pi^*}$	Inv-Wishart	0.0770	0.1249	0.4116	2.448
$\Omega_{\delta,\gamma}, \Omega_{\delta,u}, \Omega_{\delta,mp}, \Omega_{\delta,\pi^*}$	Inv-Wishart	0.0000	-0.8001	0.0003	0.8053
$\Omega_{\gamma,u}, \Omega_{\gamma,mp}, \Omega_{\gamma,\pi^*}$	Inv-Wishart	0.0001	-0.7785	-0.0024	0.7808
$\Omega_{u,mp}, \Omega_{u,\pi^*}$	Inv-Wishart	0.0000	-0.7899	-0.0027	0.7693
Ω_{mp,π^*}	Inv-Wishart	0.0000	-0.7098	0.0036	0.8009

Table 3: **Draws and time for method to reach 300 effective draws**

		Number of draws (thousands)		Time (minutes)	
		Conjugate conditionals	Random-walk Metropolis	Conjugate conditionals	Random-walk Metropolis
Data	Disturbances	Average (min,max)	Average (min,max)	Average (min,max)	Average (min,max)
Panel A. RBC model					
Actual	AR(1)	425	3365	230	538
	VAR(1)	1160	1250	729	258
Simulated	AR(1)	213	2702	115	432
		(45,785)	(70,13696)	(24,424)	(11,2191)
	VAR(1)	477	458	300	95
		(55,3045)	(100,1265)	(35,1913)	(21,261)
Panel B. NK model					
Actual	AR(1)	56	196	54	51
	VAR(1)	1329	50000	1115	18804
Simulated	AR(1)	52	801	49	208
		(44,62)	(158,5124)	(42,59)	(41,1332)
	VAR(1)	3263	36685	2738	13797
		(200,12492)	(17350,40000)	(168,10482)	(6525,15043)

Table 4: **Posterior distribution for the RBC model**

			Percentile		
Parameter	Mean	Mode	5	50	95
Panel A. Independent AR(1)s					
Economic					
γ	1.4029	1.4234	0.4970	1.2435	2.8629
θ	0.6184	0.4896	0.2632	0.5471	1.2036
Statistical					
Φ_A	0.8173	0.8106	0.7422	0.8174	0.8923
Φ_G	0.7505	0.7518	0.6713	0.7520	0.8234
Ω_A	.00014	.00014	.00012	.00014	.00017
Ω_G	0.2706	0.2475	0.1928	0.2645	0.3684
Panel B. Unrestricted VAR(1)					
Economic					
γ	0.4301	0.4304	0.2892	0.4170	0.6060
θ	4.8550	4.3184	1.9072	4.6302	8.5641
Statistical					
Φ_{AA}	0.9385	0.9355	0.9058	0.9402	0.9656
Φ_{AG}	0.0048	0.0048	0.0041	0.0049	0.0054
Φ_{GA}	-8.62	-8.26	-11.21	-8.50	-6.25
Φ_{GG}	0.8805	0.8828	0.8362	0.8811	0.9232
Ω_{AA}	.00013	.00013	.00011	.00013	.00016
Ω_{AG}	0.0084	0.0071	0.0045	0.0080	0.0138
Ω_{GG}	2.0718	1.3527	0.7942	1.6752	4.6432

Figure 1: The prior and posterior distribution for the economic parameters in the RBC model

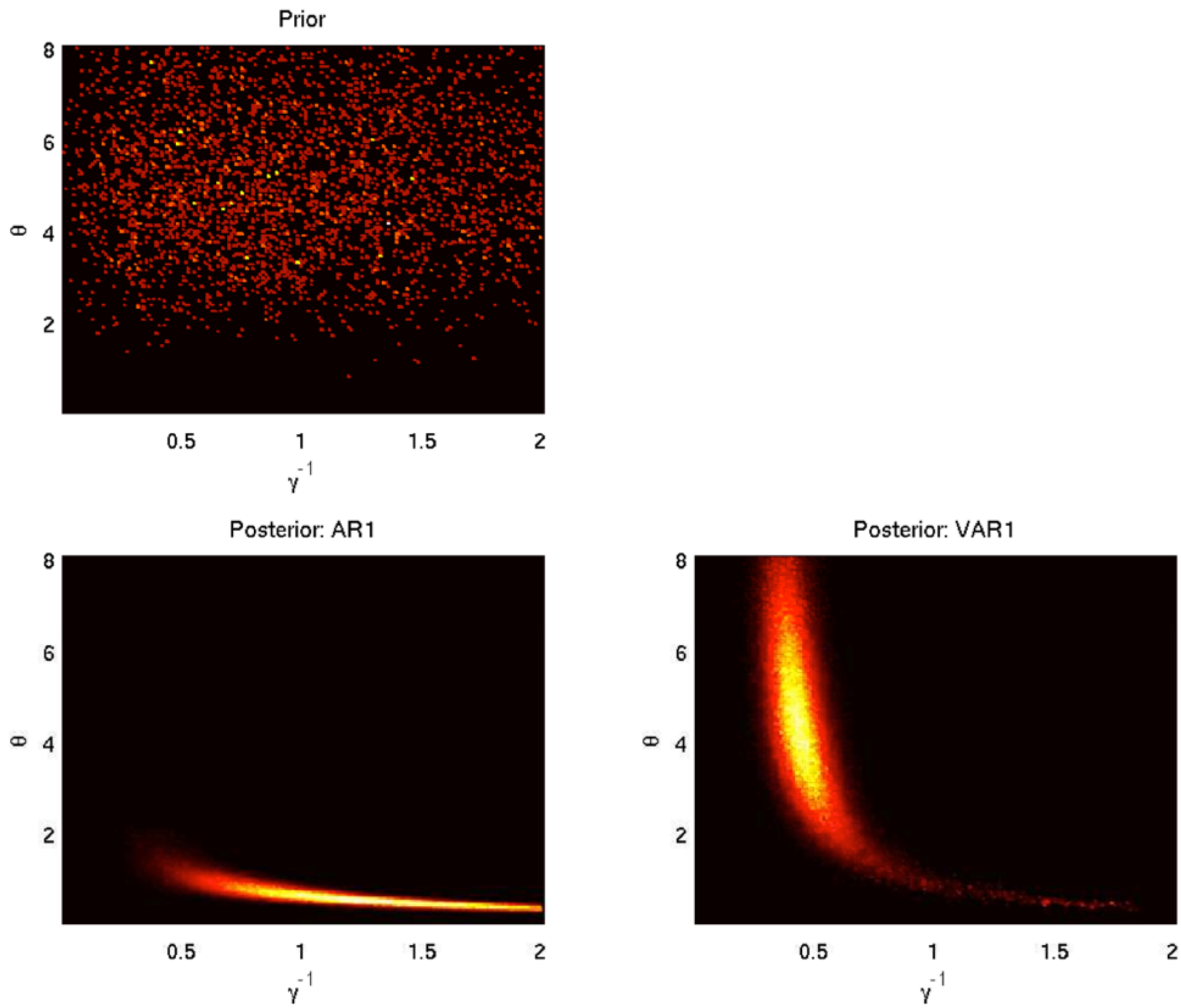
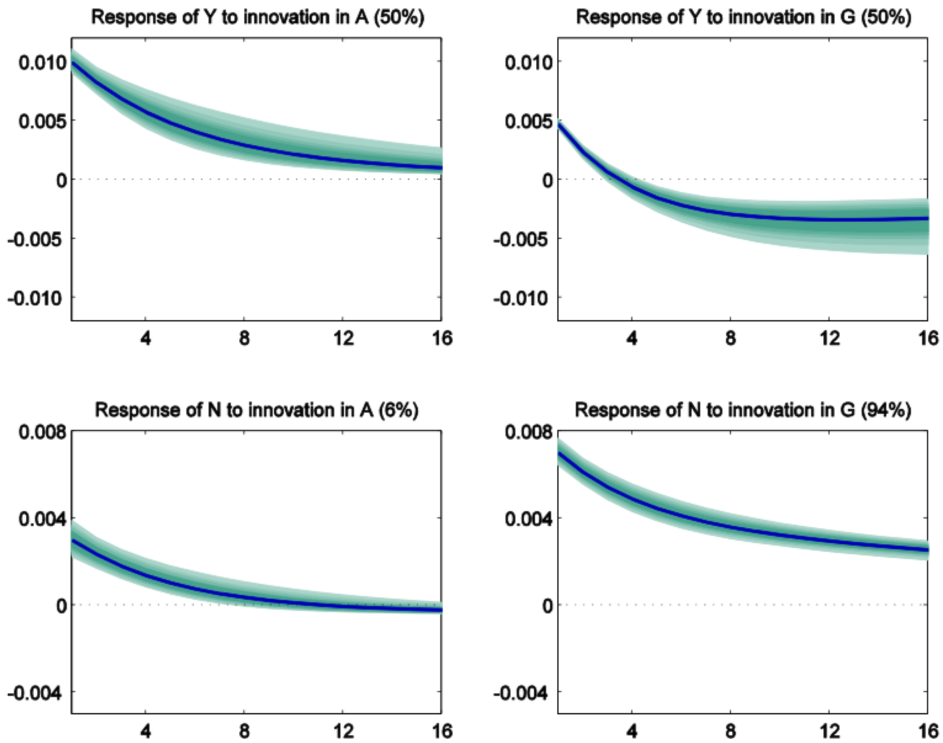


Figure 2: Impulse response functions in RBC model, median and distributions

(a) Independent AR(1)s case



(b) Unrestricted VAR(1) case

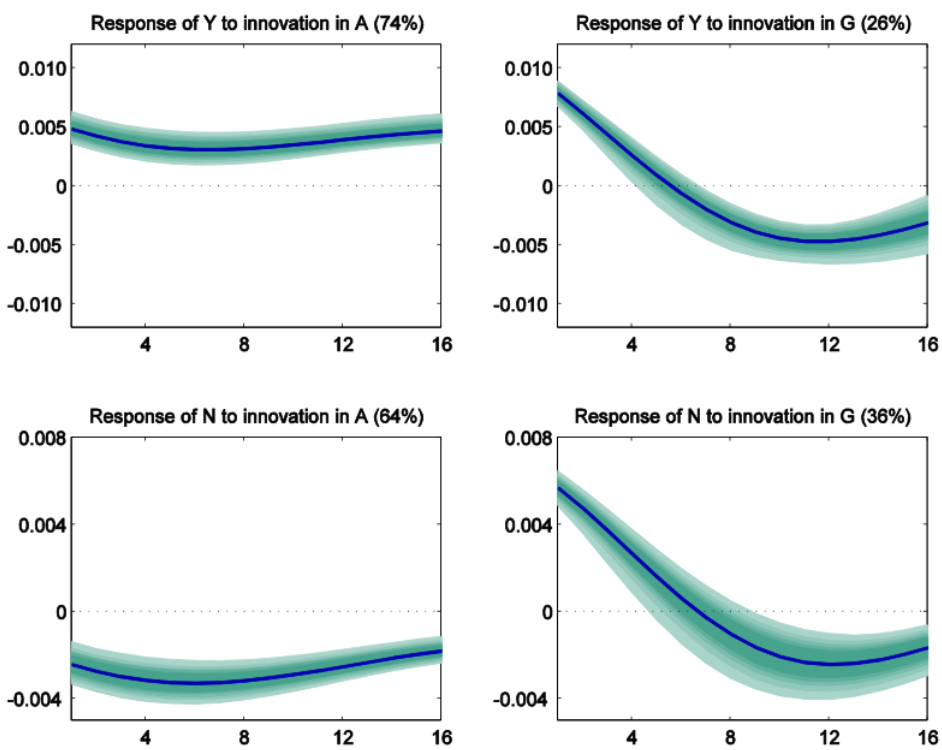


Table 5: **Posterior distribution for the NK model**

			Percentile		
Parameter	Mean	Mode	5	50	95
Panel A. Independent AR(1)s					
Economic					
ω	1.0133	0.96661	0.7057	1.0002	1.3678
ξ	0.1031	0.06033	0.0260	0.0920	0.2181
η	0.6696	0.68057	0.3386	0.6973	0.9029
ζ	0.3472	0.22557	0.1280	0.3294	0.6319
ρ	0.7254	0.73722	0.6442	0.7278	0.7989
ϕ_π	1.7096	1.64669	1.3936	1.7019	2.0553
ϕ_x	0.5237	0.54184	0.1899	0.5253	0.8529
π^*	2.3928	2.40919	1.4444	2.4047	3.2993
r^a	1.9353	1.92588	0.7282	1.9339	3.1373
γ^a	2.9983	2.99986	2.4170	2.9994	3.5743
φ_γ	—	—	2.2842	10.2847	82.5071
η_γ	—	—	0.3400	0.6916	0.8943
Statistical					
Φ_δ	0.9143	0.93016	0.8493	0.9172	0.9692
Φ_γ	0.4991	0.49924	0.4441	0.4990	0.5537
Φ_u	0.4690	0.44401	0.2970	0.4664	0.6420
Φ_{mp}	0.2187	0.21655	0.1601	0.2179	0.2801
Φ_{π^*}	0.9593	0.96115	0.9326	0.9601	0.9831
Ω_δ	1.2721	1.04357	0.6680	1.1630	2.2445
Ω_γ	0.0277	0.02592	0.0216	0.0273	0.0351
Ω_u	0.2198	0.15639	0.1227	0.1997	0.3780
Ω_{mp}	0.1332	0.11664	0.0919	0.1296	0.1864
Ω_{π^*}	0.5487	0.47846	0.2467	0.4910	1.0410
Panel B. Unrestricted VAR(1)					
Economic					
ω	1.0026	0.9328	0.6902	0.9894	1.3612
ξ	0.1013	0.0316	0.0293	0.0931	0.2011
η	0.6601	0.9080	0.3039	0.6856	0.9273

ζ	0.2194	0.1100	0.0672	0.2005	0.4359
ρ	0.7190	0.8516	0.5854	0.7232	0.8385
ϕ_π	1.6003	1.4554	1.2122	1.5972	1.9944
ϕ_x	0.5133	0.5215	0.1758	0.5134	0.8522
π^*	2.2909	2.2139	1.4672	2.3038	3.0687
r^a	2.1154	2.2204	1.0346	2.1290	3.1502
γ^a	2.9991	2.9971	2.4142	2.9989	3.5828
φ_γ	—	—	2.0400	9.6026	142.0639
η_γ	—	—	0.3014	0.6807	0.9206
Statistical					
$\Phi_{\delta,\delta}$	0.8164	0.7865	0.6389	0.8311	0.9410
$\Phi_{\delta,\gamma}$	-0.0041	0.0028	-0.3422	-0.0027	0.3326
$\Phi_{\delta,u}$	-0.0598	0.0545	-0.3934	-0.0606	0.2783
$\Phi_{\delta,mp}$	-0.0225	0.1057	-0.3815	-0.0135	0.3059
Φ_{δ,π^*}	0.0327	-0.029	-0.1248	0.0293	0.2000
$\Phi_{\gamma,\delta}$	0.0005	0.0008	-0.0095	0.0005	0.0104
$\Phi_{\gamma,\gamma}$	0.4999	0.4998	0.4731	0.5000	0.5267
$\Phi_{\gamma,u}$	-0.0004	0.0001	-0.0217	-0.0004	0.0211
$\Phi_{\gamma,mp}$	-0.0006	-0.000	-0.0251	-0.0006	0.0239
Φ_{γ,π^*}	-0.0003	-0.000	-0.0120	-0.0003	0.0114
$\Phi_{u,\delta}$	-0.0039	0.0729	-0.1669	0.0010	0.1459
$\Phi_{u,\gamma}$	0.0033	0.0007	-0.1478	0.0023	0.1575
$\Phi_{u,u}$	0.5223	0.5065	0.3756	0.5100	0.7179
$\Phi_{u,mp}$	0.0437	0.0270	-0.0967	0.0346	0.2182
Φ_{u,π^*}	0.0156	0.0263	-0.1037	0.0209	0.1189
$\Phi_{mp,\delta}$	0.0137	0.1395	-0.0864	0.0126	0.1170
$\Phi_{mp,\gamma}$	0.0016	0.0015	-0.0854	0.0014	0.0884
$\Phi_{mp,u}$	0.0247	0.0288	-0.0569	0.0207	0.1208
$\Phi_{mp,mp}$	0.2238	0.2543	0.1410	0.2202	0.3177
Φ_{mp,π^*}	-0.0212	-0.0304	-0.0834	-0.0246	0.0578
$\Phi_{\pi^*,\delta}$	0.0354	0.0940	-0.0587	0.0348	0.1288
$\Phi_{\pi^*,\gamma}$	-0.0006	0.0007	-0.1186	-0.0005	0.1173
$\Phi_{\pi^*,u}$	-0.0026	0.0211	-0.1158	-0.0028	0.1105
$\Phi_{\pi^*,mp}$	-0.0007	0.0316	-0.1178	-0.0007	0.1153
Φ_{π^*,π^*}	0.9415	0.9475	0.8798	0.9449	0.9906
$\Omega_{\delta,\delta}$	1.9394	0.4090	0.6442	1.6494	4.1824
$\Omega_{\delta,\gamma}$	-0.0018	0.0006	-0.0286	-0.0013	0.0235

$\Omega_{\delta,u}$	-0.4591	0.0763	-1.6014	-0.3203	0.2127
$\Omega_{\delta,mp}$	-0.1289	0.1303	-0.6311	-0.0664	0.1639
Ω_{δ,π^*}	0.1586	0.1046	-0.3976	0.1268	0.8140
$\Omega_{\gamma,\gamma}$	0.0113	0.0094	0.0088	0.0111	0.0143
$\Omega_{\gamma,u}$	0.0010	0.0000	-0.0099	0.0006	0.0135
$\Omega_{\gamma,mp}$	0.0005	0.0003	-0.0059	0.0004	0.0072
Ω_{γ,π^*}	-0.0002	0.0001	-0.0091	-0.0002	0.0086
$\Omega_{u,u}$	0.4196	0.0853	0.1032	0.2560	1.2694
$\Omega_{u,mp}$	0.0425	0.0195	-0.0628	0.0060	0.2790
Ω_{u,π^*}	-0.0507	0.0274	-0.3455	-0.0249	0.1580
$\Omega_{mp,mp}$	0.1253	0.0986	0.0638	0.1054	0.2560
Ω_{mp,π^*}	-0.0011	0.0360	-0.0987	0.0000	0.0902
Ω_{π^*,π^*}	0.2379	0.0878	0.0773	0.1794	0.5827

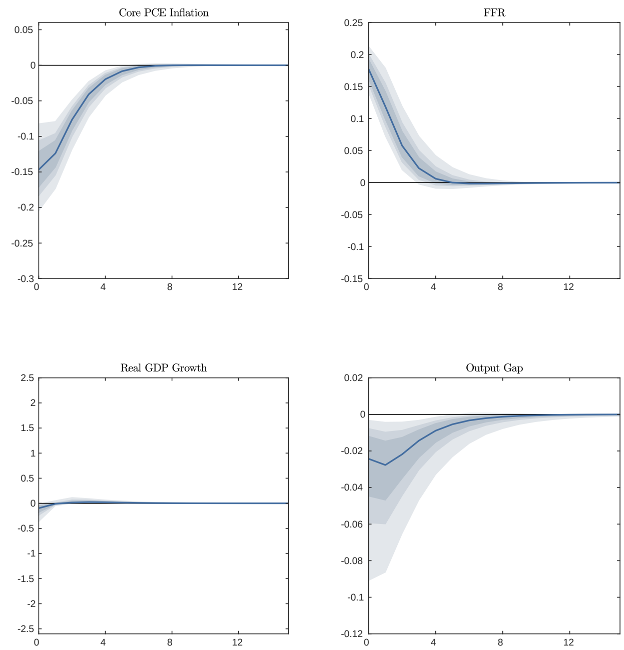
Table 6: **Estimated dynamic correlation of the disturbances in the Smets-Wouters model**

	Shock						
Variable	Total Productivity	Risk Premium	Government Spending	Investment Productivity	Monetary Policy	Price Markup	Wage Markup
Total Productivity	0.9141* (0.0348)	0.0096 (0.0239)	-0.2712* (0.0521)	-0.0922 (0.1051)	-0.0142 (0.1290)	0.0074 (0.1573)	0.0872 (0.0995)
Risk Premium	0.2099* (0.1217)	0.2397 (0.2056)	-0.1692 (0.1635)	-0.8591* (0.4121)	-1.2483* (0.6686)	0.7074 (0.6548)	0.4821 (0.4196)
Government Spending	-0.1787* (0.0394)	-0.0196 (0.0252)	0.6769* (0.0660)	0.1924 (0.1345)	0.0240 (0.1482)	0.1926 (0.1760)	0.1419 (0.1190)
Investment Productivity	0.0643* (0.0280)	-0.0324* (0.0175)	-0.0779* (0.0387)	0.6918* (0.0575)	-0.0878 (0.0964)	0.0180 (0.1008)	0.0756 (0.0912)
Monetary Policy	-0.0330* (0.0174)	-0.0335* (0.0148)	0.0192 (0.0276)	0.1143* (0.0438)	0.1838* (0.0771)	0.0011 (0.0842)	-0.0052 (0.0640)
Price Markup	-0.0058 (0.0058)	0.0014 (0.0055)	-0.0036 (0.0102)	0.0072 (0.0171)	0.0062 (0.0399)	0.6629* (0.0842)	-0.0082 (0.0222)
Wage Markup	0.0055 (0.0083)	0.0029 (0.0081)	0.0194 (0.0132)	-0.0024 (0.0261)	-0.0194 (0.0544)	-0.0035 (0.0532)	0.9422* (0.0331)

The entries are the mean and, in parenthesis, the standard error of the posterior marginal distribution of the elements of Φ in the law of motion for the disturbances: $s_t = \Phi s_{t-1} + e_t$. $A*$ is included if zero is not within the 90% posterior credible set.

Figure 3: Impulse response functions to monetary policy shock in NK model, median and distributions

(a) Independent AR(1)s case



(b) Unrestricted VAR(1) case

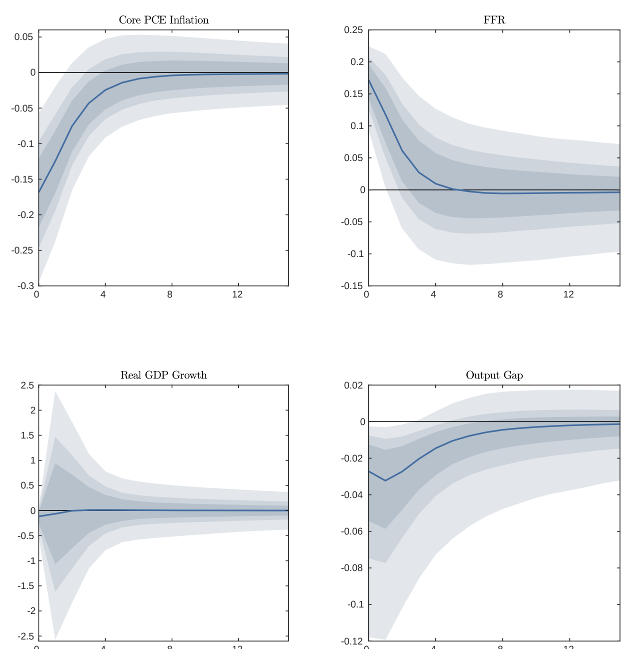


Table 7: **Variance decomposition in the Smets-Wouters model**

Variable	Shock						
	Total Productivity	Risk Premium	Government Spending	Investment Productivity	Monetary Policy	Price Markup	Wage Markup
Panel A. Independent AR(1) disturbances							
1-quarter ahead							
Output	0.016	0.289	0.475	0.1160	0.065	0.025	0.003
Hours	0.421	0.160	0.274	0.082	0.034	0.006	0.014
Real wage	0.010	0.016	0.000	0.006	0.012	0.268	0.682
Inflation	0.026	0.003	0.001	0.007	0.014	0.807	0.137
2-years ahead							
Output	0.177	0. 075	0.184	0.191	0.095	0.083	0.163
Hours	0.158	0.075	0.203	0.149	0.087	0.059	0.242
Real wage	0.098	0.018	0.000	0.061	0.055	0.269	0.474
Inflation	0.050	0.008	0.003	0.022	0.050	0.408	0.443
8-years ahead							
Output	0.200	0.023	0.134	0.072	0.033	0.034	0.474
Hours	0.064	0.026	0.170	0.066	0.033	0.026	0.596
Real wage	0.330	0.011	0.001	0.080	0.044	0.189	0.304
Inflation	0.046	0.007	0.004	0.021	0.044	0.321	0.542
Unconditional							
Output	0.133	0.014	0.208	0.043	0.019	0.020	0.489
Hours	0.047	0.015	0.257	0.041	0.019	0.015	0.558
Real wage	0.379	0.010	0.001	0.073	0.040	0.172	0.280
Inflation	0.044	0.006	0.006	0.019	0.037	0.279	0.593
Panel B. Dynamic VAR(1) disturbances							
1-quarter ahead							
Output	0.008	0.454	0.392	0.032	0.016	0.053	0.023
Hours	0.468	0.229	0.212	0.023	0.007	0.014	0.032
Real wage	0.045	0.036	0.004	0.002	0.007	0.325	0.564

Inflation	0.054	0.009	0.020	0.022	0.023	0.638	0.202
2-years ahead							
Output	0.108	0.161	0.173	0.066	0.019	0.103	0.308
Hours	0.210	0.180	0.069	0.136	0.012	0.046	0.289
Real wage	0.300	0.023	0.057	0.021	0.029	0.227	0.275
Inflation	0.058	0.020	0.067	0.056	0.063	0.288	0.386
8-years ahead							
Output	0.237	0.051	0.275	0.096	0.011	0.027	0.252
Hours	0.128	0.104	0.059	0.097	0.019	0.035	0.484
Real wage	0.452	0.046	0.199	0.120	0.011	0.059	0.069
Inflation	0.068	0.027	0.081	0.066	0.060	0.258	0.383
Unconditional							
Output	0.381	0.044	0.287	0.129	0.005	0.011	0.094
Hours	0.177	0.089	0.094	0.107	0.016	0.028	0.407
Real wage	0.456	0.047	0.258	0.142	0.005	0.019	0.036
Inflation	0.224	0.036	0.175	0.098	0.032	0.145	0.231

Figure 4: Posterior distribution of two policy-rule parameters in Smets-Wouters model with independent and correlated disturbances

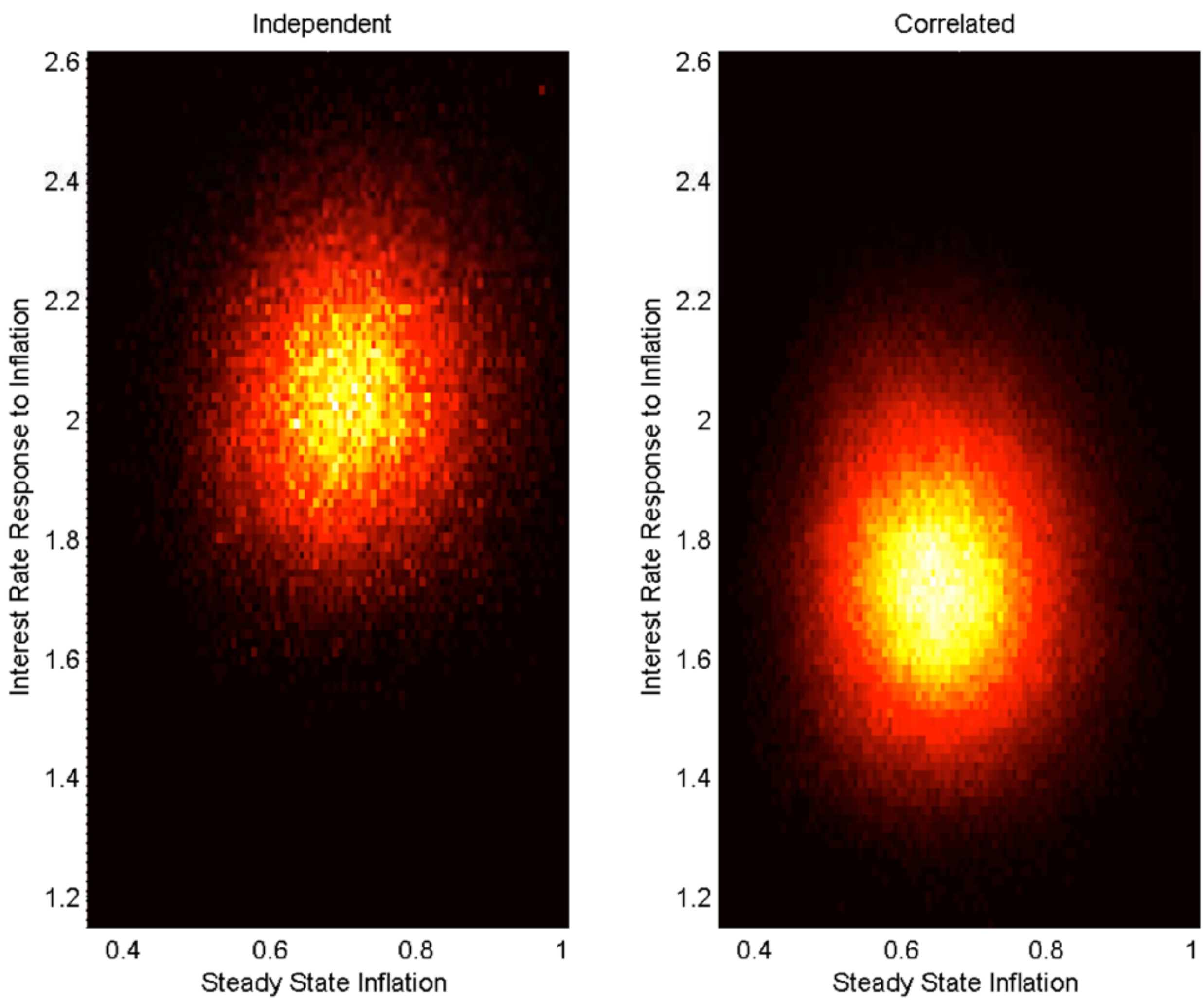


Table A.1: Prior and posterior in SW model, independent AR(1) disturbances

	Prior				Posterior					
	Dist	5%	Median	95%	Mode	Mean	SE	5%	Median	95%
γ^*	N	0.2355	0.4000	0.5645	0.3898	0.3864	0.0191	0.3523	0.3881	0.4145
l^*	N	-0.4935	0.0000	0.4935	0.0000	0.0003	0.3006	-0.4989	0.0020	0.4954
π^*	G	0.4652	0.6146	0.7931	0.6873	0.7100	0.1024	0.5454	0.7071	0.8843
$\beta^{-1}-1$	G	0.1111	0.2368	0.4339	0.1470	0.1698	0.0592	0.0830	0.1643	0.2765
ϕ	N	1.5327	4.0000	6.4673	6.1285	6.1955	1.1311	4.4047	6.1516	8.1194
σ_c	N	0.8832	1.5000	2.1168	1.4058	1.3673	0.1409	1.1508	1.3589	1.6113
λ	B	0.5242	0.7068	0.8525	0.7024	0.7083	0.0486	0.6218	0.7122	0.7807
ξ_w	B	0.3351	0.5000	0.6649	0.7056	0.6756	0.0701	0.5562	0.6788	0.7861
σ_l	N	0.7664	2.0000	3.2336	1.7248	1.7625	0.5421	0.9467	1.7220	2.7179
ξ_p	B	0.3351	0.5000	0.6649	0.7011	0.6845	0.0572	0.5850	0.6872	0.7735
ι_w	B	0.2526	0.5000	0.7474	0.5110	0.5137	0.1259	0.3061	0.5142	0.7203
ι_p	B	0.2526	0.5000	0.7474	0.2645	0.3024	0.1109	0.1438	0.2899	0.5046
ψ	B	0.2526	0.5000	0.7474	0.6195	0.6366	0.0693	0.5299	0.6326	0.7585
Φ_{SW}	N	1.0526	1.2500	1.4474	1.6617	1.6628	0.0764	1.5398	1.6608	1.7914
r_π	N	1.0888	1.5000	1.9112	1.9834	2.0435	0.1724	1.7654	2.0392	2.3341
ρ_{SW}	B	0.5701	0.7595	0.8971	0.8015	0.8008	0.0258	0.7562	0.8020	0.8405
r_y	N	0.0378	0.1200	0.2022	0.0846	0.0884	0.0207	0.0566	0.0872	0.1243
$r_{\Delta y}$	N	0.0378	0.1200	0.2022	0.2257	0.2257	0.0289	0.1788	0.2254	0.2739
α	N	0.2178	0.3000	0.3822	0.1676	0.1698	0.0179	0.1408	0.1695	0.1998
$\Phi_{A,1}$	N	0.1986	0.4964	0.7732	0.9601	0.9609	0.0139	0.9369	0.9618	0.9822
$\Phi_{B,1}$	N	0.2010	0.4959	0.7805	0.2021	0.2382	0.1478	0.0274	0.2206	0.5267
$\Phi_{G,1}$	N	0.1869	0.4994	0.7780	0.9945	0.9910	0.0062	0.9795	0.9922	0.9986
$\Phi_{EI,1}$	N	0.1957	0.4975	0.7853	0.7119	0.7147	0.0570	0.6204	0.7149	0.8089
$\Phi_{ER,1}$	N	0.1925	0.4958	0.7764	0.1698	0.1779	0.0713	0.0604	0.1787	0.2934
$\Phi_{EP,1}$	N	0.1967	0.4983	0.7772	0.7203	0.7053	0.0982	0.5365	0.7098	0.8575
$\Phi_{EW,1}$	N	0.1834	0.4979	0.7882	0.9802	0.9794	0.0098	0.9616	0.9807	0.9931
$-\Psi_{EP}$	B	0.1718	0.5000	0.8282	0.5470	0.5228	0.1363	0.2866	0.5291	0.7358
$-\Psi_{EW}$	B	0.1718	0.5000	0.8282	0.8926	0.8540	0.0641	0.7331	0.8653	0.9367
Ω_A	IG2	0.0291	0.0823	0.3889	0.2076	0.2143	0.0257	0.1758	0.2122	0.2596
Ω_B	IG2	0.0291	0.0823	0.3889	3.4472	3.9211	2.0819	1.0706	3.6287	7.8164
Ω_G	IG2	0.0291	0.0823	0.3889	0.3182	0.3285	0.0376	0.2723	0.3255	0.3946
Ω_{EI}	IG2	0.0291	0.0823	0.3889	0.2170	0.2272	0.0453	0.1621	0.2223	0.3092
Ω_{ER}	IG2	0.0291	0.0823	0.3889	0.0615	0.0648	0.0078	0.0528	0.0642	0.0786
Ω_{EP}	IG2	0.0291	0.0823	0.3889	0.0264	0.0279	0.0051	0.0203	0.0275	0.0371
Ω_{EW}	IG2	0.0291	0.0823	0.3889	0.0673	0.0701	0.0117	0.0526	0.0693	0.0906

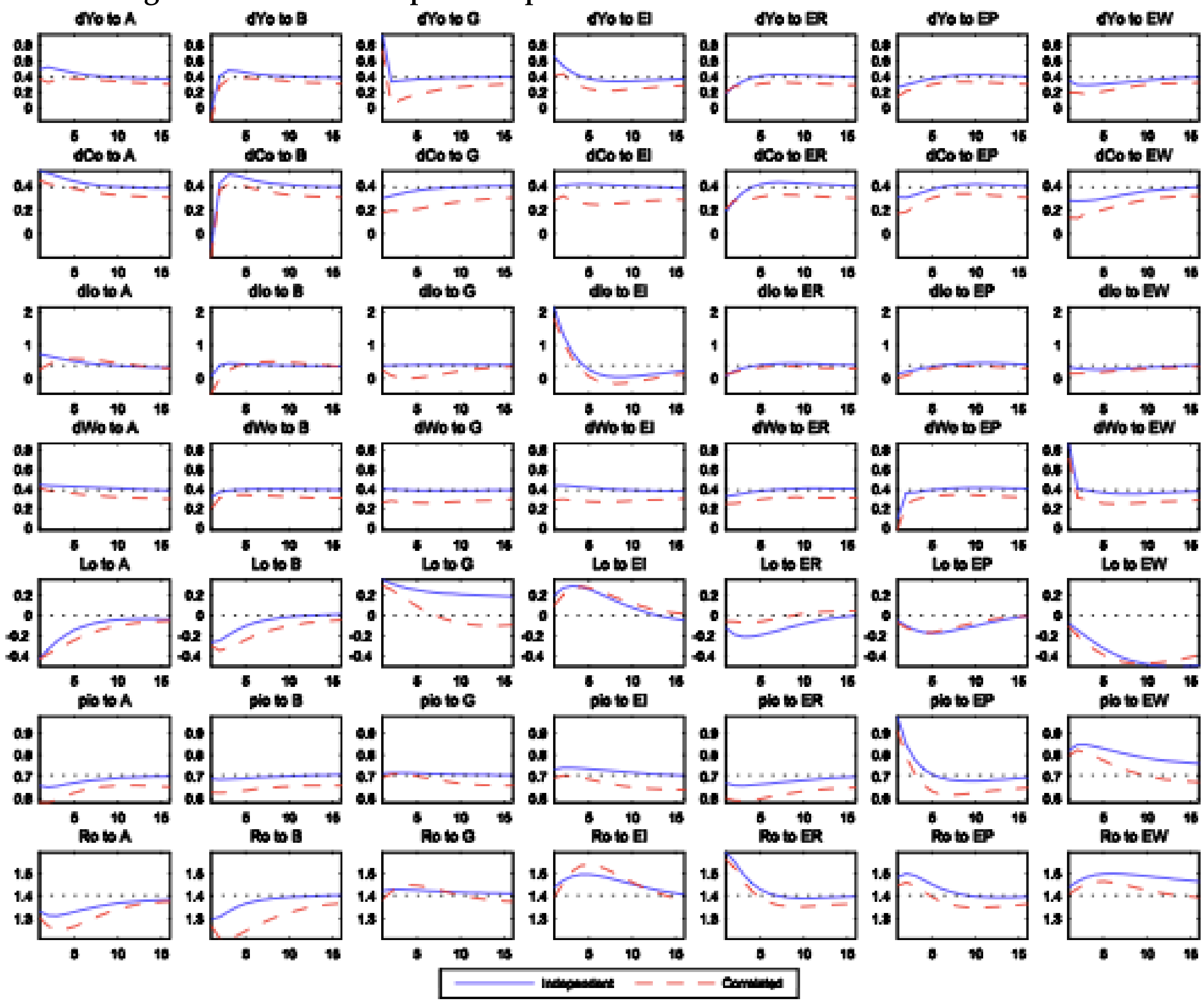
Table A.2: Prior and posterior distributions in SW model, correlated disturbances

	Prior				Posterior					
	Dist	5%	Median	95%	Mode	Mean	SE	5%	Median	95%
γ^*	N	0.2355	0.4000	0.5645	0.2753	0.2964	0.0201	0.2623	0.2976	0.3271
l^*	N	-0.4935	0.0000	0.4935	-0.0000	-0.0001	0.2976	-0.4911	-0.0003	0.4878
π^*	G	0.4652	0.6146	0.7931	0.6090	0.6559	0.1024	0.4937	0.6525	0.8295
$\beta^{-1} - 1$	G	0.1111	0.2368	0.4339	0.2335	0.2645	0.0907	0.1304	0.2563	0.4259
ϕ	N	1.5327	4.0000	6.4673	5.1223	5.3067	1.1751	3.3973	5.2952	7.2588
σ_c	N	0.8832	1.5000	2.1168	1.4438	1.5421	0.2238	1.2028	1.5300	1.9303
λ	B	0.5242	0.7068	0.8525	0.5250	0.6873	0.0629	0.5668	0.6965	0.7745
ξ_w	B	0.3351	0.5000	0.6649	0.6400	0.5441	0.0545	0.4560	0.5430	0.6356
σ_l	N	0.7664	2.0000	3.2336	0.9592	1.2453	0.5283	0.4435	1.2086	2.1817
ξ_p	B	0.3351	0.5000	0.6649	0.5150	0.5832	0.0628	0.4772	0.5845	0.6841
ι_w	B	0.2526	0.5000	0.7474	0.4756	0.5619	0.1284	0.3453	0.5644	0.7687
ι_p	B	0.2526	0.5000	0.7474	0.2083	0.2912	0.1105	0.1324	0.2789	0.4932
ψ	B	0.2526	0.5000	0.7474	0.3518	0.4891	0.0585	0.3944	0.4885	0.5857
Φ_{SW}	N	1.0526	1.2500	1.4474	1.4191	1.4946	0.0734	1.3765	1.4930	1.6180
r_π	N	1.0888	1.5000	1.9112	1.5055	1.7383	0.1887	1.4361	1.7327	2.0606
ρ_{SW}	B	0.5701	0.7595	0.8971	0.7611	0.7535	0.0325	0.6979	0.7552	0.8035
r_y	N	0.0378	0.1200	0.2022	0.0564	0.0801	0.0302	0.0330	0.0787	0.1314
$r_{\Delta y}$	N	0.0378	0.1200	0.2022	0.2250	0.1913	0.0305	0.1418	0.1909	0.2420
α	N	0.2178	0.3000	0.3822	0.0395	0.0994	0.0183	0.0713	0.0984	0.1311
$\Phi_{A,A,1}$	N	0.1787	0.4932	0.7524	0.9302	0.9141	0.0348	0.8578	0.9137	0.9719
$\Phi_{A,B,1}$	N	-0.2779	-0.0033	0.2907	-0.0088	0.0096	0.0239	-0.0366	0.0135	0.0399
$\Phi_{A,G,1}$	N	-0.2820	-0.0025	0.2848	-0.2914	-0.2712	0.0521	-0.3588	-0.2702	-0.1876
$\Phi_{A,EI,1}$	N	-0.2886	-0.0049	0.2828	-0.0961	-0.0922	0.1051	-0.2675	-0.0899	0.0753
$\Phi_{A,ER,1}$	N	-0.2817	0.0021	0.2944	0.0222	-0.0142	0.1290	-0.2228	-0.0162	0.2012
$\Phi_{A,EP,1}$	N	-0.2768	-0.0006	0.2862	-0.0588	0.0074	0.1573	-0.2537	0.0096	0.2625
$\Phi_{A,EW,1}$	N	-0.2859	0.0016	0.3046	0.1247	0.0872	0.0995	-0.0767	0.0873	0.2513
$\Phi_{B,A,1}$	N	-0.2828	0.0012	0.2923	0.0693	0.2099	0.1217	0.0385	0.1974	0.4282
$\Phi_{B,B,1}$	N	0.1942	0.4910	0.7528	0.6947	0.2397	0.2056	-0.0321	0.1961	0.6489
$\Phi_{B,G,1}$	N	-0.2874	0.0048	0.3024	-0.1308	-0.1692	0.1635	-0.4602	-0.1550	0.0707
$\Phi_{B,EI,1}$	N	-0.2864	-0.0036	0.2867	-0.4177	-0.8591	0.4121	-1.5980	-0.8148	-0.2752
$\Phi_{B,ER,1}$	N	-0.2799	0.0076	0.3013	-0.5817	-1.2483	0.6686	-2.4081	-1.2064	-0.2460
$\Phi_{B,EP,1}$	N	-0.2786	-0.0013	0.2933	0.2720	0.7074	0.6548	-0.2060	0.6163	1.9356
$\Phi_{B,EW,1}$	N	-0.2873	-0.0018	0.3019	0.1370	0.4821	0.4196	-0.1093	0.4332	1.2362

		Prior			Posterior					
	Dist	5%	Median	95%	Mode	Mean	SE	5%	Median	95%
$\Phi_{G,A,1}$	N	-0.2914	-0.0008	0.2920	-0.1899	-0.1787	0.0394	-0.2445	-0.1779	-0.1155
$\Phi_{G,B,1}$	N	-0.2791	0.0017	0.2847	-0.0630	-0.0196	0.0252	-0.0596	-0.0191	0.0187
$\Phi_{G,G,1}$	N	0.1996	0.4931	0.7622	0.7551	0.6769	0.0660	0.5686	0.6770	0.7846
$\Phi_{G,EI,1}$	N	-0.2826	0.0010	0.2724	0.4629	0.1924	0.1345	-0.0144	0.1843	0.4281
$\Phi_{G,ER,1}$	N	-0.2816	0.0036	0.2758	-0.0990	0.0240	0.1482	-0.2175	0.0228	0.2704
$\Phi_{G,EP,1}$	N	-0.2839	0.0016	0.2958	0.1392	0.1926	0.1760	-0.0969	0.1927	0.4808
$\Phi_{G,EW,1}$	N	-0.2794	-0.0026	0.2871	0.3713	0.1419	0.1190	-0.0489	0.1394	0.3405
$\Phi_{EI,A,1}$	N	-0.2904	-0.0039	0.2980	0.0595	0.0643	0.0280	0.0203	0.0632	0.1117
$\Phi_{EI,B,1}$	N	-0.2901	-0.0019	0.2762	-0.0385	-0.0324	0.0175	-0.0638	-0.0305	-0.0083
$\Phi_{EI,G,1}$	N	-0.2882	-0.0026	0.2836	-0.1085	-0.0779	0.0387	-0.1463	-0.0749	-0.0199
$\Phi_{EI,EI,1}$	N	0.1922	0.4911	0.7543	0.6312	0.6918	0.0575	0.5961	0.6928	0.7854
$\Phi_{EI,ER,1}$	N	-0.2855	0.0030	0.2898	-0.1230	-0.0878	0.0964	-0.2495	-0.0852	0.0646
$\Phi_{EI,EP,1}$	N	-0.2782	0.0022	0.2846	0.0318	0.0180	0.1008	-0.1424	0.0148	0.1882
$\Phi_{EI,EW,1}$	N	-0.2901	0.0012	0.2830	0.0213	0.0756	0.0912	-0.0741	0.0781	0.2184
$\Phi_{ER,A,1}$	N	-0.2710	0.0030	0.2851	-0.0283	-0.0330	0.0174	-0.0624	-0.0327	-0.0052
$\Phi_{ER,B,1}$	N	-0.2781	0.0038	0.2789	-0.0810	-0.0335	0.0148	-0.0620	-0.0308	-0.0149
$\Phi_{ER,G,1}$	N	-0.2786	-0.0002	0.2843	0.0075	0.0192	0.0276	-0.0260	0.0190	0.0652
$\Phi_{ER,EI,1}$	N	-0.2735	0.0015	0.2936	0.1007	0.1143	0.0438	0.0459	0.1126	0.1896
$\Phi_{ER,ER,1}$	N	0.2035	0.4908	0.7542	0.1292	0.1838	0.0771	0.0577	0.1832	0.3102
$\Phi_{ER,EP,1}$	N	-0.2715	0.0017	0.2855	-0.0099	0.0011	0.0842	-0.1391	0.0017	0.1384
$\Phi_{ER,EW,1}$	N	-0.2837	0.0003	0.2837	0.0423	-0.0052	0.0640	-0.1122	-0.0040	0.0976
$\Phi_{EP,A,1}$	N	-0.2843	-0.0005	0.2877	-0.0107	-0.0058	0.0058	-0.0157	-0.0056	0.0034
$\Phi_{EP,B,1}$	N	-0.2901	0.0049	0.2851	-0.0007	0.0014	0.0055	-0.0071	0.0012	0.0107
$\Phi_{EP,G,1}$	N	-0.2960	-0.0016	0.2891	-0.0273	-0.0036	0.0102	-0.0204	-0.0035	0.0126
$\Phi_{EP,EI,1}$	N	-0.2846	-0.0007	0.2847	-0.0007	0.0072	0.0171	-0.0207	0.0078	0.0336
$\Phi_{EP,ER,1}$	N	-0.2898	0.0014	0.2903	0.0253	0.0062	0.0399	-0.0570	0.0051	0.0735
$\Phi_{EP,EP,1}$	N	0.1988	0.4934	0.7435	0.8069	0.6629	0.0842	0.5205	0.6660	0.7962
$\Phi_{EP,EW,1}$	N	-0.2785	0.0025	0.2842	-0.0243	-0.0082	0.0222	-0.0484	-0.0057	0.0235
$\Phi_{EW,A,1}$	N	-0.2686	-0.0004	0.2937	0.0099	0.0055	0.0083	-0.0083	0.0055	0.0191
$\Phi_{EW,B,1}$	N	-0.2822	-0.0018	0.2723	-0.0008	0.0029	0.0081	-0.0091	0.0027	0.0149
$\Phi_{EW,G,1}$	N	-0.2895	0.0008	0.2891	0.0124	0.0194	0.0132	-0.0003	0.0183	0.0429
$\Phi_{EW,EI,1}$	N	-0.2707	0.0043	0.2966	-0.0247	-0.0024	0.0261	-0.0434	-0.0034	0.0427
$\Phi_{EW,ER,1}$	N	-0.2839	-0.0002	0.2809	-0.0082	-0.0194	0.0544	-0.1077	-0.0193	0.0676
$\Phi_{EW,EP,1}$	N	-0.2856	-0.0028	0.2851	0.0002	-0.0035	0.0532	-0.0854	-0.0061	0.0867
$\Phi_{EW,EW,1}$	N	0.1819	0.4911	0.7657	0.9735	0.9422	0.0331	0.8826	0.9481	0.9830

	Prior				Posterior					
	Dist	5%	Median	95%	Mode	Mean	SE	5%	Median	95%
$-\Psi_{EP}$	B	0.1718	0.5000	0.8282	0.5134	0.3397	0.1212	0.1422	0.3382	0.5421
$-\Psi_{EW}$	B	0.1718	0.5000	0.8282	0.9665	0.6739	0.1094	0.4751	0.6877	0.8264
Ω_A	IG2	0.0291	0.0823	0.3889	0.2431	0.2257	0.0284	0.1828	0.2235	0.2756
Ω_B	IG2	0.0291	0.0823	0.3889	0.4768	5.1996	3.3725	0.6226	4.8122	11.5165
Ω_G	IG2	0.0291	0.0823	0.3889	0.2563	0.2584	0.0310	0.2120	0.2562	0.3126
Ω_{EI}	IG2	0.0291	0.0823	0.3889	0.0582	0.0912	0.0334	0.0473	0.0854	0.1552
Ω_{ER}	IG2	0.0291	0.0823	0.3889	0.0497	0.0542	0.0066	0.0444	0.0537	0.0659
Ω_{EP}	IG2	0.0291	0.0823	0.3889	0.0296	0.0257	0.0052	0.0181	0.0252	0.0351
Ω_{EW}	IG2	0.0291	0.0823	0.3889	0.0855	0.0724	0.0132	0.0528	0.0712	0.0961

Figure A.3: Median impulse response functions in the Smets-Wouters model



Variables: dY is output growth, dCo is consumption growth, dIo is investment, dWo is wage growth, Lo is hours, pio is inflation, and Ro is the nominal interest rate.

Disturbances: total factor productivity (A), risk premium (B), government spending (G), investment-specific productivity (EI), nominal interest rates (ER), price markups (EP), wage markups (EW).

Table A.4: Posterior variance decompositions, Smets-Wouters model, percentiles 5, 95

	Shock						
Variable	Total productivity	Risk premium	Government spending	Investment productivity	Monetary Policy	Price markup	Wage markup
Panel A. Independent AR(1) disturbances							
1-quarter ahead							
Output	.004, .044	.234, .355	.397, .545	.072, .168	.043, .099	.016, .038	.000, .017
Hours	.349, .491	.123, .210	.229, .326	.053, .120	.021, .058	.002, .011	.006, .028
Real wage	.002, .029	.004, .049	.000, .001	.002, .013	.003, .030	.208, .341	.583, .756
Inflation	.011, .051	.001, .012	.000, .003	.001, .023	.005, .033	.680, .907	.069, .222
2-years ahead							
Output	.109, .264	.047, .137	.121, .267	.107, .299	.051, .162	.054, .124	.069, .297
Hours	.101, .226	.047, .144	.147, .269	.092, .229	.048, .145	.036, .093	.122, .403
Real wage	.028, .231	.003, .046	.001, .002	.004, .107	.024, .099	.293, .389	.332, .624
Inflation	.029, .083	.006, .026	.000, .007	.028, .065	.026, .091	.181, .544	.315, .557
8-years ahead							
Output	.111, .323	.013, .047	.066, .246	.034, .143	.014, .072	.017, .065	.315, .633
Hours	.039, .103	.014, .055	.093, .279	.034, .121	.015, .069	.014, .049	.420, .752
Real wage	.121, .590	.004, .027	.000, .002	.031, .164	.018, .087	.112, .323	.147, .520
Inflation	.023, .079	.002, .022	.001, .010	.005, .064	.021, .083	.223, .435	.408, .674
Unconditional							
Output	.048, .278	.005, .031	.051, .642	.012, .106	.005, .051	.006, .048	.201, .762
Hours	.018, .088	.005, .035	.076, .674	.012, .096	.006, .048	.005, .035	.227, .828
Real wage	.130, .708	.004, .024	.000, .005	.025, .157	.013, .084	.084, .306	.108, .512
Inflation	.018, .081	.002, .018	.002, .014	.004, .062	.015, .076	.146, .405	.429, .786
Panel B. Dynamic VAR(1) disturbances							
1-quarter ahead							
Output	.000, .067	.355, .560	.259, .488	.003, .076	.001, .060	.021, .114	.005, .073
Hours	.383, .547	.173, .303	.152, .275	.003, .050	.000, .031	.003, .041	.014, .075
Real wage	.014, .101	.006, .093	.000, .026	.000, .016	.000, .032	.219, .440	.419, .695
Inflation	.021, .104	.000, .067	.001, .062	.001, .074	.001, .086	.482, .790	.111, .308
2-years ahead							
Output	.035, .220	.082, .300	.093, .288	.017, .190	.002, .098	.028, .221	.173, .460
Hours	.110, .323	.069, .342	.035, .136	.036, .289	.001, .074	.006, .144	.145, .487
Real wage	.152, .458	.008, .090	.007, .182	.003, .116	.001, .125	.097, .404	.110, .477
Inflation	.021, .120	.001, .130	.009, .170	.005, .180	.007, .182	.187, .426	.247, .528
8-years ahead							
Output	.098, .395	.023, .121	.143, .428	.027, .225	.001, .053	.008, .084	.110, .460
Hours	.059, .250	.037, .254	.022, .154	.033, .263	.003, .099	.008, .120	.286, .671
Real wage	.311, .590	.009, .126	.094, .323	.032, .243	.001, .050	.021, .156	.028, .162
Inflation	.028, .142	.004, .124	.017, .194	.013, .184	.009, .173	.167, .377	.241, .523
Unconditional							
Output	.189, .552	.010, .126	.162, .432	.030, .254	.001, .030	.002, .048	.022, .308
Hours	.076, .391	.028, .225	.030, .250	.037, .252	.003, .085	.007, .103	.154, .636
Real wage	.309, .607	.005, .134	.135, .398	.038, .264	.000, .028	.004, .071	.008, .125
Inflation	.074, .431	.007, .116	.059, .321	.033, .203	.006, .116	.048, .267	.080, .436