On the distribution of information in the moment structure of DSGE models

Nikolay Iskrev
Bank of Portugal

There is a long tradition in macroeconomics of using selected moments of the data to determine empirically relevant values of structural parameters. This paper presents a formal approach for evaluating the implications of DSGE models for the distribution of information in the moment structure of their variables. Specifically, it shows how to address the following questions: (1) what are the efficiency gains from using more instead of fewer moments; (2) what is the efficiency loss from assigning suboptimal weights on the used moments; and (3) which particular dimensions of the data - first and second order moments in the time domain, and sets of frequencies in the frequency domain - are most informative about individual structural parameters. The analysis is based on the asymptotic properties of maximum likelihood and moment matching estimators and is simple to perform for general linearized models. A standard real business cycle model is used as an illustration.

Keywords: DSGE models, MLE, GMM, information content

JEL classification: C32, C51, C52, E32

1 Introduction

There is a long tradition in macroeconomics of using selected moments of the data to find empirically relevant values of structural parameters. One common approach is to calibrate parameters so that models match some selected moments in the data. Alternatively, there is a variety of full and limited information methods that are used to formally estimate and assess the empirical fit of full-fledged dynamic stochastic general equilibrium (DSGE) models. All of these approaches are based on the premise that DSGE models provide a complete characterization of the relationship between the parameters and the theoretical moments of the variables in the underlying model. Thus,

Contact Information: Banco de Portugal, Av. Almirante Reis 71, 1150-012, Lisboa, Portugal
e-mail: Nikolay.Iskrev@bportugal.pt; phone:+351 213130006

in principle, the moment structure provides the necessary information for calibrating or estimating the parameters of interest.

The purpose of this paper is to study how the information is distributed in the moment structure of the model variables. In general, DSGE models impose restrictions on the mean and the complete set of autocovariances $\{\boldsymbol{\Sigma}(j)\}_{j=-\infty}^{\infty}$ of the observed variables. Also, most models have implications for all frequencies of the spectrum, and not, for example, only for the business cycle frequencies. This implies that estimation based on all moments at all frequencies is more efficient than estimation based only on a part of the moments structure. Little is known, however, about the size of these efficiency gains. Furthermore, it is a common practice to use estimators that weigh the included moments in a manner that is known to be suboptimal. While it is understood that this leads to inefficient estimates, the question of how significant is the loss has largely been ignored. To examine these issues, I consider estimators that differ in the part of the moment structure they utilize and in the way in which the included moments are weighted in the objective function. Comparing the efficiency of alternative estimators, I address the following questions: (1) what are the benefits in terms of estimation precision of using more instead of fewer moments or all instead of a restricted band of frequencies; (2) what are the efficiency gains from using optimal weighting schemes; and (3) which dimensions of the data - particular unconditional moments in the time domain and frequency bands in the frequency domain - are most informative for individual deep parameters.

The answers to these questions are important to empirical researchers who wish to estimate DSGE models and have to make a choice between different estimation approaches. These include: Gaussian maximum likelihood estimation in the time domain (Hansen and Sargent (1980)) or in the frequency domain (Christiano and Vigfusson (2003)), moment matching (Christiano and Eichenbaum (1992)), VAR-based indirect inference (Smith (1993)), and impulse response matching (Rotemberg and Woodford (1997)). In general, the choice involves a trade-off between estimation efficiency and robustness of the results. One of the contributions of this paper is to provide a quantitative answer regarding the efficiency side of this trade-off. Specifically, it derives analytic expressions for the asymptotic distributions of parameter estimates for maximum likelihood (ML) in the time and frequency domains as well as for generalized method of moments (GMM) estimators with optimal and some common suboptimal weighting matrices. Evaluating the asymptotic covariance matrices for a particular DSGE model and parameter values shows the size of the asymptotic efficiency gains of the full information MLE over estimation based on restricted bands of frequencies or subsets of moments under different weighting schemes.

In addition to its purely econometric aspect, the analysis in this paper may also be useful for better understanding the economic properties of DSGE models. Economic models are build to approximate certain dimensions of observed economic behavior. However, it is often far from obvious how different economic parameters affect the stochastic properties of the model economy and how distinct these effects are from

those of other parameters. This is particularly true for large models with many, often overlapping, features, whose individual effects on the observed properties of the model may be difficult to disentangle. The extend to which such problems are exacerbated by limitations in what is observed, i.e. a set of moments or a band of frequencies, can be determined by comparing the asymptotic efficiency of full information MLE with the respective limited information approach. Straightforward analysis of the asymptotic covariance matrices shows how strong and distinct are the implications of individual structural parameters on the observed moment structure, and reveals groups of strongly related parameters.

While various different methods have been proposed and are being used to estimate DSGE models, there has been relatively little research comparing the properties of these estimators. An early contribution is West (1986) who compares the asymptotic distributions of a full information ML and a limited information instrumental variables estimators applied to a single equation from a model with rational expectations. Several studies, such as Fuhrer and Rudebusch (2004) and Jondeau at al. (2004), have used Monte Carlo experiments to examine the small sample properties of the ML and GMM estimators applied to single equations containing expectations. Closest to the present paper is Ruge-Murcia (2007) who studies the small sample properties of several different methods, including moment matching and ML, applied to the estimation of the parameters of a fully articulated DSGE model. Although Ruge-Murcia (2007) also reports asymptotic standard errors, he does not make use of the fact that the asymptotic covariance matrices of the ML and GMM estimators can be evaluated exactly.[1] The advantage of the approach proposed in the present paper over the usual Monte Carlo method is that it allows to compare the efficiency properties of different estimators directly, without having to simulate and estimate models. Consequently, it is much easier to study the effect of using different sets of observables, moments or frequencies and compare the results for a range of plausible values of the structural parameters. Furthermore, unlike the computationally intensive Monte Carlo approach, it is not limited to small-scale models with few parameters.[2] On the other hand, the Monte Carlo approach is clearly more appropriate for studying the small-sample properties and the consequences of misspecification for the performance of different estimators. These issues are outside the of scope of the present paper.

The paper is also related to the recent literature on identification in DSGE models. Iskrev (2010b) and Komunjer and Ng (2011) present conditions for identification of structural parameters from the first and second order moments. Qu and Tkachenko (2012) add a condition for identification from a subset of frequencies. Canova and Sala (2009), Iskrev (2010a), and Müller (2012) deal with the strength of parameter identification. The present paper is most similar to Iskrev (2010a) in that it treats identification and the distribution of information about the parameters as properties of the underlying

[1] Ruge-Murcia (2007) uses the Hessian of the log-likelihood to approximate the Fisher information matrix and the Newey–West procedure to estimate the optimal weighting matrix of the GMM estimator.

[2] The model in Ruge-Murcia (2007) has 5 parameters, only 3 of which are free.

economic model, and in using asymptotic covariance matrices to study these properties. In a full information setting the asymptotic covariance matrix is equal to the inverse of the Fisher information matrix, which, as Rothenberg (1971) points out, "is a measure of the amount of information about the unknown parameters available in the sample". The focus in this paper is shifted from measuring the strength of identification to uncovering the sources of identification. Furthermore, it shows how to extend the analysis in Iskrev (2010a) to a limited information setting, specifically to GMM estimation based on first and second order moments and ML estimation based on subsets of frequencies.

The importance of understanding the sources of identification of parameters in macroeconomic models has been emphasized recently by Ríos-Rull et al. (2012). In that paper, the authors argue that alternative empirical methodologies are associated with different identification assumptions which may lead to different and even conflicting findings regarding the answers to important macroeconomic questions. Using the aggregate labor supply elasticity parameter to make their argument, Ríos-Rull et al. (2012) describe how identification is achieved in the context of various calibration exercises and formal estimation using likelihood-based Bayesian methods. The present paper complements Ríos-Rull et al. (2012) by showing how to determine which particular moments of the observed variables are most informative about individual structural parameters. It does so in the framework of optimal GMM estimation using the fact that an optimal estimator weighs the utilized moments so as to maximize their information content. The relative importance of different moments is determined using the weights assigned to them in the first order conditions. Although, in general, estimation by GMM is not equivalent to full information likelihood-based estimation, it is shown that, in the context of a particular DSGE model, including sufficiently many lagged autocovariances in the GMM objective function results in that estimator being asymptotically as efficient as MLE. Thus, the results regarding the relative importance of the moments are relevant for the full information setting as well.

The paper proceeds as follows. Section 2 introduces the class of models that are the subject of the paper, and presents the main econometric results used in the analysis. The main results of the paper are presented in Section 3, which has five subsections. The first describes a standard real business cycle model which is used to illustrate the application of the proposed approach. The second studies the asymptotic efficiency gains from increasing the number of moments used in the estimation. Specifically, it compares the asymptotic efficiency of MLE to that of GMM estimators with an increasing number of optimally weighted moments. It also shows how to explain differences in the size of the gains across parameters in terms of the type of information contained in the additional moments. The third studies how the choice of a weighting matrix affects the asymptotic efficiency properties of the GMM estimator. It considers two alternatives of the optimal weighting matrix, namely, a diagonal matrix with the inverse of the variances of the moments on the diagonal, and the identity matrix. The fourth asks which particular moments are most informative about the structural parameters. It uses the first order conditions of the optimal GMM estimator to identify moments that are most important

about individual parameters, and compares the efficiency of the optimal GMM estimator with different sets of moments to identify those that are most informative about the vector of parameters as a whole. The fifth studies how information is distributed in the frequency domain by comparing the asymptotic efficiency of MLE using information from three subsets of frequencies: low, business cycle and high. It also shows how to analyze differences between parts of the spectrum in terms of the type of information they convey about the parameters. Section 4 concludes.

2 DSGE models and estimators

This section introduces the class of linearized DSGE models and describes the econometric framework used in the analysis.

2.1 General setup

I consider DSGE models expressed in terms of stationary variables and linearized around the steady state values of these variables. Such models can be expressed as follows

$$\boldsymbol{\Gamma}_0(\boldsymbol{\theta})\boldsymbol{z}_t = \boldsymbol{\Gamma}_1(\boldsymbol{\theta})\,\mathrm{E}_t\,\boldsymbol{z}_{t+1} + \boldsymbol{\Gamma}_2(\boldsymbol{\theta})\boldsymbol{z}_{t-1} + \boldsymbol{\Gamma}_3(\boldsymbol{\theta})\boldsymbol{u}_t \tag{2.1}$$

where $\boldsymbol{z}_t$ is an $m-$dimensional vector of deviations from steady states, and $\boldsymbol{u}_t$ is an n-dimensional random vector of structural shocks with $\boldsymbol{u}_t \sim i.i.d.\ \mathbb{N}(\boldsymbol{0}, \boldsymbol{I}_n)$. The elements of the matrices $\boldsymbol{\Gamma}_0$, $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$ and $\boldsymbol{\Gamma}_3$ are functions of a $k-$dimensional vector of deep parameters $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ is a point in $\boldsymbol{\Theta} \subset \mathbb{R}^k$. The parameter space $\boldsymbol{\Theta}$ is defined as the set of all theoretically admissible values of $\boldsymbol{\theta}$.

The solution of equation (2.1) can be expressed as a linear state space model

$$\boldsymbol{x}_t = \boldsymbol{s}(\boldsymbol{\theta}) + \boldsymbol{C}(\boldsymbol{\theta})\boldsymbol{z}_t \tag{2.2}$$

$$\boldsymbol{z}_t = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{z}_{t-1} + \boldsymbol{B}(\boldsymbol{\theta})\boldsymbol{u}_t \tag{2.3}$$

where $\boldsymbol{x}_t$ is a l-dimensional vector of observed variables, $\boldsymbol{s}(\boldsymbol{\theta})$ is a l-dimensional vector, $\boldsymbol{C}$ is a $l \times m$ matrix, $\boldsymbol{A}$ is a $m \times m$ matrix, and the $\boldsymbol{B}$ is a $m \times n$ matrix.

2.2 Econometric framework

I consider two types of estimators: generalized method of moments (GMM) estimators, and Gaussian maximum likelihood estimator (MLE). What these estimators have in common is that they use only information contained in the first and second order moments of the data. They differ with respects to which the particular moments are utilized and what weights are assigned to them in the objective function.

2.2.1 Generalized method of moments estimation

The state space model (2.2)-(2.3) implies that the unconditional first and second moments of $\boldsymbol{x}_t$ are given by

$$\mathrm{E}\,\boldsymbol{x}_t := \boldsymbol{\mu} = \boldsymbol{s} \tag{2.4}$$

$$\mathrm{cov}(\boldsymbol{x}_{t+h}, \boldsymbol{x}_t') := \boldsymbol{\Sigma}(h) = \begin{cases} \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{z}}(0)\boldsymbol{C}' & \text{if } h = 0 \\ \boldsymbol{C}\boldsymbol{A}^h\boldsymbol{\Sigma}_{\boldsymbol{z}}(0)\boldsymbol{C}' & \text{if } h > 0 \end{cases} \tag{2.5}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{z}}(0) := \mathrm{E}\,\boldsymbol{z}_t\boldsymbol{z}_t'$ solves the matrix equation

$$\boldsymbol{\Sigma}_{\boldsymbol{z}}(0) = \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{z}}(0)\boldsymbol{A}' + \boldsymbol{\Omega} \tag{2.6}$$

Let $\boldsymbol{\sigma}_q := [\mathrm{vech}(\boldsymbol{\Sigma}(0))', \mathrm{vec}(\boldsymbol{\Sigma}(1))', ..., \mathrm{vec}(\boldsymbol{\Sigma}(q))']'$ and define $\boldsymbol{m}(\boldsymbol{\theta}, q) := [\boldsymbol{\mu}', \boldsymbol{\sigma}_q']'$ as a $q \times l^2 + l(l+3)/2$-dimensional vector collecting the fist and second order moments up to lag q. Furthermore, let $\hat{\boldsymbol{m}}_T(q)$ be the estimate of $\boldsymbol{m}(\boldsymbol{\theta}, q)$ using a sample $\boldsymbol{X}_T := [\boldsymbol{x}_1', \ldots, \boldsymbol{x}_T']'$. A GMM estimation of $\boldsymbol{\theta}$ is based on the fact that the sample moments converge to their unconditional expected values. Specifically, the GMM estimator minimizes over the parameter space $\boldsymbol{\Theta}$ the objective function

$$Q_T(\boldsymbol{\theta}, q) := \left(\boldsymbol{m}(\boldsymbol{\theta}, q) - \hat{\boldsymbol{m}}_T(q)\right)' \boldsymbol{W}_T \left(\boldsymbol{m}(\boldsymbol{\theta}, q) - \hat{\boldsymbol{m}}_T(q)\right), \tag{2.7}$$

where $\boldsymbol{W}_T$ is a positive definite and possibly random weighting matrix converging in probability to a positive definite matrix $\boldsymbol{W}$.

Under the regularity condition stated in Hansen (1982), the maximizer $\hat{\boldsymbol{\theta}}_T(q)$ of Q_T is consistent and asymptotically normally distributed with

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T(q) - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathbb{N}\left(\boldsymbol{0}, \boldsymbol{V}_0^{\boldsymbol{\theta}}\right), \tag{2.8}$$

where $\boldsymbol{J}_0 := \partial \boldsymbol{m}(\boldsymbol{\theta}_0, q)/\partial \boldsymbol{\theta}'$, $\boldsymbol{V}_0^{\boldsymbol{\theta}} := (\boldsymbol{J}_0'\boldsymbol{W}\boldsymbol{J}_0)^{-1}\boldsymbol{J}_0'\boldsymbol{W}\boldsymbol{V}_0^{\boldsymbol{m}}\boldsymbol{W}\boldsymbol{J}_0(\boldsymbol{J}_0'\boldsymbol{W}\boldsymbol{J}_0)^{-1}$ and $\boldsymbol{V}_0^{\boldsymbol{m}}$ is the asymptotic covariance matrix of the moment condition,

$$\sqrt{T}\left(\hat{\boldsymbol{m}}_T(q) - \boldsymbol{m}(\boldsymbol{\theta}_0, q)\right) \xrightarrow{d} \mathbb{N}\left(\boldsymbol{0}, \boldsymbol{V}_0^{\boldsymbol{m}}\right) \tag{2.9}$$

For a given set of moments in $\boldsymbol{m}$, the asymptotic efficiency of the GMM estimator is determined by the choice of weighting matrix $\boldsymbol{W}$. In order to evaluate $\boldsymbol{V}^{\boldsymbol{\theta}}$, in addition to $\boldsymbol{W}$ one also needs the derivative of the theoretical moments $\boldsymbol{J}$ and the asymptotic covariance matrix of the sample moments $\boldsymbol{V}^{\boldsymbol{m}}$. The Jacobian matrix is straightforward to compute analytically as shown in Iskrev (2010b). To see how to compute $\boldsymbol{V}^{\boldsymbol{m}}$, first note that the matrix is block diagonal with blocks $\boldsymbol{V}^{\boldsymbol{\mu}}$ and $\boldsymbol{V}^{\boldsymbol{\sigma}}$ corresponding to the mean vector $\boldsymbol{\mu}$ and the vector of second order moments $\boldsymbol{\sigma}$. That is

$$\boldsymbol{V}^{\boldsymbol{\mu}} := \lim_{T\to\infty} T\,\mathrm{cov}\left(\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}(\boldsymbol{\theta})\right) \tag{2.10}$$

$$\boldsymbol{V}^{\boldsymbol{\sigma}} := \lim_{T\to\infty} T\,\mathrm{cov}\left(\hat{\boldsymbol{\sigma}}_T - \boldsymbol{\sigma}(\boldsymbol{\theta})\right) \tag{2.11}$$

It is well known that $\boldsymbol{V}^{\boldsymbol{\mu}}$ in (2.10) is equal to 2π times the spectral density of $\boldsymbol{x}_t$ evaluated at frequency zero. Therefore, using (2.2)-(2.3) we have

$$\boldsymbol{V}^{\boldsymbol{\mu}} = \sum_{h=-\infty}^{\infty} \boldsymbol{\Sigma}(h) = \boldsymbol{C}\left(\boldsymbol{I}_m - \boldsymbol{A}\right)^{-1} \boldsymbol{B}\boldsymbol{B}' \left(\boldsymbol{I}_m - \boldsymbol{A}'\right)^{-1} \boldsymbol{C}' \tag{2.12}$$

The asymptotic covariance matrix of the second order moments can be computed using the following result due to Bartlett (1955) (see the working paper version of Lund at al. (2009) for details.)

Theorem 1. *Let $\boldsymbol{x}_t$ be a process generated by the model in Section 2. If $\hat{\sigma}_{i,j}(h)$ is the sample autocovariance at lag h between the $i-th$ and the $j-th$ components of $\boldsymbol{x}$, then*

$$\begin{aligned} \lim_{T\to\infty} T \operatorname{cov}\left(\hat{\sigma}_{i_1,j_1}(p), \hat{\sigma}_{i_2,j_2}(q)\right) &= \sum_{h=-\infty}^{\infty} \big(\sigma_{i_1,i_2}(h)\sigma_{j_1,j_2}(h-p+q) \\ &+ \sigma_{i_1,j_2}(h+q)\sigma_{j_1,i_2}(h-p)\big) \end{aligned} \tag{2.13}$$

Remark. *If $\boldsymbol{u}_t$ is not Gaussian, the collection of sample autocovariances are still asymptotically normal with*

$$\lim_{T\to\infty} T \operatorname{cov}\left(\hat{\sigma}_{i_1,j_1}(p), \hat{\sigma}_{i_2,j_2}(q)\right) = \sum_{h=-\infty}^{\infty} \left(\mathrm{E}\, \boldsymbol{x}_{t,i_1}\boldsymbol{x}_{t+p,j_1}\boldsymbol{x}_{t+h,i_2}\boldsymbol{x}_{t+h+q,j_2} - \sigma_{i_1,j_1}(p)\sigma_{i_2,j_2}(q)\right)$$

Unlike $\boldsymbol{V}^{\boldsymbol{\mu}}$, which can be evaluated in closed form using the second equality in (2.12), in general $\boldsymbol{V}^{\boldsymbol{\sigma}}$ has to be approximated by truncating the infinite sum in (2.13). Note, however, that the terms involved in the summation are very easy to compute, for any lag, with the formula in (2.5). Thus, $\boldsymbol{V}^{\boldsymbol{\sigma}}$ can be approximated arbitrarily well using the result in Theorem 1.

Example Suppose that x_t is generated by an ARMA(1,1) model:

$$x_t = \phi_1 x_{t-1} + \varepsilon_t + \phi_2 \varepsilon_{t-1}, \qquad \varepsilon \sim i.i.d.\ \mathbb{N}\left(0, \sigma^2\right) \tag{2.14}$$

There are three parameters in this model, $\boldsymbol{\theta} := [\phi_1, \phi_2, \sigma]'$, which can be estimated by matching the model-implied unconditional variance and the first $q \geq 2$ autocovariances of x_t to their sample counterparts. Table 1 shows the asymptotic standard errors of the efficient GMM estimator of $\boldsymbol{\theta}$ for two different DGPs. The first DGP features a large autoregressive coefficient $\phi_1 = .9$, and a small moving average coefficient $\phi_2 = .1$; the reverse is true for the second DGP with $\phi_1 = .1$ and $\phi_2 = .5$. In both cases σ is equal to 1. As the results in the table show, the properties of the model determine the distribution of information in the autocovariance function of the x_t. Specifically, for the first DGP the standard errors of all parameters converge very quickly - after $q = 2$ for σ, and after $q = 3$ for ϕ_1 and ϕ_2. For the second DGP the standard errors of ϕ_1 and ϕ_2 converge only after $q = 10$ and $q = 9$, respectively, and the standard error of σ converges after $q = 5$. That is, autocovariances beyond lag 3 in the first case, and lag 10 in the second case, provide very little additional information about the parameters of the model.

[insert Table 1 here]

2.2.2 Maximum likelihood estimation

Time Domain

In the time domain the log-likelihood function of a sample $\boldsymbol{X}_T$ can be derived using the prediction error method whereby a sequence of one-step ahead prediction errors $\boldsymbol{e}_{t|t-1} = \boldsymbol{x}_t - \boldsymbol{s} - \boldsymbol{C}\hat{\boldsymbol{z}}_{t|t-1}$ is constructed by applying the Kalman filter to the obtain one-step ahead forecasts of the state vector $\hat{\boldsymbol{z}}_{t|t-1}$. The Gaussianity of the structural shocks implies that the conditional distribution of $\boldsymbol{e}_{t|t-1}$ is also Gaussian with zero mean and covariance matrix given by $\boldsymbol{S}_{t|t-1} = \boldsymbol{C}\boldsymbol{P}_{t|t-1}\boldsymbol{C}'$, where $\boldsymbol{P}_{t|t-1} = \mathrm{E}\left(\boldsymbol{z}_t - \hat{\boldsymbol{z}}_{t|t-1}\right)\left(\boldsymbol{z}_t - \hat{\boldsymbol{z}}_{t|t-1}\right)'$ is the conditional covariance matrix of the one-step ahead forecast, and is also obtained from the Kalman filter recursion. This implies that the log-likelihood function of the sample is given by

$$\ell_T(\boldsymbol{\theta}) = \text{const.} - \frac{1}{2}\sum_{t=1}^{T}\log(|\boldsymbol{S}_{t|t-1}|) - \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{e}_{t|t-1}'\boldsymbol{S}_{t|t-1}^{-1}\boldsymbol{e}_{t|t-1} \tag{2.15}$$

Under standard assumptions the maximizer $\tilde{\boldsymbol{\theta}}_T$ of $\ell_T(\boldsymbol{\theta})$ is consistent and asymptotically normally distributed with

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \xrightarrow{d} \mathbb{N}\left(\boldsymbol{0}, \boldsymbol{\mathcal{I}}_0^{-1}\right) \tag{2.16}$$

where $\boldsymbol{\mathcal{I}}_0$ is the asymptotic Fisher information matrix evaluated at the true $\boldsymbol{\theta}$. The computation of information matrix for Gaussian linear state space models is discussed, among others, in Zadrozny (1989), Segal and Weinstein (1989), Zadrozny and Mittnik (1994), and Klein and Neudecker (2000). In the context of DSGE model, the computation of $\boldsymbol{\mathcal{I}}_0$ is discussed in Iskrev (2008) and Iskrev (2010a).

Frequency Domain

The log-likelihood function of $\boldsymbol{X}_T$ can be approximated in the frequency domain as

$$\begin{aligned}\ell_T^{\star}(\boldsymbol{\theta}) &= \text{const.} - \sum_{j=1}^{T/2+1}\log\det(\boldsymbol{F}(\omega_j)) - \sum_{j=1}^{T/2+1}\mathrm{tr}\left(\boldsymbol{F}(\omega_j)^{-1}\hat{\boldsymbol{F}}(\omega_j)\right) \\ &- \frac{T}{2}\mathrm{tr}\left[\boldsymbol{F}(0)^{-1}\left(\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}\right)\left(\hat{\boldsymbol{\mu}}_T - \boldsymbol{\mu}\right)'\right]\end{aligned} \tag{2.17}$$

where $\boldsymbol{F}(\omega) := \sum_{\tau=-\infty}^{\infty}\boldsymbol{\Sigma}(j)\exp^{-i\omega\tau}$ is the spectral density matrix of $\boldsymbol{x}_t$, $\hat{\boldsymbol{F}}(\omega_j) := \frac{1}{T}\boldsymbol{x}(\omega_j)\boldsymbol{x}(\omega_j)'$ is the periodogram of $\boldsymbol{X}_T$, and $\boldsymbol{x}(\omega_j) := \sum_{t=1}^{T}\boldsymbol{x}_t\exp^{-i\omega_j t}$, $\omega_j = \frac{2\pi j}{T}$, $j = 1,\ldots,T$ is the Fourier transform of $\boldsymbol{X}_T$.

A general discussion of the asymptotic theory of MLE in the frequency domain is presented in Dunsmuir (1979). Under the conditions stated there the maximizer of $\ell_T^{\star}(\boldsymbol{\theta})$ is consistent and asymptotically normally distributed with

$$\sqrt{T}(\tilde{\boldsymbol{\theta}}_T^{\star} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathbb{N}\left(\boldsymbol{0}, \mathfrak{I}_0^{-1}\right) \tag{2.18}$$

where $\mathfrak{I}_0$ is the frequency domain version of the asymptotic Fisher information matrix evaluated at the true $\boldsymbol{\theta}$. Whittle (1953) provides an expression for $\mathfrak{I}$ for general linear Gaussian state space models. Some details on the evaluation of the matrix for DSGE models are presented in the Appendix.

The two information matrices, $\boldsymbol{\mathcal{I}}$ and $\mathfrak{I}$, are equivalent apart from errors due to the approximations involved in computing $\mathfrak{I}$. The frequency domain approach is nevertheless useful as it allows us to determine the distribution of information contained in the moment structure across different frequencies. For instance, we can find out whether low, high, or business cycle frequencies are most informative for a given parameter. The time domain approaches, on the other hand, can be used to determine the relative informativeness of the means, covariances, and autocovariances at different lags.

Example (continued) The ARMA model in (2.14) *can be estimated by ML, maximizing either* (2.15) *or* (2.17)*. The left panel of Table 2 shows asymptotic standard errors for the frequency domain MLE using different frequencies bands, namely, low frequencies (with period from 32 quarters to infinity), business cycle (BC) frequencies (with period from 6 to 32 quarters), and high frequencies (with period between 2 and 6 quarters). The results for the full set of frequencies, equivalent to the time domain MLE, are also shown. Note that they are the same as the converged GMM standard errors in Table 1. This supports the earlier observation that autocovariances at lags longer than what is shown in the table do not provide additional information about the parameters. As with the GMM results in Table 1, the distribution of information across frequencies changes quite substantially between the two DGPs. Specifically, under DGP 1 the low frequency band yields the lowest standard error for ϕ_1 and the business cycle frequencies are most informative about σ. Under DGP 2 the high frequencies are the most informative for both ϕ_1 and σ. In both cases the standard error of ϕ_2 is lowest in the high frequency band. Another interesting result in the table is the very large standard errors of ϕ_2 and σ in the low frequency band, and particularly under DGP 2 when all three parameters are nearly unidentifiable. It turns out that the problem has to do with the very similar roles played by ϕ_2 and σ in DGP1, and all three parameters in DGP 2, in terms of the low frequency properties of the process. This can be seen by the dramatic reduction in the standard error of ϕ_1 and ϕ_2 if σ is assumed known, as shown in the right panel of Table 2. Note also that the standard errors are affected only when a restricted part of the spectrum is used, while if estimation is based on all frequencies the standard errors of ϕ_1 and ϕ_2 are the same whether σ is estimated or not.*

[insert Table 2 here]

This example demonstrates the potential of the proposed approach to provide insights into the properties of models regarding the distribution of information in the moment structure. The results shown in Tables 1 and 2 are not particularly interesting on their own as they only illustrate what are well-known features of ARMA processes. For example, it is quite obvious that whether the lower or the higher frequencies are

more informative about the AR part depends on the persistence of the process as this determines whether there is more short term or more long term volatility. Also, as in the pure AR and MA models, most of the information about the AR part is in the first few autocovariances, while the autocovariances at longer lags contain relatively more information about the MA part. However, applying this line of reasoning in the context of large and complicated structural models, such as DSGE models, is much harder since the relationship between the deep parameters and the time series representation implied by the model is extremely complex. As will be shown in the next section, using the approach proposed in this paper is not any harder for such models than it is for the simple ARMA example.

3 An application

This section shows how to apply the econometric results presented above to study the distribution of information in the moments of a DSGE model. It is divided in five subsections. The first presents the model, the parameter values and observables used in the study. Subsections two and three compare the efficiency of ML and GMM estimators with optimal and some suboptimal weighting matrices. Subsection four seeks to identify the most informative moments in the moment structure of the observables, while the fifth considers the distribution of information in the frequency domain. The section concludes with a discussion of the consequences of having different values for some parameters and using alternative sets of observables, relegating the details to the Appendix.

3.1 Model

Consider a simple economy in which a central planner solves the following problem:

$$\max \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \exp(v_t) \left(\ln C_t - \frac{(H_t/\exp(b_t))^{1+1/\nu}}{1+1/\nu} \right) \tag{3.1}$$

subject to

$$Y_t = C_t + K_{t+1} - (1-\delta)K_t = \exp(a_t) H_t^{\alpha} K_t^{1-\alpha}, \tag{3.2}$$

where C_t is consumption, H_t is hours worked, Y_t is output, K_t is capital. There are three types of shocks in the economy: a_t is a technology shock, b_t is a labor supply shock and v_t is a intertemporal preference shock. The three exogenous shock processes are independent and evolve according to

$$\begin{aligned} a_t &= (1-\rho_a)\bar{a} + \rho_a a_{t-1} + \sigma_a \varepsilon_{a,t} \\ b_t &= (1-\rho_b)\bar{b} + \rho_b b_{t-1} - \sigma_b \varepsilon_{B,t} \\ v_t &= \rho_v v_{t-1} + \sigma_v \varepsilon_{v,t} \end{aligned} \tag{3.3}$$

where $\varepsilon_{a,t} \sim \mathbb{N}(0,1)$, $\varepsilon_{b,t} \sim \mathbb{N}(0,1)$, $\varepsilon_{v,t} \sim \mathbb{N}(0,1)$ are mutually uncorrelated.

The model has 12 parameters collected in the vector

$$\boldsymbol{\theta} := [\alpha, \beta, \delta, \nu, \bar{a}, \bar{b}, \rho_a, \rho_b, \rho_v, \sigma_a, \sigma_b, \sigma_v].$$

The subsequent analysis is carried out using the values shown in Table 3. With three observed variables, namely, the logs of Y_t, H_t and C_t, applying the rank conditions from Iskrev (2010b) shows that for $\boldsymbol{\theta}$ to be identified it is sufficient to use the mean, the contemporaneous covariance and first order autocovariance matrix of the observables. The effect of utilizing additional lagged autocovariances is studied next.

[insert Table 3 here]

3.2 MLE and GMM with optimal weighting matrix

The first set of results concerns the case where the included moments are weighted optimally, that is, GMM estimation with optimal weighting matrix and maximum likelihood estimation. The results are presented in Table 4. For GMM the set of moments includes the means, the contemporaneous covariances and autocovariances at lags $q = 1, 2, 3, 4, 5$ and 20. The standard errors of the MLE (shown in the column labelled "lags = ∞") are computed from the inverse of the asymptotic Fisher information matrix. All standard errors are normalized for a sample size of $T = 100$.

[insert Table 4 here]

The first thing to note is that, as implied by asymptotic theory, estimation efficiency increases with the number of moments. The size of the gains, however, varies substantially across parameters. In the case of ρ_b and σ_a, the differences between the asymptotic standard errors of GMM with $q = 1$ and GMM with $q = 2$ are .14% and .41%, respectively. Even compared to ML, the GMM standard errors for these two parameters are only about 1% larger when $q = 1$. The efficiency gains are more substantial for other parameters; in particular, for ρ_v and σ_v the GMM standard errors are around 20% larger with $q = 1$ compared to $q = 2$, while compared to ML, the GMM with $q = 1$ standard errors are twice as large. The last two parameters, together with β and δ , are also the only ones for which ML is substantially better than GMM as long as q is not too large. The differences between the ML and the GMM with $q = 5$ standard errors for these four parameters are between 17% and 33%, while for most of the other parameters the differences are around 5% or less. When ML is compared to GMM with $q = 20$, the standard errors for all parameters are nearly identical.

To help understand what causes the differences in the size of the gains from using a larger number of lagged autocovariances, it is useful to consider the situation in which there is only one free parameter. Table 5 reports standard errors for each parameter computed assuming that the remaining eleven parameters are known. Naturally, the

standard errors are smaller than those in Table 4. For α, $\bar{a}$, ρ_v and σ_v, the differences reach 200%. On the other extreme is σ_a, for which the difference is only about $2-3\%$. What is more interesting is that the gains from additional lagged autocovariances are much smaller than before. As in Table 4, the largest gains are with respect to ρ_v and σ_v, but now the differences between the standard errors of ML and GMM with $q=1$ are only around 1%. This suggests that lagged autocovariances are useful only when the information they provide is shared among several free parameters. Such sharing of information occurs when different parameters have similar effects on observed properties of the model variables, and, as a result, are difficult to distinguish on the basis of a set of moments. Adding additional moments may make this easier, reducing the estimation uncertainty, as in the case of ρ_v and σ_v, but is not guaranteed to do so.

[insert Table 5 here]

The above intuition can be formalized using the following factorization of the asymptotic standard error of θ_i:[3]

$$s.e.(\theta_i|q) = s.e.(\theta_i|\boldsymbol{\theta}_{-i}, q)\left(\frac{1}{\sqrt{1-\boldsymbol{\varrho}_i^2(q)}}\right) \tag{3.4}$$

The first term $s.e.(\theta_i|\boldsymbol{\theta}_{-i}, q)$ is the standard error of θ_i given the other elements in $\boldsymbol{\theta}$ and the number of moments determined by q. The second term is a multiplier which captures the effect of having other unknown parameters on the estimation uncertainty of θ_i. It is a function of $\boldsymbol{\varrho}_i(q)$, defined as the coefficient of multiple correlation between the derivative of the GMM objective function with respect to θ_i and the derivatives with respect to the other unknown parameters,

$$\boldsymbol{\varrho}_i(q) := \operatorname{corr}\left(\frac{\partial Q_T(\boldsymbol{\theta}, q)}{\partial \theta_i}, \frac{\partial Q_T(\boldsymbol{\theta}, q)}{\partial \boldsymbol{\theta}_{-i}}\right) \tag{3.5}$$

If θ_i is the only free parameter or if $\partial Q_T(\boldsymbol{\theta}, q)/\partial\theta_i$ is uncorrelated with the other derivatives, we have $\boldsymbol{\varrho}_i(q) = 0$ and $s.e.(\theta_i|q) = s.e.(\theta_i|\boldsymbol{\theta}_{-i}, q)$. On the other hand, if the effect of θ_i on the moments is similar to that of the other parameters, $\boldsymbol{\varrho}_i(q) \approx 1$ and $s.e.(\theta_i)$ is much larger than $s.e.(\theta_i|\boldsymbol{\theta}_{-i}, q)$. In the extreme case of $\boldsymbol{\varrho}_i(q) = 1$, we have several parameters, including θ_i, which are unidentifiable on the basis of the given set of moments.

Example (continued) An example of a model with $\boldsymbol{\varrho}_i = 1$ is the ARMA(1,1) process in (2.14) *when $\phi_1 = -\phi_2$. As is well-known, in that case the two parameters cannot be identified separately. Evaluating the factorization at the parameter values of DGPs 1 and 2 shows that for all parameters the value of $\boldsymbol{\varrho}_i$ is much larger under DGP 2, and this explains the much larger standard errors. Furthermore, as can be seen in Figure*

[3]The same factorization in a full information setting is introduced in Iskrev (2010a) as a tool for identification analysis. See that paper for more details.

1 for ϕ_1, the value of $\boldsymbol{\varrho}_i$ converges much faster under DGP 1 explaining the similarly faster convergence of the standard errors in that case.

[insert Figure 1 here]

Returning to the DSGE model, Table 6 shows the values of $\boldsymbol{\varrho}_i(q)$ from the factorization in (3.4) of the standard errors in Table 4. As can be expected from the earlier discussion, the largest values are for α, $\bar{a}$, ρ_v and σ_v. Furthermore, note that the value of $\boldsymbol{\varrho}_i(q)$ decreases with q and the change is more pronounced for ρ_v and σ_v. Although the change is not large, it leads to a significant decrease in the standard errors of these parameters as q grows. For example, the difference between GMM with $q = 1$ and ML standard errors for σ_v in Table 4 is largely due to a drop in $\boldsymbol{\varrho}_i(q)$ from .9996 to .9985, which translates in a change of the second term in (3.4) from 35.4 to 18.3. This, combined with the fairly constant value of $s.e.(\theta_i|\boldsymbol{\theta}_{-i}, q)$ in Table 5, results in significant drop in the σ_v's standard error.

[insert Table 6 here]

As discussed above, a large value of $\boldsymbol{\varrho}_i(q)$ implies that there is a group of parameters whose effects on the GMM or ML objective functions are difficult to distinguish from each other. Following Iskrev (2010a), we can compute correlation coefficients for small groups of parameters in order to determine the most closely related ones. This is done in Table 7, which shows groups of one to four parameters that have the highest degree of collinearity with the parameters in the first column.[4] The results are based on the objective function of GMM with $q = 1$, but do not change for other values of q or when the ML objective is used. The results reveal that the high values of $\boldsymbol{\varrho}_i(q)$ in Table 6 are, in most cases, due to very few parameters. For instance, the pairwise correlation coefficient of ρ_v and σ_v is .998; adding the remaining ten parameters increases it to .9995 in the case of ρ_v and .9996 in the case of σ_v. On the other hand, the strongest pairwise collinearity for β is only .537 (again with respect to σ_v); adding α, δ and $\bar{a}$ raises the correlation coefficient to .9967 which is indistinguishable from the overall collinearity value.

[insert Table 7 here]

3.3 GMM with suboptimal weighting matrix

Next, I examine how the choice of $\boldsymbol{W}$ affects the asymptotic efficiency properties of the GMM estimator. Two alternatives of the optimal weighting matrix are considered: a diagonal matrix with the inverse of the variances of the moments on the diagonal, and the identity matrix. These are by far the most widely used suboptimal weighting matrices in the literature.

[4] That is, instead of the coefficient of multiple correlation of θ_i with the eleven other parameters, it is computed with respect to smaller subsets of parameters.

As before, the set of moments includes the means, the contemporaneous covariances and autocovariances at lags $q = 1, 2, 3, 4, 5$ and 20. A problem with the identification of the model arises when $q = 1$ and the moments are weighted equally. Specifically, the matrix $(\boldsymbol{J}'\boldsymbol{J})$, the inverse of which enters the sandwich formula for the asymptotic covariance matrix (see Section 2.2.1), is singular. No such problem arises with the diagonal weighting matrix, the results for which are presented in the upper panel of Table 8. Further analysis shows that the identification problem is due to near-linear dependence of the column of $\boldsymbol{J}$ corresponding to $\bar{b}$ and the other columns of the Jacobian matrix. Assuming that $\bar{b}$ is known restores identification and the results are shown in the lower panel of Table 8.

[insert Table 8 here]

Comparing these results to the standard errors with optimal weights shows that, as implied by asymptotic theory, using the optimal weighting matrix is more efficient, in some cases vastly so, than using either one of the alternative weighting schemes. The largest loss of efficiency occurs with respect to σ_a, for which the standard errors are between 16 and 27 times larger when the diagonal matrix is used instead of the optimal one. Other parameters whose standard errors are much larger with either one of the suboptimal weighting matrices are α, ν, $\bar{a}$ and σ_b. Relatively smaller are the efficiency losses for ρ_b, ρ_v and σ_v. Interestingly, the asymptotic efficiency gaps for all parameters and both weighting matrices increase monotonically with the value of q, meaning that the gains from additional lagged autocovariances are smaller here than when the moments are weighted optimally. In fact, for most parameters and lags in Table 8 increasing the number of lagged autocovariances results in larger standard errors.

Comparing the results for the two suboptimal weighting matrices shows that neither one dominates the other in the sense of always resulting in smaller standard errors. Still, the identity matrix tends to be more efficient for a larger number of parameters. For six of them - $\alpha, \delta, \bar{a}, \rho_a, \sigma_a$ and σ_v it is always more efficient to assign equal weights, while for three - ν, ρ_b and σ_b, it is more efficient to weigh the moments according to their precision. In the remaining cases the results change with the number of included autocovariances.[5]

The substantial loss of efficiency when the diagonal instead of the optimal weighting matrix is used indicates that it is not sufficient to put larger weights on moments that are more precisely estimated. In fact this may be very inefficient even compared to equal weighting if the more precisely estimated moments are also highly collinear and thus provide little independent information about the parameters. Assigning larger weights on such moments results in underweighting more informative moments because of their larger variances. Since the asymptotic weighting matrix is available, it is possible to determine the degree of model-implied collinearity among different sample moments. Not

[5]To compare the results for $q = 1$ the standard errors with the diagonal $\boldsymbol{W}$ were recalculated assuming that $\bar{b}$ is known.

surprisingly, it is found that the strongest collinearity exists among autocovariances and cross-autocovariances of the same variables. For instance, the correlation between the sample estimates of $\text{cov}(y_t, y_t)$ and $\text{cov}(y_t, y_{t-h})$ is 0.9999987 for $h = 1$, and remains above .99 for $h = 20$. The sample autocovariances and the cross autocovariances (especially of y and c) are also highly collinear. Clearly, the amount of independent information in these moments is negligible and an efficient weighting scheme should take this into account, in addition to accounting for the own uncertainty of the moments.

3.4 Identifying the most informative moments

One of the main results in the previous section was that the choice of weighting matrix has a large impact on the efficiency of the moment matching estimator. While most of the information about the parameters is concentrated in the mean vector and the first few autocovariance matrices, weighting the moments optimally may be much more efficient than, for example, assigning equal weights to the same set of moments. This implies that some moments provide much more information than others, and weighing them accordingly is crucial. The goal of this section is to determine which particular first and second order moments are most informative for which deep parameters.

The theory of optimal GMM provides a natural framework for studying the relative importance of moments in estimation. Optimality means that the moments are weighted so as to minimize the covariance matrix of the estimator, or, in other words, to maximize the information content of the used moments. This suggests that the optimal weights can be used to determine the relative importance of the moments for each parameter.

For the sake of simplifying the notation, let $\boldsymbol{m}(\boldsymbol{\theta})$ denote $\boldsymbol{m}(\boldsymbol{\theta}, p)$, $\hat{\boldsymbol{m}}$ denote $\hat{\boldsymbol{m}}_T(p)$ and m_j and $\hat{m}_j$ be the j-th elements of the two vectors. Furthermore, define $L^p := p \times l^2 + l(l+3)/2$ so that both $\boldsymbol{m}$ and $\hat{\boldsymbol{m}}$ have dimension $L^p \times 1$. Consider the first-order conditions of the problem in (2.7):

$$\frac{\partial Q_T(\hat{\boldsymbol{\theta}}_{gmm})}{\partial \boldsymbol{\theta}'} = \left(\frac{\partial \boldsymbol{m}(\hat{\boldsymbol{\theta}}_{gmm})}{\partial \boldsymbol{\theta}'} \right)' \boldsymbol{W}_T \left(\boldsymbol{m}(\hat{\boldsymbol{\theta}}_{gmm}) - \hat{\boldsymbol{m}} \right) = 0 \tag{3.6}$$

The GMM estimator of $\boldsymbol{\theta}$, defined earlier as the minimizer of (2.7), can equivalently be seen as the solution of a system of k linear combinations of L^p moment conditions. The left-hand side of the i-th equation, corresponding to the partial derivative with respect to θ_i, is a weighted sum of the form:

$$\frac{\partial Q_T(\boldsymbol{\theta})}{\partial \theta_i} = \sum_{j=1}^{L^p} \upsilon_{ij} \left(m_j(\boldsymbol{\theta}) - \hat{m}_j \right) / m_j(\boldsymbol{\theta}) \tag{3.7}$$

where υ_{ij} is the $j-$th element of $(\partial \boldsymbol{m}(\boldsymbol{\theta})/\partial \theta_i)' \boldsymbol{W}_T m_j(\boldsymbol{\theta})$. The division by $m_j(\boldsymbol{\theta})$ in (3.7) is in order to make υ_{ij} independent of the scale of the moments. When the optimal weighting matrix is used, the weights υ_{ij} depend on two factors: the sensitivity

of the theoretical moments with respect to θ_i, and the asymptotic covariance matrix of the sample moments $\boldsymbol{V}^{\boldsymbol{m}}$. Intuitively speaking, the informative moments are highly sensitive to changes in the parameter and are themselves well identified from the data. The larger, in absolute value, the weight υ_{ij}, the more important it is for θ_i to match the theoretical value of moment j to its sample estimate. This suggests measuring the importance of moment $m_j(\boldsymbol{\theta})$ for parameter θ_i with

$$\omega_{ij}^{\star} := \frac{|\upsilon_{ij}^{\star}|}{\sum_{q=1}^{L^p} |\upsilon_{iq}^{\star}|}, \tag{3.8}$$

where the subscript indicates that the weights are obtained with the optimal GMM weighting matrix.

It is important to point out that $\omega_{ij}^{\star}$ is a function of $\boldsymbol{\theta}$ only and does not depend on the sample moments. Since the underlying DSGE model completely characterizes both the stochastic properties of the model variables and the relationship between structural parameters and the moments of the observables, the value of $\boldsymbol{\theta}$ is all that is needed to determine how informative the moments are at a given point in the parameter space.

Figures 2 and 3 show all moments whose relative weights $\omega_{ij}^{\star}$ in the first order conditions exceed 0.05. The results are based on GMM with $q = 20$, which means that 189 moments enter the objective function. As the figures show, very few of these moments receive significant weight, and those that do account for much of the total weight. For example, only the means of the three observables receive any weight in the first order conditions with respect to $\bar{a}$ and $\bar{b}$. For the other parameters there are between 5 and 9 moments which account for between 75% in the case of σ_b and 98% in the case of σ_a of the total weight. For all parameters except $\bar{a}$ and $\bar{b}$ it is either the variances, covariances, or first order autocovariances that receive the bulk of the weight. Finding that autocovariances at lags 2 or greater are not assigned significant weights may be puzzling given the results in Section 3.2, which suggest that there is a substantial amount of information in the higher order autocovariances. In the case of $\bar{b}$, for instance, there is a 17% drop in the standard error when autocovariances up to lag 20 instead of only the first order one are included. The explanation lies in the fact that the first order conditions with respect to a parameter θ_i are always conditional on the other parameters. That is, they are the same whether or not θ_i is the only free parameter. As can be seen in Table 5, the standard error of $\bar{b}$ would not change with the number of lags if all parameters except $\bar{b}$ were known. The information contained in the second order moments helps with the identification of $\bar{b}$ only indirectly, by providing additional information about the other parameters in the model, and in particular about ν, ρ_a, ρ_b and σ_b, which, as indicated in Table 7, are difficult to distinguish from $\bar{b}$ on the basis of the first order moments only. Similarly, the other parameters benefit from information in higher order autocovariances even though the weights assigned to them in the first order conditions are small.

[insert Figures 2 and 3 here]

The figures suggest that moments involving output (y) and consumption (c) play a dominant role in the first order conditions of almost all parameters, while the moments of hours (h) receive significant weight only in a few cases. A fuller picture of the distribution of weight across moments of the three variables is presented in Table 9, where a distinctions is made between own moments (such as $\mathrm{E}\, y_t$ and $\mathrm{cov}(y_t, y_{t-k})$) and cross moments (such as $\mathrm{cov}(y_t, h_{t-k})$) of the variables. The moments of h alone are very important for three parameters - $\bar{a}$, $\bar{b}$ and σ_b, while cross moments of h with either c, y, or both, play an important role also for ν, ρ_a, ρ_b and σ_a. The own moments of y receive a large proportion of the weight in the first order conditions of two parameters - ρ_a and σ_a, while those of c dominate in the first order conditions of α, β, δ, $\bar{a}$, ρ_b, ρ_v and σ_v. With a very few exceptions, the cross moments of y and c have significant weights in the first order conditions of almost all parameters.

[insert Table 9 here]

Using either one of the two suboptimal weighting matrices results in a quite different distribution of weight across the moments of the observables. The results can be seen in Figures **??** to **??** and Table **??** in the Appendix. The main difference is with respect to the four deep parameters, for which, when $\boldsymbol{W}$ is the diagonal or the identity matrix, almost all of the weight in the first order conditions is placed on own moments of the variables, particularly their means. With respect to the shock parameters, the distribution of weight between own and cross moments is broadly in line with the case when the optimal $\boldsymbol{W}$ is used. Another general difference is that, while the moments of y and c are very important irrespectively of the weighting scheme, using the suboptimal matrices results in a bias towards the moments of output at the expense of those of consumption. On the other hand, in all three cases the moments involving h either dominate or are very important in the first order conditions for ν, $\bar{b}$, ρ_b and σ_b.

[insert Table 10 here]

Another way of selecting the most informative moments is to compare the efficiency of the optimal GMM estimator with different sets of moments. Unlike using the weights in the first order conditions, this approach takes into account the fact that many parameters are estimated simultaneously. In principle one can consider all possible sets of moments of a given size and select the one minimizing the covariance matrix of $\boldsymbol{\theta}$ or some elements of its diagonal. For instance, since there are 12 parameters in the model, the minimum number of moments required for identification of $\boldsymbol{\theta}$ is 12. However, trying to select the best 12 out of a large number of moments quickly becomes infeasible. For example, there are more than 1 billion possible combinations of 12 out of 36 moments, the second number being the total number of moments if the means, covariances and autocovariances up to lag 3 are considered. Considering instead the

first 27 moments (i.e. moments up to autocovariances at lag 2) and using $\ln(|\boldsymbol{V}^{\boldsymbol{\theta}}|)$ to measure the estimation uncertainty for $\boldsymbol{\theta}$ as a whole, shows that no autocovariances at lag 2 appear in the optimal set of moments. That is, the optimal set of moments is the same whether the first 18 or the first 27 moments are considered. This suggest that it may be unnecessary to extend the number of moments by adding autocovariances at longer lags. The optimal selection includes all first order moments, variances and own autocovariances, together with the cross-covariances and cross-autocovariances of c and y. In other words, the excluded 6 from the first 18 moments are all of the cross-second order moments involving h. This is consistent with the earlier finding that moments of h receive relatively small weight in the GMM first order conditions. The standard errors of the parameters with this set of moments is shown in the first column of Table 10. Note that they are larger, in some cases, such as ν, much larger than the standard errors obtained with using all of the first 18 moments (see Table 4). Thus, the excluded moments, though less important, still provide useful information about the parameters. The table also shows results with other sets of moments chosen to minimize the standard errors of individual parameters. In all cases the mean of h together with at least one of the other first-order moments is included. The reason is that without two first-order moments, one of which being $\mathrm{E}\, h_t$, the steady state parameters $\bar{a}$ and $\bar{b}$ are not identified. Also note that for all parameters except σ_a it is optimal to include at least one auto or cross-auto covariance at lag 2. This shows that it is possible to improve the identification of individual parameters by using higher-order autocovariances even though the identification of $\boldsymbol{\theta}$ as a whole worsens.

One obvious shortcoming of the second method, besides the prohibitively high computational cost, is that it does not tell us which particular moments are most informative about a given parameter. The optimal selection includes at least as many moments as there are parameters and there is no obvious way of ranking them in terms of their importance. Furthermore, some of the moments in Table 10 are selected not because they are very informative about the parameter of interest, but because they help better identify other parameters and thus distinguish them from the parameter of interest. For these reasons the first order conditions approach is preferable if one wishes to know what parts of the moment structure are most informative about the individual parameters. The second approach, on the other hand, is useful for determining the optimal set of moments, of a given size, for $\boldsymbol{\theta}$ as a whole.

3.5 Distribution of information in the frequency domain

The efficiency properties of MLE discussed in section 3.2 reflect the total amount of information in the complete set of first and second order moments of the observed variables. In this section I examine how the information is distributed across different frequency bands. I use the fact that the likelihood function in (2.17), as well as the asymptotic information matrix, are constructed by summing up terms which are independent across frequencies. Therefore, to evaluate the amount of information within a

band of frequencies one has to include the terms corresponding to these frequencies and omit the ones outside the band. The inverse of the resulting information matrix is the asymptotic covariance matrix of the ML estimator based on the specified frequencies.

Three non-overlapping frequency bands are considered: low frequencies (with period from 32 quarters to infinity), business cycle (BC) frequencies (with period from 6 to 32 quarters), and high frequencies (with period between 2 and 6 quarters). Before turning to the results, it should be pointed out that two parameters - $\bar{a}$ and $\bar{b}$ are not identified unless information at frequency zero is utilized. To deal with that, I consider two cases: first, I fix the two unidentified parameters and compute the standard errors of the remaining parameters for the three frequency bands; second, I treat $\bar{a}$ and $\bar{b}$ as free but add frequency zero to the BC and high frequency bands. The results are shown in Table 11, where the standard errors based on all frequencies are shown as well. Note that in the lower panel of Table 11 the standard errors computed using the frequency domain approximation of the information matrix for all frequencies are either identical or very close to those from the time domain calculations.[6]

The results show that the information about the parameters is concentrated in the lower frequencies. The standard errors of all parameters except σ_a are smallest in the low frequency band and largest in the high frequency band. For σ_a the business cycle frequencies are slightly more informative than the low frequencies. The low frequencies are extremely important for ρ_v and σ_v, whose standard errors in the BC frequencies are several hundred times larger than in the low frequencies. Comparing the upper and the lower panels of Table 11 shows that frequency zero in particular is very important for β and δ, whose standard errors in the BC frequency band changes dramatically just by adding frequency zero. The remaining parameters benefit much less from information in frequency zero, while for ρ_v and σ_v there is almost no change when frequency zero is included.

[insert Table 11 here]

As the results show, there may be a considerable loss in efficiency if estimation is based on information from only a subset of frequencies, even if the most informative part of the spectrum is used. For more than half of the parameters the standard errors in the low frequencies are about twice as large as the standard errors with all frequencies. The loss is greatest for ν, σ_a and σ_b. On the other hand, there are parameters like ρ_v and σ_v for which almost all of the information is concentrated in the low frequencies and the loss from using only those frequencies is relatively small - 9% and 6% respectively.

The dominant role of the lower part of the spectrum is in part explained by the fact that the shocks driving the model are highly persistent.[7] As a result the observed variables are also very persistent and there is relatively little information in the higher

[6]The differences are due to the approximation of an integral by a finite sum of frequencies.

[7]Another reason is that the steady state relationships are very informative for some parameters, particularly for α, β and δ.

end of the spectrum. This information deficiency manifests itself in two ways: (1) there is little information about individual parameters because they do not affect much the behavior of the variables in those frequencies; (2) there is little information to help distinguish one parameter from the others because they affect in similar ways the properties of the variables in a given band of frequencies. The contribution of each one of these two factors can be isolated using the decomposition of the asymptotic standard errors shown in equation (3.4) and reported below for convenience.

$$s.e.(\theta_i) = s.e.(\theta_i|\boldsymbol{\theta}_{-i}) \left(\frac{1}{\sqrt{1-\boldsymbol{\varrho}_i^2}} \right)$$

As explained earlier, the first term measures the estimation uncertainty about θ_i if the remaining parameters were known; it is straightforward to show that it is inversely related to the sensitivity of the objective function, here the log-likelihood, with respect to θ_i. The second term captures the effect of having other unknown parameters on the uncertainty about θ_i; it is a function of $\boldsymbol{\varrho}_i$ which measures the degree of collinearity between the derivative of the log-likelihood with respect to θ_i and the derivatives with respect to the other elements of $\boldsymbol{\theta}$.

Example (continued) The decomposition in (3.4) *may be used to shed light on the frequency-domain results concerning the parameters of the ARMA(1,1) model shown in Table 2. Specifically, the very large standard errors in the low frequencies under DGP 2 can be explained with the particularly strong pairwise collinearity of both ϕ_1 and ϕ_2 with σ in those frequencies. This is shown in Table 12. Similarly, ϕ_2 and σ are highly collinear in the lower part of the spectrum under DGP 1, resulting in large standard errors. Fixing σ greatly reduces the standard errors of ϕ_1 and ϕ_2, although in DGP 2 they are still very large in the low frequencies, partly because ϕ_1 and ϕ_2 are highly collinear with each other in those frequencies. As the table shows, there is no correlation of ϕ_1 and ϕ_2 with σ when all frequencies are accounted for, which explains why the standard errors ϕ_1 and ϕ_2 are the same whether σ is fixed or not in the last column of Table 2.*

[insert Table 12 here]

Returning to the DSGE model, to evaluate the contribution of the two factors in (3.4) to the relative inefficiency resulting from using only a restricted part of the spectrum, I compute the ratio of each term in the decomposition for a given frequency band relative to the same term with all frequencies. The results for the case when $\bar{a}$ and $\bar{b}$ are fixed are reported in Table 13. Starting with the first term in the decomposition, the results show that, conditional on knowing the remaining parameters, the low frequencies are still the most informative about α, β and δ, but it is now the high frequencies that contain most information about all other parameters. This implies that the higher frequencies are more strongly affected by all except the first three parameters. However, as the second terms of the decomposition indicate, these effects are not very distinct in the business cycle and especially in the high frequencies, leading to a much worse overall identification

of the parameters in that part of the spectrum. For all parameters without exception the degree of collinearity is strongest in the high end of the spectrum and weakest in the low frequencies. Proceeding as in Section 3.2, with analysis of the most strongly collinear subsets of parameters, shows that it is only the degree of collinearity that changes, while the composition of the groups of parameters involved is mostly constant across frequency bands and is also very similar to what is shown in Table 7. For instance, the set of four parameters most highly collinear with ν is always that of α, δ, ρ_b and σ_b; the degree of collinearity changes from .67 with the full spectrum, to .26, .93 and .95 in the low, business cycle and high frequencies, respectively.[8] As found in Section 3.2, the large values of $\boldsymbol{\varrho}_i$ for ρ_v and σ_v are almost entirely due to the strong pairwise collinearity between these two parameters, which varies between .97 in the low frequencies, .9999998 in the business cycle frequencies, .9999999998 in the high frequencies, and .998 with the full spectrum. Detailed information about the collinearity patterns for all parameters and frequency bands is provided in the Appendix.

[insert Table 13 here]

Before concluding this section, it is worth emphasizing that the results that have been presented are conditional on, (1) the values assigned to the parameters and, (2) the set of observables. While the values shown in Table 3 are quite standard, very different values for some of the parameters can also be found in the literature. For instance, as discussed in Ríos-Rull et al. (2012), there is little consensus on the value of the Frisch elasticity of labor supply ν, with .38 being on the lower end of the estimates reported in the literature. Increasing the value of ν while keeping the other parameters unchanged does not affect much the properties of the model in terms of which moments and frequencies are most informative, how quickly the GMM standard errors converge to the ML ones, the collinearity among the parameters, etc. The only noticeable effect is on values of the asymptotic standard errors, in particular of β, δ and ρ_a, which decrease by between 10% and 20%, and of $\bar{b}$ and σ_a, which increase by 10% and 30% when ν is equal to 1.5 instead of 0.38. Regarding the set of observed variables, in addition to output, consumption and hours, one could treat as observed also investment (i), the real wage (w) and the real interest rate (r), the latter two represented in the model by the marginal products of labor and capital, respectively. This gives us six variables in total out of which we can form 20 different sets of three variables. However, many of these combinations can be ruled out either because $\boldsymbol{\theta}$ cannot be identified, or because they give exactly the same results as alternative sets of variables.[9] This leaves five sets of observables, in addition to the one that has been analysed, namely (y,r,h), (r,c,h),

[8] The GMM results (see Table 7) are somewhat different due to the fact that $\bar{a}$ and $\bar{b}$ were treated as free.)

[9] Examples of the first group are (y,c,i), which are clearly collinear, and (c,i,r), which do not contain enough information. Examples of the second group are (y,c,h), (y,i,h), (i,c,h), which yield the same standard errors.

(w,c,h), (w,i,h), and (w,c,r). The standard errors obtained with these sets of variables are, in some cases, quite different and neither one dominates in the sense of providing the smallest standard errors for all parameters. In fact, each one of them is optimal for at least one of the parameters. In terms of minimizing the uncertainty about $\boldsymbol{\theta}$ as a whole, as measured by $\ln(|\boldsymbol{V}^{\boldsymbol{\theta}}|)$, the most informative set of variables is (w,i,h) for both GMM and ML estimators; if instead the focus is only on the deep parameters (α, β, δ and ν), the best set would be (y,c,h). Apart from these differences, the results are generally similar in that the lower frequencies are most informative irrespectively of the variables, and the GMM standard errors converge fairly quickly to the MLE standard errors. The only exceptions worth mentioning are that the GMM standard errors for β, δ and $\bar{b}$ are noticeably slower to converge when either (r,c,h) or (r,c,w) are observed, and both σ_a and σ_b are somewhat better identified in the business cycle frequencies when (y,r,h) is observed.

4 Concluding remarks

The purpose of this paper has been to present a formal framework for analyzing how information about the parameters of DSGE models is distributed in the moment structure of the models' variables. The main objective of such an analysis is to determine which dimensions of the data - moments in the time domain and bands of frequencies in the frequency domain, are most useful for identifying specific deep parameters or structural shocks. The application of the presented approach may be useful to researchers who are interested in estimating models and have to make a choice among different estimation approaches or sets of observables. It may also be helpful to researchers who want to better understand the relationship between the structural features present in their models and the statistical properties of the modeled variables.

A distinct feature of the presented approach is that it uses analytical tools only. It exploits the fact that, for Gaussian models, the asymptotic covariance matrices of maximum likelihood and moment-matching estimators are either available in closed form or are easy to approximate arbitrarily well. This makes it possible to compare the efficiency properties of different estimators directly, without having to simulate and estimate the models. Extending the analysis to non-Gaussian models and moments beyond the second order is in principle straightforward, but would inevitably be computationally more demanding to implement in practice.

References

Bartlett, M. (1955): *An introduction to stochastic processes: with special reference to methods and applications*. University Press.

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

Christiano, L. J., and M. Eichenbaum (1992): "Current Real-Business-Cycle Theories and Aggregate Labor-Market Fluctuations," *American Economic Review*, 82(3), 430–50.

Christiano, L. J., and R. J. Vigfusson (2003): "Maximum likelihood in the frequency domain: the importance of time-to-plan," *Journal of Monetary Economics*, 50(4), 789–815.

Dunsmuir, W. (1979): "A Central Limit Theorem for Parameter Estimation in Stationary Vector Time Series and its Application to Models for a Signal Observed with Noise," *The Annals of Statistics*, 7(3), pp. 490–506.

Fuhrer, J. C., and G. D. Rudebusch (2004): "Estimating the Euler equation for output," *Journal of Monetary Economics*, 51(6), 1133 – 1153.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–54, available at http://ideas.repec.org/a/ecm/emetrp/v50y1982i4p1029-54.html.

Hansen, L. P., and T. J. Sargent (1980): "Formulating and estimating dynamic linear rational expectations models," *Journal of Economic Dynamics and Control*, 2(1), 7–46.

Iskrev, N. (2008): "Evaluating the information matrix in linearized DSGE models," *Economics Letters*, 99(3), 607–610.

——— (2010a): "Evaluating the strength of identification in DSGE models. An a priori approach," unpublished manuscript.

——— (2010b): "Local identification in DSGE models," *Journal of Monetary Economics*, 57(2), 189–202.

Jondeau, E., H. L. Bihan, and C. Galles (2004): "Assessing Generalized Method-of-Moments Estimates of the Federal Reserve Reaction Function," *Journal of Business & Economic Statistics*, 22, 225–239.

Klein, A., and H. Neudecker (2000): "A direct derivation of the exact Fisher information matrix of Gaussian vector state space models," *Linear Algebra and its Applications*, 321(1-3), 233–238.

Komunjer, I., and S. Ng (2011): "Dynamic Identification of Dynamic Stochastic General Equilibrium Models," *Econometrica*, 79(6), 1995–2032.

Lund, R., H. Bassily, and B. Vidakovic (2009): "Testing equality of stationary autocovariances," *Journal of Time Series Analysis*, 30(3), 332–348.

Müller, U. K. (2012): "Measuring prior sensitivity and prior informativeness in large Bayesian models," *Journal of Monetary Economics*, 59(6), 581 – 597.

Qu, Z., and D. Tkachenko (2012): "Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models," *Quantitative Economics*, 3(1), 95–132.

Ríos-Rull, J.-V., F. Schorfheide, C. Fuentes-Albero, M. Kryshko, and R. Santaeulàlia-Llopis (2012): "Methods versus Substance: Measuring the Effects of Technology Shocks on Hours," *Journal of Monetary Economics, forthcoming*.

Rotemberg, J., and M. Woodford (1997): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," in *NBER Macroeconomics Annual 1997, Volume 12*, NBER Chapters, pp. 297–361. National Bureau of Economic Research, Inc.

Rothenberg, T. J. (1971): "Identification in Parametric Models," *Econometrica*, 39(3), 577–91, available at http://ideas.repec.org/a/ecm/emetrp/v39y1971i3p577-91.html.

Ruge-Murcia, F. J. (2007): "Methods to estimate dynamic stochastic general equilibrium models," *Journal of Economic Dynamics and Control*, 31(8), 2599–2636.

Segal, M., and E. Weinstein (1989): "A new method for evaluating the log-likelihood gradient, the Hessian, and the Fisher information matrix for linear dynamic systems.," *IEEE Transactions on Information Theory*, 35(3), 682–.

Smith, A A, J. (1993): "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8(S), S63–84.

West, K. D. (1986): "Full-versus limited-information estimation of a rational-expectations model: Some numerical comparisons," *Journal of Econometrics*, 33(3), 367–385.

Whittle, P. (1953): "The Analysis of Multiple Stationary Time Series," *Journal of the Royal Statistical Society. Series B (Methodological)*, 15(1), pp. 125–139.

Zadrozny, P. A. (1989): "Analytic derivatives for estimation of linear dynamic models," *Computers and Mathematics with Applications*, 18(6-7), 539–553.

Zadrozny, P. A., and S. Mittnik (1994): "Kalman-Filtering Methods for Computing Information Matrices for Time-Invariant, Periodic, and Generally Time-Varying VARMA Models and Samples," *Computers and Mathematics with Applications*, 28, 107–119.

Tables

Table 1: ARMA(1,1), asymptotic standard errors for GMM with optimal $\boldsymbol{W}$

	DGP 1: $\phi_1 = .9$, $\phi_2 = .1$, $\sigma = 1$								
	2	3	4	5	6	7	8	9	10
ϕ_1	0.0477	0.0475	0.0475	0.0475	0.0475	0.0475	0.0475	0.0475	0.0475
ϕ_2	0.1107	0.1085	0.1085	0.1085	0.1085	0.1085	0.1085	0.1085	0.1085
σ	0.0707	0.0707	0.0707	0.0707	0.0707	0.0707	0.0707	0.0707	0.0707
	DGP 2: $\phi_1 = .1$, $\phi_2 = .5$, $\sigma = 1$								
	2	3	4	5	6	7	8	9	10
ϕ_1	0.2404	0.1958	0.1821	0.1771	0.1752	0.1745	0.1742	0.1742	0.1741
ϕ_2	0.2695	0.1883	0.1645	0.1562	0.1532	0.1521	0.1517	0.1516	0.1516
σ	0.0738	0.0711	0.0708	0.0707	0.0707	0.0707	0.0707	0.0707	0.0707

Note: The set of moments includes the variance and first order autocovariaqnce plus the autocovariances of the indicated order. The standard errors are normalized for $T = 100$.

Table 2: ARMA(1,1), asymptotic standard errors within frequency bands

DGP 1: $\phi_1 = .9$, $\phi_2 = .1$, $\sigma = 1$								
	low	BC	high	all	low	BC	high	all
ϕ_1	0.2886	0.4072	9.7245	0.0475	0.1053	0.2686	0.2179	0.0475
ϕ_2	202.4	1.9745	0.3965	0.1085	0.7204	0.1776	0.1663	0.1085
σ	181.6	1.5140	4.7925	0.0707	fixed	fixed	fixed	fixed
DGP 2: $\phi_1 = .1$, $\phi_2 = .5$, $\sigma = 1$								
	low	BC	high	all	low	BC	high	all
ϕ_1	9765	7.8548	0.4110	0.1741	39.541	0.8653	0.2479	0.1741
ϕ_2	98341	56.472	0.2038	0.1514	65.427	1.1478	0.1665	0.1514
σ	54710	29.227	0.1537	0.0707	fixed	fixed	fixed	fixed

Note: The standard errors are computed using frequency domain approximation of the asymptotic information matrix and are normalized for $T = 100$. The frequency bands are: $[0, \pi/16)$ (low), $[\pi/16, \pi/3]$ (business cycle - BC), $(\pi/3, \pi]$ (high), and $[0, \pi]$ (all). The standard errors are normalized for $T = 100$.

Table 3: Parameter values

parameter	value	description
α	0.640	labor share of output
β	0.982	discount factor
δ	0.025	depreciation rate
ν	0.380	labor supply elasticity
$\bar{a}$	0.600	steady state TFP shock
$\bar{b}$	-1.490	steady state labor supply shock
ρ_a	0.970	AR TFP shock
ρ_b	0.980	AR labor supply shock
ρ_v	0.990	AR preference shock
σ_a	0.006	std. dev. TFP shock
σ_b	0.007	std. dev. labor supply shock
σ_v	0.019	std. dev. preference shock

Table 4: Asymptotic standard errors with optimal $\boldsymbol{W}$

	lags						
param.	1	2	3	4	5	20	∞
α	0.103	0.097	0.094	0.093	0.092	0.088	0.087
β	0.017	0.015	0.013	0.013	0.012	0.011	0.011
δ	0.012	0.010	$9.15e{-3}$	$8.61e{-3}$	$8.26e{-3}$	$6.99e{-3}$	$6.73e{-3}$
ν	0.462	0.437	0.426	0.420	0.416	0.402	0.398
$\bar{a}$	0.406	0.402	0.400	0.399	0.399	0.396	0.394
$\bar{b}$	0.096	0.091	0.088	0.087	0.086	0.083	0.082
ρ_a	0.018	0.016	0.016	0.015	0.015	0.014	0.014
ρ_b	$1.81e{-2}$	$1.81e{-2}$	$1.81e{-2}$	$1.80e{-2}$	$1.80e{-2}$	$1.80e{-2}$	$1.79e{-2}$
ρ_v	0.027	0.022	0.020	0.019	0.018	0.015	0.014
σ_a	$4.40e{-4}$	$4.38e{-4}$	$4.37e{-4}$	$4.37e{-4}$	$4.37e{-4}$	$4.36e{-4}$	$4.36e{-4}$
σ_b	$8.44e{-4}$	$8.15e{-4}$	$8.02e{-4}$	$7.94e{-4}$	$7.90e{-4}$	$7.73e{-4}$	$7.69e{-4}$
σ_v	0.051	0.042	0.038	0.035	0.033	0.027	0.025

Note: The set of moments includes the means, the contemporaneous covariances, and the autocovariances of the indicated order. The standard errors are normalized for $T = 100$. $\boldsymbol{W}$ is the weighting matrix in the GMM objective function. The standard errors for MLE (lags $= \infty$) are computed using the asymptotic information matrix.

Table 5: Asymptotic standard errors assuming one free parameter

	lags						
param.	1	2	3	4	5	20	∞
α	$4.54e{-3}$	$4.54e{-3}$	$4.54e{-3}$	$4.54e{-3}$	$4.54e{-3}$	$4.54e{-3}$	$4.54e{-3}$
β	$1.35e{-3}$	$1.35e{-3}$	$1.35e{-3}$	$1.35e{-3}$	$1.35e{-3}$	$1.34e{-3}$	$1.34e{-3}$
δ	$1.08e{-3}$	$1.08e{-3}$	$1.08e{-3}$	$1.08e{-3}$	$1.08e{-3}$	$1.07e{-3}$	$1.07e{-3}$
ν	0.162	0.162	0.162	0.162	0.162	0.162	0.162
$\bar{a}$	0.020	0.020	0.020	0.020	0.020	0.020	0.020
$\bar{b}$	0.035	0.035	0.035	0.035	0.035	0.035	0.035
ρ_a	$9.48e{-3}$	$9.48e{-3}$	$9.48e{-3}$	$9.47e{-3}$	$9.47e{-3}$	$9.43e{-3}$	$9.40e{-3}$
ρ_b	$9.64e{-3}$	$9.64e{-3}$	$9.64e{-3}$	$9.64e{-3}$	$9.64e{-3}$	$9.64e{-3}$	$9.63e{-3}$
ρ_v	$8.26e{-4}$	$8.24e{-4}$	$8.23e{-4}$	$8.22e{-4}$	$8.21e{-4}$	$8.15e{-4}$	$8.12e{-4}$
σ_a	$4.24e{-4}$	$4.24e{-4}$	$4.24e{-4}$	$4.24e{-4}$	$4.24e{-4}$	$4.24e{-4}$	$4.24e{-4}$
σ_b	$4.95e{-4}$	$4.95e{-4}$	$4.95e{-4}$	$4.95e{-4}$	$4.95e{-4}$	$4.95e{-4}$	$4.95e{-4}$
σ_v	$1.36e{-3}$	$1.36e{-3}$	$1.36e{-3}$	$1.35e{-3}$	$1.35e{-3}$	$1.35e{-3}$	$1.34e{-3}$

Note: The standard errors for each parameter are computed assuming all other parameters are known.

Table 6: Collinearity components of the asymptotic standard errors

	lags						
param.	1	2	3	4	5	20	∞
α	0.9990	0.9989	0.9988	0.9988	0.9988	0.9987	0.9986
β	0.9967	0.9957	0.9950	0.9945	0.9941	0.9925	0.9920
δ	0.9959	0.9942	0.9930	0.9922	0.9915	0.9881	0.9872
ν	0.9370	0.9292	0.9253	0.9230	0.9214	0.9155	0.9138
$\bar{a}$	0.9988	0.9988	0.9988	0.9988	0.9988	0.9987	0.9987
$\bar{b}$	0.9320	0.9234	0.9191	0.9164	0.9147	0.9081	0.9061
ρ_a	0.8474	0.8178	0.8013	0.7907	0.7833	0.7542	0.7472
ρ_b	0.8462	0.8457	0.8455	0.8453	0.8452	0.8441	0.8430
ρ_v	0.9995	0.9993	0.9992	0.9991	0.9990	0.9986	0.9984
σ_a	0.2647	0.2500	0.2434	0.2396	0.2371	0.2285	0.2259
σ_b	0.8102	0.7943	0.7866	0.7821	0.7791	0.7684	0.7654
σ_v	0.9996	0.9995	0.9994	0.9993	0.9992	0.9987	0.9985

Note: The table shows the values of $\boldsymbol{\varrho}_i(q)$ from factorization (3.4) of the standard errors in Table 4.

Table 7: Strongly collinear subsets of parameters

param.	(1)	(2)	(3)	(4)
α	0.930 ($\bar{a}$)	0.9926 (β,$\bar{a}$)	0.9955 (β,$\bar{a}$,σ_v)	0.9971 (β,$\bar{a}$,$\bar{b}$,σ_v)
β	0.538 (σ_v)	0.9445 (α,$\bar{a}$)	0.9733 (α,$\bar{a}$,σ_v)	0.9816 (α,δ,$\bar{a}$,σ_v)
δ	0.844 ($\bar{a}$)	0.8994 (α,$\bar{a}$)	0.9753 (β,$\bar{a}$,σ_v)	0.9876 (β,$\bar{a}$,ρ_a,σ_v)
ν	0.515 ($\bar{b}$)	0.7060 ($\bar{b}$,σ_b)	0.8239 ($\bar{b}$,ρ_b,σ_b)	0.8389 ($\bar{b}$,ρ_a,ρ_b,σ_b)
$\bar{a}$	0.930 (α)	0.9923 (α,β)	0.9954 (α,β,δ)	0.9970 (α,β,$\bar{b}$,σ_v)
$\bar{b}$	0.515 (ν)	0.5880 (ν,σ_b)	0.6725 (ν,ρ_b,σ_b)	0.6872 (ν,ρ_a,ρ_b,σ_b)
ρ_a	0.156 (ν)	0.1862 (δ,$\bar{a}$)	0.3575 (α,δ,$\bar{a}$)	0.7104 (β,δ,$\bar{a}$,σ_v)
ρ_b	0.424 (ν)	0.4956 (ν,$\bar{b}$)	0.5999 (ν,$\bar{b}$,σ_b)	0.6153 (ν,$\bar{b}$,ρ_a,σ_b)
ρ_v	0.998 (σ_v)	0.9989 (β,σ_v)	0.9989 (β,δ,σ_v)	0.9991 (α,β,δ,σ_v)
σ_a	0.012 (α)	0.0330 (α,$\bar{a}$)	0.0997 (α,β,$\bar{a}$)	0.1275 (α,β,$\bar{a}$,σ_v)
σ_b	0.483 (ν)	0.5634 (ν,$\bar{b}$)	0.6487 (ν,$\bar{b}$,ρ_b)	0.6638 (ν,$\bar{b}$,ρ_a,ρ_b)
σ_v	0.998 (ρ_v)	0.9989 (β,ρ_v)	0.9989 (β,δ,ρ_v)	0.9992 (α,β,δ,ρ_v)

Note: The table shows groups of one to four parameters whose collinearity with the parameters in the first column is strongest among all possible groups of that size. The results are based on GMM with $q = 1$.

Table 8: Asymptotic standard errors with suboptimal $\boldsymbol{W}$

	lags					
param.	1	2	3	4	5	20
			diagonal			
α	0.733	0.745	0.751	0.756	0.760	0.823
β	0.058	0.059	0.060	0.061	0.061	0.062
δ	0.050	0.050	0.051	0.051	0.052	0.065
ν	3.453	3.522	3.601	3.678	3.753	4.811
$\bar{a}$	3.759	3.808	3.834	3.859	3.885	4.317
$\bar{b}$	0.334	0.336	0.339	0.342	0.345	0.424
ρ_a	0.106	0.107	0.108	0.109	0.110	0.137
ρ_b	0.043	0.043	0.044	0.045	0.046	0.050
ρ_v	0.043	0.043	0.044	0.045	0.045	0.065
σ_a	0.007	0.007	0.007	0.007	0.007	0.012
σ_b	0.008	0.008	0.009	0.009	0.009	0.011
σ_v	0.072	0.070	0.070	0.070	0.070	0.075
			identity			
α	0.728	0.651	0.620	0.620	0.633	0.758
β	0.063	0.052	0.049	0.052	0.053	0.066
δ	0.043	0.041	0.043	0.043	0.043	0.044
ν	2.826	3.939	3.869	3.964	4.045	4.976
$\bar{a}$	3.617	2.810	3.096	3.156	3.204	3.748
$\bar{b}$	fixed	0.283	0.316	0.343	0.348	0.432
ρ_a	0.104	0.085	0.085	0.087	0.088	0.115
ρ_b	0.041	0.049	0.053	0.054	0.055	0.060
ρ_v	0.039	0.044	0.044	0.045	0.046	0.065
σ_a	$5.65e{-3}$	$5.25e{-3}$	$5.27e{-3}$	$5.35e{-3}$	$5.48e{-3}$	$8.22e{-3}$
σ_b	0.007	0.009	0.011	0.011	0.011	0.013
σ_v	0.059	0.062	0.061	0.061	0.061	0.058

Note: The set of moments includes the means, the contemporaneous covariances, and the autocovariances of the indicated order. The standard errors are normalized for $T = 100$. $\boldsymbol{W}$ is the weighting matrix in the GMM objective function.

Table 9: Distribution of weight across variables, optimal $\boldsymbol{W}$

	own moments only			cross moments only		
param.	y	h	c	(y,h)	(y,c)	(h,c)
α	0.13	0.04	0.38	0.02	0.42	0.02
β	0.11	0.00	0.46	0.00	0.42	0.01
δ	0.13	0.01	0.42	0.01	0.42	0.01
ν	0.03	0.01	0.02	0.43	0.18	0.33
$\bar{a}$	0.27	0.46	0.27	0.00	0.00	0.00
$\bar{b}$	0.19	0.70	0.11	0.00	0.00	0.00
ρ_a	0.24	0.00	0.01	0.07	0.53	0.15
ρ_b	0.08	0.02	0.18	0.14	0.26	0.31
ρ_v	0.09	0.00	0.50	0.01	0.39	0.02
σ_a	0.55	0.10	0.00	0.32	0.02	0.00
σ_b	0.06	0.40	0.05	0.21	0.11	0.17
σ_v	0.09	0.00	0.50	0.01	0.39	0.02

Note: The table shows the relative weight on moments of the three variable in the first order conditions of GMM with $q = 20$.

Table 10: Standard errors with optimal sets of 12 moments

param.	$\boldsymbol{m}_{\boldsymbol{\theta}}$	$\boldsymbol{m}_{\alpha}$	$\boldsymbol{m}_{\beta}$	$\boldsymbol{m}_{\delta}$	$\boldsymbol{m}_{\nu}$	$\boldsymbol{m}_{\bar{a}}$	$\boldsymbol{m}_{\bar{b}}$	$\boldsymbol{m}_{\rho_a}$	$\boldsymbol{m}_{\rho_b}$	$\boldsymbol{m}_{\rho_v}$	$\boldsymbol{m}_{\sigma_a}$	$\boldsymbol{m}_{\sigma_b}$	$\boldsymbol{m}_{\sigma_v}$
α	0.324	0.141	0.311	0.666	0.882	1.654	1.574	0.677	0.869	0.446	0.556	0.882	0.479
β	0.040	0.045	0.021	0.078	0.062	0.784	0.063	0.148	0.063	0.060	1.739	0.062	0.044
δ	0.034	0.060	0.058	0.012	0.084	1.014	0.185	0.183	0.081	0.041	1.221	0.084	0.052
ν	5.528	3.098	5.704	4.011	1.167	3.145	6.555	3.548	1.168	9.219	1.505	1.167	17.170
$\bar{a}$	1.552	1.043	2.081	2.834	4.807	0.486	9.318	3.314	4.706	1.943	7.765	4.807	2.563
$\bar{b}$	0.515	0.338	0.565	0.463	0.481	2.064	0.160	0.611	0.479	0.988	4.625	0.481	1.766
ρ_a	0.086	0.097	0.077	0.105	0.187	0.172	0.426	0.033	0.189	0.049	0.818	0.187	0.053
ρ_b	0.046	0.031	0.049	0.079	0.020	0.030	0.047	0.058	0.020	0.117	0.029	0.020	0.237
ρ_v	0.058	0.377	0.045	0.044	0.392	0.721	0.157	0.114	0.407	0.028	1.879	0.392	0.048
σ_a	$8.3e{-3}$	$1.7e{-2}$	$9.4e{-3}$	$4.8e{-3}$	$1.3e{-2}$	$1.1e{-2}$	$6.0e{-2}$	$5.1e{-3}$	$1.2e{-2}$	$8.4e{-3}$	$5.7e{-4}$	$1.3e{-2}$	$7.8e{-3}$
σ_b	$1.1e{-2}$	$6.0e{-3}$	$1.1e{-2}$	$1.5e{-2}$	$1.8e{-3}$	$5.0e{-3}$	$1.6e{-2}$	$1.4e{-2}$	$1.8e{-3}$	$2.9e{-2}$	$4.0e{-3}$	$1.8e{-3}$	$6.0e{-2}$
σ_v	0.116	0.677	0.100	0.059	0.551	1.426	0.535	0.234	0.571	0.078	2.925	0.551	0.052
$\ln(\lvert\boldsymbol{V}_{\boldsymbol{\theta}}\rvert)$	-102.7	-90.8	-93.8	-93.3	-89.0	-85.4	-86.8	-82.3	-89.0	-90.0	-85.3	-89.0	-89.1

Note: The moments are selected as the best 12 of the first 27 moments in terms of minimizing either $\ln(\lvert\boldsymbol{V}^{\boldsymbol{\theta}}\rvert)$ (in the first column) or the SE of the respective parameter. All sets include $E(h_t)$ along with the following moments:

$\boldsymbol{m}_{\boldsymbol{\theta}}$: $E(y_t)$, $E(c_t)$, $cov(y_t, y_t)$, $cov(c_t, y_t)$, $cov(h_t, h_t)$, $cov(c_t, c_t)$, $cov(y_t, y_{t+1})$, $cov(c_t, y_{t+1})$, $cov(h_t, h_{t+1})$, $cov(y_t, c_{t+1})$, $cov(c_t, c_{t+1})$

$\boldsymbol{m}_{\alpha}$: $E(y_t)$, $E(c_t)$, $cov(h_t, y_t)$, $cov(h_t, h_t)$, $cov(c_t, c_t)$, $cov(c_t, h_{t+1})$, $cov(c_t, c_{t+1})$, $cov(h_t, h_{t+2})$, $cov(c_t, h_{t+2})$, $cov(h_t, c_{t+2})$, $cov(c_t, c_{t+2})$

$\boldsymbol{m}_{\beta}$: $E(y_t)$, $E(c_t)$, $cov(y_t, y_t)$, $cov(c_t, y_t)$, $cov(h_t, h_t)$, $cov(c_t, y_{t+1})$, $cov(h_t, h_{t+1})$, $cov(h_t, y_{t+2})$, $cov(c_t, h_{t+2})$, $cov(y_t, c_{t+2})$, $cov(c_t, c_{t+2})$

$\boldsymbol{m}_{\delta}$: $E(y_t)$, $E(c_t)$, $cov(y_t, y_t)$, $cov(h_t, y_t)$, $cov(h_t, h_t)$, $cov(y_t, y_{t+1})$, $cov(h_t, y_{t+1})$, $cov(c_t, c_{t+1})$, $cov(c_t, y_{t+2})$, $cov(c_t, h_{t+2})$, $cov(c_t, c_{t+2})$

$\boldsymbol{m}_{\nu}$: $E(y_t)$, $E(c_t)$, $cov(h_t, y_t)$, $cov(h_t, h_t)$, $cov(c_t, h_t)$, $cov(h_t, y_{t+1})$, $cov(c_t, y_{t+1})$, $cov(y_t, h_{t+1})$, $cov(h_t, h_{t+1})$, $cov(c_t, y_{t+2})$, $cov(y_t, c_{t+2})$

$\boldsymbol{m}_{\bar{a}}$: $E(c_t)$, $cov(h_t, y_t)$, $cov(h_t, h_t)$, $cov(c_t, h_t)$, $cov(c_t, c_t)$, $cov(h_t, h_{t+1})$, $cov(c_t, h_{t+1})$, $cov(y_t, c_{t+1})$, $cov(h_t, c_{t+1})$, $cov(c_t, c_{t+1})$, $cov(y_t, c_{t+2})$

$\boldsymbol{m}_{\bar{b}}$: $E(y_t)$, $E(c_t)$, $cov(h_t, y_t)$, $cov(c_t, h_t)$, $cov(c_t, c_t)$, $cov(y_t, y_{t+1})$, $cov(h_t, y_{t+1})$, $cov(y_t, c_{t+1})$, $cov(h_t, c_{t+1})$, $cov(y_t, h_{t+2})$, $cov(c_t, h_{t+2})$

$\boldsymbol{m}_{\rho_a}$: $E(y_t)$, $cov(y_t, y_t)$, $cov(h_t, y_t)$, $cov(c_t, y_t)$, $cov(c_t, h_t)$, $cov(y_t, y_{t+1})$, $cov(c_t, y_{t+1})$, $cov(h_t, c_{t+1})$, $cov(h_t, y_{t+2})$, $cov(c_t, y_{t+2})$, $cov(y_t, c_{t+2})$

$\boldsymbol{m}_{\rho_b}$: $E(y_t)$, $E(c_t)$, $cov(h_t, y_t)$, $cov(h_t, h_t)$, $cov(c_t, h_t)$, $cov(h_t, y_{t+1})$, $cov(y_t, h_{t+1})$, $cov(h_t, h_{t+1})$, $cov(y_t, c_{t+1})$, $cov(c_t, y_{t+2})$, $cov(y_t, c_{t+2})$

$\boldsymbol{m}_{\rho_v}$: $E(y_t)$, $E(c_t)$, $cov(c_t, h_t)$, $cov(c_t, c_t)$, $cov(h_t, y_{t+1})$, $cov(h_t, c_{t+1})$, $cov(c_t, c_{t+1})$, $cov(y_t, y_{t+2})$, $cov(c_t, y_{t+2})$, $cov(y_t, c_{t+2})$, $cov(c_t, c_{t+2})$

$\boldsymbol{m}_{\sigma_a}$: $E(c_t)$, $cov(y_t, y_t)$, $cov(h_t, y_t)$, $cov(c_t, y_t)$, $cov(h_t, h_t)$, $cov(y_t, y_{t+1})$, $cov(h_t, y_{t+1})$, $cov(c_t, y_{t+1})$, $cov(y_t, h_{t+1})$, $cov(h_t, h_{t+1})$, $cov(y_t, c_{t+1})$

$\boldsymbol{m}_{\sigma_b}$: $E(y_t)$, $E(c_t)$, $cov(h_t, y_t)$, $cov(h_t, h_t)$, $cov(c_t, h_t)$, $cov(h_t, y_{t+1})$, $cov(c_t, y_{t+1})$, $cov(y_t, h_{t+1})$, $cov(h_t, h_{t+1})$, $cov(c_t, y_{t+2})$, $cov(y_t, c_{t+2})$

$\boldsymbol{m}_{\sigma_v}$: $E(y_t)$, $E(c_t)$, $cov(c_t, c_t)$, $cov(h_t, y_{t+1})$, $cov(c_t, y_{t+1})$, $cov(c_t, h_{t+1})$, $cov(c_t, c_{t+1})$, $cov(y_t, y_{t+2})$, $cov(y_t, c_{t+2})$, $cov(h_t, c_{t+2})$, $cov(c_t, c_{t+2})$

Table 11: Asymptotic standard errors within frequency bands

param.	low	BC	high	all
		$\bar{a}$ and $\bar{b}$ fixed		
α	0.026	0.930	3.387	0.020
β	$8.08e{-3}$	0.434	1.797	$6.33e{-3}$
δ	$7.49e{-3}$	0.431	1.760	$5.82e{-3}$
ν	0.317	3.878	13.892	0.217
$\bar{a}$	fixed	fixed	fixed	fixed
$\bar{b}$	fixed	fixed	fixed	fixed
ρ_a	0.027	0.165	0.601	0.012
ρ_b	0.022	0.275	1.029	0.012
ρ_v	0.015	2.568	45.194	0.014
σ_a	$2.00e{-3}$	$1.70e{-3}$	$5.71e{-3}$	$4.25e{-4}$
σ_b	$2.31e{-3}$	$5.16e{-3}$	0.018	$5.90e{-4}$
σ_v	0.025	4.267	75.114	0.024
		frequency zero always included		
α	0.186	0.893	3.099	0.087
β	0.021	0.079	0.273	0.011
δ	$8.60e{-3}$	0.039	0.134	$6.73e{-3}$
ν	1.268	3.724	12.833	0.398
$\bar{a}$	0.835	4.329	15.002	0.394
$\bar{b}$	0.169	0.795	2.755	0.082
ρ_a	0.028	0.133	0.446	0.014
ρ_b	0.024	0.225	0.772	0.018
ρ_v	0.016	2.568	45.193	0.014
σ_a	$2.02e{-3}$	$1.62e{-3}$	$5.29e{-3}$	$4.35e{-4}$
σ_b	$2.94e{-3}$	$5.03e{-3}$	0.017	$7.69e{-4}$
σ_v	0.027	4.266	75.114	0.025

Note: The standard errors are computed using frequency domain approximation of the asymptotic information matrix and are normalized for $T = 100$. The frequency bands are: $[0, \pi/16)$ (low), $[\pi/16, \pi/3]$ (business cycle - BC), $(\pi/3, \pi]$ (high), and $[0, \pi]$ (all). $\bar{a}$ and $\bar{b}$ are identified only when frequency zero is included.

Table 12: ARMA(1,1), collinearity with σ

	DGP 1: $\phi_1 = .9$, $\phi_2 = .1$, $\sigma = 1$				DGP 2: $\phi_1 = .1$, $\phi_2 = .5$, $\sigma = 1$			
	low	BC	high	all	low	BC	high	all
ϕ_1	0.8989376	0.065288	0.9991	0	0.9999697	0.973471	0.7243	0
ϕ_2	0.9999909	0.990694	0.5908	0	0.9999992	0.999109	0.3569	0

Note: The table shows the pairwise correlation coefficients of ϕ_1 and ϕ_2 with σ.

Table 13: Decomposition of the relative inefficiency ($\bar{a}$ and $\bar{b}$ fixed)

param.	low	BC	high
α	1.3 = 1.0 × 1.3	46 = 12.0 × 3.8	169 = 7.9 × 21
β	1.3 = 1.2 × 1.1	69 = 3.5 × 20	284 = 2.3 × 124
δ	1.3 = 1.0 × 1.2	74 = 6.8 × 11	303 = 4.7 × 65
ν	1.5 = 1.8 × 0.8	18 = 2.2 × 8.0	64 = 1.4 × 45
ρ_a	2.3 = 2.3 × 1.0	14 = 2.1 × 6.6	50 = 1.3 × 38
ρ_b	1.9 = 1.9 × 1.0	24 = 2.2 × 11	89 = 1.4 × 64
ρ_v	1.1 = 4.3 × 0.3	181 = 1.9 × 94	3180 = 1.2 × 2605
σ_a	4.7 = 4.0 × 1.2	4.0 = 1.9 × 2.1	13 = 1.2 × 11
σ_b	3.9 = 4.0 × 1.0	8.8 = 1.9 × 4.6	30 = 1.2 × 24
σ_v	1.0 = 4.0 × 0.3	175 = 1.9 × 91	3072 = 1.2 × 2507

Note: The asymptotic standard errors are decomposed as (see equation (3.4))

$$s.e.(\theta_i) = s.e.(\theta_i|\boldsymbol{\theta}_{-i}) \times (1/\sqrt{1-\boldsymbol{\varrho}_i^2})$$

The table shows the ratio of each term for the particular frequency band relative to its value using the full spectrum.

Figures

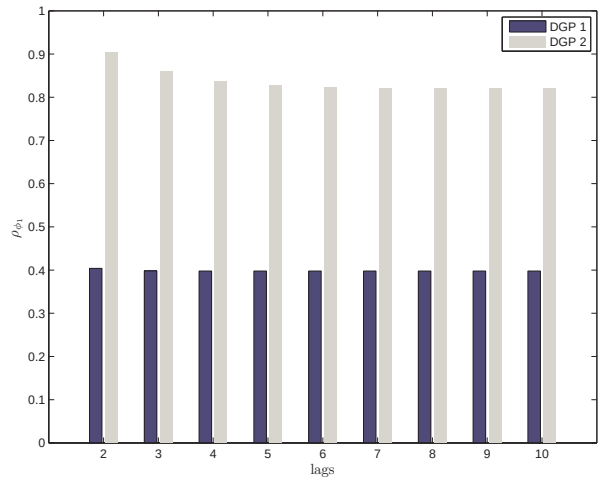

Figure 1: ARMA(1,1) model, collinearity coefficient for ϕ_1.

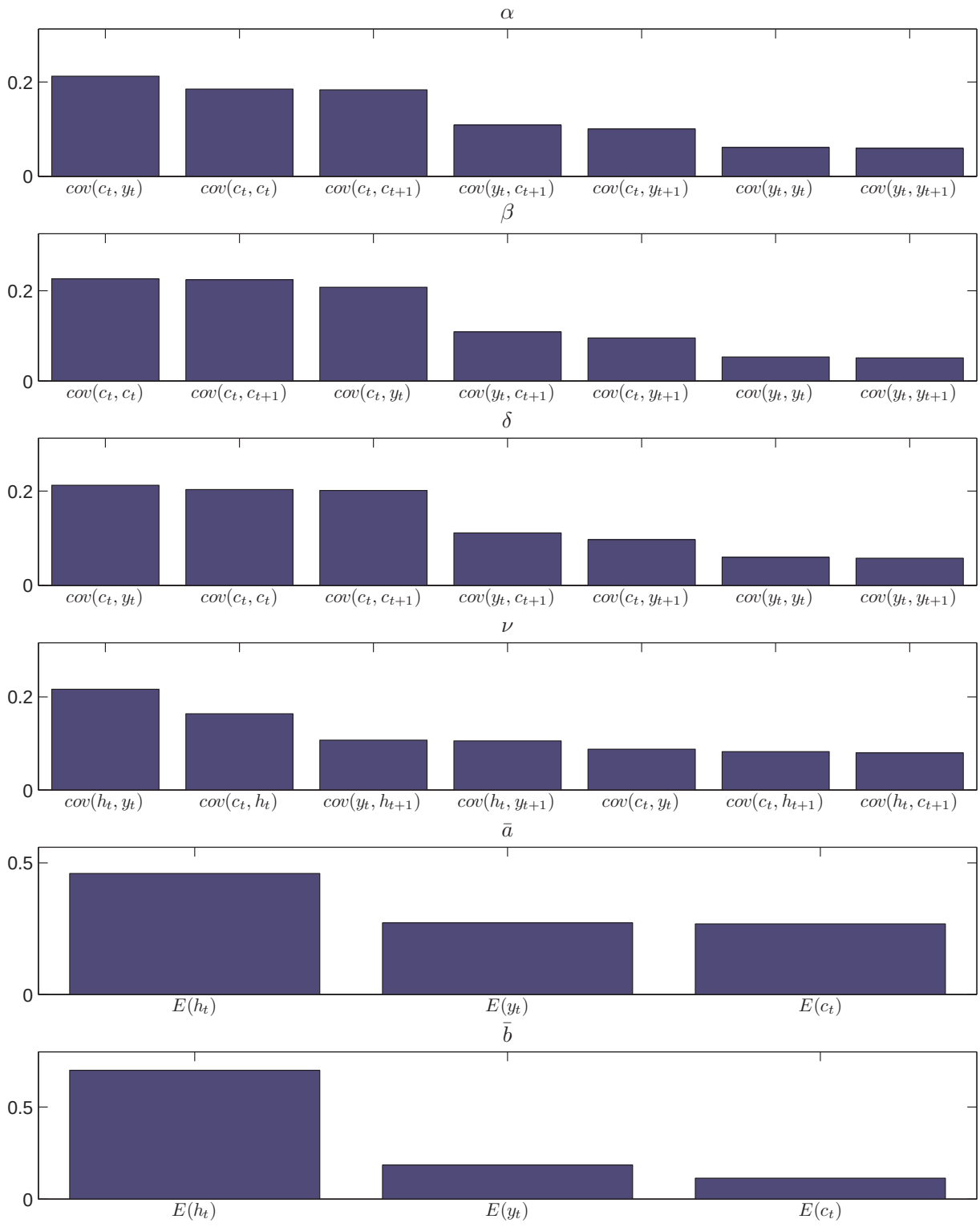

Figure 2: Distribution of weights in the first order conditions, optimal $\boldsymbol{W}$.

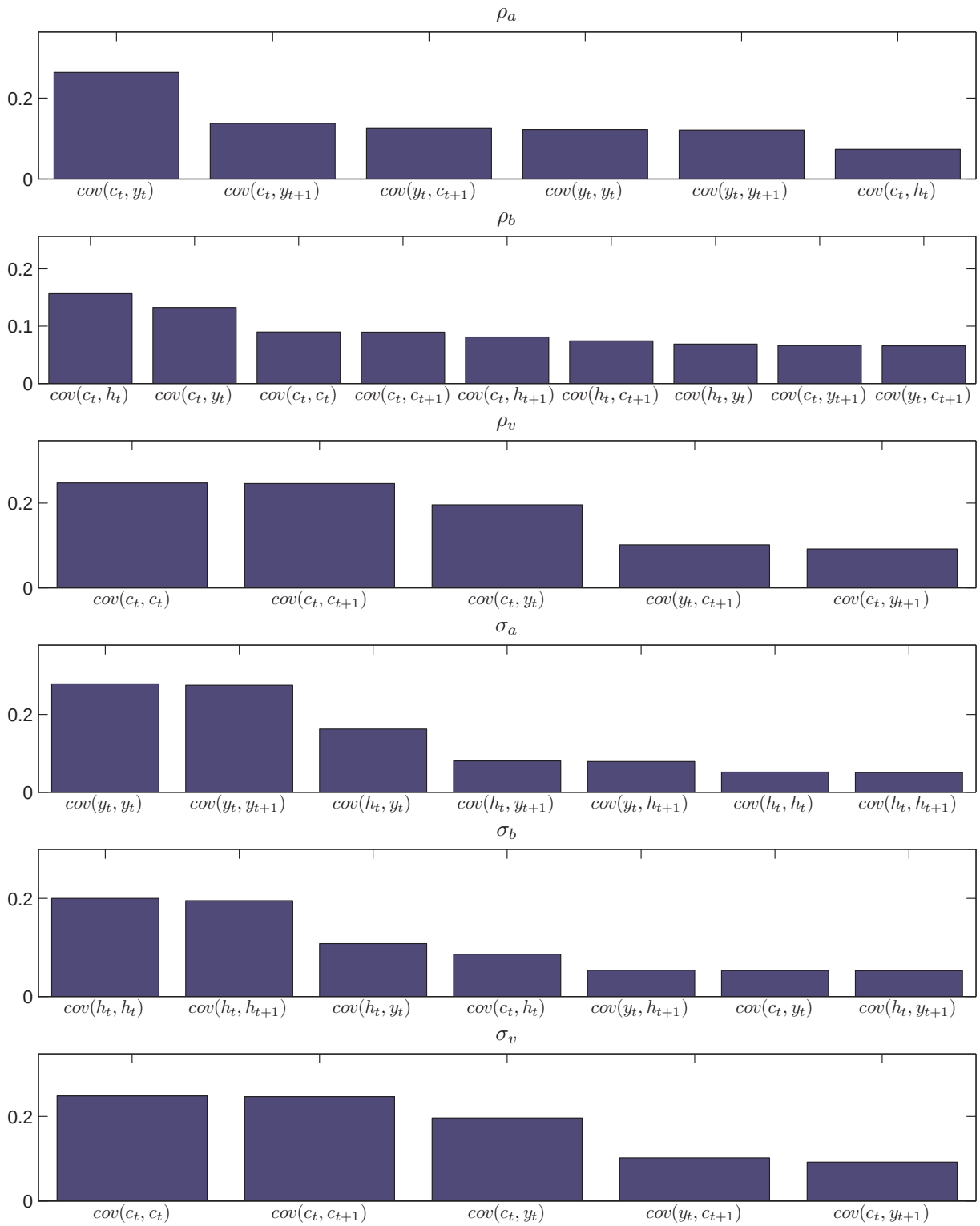

Figure 3: Distribution of weights in the first order conditions, optimal $\boldsymbol{W}$.