INNOCENZO GASPARINI INSTITUTE FOR ECONOMIC RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

DSGE models in the frequency domain

Luca Sala

Working Paper n. 504

This Version: December, 2013

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

DSGE models in the frequency domain

Luca Sala*

Università Bocconi, IGIER and Baffi Center

Abstract

We use frequency domain techniques to estimate a medium-scale DSGE model on different frequency bands. We show that goodness of fit, forecasting performance and parameter estimates vary substantially with the frequency bands over which the model is estimated.

Estimates obtained using subsets of frequencies are characterized by significantly different parameters, an indication that the model cannot match all frequencies with one set of parameters. In particular, we find that: i) the low frequency properties of the data strongly affect parameter estimates obtained in the time domain; ii) the importance of economic frictions in the model changes when different subsets of frequencies are used in estimation.

This is particularly true for the investment cost friction and habit persistence: when low frequencies are present in the estimation, the investment cost friction and habit persistence are estimated to be higher than when low frequencies are absent.

JEL Classification: C11, C32, E32

Keywords: DSGE models, frequency domain, band maximum likelihood.

*Correspondence to: Luca Sala - Department of Economics, Università Bocconi, Via Roentgen 1, 20136 Milano, Italy; Phone: +39 0258363062; Fax: +39 0258363302; e-mail: luca.sala@unibocconi.it. Carmine Gabriele provided outstanding research assistance. This project started while I was visiting Norges Bank. I wish to thank the Institution for its hospitality. I also wish to thank seminar participants at Norges Bank, where I presented a preliminary version in September 2008, for their comments. I wish to thank Domenico Giannone for providing me with the codes for the BVAR estimation. Financial support from the Baffi Center at Università Bocconi is gratefully acknowledged.

1 Introduction

DSGE models have become the benchmark tool for quantitative dynamic macroeconomics both for academics and for policy-makers. Estimated versions of these models are used routinely by policy institutions and inform policy decisions at many Central Banks around the world.

The empirical performance of these models has been studied in a number of dimensions. Among others, Smets and Wouters, 2003, analyze the fit by comparing the autocovariance functions of the estimated model and the data; Del Negro, Schorfheide, Smets and Wouters, 2007, study the forecasting performance with regard to unrestricted Bayesian VARs; Canova and Sala, 2009, focus on identification issues; Guerron-Quintana, 2010, studies the sensitivity of the results to the observables used in the estimation.

In this paper, we use a state-of-the-art medium-scale DSGE model similar to those developed by Christiano *et al.*, 2005 or Smets and Wouters, 2007, and study it in the frequency domain. We exploit the feature that the likelihood function of the model has an asymptotically equivalent representation in the frequency domain in which the observations are asymptotically orthogonal (Hannan, 1970). This orthogonality property allows us to estimate the model with Bayesian methods on various frequency bands. We then evaluate fit, parameter values and out-of-sample forecasting performance for models estimated across frequency bands.

DSGE models are typically estimated in the time domain either by combining a set of priors with the likelihood function in a Bayesian framework (An and Schorfheide, 2007) or by maximum likelihood (Ireland, 2004). This is equivalent to fitting the model over the whole spectral density. As the model generates cross-frequency restrictions, the presence in the estimation of frequencies that the model is not intended to explain may affect the estimates. Potential misspecification at some frequencies may spill to all the spectrum and contaminate the estimates (Hansen and Sargent, 1993, focus on the biases caused by misspecification of seasonal frequencies; Cogley, 2001a, focuses on the misspecification of the trend component). In general, if no misspecification is present, parameter estimates will not depend on the frequencies used. Estimation on frequency bands would just be less efficient than estimation over the entire frequency domain.

In this paper, we do not take a stand on what is the form of the potential misspecification or at what frequencies it is mainly located. Given the DSGE model that we use, we analyze whether results are sensitive to the frequency bands used in the estimation.

There are predecessors to this paper. Altug, 1989, estimates a singular real business cycle (RBC) model in the frequency domain by exploiting its representation as a dynamic factor model; Hansen and Sargent, 1993, use the likelihood function expressed in the frequency domain to study the effects of using seasonally adjusted versus seasonally unadjusted data in the estimation of rational expectations models. Watson, 1993, evaluates the fit of calibrated RBC models by comparing the spectral densities of the data and of the model. Diebold *et al.*, 1998, Berkowitz, 2001, and Cogley, 2001b, stress the importance of estimating economic models over subsets of frequencies in the presence of misspecification or measurement errors. They focus on different classes of models and use different loss functions. Christiano and Vigfusson, 2003, consider an economic model which is a simplified version of the model analyzed here, but they do not consider estimation over subsets of frequencies.
Our paper is also related to a large literature that, starting from Kydland and Prescott, 1982, estimates DSGE models on prefiltered data[1,2]. We will discuss the relations with this literature in greater detail below. We note here that our estimation method, which is based on the likelihood function computed over frequency bands, is not equivalent to any of the approaches commonly used in the empirical literature on DSGE models.

Results show that current generation DSGE models perform reasonably well when estimated over all frequencies, in comparison to first-generation RBC models analyzed in Watson, 1993. There are a number of dimensions however, in which they do not fit the data satisfactorily. These dimensions include the interaction between real and nominal variables and the labor market variables.

We show that parameter estimates and model's fit depend critically on the inclusion of few low frequencies in the estimation. When we analyze the role of various frictions over frequencies, we see that the investment cost function and the associated exogenous shock are estimated to be very important only when low frequencies are present.

In terms of forecasting performance, we compare DSGEs with BVARs, and observe that there are some gains from using models estimated on subsets of frequencies. It is nevertheless difficult to

[1]Table 1 in Ng and Gorodnichenko, 2010 provides a long list of references to papers that use different methods to filter data and/or model, with a special emphasis to the treatment of low frequency components.

[2]Standard filters used in the macroeconomic literature are the Hodrick and Prescott (HP), 1997, filter, and the band-pass filter, see Baxter and King, 1999, or Christiano and Fitzgerald, 2003.

select the best forecasting model, as the forecasting performance varies substantially, depending on the variable to be forecast and on the forecasting horizon.

The paper is organized as follows. Section 2 discusses the econometric model and compares our approach with others in the literature. Section 3 briefly presents the DSGE model. Section 4 discusses the estimation results. Section 5 deals with the forecasting exercise. Section 6 concludes.

2 The likelihood function of DSGE models in the frequency domain

In this Section, we present the well-known result that the Gaussian log-likelihood function of a state-space model has a counterpart in the frequency domain.

Given a linearized DSGE model, parameterized by the vector of parameters θ:

$$
\begin{aligned}
y_t &= Z(\theta)x_t && (1)\\
x_t &= A(\theta)x_{t-1} + B(\theta)u_t^{\theta} && (2)
\end{aligned}
$$

with $V(u_t^{\theta}) = \Sigma(\theta)$, in which the $(N \times 1)$ vector y_t in the measurement equation (1) represents the observables and the $(k \times 1)$ vector x_t in the transition equation (2) denotes the states, the $(N \times N)$ spectral density matrix for y_t can be written as:

$$
S_y(\omega,\theta) = \frac{1}{2\pi} Z(\theta)(I - A(\theta)e^{-i\omega})^{-1}B(\theta)\Sigma(\theta)B(\theta)'(I - A(\theta)'e^{i\omega})^{-1}Z(\theta)' \qquad (3)
$$

where $\omega \in [0\ 2\pi]$ denotes the frequency.

The log-likelihood function of the state space model in (1) and (2) has an asymptotic counterpart in the frequency domain (Hannan, 1970 or Harvey, 1991)[3]:

$$
\log L(\theta, I_y) \propto -\frac{1}{2}\sum_{j=1}^{T}\left\{\log\left[\det S_y(\omega_j,\theta)\right] + \mathrm{tr}\left[S_y(\omega_j,\theta)^{-1} I_y(\omega_j)\right]\right\} \qquad (4)
$$

The likelihood function depends on two arguments: the periodogram $I_y(\omega_j)$, to be defined below,

[3] For finite T, equation (4) is an approximation to the time-domain log-likelihood. As $T \to \infty$, equation (4) converges to the time-domain log-likelihood (Harvey, 1989, p. 192).

and the spectral density of the model $S_y(\omega, \theta)$, defined in (3). The part of the likelihood that depends on the data is the periodogram, $I_y(\omega_j)$:

$$I_y(\omega_j) = \frac{1}{T} q(\omega_j) q(\omega_j)' \tag{5}$$

where

$$q(\omega_j) = \sum_{t=1}^{T} y_t e^{-i\omega_j t} \qquad \omega_j = \frac{2\pi j}{T}, j = 1, \dots, T \tag{6}$$

is the *discrete Fourier transform* of the observables y_t.

The elements in the summation in equation (4) are asymptotically uncorrelated. This implies that one can include only the elements associated to the frequencies ω_j of interest, estimate the model and check what the impact of different frequencies on parameter estimates is. It is therefore useful to add to equation (4) an indicator $w(\omega_j)$ that takes value 1 if frequency ω_j is included and value 0 if it is excluded:

$$\log L^w(\theta, I_y) \propto -\frac{1}{2} \sum_{j=1}^{T} w(\omega_j) \left\{ \log \left[\det S_y(\omega_j, \theta) \right] + \operatorname{tr} \left[S_y(\omega_j, \theta)^{-1} I_y(\omega_j) \right] \right\} \tag{7}$$

This is precisely what we do in the rest of the paper by choosing indicator functions $w(\omega_j)$ that give positive weight to well-defined sets of frequencies.

2.1 Relation with the literature

We now discuss the relation between our approach and the common practice of estimating models in the time domain using prefiltered data. The literature on the topic is very large and a number of papers (Singleton, 1988, Cogley, 2001a and, more recently, Ng and Gorodnichenko, 2010, Canova and Ferroni, 2011 or Canova, 2012) have discussed under what conditions the estimation on prefiltered data delivers consistent estimates. The literature has mainly focused on the treatment of low frequency components, but the same ideas apply to any frequency.

One approach assumes a stationary economic model, specifically developed to explain business cycles. The model is matched with data detrended with a statistical filter (Smets and Wouters, 2003, Rubio-Ramirez and Rabanal, 2005). This is justified only under the (often implicit) assumption that the optimal decision rules of agents can be factorized in a trend and in a stationary part and that removing the trend does not distort the business cycle, stationary component. If this is not true, estimates will be biased (Singleton, 1988). In general, Cogley, 2001a, Doorn, 2006, and Ng and Gorodnichenko, 2010 convincingly show that estimates are severely biased when model and data are filtered with different filters.

A different approach specifies a non-stationary model and prefilters both the data and the model with the same filter. As in Cogley, 2001a, let us consider a situation in which the vector of observables is composed by $x_t = [\Delta x_{1t}\ x_{2t}]$, where x_{1t} is a vector of non-stationary variables[4] and x_{2t} is a vector of stationary variables.[5] It can be shown that the likelihood function for x_t and the likelihood function for $K(L)x_t = [S(L)\Delta x_{1t}\ H(L)x_{2t}]$, where $H(L)$ denotes the HP filter and $S(L) = H(L)/(1-L)$, are the same (Cogley, 2001a, Proposition 2)[6]. Maximum likelihood (ML henceforth) estimation where data and model are both filtered with $K(L)$ is equivalent to ML estimation on x_t. As the band spectral ML estimator in general differs from the ML estimator[7], it will also differ from the ML estimator on prefiltered data $K(L)x_t$.

3 The model

The DSGE model which we estimate shares many features with Smets and Wouters, 2007 and Christiano, Eichenbaum and Evans, 2005. The particular specification which we employ builds on Justiniano *et al.*, 2010 and Sala *et al.*, 2011. The model combines a RBC core with Keynesian features. The RBC core model features habit formation, investment adjustment costs, and variable capital utilization, while the Keynesian features include monopolistic competition in goods and labor markets, and nominal price and wage rigidities. The model also includes growth in the form of a technology shock with a unit root, as in Altig *et al.*, 2005. There are seven exogenous shocks: household preferences, labor-augmenting technology, investment-specific technology, government spending, price and wage markup, and monetary policy, all assumed to follow $AR(1)$ processes. As the model is relatively standard, we only report the stationary, log-linearized version here. Appendix A presents the model in greater detail.

[4]The vector x_{1t} could be composed either by trend-stationary or difference stationary processes.

[5]This will also be the setup of our model, in which the technology process has a unit root, which is inherited by some of the observables.

[6]The argument would be the same if $H(L)$ was a band-pass filter.

[7]As shown above, the two are asymptotically equivalent only when all frequencies are considered in the summation in equation 7.

Effective capital:

$$\hat{k}_t + \hat{\epsilon}^z_t = \hat{\nu}_t + \hat{\bar{k}}_{t-1}; \tag{8}$$

Physical capital accumulation:

$$\hat{\bar{k}}_t = \frac{1-\delta}{\gamma_z}\left[\hat{\bar{k}}_{t-1} - \hat{\epsilon}^z_t\right] + \left(1 - \frac{1-\delta}{\gamma_z}\right)\left[\hat{inv}_t + \hat{\epsilon}^{inv}_t\right]; \tag{9}$$

Marginal utility of consumption:

$$\begin{aligned}\left(1-\frac{h}{\gamma_z}\right)\left(1-\frac{\beta h}{\gamma_z}\right)\hat{\lambda}_t = \frac{h}{\gamma_z}\left[\hat{c}_{t-1} - \hat{\epsilon}^z_t\right] - \left(1+\frac{\beta h^2}{\gamma_z^2}\right)\hat{c}_t \\ +\frac{\beta h}{\gamma_z}E_t\left[\hat{c}_{t+1} + \hat{\epsilon}^z_{t+1}\right] + \left(1-\frac{h}{\gamma_z}\right)\left[\hat{\epsilon}^b_t - \frac{\beta h}{\gamma_z}E_t\hat{\epsilon}^b_{t+1}\right];\end{aligned} \tag{10}$$

Consumption Euler equation:

$$\hat{\lambda}_t = E_t\hat{\lambda}_{t+1} + \left[\hat{i}_t - E_t\hat{\pi}_{t+1}\right] - E_t\hat{\epsilon}^z_{t+1}; \tag{11}$$

Investment:

$$\hat{inv}_t = \frac{1}{1+\beta}\left[\hat{inv}_{t-1} - \hat{\epsilon}^z_t\right] + \frac{1}{\eta_k\gamma_z^2(1+\beta)}\left[\hat{q}_t + \hat{\epsilon}^{inv}_t\right] + \frac{\beta}{1+\beta}E_t\left[\hat{inv}_{t+1} + \hat{\epsilon}^z_{t+1}\right]; \tag{12}$$

Tobin's Q:

$$\hat{q}_t = \frac{\beta(1-\delta)}{\gamma_z}E_t\hat{q}_{t+1} + \left[1 - \frac{\beta(1-\delta)}{\gamma_z}\right]E_t\hat{r}^k_{t+1} - \left[\hat{i}_t - E_t\hat{\pi}_{t+1}\right]; \tag{13}$$

Capital utilization:

$$\hat{\nu}_t = \eta_\nu \hat{r}^k_t; \tag{14}$$

Production function:

$$\hat{y}_t = \frac{Y+F}{Y}\left[\alpha\hat{k}_t + (1-\alpha)\,\hat{l}_t\right]; \tag{15}$$

Labor demand:

$$\hat{w}_t = \hat{mc}_t + \alpha\hat{k}_t - \alpha\hat{l}_t; \tag{16}$$

Capital renting:

$$\hat{r}_t^k = \hat{mc}_t - (1-\alpha)\hat{k}_t + (1-\alpha)\hat{l}_t; \tag{17}$$

Phillips curve:

$$\hat{\pi}_t = \iota_b \hat{\pi}_{t-1} + \iota_o \left[\hat{mc}_t + \hat{\epsilon}_t^p\right] + \iota_f E_t \hat{\pi}_{t+1}, \tag{18}$$

where

$$\iota_b = \frac{\gamma_p}{1+\beta\gamma_p}, \quad \iota_o = \frac{(1-\beta\theta_p)(1-\theta_p)}{\theta_p(1+\beta\gamma_p)[1+(\epsilon_p-1)\xi]}, \quad \iota_f = \frac{\beta}{1+\beta\gamma_p};$$

Aggregate wage:

$$\begin{aligned} \hat{w}_t &= \gamma_b \left[\hat{w}_{t-1} - \hat{\pi}_t + \gamma_w \hat{\pi}_{t-1} - \hat{\epsilon}_t^z\right] + \gamma_o \left[\omega \hat{l}_t - \hat{\lambda}_t + \hat{\epsilon}_t^b\right] \\ &\quad + \gamma_f E_t \left[\hat{w}_{t+1} + \hat{\pi}_{t+1} - \gamma_w \hat{\pi}_t + \hat{\epsilon}_{t+1}^z\right] + \gamma_o \hat{\epsilon}_t^w, \end{aligned} \tag{19}$$

where

$$\gamma_b = \frac{1}{(1+\beta)(1+\kappa_w)}, \quad \gamma_o = \frac{\kappa_w}{1+\kappa_w}, \quad \gamma_f = \frac{\beta}{(1+\beta)(1+\kappa_w)},$$

$$\kappa_w = \frac{(1-\beta\theta_w)(1-\theta_w)}{\theta_w(1+\beta)[1+\omega\epsilon^w/(\epsilon^w-1)]};$$

Government spending:

$$\hat{g}_t = \hat{y}_t + \frac{1-g_y}{g_y}\hat{\epsilon}_t^g; \tag{20}$$

Monetary policy rule:

$$\hat{i}_t = \rho_s \hat{i}_{t-1} + (1-\rho_s)\left[\phi_\pi \hat{\pi}_t + \phi_y\left(\hat{y}_t - \hat{y}^f lex_t\right)\right] + \hat{\epsilon}_t^m; \tag{21}$$

Resource constraint:

$$\hat{y}_t = \frac{c}{y}\hat{c}_t + \frac{inv}{y}\hat{inv}_t + \frac{g}{y}\hat{g}_t + \frac{r^k k}{y}\hat{\nu}_t. \tag{22}$$

$\hat{k}_t$ denotes effective capital, $\hat{\nu}_t$ is the capital utilization rate, $\hat{\bar{k}}_{t-1}$ is the physical capital, $\hat{\lambda}_t$ is the marginal utility of consumption, $\hat{c}_t$ is consumption, $\hat{y}_t$ is output, $\hat{y}_t - \hat{y}_t^{flex}$ is the output gap, $\hat{i}_t$ is the nominal interest rate, $\hat{q}_t$ is Tobin's q, $\hat{r}_t^k$ is the rental rate of capital, $\hat{\nu}_t$ is capacity utilization, $\hat{l}_t$ is labor input, $\hat{w}_t$ is the real wage, $\hat{mc}_t$ is marginal costs, $\hat{\pi}_t$ is the inflation rate, $\hat{g}_t$ is government spending as a fraction of output, $\hat{inv}_t$ is investment. $\hat{\epsilon}_t^z$ is the technology shock, ϵ_t^m is a monetary shock $\hat{\epsilon}_t^b$ is a preference shock, $\hat{\epsilon}_t^{inv}$ is the investment shock, ϵ_t^p is a price-markup shock, ϵ_t^w is a wage mark-up shock and ϵ_t^g is a shock to government spending.

α is the capital share, δ is the depreciation rate, γ_z is the growth rate of technology, β is the discount factor, h measures the degree of habits in consumption, ω is the inverse Frisch elasticity of labor supply, θ_w and θ_p are, respectively, the Calvo wage and price parameters, η_ν is the elasticity of capital utilization, η_k is the 2^{nd} derivative of the investment cost function, γ_w and γ_p denote, respectively, wage and price indexation, ε_w and ε_p are, respectively, the steady state wage and price markups, ϕ_π is the coefficient on inflation in the Taylor rule, ϕ_y the coefficient on the output gap in the Taylor rule, ρ_s is the degree of smoothing in the Taylor rule, F denotes fixed costs in the production function, endogenously determined by a zero-profit condition, g_y is the average ratio of government spending to output and ξ is the Kimball's aggregator parameter.

4 Estimation

In this Section, we describe the details of the estimation method. Let us start with the data.

4.1 Data

The seven variables used in the estimation are: $[\Delta\log(Y_t);\ \Delta\log(I_t);\ \Delta\log(C_t);\ \Delta\log(W_t);\ H_t;\ \pi_t;\ i_t]$ and they are matched with the following quarterly U.S. data from 1960Q1 to 2010Q4: (1) output growth: the quarterly growth rate of per capita real GDP; (2) investment growth: the quarterly growth rate of per capita real private investment plus real personal consumption expenditures of durable goods; (3) consumption growth: the quarterly growth rate of per capita real personal consumption expenditures of services and nondurable goods; (4) real wage growth: the quarterly growth rate of real compensation per hour; (5) employment: hours of all persons divided by population; (6) inflation: the quarterly growth rate of the GDP deflator; and (7) the nominal interest rate: the quarterly average of the federal funds rate.

We use (100 times the) growth rates for the variables that are non-stationary in the theoretical model (output, consumption, investment, and the real wage) and demean all the variables. We write the measurement equation of the state space form to match the seven series with their model counterparts. The first 170 observations in the dataset (up to 2002Q2) are used in estimation. The last 34 (from 2002Q3 to 2010Q4) are used to perform a pseudo-out-of-sample forecasting exercise to be described below.

It is instructive to look at the spectral density of the data and at the decomposition of the variance by frequency bands. We report them in Figure 1[8]. The growth rates of output, investment and consumption share a similar spectral shape: there is a spectral peak at frequencies corresponding to cycles with a 7 years period. The decomposition of the variance[9] in low (cycles with period longer than 8 years), business cycle (cycles with period between 1 and 8 years) and high frequencies (cycles with period shorter than 1 year) on non-overlapping bands is similar for variables in first differences, $\Delta\log(Y)$, $\Delta\log(I)$, $\Delta\log(C)$ and $\Delta\log(W)$; around 10% of the variance is located at low frequencies , $40-50\%$ of the variance is at business cycle frequencies , the remainder (around $35-40\%$) is located at high frequencies.

Variables in levels behave very differently: the spectral density of hours, inflation and the interest rate does not display any peak and it is monotonically decreasing to 0 as frequency increases. This is reminiscent of the "typical spectral shape" of economic variables, noted by Granger, 1966. The variance decomposition across frequencies is consequently very different from the one of growth rates: $75-80\%$ of the variance is located at low frequencies, $20-25\%$ is at business cycle frequencies and very little is left at high frequencies.

4.2 Estimating the model over frequency bands

We consider four frequency bands[10] in estimating the model:

1. *All*; we put $w(\omega_j) = 1$ on all frequencies in equation (7);

2. *High-pass*; we put $w(\omega_j) = 1$ on frequencies $\omega_j > 2\pi/32$, corresponding to cycles with period shorter than 8 years (32 quarters). This band contains 95% of the observations;

3. Business cycle, labeled *BC*; we set $w(\omega_j) = 1$ on frequencies between $2\pi/32 < \omega_j < 2\pi/4$, corresponding to cycles with period between 1 and 8 years (between 4 and 32 quarters). This band contains 44% of the observations;

4. *Low-pass*; we set $w(\omega_j) = 1$ for $\omega_j < 2\pi/4$, corresponding to cycles with period longer than 1

[8] The spectral density of the data has been estimated by fitting a BVAR with Minnesota and sum-of-the-coefficients priors on variables in levels: $[\log(Y_t); \log(I_t); \log(C_t); \log(W_t); H_t; \pi_t; i_t]$ and transforming non-stationary variables in growth rates (for details, see Appendix C). The gray area is a 68% credible set.

[9] Recall that the integral below the spectral density is equal to the variance. The integral over frequency bands gives the variance located at those frequencies.

[10] Each frequency band corresponds to a set of weights $w(\omega_j)$ in equation (7).

year (4 quarters)). This band contains 49% of the observations.

In order to keep enough data points in the estimation, the bands are partly overlapping. As we move from experiment 2 to 4, we give less weight to high frequencies and more weight to low frequencies.

We discard frequency zero in the estimation. We do so because the model implies cointegration relations between output, consumption, investment and the real wage: the unit root variables are all driven by the common stochastic trend in technology. When the trending variables are expressed in growth rates, their spectral density at frequency zero has reduced rank and the model is stochastically singular at that frequency.

4.3 Bayesian inference

We estimate the model with Bayesian methods (see An and Schorfheide, 2007, for a comprehensive survey). For each of the four sets of frequencies discussed above, we obtain the posterior distribution, $\log L^w(\theta; I_y) + p(\theta)$, where the log-likelihood function $\log L^w(\theta; I_y)$ is defined in (7) and $p(\theta)$ defines the log-priors for the parameters to be estimated.

Few parameters are calibrated using standard values: the discount factor β is set to 0.99, the capital depreciation rate δ to 0.025, the capital share α in the Cobb-Douglas production function is set to 0.33, the average ratio of government spending to output to 0.2, the steady-state growth rate, γ_z is set to 4% per year. Finally, the sensitivity of the firm's elasticity of demand with regard to shifts in its market share, the Kimball's aggregator parameter,[11] denoted by ξ, is calibrated to 10.

I estimate the remaining 13 structural parameters: the elasticity of the utilization rate to the rental rate of capital, η_ν;[12] the elasticity of the investment adjustment cost function, η_k; the habit parameter h and the labor supply elasticity ω; the steady-state wage and price mark-ups, ϵ^w and ϵ^p; the wage and price rigidity parameters, θ_w and θ_p; the wage indexation parameters, γ_w and γ_p and the monetary policy parameters ϕ_π, ϕ_y, and ρ_s. In addition, I estimate the autoregressive parameters of the exogenous disturbances, as well as the standard deviations of the innovations.

I stick to standard prior specification and standard calibration, as I want to compare results

[11]This parameter only appears in the log-linear version of the model.

[12]Following Smets and Wouters, 2007, I define ψ_ν so that $\eta_\nu = (1 - \psi_\nu) / \psi_\nu$ and estimate ψ_ν.

with those already present in the literature. These are summarized in the second column of Table 2. Most of the priors are standard in the literature; see, for example, Justiniano *et al.*, 2010.

The utilization rate elasticity ψ_ν and the habit parameter h are both assigned Beta priors with mean 0.5 and standard deviation 0.1, while the capital adjustment cost elasticity η_k is assigned a Normal prior with mean 4 and standard deviation 1.5. The labor supply elasticity ω (the inverse of the Frisch elasticity) is given a Gamma prior with mean 2 and standard deviation 0.75.

The two Calvo parameters for wage and price adjustment, θ_w and θ_p, are assigned Beta priors with means 3/4 and 2/3, respectively, and standard deviation 0.1, while the wage and price indexation parameters γ_w and γ_p are given a Beta distribution with mean 0.5 and standard deviation 0.15. The two steady-state wage and price markups are both given Normal priors centered around 1.15, with a standard deviation of 0.05.

The coefficient ϕ_π on inflation in the monetary policy rule is given a Normal prior with mean 1.7 and standard deviation 0.3, while the coefficient ϕ_y on the output gap is given a Gamma prior with mean 0.125 and standard deviation 0.1. The coefficient on the lagged interest rate, ρ_s, is assigned a Beta prior with mean 0.75 and a standard deviation of 0.1. All these are broadly consistent with empirically estimated monetary policy rules.

All persistence parameters for the shocks are given Beta priors with mean 0.5 and standard deviation 0.1. Finally, for the standard deviations of the shock innovations, we assign Inverse Gamma priors with mean 0.15 and standard deviation 0.15.

As often done in the literature (Justiniano *et al.*, 2010), we normalize few shocks before estimation. The investment specific shock is normalized so that it has a unitary impact on physical capital. The price markup shock is normalized so that it has a unitary impact on inflation and finally the preference shock ϵ_p is normalized so to have a unitary impact on the marginal utility of consumption.

The posterior distribution of the 4 models is obtained by generating draws with the random-walk Metropolis-Hastings algorithm (for a review, see An and Schorfheide, 2007). The mode is computed with the simulated annealing algorithm (see Corana *et al.*, 1987 or Goffe *et al.*, 1994). We generate a Markov chain of 200,000 draws per each of the four frequency bands. We discard the initial 50,000 draws and retain one every five subsequent draws. We simulate two different chains for each model and check robustness of the results. We verify convergence by checking that the recursive means and variances of the Markov chain stabilized after the burn-in period with a CUMSUM statistic.

As an additional check on our estimation method, we note that the log-likelihood expressed in the frequency domain is an approximation of the likelihood in the time domain. We therefore verify that results in the time domain are similar to those in the frequency domain when all frequencies are used. We draw from the posterior obtained from the time domain with the random-walk Metropolis-Hastings algorithm, evaluating the log-likelihood in the time domain with the Kalman filter. Parameters' posterior distributions (shown in column entitled *Time* in Table 2) are very similar to those obtained using all frequencies and implied spectral densities are virtually identical, thus confirming the goodness of the approximation to the likelihood in the frequency domain.[13]

4.4 Results

Figure 2 shows the diagonal of the log-spectral density matrix[14] for the seven observables for the 4 estimated models, together with the log-spectral density of the data. Figures 3 and 4 show, respectively, the autocovariance and the autocorrelation function of the data and of the 4 estimated models. Table 1 reports the ratio between the variance of the data and the variance implied by each of the four models evaluated at the posterior median. The ratio is computed over all frequencies and also over three non-overlapping bands: low (cycles with a period longer than 8 years), business cycle (cycles with a period between 2 and 8 years) and high (cycles with a period shorter than 1 year). Each entry in Table 1 is computed as:

$$\frac{\int_{\Omega_F} S_y^{ii}(\omega, \theta_M) d\omega}{\int_{\Omega_F} S_{y,VAR}^{ii}(\omega) d\omega} \tag{23}$$

where Ω_F denotes the frequency band, $S_y(\omega, \theta_M)$ is model M spectral density, $M = [$*All*, *High-pass*, *BC*, *Low-pass*$]$ and $S_{y,VAR}(\omega)$ is the median of the spectral density plotted in Figure 1.

There are significant differences in fit among the four models.

We first discuss the results for variables that are stationary in the model and enter the estimation in levels (hours, inflation and the interest rate). Secondly, we discuss the results for variables that are non-stationary in the model and enter the estimation in growth rates ($\Delta \log(Y)$, $\Delta \log(I)$,

[13]In Appendix B we report results from a Monte Carlo experiment. We generate data from the DSGE model and evaluate the small-sample properties of the frequency domain ML estimator on different frequency bands and in comparison with the time domain version. Results show that the frequency domain approximation, both on the whole spectrum and on subsets of frequencies, is remarkably good.

[14]We show here the log-spectral density, as opposed to the spectral density, in order to make the figure clearer and easier to interpret.

$\Delta \log(C)$ and $\Delta \log(W)$.

Regarding variables in levels, models that do not include low frequencies in the estimation (*High-pass* and *BC*) generate too little volatility (the ratio between the variance generated by the model and the variance of the data is around 35% for model *BC* and 45% for model *High-pass*), especially at low frequencies (the ratio between the variance generated by the model and the variance of the data at low frequency is around 10% for model *BC* and 20% for model *High-pass*). This is often compensated by too much volatility at business cycle or high frequencies.

Models that include low frequencies (*All* and *Low-pass*) generate the right amount of volatility for π and i while they over predict the variance of hours. These models also tend to generate too much volatility at business cycle frequencies.

It is interesting to note that, in a number of cases in which the variance of the model over all frequencies matches closely the variance of the data, this hides large errors at all frequency bands, which, when integrated over all frequencies, cancel out. This is, for example, the case of the variance of the interest rate obtained with model *All*: the overall model/data variance ratio is around one (0.84), but this is obtained by averaging ratios of 0.53, 1.90 and 1.73 at low, business cycle and high frequencies, respectively.

Regarding variables in first-differences, models that include low frequencies in the estimation (*All* and *Low-pass*) tend to over predict the variance located at low and business cycle frequencies and to over predict the overall variance. Models that do not include low frequencies overall generate less volatility but still over predict the volatility of the growth rates of output and consumption. Differences among models become less relevant at high frequencies, while model *Low-pass* under predict the high frequency volatility.

Turning to comovements, Figures 3 and 4 show that the 4 estimates can again be divided into 2 groups. Models that include low frequencies generate similar autocovariances and autocorrelations but are significantly different from those of models estimated without low frequencies.

Models that do not consider low frequencies generate too little autocorrelation for the variables in levels. No model is able to match the autocovariance of $\Delta \log(Y)$ and $\Delta \log(I)$ with the variables in levels, nor the autocorrelation of hours.

Summarizing, there are two results that emerge from comparing the fit of the four models.

The first is that there are a number of dimensions in which state-of-the-art DSGE models do not fit well. These dimensions relate to the interaction between nominal and real variables and to labor market variables (especially hours). Our analysis of fit in the frequency domain is reminiscent of the results in Watson, 1993, where he showed that the spectral density of first generation RBC models was very different from the spectral density of the data, even though the often used time domain statistics showed a good fit between the model and data.

The second is that low frequencies are key in shaping the fit of the model. The two estimated models in which low frequencies are present (*All* and *Low-pass*) deliver a very similar fit. The two estimated models in which low frequencies are not present (*High-pass* and *BC*) are very similar among themselves, but are very different from models in which low frequencies are present.

Let us now analyze how do parameters change when estimated over various frequencies. Table 2 shows the posterior distribution of parameters in the 4 cases. We compare each of the three alternative models to *All*. From the posterior distribution of model *All*, we compute the marginal posterior distribution of each parameter. We then compare the parameter's median in each of the three models [*Low-pass, BC, High-pass*] to the corresponding marginal posterior of model *All* and report percentiles. Parameter values that fall in the tails of the marginal posterior distribution (with less than 1% probability) are indicated in Table 2 with two asterisks. One asterisk indicate parameter values with probability between 1% and 5%. Some parameters are hardly sensitive to the different frequencies on which they are estimated. This is the case, for example, for price stickiness, θ_p, the two markups, ϵ_p and ϵ_w. These are parameters that have been shown to be weakly identified in this class of models (see Canova and Sala, 2009 and Iskrev, 2010).

Other parameters follow a clear pattern when moving from *Low-pass* to *BC* and to *High-pass*. When less weight is given to low frequencies and more weight to high frequencies, they change monotonically. ψ_z, the elasticity of capital utilization increases from 0.68 to 0.81, η_k, the second derivative of the investment cost function, decreases from 5.1 to 0.65, ϕ_π, the monetary policy response to inflation, decreases from 1.81 to 1.28 and to 1.12; ϕ_y, the monetary policy response to the output gap, increases from 0.19 to 0.34 and to 0.56.

Some parameters of the exogenous shocks display similar patterns. ρ_b, the AR(1) parameter of the preference shock and ρ_w, the AR(1) parameter of the wage markup shock, decrease respectively from 0.77 to 0.38 and from 0.53 to 0.28 as we move from low frequencies to high frequencies. This is the same pattern for ρ_p, the AR(1) parameter of the price mark-up shock and ρ_z, the AR(1) parameter of the technology shock. As it is to be expected, as we move from high to low frequencies, the shocks become more autocorrelated in order to generate more variance at low frequencies.

The standard deviation of the innovation to the investment shock σ_i increases from 0.06 to 0.29 as we move from high to low frequencies; the volatilities of other innovations change with the frequencies but not monotonically.

From the analysis above it is difficult to understand whether the observed changes in parameters are relevant in terms of fit. If the posterior is flat in the direction of a given parameter, a large variation in the value of that parameter will have a small effect on model's fit (Canova and Sala, 2009). In order to evaluate properly model's fit sensitivity to different parameters on various frequency bands, and, at the same time, to take into account the possibility of weak identification, we evaluate the sensitivity of the value of the posterior evaluated at the median with regard to changes in the parameters. More precisely, we consider as a benchmark the value of the posterior for the *All* case evaluated at the parameter's posterior median; we then replace one parameter at a time with the median estimated in the *High-pass*, *BC* and *Low-pass* case, respectively. Table 3 reports percentage variations in the value of the posterior. Large values for a given parameter indicate that model's fit is very sensitive to the variation in that parameter. We report parameters for which the posterior changes by more than 5% in bold, and we underline parameters that generate a percentage change in the posterior between 1% and 5%.

Fit is hardly sensitive to changes in the elasticity of capital utilization and the Frisch elasticity. The indexation parameters, the Calvo probability parameters and the markups, both in wages and prices are weakly identified and they do not significantly affect fit.

Models *All* and *Low-pass* are characterized by similar parameter values and fit. There is only one structural parameter to which the model's fit is sensitive and this is η_k, the elasticity of the investment cost function. Fit is also sensitive to some shocks' variances, namely, σ_b, σ_i, σ_p and σ_w.

When comparing *All* to *High-pass*, there are 3 features of the model that matter: the investment adjustment cost, η_k, habit persistence h and the Taylor rule. The shocks that matter the most are the investment specific shock and the premium shock.

When the *BC* parameters are used, results are similar to those of the *High-pass* case. Fit is sensitive to the investment adjustment cost, η_k, to habit persistence h and to the interest rate smoothing parameter, ρ_s. The exogenous shocks to which fit is most sensitive are the investment specific shock and the premium shock.

Summarizing, the model's feature which is sensitive to the frequency bands is the investment adjustment cost and the associated investment-specific shock. When there are no low frequencies in the estimation, the investment cost is relatively low, at 0.65. When low frequencies enter the estimation, the investment costs jumps up to 5.28, in order to slow down the response of investment to changes in Tobin's q and generate more persistent dynamics. The same is true, albeit to a lesser extent, for habit persistence: when low frequencies enter the estimation, habit persistence is 0.8; when there are no low frequencies, that number is reduced to 0.7.

5 Forecasting

Let us now turn to the analysis of the forecasting performance of DSGE models. In this pseudo-out-of-sample forecasting exercise, we compute forecasts using the standard approach based on the state-space representation of the model. The objective is to evaluate if the forecasting performance is affected when parameter estimates obtained on subsets of frequencies are used in forecasting actual data. We employ two forecasting models.

The first is the actual state-space model:

$$y_t = Z(\hat{\theta}_m)x_t \tag{24}$$

$$x_t = A(\hat{\theta}_m)x_{t-1} + B(\hat{\theta}_m)u_t^{\hat{\theta}_m} \tag{25}$$

where $V(u_t^{\hat{\theta}_m}) = \Sigma(\hat{\theta}_m)$ and the index m denotes one of the four estimates obtained above.

The second model, considered as a benchmark, is a Bayesian VAR (BVAR henceforth). In selecting the priors, we combine two prior specifications. The first is the Minnesota prior (Litterman, 1996) based on the extension to a Normal-inverted Wishart, as in Kadiyala and Karlssson, 1997. The second is the sum-of-coefficient prior (Sims and Zha, 1998) in which the VAR coefficients are restricted to sum to zero. These priors have been successfully used in the literature on the forecasting performance of Bayesian VARs (Banbura *et al.*, 2010). Appendix C discusses in greater detail the BVAR specification. The BVAR is estimated with variables in levels and on the same sample over which DSGE models are estimated.

The pseudo-out-of-sample experiment is as follows.

We put ourselves at time $T = 2002Q2$, the end of the estimation period. For each of the four estimated DSGE models (*All*, *Low-pass*, *BC*, *High-pass*), we follow the *sampling the future* procedure described in Christoffel *et al.*, 2011, to construct predictive densities:

1. Draw a parameter vector θ_j from the posterior distribution;
2. Draw the state variables at time T from $x_T \sim N(x_{T|T}, P_{T|T})$, where $x_{T|T}$ is the estimate of x_T and $P_{T|T}$ is the covariance matrix of x_T, given θ_j, obtained from the Kalman filter;
3. Simulate a path for the state variables from equation (25) using the drawn value for x_T as the initial value and a sequence of structural shocks $u_{T+1}, ..., u_{T+h}$ drawn from $N(0, \Sigma(\theta_j))$;
4. Compute a path for the variables $y_{T+1}, ..., y_{T+h}$ using the measurement equation (24);
5. Repeat steps 2-4 M_1 times for a given θ_j. Repeat steps 1-5 M_2 times.

We set $M_1 = 10$ and $M_2 = 100$ and report median forecasts $h-$steps ahead $(h = 1, 2, 3, 4, 8, 12)$.

For the BVAR in levels, we draw from the posterior distribution, forecast the levels, transform the forecasts to obtain forecasts for the variables in growth rates (output, consumption, investment and the real wage) and compute median forecasts $h-$steps ahead $(h = 1, 2, 3, 4, 8, 12)$. We then add one observation and repeat the same procedure. We do not re-estimate models as we add observations.

We use 34 observations as a forecasting sample. We therefore end up with 34 one-step ahead forecasts, 33 two-steps ahead forecasts and so on, up to 23 twelve-steps ahead forecasts.

We use two measures of forecasting performance: the RMSFE (Root Mean Squared Forecast Error) for each variable and for each forecasting horizon h, and the log-determinant of the covariance matrix of forecast errors for different h, $\log|\Omega(h)|$, proposed by Doan *et al.*, 1984, where the covariance matrix of forecast errors is:

$$\Omega(h) = N_h^{-1} \sum_{t=T}^{T+N_h-1} e_{t+h|t} e'_{t+h|t} \tag{26}$$

$e_{t+h|t}$ is the h-step-ahead forecast error vector from the forecast produced at time t and N_h is the number of evaluated h-step-ahead forecasts. The eigenvectors of the forecast error covariance matrix generate linear combinations of the variables with uncorrelated forecast errors. The determinant equals the product of the eigenvalues and thereby measures the product of the forecast error variances associated with these linear combinations. The more linear combinations exist that can be predicted with small forecast error variance, the smaller the log-determinant statistic (Del Negro and Schorfheide, 2012).

5.1 Discussion

Figure 5 shows the ratio between the RMSFE of the BVAR and the RMSFE of alternative models for each forecasting horizon. Numbers above (below) one mean that the BVAR delivers a smaller (larger) RMSFE than the alternative model. Figure 6 shows the log-determinant of the covariance matrix of forecast errors for each forecasting model and for each forecasting horizon. Let us start from the RMSFE plots. There are large differences depending on the variable to be forecast and on the forecasting horizon. Concerning consumption growth, we see that all DSGE models beat the BVAR by a large amount. Within DSGEs, the best models at short forecasting horizon are *All* and *Low-pass*. As the forecasting horizon becomes longer, differences among DSGEs tend to disappear. For output and investment growth, the BVAR is the best forecasting horizon, while among DSGEs, *Low-pass* and *BC* are the best forecasting models at all horizons in the short run. As before, at long forecasting horizons, different DSGEs converge to a similar performance and are similar to the BVAR model at horizon 8 and 12. For wage growth, *High-pass* and *All* are now the best DSGE models in the short-run. The BVAR is nevertheless superior at all forecast horizon. Concerning hours, the BVAR is the best model up to 8 quarters ahead. At long horizons, DSGEs are very similar among themselves and deliver smaller RMSFE than the BVAR. Among DSGEs, at short horizons *Low-pass* and *BC* beat *All* and especially *High-pass*, that performs poorly. For nominal variables, inflation and interest rate, DSGE models outperform the BVAR at 1 and 2 steps ahead. For horizons 3, 4 and 8 the BVAR is the best model, while at $h = 12$ DSGEs and BVAR have similar performances. In general, few conclusions can be drawn. At longer forecasting horizons, all DSGEs deliver a similar performance. This is not surprising: as forecasts tend to converge to the unconditional mean, differences among models tend to disappear. The BVAR is not the best forecasting models at all horizons and for all variables. DSGE models have a good performance especially when forecasting 8 and 12 steps ahead and, at those horizons, tend to behave similarly to the BVAR for all variables. Among DSGEs, *Low-pass* generally has a very good performance when forecasting at short horizons, with the exception of the real wage. Model *High-pass* does the opposite. It does badly on all variables, with the exception of the real wage. The multivariate statistic offer similar results. At short horizons the BVAR forecasts better than DSGEs, while, at longer horizons, the advantage of the BVAR is reduced; at horizon 12, DSGEs are the best forecasting models. At short horizons, *All* and *Low-pass* are the best models, while at longer horizons, *BC* and especially *High-pass* are the best models.

6 Conclusion

In this paper we use frequency domain techniques to estimate a medium-scale DSGE model on different frequency bands. We show that goodness of fit, forecasting performance and parameter estimates vary substantially with the frequency bands over which the model is estimated.

We point to two main results. First, the information contained in low frequencies (denoted here as frequencies whose period is longer than eight years) drive the estimates obtained in the time domain (or equivalently, over all frequencies). Second, when the estimation is performed over different frequency bands, parameter estimates and model fit are significantly affected. More specifically, when low frequencies are not present in the estimation, model fit and parameter estimates change significantly.

When we analyze parameter estimates, we note that the friction that is more sensitive to frequency bands is the investment cost friction: when low frequencies are present, the investment friction is estimated to be high, slowing down the response of investment to Tobin's q, and generating persistent dynamics to match the low frequency component of real variables, such as output and investment growth; when the model is estimated over higher frequencies, this friction becomes less relevant. The same is true for habit persistence: when low frequencies are not present, habit persistence significantly decreases.

If there is no misspecification, parameter estimates should not depend (asymptotically) on frequency bands. However, when low frequencies are present, estimated parameters are different from estimated parameters when low frequencies are absent: a signal that the model cannot match all frequencies with one set of parameters. Interestingly, when low frequencies are eliminated, parameters associated to frictions that generate slow moving fluctuations become less important. We conjecture that many of the mechanisms needed in DSGE models to match the data turn out to be significant simply because models are estimated over low frequencies.

Our results show that estimating a DSGE model in the time domain (or equivalently, over all frequencies) has pros and cons. If the model is designed for business cycle analysis, and if estimates over frequency bands are different, then it is not clear whether it is reasonable to estimate it over all frequencies.

Our results also have important implications in understanding the outcomes of policy experiments conducted with DSGE models and shed a negative light on the assumed "structuralness" of parameters. If the estimated model is used to compute the welfare costs of business cycles, as in Lucas, 1987, or to evaluate the effect of a new tax on the economy, potentially different answers can be obtained, depending on the frequency used in estimating parameters.

We believe that the approach of this paper could be employed as a useful diagnostic check for estimated models. The estimation of models on all frequencies and on subsets of frequencies gives insights on the role of frictions across frequency bands and this could be employed to build improved models. We see this as an interesting topic for future research.

Finally, when the forecasting performance is analyzed, we find that there are some gains from using models estimated over subsets of frequencies, but that forecasting performance depends on the variable to be forecasted and on the forecast horizon. In general, DSGEs are competitive with BVARs when forecasting 12 steps ahead. Differences among DSGEs estimated over different frequencies tend to vanish when the forecasting horizon becomes longer, as they all converge to the unconditional mean.

Bibliography

Adolfson M, Lindé J, Villani M. 2007. Forecasting performance of an open economy DSGE model. *Econometric Reviews* **26**: 289-328.

Altig D, Christiano L, Eichenbaum M, Lindé J. 2011. Firm-specific capital, nominal rigidities and the business cycle. *Review of Economic Dynamics* **14**(2): 225-247.

Altug S. 1989. Time-to-build and aggregate fluctuations: some new evidence. *International Economic Review* **30**(4): 889-920.

An S, Schorfheide F. 2007. Bayesian analysis of DSGE models. *Econometric Reviews* **26**(2-4): 113-172.

Banbura M, Giannone D, Reichlin L. 2010. Large Bayesian vector auto regressions. *Journal of Applied Econometrics* **25**(1): 71-92.

Baxter M, King R. G. 1999. Measuring business cycles: approximate band-pass filters for economic time series. *Review of Economics and Statistics* **81**(4): 575-593.

Berkowitz J. 2001. Generalized spectral estimation of the consumption-based asset pricing model. *Journal of Econometrics* **104**: 269-288.

Calvo GA. 1983. Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics* **12**(3): 383-398.

Canova F. 2012. Bridging DSGE models and the raw data. *Barcelona Graduate School of Economics. Working Papers 635.*

Canova F, Sala L. 2009. Back to square one: identification issues in DSGE models. *Journal of Monetary Economics* **56**(4): 431-449.

Canova F, Ferroni F. 2011. Multiple filtering devices for the estimation of cyclical DSGE models. *Quantitative Economics* **2**(1): 73-98.

Christiano LJ, Fitzgerald TJ. 2003. The band pass filter. *International Economic Review* **44**(2): 435-465.

Christiano LJ, Vigfusson RJ. 2003. Maximum likelihood in the frequency domain: the importance of time-to-plan. *Journal of Monetary Economics* **50**(4): 789-815.

Christiano LJ, Eichenbaum M, Evans C. 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* **113**(1): 1-45.

Christoffel K, Coenen G, Warne A. 2011. Forecasting with DSGE models. In *Handbook of*

Economic Forecasting, Clements MP, Hendry DF (eds). Oxford University Press: Oxford.

Cogley T. 2001a. Estimating and testing rational expectations models when the trend specification is uncertain. *Journal of Economic Dynamics and Control* **25**(10): 1485-1525.

Cogley T. 2001b. A frequency decomposition of approximation errors in stochastic discount factor models. *International Economic Review* **42(2)**: 473-503.

Corana A, Marchesi M, Martini C and Ridella S. 1987. Minimizing multimodal functions of continuous variables with the 'simulated annealing algorithm'. *ACM Transactions on Mathematical Software* **13**: 262-280.

Del Negro M, Schorfheide F. 2012. DSGE model-based forecasting. *New York Fed Staff Reports, no. 554*.

Del Negro M, Schorfheide F, Smets F, Wouters F. 2007. On the fit and forecasting performance of New Keynesian models. *Journal of Business and Economic Statistics* **25**(2): 123-162.

Diebold FX, Ohanian LE, Berkowitz J. 1998. Dynamic equilibrium economies: a framework for comparing models and data. *Review of Economic Studies* **65**(3): 433-451.

Doan T. 1992. *RATS User's Manual, Version 4*. Evanston. Il.

Doan T, Litterman R, Sims C. 1984. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, **3**(1): 1-100.

Doorn, D. 2006. Consequences of Hodrick-Prescott filtering for parameter estimation in a structural model of inventory behavior. *Applied Economics* **38**(16): 1863-1876.

Erceg CJ, Henderson DW, Levin AT. 2000. Optimal monetary policy with staggered wage and price contracts. *Journal of Monetary Economics* **46**(2): 281-313.

Goffe WL, Ferrier GD, Rogers J. 1994. Global optimization of statistical functions with simulated annealing. *Journal of Econometrics* **60**(1-2): 65-99.

Granger CWJ. 1966. The typical spectral shape of an economic variable. *Econometrica* **34**(1): 150-161.

Guerron-Quintana PA. 2010. What you match does matter: the effects of data on DSGE estimation. *Journal of Applied Econometrics* **25**: 774-804.

Hannan EJ. 1970. *Multiple time series.* Wiley, John & Sons.

Hansen LP, Sargent TJ. 1993. Seasonality and approximation errors in rational expectations models. *Journal of Econometrics* **55**(1-2): 21-55.

Harvey AC. 1991. *Forecasting, structural time series models and the Kalman filter.* Cambridge

University Press: Cambridge, UK.

Hodrick RJ, Prescott EC. 1997. Postwar U.S. business cycles: an empirical investigation. *Journal of Money, Credit and Banking* **29**(1): 1-16.

Ireland PN. 2004. A method for taking models to the data. *Journal of Economic Dynamics and Control* **28**(6): 1205-1226.

Iskrev N. 2010. Local identification in DSGE models. *Journal of Monetary Economics* **57**(2): 189-202.

Justiniano A, Primiceri GE, Tambalotti A. 2010. Investment shocks and business cycles. *Journal of Monetary Economics* **57**(2): 132-145.

Kadiyala KR, Karlsson S. 1997. Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics* **12**(2): 99-132.

Kimball MS. 1995. The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit, and Banking* **27**(4): 1241-1277.

Koop G, Korobilis D. 2009. Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics* **3**(4): 267-358.

Kydland FE, Prescott EC. 1982. Time to build and aggregate fluctuations. *Econometrica* **50**(6): 1345-70.

Gorodnichenko Y, Ng S. 2010. Estimation of DSGE models when the data are persistent. *Journal of Monetary Economics* **57**(3): 325-340.

Lucas RE Jr. 1987. *Models of business cycles*. Basil Blackwell: New York.

Rabanal P, Rubio-Ramirez JF. 2005. Comparing New Keynesian models of the business cycle: a Bayesian approach. *Journal of Monetary Economics* **52**(6): 1151-1166.

Sala L, Soderstrom U, Trigari A. 2010. The output gap, the labor wedge, and the dynamic behavior of hours. *IGIER, Università Bocconi Working Paper 365.*

Sargent TJ. 1989. Two models of measurements and the investment accelerator. *Journal of Political Economy* **97**(2): 251-287.

Sims C, Zha T. 1998. Bayesian methods for dynamic multivariate models. *International Economic Review* **39**(4): 949-968.

Singleton KJ. 1988. Econometric issues in the analysis of equilibrium business cycle models. *Journal of Monetary Economics* **21**: 361-386.

Smets F, Wouters R. 2003. An estimated dynamic stochastic general equilibrium model of the

Euro area. *Journal of the European Economic Association* **1**(5): 1123-1175.

Smets F, Wouters R. 2007. Shocks and frictions in US business cycles: a Bayesian DSGE approach. *American Economic Review* **97**(3): 586-606.

Watson MW. 1993. Measures of fit for calibrated models. *Journal of Political Economy* **101**: 1011-1041.

Yun T. 1996. Nominal price rigidity, money supply endogeneity, and business cycles. *Journal of Monetary Economics* **37**(2-3): 345-370.

Model	Frequency bands			
All	*All*	*Low*	*BC*	*High*
$\Delta\log(Y)$	1.93	1.76	2.27	1.75
$\Delta\log(I)$	1.44	1.32	1.87	1.21
$\Delta\log(C)$	2.32	2.50	3.59	1.58
$\Delta\log(W)$	1.46	1.70	2.03	1.22
H	1.30	0.75	3.27	8.20
π	0.75	0.47	2.11	1.53
i	0.84	0.53	1.90	1.73
High-pass	*All*	*Low*	*BC*	*High*
$\Delta\log(Y)$	1.31	1.15	1.18	1.42
$\Delta\log(I)$	1.12	0.51	1.11	1.20
$\Delta\log(C)$	1.68	1.68	1.92	1.55
$\Delta\log(W)$	1.41	1.75	1.85	1.18
H	0.29	0.09	0.79	5.59
π	0.26	0.083	0.81	1.23
i	0.34	0.13	0.96	1.86
BC	*All*	*Low*	*BC*	*High*
$\Delta\log(Y)$	1.23	1.15	1.17	1.29
$\Delta\log(I)$	0.83	0.56	0.93	0.81
$\Delta\log(C)$	1.40	1.54	1.82	1.13
$\Delta\log(W)$	0.98	1.64	1.63	0.62
H	0.41	0.18	1.10	4.96
π	0.33	0.16	1.10	0.87
i	0.41	0.17	1.07	2.02
Low-pass	*All*	*Low*	*BC*	*High*
$\Delta\log(Y)$	1.68	1.89	1.93	1.46
$\Delta\log(I)$	0.99	1.29	1.28	0.79
$\Delta\log(C)$	1.66	2.36	2.71	0.88
$\Delta\log(W)$	1.08	1.80	1.92	0.63
H	1.44	1.03	2.95	5.98
π	0.77	0.52	2.30	0.91
i	0.94	0.61	2.10	1.95

Table 1: Model variance divided by the data variance. The ratio is computed for 7 observables (in rows) and for the 4 models (from top to bottom: *All*, *High-pass*, *BC*, *Low-pass*), over 4 non-overlapping frequency bands (in column, from left to right, *All*, *Low*, *BC*, *High*). Each entry is computed as: $\frac{\int_{\Omega_F} S_y^{ii}(\omega,\theta_M)d\omega}{\int_{\Omega_F} S_{y,VAR}^{ii}(\omega)d\omega}$, where Ω_F denotes the frequency band, $S_y(\omega,\theta_M)$ the model's M spectral density and $S_{y,VAR}(\omega)$ the median of the spectral density of the data in Figure 1 .

		Model				
Parameter	Prior	*Time*	*All*	*High-pass*	*BC*	*Low-pass*
ψ_z (elasticity of k utilization)	$Beta(0.5\ ,\ 0.1)$	0.72 [0.60;0.82]	0.73 [0.62;0.83]	0.83* [0.74;0.89]	0.78 [0.68;0.86]	0.72 [0.60;0.81]
η_k (2^{nd} derivative of I cost f.)	$\mathcal{N}(4\ ,\ 1.5)$	3.09 [2.09;4.35]	2.91 [1.96;4.13]	0.65** [0.40;1.09]	1.81* [0.93;3.67]	5.28** [3.78;7.13]
h (habit persistence)	$Beta(0.5\ ,\ 0.1)$	0.75 [0.67;0.82]	0.80 [0.74;0.85]	0.71** [0.63;0.78]	0.71** [0.63;0.78]	0.81 [0.75;0.86]
ω (inverse Frish elast.)	$\Gamma(2\ ,\ 0.75)$	2.93 [1.74;4.43]	3.02 [1.85;4.67]	2.99 [1.92;4.42]	2.56 [1.44;4.14]	2.45 [1.44;3.91]
γ_w (w indexation)	$Beta(0.5\ ,\ 0.15)$	0.69 [0.53;0.83]	0.69 [0.51;0.82]	0.38** [0.21;0.58]	0.31** [0.15;0.51]	0.58 [0.38;0.77]
γ_p (p indexation)	$Beta(0.5\ ,\ 0.15)$	0.26 [0.11;0.44]	0.21 [0.09;0.49]	0.15 [0.06;0.28]	0.25 [0.12;0.44]	0.43 [0.23;0.65]
θ_w (w stickiness)	$Beta(0.75\ ,\ 0.1)$	0.83 [0.74;0.90]	0.82 [0.74;0.89]	0.71* [0.60;0.81]	0.76 [0.63;0.87]	0.82 [0.73;0.90]
θ_p (p stickiness)	$Beta(0.66\ ,\ 0.1)$	0.79 [0.71;0.85]	0.80 [0.75;0.85]	0.85 [0.80;0.90]	0.80 [0.71;0.86]	0.78 [0.71;0.83]
ε_p (p markup)	$\mathcal{N}\ (1.15\ ,\ 0.05)$	1.36 [1.30;1.42]	1.36 [1.30;1.42]	1.33 [1.27;1.39]	1.33 [1.26;1.39]	1.33 [1.26;1.39]
ε_w (w markup)	$\mathcal{N}\ (1.15\ ,\ 0.05)$	1.14 [1.06;1.23]	1.14 [1.06;1.22]	1.16 [1.08;1.23]	1.15 [1.07;1.23]	1.14 [1.06;1.22]
ϕ_π (π in Taylor rule)	$\mathcal{N}\ (1.7\ ,\ 0.3)$	1.86 [1.56;2.22]	2.05 [1.68;2.43]	1.12** [0.72;1.58]	1.28** [0.91;1.70]	1.81 [1.48;2.18]
ϕ_y (Y gap in Taylor rule)	$\mathit{\Gamma}(0.125\ ,\ 0.1)$	0.31 [0.21;0.44]	0.31 [0.22;0.45]	0.62** [0.47;0.82]	0.34 [0.21;0.51]	0.19** [0.11;0.28]
ρ_s (i smoothing)	$Beta(0.75\ ,\ 0.1)$	0.81 [0.75;0.85]	0.80 [0.75;0.84]	0.60** [0.48;0.69]	0.53** [0.41;0.63]	0.76 [0.69;0.82]
ρ_z (technology)	$Beta(0.5\ ,\ 0.1)$	0.13 [0.07;0.22]	0.21 [0.14;0.29]	0.24 [0.16;0.33]	0.35** [0.25;0.45]	0.35** [0.25;0.46]
ρ_m (monetary)	$Beta(0.5\ ,\ 0.1)$	0.24 [0.15;0.35]	0.30 [0.22;0.40]	0.33 [0.22;0.50]	0.29 [0.18;0.42]	0.30 [0.19;0.41]
ρ_b (preference)	$Beta(0.5\ ,\ 0.1)$	0.74 [0.62;0.83]	0.67 [0.58;0.75]	0.38** [0.28;0.51]	0.51** [0.38;0.64]	0.77 [0.68;0.83]
ρ_i (I specific)	$Beta(0.5\ ,\ 0.1)$	0.50 [0.38;0.64]	0.49 [0.38;0.61]	0.50 [0.40;0.61]	0.46 [0.34;0.58]	0.48 [0.35;0.60]
ρ_p (p markup)	$Beta(0.5\ ,\ 0.1)$	0.66 [0.53;0.76]	0.65 [0.44;0.75]	0.44* [0.30;0.56]	0.57 [0.43;0.70]	0.69 [0.53;0.78]
ρ_w (w markup)	$Beta(0.5\ ,\ 0.1)$	0.29 [0.18;0.41]	0.33 [0.23;0.44]	0.28 [0.19;0.39]	0.45* [0.32;0.56]	0.53** [0.39;0.65]
ρ_g (G)	$Beta(0.5\ ,\ 0.1)$	0.96 [0.93;0.98]	0.93 [0.90;0.95]	0.87** [0.81;0.91]	0.84** [0.77;0.90]	0.91 [0.87;0.94]
σ_z (technology)	$I\Gamma(0.15\ ,\ 0.15)$	1.08 [0.99;1.18]	1.10 [1.01;1.21]	1.12 [1.02;1.23]	0.90** [0.79;1.03]	0.87** [0.77;1.00]
σ_m (monetary)	$I\Gamma(0.15\ ,\ 0.15)$	0.23 [0.21;0.26]	0.24 [0.22;0.27]	0.22* [0.19;0.24]	0.23 [0.20;0.27]	0.26 [0.23;0.31]
σ_b (preference)	$I\Gamma(0.15\ ,\ 0.15)$	0.44 [0.31;0.84]	0.71 [0.45;1.09]	0.42* [0.27;0.66]	0.30** [0.20;0.49]	0.46 [0.32;0.80]
σ_i (I specific)	$I\Gamma(0.15\ ,\ 0.15)$	0.22 [0.15;0.31]	0.21 [0.14;0.29]	0.07** [0.05;0.10]	0.12** [0.07;0.22]	0.28 [0.19;0.41]
σ_p (p markup)	$I\Gamma(0.15\ ,\ 0.15)$	0.09 [0.07;0.12]	0.10 [0.07;0.13]	0.13* [0.11;0.16]	0.08 [0.06;0.11]	0.06** [0.05;0.08]
σ_w (w markup)	$I\Gamma(0.15\ ,\ 0.15)$	0.22 [0.18;0.25]	0.21 [0.18;0.24]	0.21 [0.18;0.25]	0.13** [0.10;0.17]	0.12** [0.08;0.17]
σ_g (G)	$I\Gamma(0.15\ ,\ 0.15)$	0.36 [0.33;0.40]	0.36 [0.33;0.40]	0.35 [0.32;0.38]	0.40 [0.34;0.46]	0.41* [0.36;0.48]
Posterior value at median			1102.0	1162.8	426.4	389.6

Table 2: Parameter estimates: median. In brackets the 5% and 95% percentile of the posterior distribution. Median and posterior percentiles from one chain of 200,000 draws from a RW Metropolis algorithm. The initial 50,000 draws have been discarded. We retain 1 in every 5 subsequent draws.

	All/High-pass	*All/BC*	*All/Low-pass*
ψ_z (elasticity of k utilization)	0.15	0.04	0.00
η_k (2^{nd} derivative of I cost f.)	**11.3**	1.84	**6.17**
h (habit persistence)	3.87	4.17	0.00
ω (inverse Frish elast.)	0.00	0.03	0.03
γ_w (w indexation)	0.62	0.92	0.08
γ_p (p indexation)	0.13	0.00	0.51
θ_w (w stickiness)	0.95	0.33	0.00
θ_p (p stickiness)	0.29	0.01	0.01
ϵ_p (p markup)	0.01	0.03	0.03
ϵ_w (w markup)	0.01	0.01	0.00
ϕ_π (π in Taylor rule)	0.91	0.51	0.01
ϕ_y (Y gap in Taylor rule)	1.34	0.02	0.76
ρ_s (i smoothing)	5.42	**9.06**	0.38
ρ_z (technology)	0.02	0.48	0.48
ρ_m (monetary)	0.00	0.01	0.00
ρ_b (preference)	3.65	0.95	0.38
ρ_i (I specific)	0.00	0.02	0.00
ρ_p (p markup)	**6.37**	0.87	0.14
ρ_w (w markup)	0.11	0.50	1.46
ρ_g (G)	0.50	0.89	0.04
σ_z (technology)	0.01	0.71	0.99
σ_m (monetary)	0.19	0.02	0.14
σ_b (preference)	**5.46**	**19.6**	3.50
σ_i (I specific)	**52.1**	**6.59**	1.27
σ_p (p markup)	1.55	0.18	3.79
σ_w (w markup)	0.03	4.83	**7.04**
σ_g (G)	0.02	0.14	0.29
Overall	39.0	27.8	11.0

Table 3: Percentage change in the posterior when the corresponding parameter is varied from the median value estimated in model *All* to the median value estimated in model *High-pass* (column 2), *BC* (column 3) and *Low-pass* (column 4).

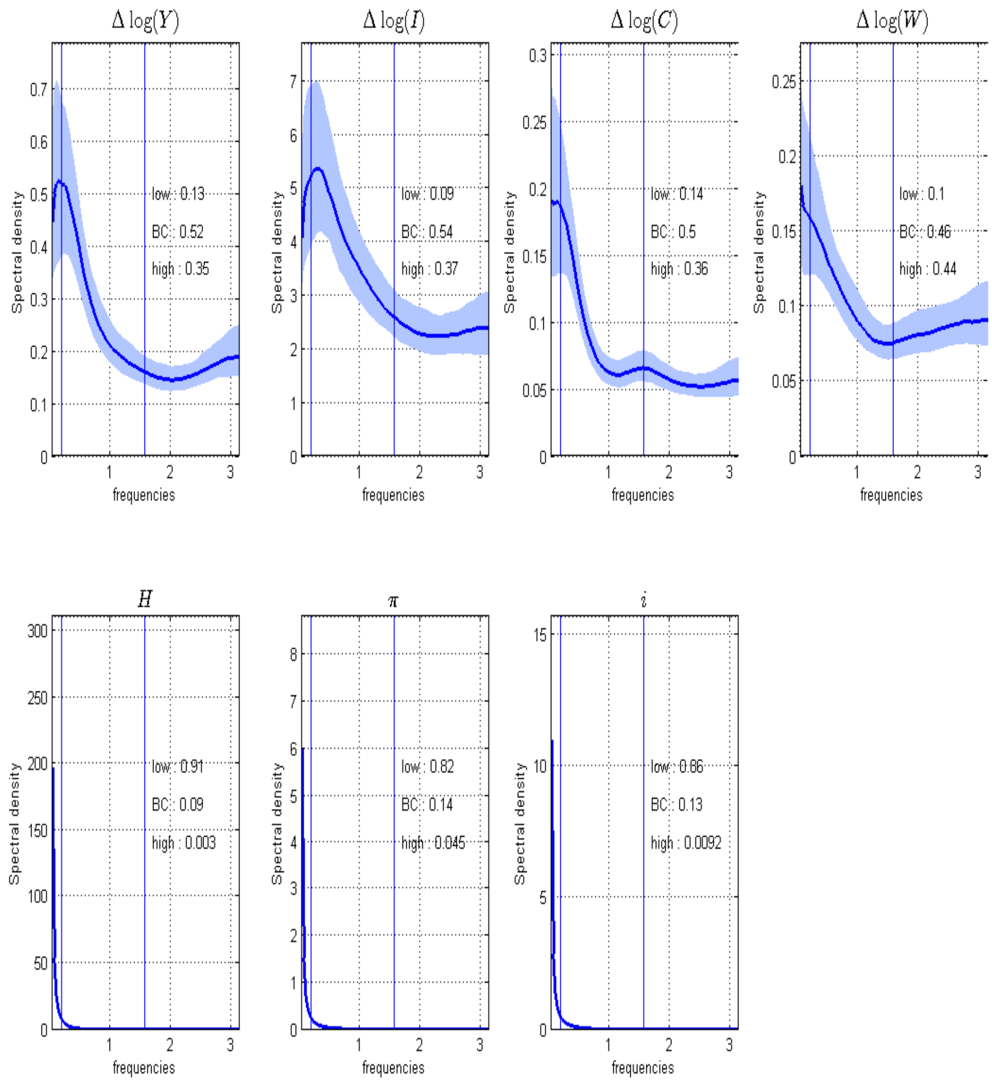

Figure 1: Data spectrum, BVAR estimates. Numbers in the figure indicate the fraction of variance located at different frequency bands - *low* denotes cycles with period $32 < per < \infty$ quarters; *BC* denotes cycles with period $4 < per < 32$ quarters; *high* denotes cycles with period $2 < per < 4$. Vertical bars separate the frequency domain in the three regions *low*, *BC* and *high*. Bold line: posterior median of the BVAR estimates. Shaded areas: 68% credible sets. For details, see Appendix C and footnote 6.

.

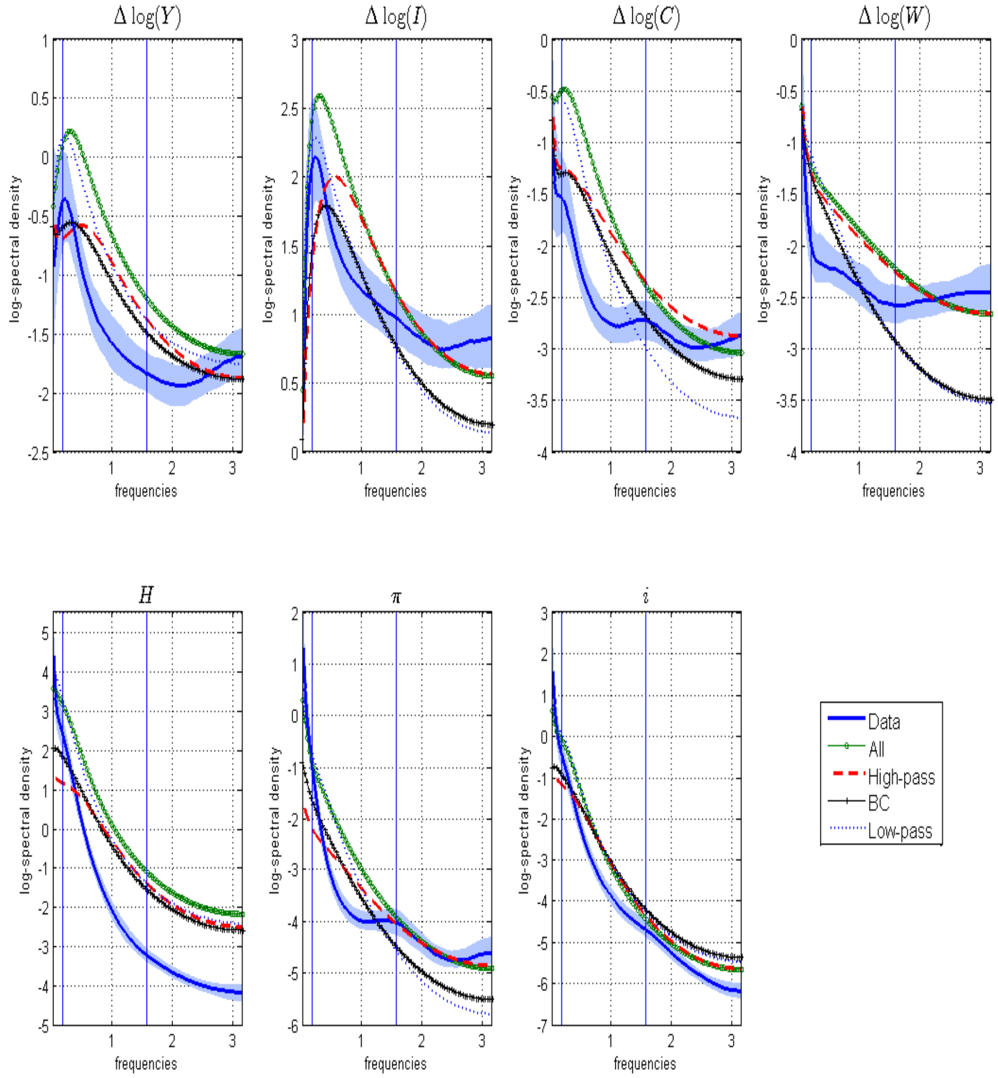

Figure 2: Log-spectra of data and models estimated on different frequency bands. Vertical bars separate the frequency domain in the three regions *low*, *BC* and *high*. *low* denotes cycles with period $32 < per < \infty$ quarters; BC denotes cycles with period $4 < per < 32$ quarters; *high* denotes cycles with period $2 < per < 4$. For details, see Appendix C and footnote 6.

.

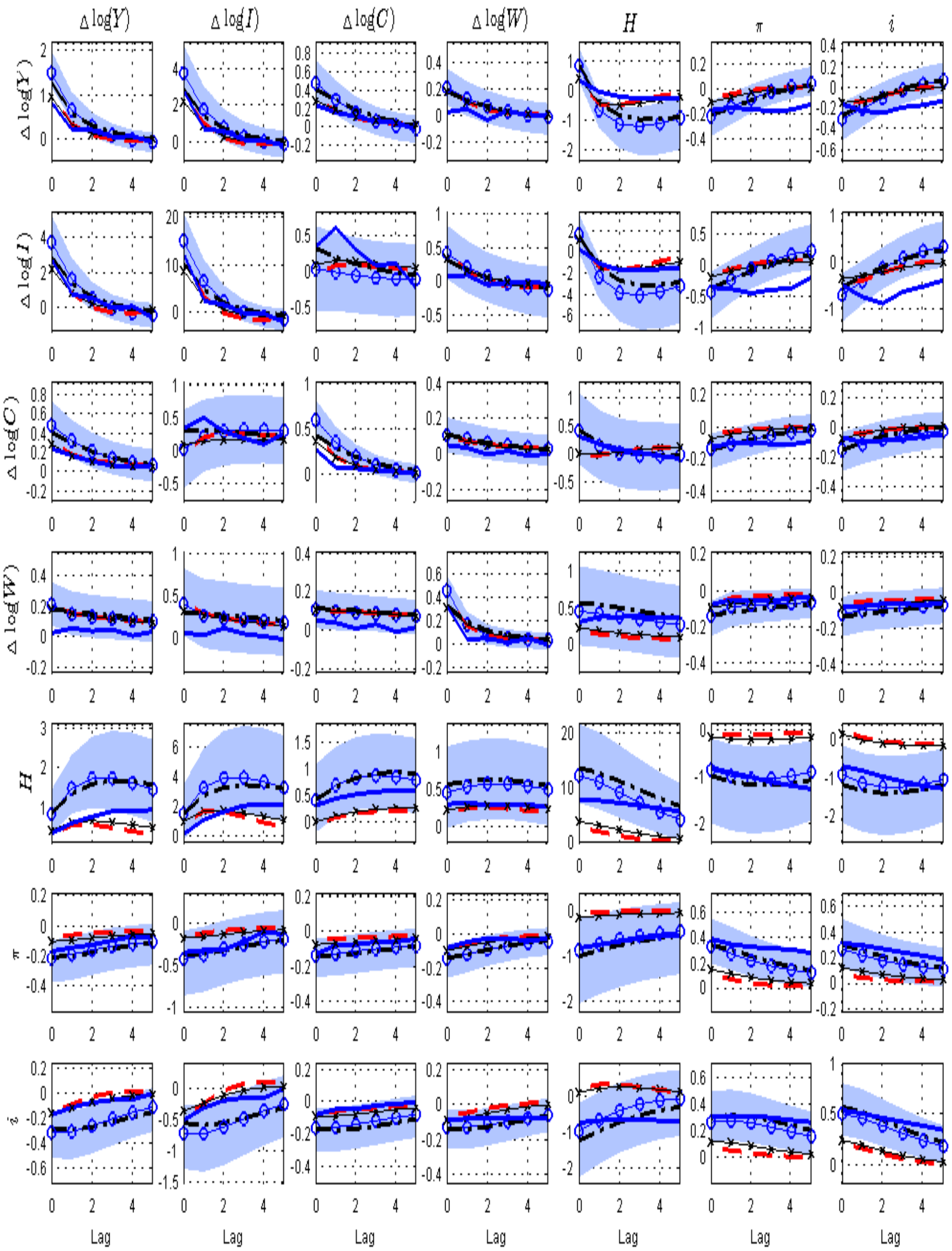

Figure 3: Autocovariance function of models estimated on different frequency bands. Bold: data; Circles: *All*; Dashed: *High-pass*; Starred: *BC*; Dotted: *Low-pass*. Shaded areas are 90% Bayesian credible sets for model *All*. The posterior distribution is obtained by drawing 100 parameters from the posterior distribution, simulating 100 samples of 170 observations (to match the dimension of the sample) for each draw and computing the autocovariance function for each sample.

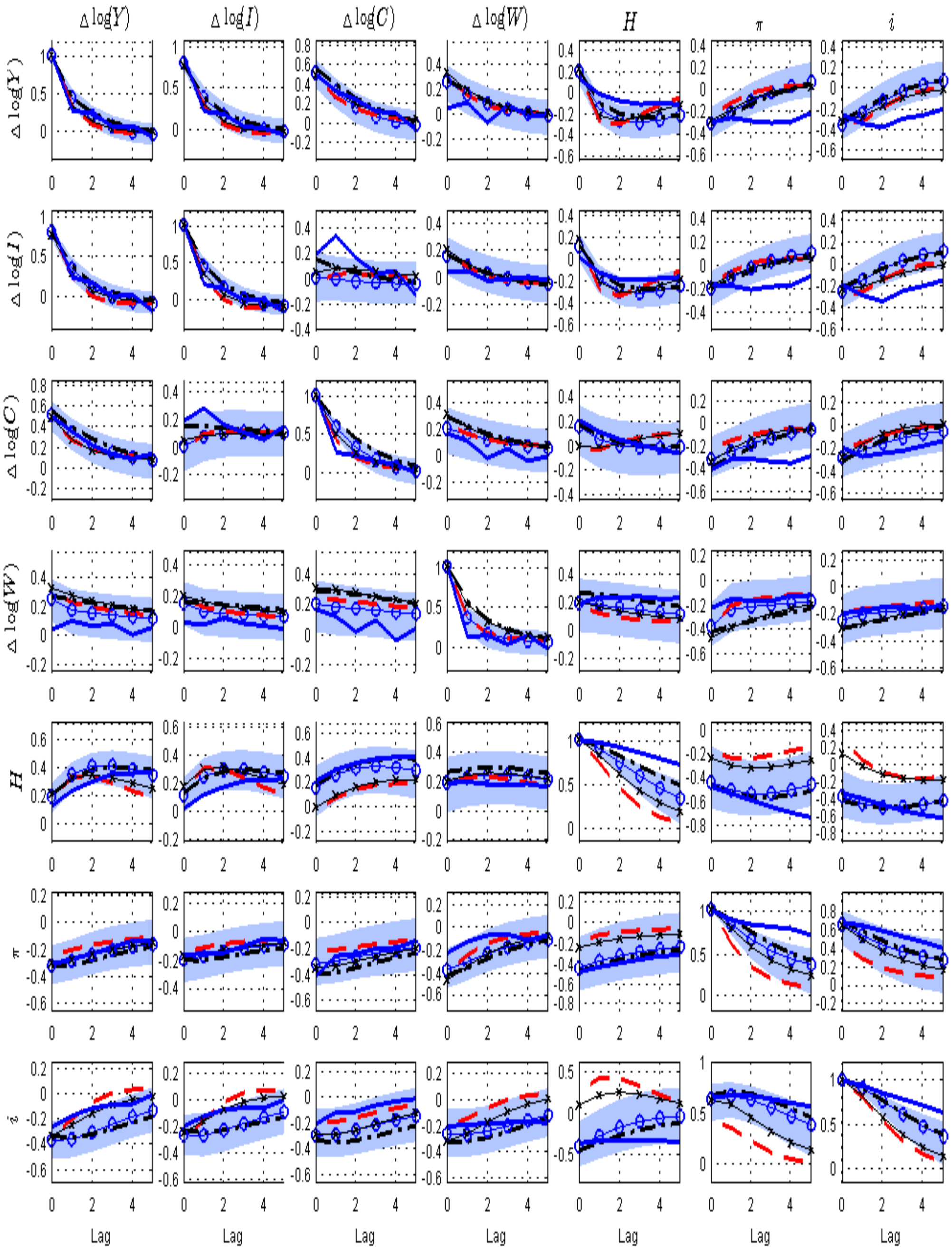

Figure 4: Autocorrelation function of models estimated on different frequency bands. Bold: data; Circles: *All*; Dashed: *High-pass*; Starred: *BC*; Dotted: *Low-pass*. Shaded areas are 90% Bayesian credible sets for model *All*. The posterior distribution is obtained by drawing 100 parameters from the posterior distribution, simulating 100 samples of 170 observations (to match the dimension of the sample) for each draw and computing the autocorrelation function for each sample.

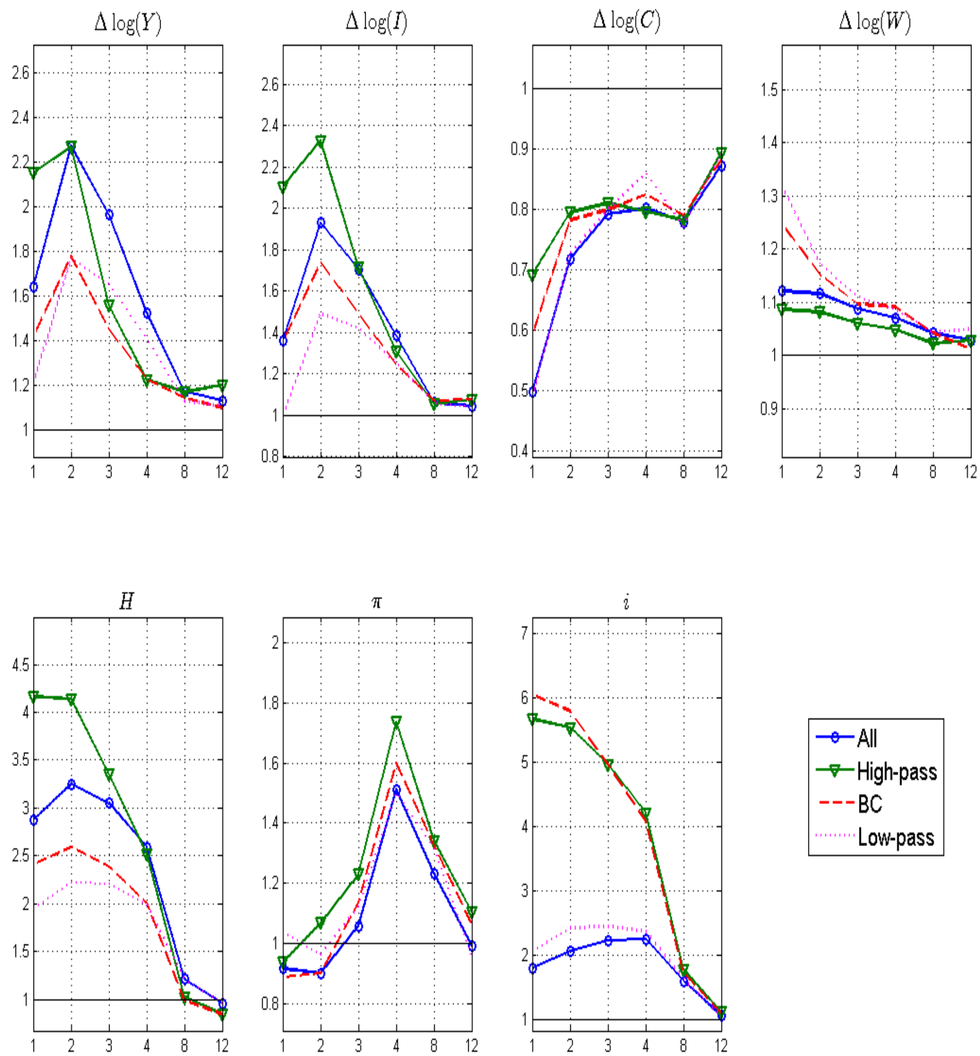

Figure 5: Mean Squared Forecast Error (MSFE). Ratio with regard to the MSFE obtained with a BVAR. The horizontal axis denotes the forecasting horizon. Numbers above one indicate that BVAR delivers a smaller MSFE.

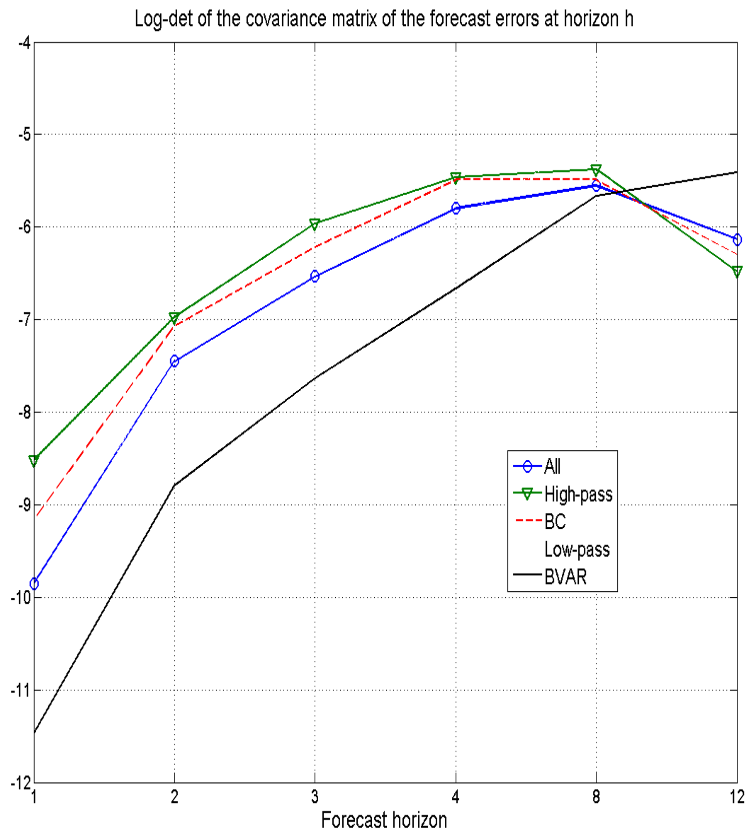

Figure 6: Log-determinant of the covariance matrix of the forecast errors. Horizontal axis denotes the forecasting horizon. Good forecasting models are characterized by a smaller value of the log-determinant.

A Appendix: the DSGE model

We present here the DSGE model in greater detail. The presentation follows closely Sala *et al.*, 2011.

A.1 Households

The economy features a continuum of households, indexed by $j \in [0,1]$. Each household consumes final goods, supplies a specific type of labor to intermediate goods firms via employment agencies, saves in one-period nominal government bonds, and accumulates physical capital through investment. It transforms physical capital to effective capital by choosing the capital utilization rate, and then rents the effective capital to intermediate goods firms.

Household j chooses consumption $C_t(j)$, labor supply $H_t(j)$, bond holdings $B_t(j)$, the rate of capital utilization ν_t, investment I_t, and physical capital $\bar{K}_t$ to maximize the intertemporal utility function

$$E_t \left\{ \sum_{s=0}^{\infty} \beta^s \epsilon_{t+s}^b \left[\log \left(C_{t+s}(j) - hC_{t+s-1}(j) \right) - \frac{H_{t+s}(j)^{1+\omega}}{1+\omega} \right] \right\} \tag{1}$$

where β is a discount factor, h measures the degree of habits in consumption, ω is the inverse Frisch elasticity of labor supply, ϵ_t^b is an intertemporal preference shock, and ϵ_t^b is a shock to the disutility of supplying labor. The intertemporal preference shock has mean unity and is assumed to follow the autoregressive process

$$\log \epsilon_t^b = \rho_b \log \epsilon_{t-1}^b + \zeta_t^b, \qquad \zeta_t^b \sim i.i.d.\ N(0, \sigma_b^2). \tag{2}$$

The capital utilization rate ν_t transforms physical capital $\bar{K}_t$ into efficient capital K_t according to

$$K_t = \nu_t \bar{K}_{t-1}, \tag{3}$$

and the efficient capital is rented to intermediate goods firms at the nominal rental rate R_t^k. The cost of capital utilization per unit of physical capital is given by $\mathcal{A}(\nu_t)$, and we assume that $\nu_t = 1$ in steady state, $\mathcal{A}(1) = 0$, and $\mathcal{A}'(1)/\mathcal{A}''(1) = \eta_\nu$, as in Christiano, Eichenbaum and Evans, 2005 and others.

Physical capital accumulates according to

$$\bar{K}_t = (1-\delta)\bar{K}_{t-1} + \epsilon_t^i \left[1 - \mathcal{S}\left(\frac{I_t}{I_{t-1}}\right)\right] I_t, \tag{4}$$

where δ is the depreciation rate of capital, ϵ_t^i is an investment-specific technology shock with mean unity, and $\mathcal{S}(\cdot)$ is an adjustment cost function which satisfies $\mathcal{S}(\gamma_z) = \mathcal{S}'(\gamma_z) = 0$ and $\mathcal{S}''(\gamma_z) = \eta_k > 0$, where γ_z is the steady-state growth rate. The investment-specific technology shock follows the process

$$\log \epsilon_t^i = \rho_i \log \epsilon_{t-1}^i + \zeta_t^i, \qquad \zeta_t^i \sim i.i.d.\ N(0, \sigma_i^2). \tag{5}$$

Let P_t be the nominal price level, R_t the one-period nominal (gross) interest rate, $A_t(j)$ the net returns from a portfolio of state-contingent securities, W_t the nominal wage, Π_t nominal lump-sum profits from ownership of firms, and T_t nominal lump-sum transfers. Household j's budget constraint is then given by

$$P_tC_t + P_tI_t + B_t = T_t + R_{t-1}B_{t-1} + A_t(j) + \Pi_t + W_t(j)H_t(j) + r_t^k \nu_t \bar{K}_{t-1} - P_t\mathcal{A}(\nu_t)\bar{K}_{t-1}. \tag{6}$$

Assuming that households have access to a complete set of state-contingent securities, consumption and asset holdings are the same for all households.

A.2 Final goods-producing firms

A perfectly competitive sector combines a continuum of intermediate goods $Y_t(i)$ indexed by $i \in [0,1]$ into a final consumption good Y_t.

Following Smets and Wouters, 2007, we assume that each firm's elasticity depends inversely on its relative market share, as in Kimball, 1995, who generalizes the standard Dixit-Stiglitz aggregator. Thus, letting $Y_t(i)$ be the quantity of output sold by retailer i and $P_t(i)$ the nominal price, final goods, denoted Y_t, are a composite of individual retail goods following

$$\int_0^1 \mathcal{G}\left(\frac{Y_t(i)}{Y_t}, \epsilon_t^p\right) di = 1, \tag{7}$$

where the function $\mathcal{G}(\cdot)$ is increasing and strictly concave with $\mathcal{G}(1) = 1$, and ϵ_t^p is a time-varying measure of substitutability across differentiated intermediate goods. This substitutability implies a time-varying (gross) markup of price over marginal cost equal to ϵ_t^p that is assumed to follow the process

$$\log \epsilon_t^p = (1-\rho_p)\log \epsilon^p + \rho_p \log \epsilon_{t-1}^p + \zeta_t^p, \qquad \zeta_t^p \sim i.i.d.\ N(0, \sigma_p^2), \tag{8}$$

where ϵ^p is the steady-state price markup.

A.3 Intermediate goods producing firms

Each firm in the intermediate goods sector produces a differentiated intermediate good i using capital and labor inputs according to the production function

$$Y_t(i) = \max\left\{K_t(i)^\alpha \left[Z_t H_t(i)\right]^{1-\alpha} - Z_t F, 0\right\}, \tag{9}$$

where α is the capital share, Z_t is a labor-augmenting productivity factor, whose growth rate $\epsilon_t^z = Z_t/Z_{t-1}$ follows a stationary exogenous process with steady-state value ϵ^z which corresponds to the economy's steady-state (gross) growth rate γ_z, and F is a fixed cost that ensures that profits are zero. The rate of technology growth is assumed to follow

$$\log \epsilon_t^z = (1-\rho_z)\log \epsilon^z + \rho_z \log \epsilon_{t-1}^z + \zeta_t^z, \qquad \zeta_t^z \sim i.i.d.\ N(0, \sigma_z^2). \tag{10}$$

Thus, technology is non-stationary in levels but stationary in growth rates, following Altig et al., 2005. We assume that capital is perfectly mobile across firms and that there is a competitive rental market for capital.

Prices of intermediate goods are set in a staggered fashion, following Calvo, 1983. Thus, only a fraction $1-\theta_p$ of firms are able to reoptimize their price in any given period.

Firms that do not re-optimize instead index their price to a combination of past inflation and steady-state inflation according to the rule

$$P_t(i) = \overline{\gamma}_p \pi_{t-1}^{\gamma_p} P_{t-1}(i), \tag{11}$$

where $\overline{\gamma}_p = \pi^{1-\gamma_p}$ is an adjustment for steady-state inflation.

A.4 The labor market

As in Erceg *et al.*, 2000, each household is a monopolistic supplier of specialized labor $H_t(j)$, which is combined by perfectly competitive employment agencies into labor services H_t according to

$$H_t = \left[\int_0^1 H_t(j)^{1/\epsilon_t^w} dj \right]^{\epsilon_t^w}, \tag{12}$$

where ϵ_t^w is a time-varying measure of substitutability across labor varieties that translates into a time-varying (gross) markup of wages over the marginal rate of substitution between consumption and leisure. The wage markup shock is assumed to follow

$$\log \epsilon_t^w = (1-\rho_w) \log \epsilon^w + \rho_w \log \epsilon_{t-1}^w + \zeta_t^w, \qquad \zeta_t^w \sim i.i.d.\ N(0, \sigma_w^2). \tag{13}$$

where ϵ^w is the steady-state wage markup.

In any given period, a fraction $1-\theta_w$ of households are able to set their wage optimally. Similar to the price indexation scheme, the remaining fraction indexes their wage to the steady-state growth rate γ_z and a combination of past inflation and steady-state inflation according to

$$W_t(j) = W_{t-1}(j) \gamma_z \pi_{t-1}^{\gamma_w} \pi^{1-\gamma_w}. \tag{14}$$

A.5 Government

The government sets public spending G_t according to

$$G_t = \left[1 - \frac{1}{\epsilon_t^g} \right] Y_t, \tag{15}$$

where ϵ_t^g is a spending shock with mean unity that follows the process

$$\log \epsilon_t^g = \rho_g \log \epsilon_{t-1}^g + \zeta_t^g, \qquad \zeta_t^g \sim i.i.d.\ N(0, \sigma_g^2). \tag{16}$$

The nominal interest rate i_t is set using the monetary policy rule[1]

[1]The monetary policy rule is specified in terms of output gap Y_t/Y_t^{flex}, defined as the deviation of output from the level under flexible prices and wages.

$$\frac{i_t}{i} = \left(\frac{i_{t-1}}{i}\right)^{\rho_s} \left[\left(\frac{\pi_t}{\pi}\right)^{\phi_\pi} \left(Y_t / Y_t^{flex}\right)^{\phi_y}\right]^{1-\rho_s} \epsilon_t^m, \tag{17}$$

where π is the steady-state level of inflation and ϵ_t^m is an AR(1) monetary policy shock which follows:

$$\log \epsilon_t^m = \rho_m \log \epsilon_{t-1}^m + \zeta_t^m, \qquad \zeta_t^m \sim i.i.d.\ N(0, \sigma_m^2). \tag{18}$$

A.6 Market clearing

Finally, to close the model, the resource constraint implies that output is equal to the sum of consumption, investment, government spending, and the capital utilization costs:

$$Y_t = C_t + I_t + G_t + \mathcal{A}(\nu_t) \bar{K}_{t-1}. \tag{19}$$

To find the steady state, the model is expressed in stationary form. Thus, for the non-stationary variables, let lower-case letters denote their value relative to the technology process Z_t:

$$y_t \equiv Y_t/Z_t, \quad k_t \equiv K_t/Z_t, \quad \bar{k}_t \equiv \bar{K}_t/Z_t, \quad i_t \equiv I_t/Z_t, \quad c_t \equiv C_t/Z_t,$$

$$g_t \equiv G_t/Z_t, \quad \lambda_t \equiv \Lambda_t Z_t, \quad w_t \equiv W_t/(Z_t P_t), \quad w_t^* \equiv W_t^*/(Z_t P_t),$$

Note that the marginal utility of consumption Λ_t will shrink as the economy grows. The wage is expressed in real terms. Also, the real rental rate of capital and real marginal cost are expressed as:

$$r_t^k \equiv R_t^k/P_t, \quad mc_t \equiv MC_t/P_t,$$

and the optimal relative price as

$$p_t^* \equiv P_t^*/P_t.$$

The stationary model is then log-linearized around the steady state.

B Appendix: Monte Carlo experiments

In this Appendix, we study the small sample performance of the frequency domain maximum likelihood estimator by running a number of Monte Carlo experiments. First, we investigate the small sample performance of the frequency domain estimator over all frequencies with regard to the time domain counterpart. Second, we study the small sample performance of the maximum likelihood over frequency bands.

Our Monte Carlo is conducted as follows. We generate 100 artificial samples of length 170 from the state-space representation of our DSGE model, parameterized with the median parameters estimated in *All*[2]. The length of the artificial samples is chosen in order to mimic the number of actual observations. For each artificial sample, we estimate parameters using the four different sets of frequencies discussed in the main text, *All*, *High-pass*, *Low-pass* and *BC*. We also estimate parameters in the time domain, evaluating the log-likelihood with the Kalman filter. Table 1 reports median estimates over the 100 artificial samples, with standard deviations in brackets. We first compare estimation over all frequencies *All* (in Column 3) with estimation in the time domain (in Column 4). Theory tells us that results obtained with these two methods should be equivalent, at least asymptotically. By comparing Columns 3 and 4 we see that results are very similar. In this case the frequency domain likelihood is a good approximation of the time domain likelihood. As maximum likelihood on a frequency band is similar to an estimation on a subsample, we evaluate the loss in efficiency that we incur when estimating on a subset of frequencies. Columns 5 to 7 show that the standard deviations of the estimates increase somewhat with regard to the *All* case, but that median estimates are still very close to the true value.

In sum, the evidence shows that, when using the DSGE model as data-generating process, maximum likelihood in the frequency domain is equivalent to maximum likelihood in the time domain, and that the precision of the estimates is still very good when estimation is performed on frequency bands.

[2]In the Monte Carlo we have calibrated the inverse of the Frish elasticity ω as it is very weakly identified.

Parameter	True	*Time*	*All*	*High-pass*	*BC*	*Low-pass*
		Model				
ψ_z (elasticity of k utilization)	0.73	0.73 [0.07]	0.74 [0.08]	0.77 [0.08]	0.76 [0.09]	0.73 [0.08]
η_k (2^{nd} derivative of I cost f.)	2.91	3.53 [1.41]	3.16 [1.56]	3.41 [1.74]	3.11 [2.09]	2.98 [1.83]
h (habit persistence)	0.80	0.81 [0.04]	0.81 [0.04]	0.81 [0.05]	0.82 [0.06]	0.81 [0.06]
γ_w (w indexation)	0.69	0.63 [0.14]	0.63 [0.14]	0.66 [0.16]	0.66 [0.18]	0.63 [0.15]
γ_p (p indexation)	0.21	0.21 [0.19]	0.21 [0.19]	0.21 [0.17]	0.24 [0.23]	0.24 [0.22]
θ_w (w stickiness)	0.82	0.87 [0.12]	0.88 [0.08]	0.86 [0.11]	0.88 [0.12]	0.87 [0.10]
θ_p (p stickiness)	0.80	0.81 [0.04]	0.81 [0.05]	0.81 [0.05]	0.81 [0.06]	0.82 [0.06]
ε_p (p markup)	1.36	1.36 [0.06]	1.35 [0.06]	1.34 [0.06]	1.35 [0.06]	1.36 [0.05]
ε_w (w markup)	1.14	1.39 [0.03]	1.37 [0.25]	1.30 [0.30]	1.33 [0.30]	1.48 [0.29]
ϕ_π (π in Taylor rule)	2.05	2.12 [0.91]	2.02 [0.82]	2.31 [0.99]	2.13 [0.97]	1.95 [0.84]
ϕ_y (Y gap in Taylor rule)	0.31	0.34 [0.22]	0.30 [0.21]	0.35 [0.24]	0.33 [0.24]	0.30 [0.20]
ρ_s (i smoothing)	0.80	0.81 [0.07]	0.79 [0.07]	0.80 [0.08]	0.79 [0.09]	0.79 [0.08]
ρ_z (technology)	0.21	0.21 [0.08]	0.18 [0.08]	0.19 [0.08]	0.21 [0.14]	0.19 [0.13]
ρ_m (monetary)	0.3	0.31 [0.09]	0.30 [0.10]	0.28 [0.1]	0.29 [0.20]	0.28 [0.18]
ρ_b (preference)	0.67	0.69 [0.09]	0.68 [0.10]	0.64 [0.14]	0.62 [0.15]	0.69 [0.13]
ρ_i (I specific)	0.49	0.46 [0.08]	0.47 [0.08]	0.48 [0.10]	0.47 [0.15]	0.47 [0.12]
ρ_p (p markup)	0.65	0.67 [0.19]	0.66 [0.18]	0.67 [0.18]	0.64 [0.24]	0.65 [0.20]
ρ_w (w markup)	0.33	0.32 [0.08]	0.32 [0.08]	0.33 [0.08]	0.29 [0.15]	0.31 [0.14]
ρ_g (G)	0.93	0.93 [0.04]	0.94 [0.04]	0.90 [0.11]	0.86 [0.12]	0.93 [0.05]
σ_z (technology)	1.10	1.09 [0.07]	1.14 [0.11]	1.15 [0.11]	1.13 [0.15]	1.15 [0.14]
σ_m (monetary)	0.24	0.24 [0.01]	0.25 [0.02]	0.25 [0.02]	0.26 [0.03]	0.25 [0.03]
σ_b (preference)	0.71	0.73 [0.44]	0.79 [0.86]	0.82 [1.39]	0.96 [1.47]	0.76 [0.93]
σ_i (I specific)	0.21	0.24 [0.09]	0.23 [0.10]	0.24 [0.11]	0.22 [0.16]	0.22 [0.15]
σ_p (p markup)	0.10	0.10 [0.03]	0.10 [0.03]	0.10 [0.03]	0.10 [0.04]	0.10 [0.03]
σ_w (w markup)	0.21	0.20 [0.02]	0.21 [0.02]	0.21 [0.03]	0.23 [0.06]	0.22 [0.06]
σ_g (G)	0.36	0.36 [0.02]	0.37 [0.02]	0.36 [0.02]	0.36 [0.03]	0.36 [0.03]

Table 1: Monte Carlo results. 100 simulations with sample size $T = 170$. Column 2 reports the true values. Each column displays median estimates. Standard deviations in brackets.

C Appendix: the BVAR model

In this Appendix, we discuss the BVAR specification.

We combine two common prior specifications. The first is the Minnesota prior (Litterman, 1996), based on the extension to a Normal-inverted Wishart, as in Kadiyala and Karlssson, 1997. The second is the sum-of-coefficient prior (Sims and Zha, 1998) in which the VAR coefficients are restricted to sum to zero. The presentation in this Appendix follows Banbura *et al.*, 2011.

Let us write the VAR model for the N variables Y_t in levels as:

$$Y_t = c + A_1 Y_{t-1} + ... + A_p Y_{t-p} + u_t \tag{20}$$

or, in its error correction form:

$$\Delta Y_t = c - (I - A_1 - A_2 - ... - A_p) Y_{t-1} + B_1 \Delta Y_{t-1} + ... + B_{p-1} \Delta Y_{t-p+1} + u_t \tag{21}$$

The Minnesota prior assumes that:

$$E[(A_p)_{ij}] = \begin{cases} \delta_i & j = i, p = 1 \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

$$V[(A_p)_{ij}] = \begin{cases} \dfrac{\lambda_0^2}{p^2} & j = i \\ \dfrac{\lambda_0^2 \lambda_c \sigma_i^2}{p^2 \sigma_j^2} & \text{otherwise} \end{cases} \tag{23}$$

where the parameters $A_1, ...A_p$ are assumed to be a priori independent. The prior on the intercept is diffuse.

The Minnesota prior assumes that the equations in the VAR are tilted towards an AR(1) with coefficient δ_i. If $\delta_i = 1$ this implies that variable i follows a random walk with drift.

The specification in (23) assumes that the prior variance of the coefficients is inversely related to the lag (the coefficient $1/p^2$). The parameter λ_0 captures the tightness of the prior and the parameter λ_c controls the cross equation tightness of the prior. The scaling factor σ_i^2/σ_j^2 takes into account that variables may have different scales.

Let us now focus on the ECM representation (21). Note that in a VAR in first-differences, the restriction $(I - A_1 - A_2 - ... - A_p) = 0$ is satisfied. The sum of coefficients prior (Sims and Zha, 1997) forces the matrix $\Pi = (I - A_1 - A_2 - ... - A_p)$ to shrink to 0, and therefore tilts the VAR in levels towards a VAR in first-differences.

The VAR model in equation (20) can be written as:

$$Y = X\beta + U \tag{24}$$

where $Y = [Y_1 ... Y_T]'$, $X = [X_1 ... X_T]'$ and $X = [Y'_{t-1} ... Y'_{t-p} \;\; 1]'$, $U = [u_1 ... u_T]'$, $\beta = [A_1 ... A_p \;\; c]'$.

Following Kadiyala and Karlsson, 1997, the Normal-inverted Wishart prior has the form:

$$\begin{aligned} vec(\beta)|\Psi &\sim N(vec(\beta_0), \Psi \otimes \Omega_0) \\ \Psi &\sim IW(S_0, \alpha_0) \end{aligned} \tag{25}$$

The prior parameters β_0, Ω_0, S_0 and α_0 in (25) are chosen so that prior expectations and variances of β coincide with those in equations (22) and (23) and that the prior covariance matrix of the residuals is diagonal, fixed and known: $\Psi = \Sigma$, with $\Sigma = diag[\sigma_1^2, ..., \sigma_N^2]$.[3]

Dummy observations can be used to implement the priors. As shown by Sims and Zha, 1998 and Banbura *et al.*, 2010, adding T_d observations Y_d and X_d to model (24) is equivalent to imposing the Normal-inverted Wishart prior (25) with $\beta_0 = (X'_d X_d)^{-1} X'_d Y_d$, $\Omega_0 = (X'_d X_d)^{-1}$, $U_d = (Y_d - X_d B_0)$, $S_0 = U'_d U_d$, and $\alpha_0 = T_d - (Np + 1)$. The dummy observations are:

$$Y_d = \begin{bmatrix} (1/\lambda_0) diag[\delta_1 \sigma_1, ..., \delta_N \sigma_N] \\ 0_{N(p-1) \times N} \\ diag[\sigma_1, ..., \sigma_N] \\ 0_{1 \times N} \end{bmatrix} \tag{26}$$

$$X_d = \begin{bmatrix} (1/\lambda_0)(J_p \otimes diag[\sigma_1, ..., \sigma_N]) & 0_{Np \times 1} \\ 0_{N \times Np} & 0_{N \times 1} \\ 0_{1 \times Np} & \epsilon \end{bmatrix} \tag{27}$$

where $J_p = diag[1, ..., p]$ and ϵ is a very small number that implements an improper prior on

[3] The condition that allows the covariance matrix of $vec(\beta)|\Psi$ to be equal to $(\Psi \otimes \Omega_0)$ is: $\lambda_c = 1$. This is assumed throughout.

the constant c. Although the parameters should be set using only prior information, we follow Litterman, 1986, Sims and Zha, 1998 and Banbura *et al.*, 2010 and set the σ_i^2 equal to the residual variance obtained from an autoregression of order p for the variables Y_{it}.

The sum of coefficients prior can be implemented by adding N dummy observations $(1/\tau)diag[\delta_1\mu_1, ..., \delta_N\mu_N]$ to Y_d and $\left[(1/\tau)(i_p' \otimes diag[\delta_1\mu_1, ..., \delta_N\mu_N]) \;\; 0_{N\times 1}\right]'$ to X_d.

The parameter μ_i is equal to the sample mean of the variable Y_i (as in Sims and Zha, 1998) and i_p is a $p \times 1$ unit vector. The total number of dummy observations is now $T_d = N(p+2)+1$.

The regression model augmented with the dummy observations is:

$$Y_* = X_*\beta_* + U_* \tag{28}$$

where $Y_* = [Y' \; Y_d']'$, $X_* = [X' \; X_d']'$, $U_* = [U' \; U_d']'$ and $\beta_* = [A_1 ... A_p \; c]'$ implies that the posterior distribution has the form:

$$\begin{gathered} vec(\beta_*)|\Psi, Y \sim N(vec(\hat{\beta}_*), \Psi \otimes (X_*'X_*)^{-1}) \\ \Psi|Y \sim IW(\hat{\Sigma}_*, T_d + T + 2 - (Np+1)) \end{gathered} \tag{29}$$

where $\hat{\beta}_* = (X_*'X_*)^{-1}X_*'Y_*$ and $\hat{\Sigma}_* = (Y_* - X_*\hat{\beta}_*)'(Y_* - X_*\hat{\beta}_*)$.

We estimate the forecasting BVAR with variables in levels with $p = 1$, as selected by the Schwarz Bayesian information criterion (BIC). We set $\delta_i = 1$ for output, investment, consumption and the real wage and $\delta_i = 0.9$ for hours, inflation and the nominal interest rate, those variables that are stationary in the DSGE model, reflecting a prior belief that they exhibit a fair degree of persistence (Koop and Korobilis, 2009)[4]. We set $\lambda_0 = 0.12$ as in Kadiyala and Karlsson, 1997[5], $\tau = 10\lambda_0$ as in Banbura *et al.*, 2010 and Christoffel *et al.*, 2011.[6]

[4]Banbura *et al.*, 2010, set $\delta_i = 0$ for variables characterized by substantial mean reversion

[5]We have also experimented with $\lambda_0 = 0.262$ and $\lambda_0 = 0.108$ as suggested in Banbura, Giannone and Reichlin, 2010. Forecasting results are unchanged.

[6]Figure 1 in the paper displays estimates for the spectral densities of the observables $[\Delta\log(Y_t);\ \Delta\log(I_t);\ \Delta\log(C_t);\ \Delta\log(W_t);\ H_t;\ \pi_t;\ i_t]$. Those estimates have been obtained by fitting a BVAR with Minnesota and sum-of-coefficients priors on the levels: $[\log(Y_t);\ \log(I_t);\ \log(C_t);\ \log(W_t);\ H_t;\ \pi_t;\ i_t]$ and differentiating variables expressed in growth rates. p is set to 4 and $\lambda_0 = 0.12$. Shaded areas show 68% credible sets.