OUP Handbook on Computational Economics and Finance

Shu-Heng Chen
Department of Economics, National Chengchi University, Taipei, Taiwan 116

Mak Kaboudan
School of Business, University of Redlands, Redlands, CA, USA 92373

Ye-Rong Du
Department of Economics, National Chengchi University, Taipei, Taiwan 116

OXFORD
UNIVERSITY PRESS

1

Dynamic stochastic general equilibrium models: A computational perspective

Michel Juillard

Bank of France, 31 rue Croix des Petits Champs, 75001 Paris

michel.juillard@mjui.fr

ABSTRACT

DSGE models have become popular in macroeconomics but the combination of nonlinear microeconomic behavior of the agents and model consistent expectations raise intricate computational issues. This chapter reviews solution methods and estimation of DSGE models. Perfect foresight deterministic models can easily be solved with a great degree of accuracy. In practice, medium size stochastic models can only be solved by local approximation or perturbation approach. The Bayesian approach to estimation is privileged. It provides a convenient way to communicate both the prior information available to the econometrician and new information revealed by the data. This chapter focuses on methods frequently used in applied work rather than aiming at being exhaustive.

1.1 Introduction

Dynamic stochastic general equilibrium models (DSGE) have become very popular in applied macroeconomics, both in academics and in policy institutions. This chapter reviews the methods that are currently used to solve them and estimate their parameters.

DSGE modeling adopts the methodology developed by Kydland and Prescott (1982), at the origin of real business cycle analysis. It views macroeconomic models as simplified representations of reality built on behavior of representative agents that is consistent with microeconomic theory. Nowadays, DSGE models take into account a large number of frictions, real or nominal, borrowed from New Keynesian economics.

Depending on the aim of the model, nominal rigidities and money transmission mechanisms, labor market, fiscal policy or open economy aspects are emphasized and the corresponding mechanisms developed.

However, whatever the focus of a particular model, common features are present that define a class of models for which a particular methodology has been developed.

In this class of model, most agents (households, firms, financial intermediaries,...) take their decisions while considering inter-temporal objective functions: life-time welfare for households, investment value for the firms, and so on. Agents must therefore solve dynamic optimization problems.

Given the specification of utility and technological constraints usually used in microeconomics the resulting models are non-linear. Because of the necessity to solve dynamic optimization problems, future values of some variables matter for current decisions. In a stochastic world, these future values are unknown and agents must form expectations about the future. The hypothesis of *rational expectations* according to which agents form expectations that are consistent with the conditional expectations derived from the model (see Muth, 1961) provide a convenient solution but not very realistic solution in absence of precise knowledge on the actual process of expectation formation.

All that leads to mathematical models that take the form of non-linear stochastic difference equations. Solving such models is not easy and sophisticated numerical techniques must be used. In what follows, we present popular algorithms to find approximate solutions for such models, both in stochastic and deterministic cases. Efficient algorithms exist for the limiting case where there is no future uncertainty and these algorithms can be used to study separately the full implication of non-linearities in the model in absence of stochastic components.

The estimation of the parameters of DSGE models is currently mostly done on the basis of a linear approximation of the model, even if several authors have attempted to estimate non-linear approximation with various versions of the particle filter (see for example Amisano and Tristani, 2010; Anreasen, 2011). Even with linear approximation, estimation of DSGE model remains computationally very intensive, requiring to solve repeatedly the model and to compute its log-likelihood with the Kalman filter.

It is possible to estimate DSGE models by the maximum likelihood, but we advocate rather a Bayesian approach as a way to make explicit, in the estimation process, the use of *a priori* information, to mitigate the problems arising from lack of identification of some parameters and to address misspecification of the model in some direction.

In the second Section, we present a generic version of the model, to fix notations to be used later. The solution of perfect foresight models is discussed in Section 1.3 and for stochastic models in Section 1.4. Estimation is presented in Section 1.5. We present a list of software products implementing these methods in Section 1.6 and conclude with directions for future work.

1.2 A generic model

A DSGE model can, in general, be represented as a set of non–linear equations. The unknowns of these equations are the endogenous variables. The equations relate the

current value of the endogenous variables to their future values, because of expectations and to their past values, to express inertia. The system is affected by external influences that are described by exogenous variables.

The dynamics of such systems can be studied first by abstracting of all uncertainty and making the extreme assumption that agents know exactly the future. One speaks then of *perfect foresight model*. Perfect foresight, deterministic, models, even of large size, can be studied with great accuracy and with simpler methods than stochastic ones.

Formally, we write a perfect foresight model as

$$f(y_{t+1}, y_t, y_{t-1}, u_t) = 0$$

where $f()$ is a vector of $n$functions,$\mathcal{R}^{3n+p} \rightarrow \mathcal{R}^n$, y is a vector of n endogenous variables, that can appear in the model in the current period, t, as well as in the next period $t+1$ and in the previous period $t-1$. u_t is a vector of p exogenous variables. Expressing the equations of the model as functions equal to zero helps the mathematical treatment.

In general, endogenous variables may appear in a model with leads or lags of more than one period and exogenous variables at periods other than the current one may also enter the equations. However, with the addition of adequate auxiliary variables and equations, it is always possible to write more complicated models in this canonical form. The algorithm to do so is discussed in Broze et al. (1990). In a given model, not all variables are necessarily present in previous, current and future periods. If it is important to take this fact into account for efficient implementation of computer code, it simplifies the exposition of the solution algorithms to abstract from it, without departure from generality.

When the model is stochastic, the exogenous variables are zero mean random variables. Note that this assumption is compatible with a large class of models but excludes models where exogenous processes contain an exogenous change in mean such as increase in life expectancy or policy change. When exogenous variables are stochastic, the endogenous are random as well and, in current period t, it is not possible to know the exact future values of endogenous variables in $t+1$, only their conditional distribution, given the information available in t. With the rational expectations hypothesis, the equations of the model hold under conditional expectation:

$$\mathbb{E}\left\{ f(y_{t+1}, y_t, y_{t-1}, u_t) \,|\Omega_t \right\} = 0,$$

where Ω_t is the information set available at period t. In what follows, we assume that shocks u_t are observed at the beginning of period t and that the state of the system is described by y_{t-1}. Therefore, we define the information set available to the agents at the beginning of period t as,

$$\Omega_t = \{u_t, y_{t-1}, y_{t-2}, \ldots\}.$$

This convention is arbitrary and another one could be used instead, but such a convention is necessary to fully specify a given model. Now we use the lighter but equivalent

notation

$$\mathbb{E}_t \left\{ f(y_{t+1}, y_t, y_{t-1}, u_t) \right\} = 0.$$

We make the following restrictive assumptions on the shocks u_t:

$$\begin{aligned} \mathbb{E}(u_t) &= 0 \\ \mathbb{E}(u_t u_\tau') &= 0 \quad t \neq \tau \end{aligned}$$

We take into account possible correlation between the shocks, but exclude serial correlation. Note that auto-correlated processes can be accommodated by adding auxiliary endogenous variables. In that case, the random shock is the innovation of the auto-correlated process.

1.2.1 The nature of the solution

Despite the fact that the deterministic and the stochastic versions of the model are very close, there is an important difference that has consequences for the solution strategy.

In the deterministic case, it is a perfect foresight model where all information about the future values of the exogenous variables is known at the time of computation. On the contrary, in the stochastic case, the realizations of the exogenous variables are only learned period by period.

In the perfect foresight case, it is therefore possible to compute at once the trajectory of the endogenous variables. In the stochastic case, this approach is not available because the value of the shocks are only learned at the beginning of each period. Instead, the solution must take the form of a *solution function* that specifies how endogenous variables y_t are set as a function of the previous state y_{t-1} and the shocks observed at the beginning of period t:

$$y_t = g\left(y_{t-1}, u_t\right).$$

In most cases, there is no closed form expression for function $g()$. It is necessary to use numerical methods to approximate this unknown function. On the basis of the Implicit Function Theorem, Jin and Judd (2002) discuss the conditions for the existence of a unique solution function in the neighborhood of the steady state.

It is well known that rational expectation models entails a multiplicity of solutions, many of them taking the form of self-fulfilling dynamics (see for example Farmer and Woodford, 1997). Most research have focused on models that, after a shock, display a single trajectory back to steady state. Note that this convergence is only asymptotic. In such cases, agents are supposed to be able to coordinate their individual expectations on this single trajectory. Much of the literature on DSGE models has been following this approach but attention has also been given to models with a multiplicity of stable solutions and possible sunspots as in Lubik and Schorfheide (2004).

1.3 Solving perfect foresight models

In the perfect foresight case, the only approximation that we make is that convergence back to the steady state takes place after a finite number of periods rather than asymptotically. The larger number of periods one considers, the more innocuous the approximation.

One can then represent the problem as a two boundary value problem where the initial value of variables appearing with a lag in the model are given by initial conditions and the final value of variables appearing with a lead in the model are set at their steady state value.

When one stacks the equations for all the T periods of the simulation as well as the initial and the terminal conditions, one obtains a large set of nonlinear equations:

$$f\left(y_{t+1}, y_t, y_{t-1}, u_t\right) = 0 \qquad t = 1, \ldots, T$$

with initial conditions, $y_0 = y^\star$, and terminal conditions, $y_{T+1} = \bar{y}$, the steady state.

Until Laffargue (1990), the literature considers that for large macroeconomic models, the size of the nonlinear system would be too large to use Newton method to solve it. Consider for example, a multi–country model with ten countries where the one–country model counts 40 equations and that one simulates over 100 periods. The resulting system of nonlinear equations would count 40,000 equations and the Jacobian matrix 1.6 billion elements.

Such large problems seemed then better attacked by first order iterative methods such as Gauss–Seidel as in Fair and Taylor (1983) or Gilli and Pauletto (1997).

On the contrary, Laffargue (1990) and Boucekkine (1995) show that the large Jacobian of the stacked system has a particular structure that can be exploited to solve efficiently the linear problem at the heart of Newton's method.

The vectors of endogenous variables in each period, y_t, can be stacked in a single large vector Y such that

$$Y = \left[\, y_1' \ \ldots \ y_T' \,\right]'.$$

and the entire system for all T periods can be written as $\mathcal{F}(Y) = 0$. Using the Newton method to solve this system of equations entails starting with a guess $Y^{(0)}$ and obtaining iteratively a series of $Y^{(k)}$, such that

$$\left.\frac{\partial \mathcal{F}}{\partial Y}\right|_{Y=Y^{(k-1)}} \Delta Y^{(k)} = -\mathcal{F}(Y^{(k-1)}),$$

and

$$Y^{(k)} = \Delta Y^{(k)} + Y^{(k-1)}.$$

The iterations are repeated until $\mathcal{F}(Y^{(k)})$ or $||x^{(k)} - x^{(k-1)}||$ are small enough. As mentioned above, a practical difficulty arises when the size of the Jacobian matrix $\frac{\partial \mathcal{F}}{\partial Y}$ is very large.

As remarked by Laffargue (1990), given the dynamic nature of the system, and that, in each period, the current variables depend only on the value of the variables in the previous and in the next period, this Jacobian matrix has a block tridiagonal

structure:

$$\begin{bmatrix} f_{2,1} & f_{1,1} & & & & \\ & \ddots & \ddots & & & \\ \vdots & f_{3,t} & f_{2,t} & f_{1,t} & \vdots & \\ & & \ddots & \ddots & & \\ & & & f_{3,T} & f_{2,T} \end{bmatrix} \begin{bmatrix} \Delta y_1^{(k)} \\ \Delta y_2^{(k)} \\ \vdots \\ \Delta y_{t-1}^{(k)} \\ \Delta y_t^{(k)} \\ \Delta y_{t+1}^{(k)} \\ \vdots \\ \Delta y_{T-1}^{(k)} \\ \Delta y_T^{(k)} \end{bmatrix} = \begin{bmatrix} -f_1\left(y_2^{(k-1)}, y_1^{(k-1)}, y^\star, u_1\right) \\ \vdots \\ -f_t\left(y_{t+1}^{(k-1)}, y_t^{(k-1)}, y_{t-1}^{(k-1)}, u_t\right) \\ \vdots \\ -f_T\left(\bar{y}, y_T^{(k-1)}, y_{T-1}^{(k-1)}, u_T\right) \end{bmatrix}.$$

The fact that the partial derivatives with respect to the state variables appear below the main diagonal follows directly from the fact that they are indeed predetermined variables. This particular structure suggests that it is possible to triangularize the Jacobian by solving T linear problems of the size of the model for one period and then to find the improvement vector to the solution of the whole system through backward substitution.

For example, after triangularization in period t, the system looks like

$$\begin{bmatrix} I & M_1 & & & & \\ & \ddots & \ddots & & & \\ & & I & M_t & & \\ & & f_{3,t+1} & f_{2,t+1} & f_{1,t+1} & \\ & & & \ddots & \ddots & \\ & & & & f_{3,T} & f_{2,T} \end{bmatrix} \begin{bmatrix} \Delta y_1^{(k)} \\ \Delta y_2^{(k)} \\ \vdots \\ \Delta y_{t-1}^{(k)} \\ \Delta y_t^{(k)} \\ \Delta y_{t+1}^{(k)} \\ \vdots \\ \Delta y_{T-1}^{(k)} \\ \Delta y_T^{(k)} \end{bmatrix} \begin{bmatrix} d_0 \\ d_1 \\ d_2 \\ \vdots \\ d_t \\ -f_{t+1}\left(y_{t+2}^{(k-1)}, y_{t+1}^{(k-1)}, y_t^{(k-1)}, u_t\right) \\ \vdots \\ -f_T\left(\bar{y}, y_T^{(k-1)}, y_{T-1}^{(k-1)}, u_T\right) \end{bmatrix}$$

The triangularization obeys to the following recursive rules:

$$\begin{aligned} M_1 &= f_{2,1}^{-1} f_{1,1} \\ d_1 &= -f_{2,1}^{-1} f_1^{(k-1)}\left(y_2^{(k-1)}, y_1^{(k-1)}, y^\star, u_1\right) \end{aligned}$$

then

$$\begin{aligned} M_t &= \left(f_{2,t} - f_{3,t} M_{t-1}\right)^{-1} f_{1,t} \qquad t = 2, \ldots, T \\ d_t &= -\left(f_{2,t} - f_{3,t} M_{t-1}\right)^{-1} \left(f_{3,t} d_{t-1} + f_t^{(k-1)}\left(y_{t+1}^{(k-1)}, y_t^{(k-1)}, y_{t-1}^{(k-1)}, u_t\right)\right) \qquad t = 2, \ldots, T \end{aligned}$$

The values of $\Delta y_t^{(k)}$ are then obtained by backward substitution, starting with $\Delta y_T^{(k)}$:

$$\begin{aligned}\Delta y_T^{(k)} &= d_T \\ \Delta y_t^{(k)} &= d_t - M_t \Delta y_{t+1}^{(k)} \qquad\qquad t = T-1, \ldots, 1\end{aligned}$$

Note that this approach at solving a large two-point boundary value problem starts showing its age. It was developed in the mid-90's when a PC had no more than 64MB RAM. In 2014, when 4GB RAM or more is the norm, the linear problem that is at the core of Newton's method can be more efficiently handled simply using sparse matrix code and storing at once all the non-zero elements of the Jacobian of the entire stacked non-linear model.

As stated above, the only simplifying assumption, in this approach, is to consider that, after a shock, the steady state is reached in finite time, rather than only asymptotically. As usually, one is more interested by the trajectory at the beginning of the simulation, around the time when shocks are hitting the economy, it is easy to verify whether the beginning of the simulation is affected by varying the horizon, T.

It is also possible to consider alternative terminal conditions such as $y_T = y_{T+1}$, or aiming at the trajectory resulting from a linear approximation of the model, among others.

1.4 Solving stochastic models

As stated above, the stochastic version of the general model is pretty similar:

$$\mathbb{E}\left\{f\left(y_{t+1}, y_t, y_{t-1}, u_t | u_t, y_{t-1}, y_{t-2}, \ldots\right)\right\} = 0 \tag{1.1}$$

the exogenous variables, u_t, are now stochastic variables, and, because the future value of endogenous variables, y_{t+1}, will be affected by u_{t+1} that are still unknown in period t, the equation can only hold under conditional expectation.

Because only y_{t-1} affects the dynamics of the model, it is sufficient to retain u_t and y_{t-1} in the information set for period t[1].

Obviously, at a given dates, future shocks are unknown and it is not possible to compute numerical trajectories as in the case of perfect foresight models. It is necessary to change the focus of inquiry toward the *decision* rules used to set y_t as a function of the previous state of the system and current shocks in a way consistent with equation (1.1). In the stochastic case, we are required to search for an unknown function. It turns out that only in very few cases this solution function has an analytic expression. It is therefore necessary to use numerical methods to provide an approximation of the solution function.

[1]It is frequent in the macroeconomic literature to make different assumptions concerning the information set. For example, the assumption made here is consistent with stock of capital on an end of period basis. When one considers stock of capital on a beginning of period basis, then stock of capital at the current period is predetermined and enters the information set.

When using existing software for solving a DSGE model, it is necessary to be attentive to the assumption used in that software concerning the information set and rewrite the model in manner consistent with this assumption.

Several methods exist in the literature to compute approximation of the solution function. Discretization of the state space and iterations on the policy function provide an approximation in tabular form, global methods such as projection methods control the quality of approximation over the entire support of the model while the perturbation approach provides local approximation around a given point. Early surveys of these methods appear in Taylor and Uhlig (1990) and Judd (1996).

For medium to large size models, the most frequently used method is perturbation. It is possible to solve models with several hundreds of equations easily at first or second order. At first order, the perturbation approach is identical to linearization that has been the dominant method used since the inception of RBC analysis. An early survey of methods used to solve dynamic linear economies can be found in Anderson et al. (1996).

It is important to note that the approximate solution computed by perturbation method doesn't depend on the entire distribution of the stochastic shocks but only on as many moments as the order of approximation. We write Σ_u the covariance matrix of the shocks and $\Sigma_u^{(k)}$ the tensor containing the kth moments of this distribution.

It is useful to introduce the *stochastic scale variable*, σ, in the model in order to take into account the effect of future uncertainty on today's decisions. We introduce also the auxiliary random variables ε_t such that

$$u_{t+1} = \sigma\varepsilon_{t+1}.$$

When $\sigma = 0$, there is no uncertainty concerning the future. The moments of ε_t are consistent with the moments of shocks u_t: $\Sigma_u^{(k)} = \sigma^k \Sigma_\varepsilon^{(k)}$.

In the perturbation approach, it is necessary to include the stochastic scale variable as an argument of the solution function:

$$y_t = g\left(y_{t-1}, u_t, \sigma\right).$$

The stochastic scale doesn't play a role at first order, but it appears when deriving the solution at higher orders.

Using the solution function, it is possible to replace y_t and y_{t+1} in the original model and define an equivalent model $F()$ that depends only on y_{t-1}, u_t, ε_{t+1} and σ:

$$\begin{aligned} y_{t+1} &= g(y_t, u_{t+1}, \sigma) \\ &= g(g(y_{t-1}, u_t, \sigma), u_{t+1}, \sigma) \\ F(y_{t-1}, u_t, \varepsilon_{t+1}, \sigma) &= f(g(g(y_{t-1}, u_t, \sigma), \sigma\varepsilon_{t+1}, \sigma), g(y_{t-1}, u_t, \sigma), y_{t-1}, u_t) \end{aligned}$$

and

$$\mathbb{E}_t\left\{F(y_{t-1}, u_t, \varepsilon_{t+1}, \sigma)\right\} = 0. \tag{1.2}$$

It is worthwhile to underscore the different roles played by the exogenous shocks depending on whether they are already realized, u_t, or still to be expected, $u_{t+1} = \sigma\varepsilon_{t+1}$. Once the shocks are observed, they are just additional elements of the state space. When there are random shocks to happen in the future, they contribute to the uncertainty faced by the agents. Their rational decision will be based on the expected

value of future development. Replacing future shocks, u_{t+1}, by the stochastic scale variable and auxiliary shocks, $\sigma\varepsilon_{t+1}$, it is possible to take future uncertainty into account in the perturbation approach.

Based on the partial derivatives of $\mathbb{E}_t\{F(y_{t-1}, u_t, \varepsilon_{t+1}, \sigma)\}$, evaluated at the deterministic steady state, we will recover the partial derivatives of unknown function $g(y_{t-1}, u_t, \sigma)$. It is useful to distinguish two types of perturbation that take place simultaneously:

1. for state space points away, but in the neighborhood, of the deterministic steady state, by considering variations in y_{t-1} and u_t
2. away from a deterministic model towards a stochastic one, by increasing the stochastic scale of the model from $\sigma = 0$ to a positive value.

The deterministic steady state, $\bar{y}$, is formally defined by

$$f(\bar{y}, \bar{y}, \bar{y}, 0) = 0.$$

A model can have several steady states, but only one of them will be used for a local approximation.

Furthermore, the decision rule evaluated at the deterministic steady state must verify, in absence of shocks and future uncertainty,

$$\bar{y} = g(\bar{y}, 0, 0).$$

The deterministic steady state is found by solving a set of non-linear equations. Because, in practice, the steady state needs to be computed repeatedly for a great many values of the parameters in estimation, it is best to use an analytic solution when one is available, or to use analytic substitution to reduce the size of the non–linear problem to be solved.

The perturbation approach starts with a Taylor expansion of the original model. It is necessary to proceed order by order: the first order approximation of the solution function will enter in the second order approximation; the first and the second order solution in the computation of the third order and so on.

A first order expansion of the decision rule takes the form

$$y_t \approx \bar{y} + g_y \hat{y}_{t-1} + g_u u_t$$

where $\hat{y}_{t-1} = y_{t-1} - \bar{y}$ and the first order derivatives of function $g(y_{t-1}, u_t)$, contained in matrices g_y and g_u are unknown. Our task is to recover them from the the first order expansion of the original model:

$$\begin{aligned}
&\mathbb{E}_t\left\{F^{(1)}(y_{t-1}, u_t, \varepsilon_{t+1}, \sigma)\right\} \\
&= \mathbb{E}_t\Big\{f(\bar{y}, \bar{y}, \bar{y}, 0) + f_{y_+}\left(g_y\left(g_y\hat{y} + g_u u + g_\sigma\sigma\right) + g_u\sigma\varepsilon' + g_\sigma\sigma\right) \\
&\quad + f_{y_0}\left(g_y\hat{y} + g_u u + g_\sigma\sigma\right) + f_{y_-}\hat{y} + f_u u\Big\} \\
&= 0.
\end{aligned}$$

Here, we introduce the following notations: $\hat{y} = y_{t-1} - \bar{y}$, $u = u_t$, $\varepsilon' = \varepsilon_{t+1}$, $f_{y_+} = \frac{\partial f}{\partial y_{t+1}}$, $f_{y_0} = \frac{\partial f}{\partial y_t}$, $f_{y_-} = \frac{\partial f}{\partial y_{t-1}}$, $f_u = \frac{\partial f}{\partial u_t}$, $g_y = \frac{\partial g}{\partial y_{t-1}}$, $g_u = \frac{\partial g}{\partial u_t}$, $g_\sigma = \frac{\partial g}{\partial \sigma}$.

It is easy to compute the conditional expectation. Evaluated at the deterministic steady state, all partial derivatives are deterministic as well. The expectation being a linear operator, it is distributed over all the terms and reduces to $\mathbb{E}_t\{\varepsilon'\} = 0$. This disappearance of future shocks is a manifestation of the property of certainty equivalence in linear(–ized) models.

We are now faced with a deterministic equation:

$$\begin{aligned}
&\mathbb{E}_t\left\{F^{(1)}(y_{t-1}, u_t, \varepsilon_{t+1}, \sigma)\right\} \\
&= f(\bar{y}, \bar{y}, \bar{y}, 0) + f_{y_+}\left(g_y\left(g_y\hat{y} + g_u u + g_\sigma\sigma\right) + g_\sigma\sigma\right) \\
&\quad + f_{y_0}\left(g_y\hat{y} + g_u u + g_\sigma\sigma\right) + f_{y_-}\hat{y} + f_u u\Big\} \\
&= \left(f_{y_+}g_y g_y + f_{y_0}g_y + f_{y_-}\right)\hat{y} + \left(f_{y_+}g_y g_u + f_{y_0}g_u + f_u\right)u \\
&\quad + \left(f_{y_+}g_y g_\sigma + f_{y_0}g_\sigma\right)\sigma \\
&= 0.
\end{aligned}$$

Because the equation must hold for any value of $\hat{y}$, u and σ, it must be that

$$\left(f_{y_+}g_y g_y + f_{y_0}g_y + f_{y_-}\right) = 0, \tag{1.3}$$

$$\left(f_{y_+}g_y g_u + f_{y_0}g_u + f_u\right) = 0, \tag{1.4}$$

$$\left(f_{y_+}g_y g_\sigma + f_{y_0}g_\sigma\right) = 0, \tag{1.5}$$

These equations will let us recover unknown g_y, g_u and g_σ, respectively.

1.4.0.1 Recovering g_y.

The first condition reveals a particular difficulty as g_y appears in a matrix polynomial equation.

$$\left(f_{y_+}g_y g_y + f_{y_0}g_y + f_{y_-}\right)\hat{y} = 0 \tag{1.6}$$

Several approaches have been proposed in the literature. One of the most robust and efficient is as follows.

First, rewrite equation 1.6 using a state space representation:

$$\begin{bmatrix} 0 & f_{y_+} \\ I & 0 \end{bmatrix}\begin{bmatrix} I \\ g_y \end{bmatrix} g_y\hat{y} = \begin{bmatrix} -f_{y_-} & -f_{y_0} \\ 0 & I \end{bmatrix}\begin{bmatrix} I \\ g_y \end{bmatrix}\hat{y} \tag{1.7}$$

or, using the fact that, in absence of shocks and at first order,

$$\begin{aligned}
\hat{y}_t &= g_y\hat{y}_{t-1} \\
\hat{y}_{t+1} &= g_y g_y\hat{y}_{t-1}.
\end{aligned}$$

$$\begin{bmatrix} 0 & f_{y_+} \\ I & 0 \end{bmatrix} \begin{bmatrix} \hat{y}_t \\ \hat{y}_{t+1} \end{bmatrix} = \begin{bmatrix} -f_{y_-} & -f_{y_0} \\ 0 & I \end{bmatrix} \begin{bmatrix} \hat{y}_{t-1} \\ y_t \end{bmatrix} \tag{1.8}$$

Note that the lower sub matrix block of the coefficient matrices imposes that the upper half of the right hand side state vector be equal to the lower half of the left hand side one.

Given that only $y_{t-1} - \bar{y}$ is fixed by initial conditions in dynamical system (1.8), the dynamics are obviously under determined. This should not come at a surprise as it is well known that rational expectation models admit many solutions, most of them with self–fulfilling dynamics.

The literature on DSGE models focuses on models that have a unique stable dynamics, meaning that after a shock, there is a single trajectory back to equilibrium. The existence of a unique stable trajectory makes it easier to postulate that agents are able to coordinate their expectation on a single trajectory for the economy.

We therefore use the requirement of a single stable trajectory as a selection device to isolate one solution for g_y. Studying the stability of a linear dynamic system requires analyzing its eigenvalues. However, the presence of the D matrix on the left hand side, makes computing the eigenvalues more complicated, particularly as, in many applications, this matrix may be singular.

The theory of generalized eigenvalues and the real generalized Schur decomposition, see Golub and van Loan (1996), provides a way to handle this problem.

The real generalized Schur decomposition stipulates that for a pencil formed by two real n x n matrices, there exist orthonormal matrices Q and Z such that

$$S = QEZ$$
$$T = QDZ$$

and S is upper–triangular and T is quasi upper–triangular, with $Q'Q = Z'Z = I$. A quasi–triangular matrix is a block triangular matrix, with either 1x1 or 2x2 blocks on the main diagonal. The scalar blocks are associated with real eigenvalues and the 2x2 blocks with complex ones. The algorithm necessary to perform the generalized Schur decomposition is often referred to as the QZ algorithm and is available in several linear algebra libraries and matrix programming languages such as Gauss, Matlab, Octave or Scilab.

The generalized Schur decomposition permits the computation of the generalized eigenvalue problem that solves

$$\lambda_i D x_i = E x_i,$$

When a diagonal block of matrix S is a scalar, $S_{i,i}$, the generalized eigenvalue is obtain is following manner:

$$\lambda_i = \begin{cases} \frac{S_{i,i}}{T_{i,i}} & \text{if } T_{i,i} \neq 0 \\ +\infty & \text{if } T_{i,i} = 0 \text{ and } S_{i,i} > 0 \\ -\infty & \text{if } T_{i,i} = 0 \text{ and } S_{i,i} < 0 \\ \text{any } c \in \mathbb{C} & \text{if } T_{i,i} = 0 \text{ and } S_{i,i} = 0. \end{cases}$$

In the last case, any complex number is generalized eigenvalues of pencil $< D, E >$. This obviously creates a problem for the stability analysis. However, this case only occurs when the model is singular, when one equation can be expressed as linear combination of the other ones. It is nevertheless important for the software to check for this case as it is an easy mistake to make when writing a complex model.

The algorithm is such that when a diagonal block of matrix S is a 2x2 matrix of the form

$$\begin{bmatrix} S_{ii} & S_{i,i+1} \\ S_{i+1,i} & S_{i+1,i+1} \end{bmatrix},$$

the corresponding block of matrix T is a diagonal matrix ,

$$(S_{i,i}T_{i+1,i+1} - S_{i+1,i+1}T_{i,i})^2 < -4S_{i+1,i}S_{i+1,i}T_{i,i}T_{i+1,i+1},$$

and there is a pair of conjugate eigenvalues:

$$\lambda_i, \lambda_{i+1} = \frac{S_{ii}T_{i+1,i+1} + S_{i+1,i+1}T_{i,i} \pm \sqrt{(S_{i,i}T_{i+1,i+1} - S_{i+1,i+1}T_{i,i})^2 + 4S_{i+1,i}S_{i+1,i}T_{i,i}T_{i+1,i+1}}}{2T_{i,i}T_{i+1,i+1}} \tag{1.9}$$

In any case, the theory of generalized eigenvalues is an elegant way of solving the problem created by the possibility of the D matrix to be singular: it introduces the notion of infinite eigenvalue. From the point of view of the analysis of the dynamics of a linear system, it is obvious that infinite eigenvalues, positive or negative, must be treated as explosive roots[2].

The next step is to apply the real generalized Schur decomposition to the linear dynamic system while partitioning it between stable and unstable components:

$$\begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \begin{bmatrix} Z'_{11} & Z'_{21} \\ Z'_{12} & Z'_{22} \end{bmatrix} \begin{bmatrix} I \\ g_y \end{bmatrix} g_y\hat{y} = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} \begin{bmatrix} Z'_{11} & Z'_{21} \\ Z'_{12} & Z'_{22} \end{bmatrix} \begin{bmatrix} I \\ g_y \end{bmatrix} \hat{y}$$

The partitioning is such that S_{11} and T_{11} have stable eigenvalues and S_{22} and T_{22}, explosive ones. The rows of the Z matrix are in turn partitioned so as to be conformable with $\begin{bmatrix} I \\ g_y \end{bmatrix}$.

The only way to cancel the influence of explosive roots on the dynamics and to obtain a stable trajectory is to impose

$$Z'_{21} + Z'_{22}g_y = 0,$$

or

$$g_y = -\left(Z_{22}^{-1}\right)' Z'_{21}. \tag{1.10}$$

A unique stable trajectory exists if and only if Z_{22} is non-singular: there must be

[2]The additional complexity introduced by the emergence of quasi–triangular matrices in the real generalized Schur decomposition is the price being paid to remain in the set of real numbers. From a computer implementation point of view, it is simpler and more efficient than having to use computations with complex numbers.

as many roots larger than one in modulus as there are forward–looking variables in the model and the rank condition must be satisfied. This corresponds to Blanchard and Kahn conditions for existence and unicity of a stable equilibrium (Blanchard and Kahn, 1980).

When the condition is satisfied, equation (1.10) provides the determination of g_y.

Determining g_y, while selecting the stable trajectory, is the most mathematically involved step in the solution of linear rational expectation models. Recovering g_u and g_σ is in turn much simpler.

1.4.0.2 Recovering g_u.
Given g_y, the solution for g_u is directly obtained from equation (1.4):

$$f_{y_+} g_y g_u + f_{y_0} g_u + f_u = 0$$

and

$$g_u = -\left(f_{y_+} g_y + f_{y_0}\right)^{-1} f_u.$$

1.4.0.3 Recovering g_σ.
Equation (1.5) provides the necessary condition to establish that g_σ is always null:

$$f_{y_+} g_y g_\sigma + f_{y_0} g_\sigma = 0$$

is homogeneous and

$$g_\sigma = 0.$$

This is yet another manifestation of the certainty equivalence property of first order approximation.

1.4.0.4 First order approximated decision function.
Putting everything together, the first order approximation of the solution function $g()$ takes the form

$$y_t = \bar{y} + g_y \hat{y}_{t-1} + g_u u_t. \tag{1.11}$$

It is a $VAR(1)$ model, but the coefficient matrices g_y and g_u are constrained by the structural parameters, the specification of the equations of the original non–linear model, the rational expectation hypothesis and the selection of a stable dynamics.

However, the form of the first–order approximated solution let us use all the usual tools developed for the analysis of VAR models (see, for example Hamilton, 1994).

In particular, the first and second theoretical moments are derived as

$$\begin{aligned} \mathbb{E}\left\{y_t\right\} &= \bar{y}, \\ \Sigma_y &= g_y \Sigma_y g_y' + \sigma^2 g_u \Sigma_u g_u' \end{aligned}$$

where Σ_y is the unconditional variance of endogenous variables y_t. The variance is determined by a Lyapunov equation that is best solved by a specialized algorithm (Bini et al., 2012).

To the extend that DSGE models are used to analyze fluctuations at business cycle frequencies, the moments are often compared to empirical moments in the data after

detrending by the Hodrick–Prescott filter. Uhlig (1999) provides formulas to compute theoretical variances after removing the H-P trend.

Impulse responses functions (IRFs) can be evaluated directly, simply by running forward equation (1.11), with u_1 equal to the deterministic impulse and $u_t = 0$, for $t > 1$. This provides an *average* IRF where it is the effect of random shocks after the first period that is averaged. Because the model is linear, it is equivalent to consider the average effect of future shocks or the average shock equal to zero. Note also that, in a linear model, the IRF is independent from the initial position at which the system sits when the deterministic impulse is hitting.

1.4.1 Second order approximation of the model

A second order approximation brings two changes in comparison with a first order approximation: first, the decision rules have the shape of parabolic curves instead of straight lines and, most importantly, the certainty equivalence is broken.

In most cases, fitting the decision rules with parabolic curves instead of straight lines brings only a moderate benefit and it is not true that a second order approximation is always more accurate than a first order one. Remember also that the Taylor expansion of a function diverges outside its ratio of convergence (Judd, 1998).

Breaking the certainty equivalence is the most interesting qualitative benefit of going to second order. It permits to address issues related to attitude toward risk, precautionary motive and risk premium, although in a very elementary manner: at second order, the risk premium is a constant. If one wants a risk premium that varies with the state of the system, it is necessary to consider least a third order approximation.

Considering a second order approximation is a natural step in a perturbation approach. It has been discussed in Collard and Juillard (2001), Kim et al. (2008), Sims (2000), Gomme and Klein (2011). The computation of a second order approximation is done on the basis of the first order approximation, adding the second order Taylor coefficients to the solution function. As for the derivation of the first order approximation, we start with the second order approximation of the original model. However, the derivation is mathematically simpler, because the selection of the locally stable trajectory has been done at order 1.

A second order approximation of model (1.2) is given by

$$\begin{aligned}\mathbb{E}_t\left\{F^{(2)}(y_{t-1},u_t,\varepsilon_{t+1},\sigma)\right\} &= \mathbb{E}_t\Big\{F^{(1)}(y_{t-1},u_t,\varepsilon_{t+1},\sigma)\\ &+0.5\Big(F_{y\_y\_}(\hat{y}\otimes\hat{y})+F_{uu}(u\otimes u)+F_{u'u'}\sigma^2(\epsilon'\otimes\epsilon')+F_{\sigma\sigma}\sigma^2\Big)+F_{y\_u}(\hat{y}\otimes u)\\ &+F_{y\_u'}(\hat{y}\otimes\sigma\epsilon')+F_{y\_\sigma}\hat{y}\sigma+F_{uu'}(u\otimes\sigma\epsilon')+F_{u\sigma}u\sigma+F_{u'\sigma}\sigma\epsilon'\sigma\Big\}\end{aligned}$$

where $F^{(1)}(y_{t-1},u_t,\varepsilon_{t+1},\sigma)$ represents the first order approximation in a compact manner. From the derivation of the first order approximation, we know that $\mathbb{E}_t\left\{F^{(1)}(y_{t-1},u_t,\varepsilon_{t+1},\sigma)\right\} = 0$.

The second order derivatives of the vector of functions, $F()$, are represented in the following manner:

$$\frac{\partial^2 F}{\partial x \partial x} = \begin{bmatrix} \frac{\partial^2 F_1}{\partial x_1 \partial x_1} & \frac{\partial^2 F_1}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 F_1}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 F_1}{\partial x_n \partial x_n} \\ \frac{\partial^2 F_2}{\partial x_1 \partial x_1} & \frac{\partial^2 F_2}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 F_2}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 F_2}{\partial x_n \partial x_n} \\ \vdots & \vdots \ddots & \vdots & \ddots & \vdots & \\ \frac{\partial^2 F_m}{\partial x_1 \partial x_1} & \frac{\partial^2 F_m}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 F_m}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 F_m}{\partial x_n \partial x_n} \end{bmatrix}$$

It is easy to reduce the conditional expectation, but contrarily to what happens at first order, the variance of future shocks remains after simplification:

$$\begin{aligned} \mathbb{E}_t \left\{ F^{(2)}(y_{t-1}, u_t, \varepsilon_{t+1}, \sigma) \right\} &= \mathbb{E}_t \Big\{ F^{(1)}(y_{t-1}, u_t, \varepsilon_{t+1}, \sigma) + F_{y\_u}(\hat{y} \otimes u) \\ &\quad + 0.5 \Big[F_{y\_y\_}(\hat{y} \otimes \hat{y}) + F_{uu}(u \otimes u) + F_{u'u'}\sigma^2(\epsilon' \otimes \epsilon') + F_{\sigma\sigma}\sigma^2 \Big] \\ &\quad + F_{y\_u'}(\hat{y} \otimes \sigma\epsilon') + F_{y\_\sigma}\hat{y}\sigma + F_{uu'}(u \otimes \sigma\epsilon') + F_{u\sigma}u\sigma + F_{u'\sigma}\sigma\epsilon'\sigma \Big\} \\ &= 0.5 \Big[F_{y\_y\_}(\hat{y} \otimes \hat{y}) + F_{uu}(u \otimes u) + \left(F_{u'u'}\vec{\Sigma}_\varepsilon + F_{\sigma\sigma} \right) \sigma^2 \Big] + F_{y\_u}(\hat{y} \otimes u) \\ &\quad + F_{y\_\sigma}\hat{y}\sigma + F_{u\sigma}u\sigma \\ &= 0 \end{aligned}$$

where $\vec{\Sigma}_\varepsilon$ represents the vectorization of the covariance matrix of the auxiliary shocks, with the columns stacked on top of each other.

The only way for the above equation to be satisfied is when

$$\begin{aligned} F_{yy} &= 0 \\ F_{yu} &= 0 \\ F_{uu} &= 0 \\ F_{y\_\sigma} &= 0 \\ F_{u\sigma} &= 0 \\ F_{u'u'}\vec{\Sigma}_\varepsilon + F_{\sigma\sigma} &= 0 \end{aligned}$$

Each of these partial derivatives of function $F()$ represents in fact the second order derivatives of composition of the original function $f()$ and one or two instances of the solution function $g()$. The fact that each of the above partial derivatives must be equal to zero provides the restrictions needed to recover the second order partial derivatives of the solution function.

The second order derivative of the composition of two functions plays an important role in what follows. Let's consider the composition of two functions:

$$\begin{aligned} y &= z(s) \\ f(y) &= f(z(s)) \end{aligned}$$

then,

$$\frac{\partial^2 f}{\partial s \partial s} = \frac{\partial f}{\partial y}\frac{\partial^2 g}{\partial s \partial s} + \frac{\partial^2 f}{\partial y \partial y}\left(\frac{\partial g}{\partial s} \otimes \frac{\partial g}{\partial s}\right).$$

It is worth noting that the second order derivatives of vector of functions $g()$ appear in a linear manner in the final result, simply pre–multiplied by the Jacobian matrix of functions $f()$.

1.4.2 Recovering g_{yy}

The second order derivatives of the solution function with respect to the endogenous state variables, g_{yy}, can be recovered from $F_{y_- y_-} = 0$. When one unrolls this expression, one obtains

$$\begin{aligned} F_{y_- y_-} &= f_{y_+}\left(g_{yy}(g_y \otimes g_y) + g_y g_{yy}\right) + f_{y_0} g_{yy} + B_1 \\ &= 0 \end{aligned}$$

where B_1 is a term that doesn't contain the unknown second order derivatives of function $g()$, but only first order derivatives of $g()$ and first and second order derivatives of $f()$. It is therefore possible to evaluate B_1 on the basis of the specification of the original equations and the results from first order approximation.

This equation can be rearranged:

$$\left(f_{y_+} g_y + f_{y_0}\right) g_{yy} + f_{y_+} g_{yy}(g_y \otimes g_y) = -B_1$$

It is linear in the unknown matrix g_{yy}, but, given its form, can't be solved efficiently by usual algorithms for linear problems. Kamenik (2005) proposes an efficient algorithm for this type of equation. As noted earlier, matrix $f_{y_+} g_y + f_{y_0}$ is invertible under regular assumptions.

1.4.3 Recovering g_{yu}

Once g_{yy} is known, its value can be used to determine g_{yu} from $F_{yu} = 0$. Developing the latter gives:

$$\begin{aligned} F_{y_- u} &= f_{y_+}\left(g_{yy}(g_y \otimes g_u) + g_y g_{yu}\right) + f_{y_0} g_{yu} + B_2 \\ &= 0 \end{aligned}$$

where B_2 is again a term that doesn't contain second order derivatives of $g()$.

This is a standard linear problem and

$$g_{yu} = -\left(f_{y_+} g_y + f_{y_0}\right)^{-1}\left(B_2 + f_{y_+} g_{yy}(g_y \otimes g_u)\right)$$

1.4.4 Recovering g_{uu}

The procedure for recovering g_{uu} is very similar, using $F_{uu} = 0$:

$$\begin{aligned} F_{uu} &= f_{y_+}\left(g_{yy}(g_u \otimes g_u) + g_y g_{uu}\right) + f_{y_0} g_{uu} + B_3 \\ &= 0 \end{aligned}$$

where B_3 is a term that doesn't contain second order derivatives of $g()$. This is a standard linear problem and

$$g_{uu} = -\left(f_{y_+} g_y + f_{y_0}\right)^{-1} \left(B_3 + f_{y_+} g_{yy} (g_u \otimes g_u)\right)$$

1.4.5 Recovering $g_{y\sigma}$, $g_{u\sigma}$

As for first order, the partial cross–derivatives with only one occurrence of the stochastic scale σ are null. The result is derived from $F_{y\sigma} = 0$ and $F_{u\sigma} = 0$ and uses the fact that $g_\sigma = 0$,

$$\begin{aligned} F_{y\sigma} &= f_{y_+} g_y g_{y\sigma} + f_{y_0} g_{y\sigma} \\ &= 0 \\ F_{u\sigma} &= f_{y_+} g_y g_{u\sigma} + f_{y_0} g_{u\sigma} \\ &= 0 \end{aligned}$$

Then,

$$g_{y\sigma} = g_{u\sigma} = 0.$$

1.4.6 Recovering $g_{\sigma\sigma}$

Future uncertainty affects current decisions through the second derivative with respect to the stochastic scale of the model, $g_{\sigma\sigma}$. It is recovered from $\left(F_{u'u'} \vec{\Sigma}_\varepsilon + F_{\sigma\sigma}\right) \sigma^2 = 0$

$$\begin{aligned} F_{\sigma\sigma} + F_{u'u'} \Sigma_\epsilon &= f_{y_+} \left(g_{\sigma\sigma} + g_y g_{\sigma\sigma}\right) + f_{y_0} g_{\sigma\sigma} + \left(f_{y_+ y_+} (g_u \otimes g_u) + f_{y_+} g_{uu}\right) \vec{\Sigma}_\epsilon \\ &= 0 \end{aligned}$$

taking into account that $g_\sigma = 0$. Note that g_{uu} must have been determined before deriving $g_{\sigma\sigma}$.

This is astandard linear problem:

$$g_{\sigma\sigma} = -\left(f_{y_+} (I + g_y) + f_{y_0}\right)^{-1} \left(f_{y_+ y_+} (g_u \otimes g_u) + f_{y_+} g_{uu}\right) \vec{\Sigma}_\epsilon$$

1.4.7 Approximated second order decision functions

The second order approximation of the solution function, $g()$, is given by

$$y_t = \bar{y} + 0.5 g_{\sigma\sigma} \sigma^2 + g_y \hat{y}_{t-1} + g_u u_t + 0.5 \left[g_{yy} (\hat{y}_{t-1} \otimes \hat{y}_{t-1}) + g_{uu} (u_t \otimes u_t)\right] + g_{yu} (\hat{y}_{t-1} \otimes u_t)$$

Remember that σ and Σ_ε were introduced as auxiliary devices to take into account the effect of future uncertainty in the derivation of the approximated solution by a perturbation method. They are related to the variance of the original shocks by $\Sigma_u = \sigma^2 \Sigma_\varepsilon$. It is therefore always possible to choose $\Sigma_\varepsilon = \Sigma_u$ and have $\sigma = 1$.

There exist no close form solution for the moments of the endogenous variables when approximated at second order as each moment depends on all the moments of higher order. As suggested by Kim et al. (2008), it is however possible to compute

a second order approximation of these moments, by ignoring the contribution of moments higher than 2:

$$\begin{aligned}\Sigma_y &= g_y \Sigma_y g_y' + \sigma^2 g_u \Sigma_\epsilon g_u' \\ \mathbb{E}\{y_t\} &= \bar{y} + \frac{1}{2}(I - g_y)^{-1}\left(g_{\sigma\sigma} + g_{yy}\vec{\Sigma}_y + g_{uu}\vec{\Sigma}_\epsilon\right)\end{aligned}$$

The formula for the variance, Σ_y, depends only on the first derivatives of the solution function, $g()$. It is therefore the same as the variance derived for the first order approximation of the solution function. On the contrary, the unconditional mean of endogenous variables is affected by the variance of y and u and $g_{\sigma\sigma}$. It is different from the mean obtained on the basis of a first order approximation

1.4.8 Higher–order approximation

Computing higher-order approximations doesn't present additional mathematical difficulties compared with approximation at second order. The only computational difficulty is with the management of a very large number of derivatives.

The core of the procedure is provided by the Faà di Bruno formula for the k-th order of the composition of two functions in the multivariate case (Ma, 2009). As above, let's consider $f(y) = f(z(s))$. Given their high number of dimensions, we represent derivatives of arbitrary order as tensors

$$\begin{aligned}\frac{\partial^j f^i}{\partial s_{\alpha_1} \dots \partial s_{\alpha_j}} &= \left[F^i_{s^j}\right]_{\alpha_1 \dots \alpha_j}, \\ \frac{\partial^\ell f^i}{\partial y_{\beta_1}, \dots \partial y_{\beta_\ell}} &= \left[F^i_{y^\ell}\right]_{\beta_1 \dots \beta_\ell}, \\ \frac{\partial^k z^\eta}{\partial s_{\gamma_1}, \dots \partial s_{\gamma_k}} &= \left[z_{s^k}\right]^\eta_{\gamma_1 \dots \gamma_k},\end{aligned}$$

and, following Einstein notation, we use the following convention to indicate a sum of products along identical indices appearing first as subscript then as superscript of a tensor ($\beta_i, \dots, \beta_j$ in the following example):

$$[x]^{\alpha_1,\dots,\alpha_i}_{\beta_1,\dots,\beta_j}\,[y]^{\beta_1,\dots,\beta_j}_{\gamma_1,\dots,\gamma_k} = \sum_{\beta_1} \cdots \sum_{\beta_j} [x]^{\alpha_1,\dots,\alpha_i}_{\beta_1,\dots,\beta_j}\,[y]^{\beta_1,\dots,\beta_j}_{\gamma_1,\dots,\gamma_k}.$$

The partial derivatives of $f^i(s)$ with respect to s is written as a function of the partial derivatives of $f^i()$ with respect to y and the partial derivatives of $z()$ with respect to s:

$$[f_{s^j}]^i_{\alpha_1 \dots \alpha_j} = \sum_{l=1}^{j} [f_{z^l}]^i_{\beta_1 \dots \beta_l} \sum_{c \in \mathcal{M}_{l,j}} \prod_{m=1}^{l} [z_{s^{|c_m|}}]^{\beta_m}_{\boldsymbol{\alpha}(c_m)}$$

where $\mathcal{M}_{l,j}$ is the set of all partitions of the set of j indices with l classes, $|.|$ is the cardinality of a set, c_m is m-th class of partition c, and $\boldsymbol{\alpha}(c_m)$ is a sequence of α's indexed by c_m. Note that $\mathcal{M}_{1,j} = \{\{1, \dots, j\}\}$ and $\mathcal{M}_{j,j} = \{\{1\}, \{2\}, \dots, \{j\}\}$.

The formula can easily be unfolded by hand for second or third order. For higher order, the algorithm must be implemented in computer code, but it only requires loops of sums of products and the computation of all partitions of a set of indices [3].

As already noted in the approximation at 2nd order, the highest order derivative z_{s^j} always enters the expression linearly and is simply pre-multiplied by the Jacobian matrix f_z.

Our models involves the composition of the original equation and two instances of the decision function. In order to recover the kth order derivatives of the decision function, g_{y^k}, it is necessary to solve the following equation:

$$\left(f_{y_+} g_y + f_{y_0}\right) g_{y^k} + f_{y_+} g_{y^k} g_y^{\otimes k} = -B$$

where $g_y^{\otimes k}$ is the kth Kronecker power of matrix g_y and B is a term that doesn't contain the unknown k-order derivatives of function $g()$, but only lower order derivatives of $g()$ and first to k-order derivatives of $f()$. It is therefore possible to evaluate B on the basis of the specification of the original equations and the results from lower order approximation.

The other k-order derivatives are solved for in analogous manner.

1.4.9 Assessing accuracy

As one obtains an approximated value of the solution function, it is important to assess the accuracy of this approximation. Ideally, one would like to be able to compare the approximated solution to the true solution or an approximated solution delivered by a method known to be more accurate than local approximation. As we discussed above, such solutions are only available for small models.

Judd (1992) suggests to perform error analysis by plugging the approximate solution, $\hat{g}()$, into the original model 1.1:

$$\varepsilon_t = \mathbb{E}_t\Big[f\left(\hat{g}\left(\hat{g}\left(y_{t-1}, u_t, \sigma\right), u_{t+1}, \sigma\right), g\left(y_{t-1}, u_t, \sigma\right), u_t\right)\Big].$$

where u_{t+1} is random from the point of view of the conditional expectation at period t. The conditional expectation must be computed by numerical integration: for example, by quadrature formula when there is a small number of shocks, or by monomial rule or quasi Monte Carlo integration, for a larger number of shocks.

When it is possible to specify the equations of the model in such a way that the error of an equation is expressed in an interpretable unit, it provides a scale on which evaluating the relative importance of errors. Judd (1992) uses the example of the Euler equation for the consumption choice of the household that can be written so that the error appears in units of consumption.

[3] `dynare++`, written by Ondra Kamenik, available at http://www.dynare.org/documentation-and-support/dynarepp, and `perturbationAIM`, written by Eric Swanson, Gary Anderson and Andrew Levin, available at http://www.ericswanson.us/perturbation.html use such a formula to compute solutions of a DSGE model at arbitrary order.

1.5 Estimation

The above discussion of solution techniques for DSGE models assumed that the value of the model parameters was known. In practice, this knowledge can only be inferred from observation of the data.

In earlier real business cycle tradition, following Kydland and Prescott (1982), parameters were calibrated. A main idea of the *calibration* approach is to choose parameter values from microeconometric studies and to fix free parameters so as to reproduce moments of interest in the aggregate data. See Kydland and Prescott (1996); Hansen and Heckmanm (1996); Sims (1996) for a critical discussion of this approach.

Calibration methodology has the advantage to explicit the focus of analysis to some aspect of the data that the model must reproduce. Its major shortcoming is probably the absence of measure of uncertainty surrounding the chosen calibration.

The Bayesian paradigm proposes an formal way to track *a priori* information that is used in estimation and it is not surprising it became in recent years the dominant approach in quantitative macroeconomics. Canova (2007), DeJong and Dave (2011), Geweke (2005) provide in depth presentations of this approach. Schorfheide (2000) is one of the first applications of Bayesian methodology to the estimation of a DSGE model and An and Schorfheide (2007) provides a detailed discussion of the topic.

Because the use of informative priors sidesteps the issue of identification, it facilitates in practice computation, avoiding problems encountered in numerical optimization to find maximum likelihood when a parameter is weakly identified by the data.

From a methodological point of view, one can consider that the Bayesian approach builds a bridge between calibration and classical estimation: using very tight priors is equivalent to calibrating a model while uninformative priors provide results similar to classical estimation.

Uncertainty and *a priori* knowledge about the model and its parameters are described by the *prior* probability distribution. Confrontation with the data leads to a revision of these probabilities in the form of the *posterior* probability distribution.

The Bayesian approach implies several steps. First choosing the prior density for the parameters. This requires care, because it is not always obvious how to translate informal *a priori* information into a probability distribution and, in general, the specification of the priors has an influence on the results.

The second step is the computation of the posterior distribution. It is very demanding computationally. As a DSGE estimated model is nonlinear in the parameters, there is no hope for conjugate priors and the shape of the posterior distribution is unknown. It can only be recovered by simulation, using Monte Carlo Markov Chain (MCMC) methods. Often, the simulation of the posterior distribution is preceded by the computation of the posterior mode that requires numerical optimization.

When one has obtained MCMC generated sample of draws from the posterior, it is possible to compute point estimates by minimizing an appropriate loss function and corresponding confidence regions.The MCMC sample is also used to compute the marginal density of the model, that is used for model comparison, and the posterior distribution of various results of the model such as IRFs or forecasts.

In order to fix ideas, let's write the prior density of the estimated parameters of the

model as $p(\theta_{A|A})$ where A represents the model and θ_A, the estimated parameters of that model. It helps to keep an index of the model in order to, later, engage in model comparison. The vector of estimated parameters, θ_A, may contain not only structural parameters of the model, but also the parameters describing the distribution of the shocks in that model.

The prior density describes *a priori* beliefs, before considering the data. In the DSGE literature, traditional sources of prior information are microeconomic estimations, previous studies or studies conducted on similar countries. This information helps typically to set the center of the prior distribution for a given parameter. The determination of the dispersion of the prior probability is more subjective and quantifies the uncertainty attached to the prior information.

The model itself specifies the probability distribution of a sequence of observable variables, conditional on the value of the parameters, $p(Y_T|\theta_A, A)$, where Y_T represents the sequence $y_1, \ldots, y_T$. As we are dealing with a dynamic model, this density can be written as the product of a sequence of conditional densities:

$$p(Y_T|\theta_A, A) = p(y_1|\theta_A, A) \prod_{t=1}^{T} p(y_t|Y_{t-1}, \theta_A, A).$$

Once we dispose of a sample of observations, Y_T, it is possible to define the *likelihood* of the model as a function of the estimated parameters, conditional on the value of the observed variables:

$$\mathcal{L}(\theta_A|Y_T, A) = p(Y_T|\theta_A, A).$$

Using Bayes theorem, one obtains the posterior distribution of the estimated parameters:

$$p(\theta_A|Y_T, A) = \frac{p(\theta_A|A)p(Y_T|\theta_A, A)}{\int_{\Theta_A} p(Y_T, \theta_A|A)d\theta_A}$$

The posterior distribution expresses how the prior information is combined with the information obtained from the data to provide an updated distribution of possible values for the estimated parameters.

The denominator of the posterior is a scalar, the *marginal density*, that plays the role of a normalizing factor. We write

$$\begin{aligned} p(Y_T|A) &= \int_{\Theta_A} p(Y_T, \theta_A|A)d\theta_A \\ &= \int_{\Theta_A} p(Y_T|\theta_A, A)p(\theta_A|A)d\theta_A \end{aligned}$$

The marginal density is useful for model comparison, but its knowledge is not required for several other computations such as computing the posterior mode, running the MCMC simulation, or computing the posterior mean. In such cases, it is sufficient to evaluate the posterior density *kernel*:

$$p(\theta_A|A)p(Y_T|\theta_A, A) \propto p(\theta_A|Y_T, A).$$

The essential output of the Bayesian methodology is to establish the posterior distribution of the estimated parameters. However, this multi–dimensional distribution may be too much information to handle for the user of the model and there is the need to convey the results of estimation in the form of point estimates.

Given the posterior density of the parameters and the loss function of the model's user, a point estimate is determined by

$$\widehat{\theta}_A = \arg\min_{\widetilde{\theta}_A} \int_{\Theta_A} L(\widetilde{\theta}_A, \theta_A) p(\theta_A|Y_T, A) d\theta_A.$$

It minimizes the expected loss over the posterior distribution. The loss itself is defined as the loss incurred by retaining $\widetilde{\theta}_A$ as point estimate when the true parameter value is θ_A.

In economics, it is often difficult to establish the exact loss function in the context of model estimation. However, there exist general results that guide common practice:

- when the loss function is quadratic, the posterior mean minimizes expected loss;
- when the loss function is proportional to the absolute value of the difference between estimate and true value of the parameter, the posterior median minimizes this loss function;
- the posterior mode minimizes a loss function of the form 0 or 1: when the estimate coincides with the true value of the parameter the loss is null and the loss is constant for all other values.

This justifies the common usage of reporting the posterior mean of the parameters.

It is also useful to be able to communicate the uncertainty surrounding a point estimate. This is done with *credible sets*, that take into account that the posterior distribution is not necessarily symmetrical:

$$P(\theta \in C) = \int_C p(\theta|Y_T, A) d\theta = 1 - \alpha$$

is a $100(1-\alpha)\%$ credible set for θ with respect to $p(\theta|Y, A)$. Obviously, there is an infinity of such sets for a given distribution. It makes sense to choose the most likely.

A $100(1-\alpha)\%$ *highest probability density* (HPD) credible set for θ with respect to $p(\theta|Y_T, A)$ is a $100(1-\alpha)\%$ credible set with the property

$$p(\theta_1|Y_T, A) \geq p(\theta_2|Y_T, A) \quad \forall \theta_1 \in C \text{ and } \forall \theta_2 \in \bar{C}$$

where $\bar{C}$ represents the complement to C. When the distribution is uni-modal, the credible set is unique.

Estimating parameters is an important part of empirical research. It permits in particular to quantify the intensity of a given economic mechanism. But, it is rarely the end of story. Based on the estimated value of parameters, one is also interested in other quantifiable results from a model, such as moments of the endogenous variables, variance decomposition, IRFs, shocks decomposition or forecasts. All these objects are conditional on the value of the parameters and, for the two last ones, on the observed variables. Very abstractly, these post-estimation computations can be represented as a

function of parameters and observations, $\widetilde{Y} = h(Y_T, \theta)$, where $\widetilde{Y}$ can be either a scalar, a vector or a matrix, depending on the actual computation. Given the uncertainty surrounding parameter estimates, it is legitimate to consider the posterior distribution of such derived statistics.

The posterior predictive density is given by

$$\begin{aligned} p(\tilde{Y}|Y_T, A) &= \int_{\Theta_A} p(\tilde{Y}, \theta_A|Y_T, A) d\theta_A \\ &= \int_{\Theta_A} p(\tilde{Y}|\theta_A, Y_T, A) p(\theta_A|Y_T, A) d\theta_A \end{aligned}$$

1.5.1 Model comparison

Comparing two models is done on the basis of the comparison of their marginal density. When the investigator has a prior on the relative likelihood of one model versus another, the comparison should be done on the basis of the ratio of posterior probabilities or posterior odds ratio. When she considers that all the models under consideration are equally likely, the comparison can be done simply with the Bayes factor.

The ratio of posterior probabilities of two models is

$$\frac{P(A_j|Y_T)}{P(A_k|Y_T)} = \frac{P(A_j)}{P(A_k)} \frac{p(Y_T|A_j)}{p(Y_T|A_k)}$$

In favor of the model A_j versus the model A_k:

- the *prior odds ratio* is $P(A_j)/P(A_k)$
- the *Bayes factor* is $p(Y_T|A_j)/p(Y_T|A_k)$
- the *posterior odds ratio* is $P(A_j|Y_T)/P(A_k|Y_T)$

The interpretation of posterior odds ratio may be delicate. Jeffreys (1961) proposes the following scale for a posterior odds ratio in favor of a model:

$1 - 3$: the evidence is barely worth mentioning

$3 - 10$: the evidence is substantial

$10 - 30$: the evidence is strong

$30 - 100$: the evidence is very strong

> 100 : the evidence is decisive

1.5.2 Bayesian estimation of DSGE models

The application of the Bayesian methodology to the estimation of DSGE models raises a few issues linked to the adaptation of the concepts presented above to the DSGE context.

1.5.2.1 Priors.

Estimated parameters are the parameters of the structural models, but also the standard deviation of structural shocks or measurement errors and, sometimes, their correlation.

Independent priors are specified for each of these parameters as well as the implicit constraint that the value of the parameters must be such that Blanchard and Kahn condition for the existence of a unique stable trajectory is satisfied. It is important that the priors for individual parameters be chosen in such a way as to minimize the set of parameter values excluded by the constraint of a unique, stable trajectory because the existence of a large *hole* in the parameter space specified by the individual priors makes finding the posterior mode and running the MCMC algorithm much more complicate. It also creates problems for the numerical integration necessary to compute the marginal density of the model.

However, some authors have tried to estimate a model while selecting solutions in the indeterminacy region. See, for example, Lubik and Schorfheide (2004).

The most common in the literature is to use independent priors for parameters. However, such choice is often not without consequences on the estimation results. Alternatively, Del Negro and Schorfheide (2008), for example, derive joint priors for the parameters that affect the steady state of the model.

1.5.2.2 Likelihood.
From a statistical point of view, estimating a DSGE model is estimating an unobserved component model: not all the variables of the DSGE model are indeed observed. In fact, because, in general, DSGE models have more endogenous variables than stochastic shocks, some variables are linked by deterministic relationships. It does not make sense to include several co–dependent variables in the list of observed variables.

In fact, unless the variables co–dependent in the model are also linked by a deterministic relationship in the real world, the relationship embodied in the model will not be reflected in the observed variables without the model providing a stochastic shock to account for this discrepancy. This is the problem of stochastic singularity (Sargent, 1989).

The unobserved components framework suggests to use a state space representation for the estimated model (Soderlind, 1999). The measurement equation is

$$y_t^\star = \bar{y}^\star + M\hat{y}_t + \eta_t$$

where $y_t^\star$ is the vector of observed variables in period t, $\bar{y}^\star$, the steady state value of the observed variables. M is a selection matrix, $\hat{y}_t$ is the vector of centered endogenous variables in the model and η_t a vector of measurement errors.

The transition equation is simply given by the first order approximation of the model:

$$\hat{y}_t = g_y(\theta)\hat{y}_{t-1} + g_u(\theta)u_t$$

where u_t is a vector of structural shocks and $g_y(\theta)$ and $g_u(\theta)$ are the matrices of reduced form coefficients obtained via the real generalized Schur decomposition. Note that the reduced form coefficients are non-linear functions of the structural parameters.

We further assume that

$$\begin{aligned}\mathbb{E}\left(u_t u_t'\right) &= Q,\\ \mathbb{E}\left(\eta_t \eta_t'\right) &= V,\\ \mathbb{E}\left(u_t \eta_t'\right) &= 0.\end{aligned}$$

1.5.2.3 The Kalman filter.
Given the state space representation introduced above, the Kalman filter computes recursively, for $t = 1, \ldots, T$:

$$\begin{aligned}v_t &= y_t^* - \bar{y}^\star - M\hat{y}_{t|t-1},\\ F_t &= M P_{t|t-1} M' + V,\\ K_t &= P_{t|t-1} g_y' F_t^{-1},\\ \hat{y}_{t+1|t} &= g_y\left(\hat{y}_{t|t-1} + K_t v_t\right),\\ P_{t+1|t} &= g_y\left(I - K_t M\right) P_{t|t-1} g_y' + g_u Q g_u',\end{aligned}$$

where $g_y = g_y(\theta)$ and $g_u = g_u(\theta)$ and with $y_{1|0}$ and $P_{1|0}$ given. Here $\hat{y}_{t|t-1}$ is the one period ahead forecast of endogenous variables, conditional on the information contained in observed variables until period $t-1$.

The log-likelihood is obtained on the basis of the one-step ahead forecast errors, v_t, and the corresponding co-variance matrix, F_t:

$$\ln L\left(\theta | Y_T^*\right) = -\frac{Tk}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln|F_t| - \frac{1}{2} v_t' F_t^{-1} v_t,$$

where k is the number of estimated parameters. The logarithm of the posterior density is then easily computed by adding the logarithm of the prior density.

The posterior mode is usually computed numerically by hill climbing methods, but may be difficult to compute in practice.

1.5.2.4 Metropolis algorithm.
As already mentioned, the posterior density function of DSGE models is not analytic. It must be recovered by MCMC algorithm.

The posterior density of DSGE models doesn't have enough structure to make it possible to use Gibbs sampling and the algorithm of choice in practice is Metropolis algorithm.

A common implementation of the algorithm in our context is as follows. We choose first as proposal distribution a multi-normal density with co-variance matrix, Σ_{mode}, proportional to the one inferred from the Hessian matrix at the mode of the posterior density. An other choice of proposal is possible. See for example Chib and Ramamurthy (2010) for an alternative approach.

The Metropolis algorithm is made of the following steps:

1. Draw a starting point θ° with $p(\theta) > 0$ from a starting distribution $p^\circ(\theta)$.
2. For $t = 1, 2, \ldots$

(a) Draw a *proposal* θ^* from a *jumping* distribution:

$$J(\theta^*|\theta^{t-1}) = N(\theta^{t-1}, c\Sigma_{\text{mode}})$$

(b) Compute the acceptance ratio

$$r = \frac{p(\theta^*)}{p(\theta^{t-1})}$$

(c) Set

$$\theta^t = \begin{cases} \theta^* & \text{with probability } \min(r, 1) \\ \theta^{t-1} & \text{otherwise.} \end{cases}$$

The random sample generated by the Metropolis algorithm depends upon initial conditions. It is necessary to drop the initial part of the sample before proceeding with analysis. Dropping the first 30% or 50% of the sample is common in the literature.

Intuitively, one can see that an average acceptance rate very high or very low is not desirable. A high average acceptance rate means that the posterior density value at the proposal point is often close to the posterior density at the current point. The proposal point must not be very far from the current point. The Metropolis algorithm is making small steps, and traveling the entire distribution will take a long time.

On the other hand, when the proposal point is very far away from the current point, chances are that the proposal point is in the tail of the distribution and the proposal point is rarely accepted: the average acceptance rate is very low and, again, the Metropolis algorithm will take a long time to travel the distribution. The current consensus view is that aiming for an average acceptance rate close to 25% is nearly optimal (Roberts and Rosenthal, 2001).

The scale factor of the co-variance matrix of the proposal, c, helps tuning the average acceptance rate: increasing the size of this co-variance matrix leads to a smaller average acceptance rate.

It is difficult to know *a priori* how many iterations of the Metropolis algorithm are necessary before one can consider that the generated sample is representative of the target distribution. Various diagnostic tests are proposed in the literature to assess whether convergence is reached (Mengersen et al., 1999).

1.5.2.5 Numerical integration.
Computing point estimates such as the mean value of the estimated parameters under the posterior distribution and other statistics or computing the marginal data density require to compute a multi–dimensional integral involving the posterior density for which we don't have an analytic formula.

Once we have at our disposal a sample of N points, drawn in the posterior distribution thanks to the Metropolis algorithm, it is easy to compute the mean of the parameters or of a function of the parameters. It is simply the average mean of the function of the parameters at each point of the sample:

$$\mathbb{E}(h(\theta_A)) = \int_{\Theta_A} h(\theta_A) p(\theta_A|Y_T, A) d\theta_A$$
$$\approx \frac{1}{N} \sum_{k=1}^{N} h(\theta_A^k)$$

where θ_A^k is drawn from $p(\theta_A|Y_T, A)$.

Computing the marginal density of the model

$$\int_{\Theta_A} p(y|\theta_A, A) p(\theta_A|A) d\theta_A$$

turns out to be more involved numerically.

The first approach is to use the normal approximation provided by Laplace's method. It can be computed after having determined the posterior mode:

$$\hat{p}(Y_T|A) = (2\pi)^{\frac{k}{2}} |\Sigma_{\theta^M}|^{-\frac{1}{2}} p(\theta_A^M|Y_T, A) p(\theta_A^M|A)$$

where θ_A^M is the posterior mode and k the number of estimated parameters (the size of vector θ_A). The covariance matrix Σ_{θ_M} is derived from the inverse of the Hessian of the posterior distribution evaluated at its mode.

A second approach, referred to as modified harmonic mean, is proposed by Geweke (1999) and makes use of the Metropolis sample:

$$p(Y_T|A) = \int_{\theta_A} p(\theta_A|Y_T, A) p(\theta_A|A) d\theta_A$$
$$\hat{p}(Y_T|A) = \left[\frac{1}{n} \sum_{i=1}^{n} \frac{f(\theta_A^{(i)})}{p(\theta_A^{(i)}|Y_T, A) p(\theta_A^{(i)}|A)} \right]^{-1}$$
$$f(\theta) = p^{-1} (2\pi)^{\frac{k}{2}} |\Sigma_\theta|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} (\theta - \overline{\theta})' {\Sigma_\theta}^{-1} (\theta - \overline{\theta}) \right\}$$
$$\times \left\{ (\theta - \overline{\theta})' {\Sigma_\theta}^{-1} (\theta - \overline{\theta}) \leq F^{-1}_{\chi^2_k(p)} \right\}$$

with p, an arbitrary probability and, k, the number of estimated parameters. In practice, the computation is done for several values of the threshold probability, p. The fact that the result remains close when p is varying is taken as a sign of the robustness of the computation.

1.6 Available software

Several software products, some free, some commercial, implement the algorithms described above. Here is a partial list:

Name	Reference and web site
AIM	Anderson and Moore (1985)
	http://www.federalreserve.gov/pubs/oss/oss4/about.html
Dynare	Adjemian et al. (2013)
	http://www.dynare.org
dynare++	Kamenik (2011)
	http://www.dynare.org
Iris toolbox	IRIS Solution Team (2013)
	http://code.google.com/p/iris-toolbox-project
JBenge	Winschel and Krätzig (2010)
	http://jbendge.sourceforge.net
PerturbationAIM	Swanson et al. (2005)
	http://www.ericswanson.us/perturbation.html
RATS	www.estima.com
Schmitt-Grohé and Uribe	Schmitt-Grohe and Uribe (2004)
	http://www.columbia.edu/~mu2166/2nd_order.htm
TROLL	http://www.intex.com/troll
Uhlig's toolkit	Uhlig (1999)
	http://www2.wiwi.hu-berlin.de/institute/wpol/html/toolkit.htm
WinSolve	Pierse (2007)
	http://winsolve.surrey.ac.uk
YADA	Warne (2013)
	http://www.texlips.net/yada

1.7 New directions

DSGE modeling is a field that advances rapidly with continuous innovations. While linear approximation of the models and estimation of these linearized models seemed sufficient to describe the functioning of the economy in normal times, several recent developments make new demands on methods used to solve and estimate DSGE models.

Several important non-linear mechanisms have been put in focus by the Great Recession, such as the zero lower bound for nominal interest rate or debt deflation and sudden stops (Mendoza and Yue, 2011). These developments renewed interest for non–linear solution and estimation methods.

The necessity to integrate financial aspects into the models requires to move away from a unique representative agents. Introducing a discrete number of agents with different characteristics only calls for bigger models, not different solution methods, but dealing with the distribution of an infinite number of agents is a much more complex issue. Krusell and Smith (1998) and, more recently, Algan et al. (2010), Den Haan and Rendahl (2010), Kim et al. (2010), Maliar et al. (2010), Reiter (2010), and Young (2010) attempt to provide solution for the type of heterogeneous agent models where the distribution of agents becomes a state variable.

With the multiplication of questions that are addressed with DSGE models, the size of models increased as well. Nowadays, large multi–country models developed at

international institutions, such as EAGLE at the European System of Central Banks (Gomes et al., 2010), GIMF at the IMF (Kumhof et al., 2010), or QUEST at the European Commission (Ratto et al., 2009) have more than a thousand equations and it is still necessary to develop faster solution algorithms and implementations and, more importantly, faster estimation methods as it is the current bottleneck.

The arrival of new, massively parallel, hardware such as GPUs on the desktop of economists pushes back the frontier of computing and open new perspective, but many algorithms need to be reconsidered to take advantage of parallel computing.

References

Adjemian, S., H. Bastani, F. Karamé, M. Juillard, J. Maih, F. Mihoubi, G. Perendia, J. Pfeifer, M. Ratto, and S. Villemot (2013). Dynare: Reference manual version 4. Dynare Working Papers 1, CEPREMAP.

Algan, Y., O. Allais, and W. J. Den Haan (2010). Solving the incomplete markets model with aggregate uncertainty using parameterized cross-sectional distributions. *Journal of Economic Dynamics and Control 34*(1), 59–68.

Amisano, G. and O. Tristani (2010). Euro area inflation persistence in an estimated nonlinear dsge model. *Journal of Economic Dynamics and Control 34*(10), 1837–1858.

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

Anderson, E., L. Hansen, E. McGrattan, and T. Sargent (1996). Mechanics of forming and estimating dynamic linear economies. In H. Amman, D. Kendrick, and J. Rust (Eds.), *Handbook of Computational Economics*, pp. 171–252. North Holland Press.

Anderson, G. and G. Moore (1985). A linear algebraic procedure for solving linear perfect foresight models. *Economics Letters 17*(3), 247–252.

Anreasen, M. (2011). Non-linear DSGE models and the optimized central difference particle filter. *Journal of Economic Dynamics and Control 35*(10), 127–154.

Bini, D. A., B. Iannazzo, and B. Meini (2012). *Numerical solution of algebraic Riccati equations*. Philadelphia, PA: SIAM.

Blanchard, O. J. and C. M. Kahn (1980). The solution of linear difference models under rational expectations. *Econometrica 48*(5), 1305–1312.

Boucekkine, R. (1995). An alternative methodology for solving nonlinear forward-looking models. *Journal of Economic Dynamics and Control 19*(4), 711–734.

Broze, L., C. Gouriéroux, and A. Szafarz (1990). *Reduced forms of rational expectations models*. Chur: Harwood Academic Publishers.

Canova, F. (2007). *Methods for applied macroeconomic research*. Princeton University Press.

Chib, S. and S. Ramamurthy (2010). Tailored randomized block mcmc methods with application to dsge models. *Journal of Econometrics 155*, 19–38.

Collard, F. and M. Juillard (2001). Accuracy of stochastic perturbation methods: The case of asset pricing models. *Journal of Economic Dynamics and Control 25*, 979–999.

DeJong, D. N. and C. Dave (2011). *Structural macroeconometrics*. Princeton University Press.

Del Negro, M. and F. Schorfheide (2008). Forming priors for dsge models (and how it affects the assessment of nominal rigidities). *Journal of Monetary Economics 55*(7),

1191–1208.

Den Haan, W. J. and P. Rendahl (2010). Solving the incomplete markets model with aggregate uncertainty using explicit aggregation. *Journal of Economic Dynamics and Control 34*(1), 69–78.

Fair, R. C. and J. B. Taylor (1983). Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models. *Econometrica 51*(4), 1169–1185.

Farmer, R. E. and M. Woodford (1997, December). Self-fulfilling prophecies and the business cycle. *Macroeconomic Dynamics 1*(4), 740–769.

Geweke, J. (1999). Using simulation methods for Bayesian econometric models: Inference, development, and communication. *Econometric Reviews 18*(1), 1–73.

Geweke, J. (2005). *Contemporary Bayesian econometrics and statistics*. John Wiley.

Gilli, M. and G. Pauletto (1997). Sparse direct methods for model simulation. *Journal of Economic Dynamics and Control 21*, 1093–1111.

Golub, G. H. and C. F. van Loan (1996). *Matrix computations, 3rd edition*. Baltimore, MA: Johns Hopkins University Press.

Gomes, S., P. Jacquinot, and M. Pisani (2010, May). The eagle. a model for policy analysis of macroeconomic interdependence in the euro area. Working Paper Series 1195, European Central Bank.

Gomme, P. and P. Klein (2011). Second-order approximation of dynamic models without the use of tensors. *Journal of Economic Dynamics and Control 35*(4), 604–615.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Hansen, L. P. and J. J. Heckmanm (1996). The empirical foundations of calibration. *Journal of Economic Perspectives 10*(1), 87104.

IRIS Solution Team (2013). IRIS Toolbox reference manual. unpublished.

Jeffreys, H. (1961). *The theory of probability (3rd ed.)*. Oxford, UK: Oxford University Press.

Jin, H. and K. Judd (2002). Perturbation methods for general dynamic stochastic models. Working paper, Stanford University.

Judd, K. (1992). Projection methods for solving aggregate growth models. *Journal of Economic Theory 58*, 410–452.

Judd, K. (1996). Approximation, perturbation, and projection methods in economic analysis. In H. Amman, D. Kendrick, and J. Rust (Eds.), *Handbook of Computational Economics*, Chapter 12, pp. 511–585. North Holland Press.

Judd, K. (1998). *Numerical methods in economics*. Cambridge, Massachussets: MIT Press.

Kamenik, O. (2005). Solving SDGE models: A new algorithm for the Sylvester equation. *Computational Economics 25*, 167–187.

Kamenik, O. (2011). DSGE models with Dynare++: A tutorial. unpublished.

Kim, J., S. Kim, E. Schaumburg, and C. Sims (2008). Calculating and using second–order accurate solutions of discrete time dynamic equilibrium models. *Journal of Economic Dynamic and Control 32*, 3397–3414.

Kim, S. H., R. Kollmann, and J. Kim (2010). Solving the incomplete market model with aggregate uncertainty using a perturbation method. *Journal of Economic Dynamics and Control 34*(1), 50–58.

Krusell, P. and A. A. Smith (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy 106*(5), 867–896.

Kumhof, M., D. Laxton, D. Muir, and S. Mursula (2010). The global integrated monetary and fiscal model (GIMF): Theoretical structure. Working Paper Series 10/34, International Monetary Fund.

Kydland, F. and E. Prescott (1982). Time–to–build and aggregate fluctuations. *Econometrica 50*, 1345–1370.

Kydland, F. E. and E. C. Prescott (1996). The computational experiment: An econometric tool. *Journal of Economic Perspectives 10*(1), 69–85.

Laffargue, J.-P. (1990). Résolution d'un modèle macroéconomique avec anticipations rationnelles. *Annales d'Economie et de Statistique* (17), 97–119.

Lubik, T. and F. Schorfheide (2004). Testing for indeterminacy: An application to U.S. monetary policy. *American Economic Review 94*(1), 190–217.

Ma, T. W. (2009). Higher chain formula proved by combinatorics. *Electronic Journal of Combinatorics 16*(1), N21.

Maliar, L., S. Maliar, and F. Valli (2010). Solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm. *Journal of Economic Dynamics and Control 34*(1), 42–49.

Mendoza, E. G. and V. Z. Yue (2011). A general equilibrium model of sovereign default and business cycles. NBER Working Papers 17151, National Bureau of Economic Research, Inc.

Mengersen, K. L., C. P. Robert, and C. Guihenneuc-Jouyaux (1999). MCMC convergence diagnostics: A review. *Bayesian Statistics 6*, 415–440.

Muth, J. (1961). Rational expectations and the theory of price movements. *Econometrica 29*, 315335.

Pierse, R. (2007). WinSolve version 4: An introductory guide. unpublished.

Ratto, M., W. Roeger, and J. in't Veld (2009). QUEST III: An estimated open-economy dsge model of the euro area with fiscal and monetary policy. *Economic Modelling 26*, 222–233.

Reiter, M. (2010). Solving the incomplete markets model with aggregate uncertainty by backward induction. *Journal of Economic Dynamics and Control 34*(1), 28–35.

Roberts, G. and J. Rosenthal (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statistical Science 16*(4), 351–367.

Sargent, T. (1989). Two models of measurement and the investment accelerator. *Journal of Political Economy 97*(2), 251–287.

Schmitt-Grohe, S. and M. Uribe (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control 28*, 755–775.

Schorfheide, F. (2000). Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics 15*(6), 645–670.

Sims, C. (2000). Algorithm and software for second order expansion of DSGE models. Working paper, Princeton University.

Sims, C. A. (1996). Macroeconomics and methodology. *Journal of Economic Perspectives 10*(1), 105–120.

Soderlind, P. (1999). Solution and estimation of re macromodels with optimal policy. *European Economic Review 43*(4-6), 813–823.

Swanson, E., G. Anderson, and A. Levine (2005). Higher-order perturbation solutions to dynamic, discrete-time rational expectations models. unpublished.

Taylor, J. B. and H. Uhlig (1990). Solving nonlinear stochastic growth models: A comparison of alternative solution methods. *Journal of Business and Economic Statistics 8*(1), 1–17.

Uhlig, H. (1999). A toolkit for analysing nonlinear dynamic stochastic models easily. In R. Marimon and A. Scott (Eds.), *Computational methods for the study of dynamic economies*, pp. 30–61. Oxford, UK: Oxford University Press.

Warne, A. (2013). YADA manual: Computational details. unpublished.

Winschel, V. and M. Krätzig (2010). Solving, estimating, and selecting nonlinear dynamic models without the curse of dimensionality. *Econometrica 78*(2), 803–821.

Young, E. R. (2010). Solving the incomplete markets model with aggregate uncertainty using the Krusell-Smith algorithm and non-stochastic simulations. *Journal of Economic Dynamics and Control 34*(1), 36–41.