ABSTRACT

Title of Document: DECISION MAKING UNDER UNCERTAINTY AND BOUNDED RATIONALITY

Alin Tavi Mirestean, Doctor of Philosophy, 2005

Directed By: Professor Ingmar Prucha, Department of Economics

In an attempt to capture the complexity of the economic system many economists were led to the formulation of complex nonlinear rational expectations models that in many cases can not be solved analytically. In such cases, numerical methods need to be employed. In chapter one I review several numerical methods that have been used in the economic literature to solve non-linear rational expectations models. I provide a classification of these methodologies and point out their strengths and weaknesses. I conclude by discussing several approaches used to measure accuracy of numerical methods.

In the presence of uncertainty, the multistage stochastic optimization literature has advanced the idea of decomposing a multiperiod optimization problem into many subproblems, each corresponding to a scenario. Finding a solution to the original problem involves aggregating in some form the solutions to each scenario and hence its name, scenario aggregation. In chapter two, I study the viability of scenario aggregation methodology for solving rational expectation models. Specifically, I apply the scenario aggregation method to obtain a solution to a finite horizon life

cycle model of consumption. I discuss the characteristics of the methodology and compare its solution to the analytical solution of the model.

A growing literature in macroeconomics is tweaking the unbounded rationality assumption in an attempt to find alternative approaches to modeling the decision making process, that may explain observed facts better or easier. Following this line of research, in chapter three, I study the impact of bounded rationality on the level of precautionary savings in a finite horizon life-cycle model of consumption. I introduce bounded rationality by assuming that the consumer does not have either the resources or the sophistication to consider all possible future events and to optimize accordingly over a long horizon. Consequently, he focuses on choosing a consumption plan over a short span by considering a limited number of possible scenarios. While under these assumptions the level of precautionary saving in many cases is below the level that a rational expectations model would predict, there are also parameterizations of the model for which the reverse is true.

DECISION MAKING UNDER UNCERTAINTY AND BOUNDED RATIONALITY

By

Alin Tavi Mirestean

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:
Professor Ingmar Prucha, Chair
Professor Michael Binder
Professor Roger Betancourt
Professor John Chao
Professor Michael Fu

Dedication

To my family for their understanding and patience

Table of Contents

Dedication ii
Table of Contents iii
Chapter I. Review of Methods Used for Solving Non-Linear Rational Expectations Models 1
I.1. Introduction 1
I.2. Generic Model 3
I.3. Using Certainty Equivalence; The Extended Path Method 6
I.3.1. Example 7
I.3.2. Notes on Certainty Equivalence Methods 10
I.4. Local Approximation and Perturbation Methods 11
I.4.1. Regular and General Perturbation Methods 11
I.4.2. Example 13
I.4.3. Flavors of Perturbation Methods 15
I.4.4. Alternative Local Approximation Methods 16
I.4.5. Notes on Local Approximation Methods 18
I.5. Discrete State-Space Methods 19
I.5.1. Example. Discrete State-Space Approximation Using Value-Function Iteration 20
I.5.2. Fredholm Equations and Numerical Quadratures 21
I.5.3. Example. Using Quadrature Approximations 24
I.5.4. Notes on Discrete State-Space Methods 26
I.6. Projection Methods 27
I.6.1. The Concept of Projection Methods 28
I.6.2. Parameterized Expectations 39
I.6.3. Notes on Projection Methods 42
I.7. Comparing Numerical Methods: Accuracy and Computational Burden 44
I.8. Concluding Remarks 47
Chapter II. Using Scenario Aggregation Method to Solve a Finite Horizon Life Cycle Model of Consumption 49
II.1. Introduction 49
II.2. A Simple Life-Cycle Model with Precautionary Saving 50
II.3. The Concept of Scenarios 52
II.3.1. The Problem 52
II.3.2. Scenarios and the Event Tree 53
II.4. Scenario Aggregation 57
II.5. The Progressive Hedging Algorithm 60
II.5.1. Description of the Progressive Hedging Algorithm 61
II.6. Using Scenario Aggregation to Solve a Finite Horizon Life Cycle Model 63
II.6.1. The Algorithm 65
II.6.2. Simulation Results 68
II.6.3. The Role of the Penalty Parameter 72
II.6.4. More simulations 74
II.7. Final Remarks 76

Chapter III. Impact of Bounded Rationality on the Magnitude of Precautionary Saving 77
III.1. Introduction 77
III.2. Empirical Results on Precautionary Saving 80
III.3. The Model 82
III.3.1. Rule 1 87
III.3.2. Rule 2 96
III.3.3. Rule 3 104
III.4. Final Remarks 110
Appendices 112
Appendix A. Technical notes to chapter 2 112
Appendix A1. Definitions for Scenarios, Equivalence Classes and Associated Probabilities 112
Appendix A2. Description of the Scenario Aggregation Theory 115
Appendix A3. Solution to a Scenario Subproblem 118
Appendix B. Technical notes to chapter 3 121
Appendix B1. Analytical Solution for a Scenario with Deterministic Interest Rate 121
Appendix B2. Details on the Assumptions in Rule 1 124
Appendix B3. Details on the Assumptions in Rule 2 125

Chapter I. Review of Methods Used for Solving Non-Linear Rational Expectations Models

I.1. Introduction

Limitations faced by most linear macroeconomic models coupled with the growing importance of rational expectations have led many economists, in an attempt to capture the complexity of the economic system, to turn to non-linear rational expectation models. Since the majority of these models can not be solved analytically, researchers have to employ numerical methods in order to be able to compute a solution. Consequently, the use of numerical methods for solving nonlinear rational expectations models has been growing substantially in recent years.

For the past decade, several strategies have been used to compute the solutions to nonlinear rational expectations models. The available numerical methods have several common features as well as differences, and depending on the criteria used, they may be grouped in various ways. Following is an ad-hoc categorization[1] that will be used throughout this chapter.

The first group of methods I consider has as a common feature the fact that the assumption of certainty equivalence is used at some point in the computation of the solution.

[1] This classification draws on Binder et al. (2000), Burnside (1999.), Marcet et al. (1999), McGrattan (1999), Novales et al. (1999), Uhlig (1999) and Judd (1992, 1998).

The second group of methods has as a common denominator the use of a discrete state space, or the discretization of an otherwise continuous space of the state variables[2]. The methods falling into this category are often referred to as discrete state-space methods. They work well for models with a low number of state variables.

The next set of methods is generically known as the class of perturbation methods. Since perturbation methods make heavy use of local approximations, in this presentation, I group them along with some other techniques that use local approximations under the heading of local approximations and perturbation methods.

The fourth group, labeled here as projection methods consists of a collection of methodologies that approximate the true value of the conditional expectations of nonlinear functions with some finite parameterization and then evaluate the initially undetermined parameters. Several methods included in this group have recently become very popular in solving nonlinear rational expectations models containing a relatively small number of state variables[3].

The layout of the chapter contains the presentation of a generic non-linear rational expectations model followed by a description of the methods mentioned above. Throughout the chapter, special cases of the model described in section 2 are used to show how one can apply the methods discussed here.

---

[2] Examples include Baxter et al. (1990), Christiano (1990a, 1990b), Coleman (1990), Tauchen (1990) and Taylor and Uhlig (1990), Tauchen and Hussey (1991), Deaton and Laroque (1992), and Rust (1996)

[3] This approach is used, for example, by Binder et al. (2000), Christiano and Fisher (2000), Judd (1992) and Miranda and Rui (1997).

I.2. Generic Model

I start by presenting a generic model in discrete time that will be used along the way to exemplify the application of some of the methods discussed in this chapter. I assume that the problem consists of maximizing the expected present discounted value of an objective function:

$$\max_{u_t} E\left\{\sum_{t=0}^{\infty}\beta^t \pi\left(u_t\right)\middle|\Omega_0\right\} \tag{1.2.1}$$

subject to

$$x_t = h\left(x_{t-1}, u_t, y_t\right) \tag{1.2.2}$$

$$f(x_t, x_{t-1}) \geq 0 \tag{1.2.3}$$

where u_t and x_t denote the values of the control and state variables u and x respectively, at the beginning of period t. y_t is a vector of forcing variables, $\beta \in (0,1)$ a constant discount factor while π represents the objective function. I further assume that $\pi\left(\cdot\right)$ is twice continuously differentiable, strictly increasing, and strictly concave with respect to u_t. $E\left(\cdot\middle|\Omega_0\right)$ denotes the mathematical expectations operator, conditional on the information set at the beginning of period 0, Ω_0. At any point in time, t, the information set is given by $\Omega_t = \left\{u_t, u_{t-1}, ...; x_t, x_{t-1}, ...; y_t, y_{t-1}, ...\right\}$[4]. Finally, y_t is assumed to be generated by a first-order process

$$y_t = q\left(y_{t-1}, z_t\right), \tag{1.2.4}$$

[4] The elements of the information set point to the fact that variables become known at the beginning of the period. During the chapter this assumption may change to allow for an easier setup of the problem.

where the elements of z_t are distributed independently and identically across t and are drawn from a distribution with a finite number of parameters.

The preceding generic optimization problem covers various examples of models in economics, including the life-cycle model of consumption under uncertainty with or without liquidity constraints, stochastic growth model with or without irreversible investment and certain versions of asset pricing models. The present specification does not cover models that have more than one control variable. However, some of the techniques presented in this chapter could be used to solve such models.

If the underlying assumptions are such that the Bellman principle holds, one can use the Bellman equation method to solve the dynamic programming problem. The Bellman equation for the problem described by (1.2.1) - (1.2.2) is given by

$$V\left(x_t, y_t\right) = \max_{u_t} \left\{ \pi\left(u_t\right) + \beta E\left[V\left(h\left(x_t, u_{t+1}, y_{t+1}\right), y_{t+1}\right) | \Omega_t \right] \right\} \tag{1.2.5}$$

where $V(\cdot)$ is the value function. An alternative way to solve the model is to use the Euler equation method. If u can be expressed as a function of x, i.e. $u_t = g\left(x_t, x_{t-1}, y_t\right)$, the Euler equation for period t for the same problem is:

$$\begin{aligned} &\pi_u^{'}\left[g\left(x_t, x_{t-1}, y_t\right)\right] g_{x_t}^{'}\left(x_t, x_{t-1}, y_t\right) + \\ &\qquad + \beta E\left\{ \pi_u^{'}\left[g\left(x_t, x_{t+1}, y_{t+1}\right)\right] g_{x_t}^{'}\left(x_t, x_{t+1}, y_{t+1}\right) | \Omega_t \right\} = 0 \end{aligned} \tag{1.2.6}$$

So far, it has been assumed that the inequality constraint was not binding. If one considers the possibility of constraint (1.2.3) being binding, then one must employ either the Kuhn-Tucker method or the penalty function method. In the case of the former, the Euler equation for period t becomes:

$$\begin{aligned} &\pi_u^{'}\left[g\left(x_t, x_{t-1}, y_t\right)\right] g_{x_t}^{'}\left(x_t, x_{t-1}, y_t\right) + \mu_t f_{x_t}^{'}\left(x_t, x_{t-1}\right) + \mu_{t+1} f_{x_t}^{'}\left(x_t, x_{t+1}\right) + \\ &\qquad + \beta E\left\{ \pi_u^{'}\left[g\left(x_t, x_{t+1}, y_{t+1}\right)\right] g_{x_t}^{'}\left(x_t, x_{t+1}, y_{t+1}\right) | \Omega_t \right\} = 0 \end{aligned} \tag{1.2.7}$$

where μ_t and μ_{t+1} are Lagrange multipliers. The additional Kuhn-Tucker conditions are given by:

$$\mu_t \geq 0, \quad f\left(x_t, x_{t-1}\right) \geq 0, \quad \mu_t f\left(x_t, x_{t-1}\right) = 0 \tag{1.2.8}$$

Alternatively, one can use penalty methods to account for the inequality constraint. One approach is to modify the objective function by introducing a penalty term[5]. Then the new objective function becomes:

$$E\left\{\sum_{t=0}^{\infty} \beta^t \left[\pi_t\left(u_t\right) + \mu \min\left(f\left(x_t, x_{t-1}\right), 0\right)^3\right] | \Omega_0\right\}$$

where μ is the penalty parameter. Consequently, the Bellman equation is given by:

$$V\left(x_t, y_t\right) = \max_{u_t}\left\{\pi\left(u_t\right) + \mu \min\left(f\left(x_t, x_{t-1}\right), 0\right)^3 + \beta E\left[V\left(h\left(x_t, u_{t+1}, y_{t+1}\right), y_{t+1}\right) | \Omega_t\right]\right\} \tag{1.2.9}$$

Let $u_t^* = d\left(x_t, y_t\right)$ denote the solution of the problem. When an analytical solution for $d(\cdot)$ can not be computed, numerical techniques need to be used. Three main approaches have been used in the literature to solve the problem (1.2.1) - (1.2.4) and to obtain an approximation of the solution. First approach consists of modifying the specification of the problem (1.2.1) - (1.2.2) so that it becomes easier to solve, as is the case with the linear quadratic approximation[6]. Second approach is to employ methods that seek to approximate the value and policy functions by using the Bellman equation[7].

[5] This approach is used by McGrattan (1990).

[6] This approach has been used, among others, by Christiano (1990b) and McGrattan (1990).

[7] Examples of this approach are: Christiano (1990a), Rust (1997), Santos and Vigo (1998), Tauchen (1990).

Finally, the third approach focuses on approximating certain terms appearing in the Euler equation such as decision functions or expectations[8].

These approaches have shaped the design of numerical algorithms used in solving dynamic non-linear rational expectation models. In the next few sections, I will present several of the numerical methods employed by researchers in their attempt to solve functional equations such as the Euler and Bellman equations (1.2.5) - (1.2.9) presented above.

I.3. Using Certainty Equivalence; The Extended Path Method

Certainty equivalence has been used especially for its convenience since it may allow researchers to compute an analytical solution for their models. It has also been used to compute the steady state of a model as a prerequisite for applying some linearization or log-linearization around its equilibrium state[9] or to provide a starting point for more complex algorithms[10]. One methodology that received a lot of attention in the literature is the *extended path* method developed by Fair and Taylor (1983). Solving a model such as (1.2.1) - (1.2.3) usually leads to a functional equation such as a Bellman or an Euler equation.

---

[8] Examples of this approach are Binder et al. (2000), Christiano and Fisher (2000), Judd (1992), Marcet (1994), Mc-Grattan (1996).

[9] This is the case in the linear quadratic approach where the law of motion is linearized and the objective function is replaced by a quadratic approximation around the deterministic steady state.

[10] Certainty equivalence has also been used to provide starting values or temporary values in algorithms used to solve models leading to nonlinear stochastic equations as in early work by Chow (1973, 1976), Bitros and Kelejian (1976) and Prucha and Nadiri (1984).

Let

$$F\left(x_t, x_{t-1}, u_t, u_{t-1}, y_t, y_{t-1}, E_t\left\{\pi'\left[h\left(x_t, x_{t+1}, y_{t+1}\right)\right]h'_{x_t}\right\}, E_t\pi\left[h\left(x_t, x_{t+1}, y_{t+1}\right)\right]\right) = 0 \tag{1.3.1}$$

denote such a functional equation for period t. As before, x_t is the state variable, u_t is the control variable, y_t is a vector of forcing variables, $\pi(\cdot)$ is the objective function, π' is the derivative of π with respect to the control variable, and E_t is the conditional expectations operator based on information available through period t. F is a function that may be nonlinear in variables and expectations. For numerous models if the expectations terms appearing in F were known, (1.3.1) could be easily solved. Since that is not the case, the approach of the extended path method is to first set current and future values of the forcing variables to their expected values. This is equivalent to assuming that all future values of z_t in equation (1.2.4) are zero. Then equation (1.3.1) becomes:

$$F\left(x_t, x_{t-1}, u_t, u_{t-1}, y_t, y_{t-1}, \pi'\left[h\left(x_t, x_{t+1}, E_t y_{t+1}\right)\right]h'_{x_t}, \pi\left[h\left(x_t, x_{t+1}, E_t y_{t+1}\right)\right], \ldots\right) = 0 \tag{1.3.2}$$

Then, the idea is to expand the horizon and iterate over solution paths. Let us consider an example to see how this method can be applied.

I.3.1. Example[11]

Consider the following problem where the social planner or a representative agent maximizes an objective function

$$\max_{u_t} E\left\{\sum_{t=0}^{\infty} \beta^t \pi\left(u_t\right) \middle| \Omega_0\right\} \tag{1.3.3}$$

[11] The application of the extended path method in this example draws to some extent on the model presented in Gagnon (1990).

subject to

$$x_t = h\left(x_{t-1}, u_t, y_t\right) \tag{1.3.4}$$

where y_t is a Gaussian $AR(1)$ process with the law of motion $y_t = \rho y_{t-1} + z_t$ where z_t is i.i.d. $N\left(0, \sigma^2\right)$. It is further assumed that u can be expressed as a function of x, i.e. $u_t = g\left(x_t, x_{t-1}, y_t\right)$. Then the Euler equation for period t is:

$$\begin{aligned} 0 = {} & \pi'\left[g\left(x_t, x_{t-1}, y_t\right)\right] \cdot g'_{x_t}\left(x_t, x_{t-1}, y_t\right) \\ & + \beta E\left\{\pi'\left[g\left(x_{t+1}, x_t, y_{t+1}\right)\right] \cdot g'_{x_t}\left(x_{t+1}, x_t, y_{t+1}\right) \middle| \Omega_t\right\} \end{aligned} \tag{1.3.5}$$

If the expectation term were known in equation (1.3.5), it would be easy to find a solution. The idea of the extended path method is to expand the horizon and then iterate over solution paths. As in Fair and Taylor (1983), I consider the horizon $t, ..., t+k+1$ and assume that x_{t-1} and y_{t-1} are given and that $z_{t+s} = 0$ for $s = 1, ..., k+1$. Following is an algorithm that would implement the extended path methodology. The first step is to choose initial values for x_{t+s} and y_{t+s} for $s = 1, ..., k+1$ and denote them by $\hat{x}_{t+s}$ and $\hat{y}_{t+s}$. Then, for period t, the Euler equation becomes:

$$\begin{aligned} 0 = {} & \pi'\left[g\left(x_t, x_{t-1}, y_t\right)\right] \cdot g'_{x_t}\left(x_t, x_{t-1}, y_t\right) \\ & + \beta\pi'\left[g\left(\hat{x}_{t+1}, x_t, \hat{y}_{t+1}\right)\right] \cdot g'_{x_t}\left(\hat{x}_{t+1}, x_t, \hat{y}_{t+1}\right) \end{aligned} \tag{1.3.6}$$

Similarly, for period $t+s$, the Euler equation is given by:

$$\begin{aligned} 0 = {} & \pi'\left[g\left(x_{t+s}, x_{t+s-1}, y_{t+s}\right)\right] \cdot g'_{x_{t+s}}\left(x_{t+s}, x_{t+s-1}, y_{t+s}\right) \\ & + \beta\pi'\left[g\left(\hat{x}_{t+s+1}, x_{t+s}, \hat{y}_{t+s+1}\right)\right] \cdot g'_{x_{t+s}}\left(\hat{x}_{t+1}, x_{t+s}, \hat{y}_{t+s+1}\right) \end{aligned} \tag{1.3.7}$$

In addition,

$$y_{t+s} = \rho y_{t+s-1} + z_{t+s} \tag{1.3.8}$$

$$u_{t+s} = g\left(x_{t+s}, x_{t+s-1}, y_{t+s}\right) \tag{1.3.9}$$

Therefore, for period $t+s$, equations (1.3.7) - (1.3.9) define a system where x_{t+s-1}, y_{t+s-1}, $\hat{x}_{t+s+1}$, $\hat{y}_{t+s+1}$ are known so one can determine the unknowns x_{t+s}, y_{t+s} and u_{t+s}. Let x_{t+s}^{j}, y_{t+s}^{j} and u_{t+s}^{j} denote the solutions of the system for $s=0,...,k+1$, where j represents the iteration for a fixed horizon, in this case $t,...,t+k+1$. If the solutions $\left\{x_{t+s}^{j}\right\}_{s=0}^{k+1}$, $\left\{y_{t+s}^{j}\right\}_{s=0}^{k+1}$ and $\left\{u_{t+s}^{j}\right\}_{s=0}^{k+1}$ obtained in iteration j are not satisfactory then proceed with the next iteration where $\left\{\hat{x}_{t+s}^{j+1}\right\}_{s=1}^{k+1}=\left\{x_{t+s}^{j}\right\}_{s=1}^{k+1}$, $\left\{\hat{y}_{t+s}^{j+1}\right\}_{s=1}^{k+1}=\left\{y_{t+s}^{j}\right\}_{s=1}^{k+1}$. Notice that the horizon remains the same for iteration $j+1$. The iterations will continue until a satisfactory solution is obtained. At this point, the methodology calls for the extension of the horizon without modifying the starting period. Fair and Taylor extend the horizon by a number of periods that is limited to the number of endogenous variables. This is in essence an ad-hoc rule. In the present example, the horizon is extended by 2 periods, that is, $t,...,t+k+3$. The same steps are followed for the new horizon with the exception of the end criterion, which should consist of a comparison between the last obtained solution, using the $t,...,t+k+3$ horizon, and the solution provided using the previous horizon, $t,...,t+k+1$. The expansion of the horizon continues until a satisfactory solution is obtained. At that point, the procedure will start over with a new starting period and a new horizon. In our example the next starting period should be $t+1$ and the initial horizon $t+1,...,t+k+2$.

One of the less mentioned caveats of this method is that no general convergence proofs for the algorithm are available. In addition, the method relies on the certainty equivalence assumption even though the model is nonlinear. Since expectations of functions are treated as functions of the expectations in future periods in equation (1.3.2),

the solution is only approximate unless function F is linear. This assumption is similar to the one used in the case of linear-quadratic approximation to rational expectations models that has been proposed, for example, by Kydland and Prescott (1982).

In the spirit of Fair and Taylor, Fuhrer and Bleakley (1996), following an algorithm from an unpublished paper by Anderson and Moore (1986), sketch a methodology for finding the solution for nonlinear dynamic rational expectations models.

I.3.2. Notes on Certainty Equivalence Methods

All the methods that use certainty equivalence either as a main step or as a preliminary step in finding a solution, incur an approximation error due to the assumption of perfect foresight. The magnitude of this error depends on the degree of nonlinearity of the model being solved. Fair (2003), while acknowledging its limitations, argues that the use of certainty equivalence may provide good approximations for many macroeconometric models.

In the case of the extended path algorithm, the error propagates through each level of iteration and therefore it forces the use of strong convergence criteria. Due to this fact, the extended path algorithm tends to be computationally intensive. Other methodologies that only use certainty equivalence as a preliminary step as in the case of linearization methods or linear quadratic approaches are not subject to the same computational burden.

In conclusion, while there are cases where certainty equivalence may be used to obtain good approximations, one needs to be careful when using this methodology since there are no guarantees when it comes to accuracy.

I.4. Local Approximation and Perturbation Methods

Economic modeling problems have used a variety of approximation methods in the absence of a closed form solution. One of the most used approximation methods, coming in different flavors, is the local approximation. In particular, the first order approximation has been extensively used in economic modeling. Formally, a function $a(x)$ is a first order approximation of $b(x)$ around x_0 if $a(x_0) = b(x_0)$ and the derivatives at x_0 are the same, $a'(x_0) = b'(x_0)$. In certain instances, first order approximations may not be enough so one would have to compute higher order approximations. Perturbation methods often use high order local approximation and therefore rely heavily on two very well own theorems, Taylor's theorem and implicit function theorem.

I.4.1. Regular and General Perturbation Methods

Perturbation methods are formally addressed by Judd (1998). In this section, following Judd's framework, I try to highlight the basic idea of regular perturbation methods. I start by assuming that the Euler equation of the model under consideration is given by:

$$F(u, \varepsilon) = 0 \tag{1.4.1}$$

where $u(\varepsilon)$ is the policy I want to solve for and ε is a parameter. Further on, I assume that a solution to (1.4.1) exists, that F is differentiable, $u(\varepsilon)$ is a smooth function and $u(0)$ can be easily determined or is known. Differentiating equation (1.4.1) leads to:

$$F_u(u(\varepsilon), \varepsilon)u'(\varepsilon) + F_\varepsilon(u(\varepsilon), \varepsilon) = 0 \tag{1.4.2}$$

Making $\varepsilon = 0$ in equation (1.4.2) allows one to compute $u'(0)$:

$$u'(0) = -\frac{F_\varepsilon\left(u(0),0\right)}{F_u\left(u(0),0\right)} \tag{1.4.3}$$

The necessary condition for the computation of $u'(0)$ is that $F_u\left(u(0),0\right) \neq 0$. Assuming that indeed $F_u\left(u(0),0\right) \neq 0$, it means that now $u'(0)$ is known and one can compute the first order Taylor expansion, of $u(\varepsilon)$ around $\varepsilon = 0$:

$$u(\varepsilon) \cong u(0) - \frac{F_\varepsilon\left(u(0),0\right)}{F_u\left(u(0),0\right)}\varepsilon \tag{1.4.4}$$

This is a linear approximation of $u(\varepsilon)$ around $\varepsilon = 0$. In order to be able to compute higher order approximations of $u(\varepsilon)$ one needs to know at least the value of $u''(0)$. That can be found by differentiating (1.4.2):

$$u''(0) = -\frac{F_{uu}\left(u(0),0\right)\left(u'(0)\right)^2 + 2F_{u\varepsilon}\left(u(0),0\right)u'(0) + F_{\varepsilon\varepsilon}\left(u(0),0\right)}{F_u\left(u(0),0\right)} \tag{1.4.5}$$

The necessary condition for the computation of $u''(0)$ is, once again, that $F_u\left(u(0),0\right) \neq 0$. In addition, second order derivatives shall exist. Then the second order approximation of $u(\varepsilon)$ around $\varepsilon = 0$ is given by:

$$u(\varepsilon) \cong u(0) - \frac{F_\varepsilon\left(u(0),0\right)}{F_u\left(u(0),0\right)}\varepsilon - \frac{1}{2}\varepsilon^2 \frac{F_{uu}\left(u(0),0\right)\left(u'(0)\right)^2 + 2F_{u\varepsilon}\left(u(0),0\right)u'(0) + F_{\varepsilon\varepsilon}\left(u(0),0\right)}{F_u\left(u(0),0\right)}$$

In general, higher order approximations of $u(\varepsilon)$ can be computed if higher derivatives of $F(u,\varepsilon)$ with respect with u exist and if $F_u\left(u(0),0\right) \neq 0$. The advantage of regular perturbation methods based on an implicit function formulation is that one

directly computes the Taylor expansions in terms of whatever variables one wants to use, and that expansion is the best possible asymptotically.

I.4.2. Example

Consider the following optimization problem

$$\max_{u_t} E\left\{\sum_{t=0}^{\infty} \beta^t \pi\left(u_t\right) | \Omega_0\right\} \tag{1.4.6}$$

subject to

$$x_t = h\left(x_{t-1}, u_{t-1}, y_t\right) \tag{1.4.7}$$

with $y_t = y_{t-1} + \varepsilon z_t$, where u_t is the control variable, x_t is the state variable, ε is a scalar parameter and z_t is a stochastic variable drawn from a distribution with zero mean and unit variance. x_t, u_t, ε and z_t are all scalars. The Bellman equation is given by:

$$V(x_t) = \max_{u_t}\left\{\pi\left(u_t\right) + \beta E\left[V\left(h\left(x_t, u_{t+1}, \varepsilon z_{t+1}\right)\right) | \Omega_t\right]\right\} \tag{1.4.8}$$

Then the first order condition is:

$$0 = \pi_u\left(u_t\right) + \beta E\left[V'\left(h\left(x_t, u_t, \varepsilon z_{t+1}\right)\right) h_u\left(x_t, u_t, \varepsilon z_{t+1}\right)\right] \tag{1.4.9}$$

Differentiating the Bellman equation with respect to x_t, one obtains:

$$V'(x_t) = \beta E\left[V'\left(h\left(x_t, u_t, \varepsilon z_{t+1}\right)\right) h_x\left(x_t, u_t, \varepsilon z_{t+1}\right)\right] \tag{1.4.10}$$

Let the control law $U\left(x, \varepsilon\right)$ be the solution of this problem. Then the above equation becomes:

$$V'(x) = \beta E\left[V'\left(h\left(x, U\left(x, \varepsilon\right), \varepsilon z\right)\right) h_x\right]$$

The idea is to first solve for steady state in the deterministic case, which here is equivalent to $\varepsilon = 0$, and then find a Taylor expansion for $U\left(x, \varepsilon\right)$ around $\varepsilon = 0$.

Assuming that there exists a steady state defined by (x^*, u^*) such that $x^* = h(x^*, u^*)$, one can use the following system to obtain steady state solutions:

$$x^* = h(x^*, u^*) \tag{1.4.11}$$

$$0 = \pi_u(u^*) + \beta V'(h(x^*, u^*))h_u(x^*, u^*) \tag{1.4.12}$$

$$V'(x^*) = \beta V'(h(x^*, u^*))h_x(x^*, u^*) \tag{1.4.13}$$

$$V(x^*) = \pi(u^*) + \beta V(x^*) \tag{1.4.14}$$

Further assuming local uniqueness and stability for the steady state, equations (1.4.11)-(1.4.14) provide the solutions for the four steady state quantities $x^*, u^*, V(x^*)$, and $V'(x^*)$. Given that the time subscript for all variables is the same, I drop it for the moment. Going back to equations (1.4.9) - (1.4.10), in the deterministic case, that is, for $\varepsilon = 0$, one obtains:

$$0 = \pi_u(U(x)) + \beta V'[h(x, U(x))]h_u(x, U(x)) \tag{1.4.15}$$

$$V'(x) = \beta V'[h(x, U(x))]h_x(x, U(x)) \tag{1.4.16}$$

Differentiating (1.4.15) and (1.4.16) with respect to x yields

$$0 = \pi_{uu}U'_x + \beta V''(h)(h_x + h_u U'_x)h_u + \beta V'(h)(h_{ux} + h_{uu}U'_x) \tag{1.4.17}$$

$$V'' = \beta V''(h)(h_x + h_u U'_x)h_x + \beta V'(h)(h_{xx} + h_{xu}U'_x) \tag{1.4.18}$$

Therefore, the steady state version of the system (1.4.17) - (1.4.18) is given by:

$$\begin{aligned} 0 = {} & \pi_{uu}(x^*, u^*)U'_x(x^*) + \beta V''(x^*)[h_x(x^*, u^*) \\ & + h_u(x^*, u^*)U'_x(x^*)]h_u(x^*, u^*) + \beta V'(x^*)[h_{ux}(x^*, u^*) + h_{uu}(x^*, u^*)U'_x(x^*)] \end{aligned} \tag{1.4.19}$$

$$\begin{aligned} V''(x^*) = {} & \beta V''(x^*)[h_x(x^*, u^*) + h_u(x^*, u^*)U'_x(x^*)]h_x(x^*, u^*) \\ & + \beta V'(x^*)[h_{xx}(x^*, u^*) + h_{xu}(x^*, u^*)U'_x(x^*)] \end{aligned} \tag{1.4.20}$$

These equations define a quadratic system for the unknowns $V''(x^*)$ and $U'_x(x^*)$.

Going back to the stochastic case, the first order condition with respect to u is given by:

$$0 = \pi_u\left(U\left(x,\varepsilon\right)\right) + \beta E\left\{V'\left(h\left(x,U\left(x,\varepsilon\right),\varepsilon z_{t+1}\right)\right)h_u\left(x,U\left(x,\varepsilon\right),\varepsilon z_{t+1}\right)|\,\Omega_t\right\} \quad (1.4.21)$$

Taking the derivative of the Bellman equation with respect to x yields:

$$V'(x) = \beta E\left\{V'\left(h\left(x,U\left(x,\varepsilon\right),\varepsilon z_{t+1}\right)\right)h_x\left(x,U\left(x,\varepsilon\right),\varepsilon z_{t+1}\right)|\,\Omega_t\right\} \quad (1.4.22)$$

In order to obtain a local approximation of the control law around $\varepsilon = 0$, its derivatives with respect to ε must exist and be known. To find these values one needs to differentiate equations (1.4.21) - (1.4.22) with respect to ε, make $\varepsilon = 0$ and solve the resulting system for the values of the derivatives of U with respect to ε when $\varepsilon = 0$, i.e., for $U_\varepsilon'\left(x^*,0\right)$. Once that value is found, one can compute a Taylor expansion for $U\left(x,\varepsilon\right)$ around $\left(x^*,0\right)$.

If the model requires the addition of an inequality constraint such as (1.2.3) which could be the representation of a liquidity constraint or a gross investment constraint, the Bellman equation (1.4.8) becomes:

$$V(x_t) = \max_{u_t}\left\{\pi\left(u_t\right) + \mu\min\left(f\left(x_t,x_{t-1}\right),0\right)^3 + \beta E\left[V\left(h\left(x_t,u_t,\varepsilon z_t\right)\right)|\,\Omega_t\right]\right\} \quad (1.4.23)$$

where μ is the penalty parameter.

I.4.3. Flavors of Perturbation Methods

Economic modeling problems have used a variety of approximation methods that may be characterized as perturbation methods. The most common use of perturbation methods is the method of linearization around the steady state. Such linearization provides a description on how a dynamical system evolves near its steady state. It has often been used to compute the reaction of a system to shocks. While the first-order

perturbation method exactly corresponds to the solution obtained by standard linearization of first-order conditions, one well known drawback of such a solution, especially in the case of asset pricing models, is that it does not take advantage of any piece of information contained in the distribution of the shocks. Collard and Juillard (2001) use higher order perturbation methods and apply a fixed-point algorithm, which they call "bias reduction procedure", to capture the fact that the policy function depends on the variance of the underlying shocks. Similarly, Schmitt-Grohé and Uribe (2004) derive a second-order approximation to the policy function of a general class of dynamic, discrete-time, rational expectations models using a perturbation method that incorporates a scale parameter for the standard deviations of the exogenous shocks as an argument of the policy function.

I.4.4. Alternative Local Approximation Methods

There are also certain local approximations techniques used in the literature that may look like perturbation methods when in fact they are not. One frequently used approach is to find the deterministic steady state and then to replace the original nonlinear problem with a linear-quadratic problem that is similar to the original problem. The linear-quadratic problem can then be solved using standard methods. This method differs from the perturbation method in that the idea here is to replace the nonlinear problem with a linear-quadratic problem, whereas the perturbation approach focuses on computing derivatives of the nonlinear problem. Let me consider again the problem defined by equations (1.2.1) - (1.2.2). The idea is to approximate the original problem by a

combination of a quadratic objective and a linear constraint, which would take the following form:

$$\max_{u_t} E\left\{\sum_{t=0}^{\infty}\beta^t\left(Q+Wu_t+Ru_t^2\right)|\,\Omega_0\right\} \tag{1.4.24}$$

$$\text{s.t. } x_t = Ax_{t-1}+Bu_t+Cy_t+D \tag{1.4.25}$$

where Q, R, W, A, B, C and D are scalars.

In order to obtain the new specification, the first step is to compute the steady state for the deterministic problem (which means $z_t = 0$ in equation (1.2.4)). Therefore, one has to formulate the Lagrangian:

$$\mathcal{L} = \sum_{t=0}^{\infty}\beta^t\left\{\pi\left(u_t\right)-\lambda_t\left[x_t-h\left(x_{t-1},u_t,y_0\right)\right]\right\} \tag{1.4.26}$$

The first order conditions for (1.4.26) is a system of 3 equations with unknowns x,u and λ. The solution of the system represents the steady state, $\left(x^*,u^*,\lambda^*\right)$. The next step is to take the second order Taylor expansion for $\pi\left(u_t\right)$ and first order Taylor expansion for $h\left(x_{t-1},u_t,y_t\right)$ around $\left(x^*,u^*,y_0\right)$. Thus,

$$\pi\left(u_t\right) = \pi\left(u^*\right)+\pi'\left(u^*\right)\left(u_t-u^*\right)+\pi''\left(u^*\right)\frac{\left(u_t-u^*\right)^2}{2} \tag{1.4.27}$$

$$\begin{aligned} h\left(x_{t-1},u_t,y_t\right) = {} & h\left(x^*,u^*,y_0\right)+h_x'\left(x^*,u^*,y_0\right)\left(x_{t-1}-x^*\right)+ \\ & +h_u'\left(x^*,u^*,y_0\right)\left(u_t-u^*\right)+h_y'\left(x^*,u^*,y_0\right)\left(y_t-y_0\right) \end{aligned} \tag{1.4.28}$$

These expansions allow one to identify the parameters Q, R, W, A, B, C and D. Specifically,

$$\begin{aligned} Q &= \pi\left(u^*\right)-\pi'\left(u^*\right)u^*+\pi''\left(u^*\right)\frac{u^{*2}}{2} \\ W &= \pi'\left(u^*\right)-\pi''\left(u^*\right)u^* \qquad R = \frac{\pi''\left(u^*\right)}{2} \end{aligned} \tag{1.4.29}$$

$$A = h_x'\left(x^*, u^*, y_0\right) \quad B = h_u'\left(x^*, u^*, y_0\right) \quad C = h_y'\left(x^*, u^*, y_0\right)$$
$$D = h\left(x^*, u^*, y_0\right) - h_x'\left(x^*, u^*, y_0\right)x^* - h_u'\left(x^*, u^*, y_0\right)u^* - h_y'\left(x^*, u^*, y_0\right)y_0 \tag{1.4.30}$$

Once the parameters have been identified, the problem can be written in the form described by (1.4.24) and (1.4.25) which has a quadratic objective function and linear constraints[12].

If the model needs to account for an additional inequality constraint such as (1.2.3), the Lagrangian (1.4.26) becomes

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \left\{ \pi\left(u_t\right) - \lambda_t \left[x_t - h\left(x_{t-1}, u_t, y_0\right) \right] + \mu_t f\left(x_t, x_{t-1}\right) \right\} \tag{1.4.31}$$

and the additional Kuhn-Tucker conditions have to be taken into account.

I.4.5. Notes on Local Approximation Methods

The perturbation methods provide a good alternative for dealing with the major drawback of the method of linearization around steady state, that is, its lack of accuracy in the case of high volatility of shocks or high curvature of the objective function. While the first order perturbation method coincides with the standard linearization, the higher order perturbation methods offer a much higher accuracy[13].

Some of the local approximation implementations such as the linear-quadratic method [14] do fairly well when it comes to modeling movements of quantities, but not as

---

[12] There are some other variations of this approach used in the literature such as Christiano (1990b).

[13] See Collard and Juillard (2001) for a study on the accuracy of perturbation methods in the case of an asset-pricing model.

[14] Dotsey and Mao (1992), Christiano (1990b) and McGrattan (1990) have documented the quality of some implementations of the macroeconomic linear-quadratic approach.

well with asset prices. The reason behind this result is that approximation of quantity movements depends only on linear-quadratic terms whereas asset-pricing movements are more likely to involve higher-order terms.

I.5. Discrete State-Space Methods[15]

These methods can be applied in several situations. In the case where the state space of the model is given by a finite set of discrete points these methods may provide an "exact" solution[16]. In addition, these methods are frequently applied by discretizing an otherwise continuous state space. The use of discrete state-space methods in models with a continuous state space is based on the result[17] that the fixed point of a discretized dynamic programming problem may converge point wise to its continuous equivalent[18].

The discrete state-space methods sometimes prove to be a useful alternative to linearization and log-linear approximations to the first order necessary conditions, especially for certain model specifications.

---

[15] This section draws heavily on Burnside (1999) and on Tauchen and Hussey (1991)

[16] This may be the case in models without endogenous state variables, especially when there is only one state variable that follows a simple finite state process. Examples are Mehra and Prescott (1985) and Cecchetti, Lam and Mark (1993).

[17] As documented in Burnside (1999), Atkinson (1976) and Baker (1977) present convergence results related to the use of discrete state spaces to solve integral equations. Results concerning pointwise and absolute convergence of solutions to asset pricing models obtained using discrete state spaces are presented in Tauchen and Hussey (1991) and Burnside (1993).

[18] The procedure employed by discrete state-space methods in models with a continuous state space is sometimes referred to as 'brute force discretization'.

I.5.1. Example. Discrete State-Space Approximation Using Value-Function Iteration

As before, I consider the following maximization problem:

$$\max_{u_t} E\left\{\sum_{t=0}^{\infty}\beta^t\pi\left(u_t\right)|\,\Omega_0\right\} \tag{1.5.1}$$

subject to

$$x_{t+1}=h\left(x_t,u_t,y_t\right) \tag{1.5.2}$$

where y_t is a realization from an n-state Markov chain, u_t is the control variable and x_t is the state variable. Let $\mathcal{Y}=\left\{\mathcal{Y}_1,\mathcal{Y}_2,...,\mathcal{Y}_n\right\}$ be the set of all possible realizations for y_t. In order to be able to apply the above mentioned methodology one has to establish a grid for the state variable. Let the ordered set $\mathfrak{X}=\left\{\mathfrak{X}_1,\mathfrak{X}_2,...,\mathfrak{X}_k\right\}$ be the grid for x_t. Assuming that the control variable u_t can be explicitly determined from equation (1.5.2) as a function of x_t, x_{t+1} and y_t, then the dynamic programming problem can be expressed as:

$$V(x_t,y_t)=\max_{x_{t+1}\in\mathfrak{X}}\left\{\pi\left(x_t,x_{t+1},y_t\right)+\beta E\left[V\left(x_{t+1},y_{t+1}\right)|\,\Omega_t\right]\right\} \tag{1.5.3}$$

Let $\mathcal{H}(x_t,y_t)$ be the Cartesian product of $\mathcal{Y}$ and $\mathfrak{X}$, that is, the set of all possible $m=n\cdot k$ pairs (x_i,y_j). Formally, $\mathcal{H}(x_t,y_t)=\left\{(x_i,y_j)\,|\,x_i\in\mathfrak{X}\subset\Re^k \text{ and } y_j\in\mathcal{Y}\subset\Re^n\right\}$. Hence $\mathcal{H}(x_t,y_t)\subset\Re^k\times\Re^n=\Re^m$. If equation (1.5.3) is discretized using the grid given by $\mathcal{H}(x_t,y_t)$ one can think of function $V\left(\cdot\right)$ as a point in $\Re^m$. Similarly, the expression $\pi\left(x_t,x_{t+1},y_t\right)+\beta E\left(V\left(x_{t+1},y_{t+1}\right)|\,\Omega_t\right)$ can be thought of as a mapping M from $\Re^m$ into $\Re^m$. In this context $V\left(\cdot\right)$ is a fixed point for M, that is, $V=M\left(V\right)$. One of the methods commonly used to solve for the fixed point in these situations is the value function iteration.

In order to solve the maximization problem one can use various algorithms. The algorithm I am going to present follows, to some degree, Christiano (1990a). Let $S^j\left(\mathfrak{X}_p, \mathcal{Y}_q\right)$ be the value of x_{t+1} that maximizes $M\left(V_j\right)$ for given values of x_t and y_t, $\left(x_t, y_t\right) = \left(\mathfrak{X}_p, \mathcal{Y}_q\right) \subset \mathcal{H}$. Formally,

$$S_{t+1}^j\left(\mathfrak{X}_p, \mathcal{Y}_q\right) = \arg\max_{x_{t+1} \in \mathfrak{X}} \left\{\pi\left(\mathfrak{X}_p, x_{t+1}, \mathcal{Y}_q\right) + \beta E\left[V_j\left(x_{t+1}, y_{t+1}\right) | \Omega_t\right]\right\} \tag{1.5.4}$$

where j represents the iteration. The idea is to go through all the possible values for x_{t+1}, that is, the set $\mathfrak{X}$, and find the value that maximizes the right hand side of (1.5.4). That will become the value assigned to $S^j\left(\mathfrak{X}_p, \mathcal{Y}_q\right)$. Then the procedure will be repeated for a different value of the pair x_t and y_t belonging to set $\mathcal{H}(x_t, y_t)$ and, finally, a global maximum will be found. The exposition of the algorithm so far implies an exhaustive search of the grid. The speed of the algorithm can be improved by choosing a starting point for the search in every iteration and continue the search only until the first decrease in the value function is encountered[19]. The decision rule for u_t can then be derived by substituting S_{t+1} for x_{t+1} in the law of motion.

I.5.2. Fredholm Equations and Numerical Quadratures

Let me consider the model specified by (1.2.1) - (1.2.2). Then the Bellman equation is given by:

$$V\left(x_t, y_t\right) = \max_{u_t} \left\{\pi\left(u_t\right) + \beta E\left[V\left(x_{t+1}, y_{t+1}\right) | \Omega_t\right]\right\} \tag{1.5.5}$$

[19] This change in the algorithm, as presented by Christiano (1990a), is valid only when the value function is globally concave.

If y_t follows a process such as (1.2.4), one can rewrite the conditional expectation and consequently the whole equation (1.5.5) as:

$$V(x_t, y_t) = \max_{u_t} \left\{ \pi(u_t) + \beta \int V(x_{t+1}, y_{t+1}) q(y_{t+1} \mid y_t) dy_{t+1} \right\} \tag{1.5.6}$$

In the above equation, the term needing approximation is the integral

$$\int V(x_{t+1}, y_{t+1}) q(y_{t+1} \mid y_t) dy_{t+1}$$

If $V(x_{t+1}, y_{t+1})$ is continuous in y_{t+1} for every x, the integral can be replaced by an N-point quadrature approximation. An N-point quadrature method is based on the notion that one can find some points $y_{i,N}$ and some weights $w_{i,N}$ in order to obtain the following approximation

$$\sum_{i=1}^{N} V(x_{t+1}, y_{i.N}) w_{i,N} \approx \int_Y V(x_{t+1}, y_{t+1}) q(y_{t+1} \mid y_t) dy_{t+1} \tag{1.5.7}$$

where the points $y_{i,N} \in Y, i = 1, \ldots, N$, are chosen according to some rule, while the weight given to each point, $w_{i,N}$, relates to the density function $q(y)$ in the neighborhood of those points. In general, a quadrature method requires a rule for choosing the points, $y_{i,N}$, and a rule for choosing the weights, $w_{i,N}$. The abscissa $y_{i,N}$ and weights $w_{i,N}$ depend only on the density $q(y)$, and not directly on the function V.

Quadrature methods differ in their choice of nodes and weights. Possible choices are Newton-Cotes, Gauss, Gauss-Legendre and Gauss-Hermite approximations. For a classical *N*-point Gauss rule along the real line, the abscissa $y_{i,N}$ and weights $w_{i,N}$ are determined by forcing the rule to be exact for all polynomials of degree less than or equal to $2N-1$.

For most rational expectation models, integral equations are a very common occurrence both in Bellman equations such as (1.5.6), as well as in Euler equations. One of the most common forms of integral equations mentioned in the literature is the Fredholm equation[20]. Therefore, in this section I will present an algorithm similar to the one used by Tauchen and Hussey (1991) for solving such equation.

Now let me assume for a moment that the Euler equation of the model is given by a Fredholm equation of the second kind:

$$v(y_t) = \int \psi(y_{t+1}, y_t) v(y_{t+1}) q(y_{t+1} | y_t) dy_{t+1} + \gamma(y_t) \qquad (1.5.8)$$

where y_t is an n-dimensional vector of variables, E_t is the conditional expectations operator based on information available through period t, and $\psi(y_t, y_{t+1})$ and $\gamma(y_t)$ are functions of y_t and y_{t+1} that depend upon the specific structure of the economic model, and where $v(y_t)$ is the solution function of the model. The process $\{y_t\}$ is characterized by a conditional density, $q(y_{t+1} | y_t)$.

Following Tauchen and Hussey (1991), let the $T[\cdot]$ operator define the integral term in equation (1.5.8). Then (1.5.8) can be written as:

$$v = T[v] + \gamma \qquad (1.5.9)$$

Under regularity conditions, the operator $[I - T]^{-1}$ exists, where I denotes the identity operator, and the exact solution is:

$$v = [I - T]^{-1} \gamma \qquad (1.5.10)$$

[20] One example where this form of integral equation appears is a version of the asset pricing model. See Tauchen and Hussey (1991) and Burnside (1999) for more details.

An approximate solution is obtained using T_N in place of T, where T_N is an approximation of T using quadrature methods for large N. Then $[I - T_N]$ can be inverted.

$$v_N = [I - T_N]^{-1}\gamma \tag{1.5.11}$$

In some cases, the function γ is of the form $\gamma = T[\gamma_0]$ and then the approximate solution is taken as $[I - T_N]^{-1}T_N[\gamma_0]$.

I.5.3. Example. Using Quadrature Approximations

This is an example of discrete state-space approximation using quadrature approximations and value-function iterations. I consider a similar model to the one described in section I.5.1 with the difference that y_t is a Gaussian $AR(1)$ process as opposed to a Markov chain. Again, the representative agent solves the following optimization problem

$$\max_{u_t} E\left\{\sum_{t=0}^{\infty}\beta^t \pi(u_t)\middle|\Omega_0\right\} \tag{1.5.12}$$

subject to

$$x_{t+1} = h(x_t, u_t, y_t) \tag{1.5.13}$$

where y_t is a Gaussian $AR(1)$ process with the law of motion $y_t = \rho y_{t-1} + z_t$ where z_t is i.i.d. $N(0, \sigma^2)$. I assume that u_t can be expressed as a function of x, i.e. $u_t = g(x_t, x_{t+1}, y_t)$. Then the Bellman equation for the dynamic programming is given by

$$V(x_t, y_t) = \max_{x_{t+1}} \pi\left(g(x_t, x_{t+1}, y_t)\right) + \beta E\left\{V(x_{t+1}, y_{t+1})\middle|\Omega_{t+1}\right\} \tag{1.5.14}$$

Writing the expectation term explicitly, equation (1.5.14) becomes:

$$V\left(x_t, y_t\right) = \max_{x_{t+1}} \pi\left(g\left(x_t, x_{t+1}, y_t\right)\right) + \beta \int V\left(x_{t+1}, y_{+1}\right) f\left(y_{t+1} \middle| y_t\right) dy_{t+1} \qquad (1.5.15)$$

where

$$y_{t+1} = \rho y_t + z_{t+1} \qquad (1.5.16)$$

To convert the dynamic programming problem in (1.5.15) to one involving discrete state spaces one needs first to approximate the law of motion of y_t using a discrete state-space process. That is, redefine y_t to be a process which lies in a set $Y = \left\{y_{i,N}\right\}_{i=1}^{N}$ with $y_{i,N} = \sigma \overline{a}_{i,N}$, where $\left\{\overline{a}_{i,N}\right\}_{i=1}^{N}$ is the set of quadrature points corresponding to an *N*-point rule for a standard normal distribution[21]. Let the probability that $y_{t+1} = y_{i,N}$ conditional on $y_t = y_{j,N}$ be given by

$$p_{ji} = \frac{f\left(y_{i,N} \middle| y_{j,N}\right)}{f\left(y_{i,N} \middle| 0\right)} \frac{\overline{w}_{i,N}}{s_j} \qquad (1.5.17)$$

where

$$s_j = \sum_{i=1}^{N} \frac{f\left(y_{i,N} \middle| y_{j,N}\right)}{f\left(y_{i,N} \middle| 0\right)} \overline{w}_{i,N} \qquad (1.5.18)$$

and $\left\{\overline{w}_{i,N}\right\}_{i=1}^{N}$ are the quadrature weights as described in section I.5.2.. With this approximation, the Bellman equation can be written as:

$$\overline{V}\left(x_t, y_t\right) = \max_{x_{t+1}} \left\{ \pi\left(g\left(x_t, x_{t+1}, y_j\right)\right) + \beta \sum_{i=1}^{N} \overline{V}\left(x_{t+1}, y_i\right) p_{ji} \right\} \qquad (1.5.19)$$

given $y_t = y_j$, $j = 1, ..., N$.

[21] This is in fact the approach used by Tauchen and Hussey (1991) and Burnside (1999), among others.

The next step is to replace the state space by a discrete domain $\mathfrak{X}$ from which the solution is chosen. There is no universal recipe for choosing a discrete domain and therefore it is usually done on a priori knowledge of possible values of the state variable[22]. The maximization problem can now be solved by value function iteration as presented in section I.5.1..

I.5.4. Notes on Discrete State-Space Methods

Discrete state-space methods tend to work well for models with a low number of state variables. As the number of variables increases, this approach becomes numerically intractable, suffering from what the literature usually refers to as the curse of dimensionality. In addition, as pointed out in Baxter et al. (1990), when the method is used to solve continuous models there are two sources of approximation error. One is due to forcing a discrete grid on continuous state variables and second from using a discrete approximation of the true distribution of the underlying shocks. There are also instances where the use of discrete state-space methods is entirely inappropriate since the discretization process transforms an infinite state space into a finite one and in the process is changing the information structure. This may not be an issue in most models, but it definitely has an impact in models with partially revealing rational expectations equilibria[23].

[22] See Tauchen (1990) for an example.

[23] See Judd (1998) pp. 578-581 for an example.

I.6. Projection Methods[24]

As opposed to the previously presented numerical methods, the techniques that are going to be presented in this section have a high degree of generality. Projection methods appear to be applicable to solving a wide variety of economic problems. In fact, projection methods can be described as general numerical methods that make use of global approximation techniques[25] to solve equations involving unknown functions.

The idea is to replace the quantity that needs to be approximated by parameterized functions with arbitrary coefficients that are to be determined later on[26], or to represent the approximate solution to the functional equation as a linear combination of known basis functions whose coefficients need to be determined[27]. In either case, there are coefficients to be computed in order to obtain the approximate solution. These coefficients are found by minimizing some form of a residual function.

Further on, a step by step description of the general projection method is presented, followed by a discussion of the parameterized expectations approach.

---

[24] I borrow this terminology from Judd (1992, 1998). These methods are also called weighted residual methods by some authors (for example Rust (1996), McGrattan (1999), Binder et al. (2000)). In fact, one can argue that weighted residual methods are just a subset of the projection methods with a given norm and inner product.

[25] In some cases local approximations are used on subsets of the original domain and then they are pieced together to give a global approximation. One such case is the finite element method.

[26] See Marcet and Marshall (1994a), Marcet and Lorenzoni (1999), Wright and Williams (1982a, 1982b, 1984) and Miranda and Helmberger (1988)

[27] See McGrattan (1999)

I.6.1. The Concept of Projection Methods

Suppose that the functional equation can be described by:

$$F(d) = 0 \tag{1.6.1}$$

where F is a continuous map, $F : C_1 \to C_2$ with C_1 and C_2 complete normed function spaces and $d : D \subset \Re^k \to \Re^m$ is the solution to the optimization problem. More generally, d is a list of functions that enter in the equations that define the equilibrium of a model, such as decision rules, value functions, and conditional expectations functions, while the F operator expresses equilibrium conditions such as Euler equations or Bellman equations.

I.6.1.1. Defining the Problem

The problem is to find $d : D \subset \Re^k \to \Re^m$ that satisfies equation (1.6.1). This translates into finding an approximation $\hat{d}(x;\theta)$ which depends on a finite-dimensional vector of parameters $\theta = [\theta_1, \theta_2, \ldots, \theta_n]$ such that $F\left(\hat{d}\left(x;\theta\right)\right)$ is as close as possible to zero.

I.6.1.1.1. Example[28]

Consider the following finite horizon problem where the social planner or a representative agent maximizes

$$E\left\{\sum_{t=0}^{T} \beta^t \pi\left(u_t\right) \middle| \Omega_0\right\} \tag{1.6.2}$$

subject to

$$x_t = h\left(x_{t-1}, u_t, y_t\right) \tag{1.6.3}$$

[28] The example in section I.6.1 draws heavily on Binder et al. (2000)

with x_0 and x_T given. y_t is an $AR(1)$ process with the law of motion

$$y_t = \rho y_{t-1} + z_t \tag{1.6.4}$$

and z_t are i.i.d. with $z_t \sim N\left(0, \sigma_y^2\right)$. I assume that u can be expressed as a function of x, i.e. $u_t = g\left(x_{t-1}, x_t, y_t\right)$. Then the Euler equation for period $T-1$ is given by

$$\begin{aligned} 0 = {} & \pi'\left(g\left(x_{T-2}, x_{T-1}, y_{T-1}\right)\right) \cdot g'_{x_{T-1}}\left(x_{T-2}, x_{T-1}, y_{T-1}\right) \\ & + \beta E\left\{\pi'\left(g\left(x_{T-1}, x_T, y_T\right)\right) \cdot g'_{x_{T-1}}\left(x_{T-1}, x_T, y_T\right) \middle| \Omega_{T-1}\right\} \end{aligned} \tag{1.6.5}$$

Let the optimal decision rule for x_{T-1} be given by $x_{T-1}^* = d_{T-1}\left(x_{T-2}, y_{T-1}\right)$ where $d(\cdot)$ is a smooth function. The projection methodology consists of approximating $d(\cdot)$ by $\hat{d}(\cdot, \theta)$, where θ represents an unknown parameter matrix. The unknown parameters are computed such that the Euler equation also holds for $\hat{d}(\cdot, \theta)$.

Further on in this section I present the necessary steps one needs to take when applying the projection methods, drawing heavily on the formalization provided by Judd (1998)[29]. As I mentioned above, the methodology consists of finding an approximation $\hat{d}(x;\theta)$ such that $F\left(\hat{d}(x;\theta)\right)$ is as close as possible to zero. It becomes obvious that there are a few issues that need to be addressed: what form of approximation to choose for $\hat{d}(x;\theta)$; does the operator F need to be approximated; what does one understand by, or in other words, what is the formal representation of "as close as possible to zero".

[29] Judd provides a five step check list for applying the projection methods.

I.6.1.2. Finding a Functional Form

The first step comes quite naturally from the need to address the question on how to represent $d(x;\theta)$. In general $\hat{d}$ is defined as a finite linear combination of *basis functions*, $\varphi_i(x),\ i=0,\ldots,n$:

$$\hat{d}(x;\theta)=\varphi_0(x)+\sum_{i=1}^{n}\theta_i\varphi_i(x) \tag{1.6.6}$$

Therefore, the first step consists of choosing a basis over C_1.

Functions $\varphi_i(x),\ i=0,\ldots,n$ are typically simple functions. Standard examples of basis functions include simple polynomials (such as $\varphi_0(x)=1,\ \varphi_i(x)=x^i$), orthogonal polynomials (for example, Chebyshev polynomials), and piecewise linear functions. Choosing a basis is not a straightforward task. For example, ordinary polynomials are sometimes adequate in simple cases where they may provide a good solution with only a few terms. However, since they are not orthogonal on R^+ and they are all monotonically increasing and positive for $x\in R^+$, for x big enough, they are almost indistinguishable and hence they tend to reduce numerical accuracy[30]. Consequently, orthogonal bases are usually preferred to avoid the shortcomings just mentioned.

One of the more popular orthogonal bases is formed by Chebyshev polynomials. They constitute a set of orthogonal polynomials with respect to the weight function

[30] In order to solve for the unknown coefficients θ_i one needs to solve linear systems of equations. The accuracy of these solutions depends on the properties of the matrices involved in the computation, i.e. linear independence of rows and columns. Due to the properties already mentioned, regular polynomials tend to lead to ill-conditioned matrices.

$\omega(x) = 1/\sqrt{1-x^2}$, that is, $\int_{-1}^{1} p_i(x)p_j(x)\omega(x)dx = 0$ for all $i \neq j$. Chebyshev polynomials are defined on the closed interval $[-1,\ 1]$ and can be computed recursively as follows:

$$p_i(x) = 2xp_{i-1}(x) - p_{i-2}(x),\ i = 2,\ 3,\ 4,\ldots \tag{1.6.7}$$

with $p_0(x) = 1$ and $p_1(x) = x$ or, non-recursively, as:

$$p_i(x) = \cos\left(i\ \arccos\left(x\right)\right) \tag{1.6.8}$$

Another set of possible basis functions that can be used to construct a piecewise linear representation for $\hat{d}$ is given by:

$$\varphi_i(x) = \begin{cases} \dfrac{x - x_{i-1}}{x_i - x_{i-1}} & if\ \ x \in [x_{i-1}, x_i] \\ \dfrac{x_{i+1} - x}{x_{i+1} - x_i} & if\ \ x \in [x_i, x_{i+1}] \\ 0 & elsewhere \end{cases} \tag{1.6.9}$$

The points $x_i,\ i = 1,\ldots,n$ that divide the domain $D \subset \Re$ need not be equally spaced. If, for example, it is known that the function to be approximated has large gradients or kinks in certain places then the subdivisions can be smaller and clustered in those regions. On the other hand, in areas where the function is near-linear the subdivisions can be larger and hence fewer.

Once the basis is chosen, the next step is to choose how many terms and consequently how many parameters the functional form will have. In general, if the choice of the basis is good, the higher the number of terms the better the approximations. However, due to the fact that the more terms are chosen the more parameters have to be computed, one should choose the smallest number of terms, n, that yields an acceptable

approximation. One possible approach is to begin with a small n and then increase its value until some approximation threshold is reached.

I.6.1.2.1. Example

Going back to the model defined by equations (1.6.2) and (1.6.3) the next step is choosing a base. I assume that Chebyshev polynomials are used in constructing the functional form for $\hat{d}_{T-1}(\cdot,\theta)$. Then:

$$\hat{d}_{T-1}(x_{T-2}, y_{T-1}; \theta_{T-1}) = \sum_{s=1}^{n_{x,T-1}} \sum_{q=1}^{n_{y,T-1}} \theta_{T-1,sq}\, p_{s-1}(\tilde{x}_{T-1})\, p_{q-1}(\tilde{y}_{T-1}) \tag{1.6.10}$$

where $\theta_{T-1,sq}$ is the (s,q) element of θ_{T-1}, $p_l(\cdot)$ is the l-th order Chebyshev polynomial as defined in (1.6.7) - (1.6.8), while $n_{x,T-1}$ and $n_{y,T-1}$ are the maximum order of the Chebyshev polynomials assumed for $\tilde{x}_{T-1}$ and $\tilde{y}_{T-1}$ respectively. In order to restrict the domain of the polynomials to the unit interval the following transformation is applied:

$$\tilde{x}_{T-1} = 2\frac{x_{T-1} - x_{T-1}^{\min}}{x_{T-1}^{\max} - x_{T-1}^{\min}} - 1 \tag{1.6.11}$$

$$\tilde{y}_{T-1} = 2\frac{\tilde{y}_{T-1} - y_{T-1}^{\min}}{y_{T-1}^{\max} - y_{T-1}^{\min}} - 1 \tag{1.6.12}$$

I.6.1.3. Choosing a Residual Function

In many cases, computing $F(\hat{d})$ may require the use of numerical approximations such as when $F(d)$ involves integration of d. In those cases, the F operator has to be approximated. In addition, once the methodology for approximating d and F has been

established, one needs to choose a residual function. Therefore, the third step consists of defining the residual function and an approximation criterion. Let

$$R(x;\theta) \equiv \hat{F}(\hat{d}(\cdot,\theta))(x) \tag{1.6.13}$$

be the residual function. At this point, a decision has to be made on how an acceptable approximation is defined. That is accomplished by choosing an approximation criterion. One choice is to compute the sum of squared residuals, $\|R(\cdot;\theta)\| \equiv \langle R(\cdot;\theta), R(\cdot;\theta)\rangle$ and then determine θ such that $\|R(\cdot;\theta)\|$ is minimized. An alternative would be to choose a collection of *n test functions* in C_2, $p_i : D \to R^m, i = 1,...,n$, and for each guess of θ to compute the *n* projections, $P_i(\cdot) \equiv \langle R(\cdot;\theta), p_i(\cdot)\rangle$[31]. It is obvious that this step creates the projections that will be used to determine the value of the unknown coefficients, θ. Another popular choice in the literature is the weighted residual criterion defined as[32]:

$$\int_D \psi_i(x) R(x;\theta) dx = 0, \quad i = 1,\ldots,n \tag{1.6.14}$$

where $\psi_i(x),\ i = 1,\ldots,n$ are weight functions. Alternatively, the set of equations (1.6.14) can be written as

$$\int_D \omega(x) R(x;\theta) dx = 0 \tag{1.6.15}$$

where D is the domain for function d, $\omega(x) = \sum_{i=1}^{n} \omega_i \psi_i(x)$ and (1.6.15) must hold for any non-zero weights $\omega_i,\ i = 1,\ldots,n$. Therefore, the method sets a weighted integral of $R(x;\theta)$ to zero as the criterion for determining θ.

---

[31] The choice of the criterion gives the method its name. That is why in the literature the method appears both under the name "projection method" and "weighted residual method".

[32] See McGrattan (1999).

I.6.1.3.1. Example

Going back to the example, recall that Chebyshev polynomials were used in constructing the functional form for $\hat{d}_{T-1}(\cdot,\theta)$:

$$\hat{d}_{T-1}(x_{T-2}, y_{T-1}; \theta_{T-1}) = \sum_{s=1}^{n_{x,T-1}} \sum_{q=1}^{n_{y,T-1}} \theta_{T-1,sq} p_{s-1}(\tilde{x}_{T-1}) p_{q-1}(\tilde{y}_{T-1})$$

As mentioned above, the Euler equation (1.6.5) needs to hold for $\hat{d}(\cdot,\theta)$. Therefore, its right hand side is a prime candidate for defining the residuals function. Let $v_{T-1} = \begin{pmatrix} x_{T-2} \\ y_{T-1} \end{pmatrix}$.

With this notation, the residual function is given by:

$$\begin{aligned} &R_{T-1}\left[v_{T-1}; \hat{d}_{T-1}(v_{T-1}; \theta_{T-1}) \right] = \\ &\quad g\left(v_{T-1}, \hat{d}_{T-1}(v_{T-1}; \theta_{T-1}), y_{T-1}\right) \cdot g'_{x_{T-1}}\left(v_{T-1}, \hat{d}_{T-1}(v_{T-1}; \theta_{T-1}), y_{T-1}\right) \\ &\quad + (\pi')^{-1} \left\{ E\left[\beta \pi' g\left(\hat{d}_{T-1}(v_{T-1}; \theta_{T-1}), x_T, y_T\right) \cdot g'_{x_{T-1}}\left(\hat{d}_{T-1}(v_{T-1}; \theta_{T-1}), x_T, y_T\right) \right] \right\} \end{aligned} \tag{1.6.16}$$

Then the criterion for computing $\tilde{\theta}_{T-1}$ is given by the weighted residual integral equation:

$$\int_{v_{T-1}} R_{T-1}\left[v_{T-1}; \hat{d}_{T-1}(v_{T-1}; \hat{\theta}_{T-1}) \right] \bar{W}(v_{T-1}) dv_{T-1} = 0 \tag{1.6.17}$$

where $\bar{W}$ is a weighting function. In the next section it will become clear why the choice of $\bar{W}$ is important in the computation of $\tilde{\theta}_{T-1}$.

I.6.1.4. Methods Used for Estimating the Parameters

Evidently, the next step is to find $\theta \in R^n$ that minimizes the chosen criterion. In order to determine the coefficients $\theta_1, \ldots, \theta_n$ several methods can be used, depending on the criterion chosen.

If the projection criterion is chosen, finding the n components of θ means solving the n equations $\langle R(x,\theta), p_i \rangle = 0$ for some specified collection of test functions, p_i. The choice of the test functions p_i defines the implementation of the projection method. In the *least squares* implementation the projection directions are given by the gradients of the residual function. Therefore, the problem is reduced to solving the nonlinear set of equations generated by $\left\langle R(x,\theta), \frac{\partial R(x,\theta)}{\partial \theta_i} \right\rangle = 0 \quad i = 1,...,n$.

One alternative is to choose the first n elements of the basis Φ, that is, $\varphi_i(x)\ i = 1,...,n$, as the weight functions, $\psi_i(x),\ i = 1,\ldots,n$. In other words, n elements of the basis used to approximate $\hat{d}(x;\theta)$ are also used as test functions to define the projection direction, $\psi_i(x) = \varphi_i(x),\ i = 1,\ldots,n$. This technique is known as the Galerkin method. As a result of this choice, the Galerkin method forces the residual to be orthogonal to each of the basis functions. Therefore θ is chosen to solve the following set of equations:

$$P_i(\theta) = \langle R(x,\theta), \varphi_i(x) \rangle = 0 \quad i = 1,...,n \tag{1.6.18}$$

As long as the basis functions are chosen from a complete set of functions, system (1.6.18) provides the exact solution, given that enough terms are included. If the basis consists of monomials, the method is also known as the method of moments. Then θ is the solution to the system:

$$P_i(\theta) = \langle R(x,\theta), x^{i-1} \rangle = 0 \quad i = 1,...,n \tag{1.6.19}$$

The collocation method chooses θ so that the functional equation holds exactly at n fixed points, x_i, called the collocation points. That is, θ is the solution to:

$$R(x_i;\theta)=0,\quad i=1,...,n \tag{1.6.20}$$

where $\{x_i\}_{i=1}^{n}$ are n fixed points from D. It is easy to see that this is a special case of the projection approach, since $\langle R(x;\theta),\delta(x-x_i)\rangle = R(x_i;\theta)$, where $\delta(x-x_i)$ is the Dirac function at x_i. If the collocation points x_i are chosen as the n roots of the n^{th} orthogonal polynomial basis element and the basis elements are orthogonal with respect to the inner product, the method is called orthogonal collocation. The Chebyshev polynomial basis is a very popular choice for an orthogonal collocation method.

I.6.1.4.1. Example

Going back to the example, it was established that the criterion for computing $\tilde{\theta}_{T-1}$ is given by the following integral equation:

$$\int_{v_{T-1}} R_{T-1}\left[v_{T-1};\hat{d}_{T-1}(v_{T-1};\hat{\theta}_{T-1})\right]\overline{W}(v_{T-1})dv_{T-1}=0$$

As discussed in this section, given this criterion, the collocation method is a sensible choice for computing $\tilde{\theta}_{T-1}$. Then the choice for the weighting functions, as used in Binder et al. (2000), is the $n_{x,T-1}, n_{y,T-1}$ Dirac delta functions $\delta\left(x_{T-1}-x_{T-1}^{i}, y_{T-1}-y_{T-1}^{i}\right)$, where x_{T-1}^{i} and y_{T-1}^{i} are chosen such that $\tilde{x}_{T-1}^{i}$ and $\tilde{y}_{T-1}^{i}$ are the $n_{x,T-1}$ and $n_{y,T-1}$ zeros of the Chebyshev polynomials forming the basis of the approximation $\hat{d}_{T-1}\left(v_{T-1};\theta_{T-1}\right)$. The zeros for the Chebyshev polynomials are given by

$$\tilde{v}_{T-1}^{i}=\begin{pmatrix}\cos\dfrac{(2i-1)\pi}{2n_{x,T-1}}\\ \cos\dfrac{(2i-1)\pi}{2n_{y,T-1}}\end{pmatrix} \tag{1.6.21}$$

Then the integral equation can be reduced to:

$$R_{T-1}\left(v_{T-1}^{ij};\hat{d}_{T-1}^{ij}\right)=0 \tag{1.6.22}$$

for all

$$v_{T-1}^{ij}=\left(x_{T-1}^{i},y_{T-1}^{j}\right),\quad i=1,2,...,n_{x,T-1},\quad j=1,2,...,n_{y,T-1} \tag{1.6.23}$$

and

$$\hat{d}_{T-1}^{ij}=\hat{d}_{T-1}\left(v_{T-1}^{ij};\hat{\theta}_{T-1}\right) \tag{1.6.24}$$

The discrete orthogonality of Chebyshev polynomials implies that:

$$\sum_{i=1}^{n_{x,T-1}}\sum_{j=1}^{n_{y,T-1}}\left[p_{w-1}\left(\tilde{x}_{T-1}^{i}\right)p_{p-1}\left(\tilde{y}_{T-1}^{j}\right)\right]\left[p_{s-1}\left(\tilde{x}_{T-1}^{i}\right)p_{q-1}\left(\tilde{y}_{T-1}^{j}\right)\right]=0 \tag{1.6.25}$$

for $w\neq s$ and /or $p\neq q$, and

$$\sum_{i=1}^{n_{x,T-1}}\sum_{j=1}^{n_{y,T-1}}\left[p_{w-1}\left(\tilde{x}_{T-1}^{i}\right)p_{p-1}\left(\tilde{y}_{T-1}^{j}\right)\right]\left[p_{s-1}\left(\tilde{x}_{T-1}^{i}\right)p_{q-1}\left(\tilde{y}_{T-1}^{j}\right)\right]=c_{sq}\left(n_{x,T-1},n_{y,T-1}\right) \tag{1.6.26}$$

for $w=s$ and $p=q$, with

$$c_{sq}\left(n_{x,T-1},n_{y,T-1}\right)=\begin{cases}n_{x,T-1}n_{y,T-1}, & w=s=p=q=1\\ \left(n_{x,T-1}n_{y,T-1}\right)/2, & \begin{cases}w=s=1\ and\ p=q\neq 1,\\ or\\ w=s\neq 1\ and\ p=q=1,\end{cases}\\ \left(n_{x,T-1}n_{y,T-1}\right)/4, & w=s\neq 1\ and\ p=q\neq 1\end{cases} \tag{1.6.27}$$

Then θ is given by:

$$\begin{aligned}\hat{\theta}_{T-1,sq}=&\frac{1}{c_{sq}\left(n_{x,T-1},n_{y,T-1}\right)}\sum_{i=1}^{n_{x,T-1}}\sum_{j=1}^{n_{y,T-1}}p_{s-1}\left(\tilde{x}_{T-1}^{i}\right)p_{q-1}\left(\tilde{y}_{T-1}^{j}\right)\cdot\left[g\left(v_{T-1},\hat{d}_{T-1}^{ij},y_{T-1}\right)\cdot g'_{x_{T-2}}\left(v_{T-1},\hat{d}_{T-1}^{ij},y_{T-1}\right)\right.\\ &\left.+\left(\tilde{\pi}'\right)^{-1}\left(E\left\{\beta\pi'\left[g\left(\hat{d}_{T-1}^{ij},x_{T},y_{T}\right)\cdot g'_{x_{T-1}}\left(\hat{d}_{T-1}^{ij},x_{T},y_{T}\right)\right]\middle|v_{T-1}\right\}\right)\right]\end{aligned} \tag{1.6.28}$$

for $s=1,2,...,n_{x,T-1}$, $q=1,2,...,n_{y,T-1}$.

The conditional expectation from the above equation needs to be computed numerically. In order to compute the integral one can use some of the quadrature methods such as the Gauss quadrature presented in section I.5.2. All that remains is to solve equation (1.6.28) for $\theta_{T-1,sq}$, $s=1,2,...,n_{x,T-1}$, $q=1,2,...,n_{y,T-1}$. Once $\hat{d}_{T-1}\left(v_{T-1};\hat{\theta}_{T-1}\right)$ is computed, one can proceed recursively backwards to period $T-2$. Note that $x_{T-1}^{*}=\hat{d}_{T-1}\left(v_{T-1};\hat{\theta}_{T-1}\right)$ will be used in the definition of $R_{T-2}\left(v_{T-2}^{ij};\hat{d}_{T-2}^{ij}\right)$. The computation of $\hat{\theta}_{T-2}$ can now follow the same logic as the computation of $\hat{\theta}_{T-1}$.

So far the flavors of the projection methodology have been categorized either with respect to the choice of the approximation criterion or with respect to the method employed for estimating the parameters. The choice of basis functions for the representation in (1.6.6) can be used to further divide projection methods into two categories: spectral methods and finite-element methods. Spectral methods use basis functions that are smooth and non-zero on most of the domain of x such as Chebyshev polynomials and the same functions are used on all regions of the state space. Finite-element methods use basis functions that are equal to zero on most of the domain and non-zero on only a few subdivisions of the domain of x (these are in general piecewise linear functions such as those defined in (1.6.9)) and they provide different approximations in different regions of the state space. For problems with many state variables, there are typically many coefficients to compute and it implies the inversion of a large, dense matrix. With the finite-element method, however, the same matrix is sparse and its structure can typically be exploited. For the above-mentioned reasons McGrattan

(1996, 1999) argues that a finite-element method is better suited to problems in which the solution is nonlinear or kinked in certain regions.

I.6.2. Parameterized Expectations

While Marcet (1988) is largely credited in the literature with the introduction of the parameterized expectations approach, Christiano and Fisher (2000) point out that the underlying idea of parameterized expectations seems to have surfaced earlier in the work of Wright and Williams (1982a, 1982b, 1984), and then in the work of Miranda and Helmberger (1988). Marcet (1988)[33] implemented a variation of that idea and the approach finally caught on with the publication of Den Haan and Marcet (1990).

In this section, I will concentrate on what Christiano and Fisher (2000) call the conventional parameterized expectations approach due to Marcet (1988). While one may argue that this methodology does not belong under the label of projection methods, I believe that it can be viewed as a special case of projection methods by virtue of its use of parameterized functions to approximate an unknown quantity, of an implicit choice of a residual function and an approximation criterion similar to projection methods. In addition, the techniques used to estimate the parameters are also common to projection methods. The assumption is that the functional equation has the following form:

$$g\left(E_t\left[\phi\left(\eta_{t+1},\eta_t\right)\right],\eta_{t-1},\eta_t,z_t\right)=0 \tag{1.6.29}$$

where η_t includes all the endogenous and exogenous variables and z_t is a vector of exogenous shocks. As it has been repeatedly asserted in this chapter, the reason why

[33] For more information of this variant of the parameterized expectations approach, see the references cited in Marcet and Marshall (1994b).

many dynamic models are difficult to solve is that conditional expectations often appear in the equilibrium conditions. The assumption under which this methodology operates is that conditional expectations are a time-invariant function ε of some state variables:

$$\varepsilon(\upsilon_t) = E_t\left[\phi(\eta_{t+1}, \eta_t)\right] \tag{1.6.30}$$

where $E_t\left[\phi(\eta_{t+1}, \eta_t)\right] = E\left[\phi(\eta_{t+1}, \eta_t)\middle|\upsilon_t\right]$ is the conditional expectation based on the available information at time t, $\upsilon_t \in R^l$ where υ_t is a subset of (η_{t-1}, z_t). As Marcet and Lorenzoni (1999) point out, a key property of ε is that under rational expectations, if agents use ε to form their decisions, the series generated is such that ε is precisely the best predictor of the future variables inside the conditional expectations. So, if ε were known, one could easily simulate the model and check whether this is actually the conditional expectation.

The basic approach of Marcet and Marshall (1994a) is to substitute the conditional expectations in equation (1.6.29) by parameterized functions of the state variables with arbitrary coefficients. Then (1.6.29) is used to generate simulations for υ_t consistent with the parameterized expectations. With these simulations, one can iterate on the parameterized expectations until they are consistent with the solution they generate. In this fashion, the process of estimating the parameters is reduced to a fixed-point problem.

I.6.2.1. Example

Consider again the model specified by (1.6.2) - (1.6.3) with the Euler equation for period t given by:

$$
\begin{aligned}
0 = {} & \pi'\left(g\left(x_t, x_{t-1}, y_t\right)\right) \cdot g'_{x_t}\left(x_t, x_{t-1}, y_t\right) \\
& + \beta E\left\{\pi'\left(g\left(x_{t+1}, x_t, y_{t+1}\right)\right) \cdot g'_{x_t}\left(x_{t+1}, x_t, y_{t+1}\right) \middle| \Omega_t\right\}
\end{aligned} \tag{1.6.31}
$$

The idea is to substitute

$$
E_t\left\{\pi'\left(g\left(x_{t+1}, x_t, y_{t+1}\right)\right) \cdot g'_{x_t}\left(x_{t+1}, x_t, y_{t+1}\right)\right\}
$$

by a parameterized function $\psi\left(x_{t-1}, y_t; \theta\right)$ where θ is a vector of parameters. For simplicity, let the function ψ be given by:

$$
\psi_t\left(x_{t-1}, y_t; \theta_1, \theta_2\right) = \theta_1 x_{t-1} + \theta_2 y_t \tag{1.6.32}
$$

The next step is to generate a series $\left\{z_t\right\}_{t=1}^{T}$ as draws from a Gaussian distribution and to choose starting values for the elements of θ, θ_i^0, $i = 1, 2$. Then, for $\hat{\theta}_i = \theta_i^0$ and assuming that the initial values for x_t and y_t, that is, x_{-1} and y_0 are given, one can use the following system

$$
\begin{aligned}
& \pi'\left(g\left(x_t, x_{t-1}, y_t\right)\right) \cdot g'_{x_t}\left(x_t, x_{t-1}, y_t\right) + \hat{\theta}_1 x_{t-1} + \hat{\theta}_2 y_t = 0 \quad \text{for } t = 0, ..., T-1 \\
& x_t = h\left(x_{t-1}, u_t, y_t\right) \qquad\qquad \text{for } t = 0, ..., T, \text{ with } x_{-1} \text{ given} \\
& y_t = \rho y_{t-1} + z_t \qquad\qquad \text{for } t = 1, ..., T, \text{ with } y_0 \text{ given}
\end{aligned} \tag{1.6.33}
$$

to generate series $\left\{\hat{x}_t^j\right\}_{t=0}^{T}$, $\left\{\hat{y}_t^j\right\}_{t=1}^{T}$ and $\left\{\hat{u}_t^j\right\}_{t=0}^{T}$ where j represents the iteration. In order to estimate the parameters θ, proponents of this methodology run a regression of

$$
\Upsilon_t^j\left(\hat{\theta}^j\right) = \pi'\left(g\left(\hat{x}_t^j, \hat{x}_{t-1}^j, \hat{y}_t^j\right)\right) \cdot g'_{x_t}\left(\hat{x}_t^j, \hat{x}_{t-1}^j, \hat{y}_t^j\right) \tag{1.6.34}
$$

on ψ_t. Formally, the regression can be written as:

$$
\Upsilon_t^j\left(\hat{\theta}^j\right) = a_1 \hat{x}_{t-1}^j + a_2 \hat{y}_t^j + \xi_t
$$

where ξ_t is the error term. The estimates for a_1 and a_2 provide a new set of values for θ for the next iteration. With those values new series will be generated for $\left\{\hat{x}_t^{j+1}\right\}_{t=0}^{T}$ and

$\left\{\hat{u}_t^{j+1}\right\}_{t=0}^{T}$. In this particular case, there is no need to generate new series for $\left\{\hat{y}_t^{j+1}\right\}_{t=1}^{T}$ if the same vector of shocks $\left\{z_t\right\}_{t=1}^{T}$ is used. In addition, note that a_1 and a_2 are in fact functions of $\hat{\theta}$. Specifically, for iteration j, the vector of parameters a is a function of $\hat{\theta}^j$, $a = G\left(\hat{\theta}^j\right)$. Hence the final step is to find the fixed point $\theta = G(\theta)$. One approach suggested by Marcet and Lorenzoni (1999) is to compute the values of $\hat{\theta}$ for iteration $j+1$ using the following expression $\hat{\theta}^{j+1} = (1-b)\hat{\theta}^j + bG\left(\hat{\theta}^j\right)$ where $b > 0$. The iteration process should stop when $\hat{\theta}^j$ and $G\left(\hat{\theta}^j\right)$ are sufficiently close.

I.6.3. Notes on Projection Methods

As Judd (1992) points out, the advantage of the projection method framework is that one can easily generate several different implementations by choosing among different basis, residual functions or methods for estimating the parameters. Obviously, the many choices also imply some trade-offs among speed, accuracy, and reliability. For example, the orthogonal collocation method tends to be faster than the Galerkin method, while the Galerkin method tends to offer more accuracy[34].

The generality of the projection techniques can also be seen from the fact that even methods that discretize the state space can be thought of as projection methods that are using step function bases.

While throughout this section I emphasized the wide applicability of projection methods, there is an aspect that has been overshadowed. Recall that the idea is to replace

[34] See Judd (1992) for more details.

the quantity that needs to be approximated by parameterized functions (basis functions $\varphi_i(x)$) with arbitrary coefficients (a_i). In projection methods, the coefficients are chosen to be the best possible choices relative to the basis $\varphi_i(x)$ and relative to some criterion. However, the bases are usually chosen to satisfy some general criteria, such as smoothness and orthogonality conditions. Such bases may be good but very rarely are they the best possible for the problem under consideration.

An important advantage of parameterized expectations approach is that, for specific models, it may implicitly deal with the presence of inequality constraints eliminating the need to constantly check whether the Kuhn-Tucker conditions are satisfied[35].

A key component of the conventional parameterized expectations approach presented in this section is a cumbersome nonlinear regression step. The regression step implies simulations involving a huge amount of synthetic data points. The problem with this approach is that it inefficiently concentrates on a residual amount that is obtained from visiting only high probability points of the invariant distribution of the model. As Pointed out by Judd (1992) and Christiano and Fisher (2000), it is important to consider the tail areas of the distribution as well. Christiano and Fisher (2000) offer a modified version of the parameterized expectations approach that they call the Chebyshev parameterized expectations approach, specifically designed to eliminate the shortcoming discussed above. In fact, Christiano and Fisher (2000) explicitly transform the parameterized expectations approach into a projection method that they refer to as the weighted residual parameterized expectations approach. As mentioned above, expressing

[35] See Christiano and Fisher (2000) for details.

the parameterized expectations approach as a projection method opens the door to a variety of possible implementations.[36].

I.7. Comparing Numerical Methods: Accuracy and Computational Burden

It is difficult to define the global criteria of success for numerical methods. Accuracy is in general at the top of the checklist in defining a good numerical method. However, it may not always be the most important criterion when choosing a numerical method. For example, even though a method may not provide the best approximation for the policy function, it may still be preferred to other methods as long as the loss in accuracy relative to the policy function does not affect too much the value of the objective function. In such cases, speed or ease of implementation may take precedence.

There does not seem to be a general agreement in the literature on how to evaluate the accuracy of numerical methods. Consequently, a number of criteria have been proposed in order to asses the performance of numerical algorithms.

One widely used strategy for determining accuracy is to test the outcome of a computational algorithm in a particular case where the model displays an analytical solution. For example, Collard and Juillard (2001) use an average relative error and a maximal relative error criterion in order to asses the accuracy of several numerical methods. While this approach may be useful for certain specifications, the problem is that for alternative parameterizations of the model the approximation error of the computed decision and value functions may change substantially. Changes in the curvature of the objective function and in the discount factor are the usual culprits in influencing

[36] In fact, Christiano and Fisher (2000) provide two other modified versions of the parameterized expectations approach (PEA): PEA Galerkin and PEA collocation.

considerably the accuracy of the algorithm. Collard and Juillard (2001) determine that for an asset pricing model the Galerkin method using fourth order Chebyshev polynomials clearly outperforms linearization methods as well as lower order perturbation methods. However, higher order (order four and higher) perturbation methods prove to be quite accurate.

Another strategy used for analyzing the accuracy of numerical methods is to look at the residuals of the Euler equation. This seems like a natural choice especially for approaches that are based on approximating certain terms entering, or the whole, Euler equation[37].

A procedure for checking accuracy of numerical solutions based on the Euler equation residuals was proposed by den Haan and Marcet (1990, 1994). It consists of a test for the orthogonality of the Euler equation residuals over current and past information. The idea behind this test is to compute simulated time series for all the choice and state variables as well as Euler equation residuals, based on a candidate approximation. Then, using estimated values of the coefficients resulting from regressing the Euler equation residuals on lagged simulated time series, one can construct measures of accuracy. As pointed out by Santos (2000), the problem with this approach is that orthogonal Euler equation residuals may be compatible with large deviations from the optimal policy. In addition, as referenced by Judd (1992), Klenow (1991) found that the procedure failed to reject candidate solutions that resulted in relatively high errors for the choice variable while rejecting solutions resulting in occasional high large errors but without any discernible pattern.

[37] For a detailed discussion on criteria involving Euler equation residuals, please see Reiter (2000) and Santos (2000).

Judd (1992, 1998) suggested an alternative test that consists of computing a one period optimization error relative to the decision rule. The error is obtained by dividing the current residual of the Euler equation to the value of next period's decision function. Subsequently, two different norms are applied to the error term: one gives the average and the other supplies the maximum.

In a study aimed at comparing various approximation methods, Taylor and Uhlig (1990) found that performance varies greatly depending on the criterion used for assessing accuracy. For example, the decision rules indicated that some of the easier to implement methods such as the linear-quadratic method and the extended-path method were fairly close to the "exact" decision rule[38] as given by the quadrature-value-function-grid method of Tauchen (1990) or the Euler-equation grid method of Coleman (1990). However, neither the linear-quadratic nor the extended-path method performed well when using the martingale-difference tests for the Euler-equation residual. Not surprisingly, the parameterized expectations approach performed well when using the den Haan and Marcet criterion but not as well when measured against the exact decision rule.

While accuracy is very important, computational time may also play an important role in the eyes of some researchers. While the extended-path method has relatively low cost when compared to grid methods, it is fair to state that both grid methods and the extended-path method are computationally quite involved, whereas linear-quadratic methods are typically quite fast. Most projection methods also fare well in terms of

[38] Solutions obtained through discretization methods are sometimes referred to as "exact". The reason behind this labeling is that models obtained as a result of discretization may be solved exactly by finite-state dynamic programming methods. However, one has to keep in mind that reducing a continuous-state problem to a finite-state problem still involves an approximation error.

computational burden when compared to discretization methods or even parameterized expectations methods. As the state space increases, discretization methods suffer heavily from the curse of dimensionality.

The fact that none of the methods outperforms the others does not mean that every method could be applied to any model out there with a good degree of success[39]. One has to use good judgment when deciding on using a certain numerical method.

I.8. Concluding Remarks

As it has become clear over the course of this chapter, there are quite a few methodologies available for solving non-linear rational expectations models. However, if one looks closer, it becomes obvious that all methods share some common elements. For example, certainty equivalence is at the core of the extended path method but it can also be used in perturbation methods to find the equilibrium of a (deterministic) system similar to the one under investigation. The discrete state space approach can be viewed as a projection method with step functions as a basis. Similarly, the first order perturbation method is nothing more than a simple linearization around steady state. In addition, the parameterized expectations approach can be easily transformed into a projection method. Moreover, since all the functional equations for rational expectations models imply the existence of some integrals, the quadrature approximation may make an appearance in almost every methodology.

[39] Judd (1998) contains an example of a partially revealing rational expectations problem which cannot be solved by discretizing the state space, but which can be approximated by more general projection methods.

Several studies have tried to asses the performance of these numerical methods. However, even for relatively simple models their performance may vary greatly[40]. Despite all of their sophistication, none of these methods can consistently outperform the others.

Even comparing the methods is not a walk in the park. Several authors including Judd (1992), Den Haan and Marcet (1994), Collard and Juillard (2001), Santos (2000) and Reiter (2000) proposed different criteria for evaluating the performance of numerical solutions. Unfortunately, each criterion has its caveats and it has to be applied selectively, based on the specificity of the model under investigation. Therefore, one has to choose carefully the proper methodology when in need of numerical solutions.

[40] See the studies by Taylor and Uhlig (1990), Judd (1992), Rust (1997), Christiano and Fischer (2000), Santos (2000), Collard and Juillard (2001), Fair (2003), Schmitt-Grohé and Uribe (2004).

Chapter II. Using Scenario Aggregation Method to Solve a Finite Horizon Life Cycle Model of Consumption

II.1. Introduction

Multistage optimization problems are a very common occurrence in the economic literature. While there exist other approaches to solving such problems, many economic models involving intertemporal optimizing agents assume that the representative agent chooses its actions as a result of solving some dynamic programming problem. Lately, an increasing number of researchers have investigated alternative approaches to modeling the representative agent, in an attempt to find one that may explain observed facts better or easier. Following the same line of research, I explore the suitability of scenario aggregation method as an alternative to describe the decision making process of an optimizing agent in economic models. The idea is that this methodology offers a different approach that might be more consistent with the observation that agents are more likely to behave like chess players, making decisions based only on a subset of all possible outcomes and using a relatively short horizon[41]. The advantage of scenario aggregation methodology is that, while it presents attractive features for use in models assuming bounded rationality, it can also be seen as an alternative numerical method that can be used for obtaining approximate solutions for rational expectation models. Therefore, I start by studying in this chapter the viability of the scenario aggregation method, as

[41] In the next chapter I will focus more on the length of the span over which the decision making process takes place.

presented by Rockafellar and Wets (1991), to provide a good approximation for the optimal solution of a simple finite horizon life-cycle model of consumption with precautionary savings. In the next chapter, I will use scenario aggregation to model the decision making of the rationally bounded consumer.

The layout of this chapter is as follows. First, I present the setup of a simple life-cycle consumption model with precautionary saving. Then, I introduce the notion of scenarios followed by a description of the aggregation method. Next, I introduce the progressive hedging algorithm followed by its application to a finite horizon life-cycle consumption model. Then, I present simulation results and conclude the chapter with final remarks.

II.2. A Simple Life-Cycle Model with Precautionary Saving

I consider the following version of a life-cycle model. Suppose an individual agent is faced with the following intertemporal optimization problem:

$$\max_{\{c_t\}_{t=0}^{T}} E\left[\sum_{t=0}^{T} \beta^t F_t\left(c_t\right) \mid I_0\right] \tag{2.2.1}$$

where F_t is a utility function which has the typical properties assumed in the literature, i.e. it is twice differentiable, it is increasing with consumption and exhibits negative second derivative. The information set I_0 contains the level of consumption, assets, labor income and interest rate for period zero and all previous periods.

Maximization is subject to the following transition equation:

$$A_t = \left(1 + r_t\right) A_{t-1} + y_t - c_t, \qquad t = 0,1,...,T-1, \tag{2.2.2}$$

$$A_t \geq -b, \quad \text{with } A_{-1}, A_T \text{ given} \tag{2.2.3}$$

where A_t represents the level of assets at the beginning of period t, y_t the labor income at time t, and c_t represents consumption in period t. The initial and terminal conditions, A_{-1} and A_T, are given. Uncertainty is introduced in the model through the labor income. The realizations of the labor income are described by the following process:

$$y_t = y_{t-1} + \xi_t, \quad t = 1,...,T, \text{ with } y_o \text{ given} \tag{2.2.4}$$

and ξ_t being drawn from a normal distribution, $\xi_t \sim N\left(0, \sigma_y^2\right)$. For now, I will not make any particular assumption about the process generating the interest rate, r_t. Therefore, to summarize the model, a representative consumer derives utility in period t from consuming c_t, discounts future utility at a rate β and wants, in period zero, to maximize his present discounted value of future utilities for a horizon of $T+1$ periods. At the beginning of each period t the consumer receives a stochastic labor income y_t, and based on the return on his assets A_{t-1}, from the beginning of period $t-1$ to the beginning of period t, he chooses the consumption level c_t, and thus determines the level of assets A_t according to equation (2.2.2).

Of particular importance in this problem is the random variable ξ_t. In the standard formulation of the problem, ξ_t is assumed to be distributed normally with mean zero and some variance σ_y^2. Instead of making the standard assumption, if I assume that ξ_t's sample space has only a few elements, then the optimization problem (2.2.1) - (2.2.4) is a perfect candidate for being solved using the scenario aggregation method. Let me assume for the moment that the sample space is given by $\{\omega_1, \omega_2, ..., \omega_n\}$ with the associated probabilities $\{p_1, p_2, ..., p_n\}$. If S is the set of all scenarios then its cardinal is

given by n^T. It is obvious that as the sample space for the forcing variable increases, the number of scenarios increases proportional with power T. Therefore, applying the scenario aggregation method to find an approximate solution for this problem may only be feasible when T and n are relatively small. In the next chapter, I will present a solution for T relatively large.

II.3. The Concept of Scenarios

II.3.1. The Problem

In this section, I formally introduce a multistage optimization problem and then, in the following sections, I will present the idea of scenario aggregation and how it can be applied to such a problem.

The multistage stochastic optimization problem consists of minimizing an objective function, $F : R^m \to R$ subject to some constraints, which usually describe the dynamic links between stages.

The objective function F is time separable and is given by a sum of functions, $F = \sum_{t=0}^{T} F_t$ with each function F_t, $F_t : R^m \to R$ corresponding to stage t of the optimization problem. These functions depend on a set of variables u_t, which in turn represent the decisions that need to be made at each stage t. For simplicity I assume that u_t is a $m_u \times 1$ vector, with m_u independent of t, that is, the same number of decisions is to be made at each stage.

If $U(t)$ represents the set of all feasible actions at stage t, then u_t has to be part of the set $U(t)$, that is, $u_t \subset U(t), \quad t = 0,...,T, \quad U(t) \subseteq R^{m_u}$. The temporal dimension of the problem is characterized by stages t and state variables $X(t)$.

The link between stages is given by:

$$x_{t+1} = G_t(x_t, u_t, u_{t+1}).$$

Hence, the problem can be formulated as:

$$\min E\left(\sum_{t=0}^{T} F_t\left(x_t, u_t\right) | I_0 \right) \tag{2.3.1}$$

subject to:

$$x_{t+1} = G_t(x_t, u_t, u_{t+1}, \xi_t) \tag{2.3.2}$$

where I_0 is the information set at time $t = 0$ and ξ_t is the forcing variable.

In the next few sections, I will present the concept of scenarios as well as possible decomposition methods along with the idea of scenario aggregation.

II.3.2. Scenarios and the Event Tree

In this section, I present an intuitive description for the concept of scenarios. A formal description is presented in Appendix, section A1. Suppose the world can be described at each point in time by the vector of state variables x_t. In the case of a multistage optimization problem, let u_t denote the control variable and let ξ_t be the forcing variable. I assume that an agent makes decisions reflected in the control variable u_t. For simplicity let ξ_t be a random variable witch can take two values ξ_a and ξ_b with probabilities p_a and $1 - p_a$.

If the horizon has $T+1$ time periods and $\{\xi_a, \xi_b\}$ is the set of possible realizations for ξ_t then the sequence

$$\xi^s = \left(\xi_0^s, \xi_1^s, \ldots, \xi_T^s\right)$$

is called a scenario[42]. From now on, for notation simplification, I will refer to a scenario s simply by ξ^s or by the index s. Given that the set of all realizations for ξ_t is finite, one can define an event tree $\{N, A\}$ characterized by the set of nodes N and the set of arcs A. In this representation, the nodes of the tree are decision points and the arcs are realizations of the forcing variables. The arcs join nodes from consecutive levels such that a node n_t^j at level t is linked to N_{t+1} nodes, $n_{t+1}^k, k = 1,..., N_{t+1}$ at level $t+1$. In Figure 1 I represent such a tree for a span of $T = 3$ periods. As mentioned above, the forcing variable takes only two values, $\{\xi_a, \xi_b\}$ and hence the tree has 15 nodes. The arcs that join nodes from consecutive levels represent realizations of the forcing variable and are labeled accordingly.

The set of nodes N can be divided into subsets corresponding to each level (period). Suppose that at time t there are N_t nodes. For example, for $t = 1$, there are two nodes, node2 and node3. The arcs reaching these two nodes belong each to several scenarios s. The bundle of scenarios that go through one node plays a very important role in the decomposition as well as in the aggregation process. The term equivalence class has been used in the literature to describe the set of scenarios going through a particular node.

[42] Other definitions of scenarios can be found in Helgason and Wallace (1991a, 1991b) and Rosa and Ruszczynski (1994).

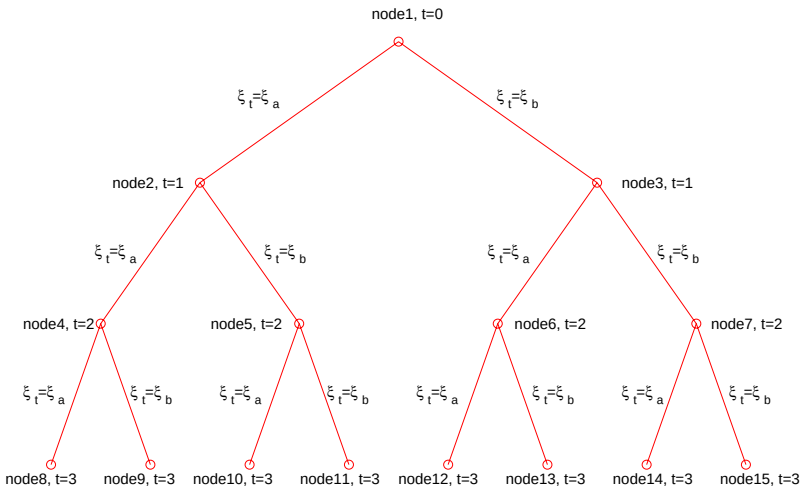

Figure 1 Event tree

By definition, an equivalence class at time t is the set of all scenarios having the first $t+1$ realizations common. As mentioned in the above description of the event tree, at time t there are N_t nodes. Every node is associated with an equivalence class. Then, the number of distinct equivalence classes at time t is also N_t.

In Figure 2 one can see that for $t=1$ there are two nodes and consequently two equivalence classes, $\{s_1, s_2, s_3, s_4\}$ and $\{s_5, s_6, s_7, s_8\}$. The number of elements of an equivalence class is given by the number of leaves stemming from the node associated with it. In this example, the number of leaves stemming from both nodes is four, which is also the number of scenarios belonging to each class.

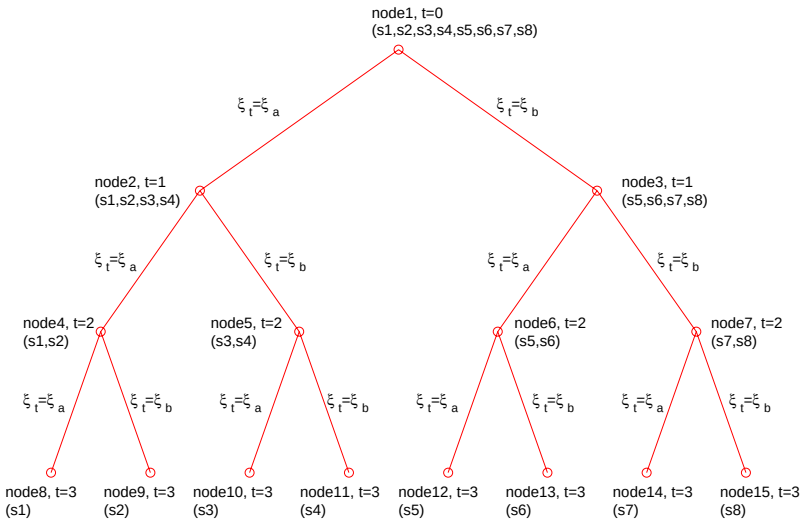

Figure 2 Equivalence classes

The transition from a state at time t to one at time $t+1$ is governed by the control variable u_t but is also dependent on the realization of the forcing variable, that is, on a particular scenario s. Since scenarios will be viewed in terms of a stochastic vector ξ with stochastic components $\xi_0^s, \xi_1^s, \ldots, \xi_T^s$, it is natural to attach probabilities to each scenario. I denote the probability of a particular realization of a scenario, s, with

$$p(s) = prob(\xi^s).$$

Let us consider the case of the event trees represented in Figure 1 and Figure 2 and assume the probability of realization ξ_a is $prob(\xi_t = \xi_a) = p_a$ while the probability of realization ξ_b, is $prob(\xi_t = \xi_b) = p_b$, with $p_a + p_b = 1$. Then, due to independence

across time, one can compute the probability of realization for scenario s_1, $prob(\xi^s = s_1) = p_a^3$. Similarly, the probability of realization for scenario s_2 is $prob(\xi^s = s_2) = p_a^2 p_b$, or $prob(\xi^s = s_2) = p_a^2 (1 - p_a)$.

Further on, I define[43] the probabilities associated with a scenario conditional upon belonging to a certain equivalence class at time t. For example, the probability associated with scenario s_1, conditional on s_1 belonging to equivalence class $\{s_1, s_2, s_3, s_4\}$ is given by $prob\left(s_1 \mid s_1 \in \{s_1, s_2, s_3, s_4\}\right) = p_a^2$

II.4. Scenario Aggregation[44]

In this section, I will show how a solution can be obtained by using special decomposition methods, which exploit the structure of the problem by splitting it into manageable pieces, and then aggregate their solutions. In the multistage stochastic optimization literature, there are two groups of methods that have been discussed: primal decomposition methods that work with subproblems that are assigned to time stages[45] and dual methods, in which subproblems correspond to scenarios[46]. Most of the methods, regardless of which group belong to, use the general theory of augmented Lagrangian decomposition. In this chapter I will concentrate on a methodology that belongs to the second group and has been derived from the work of Rockafellar and Wets (1991).

---

[43] For a more formal definition, see the Appendix, section A1.

[44] Section A2 in the Appendix offers a more formal description of scenario aggregation.

[45] See the work of Birge (1985), Ruszczynski (1986, 1993), Van Slyke and Wets (1969).

[46] See the work of Mulvey and Ruszczynski (1992), Rockafellar and Wets (1991), Ruszczynski (1989), Wets (1988).

Let us assume for a moment that the original problem can be decomposed into subproblems, each corresponding to a scenario. Then the subproblems can be described as:

$$\min_{u_t \in U_t \subseteq R^{m_u}} \sum_{t=1}^{T} F_t\left(x_t^s, u_t^s\right), \qquad s \in S \tag{2.4.1}$$

where u_t^s and x_t^s are the control and the state variable respectively, conditional on the realization of scenario s while S is a finite, relatively small set of scenarios. Moreover, suppose that each individual subproblem can be solved relatively easy. The question then becomes how to blend the individual solutions into a global optimal solution. Let the term policy[47] describe a set of chosen control variables for each scenario and indexed by the time dimension.

The policy function has to satisfy certain constraints if two different scenarios s and s' are indistinguishable at time t on information available about them at the time. Then $u_t^s = u_t^{s'}$, that is, a policy can not require different actions at time t relative to scenarios s and s' if there is no way to tell at time t which of the two scenarios will be followed. In the literature, this constraint is sometimes referred to as the non-anticipativity constraint. Going back to Figure 2, for $t = 1$, if the realization of ξ_t is ξ_a, the decision maker will find himself at the decision point node2. There are four scenarios that pass through node2 and the non-anticipativity constraint requires that only one decision be made at that point since the four scenarios are indistinguishable. A policy is

[47] A formal description of the policy function is presented in Appendix.

defined as implementable if it satisfies the non-anticipativity constraint, that is, u_t must be the same for all scenarios that have common past and present[48].

In addition, a policy has to be admissible. A policy is admissible if it always satisfies the constraints imposed by the definition of the problem. It is clear that not all admissible policies are also implementable.

By definition, a contingent policy[49] is the solution, u^s, to a scenario subproblem. It is obvious that a contingent policy is always admissible but not necessarily implementable. Therefore, the goal is to find a policy that is both admissible and implementable. Such a policy is referred to as a feasible policy. One way to create a feasible policy from a set on contingent policies is to assign weights (or probabilities) to each scenario and then aggregate the contingent policies according to these weights.

The question that the scenario aggregation methodology answers is how to obtain the optimal solution $\overline{U}$ from a collection of implementable policies $\hat{U}$. In this chapter, I will present a version of the progressive hedging algorithm originally developed by Rockafellar and Wets (1991).

[48] For certain problems the non-anticipativity constraint can also be defined in terms of the state variable, that is, $x_t(\omega)$ must be the same for all scenarios that have common past and present.

[49] I borrow this term from Rockafeller and Wets (1991).

II.5. The Progressive Hedging Algorithm

The algorithm is based on the principle of progressive hedging[50] which consists of starting with an implementable policy and creating sequences of improved policies in an attempt to reach the optimal policy.

Let us go back to the definition of an implementable policy. By computing

$$\hat{u}_t^s = \sum_{s' \in \{s^t\}_i} p\left(s' \middle| \{s^t\}_i\right) u_t^{s'} = E\left(u_t^{s'} \mid \{s^t\}_i\right) \quad \text{for all } s \in \{s^t\}_i \tag{2.5.1}$$

for all scenarios $s \in S$ and all periods $t = 1,...,T$, one creates a starting collection of implementable policies, denoted by $\hat{U}^0$. In equation (2.5.1) E represents the expectation operator. Therefore, in order to obtain an initial collection of implementable policies one should first compute some contingent policies for each scenario and then apply the expectation operator for each period t and each scenario s conditional on it belonging to the corresponding equivalence class, $\{s^t\}_i$.

The progressive hedging algorithm finds a path from $\hat{U}^0$, the set of implementable policies, to $\overline{U}$, the set of optimal policies, by solving a sequence of problems in which the scenarios subproblems are not the original ones, but a modified version of those by including some penalty terms. The algorithm is an iterative process starting from $\hat{U}^0$ and computing at each iteration k a collection of contingent policies U^k which are then aggregated into a collection of implementable policies $\hat{U}^k$ that are supposed to converge to the optimal solution $\overline{U}$. The contingent policies U^k are found as optimal solutions to the modified scenario subproblems:

[50] This term was coined by Rockafellar and Wets (1991). The idea is based on the theory of the proximal point algorithm in nonlinear programming.

$$\min F^s\left(x^s, u^s\right) + w^s u^s + \frac{1}{2}\rho\left|u^s - \hat{u}^s\right|^2 \tag{2.5.2}$$

where $|\cdot|$ is the ordinary Euclidian norm, ρ is a penalty parameter and w^s is an information price[51] . The use of ρ is justified by the fact that the new contingent policy should not depart too much from the implementable policy found in the previous iteration. The modified scenario subproblems (2.5.2) have the form of an augmented Lagrangian.

In the next subsection, I present a detailed description of the progressive hedging algorithm, which uses subproblems in the form of an augmented Lagrangian as shown above.

II.5.1. Description of the Progressive Hedging Algorithm

The optimal solution of the problem described by equations (2.3.1) - (2.3.2), $\overline{U}$, represents the best response an optimizing agent can come up with in the presence of uncertainty. An advantage of this algorithm is that one does not necessarily need to solve subproblems (2.5.2) exactly. A good approximation[52] of the solution is enough in allowing one to solve for the solution of the global problem.

Let U^k denote a collection of admissible policies and W^k a collection of information prices corresponding to iteration k. The progressive hedging algorithm, as designed by Rockafellar and Wets (1991), consists of the following steps:

---

[51] I borrow this term from Rockafellar and Wets (1991).

[52] One can envision transforming the scenario subproblems into quadratic problems by using second order Taylor approximations.

Step 0. Choose a value for ρ, W^0 and for U^0. The value of ρ may remain constant throughout the algorithm but it can also be adjusted from iteration to iteration[53]. Changing the value of ρ may improve the speed of convergence. Throughout this chapter, I will consider ρ as being constant. U^0 can be composed of the contingent policies $u^{s(0)} = \left(u_1^{s(0)}, u_1^{s(0)}, ..., u_T^{s(0)}\right)$ obtained from solving all the scenarios subproblems, whether modified or not. W^0 can be initialized to zero, $W^0 = 0$. Calculate the collection of implementable policies, $\hat{U}^0 = JU^0$, where J is the aggregation operator[54].

Step 1. For every scenario $s \in S$, solve the subproblem:

$$\min \sum_{t=1}^{T} \left[F_t^s \left(x_t^s, u_t^s\right) + w_t^s u_t^s + \frac{1}{2}\rho \left| u_t^s - \hat{u}_t^s \right|^2 \right] \tag{2.5.3}$$

For iteration $k+1$, let $u^{s(k+1)} = \left(u_1^{s(k+1)}, u_2^{s(k+1)}, ..., u_T^{s(k+1)}\right)$ denote the solution to the subproblem corresponding to scenario s. This contingent policy is admissible but not necessarily implementable. Let U^{k+1} be the collection of all contingent policies $u^{s(k+1)}$.

Step2. Calculate the collection of implementable policies, $\hat{U}^{k+1} = JU^{k+1}$. While these policies are implementable, they are not necessarily admissible in some cases[55]. If the policies obtained are deemed a good approximation, the algorithm can stop. A stopping criterion should be employed in this step.

---

[53] See Rockafeller and Wets (1991) and Helgason and Wallace (1991a, 1991b) for a discussion on the values of ρ. Rosa and Ruszczynski (1994) also provide some algorithm for updating similar penalty parameters.

[54] See the appendix for more details on the aggregation operator.

[55] Contingent policies are always admissible. If the domain of admissible policies is convex then any linear combination of the contingent policies will also belong to that domain. As noted above, by definition, the aggregation operator is linear. Therefore, for a convex problem the implementable policies computed in step 1 are also admissible.

Step3. Update the collection of information prices W^{k+1} by the following rule:

$$W^{k+1} = W^k + \rho\left(U^k - \hat{U}^k\right) \tag{2.5.4}$$

For each scenario $s \in S$ rule (2.5.4) translates into:

$$w_t^{s(k+1)} = w_t^{s(k)} + \rho\left(u_t^{s(k)} - \hat{u}_t^{s(k)}\right) \quad \text{for} \quad t = 1,...,T \tag{2.5.5}$$

This updating rule is derived from the augmented Lagrangian theory. In principle, the rule can be changed with something else as long as the decomposition properties are not altered.

Step 4. Reassign $k := k+1$ and go back to step one.

Next, I investigate how this methodology can be applied to a type of dynamic programming problem closed to what is often employed by economists for their models.

II.6. Using Scenario Aggregation to Solve a Finite Horizon Life Cycle Model

In this section, I will take a closer look at the viability of scenario aggregation in approximating a rational expectations model. I choose a standard finite horizon life cycle model that has an analytical solution, which will be used as a benchmark for the performance of the scenario aggregation method.

I start by presenting an algorithm for solving the problem given by (2.2.1) - (2.2.4) under the assumption that the length of the horizon, T, and the number of realizations of the forcing variable, n, are relatively small. The algorithm used is similar to that developed by Rockefeller and Wets (1991). As mentioned above, the idea is to split the problem into many smaller problems based on scenario decomposition and solve those problems iteratively imposing the non-anticipativity constraint. For computational convenience, I will reformulate the problem (2.2.1) - (2.2.4) as a minimization rather than

maximization. Hence, for each scenario $s \in S$, represented by the sequence of realizations $y^s = \left(y_0^s, y_1^s, \ldots, y_T^s\right)$, the problem becomes:

$$\min_{c_t} \left\{ \sum_{t=0}^{T} \beta^t \left[-F_t\left(c_t^s\right) + w_t^s c_t^s + \frac{1}{2}\rho\left(c_t^s - \underline{c}_t^{\,s}\right)^2 \right] \right\} \tag{2.6.1}$$

subject to

$$A_t^s = \left(1 + r_t^s\right) A_{t-1}^s + y_t^s - c_t^s, \quad t = 0,1,...,T \tag{2.6.2}$$

Expressing c_t^s and $\underline{c}_t^{\,s}$ as a function of A_t^s and $\underline{A}_t^s$, the augmented Lagrangian function, for a fixed scenario s, becomes:

$$\begin{aligned} L = \sum_{t=0}^{T} \beta^t \Big\{ & -F_t\left[\left(1 + r_t^s\right) A_{t-1}^s + y_t^s - A_t^s\right] + w_t^s\left[\left(1 + r_t^s\right) A_{t-1}^s + y_t^s - A_t^s\right] + \\ & + \frac{1}{2}\rho\left[\left(1 + r_t^s\right) A_{t-1}^s + y_t^s - A_t^s - \left(\left(1 + r_t^s\right)\underline{A}_{t-1}^s + y_t^s - \underline{A}_t^s\right)\right]^2 \Big\} \end{aligned} \tag{2.6.3}$$

All the underlined variables in the above equations represent implementable policies or states derived from applying implementable policies.

Before going through the steps of the algorithm, I will make a few assumptions about the functional form of the utility function as well as about the interest rate. First, it is assumed that preferences are described by a negative exponential utility function. Hence:

$$F_t\left(c_t\right) = -\frac{1}{\theta}\exp\left(-\theta c_t\right) \tag{2.6.4}$$

where θ is the risk aversion coefficient. Secondly, the interest rate, r_t, is taken to be constant. Finally, the distribution of the forcing variable is approximated by a discrete counterpart. The realizations as well as the associated probabilities are obtained using a Gauss-Hermite quadrature and matching the moments up to order two. The number of points used to approximate the original distribution determines the number of scenarios.

By decomposing the original problem into scenarios, the subproblems become deterministic versions of the original model.

II.6.1. The Algorithm

Given the assumptions made in the previous section, problem (2.6.1) becomes:

$$\min_{c_t}\left\{\sum_{t=0}^{T}\beta^t\left[\frac{1}{\theta}\exp\left(-\theta c_t^s\right)+w_t^s c_t^s+\frac{1}{2}\rho\left(c_t^s-\underline{c}_t^{\,s}\right)^2\right]\right\} \tag{2.6.5}$$

Consequently the Lagrangian for scenario s is:

$$\begin{aligned}L=\sum_{t=0}^{T}\beta^t\Bigg\{&\frac{1}{\theta}\exp\left[-\theta\left(\left(1+r\right)A_{t-1}^s+y_t^s-A_t^s\right)\right]+w_t^s\left[\left(1+r\right)A_{t-1}^s+y_t^s-A_t^s\right]+\\&+\frac{1}{2}\rho\left[\left(1+r\right)A_{t-1}^s+y_t^s-A_t^s-\left(\left(1+r\right)\underline{A}_{t-1}+y_t^s-\underline{A}_t\right)\right]^2\Bigg\}\end{aligned} \tag{2.6.6}$$

Since the consumption variable was replaced by a function of the asset level, the algorithm will be presented in terms of solving for the level of assets.

<u>*Step 0*</u>. Initialization: Set $w_t^s=0$ for all stages t and scenarios s. Choose a value for ρ that remains constant throughout the algorithm, let it be $\rho=5$. Later on, in this chapter, I will discuss the impact the value of ρ has on the convergence process. At this point, one needs a first set of policies. The convergence process, and implicitly the speed of the algorithm, is impacted by the choice of the first set of policies.

One suggestion made in the literature by Helgason and Wallace (1991a, 1991b) is to use the solution to the deterministic version of the model. This would amount to using the certainty equivalence solution in this case. I will first implement the algorithm using as starting point the certainty equivalence solution and then I will take advantage of the fact that for certain specifications of the model each scenario subproblem has an exact solution. I will then compare the convergence properties of the algorithm in these two cases.

Let $\left\{c_t^{ceq}\right\}_{t=0}^{T}$ denote the solution to the deterministic problem. Then, using the transition equation (2.6.2) one can compute the level of assets for each scenario s, $A^{s(0)} = \left\{A_0^{s(0)}, A_1^{s(0)}, ..., A_{T-1}^{s(0)}\right\}$. Next, it becomes possible to compute the implementable states $\underline{A}^{(0)} = \left\{\underline{A}_0^{(0)}, \underline{A}_1^{(0)}, ..., \underline{A}_{T-1}^{(0)}\right\}$ as a weighted average of $A_0^{s(0)}$ corresponding to all scenarios s, using as weights the probabilities of realization for each scenario. Alternatively, one can compute the first set of contingent policies by solving a deterministic life cycle consumption model for each scenario s:

$$\min_{A_t^s} \sum_{t=0}^{T} \beta^t \frac{1}{\theta} \exp\left\{-\theta\left[\left(1+r\right)A_{t-1}^s + y_t^s - A_t^s\right]\right\} \tag{2.6.7}$$

with $A_{-1}^{(s)}$ and $A_T^{(s)}$ given. As before, let $A^{s(0)} = \left\{A_0^{s(0)}, A_1^{s(0)}, ..., A_{T-1}^{s(0)}\right\}$ denote the solution to this problem. This solution is admissible but not implementable. The implementable solution for each period t, $\underline{A}_t^0$, is computed as the weighted average of all the contingent solutions for period t, $A_t^{s(0)}$, with the weights being given by the probability of realization for each particular scenario s.

<u>*Step 1.*</u> For every scenario $s \in S$, solve the subproblem:

$$\begin{aligned}\min_{A_t^s} \sum_{t=0}^{T} \beta^t \Bigg\{ & \frac{1}{\theta} \exp\left\{-\theta\left[\left(1+r\right)A_{t-1}^s + y_t^s - A_t^s\right]\right\} + \\ & + W_t^s\left[\left(1+r\right)A_{t-1}^s + y_t^s - A_t^s\right] + \\ & + \frac{1}{2}\rho\left\{\left(1+r\right)A_{t-1}^s + y_t^s - A_t^s - \left[\left(1+r\right)\underline{A}_{t-1} + y_t^s - \underline{A}_t\right]\right\}^2 \Bigg\}\end{aligned} \tag{2.6.8}$$

A detailed description of how the solution is computed can be found in the Appendix. The advantage of the scenario aggregation method is that the solution to problem (2.6.8) does not have to be computed exactly.

Let $A^{s(k)} = \left\{ A_0^{s(k)}, A_1^{s(k)}, ..., A_{T-1}^{s(k)} \right\}$ denote the contingent solution to this problem, where k denotes the iteration. Based on this solution I also compute the consumption path for each scenario, $c^{s(k)}$. This solution is admissible but not implementable and therefore the next step is to compute the implementable solution based on the contingent solutions $A^{s(k)}$.

Step 2. First, compute the implementable states $\underline{A}^k$. As it was mentioned in step 0, $\underline{A}_t^k$ is computed as the weighted average of all the contingent solutions for period t, $A_t^{s(k)}$, with the weights being given by the probability of realization for each particular scenario s. Since the space of the solutions for the problem being solved is convex, the implementable solution is also admissible. At this point, if solution $\underline{A}^k$ is considered good enough, the algorithm can stop and $\underline{A}^k$ becomes officially the solution of the problem described by (2.2.1) - (2.2.4). In order to make a decision on the viability of $\underline{A}^k$ as the optimal solution, one needs to define a stopping criterion. Based on the value of $\underline{A}^k$ I compute the implementable consumption path $\underline{c}^k$ and then use the following error sequence[56]:

$$\varepsilon^{(k)} = \sum_{t=0}^{T} \beta^t \left[\left(\underline{c}_t^{(k)} - \underline{c}_t^{(k-1)} \right)^2 + \left(\underline{A}_t^{(k)} - \underline{A}_t^{(k-1)} \right)^2 \right] \qquad (2.6.9)$$

where k is the iteration number. The termination criterion is $\varepsilon^{(k)} < d$ where d is arbitrarily chosen. In the next section, I will discuss the importance of the stopping criterion in determining the accuracy of the method.

[56] This is similar to what Helgason and Wallace (1990a) proposed. Later on in this chapter we will discuss the impact the choice of the value for d has on the results.

Step 3. For $t = 0,1,...,T$ and all scenarios s update the information prices:

$$w_t^{s(k+1)} = w_t^{s(k)} + \rho\left[(1+r)\left(A_{t-1}^{s(k)} - \underline{A}_{t-1}^{(k)}\right) - \left(A_t^{s(k)} - \underline{A}_t^{(k)}\right)\right] \quad \text{for} \quad t = 1,...,T$$

Step 4. Reassign $k := k+1$ and go back to step one.

II.6.2. Simulation Results

In this section, I present a brief picture of the results obtained by the implementation of the scenario aggregation method compared to the analytical solution. These results show that the numerical approximation obtained through scenario aggregation is close to the analytical solution for certain parameterizations of the model. In order to asses the accuracy of the scenario aggregation method I will use several criteria put forward in the literature. First, I compare the decision rule, i.e. the consumption path obtained through scenario aggregation with the values obtained from the analytical solution. In this context, I use two relative criteria similar to what Collard and Juillard (2001) use. One, E_R^a, gives the average departure from the analytical solution and is defined as:

$$E_R^a = \frac{1}{T+1}\sum_{t=0}^{T}\left|\frac{c_t^* - \underline{c}_t}{c_t^*}\right| \tag{2.6.10}$$

The other, E_R^m, represents the maximal relative error and is defined as:

$$E_R^m = \max\left\{\left|\frac{c_t^* - \underline{c}_t}{c_t^*}\right|\right\}_{t=0}^{T} \tag{2.6.11}$$

where c_t^* is the analytical solution and $\underline{c}_t$ is the value obtained through scenario aggregation. Alternatively, since the problem is ultimately solved in terms of the level of assets, the two criteria could also be expressed using the level of assets:

$$E_R^a = \frac{1}{T}\sum_{t=0}^{T-1}\left|\frac{A_t^* - \underline{A}_t}{A_t^*}\right|,\ E_R^m = \max\left\{\left|\frac{A_t^* - \underline{A}_t}{A_t^*}\right|\right\}_{t=0}^{T-1}$$

where A_t^* is given by the analytical solution and $\underline{A}_t$ by the scenario aggregation. Even though the scenario aggregation methodology does not use the Euler equation in obtaining the solution, I will use the Euler equation based criteria proposed by Judd (1998) as an alternative for determining the accuracy of the approximation. The criterion is defined as a one period optimization error relative to the decision rule. The measure is obtained by dividing the current residual of the Euler equation to the value of next period's decision function. Subsequently, two different norms are applied to the error term: one, E_E^a, gives the average and the other, E_E^m, supplies the maximum. Judd (1998) labeled these criteria as measures of bounded rationality.

The simulations were done using the following common set of parameter values: the discount factor $\beta = 0.96$; the initial and terminal values for the level of assets $A_{-1} = 500$ and $A_T = 1000$; the income generating process has a starting value of $y_0 = 200$. In addition, the interest rate is assumed deterministic. I used two values for the interest rate, $r = 0.04$ and $r = 0.06$. The distribution of the forcing variable was approximated by a 3 point discrete distribution. As I mentioned in the description of the progressive hedging algorithm, a few factors can influence the performance of the scenario aggregation method. Let us first look at how the starting values and stopping criterion influence the results.

II.6.2.1. Starting Values and Stopping Criterion

As I mentioned above, the starting values and the stopping criterion are very important elements in the implementation of the algorithm. I consider for the moment that the starting values are given by the certainty equivalence solution of the life cycle consumption model. I analyze the case where the value for the coefficient of risk aversion is $\theta = 0.01$, the variance for the income process is $\sigma_y^2 = 100$ and the interest rate is $r = 0.06$. The stopping criterion is given by the sequence $\varepsilon^{(k)}$ as defined in (2.6.9) and I arbitrarily choose $d = 0.004$. Therefore when $\varepsilon^{(k)}$ becomes smaller than $d = 0.004$ I stop and declare the solution obtained in iteration k as the solution to the problem described by (2.2.1) - (2.2.4). In Table 1 I provide the values for the accuracy measures discussed above, using the level of assets, as opposed to the level of consumption. One can see that the approximation to the analytical solution obtained by stopping when $\varepsilon^{(k)}$ is smaller than the arbitrarily chosen d is very good.

Table 1. Accuracy measures for d=.004

θ				
σ_y^2	0.01			
	E_R^a	E_R^m	E_E^a	E_E^m
100	0.001445515	0.002392885	0.000005019	0.000008735

The results presented in Table 1 are obtained after 159 iterations. Next, I will look at the behavior of the sequence $\varepsilon^{(k)}$ for the case presented above.

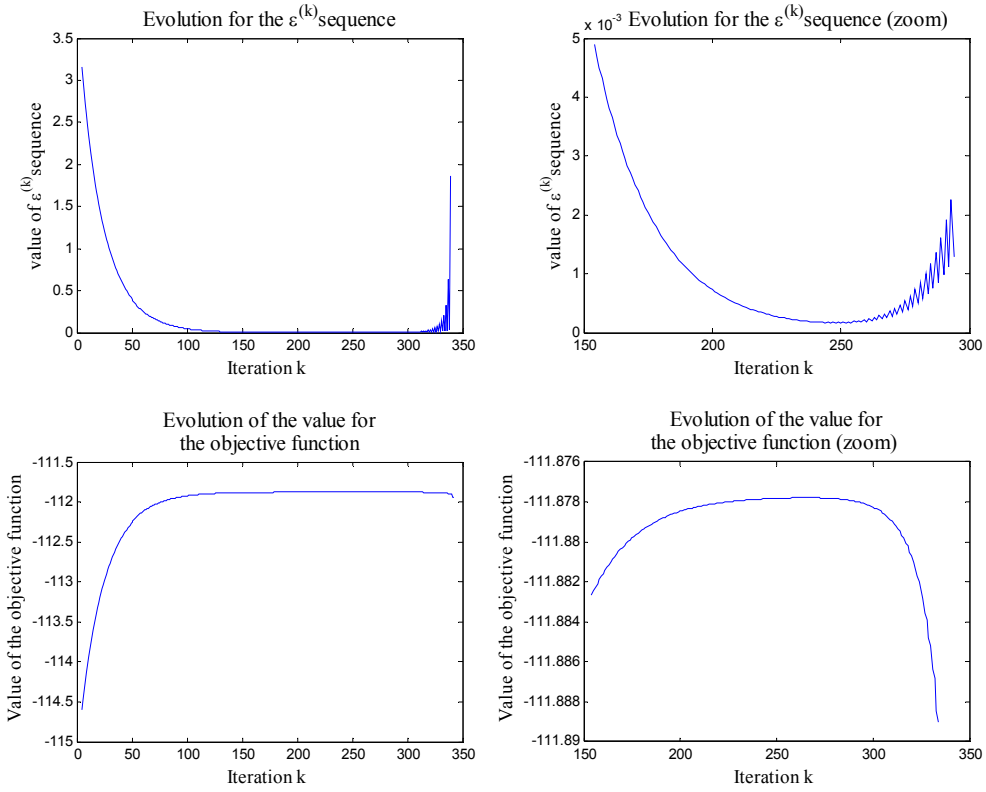

Figure 3. Evolution of the $\varepsilon^{(k)}$ sequence and the value of the objective for $\theta = .01$ and $\sigma_y^2 = 100$

One can see in Figure 3 that the value for sequence $\varepsilon^{(k)}$ continues to decrease until iteration 250 when it attains the minimum value. At the same time, the value of the objective continues to increase until iteration 266 when it attains its maximum. It is worth noting that the value of the objective is computed as in equation (2.6.12). Based on these observations one may elect to choose as stopping criterion the point where $\varepsilon^{(k)}$ attains its minimum or when the objective function attains its maximum as opposed to an arbitrary value d. Next, I look at how close is the approximation to the analytical solution when

using these criteria. In Table 2 one can see that there is not much difference between the last two criteria when compared to the analytical solution. The only difference is that the value of the expected utility is marginally higher in the second case.

Table 2. Accuracy measures for various stopping criteria

θ					
σ_y^2	0.01				
	E_R^a	E_R^m	E_E^a	E_E^m	Stopping criterion
100	0.001445515	0.002392885	0.000005019	0.000008735	Arbitrary $d = 0.004$
100	0.002137894	0.002691210	0.000007190	0.000013733	Minimum of $\varepsilon^{(k)}$
100	0.002137894	0.002691210	0.000007190	0.000013733	Maximum objective

A somewhat interesting result is that the ad-hoc stopping criterion $d = 0.004$ leads to a better approximation of the analytical solution. This is explained by the fact that the progressive hedging algorithm leads to the solution that would be obtained through the aggregation of the exact solutions for every scenario. Here the starting point is the certainty equivalent solution and the path to convergence, at some point, is very close to the analytical solution.

II.6.3. The Role of the Penalty Parameter

In the implementation of the progressive hedging algorithm, I chose the penalty parameter to be constant. Its role is to keep the contingent solution for each iteration close to the previous implementable policy. However, its value also has an impact on the speed of convergence. I will now consider the previous parameterization of the model and I am

going to change the value of the penalty parameter to see how it changes the speed of convergence. In Figure 4 one can see that as ρ increases so does the number of iterations needed to achieve convergence. While a higher value of the penalty parameter helps the convergence of contingent policies to the implementable policy, it also slows the global convergence process, requiring more iterations.

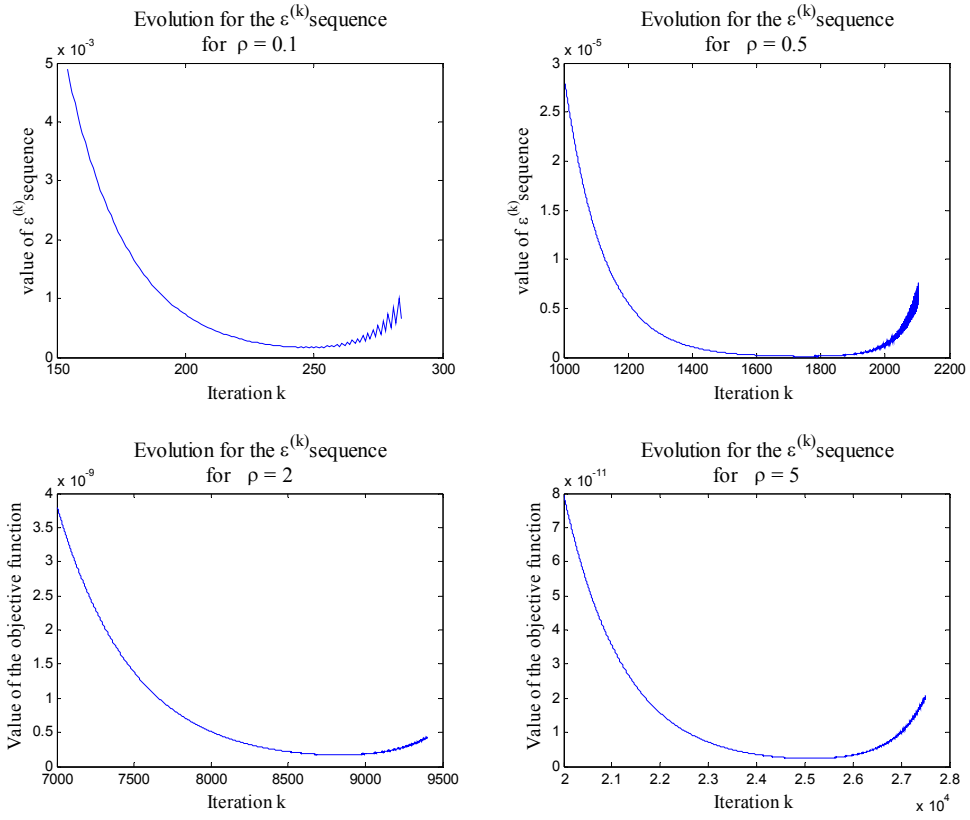

Figure 4. Convergence for different values of the penalty parameter.

For $\rho = 0.1$, 250 iterations are needed to achieve convergence, while for $\rho = 0.5$, 1780 iterations are needed. For higher values, such as $\rho = 5$, the number of iterations needed to achieve convergence increases to over 25000 iterations.

II.6.4. More simulations

In this section I investigate how close the scenario aggregation solution is to the analytical solution for various parameters. Table 3 shows the values for the four criteria enumerated above for different values of the coefficient of risk aversion and of the variance of the random variable entering the income process. All the simulations whose results are presented in Table 3 were done using a three point approximation of the distribution of the random variable entering the income process. The relative measures are computed using the level of assets.

Table 3. Accuracy measures for various parameters when interest rate r=0.04

	θ											
σ_y^2	0.01				0.05				0.1			
	E_R^a	E_R^m	E_E^a	E_E^m	E_R^a	E_R^m	E_E^a	E_E^m	E_R^a	E_R^m	E_E^a	E_E^m
1	.0000	.0000	.0000	.0000	.0001	.0001	.0000	.0000	.0002	.0003	.0000	.0000
4	.0000	.0001	.0000	.0000	.0004	.0005	.0000	.0000	.0009	.0011	.0000	.0000
25	.0005	.0007	.0000	.0000	.0029	.0037	.0000	.0000	.0058	.0074	.0000	.0000
100	.0023	.0029	.0000	.0000	.0116	.0147	.0000	.0000	.0230	.0290	.0000	.0000

For lower values of the coefficient of risk aversion the approximation is relatively good. As the coefficient of risk aversion increases in tandem with the variance of the income process, the accuracy suffers when looking at relative measures. The Euler equation measure still indicates a very good approximation.

Let us now look at how this approximation affects the value of the original objective, i.e. the expected discounted utility over the lifetime horizon. Table 4 shows the

ratio of the expected utilities for the whole horizon with the scenario aggregation as the as the denominator and the analytical solution as the numerator.

Table 4. The ratio of lifetime expected utilities $\frac{F_{as}}{F_{sc}}$

θ				
σ_y^2	0.01	0.03	0.05	0.1
1	1.00000	1.00000	1.00000	1.00003
4	1.00000	1.00001	1.00003	1.00058
25	1.00000	1.00002	1.00141	1.02027
100	1.00003	1.00051	1.02273	1.39364

The discounted utilities are computed as in the original formulation of the problem:

$$F_{sc} = \frac{1}{N}\sum_{i=1}^{N}\left\{-\sum_{t=0}^{T}\beta^t\left[\frac{1}{\theta}\exp\left(-\theta \underline{c}_t^i\right)\right]\right\} \tag{2.6.12}$$

and

$$F_{as} = \frac{1}{N}\sum_{i=1}^{N}\left\{-\sum_{t=0}^{T}\beta^t\left[\frac{1}{\theta}\exp\left(-\theta c_t^{i*}\right)\right]\right\} \tag{2.6.13}$$

where N is the number of simulations, F_{sc} is the discounted utility obtained with scenario aggregation and F_{as} is the discounted utility obtained with the analytical solution. In this formulation, both quantities are negative so their ratio is positive. Note however that the initial formulation of the problem using the objective function specified in (2.6.12) and (2.6.13) was a maximization. Therefore, higher ratio in Table 4 means that the solution obtained through scenario aggregation leads to higher discounted lifetime utility than the analytical solution. I simulate 2000 realizations of the income process and then I average the discounted utilities over this sample. The result shows that the solution

obtained through scenario aggregation leads to higher overall expected utility as the coefficient of risk aversion increases. This is explained by the fact that the level of consumption in the first few periods is higher in the case of scenario aggregation. In the context of a short horizon, this leads to higher levels of discounted utility.

II.7. Final Remarks

The results show that scenario aggregation can be used to provide a good approximation to the solution of a life-cycle model for certain values of the parameters. There are a few remarks to be made regarding the convergence. As pointed out earlier in this chapter the value of ρ has an impact on the speed of convergence. Higher values of ρ lead to faster convergence of the contingent policies towards an implementable policy but that also means that the overall convergence is slower and hence it impacts the accuracy if an ad-hoc stopping criterion is used. Therefore, one needs to choose carefully the values of the ad-hoc parameters. On the other hand, if the scenario problems have an exact solution then the final implementable policy can be obtained through a simple weighted average with the weights being the probabilities of realization for each scenario.

Chapter III. Impact of Bounded Rationality[57] on the Magnitude of Precautionary Saving

III.1. Introduction

It is fair to say that nowadays the assumption of rational expectations has become routine in most economic models. Recently, however, there has been an increasing number of papers, such as Gali et al. (2004), Allen and Carroll (2001), Krusell and Smith (1996), that have modeled consumers using assumptions that depart from the standard rational expectations paradigm. Although they are not explicitly identified as modeling bounded rationality, these assumptions clearly take a bite from the unbounded rationality, which is the standard endowment of the representative agent. The practice of imposing limits on the rationality of agents in economic models is part of the attempts made in the literature to circumvent some of the limitations associated with the rational expectations assumption. Aware of its shortcomings, even some of the most ardent supporters[58] of the rational expectations paradigm have been looking for possible alterations of the standard set of assumptions. As a result, a growing literature in macroeconomics is tweaking the unbounded rationality assumption resulting in alternative approaches that are usually presented under the umbrella of bounded rationality.

---

[57] The concept of bounded rationality in this chapter should be understood as a set of assumptions that departs from the usual rational expectation paradigm. Its meaning will become clear later in the chapter when the underlying assumptions are spelled out.

[58] Sargent (1993) for example, identifies several areas in which bounded rationality can potentially help, such as equilibrium selection in the case of multiple possible equilibria and behavior under "regime changes".

One may ask why is there a need to even consider bounded rationality. First, individual rationality tests led various researchers to "hypothesize that subjects make systematic errors by using ... rules of thumb which fail to accommodate the full logic of a decision" (J. Conlisk, 1996). Secondly, some models assuming rational expectations fail to explain observed facts, or their results may not match empirical evidence. Since most of the time models include other hypotheses besides the unbounded rationality assumption, the inability of such models to explain certain observed facts could not be blamed solely on rational expectations. Yet, it is worth investigating whether bounded rationality plays an important role in such cases. Finally, as Allen and Carroll (2001) point out, even when results of models assuming rational expectations match the data, it is still worth asking the question of how can an average individual find the solution to complex optimization problems that until recently economists could not solve. To summarize, the main idea behind this literature is to investigate what happens if one changes the assumption that agents being modeled have a deeper understanding of the economy than researchers do, as most rational expectations theories assume. Therefore, instead of using rational expectations, it is assumed that economic agents make decisions behaving in a rational manner but being constrained by the availability of data and their ability to process the available information.

While the vast literature on bounded rationality continues to grow, there is yet to be found an agreed upon approach to modeling rationally bounded economic agents. Among the myriad of methods being used, one can identify decision theory, simulation-based models, artificial intelligence based methodologies such as neural networks and genetic algorithms, evolutionary models drawing their roots from biology, behavioral

models, learning models and so on. Since there is no standard approach to modeling bounded rationality, most of the current research focuses on investigating the importance of imposing limits on rationality, as well as on choosing the methods to be used in a particular context. When modeling consumers, the method of choice so far seems to be the assumption that they follow some rules of thumb[59]. Instead of imposing some rules of thumb, my approach in modeling bounded rationality focuses on the decision making process. I borrow the idea of scenario aggregation from the multistage optimization literature and I adapt it to fit, what I believe to be, a reasonable description of the decision making process for a representative consumer. Besides the decision making process per se, I also add a few other elements of bounded rationality that have to do with the ability to gather and process information.

In the previous chapter, the method of scenario aggregation was introduced as an alternative method for solving non-linear rational expectation models. Even though it performs well in certain circumstances, the real advantage of the scenario aggregation lays in a different area. Its structure presents itself as a natural way to describe the process through which a rationally bounded agent, faced with uncertainty, makes his decision. In this chapter, I consider several versions of a life-cycle consumption model with the purpose of investigating how the magnitude of precautionary saving changes with the underlying assumptions on the (bounded) rationality of the consumer.

[59] Some of the examples are Gali et al. (2004), Allen and Carroll (2001), Lettau and Uhlig (1999) and Ingram (1990).

III.2. Empirical Results on Precautionary Saving

There seems to be little agreement in the empirical literature on precautionary saving, especially when it comes to its relationship to uncertainty. Skinner (1988) found that saving was lower than average for certain groups[60] of households that are perceived to have higher than average income uncertainty. In the same camp, Guiso, Jappelli and Terlizzese (1992), using data from the 1989 Italian Survey of Household Income and Wealth, found little correlation between the level of future income uncertainty and the level of consumption[61]. In addition, Dynan (1993), using data from the Consumer Expenditure Survey, estimated the coefficient of relative prudence and found it to be "too small to be consistent with widely accepted beliefs about risk aversion".

On the other hand, Dardanoni (1991) basing his analysis on the 1984 cross-section of the UK FES (Family Expenditure Survey) suggested that the majority of saving in the sample arises for precautionary motives. He found that average consumption across occupation and industry groups was negatively related to the within group variance of income. Carroll (1994) found that income uncertainty was statistically important in regressions of current consumption on current income, future income and uncertainty. Using UK FES data, Merrigan and Normandin (1996) estimated a model where expected consumption growth is a function of expected squared consumption growth and demographic variables and their results, based on the period 1968-1986,

[60] Specifically, the groups identified were farmers and self-employed.

[61] In fact the study on Italian consumers did find that consumption was marginally lower while wealth was marginally higher for those who were facing higher income uncertainty in the near future.

indicate that precautionary saving is an important part of household behavior. Miles (1997), using several years of cross-sections of the UK micro data and regressing consumption on several proxies for permanent income and uncertainty, found that, for each cross-section, the latter variable played a statistically significant role in determining consumption. In a study trying to measure the impact of income uncertainty on household wealth, Carroll and Samwick (1997), using the Panel Study of Income Dynamics, found that about a third of the wealth is attributable to greater uncertainty. Later on, Banks et al. (2001), exploiting not only the cross-sectional, but also the time-series dimension of their data set, find that section specific income uncertainty as opposed to aggregate income uncertainty plays a role in precautionary saving. Finally, Guariglia (2001) finds that various measures of income uncertainty have a statistically significant effect on savings decisions.

In this chapter, I am going to show that, by introducing bounded rationality in a standard life cycle model, one can increase the richness of the possible results. Even if the setup of the model would imply the existence of precautionary savings, under certain parameter values and rules followed by consumers, the precautionary saving is apparently almost inexistent. As opposed to most of the literature[62] studying precautionary savings, I introduce uncertainty in the interest rate, beside income uncertainty. In this context, the size of precautionary saving no longer depends exclusively on income uncertainty.

[62] A notable exception is Binder et al. (2000).

III.3. The Model

I start this section by presenting the formulation of a standard finite horizon life-cycle consumption model. Then I will introduce a form of bounded rationality[63] and investigate the path for consumption and savings.

Consider the finite-horizon life-cycle model under negative exponential utility. Suppose an individual agent is faced with the following intertemporal optimization problem:

$$\max_{\{c_t\}_{t=0}^{T}} E\left[\sum_{t=0}^{T} -\beta^t \frac{1}{\theta}\exp\left(-\theta c_t\right) | I_0\right] \tag{3.3.1}$$

subject to

$$A_t = \left(1+r_t\right)A_{t-1} + y_t - c_t, \qquad t = 0,1,...,T-1, \tag{3.3.2}$$

$$A_t \geq -b, \quad \text{with } A_{-1}, A_T \text{ given} \tag{3.3.3}$$

where θ is the coefficient of risk aversion, A_t represents the level of assets at the beginning of period t, y_t the labor income at time t, and c_t represents consumption in period t. The initial and terminal conditions, A_{-1} and A_T are given. The information set I_0 contains the level of consumption, assets, labor income and interest rate for period zero and all previous periods. The labor income is assumed to follow an arithmetic random walk:

[63] As it was already mentioned above, the approach in defining bounded rationality in this chapter has some similarities to the approach followed by Lettau and Uhlig (1999) in the sense that several rules are used to account for the inability of the boundedly rational agent to optimize over long horizons.

$$y_t = y_{t-1} + \xi_t, \quad t = 1,...,T, \text{ with } y_o \text{ given} \tag{3.3.4}$$

and ξ_t being drawn from a normal distribution, $\xi_t \sim N\left(0, \sigma_y^2\right)$. When the interest rate is deterministic, this problem has an analytical solution[64]. However, if the interest rate is stochastic, the solution of this finite horizon life cycle model becomes more complicated and it can not be computed analytically. For now, I will not make any particular assumption about the process generating the interest rate. Therefore, to summarize the model, a representative consumer derives utility in period t from consuming c_t, discounts future utility at a rate β and wants, in period zero, to maximize his present discounted value of future utilities for a horizon of $T+1$ periods. At the beginning of each period t the consumer receives a stochastic labor income y_t, finds out the return r_t on his assets A_{t-1}, from the beginning of period $t-1$ to the beginning of period t, and, by choosing c_t, determines the level of assets A_t according to equation (3.3.2).

Now, I introduce a rationally bounded agent in the following way. First, I assume that the agent does not have either the resources or the sophistication to be able to optimize over a long horizon. For example, if the agent enters the labor force at time zero and faces the problem described by (3.3.1) - (3.3.4) over a time span extending until his retirement, let it be period T, the assumption is that the agent does not have the ability to optimally choose, at time zero, a consumption plan over that span. Instead, he focuses on choosing a consumption plan over a shorter horizon, let it be T_h+1 periods.

Secondly, because of his limited ability to process large amounts of information he repeats this process every period in order to take advantage of any new available

[64] See the appendix for a detailed description of the analytical solution.

information. This idea of a shorter and shifting optimization horizon is similar to the approach taken by Prucha and Nadiri[65] (1984, 1986, and 1991). Now, the question is how an individual who lacks sophistication, can optimally[66] choose a consumption plan even for a short time span. In order to model the decision process I make use of the scenario aggregation method. Under this assumption, the agent evaluates several possible paths based on the realization of the forcing variables specified in the model. By assigning probabilities to each of the possible paths, the agent is in the position to aggregate the scenarios (paths), i.e., to compute the expected value for his decision.

In order to be able to use the scenario aggregation method, the forcing variables need to have a discrete distribution but in the model presented above, they are described as being drawn from a normal distribution. This leads to the third element that can be brought under the umbrella of bounded rationality. Since the agent has limited computational ability, the distribution of the forcing variable is approximated by a discrete distribution with the same mean and variance as the original distribution. This approximation does not necessarily have to be viewed as a bounded rationality element since similar approaches have been employed repeatedly in numerical solutions using state space discretization[67].

Given the assumptions made about the abilities of the rationally bounded representative agent, I will now go through the details of solving the problem described

---

[65] In their work, a finite and shifting optimization horizon is used to approximate an infinite horizon model.

[66] Optimality here means the best possible solution given the level of ability.

[67] Tauchen, among others, used this kind of approximation on various occasions, such as Tauchen (1990), Tauchen and Hussey (1991).

by equations (3.3.1) - (3.3.4). Hence, at every point in time, t, the agent solves the problem:

$$\max_{\{c_{t+\tau}\}_{\tau=0}^{T_h}} E\left[\sum_{\tau=0}^{T_h} -\beta^{\tau}\frac{1}{\theta}\exp\left(-\theta c_{t+\tau}\right) | I_t\right] \text{ for } t = 0,1,...,T-T_h \tag{3.3.5}$$

or

$$\max_{\{c_{t+\tau}\}_{\tau=0}^{T-t}} E\left[\sum_{\tau=0}^{T-t} -\beta^{\tau}\frac{1}{\theta}\exp\left(-\theta c_{t+\tau}\right) | I_t\right] \text{ for } t = T-T_h+1,...,T-1 \tag{3.3.6}$$

subject to

$$\begin{aligned} A_{t+\tau} &= \left(1+r_{t+\tau}\right)A_{t+\tau-1} + y_{t+\tau} - c_{t+\tau}, \\ &\qquad t = 0,1,...,T-1, \quad \tau = 0,...,\min\left(T_h, T-t\right) \end{aligned} \tag{3.3.7}$$

$$\text{with } A_{-1},\ A_{t-1}, A_{t+T_h} \text{ and } A_T \text{ given} \tag{3.3.8}$$

where $A_{t+\tau}$ represents the level of assets at the beginning of period $t+\tau$, $y_{t+\tau}$ the labor income at time $t+\tau$, and $c_{t+\tau}$ represents consumption in period $t+\tau$. The initial and terminal conditions, $A_{-1}, A_{t-1}, A_{t+T_h}$ and A_T are given. The information set I_t contains the level of consumption, assets, labor income and interest rate for period t and all previous periods. The labor income is assumed to follow an arithmetic random walk:

$$\begin{aligned} y_{t+\tau} &= y_{t+\tau-1} + \xi^b_{t+\tau}, \quad t = 1,...,T, \quad \tau = 0,...,\min\left(T_h, T-t\right) \\ &\text{with } y_0 \text{ given} \end{aligned} \tag{3.3.9}$$

$\xi^b_{t+\tau}$ being drawn from a discrete distribution, $D\left(0,\sigma_y^2\right)$ with a small number of realizations.

In making the above assumptions, the belief is that they would better describe the way individuals make decisions in real life. It is often the case that plans are made for shorter horizons, but not entirely forgetting about the big picture.

Recalling the results of Skinner (1988) who found that saving was lower than average for farmers and self employed, groups that are otherwise perceived to have higher than average income uncertainty, one can assume that planning for those groups does not follow the recipe given by the standard life cycle model. Given the high level of uncertainty, I believe it would be more appropriate to model these consumers as if they plan their consumption path only for a short period of time and then reevaluate. This would be consistent with the fact that farmers change their crop on a cycle of several years and may be influenced by the fluctuations in the commodities markets and other government regulations. Similarly, some among the self employed are likely to have short term contracts and are more prone to reevaluate their strategy on a high frequency basis. Therefore, the model above seems like a good description on how the decision making process works. The only detail that remains to be decided is how the consumer chooses the short horizon terminal condition, that is, the level of assets, or the wealth. For this purpose, I propose three different rules and I investigate their effect on the saving behavior.

So far, no assumption has been made about the process governing the realizations of the interest rate. From now on, I assume that the interest rate is also described by an arithmetic random walk:

$$r_t = r_{t-1} + \upsilon_t, \quad t = 1,...,T, \text{ with } r_o \text{ given} \tag{3.3.10}$$

Since in this formulation the problem does not have an analytical solution, the classical approach would be to employ numerical methods in order to describe the path of consumption, even for a very short horizon. In order to find the solution corresponding to the model incorporating the bounded rationality assumption I will use the scenario

aggregation[68] methodology. Then I will compare this solution with the numerical solution[69] that would result from the rational expectation version of the model when optimizing over the whole T period horizon.

III.3.1. Rule 1

Under rule 1, the consumer considers several possible scenarios for a short horizon and assumes that for later periods certainty equivalence holds. In this context, he makes a decision for the current period and moves on to the next period when he observes the realization of the forcing variables. Then he repeats the process by making a decision based on considering all the relevant scenarios for the near future and assuming certainty equivalence for the distant future. Hence, the decision making process takes place every period. More precisely, when optimizing in period t, the consumer considers all the scenarios in the tree event determined by the realizations of the forcing variable for the first T_h periods. From period $t+T_h$ he considers that certainty equivalence holds for the remaining $T-t-T_h$ periods. This translates specifically to considering that income and interest rate are frozen for each existing scenario for the remaining $T-t-T_h$ periods. To be more specific, for time $t=0$, the consumer considers all the scenarios available in the event tree for the first T_h periods and assumes certainty equivalence for

[68] Since an analytical solution can be obtained when income follows an arithmetic random walk and interest rate is deterministic, it is not necessary to discretize both forcing variables, but only the interest rate. This approach reduces considerably the computational burden. A short description on the methodology used along with the solution for one scenario with deterministic, interest rate is presented in the appendix. More details on the scenario aggregation methodology can be found in the second chapter.

[69] The numerical solution is obtained using projection methods and is due to Binder et al. (2000).

the remaining $T - T_h$ periods. When it advances to period $t = 1$, he optimizes again considering all the scenarios available in the tree event for periods $1, 2, ..., T_h + 1$ and assumes certainty equivalence for the remaining $T - T_h - 1$ periods.

In fact, this rule can be considered as an extension to the scenario aggregation method in order to avoid the dimensionality curse. One may recall that due to its structure, the number of scenarios in the scenario aggregation method increases exponentially with the number of periods. In effect, this rule is limiting the number of scenarios considered and it is consistent with a rationally bounded decision maker who can only consider a limited and, most likely, low number of possible scenarios.

Following are some graphical representations of the simulations for rule 1. Each graph contains the values for the coefficient of risk aversion, θ. The graphs also contain the numerical solution and, for comparison purposes, the evolution of assets if the solution were computed in the case of certainty equivalence. I first consider a group of 12 cases varying certain parameters of the model. For all simulations in this group, the total number of periods considered is $T = 40$ and the optimizing horizon is $T_h = 6$. The starting level of income is $y_0 = 200$, the initial level of assets is $A_{-1} = 500$ while the terminal value is $A_T = 1000$. The discount factor is $\beta = 0.96$, the starting value for the interest rate, $r_0 = 0.06$ while the standard deviation for the interest process is given by $\sigma_r = 0.0025$. I use a discrete distribution with three possible realizations to approximate the original distribution of the forcing variable and that implies that in each period t, for $t \leq T - T_h = 34$, the optimization process goes over $3^{T_h} = 729$ scenarios. For periods $34 = T - T_h < t \leq T - 1 = 39$ the number of scenarios considered decreases to 3^{T-t}. The

parameters that are changing in the simulations are the variance for the income process and the coefficient of risk aversion. I consider all cases obtained combining three values for the standard deviation of income, $\sigma_y \in \{1,\ 5,\ 10\}$ and four values for the coefficient of risk aversion, $\theta \in \{0.005,\ 0.01,\ 0.05,\ 0.1\}$. The results presented in this section as well as for the rest of the chapter are based on 1000 simulations. This means that for both the income generating process and the interest rate generating process, I consider 1000 realizations for each period. The decision to use only 1000 realizations was based on the observation that the sample drawn provided a good representation of the arithmetic random walk process assumed in the model. Specifically, both the mean and the standard deviation of the sample were close to their theoretical values.

Some general results have emerged from all these simulations. First, the path for the level of assets for the solution obtained in the bounded rationality case always lies below the path for the level of assets for the numeric al solution obtained in the rational expectation case. Consequently, the consumption path in the bounded rationality case starts with values of consumption higher than in the rational expectations case. Eventually the paths cross and the consumption level in the rational expectations case ends up being higher toward the end of the horizon.

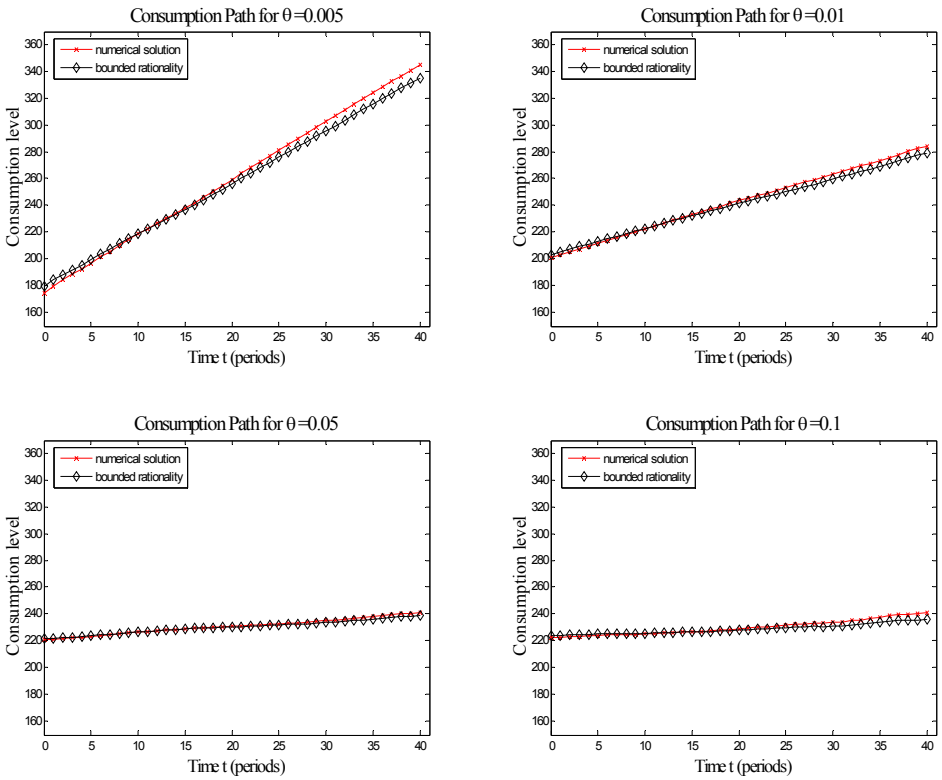

Figure 5. Consumption paths for $\sigma_y = 1$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

One can see in Figure 5 that consumption is increasing over time for both solutions, with the steepest path corresponding to the lowest value of the coefficient of risk aversion.

When looking at the asset path for the same value of the standard deviation of the income process, one notices in Figure 6 that the level of saving in the certainty equivalence case is mostly higher than the level of saving obtained in the bounded rationality case as well as under the rational expectations assumption.

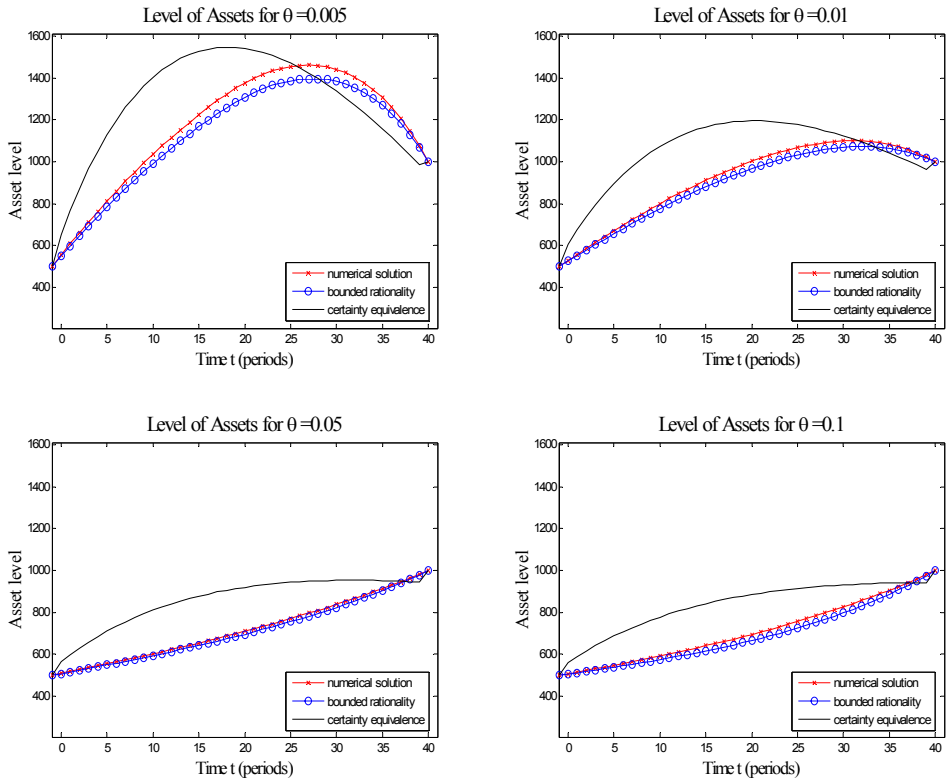

Figure 6. Asset paths for $\sigma_y = 1$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

While for lower levels of the coefficient of risk aversion $\theta \in \{0.005,\ 0.01\}$, the asset path obtained assuming certainty equivalence crosses under the other two paths in the later part of the horizon, the same is not true for higher values of the coefficient of risk aversion, $\theta \in \{0.05,\ 0.1\}$.

It is not only the relative position of the three paths that changes in the context of an increasing coefficient of risk aversion, but also the absolute size of the level of savings. Moreover, the shape of the paths for both the rational expectation and bounded rationality case changes from concave to convex.

I present now a new set of simulations with the standard deviation of income being increased to $\sigma_y = 5$. One can see in Figure 7 that the consumption paths for

$\theta \in \{0.005,\ 0.01\}$ are not much different from those presented in Figure 5 while for higher values of the risk aversion coefficient, $\theta \in \{0.05,\ 0.1\}$, the consumption paths are steeper than in the previous case.

Looking now at the level of savings, one notices in Figure 8 a similar change to that observed in the case of consumption. While not much has changed for the lower values of the coefficient for risk aversion, the asset paths for higher values of the risk aversion coefficient, $\theta \in \{0.05,\ 0.1\}$, have changed, effectively becoming concave, as opposed to convex in the previous case. Besides the concavity change, one can observe that for $\theta = 0.1$ the level of assets resulting from the numerical approximation of the rational expectations model is higher than in the case of certainty equivalence for the bigger part of the lifetime horizon.

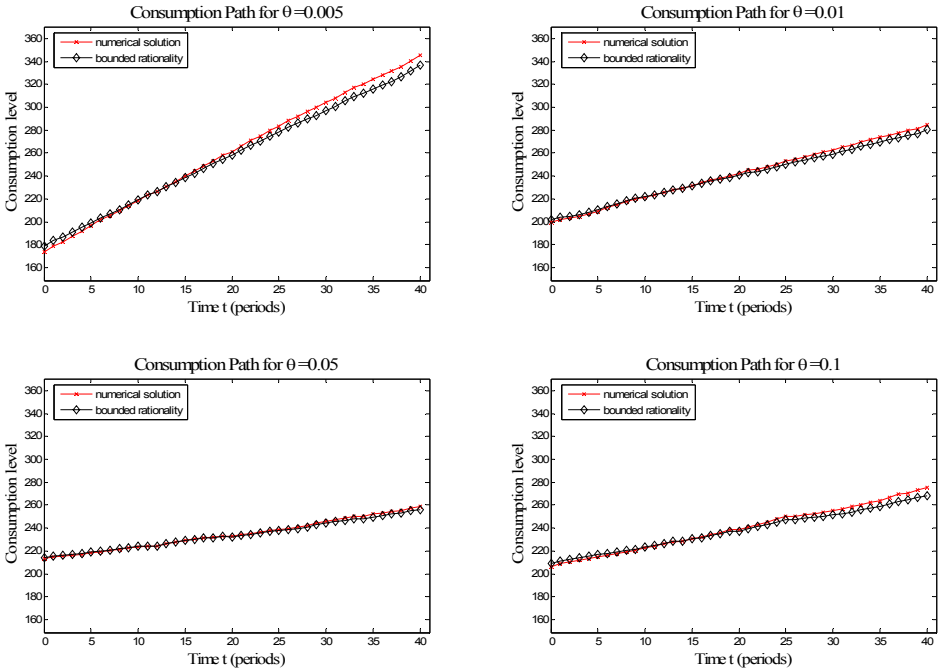

Figure 7. Consumption paths for $\sigma_y = 5$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

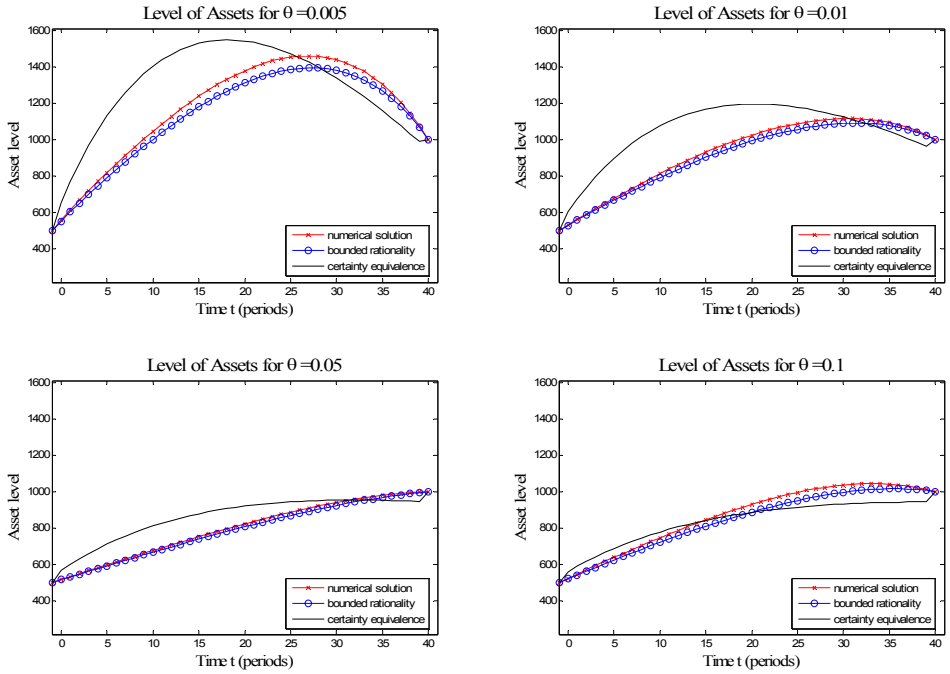

Figure 8. Asset paths for $\sigma_y = 5$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

By raising the variance of the income again, one can see in Figure 9 that the path for consumption becomes a lot steeper for $\theta \in \{0.05,\ 0.1\}$. On the other hand, there seems to be little change in the consumption pattern for $\theta = 0.005$.

On the savings front, the level of precautionary saving increases tremendously for the highest coefficient of risk aversion, $\theta = 0.1$, and quite substantially for $\theta = 0.05$. Consequently, in these two cases, the level of savings for the rational expectation model, as well as the bounded rationality version, becomes noticeably higher than what certainty equivalence produces. Yet, the level of savings continues to be higher for the much lower coefficient of risk aversion, $\theta = 0.005$, when compared with the savings pattern for $\theta = 0.01$ and $\theta = 0.05$.

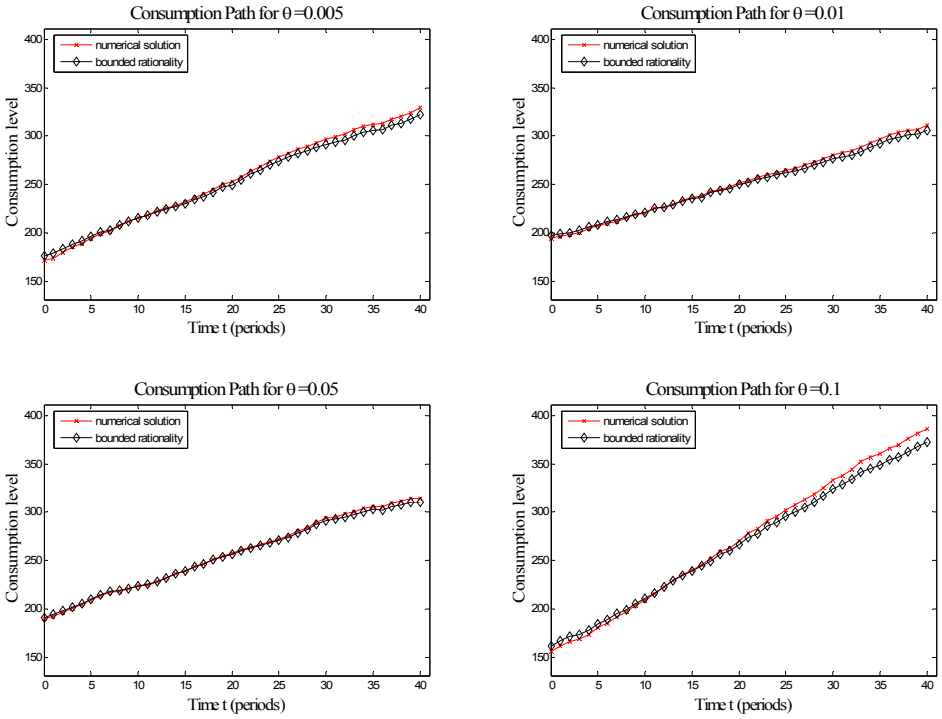

Figure 9. Consumption paths for $\sigma_y = 10$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

Another interesting observation is that if one compares the level of savings from the panel corresponding to $\theta = 0.05$ and $\sigma_y = 10$ in Figure 10, to the level of savings from the panel corresponding to $\theta = 0.005$ and $\sigma_y = 1$ in Figure 6, the two are almost the same, if not the later higher. This is to say that for values of coefficient of risk aversion and of standard deviation for income ten times as high as the ones in Figure 6, the level of precautionary saving is almost unchanged.

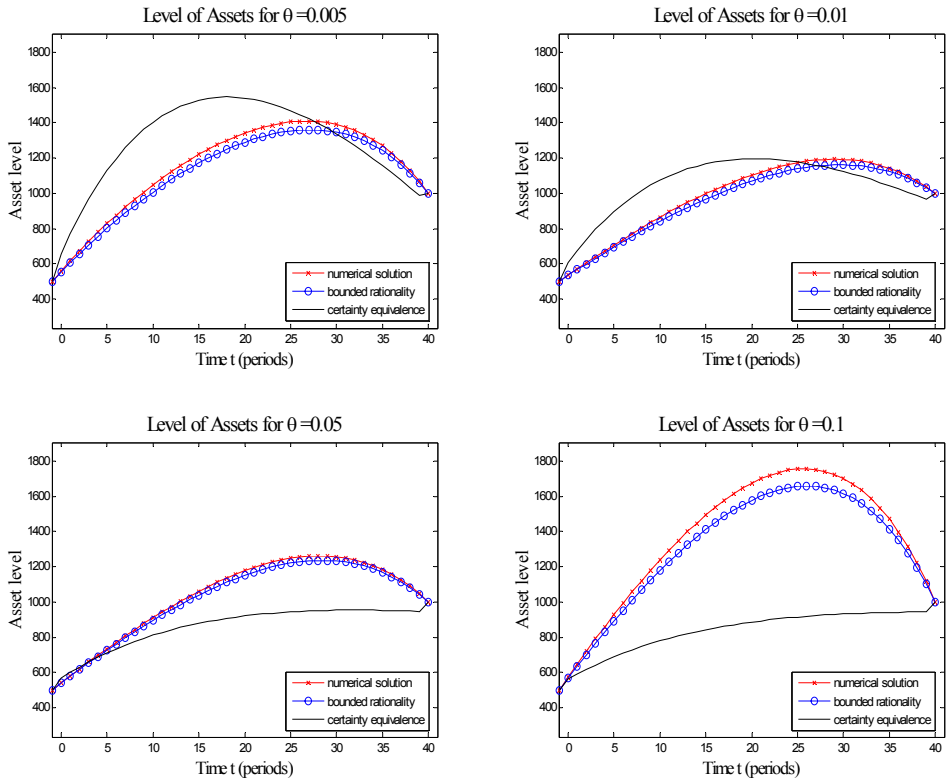

Figure 10. Asset paths for $\sigma_y = 10$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

As a general observation, it seems that the level of precautionary saving derived from the rational expectation model is consistently higher, even if not by high margins, than the level of savings obtained in the case of bounded rationality. For consumption, the paths can be steeper or flatter but the general allure remains the same. The rationally bounded consumer tends to start with a higher consumption while after a few periods the unboundedly rational consumer tends to take over and continue to consume more until the end of the horizon.

III.3.2. Rule 2

Under rule 2, the consumer considers all the relevant scenarios for the immediate short horizon and then, for the later periods, he only takes in account what I call the extreme cases. Rule 2 is similar to rule 1 in the way the decision maker emphasizes the importance of scenarios only for the short term horizon. The difference is that under rule 2, rather than assuming certainty equivalence for the later periods, the consumer considers the extreme case scenarios as a way of hedging against uncertainty in the distant future. More precisely, when optimizing in period t, the consumer considers all the scenarios in the event tree determined by the realizations of the forcing variable for the first T_h periods but then he becomes selective and only considers the extreme cases[70] for the remaining $T-t-T_h$ periods. To be more specific, for time $t=0$, the consumer considers all the scenarios available in the event tree for the first T_h periods and only the extreme cases for the remaining $T-T_h$ periods. When it advances to period $t=1$, he optimizes again considering all the scenarios available in the tree event for periods $1,2,...,T_h+1$ and only the extreme cases for the remaining $T-T_h-1$ periods.

In fact, this rule can also be considered as an extension to the scenario aggregation method in an attempt order to avoid the dimensionality curse. One may recall that due to its structure, the number of scenarios in the scenario aggregation method increases exponentially with the number of periods. This rule is in fact limiting the number of scenarios considered by trying to keep intact the possible variation in the forcing variable. As opposed to rule 1 where from time $t+T_h$ the assumption is that the

[70] The notion of extreme cases covers scenarios for which the realization of the forcing variable remains the same. For more details see section 0 in the appendix.

forcing variable keeps its unconditional mean value, that is, zero, until the end of the horizon, this rule expands the number of scenarios by adding all the extreme case scenarios stemming from the nodes existent at time $t + T_h$. This expansion can also be seen as the equivalent of placing more weight on the tails of the original distribution of the forcing variable. This rule is consistent with a rationally bounded decision maker who can only consider a limited and, most likely, low number of possible scenarios but wants to account for the variance in the forcing variable in the later periods of the optimization horizon.

Following are some graphical representations of the simulations for rule 2. The graphs depicting the consumption paths contain the bounded rationality solution as well as the numerical solution. For comparison purposes, the graph panels containing the evolution of assets display the savings pattern resulting from the solution obtained in the case of certainty equivalence on top of the solutions for the rational expectations and the bounded rationality models.

As in the case of rule 1, one can see in Figure 11 that consumption is increasing over time for both solutions, with the steepest path corresponding to the lowest value of the coefficient of risk aversion.

As opposed to the previous rule, the rationally bounded consumer does not always start with a higher level of consumption. In fact, in this panel, for $\theta = 0.05$ and $\theta = 0.1$, the solution of the rational expectations model has higher starting values for consumption.

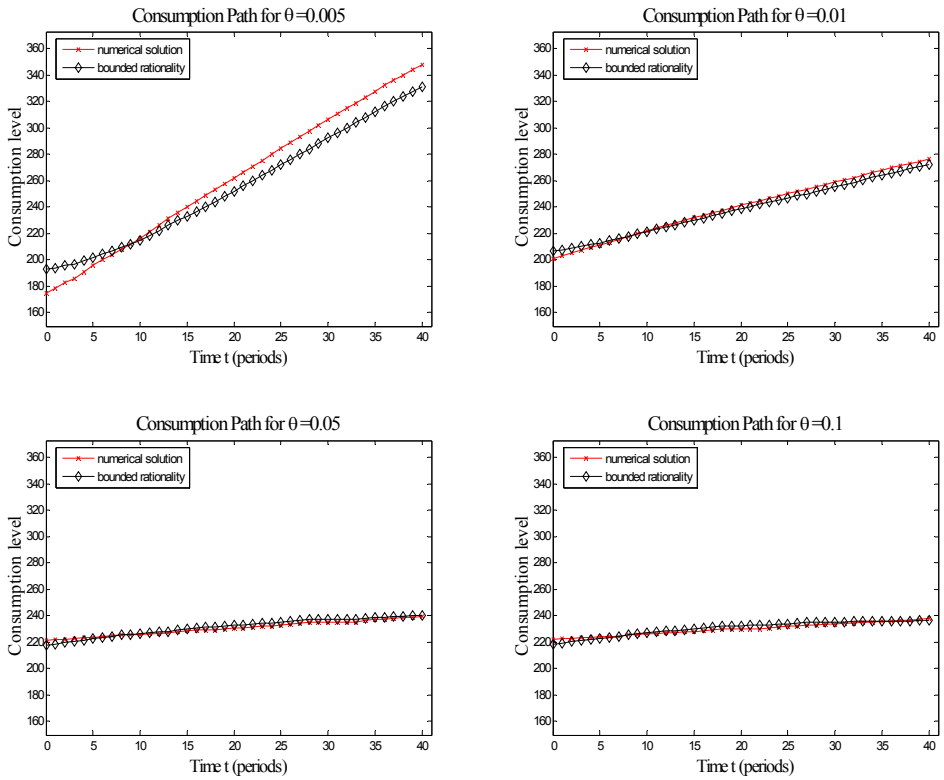

Figure 11. Consumption paths for $\sigma_y = 1$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

Looking at the asset paths for the same value of the standard deviation of the income process, one can notice in Figure 12 that the level of saving in the certainty equivalence case is mostly higher than the level of saving obtained in the bounded rationality case as well as under the rational expectations assumption. While for lower levels of the coefficient of risk aversion $\theta \in \{0.005,\ 0.01\}$, the asset path obtained assuming certainty equivalence crosses under the other two paths in the later part of the horizon, the same is not true for higher values of the coefficient of risk aversion. For $\theta \in \{0.05,\ 0.1\}$ there is only one period, the one next to last, when the level of savings under certainty equivalence is lower than in the other two cases.

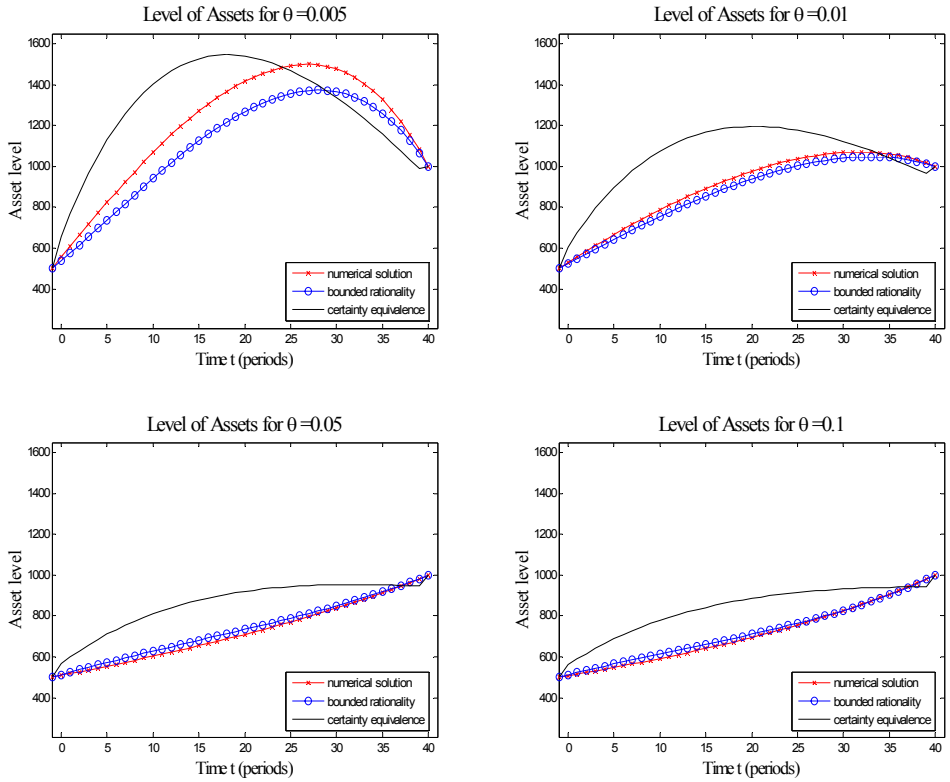

Figure 12. Asset paths for $\sigma_y = 1$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

As it was the case with rule 1, an increase in the coefficient of risk aversion results in a decrease of the absolute size of the level of savings. Moreover, the shape of the paths for both the rational expectation and bounded rationality cases changes from concave to convex. As opposed to rule 1, for $\theta \in \{0.05,\ 0.1\}$ the level of savings under bounded rationality is higher than under rational expectations.

The next set of simulations has the standard deviation of income increased to $\sigma_y = 5$. The consumption paths for $\theta \in \{0.005,\ 0.01\}$ in Figure 13 are not much different from those presented in Figure 11 while for higher values of the risk aversion coefficient, $\theta \in \{0.05,\ 0.1\}$, the consumption paths are steeper than in the previous case.

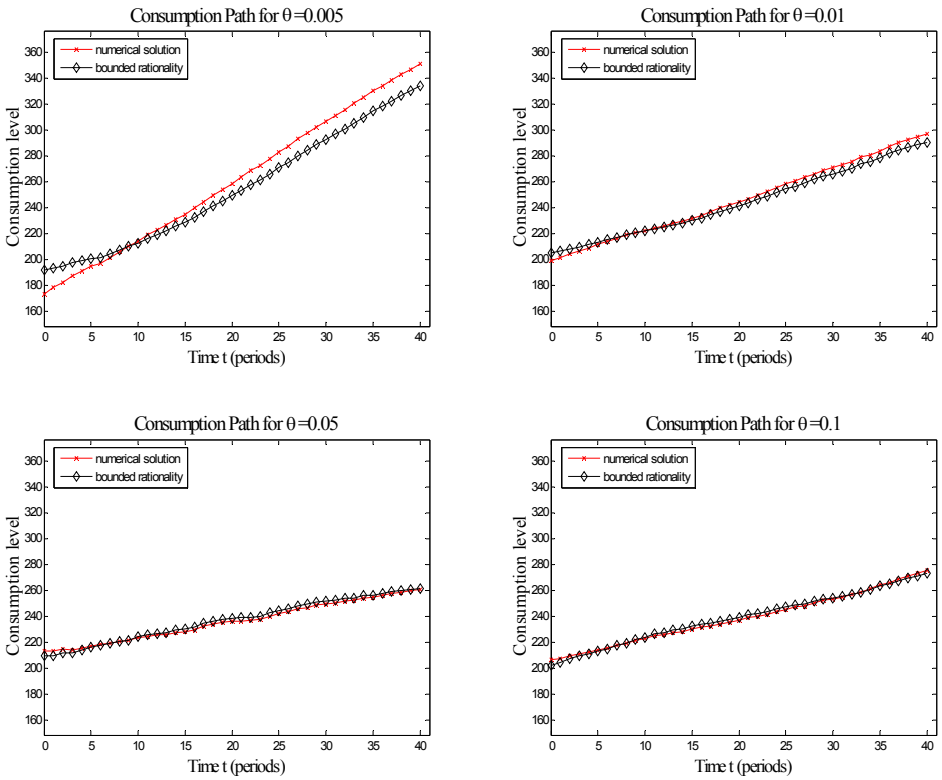

Figure 13. Consumption paths for $\sigma_y = 5$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

For the level of savings, the change is similar to that observed in the case of consumption. In Figure 14 one can see that, while not much has changed for the lower values of the coefficient for risk aversion, the asset paths for higher values of the risk aversion coefficient, $\theta \in \{0.05,\ 0.1\}$, have changed, effectively becoming concave, as opposed to convex in the previous case. Besides the concavity change, one can observe that for $\theta = 0.1$ the level of assets resulting from the numerical approximation of the rational expectations model is higher than in the case of certainty equivalence for the bigger part of the lifetime horizon.

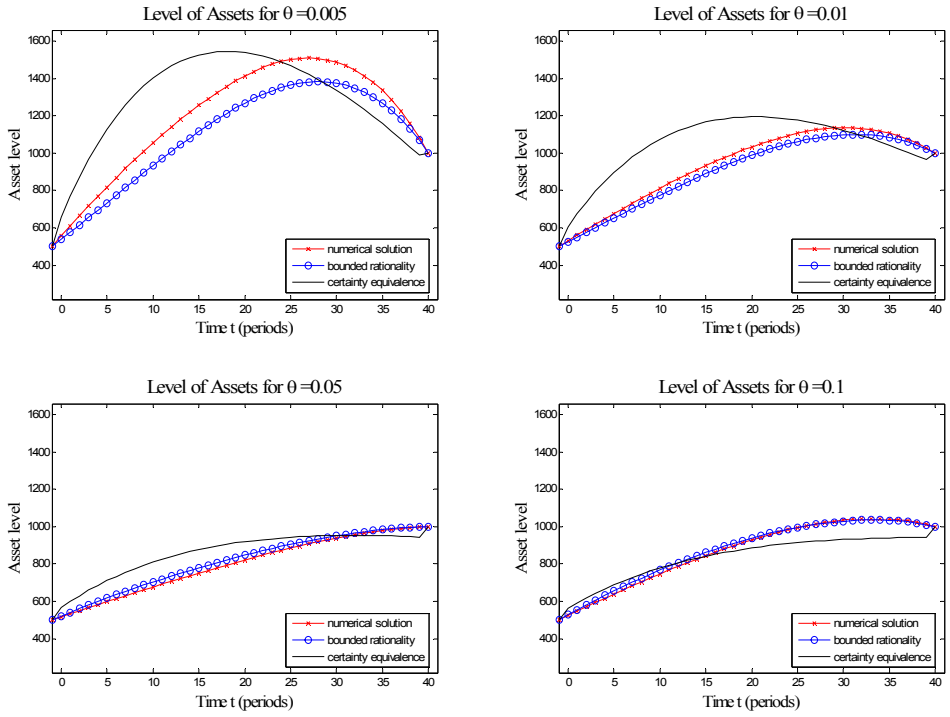

Figure 14. Asset paths for $\sigma_y = 5$, $r_0 = 0.06$ **and** $\sigma_r = 0.0025$.

For a yet higher variance of income, one can notice in Figure 15 that the path for consumption becomes a lot steeper for $\theta \in \{0.05,\ 0.1\}$. On the other hand, there seems to be little change in the consumption pattern for $\theta = 0.005$. On the savings front, the level of precautionary saving increases tremendously for the highest value of the coefficient of risk aversion considered here, $\theta = 0.1$, and quite substantially for $\theta = 0.05$. As it can be easily seen in Figure 16, in these two cases, the level of savings for the rational expectation model, as well as the bounded rationality version, becomes noticeably higher than what certainty equivalence produces.

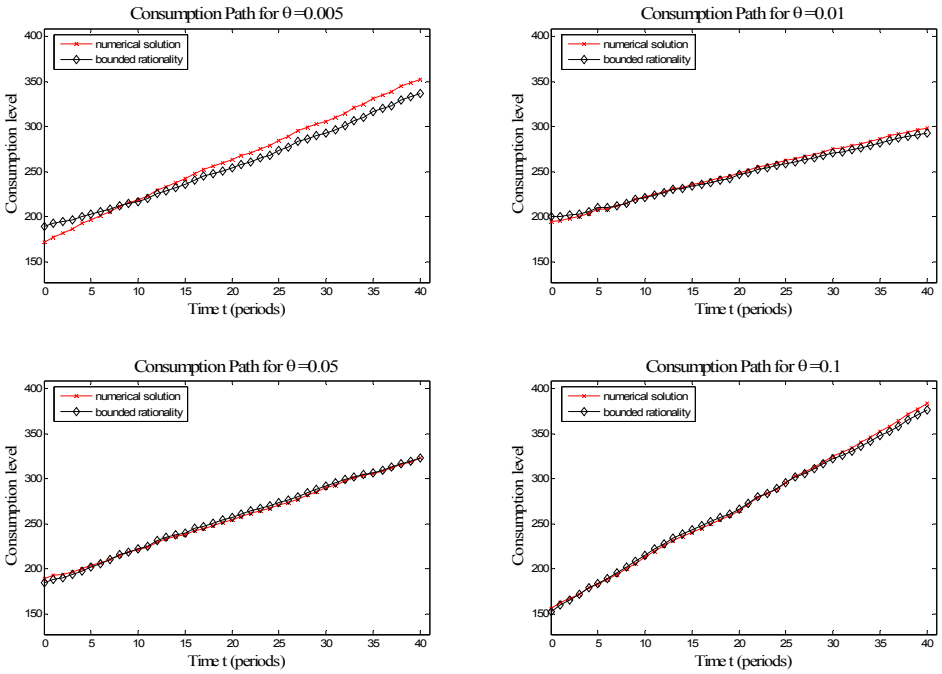

Figure 15. Consumption paths for $\sigma_y = 10$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

Yet, the level of savings continues to be higher for the much lower coefficient of risk aversion, $\theta = 0.005$, when compared with the savings pattern for $\theta = 0.01$ and $\theta = 0.05$.

As in the case of rule 1, comparing the level of savings from the panel corresponding to $\theta = 0.05$ and $\sigma_y = 10$ in Figure 16, to the level of savings from the panel corresponding to $\theta = 0.005$ and $\sigma_y = 1$ in Figure 12, leads to the observation that the two are almost the same. This is to say that for values of the coefficient of risk aversion and of standard deviation for income ten times as high as the ones in Figure 12, the level of precautionary saving is almost unchanged.

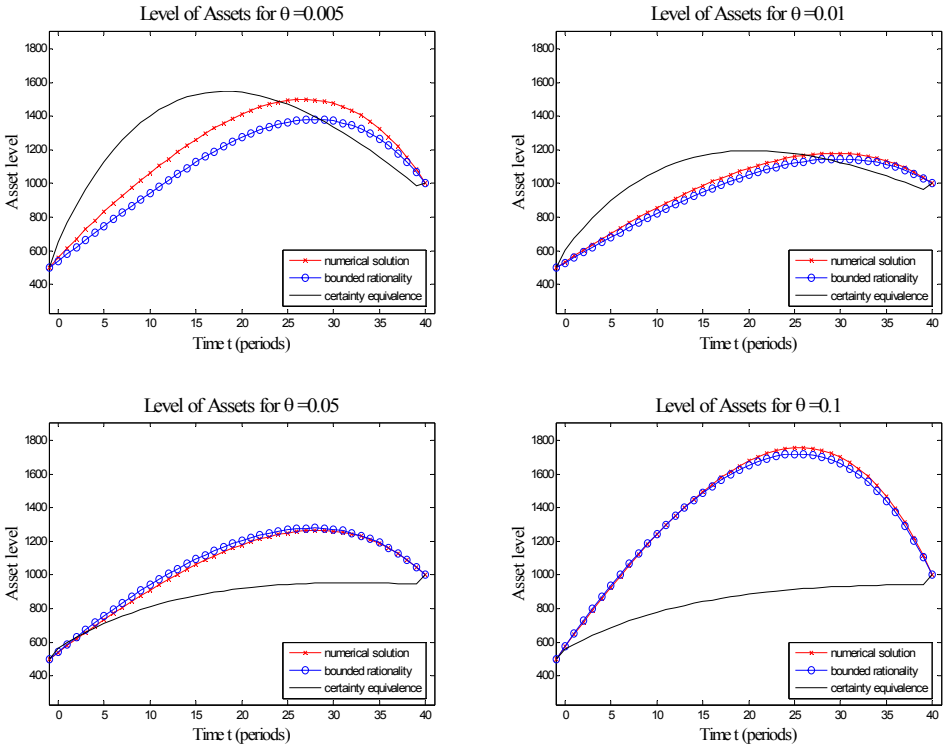

Figure 16. Asset paths for $\sigma_y = 10$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

As in the case of rule 1 the level of savings under bounded rationality is fairly close to the level of precautionary saving derived from the rational expectation model. However, in contrast to rule 1, the relative size depends on the parameters of the model and hence the level of precautionary saving derived from the rational expectation model is no longer consistently higher when compared to the level of savings obtained in the case of bounded rationality. Consequently, the rationally bounded consumer no longer starts consistently with a higher consumption level.

III.3.3. Rule 3

In this section, I will consider a simpler rule than the previous two, meaning that the level of wealth A_{t+T_h} is chosen such that, given the number of periods left until time T, a constant growth rate would insure that the final level of wealth is A_T.

Following are some graphical representations of the simulations for rule 3. All the graphs contain a representation of the numerical solution and, for comparison purposes, the graphs detailing the evolution for the level of assets also contain the certainty equivalent solution.

The simulations for rule 3 use the same values of the parameters as in the previous two sections. Consequently, the numerical solution for the rational expectations model exhibits the same characteristics as discussed before. Therefore, when presenting the results in this section I will concentrate on the solution derived from assuming bounded rationality.

As one can see in Figure 17, the consumption paths have kept their upward slope but for lower values of the coefficient of risk aversion, the difference between the rational expectation and bounded rationality solutions is considerably higher than for the previous two rules. The difference can be clearly seen in the picture, with the rationally bounded consumer consuming more in the beginning while the unboundedly rational consumers consumes more from the 12^{th} period until the end of the horizon. On the other hand, for higher values of the coefficient of risk aversion, consumption paths are almost indistinguishable.

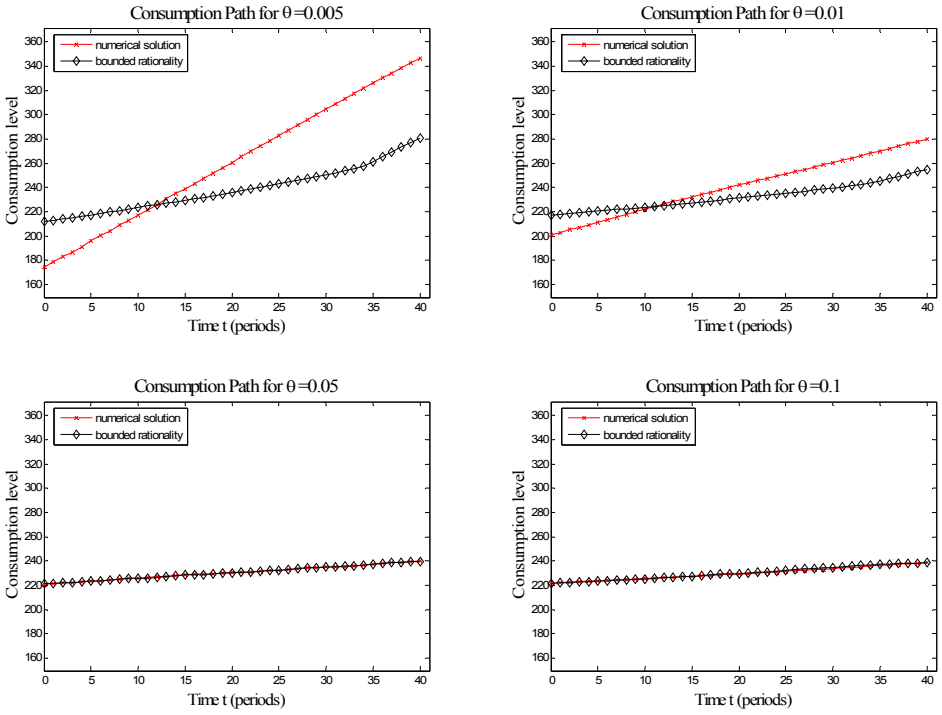

Figure 17. Consumption paths for $\sigma_y = 1$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

Looking at the asset paths in Figure 18 one will notice that, for low values of the coefficient of risk aversion, the bounded rationality assumption leads to much lower levels of precautionary saving than in the case of rational expectations or certainty equivalence. However, the surprising result is that for higher values of the coefficient of risk aversion, there is almost no difference between the level of savings under rational expectations and bounded rationality.

By increasing the standard deviation of income to $\sigma_y = 5$, one can see in Figure 19 a clear difference between the consumption paths for bounded rationality and rational expectations for all levels of risk aversion. As before, the two consumption paths have an upward slope with the rational expectation solution being the steeper one.

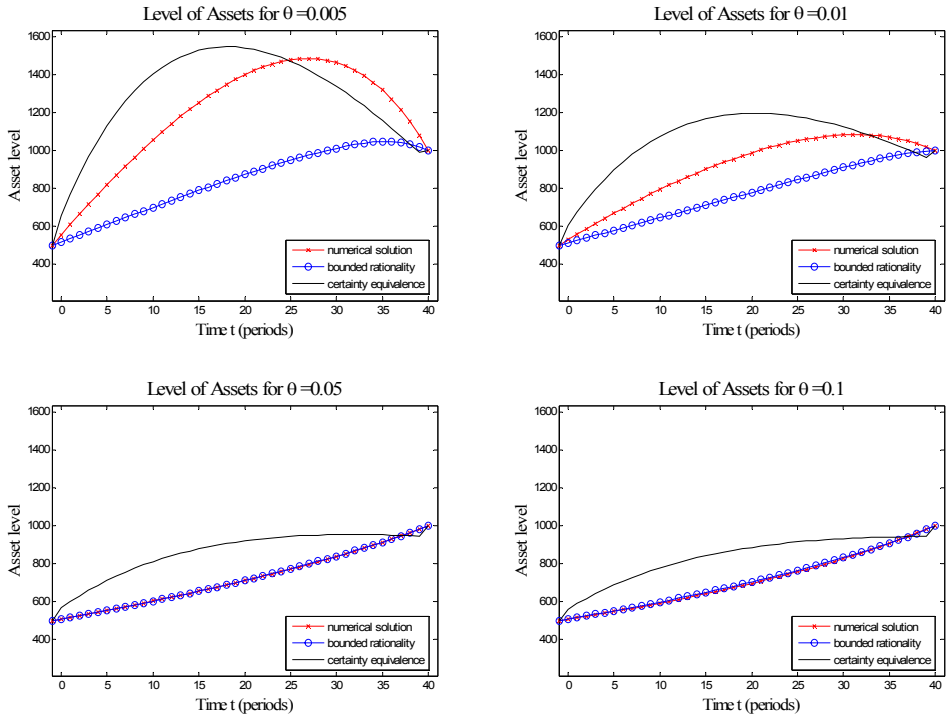

Figure 18. Asset paths for $\sigma_y = 1$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

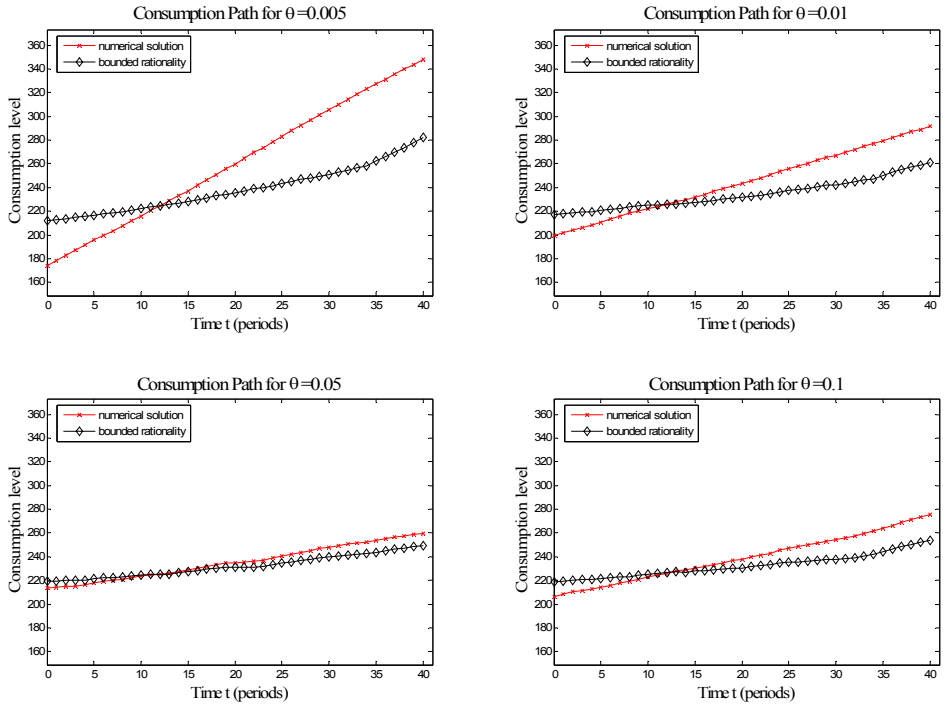

Figure 19. Consumption paths for $\sigma_y = 5$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

The asset paths represented in Figure 20 show clearly a higher level of precautionary saving in the case of rational expectations. The path corresponding to certainty equivalence produces higher levels of saving than the bounded rationality path.

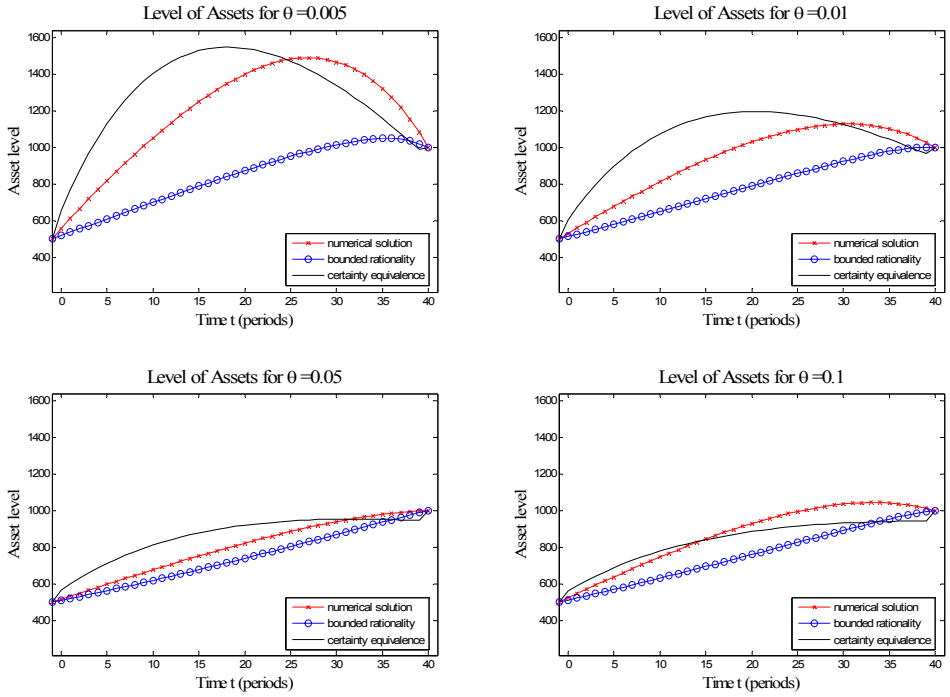

Figure 20. Asset paths for $\sigma_y = 5$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

Increasing again the standard deviation for income to $\sigma_y = 10$, one will notice in Figure 21 that there is not much change in the paths for consumption at low levels of risk aversion. However, the slope of consumption for $\theta = 0.1$ increases quite a lot.

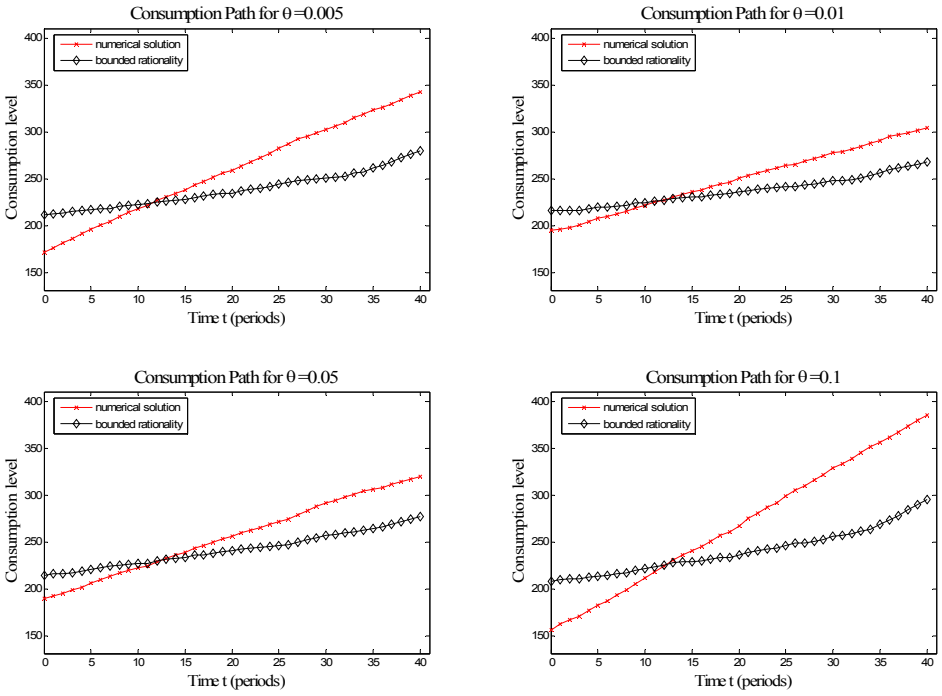

Figure 21. Consumption paths for $\sigma_y = 10$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

On the saving side, one can see in Figure 22 that for the highest coefficient of risk aversion, the rational expectations solution provides a much higher level of savings, while the rationally bounded consumer still saves less than in the case of certainty equivalence for $\theta = 0.01$.

While the level of precautionary saving depends heavily on the parameter values of the model for the unboundedly rational consumer, the same can not be said for the rationally bounded consumer in the case of rule 3. The asset path for the rationally bounded consumer is barely concave and increasing the variance of income does not seem to create the same type of changes as the ones observed for the fully rational consumer. This behavior is the result of optimizing for only short periods of time coupled with the fact that the intermediary asset level targets are chosen assuming a constant growth rate.

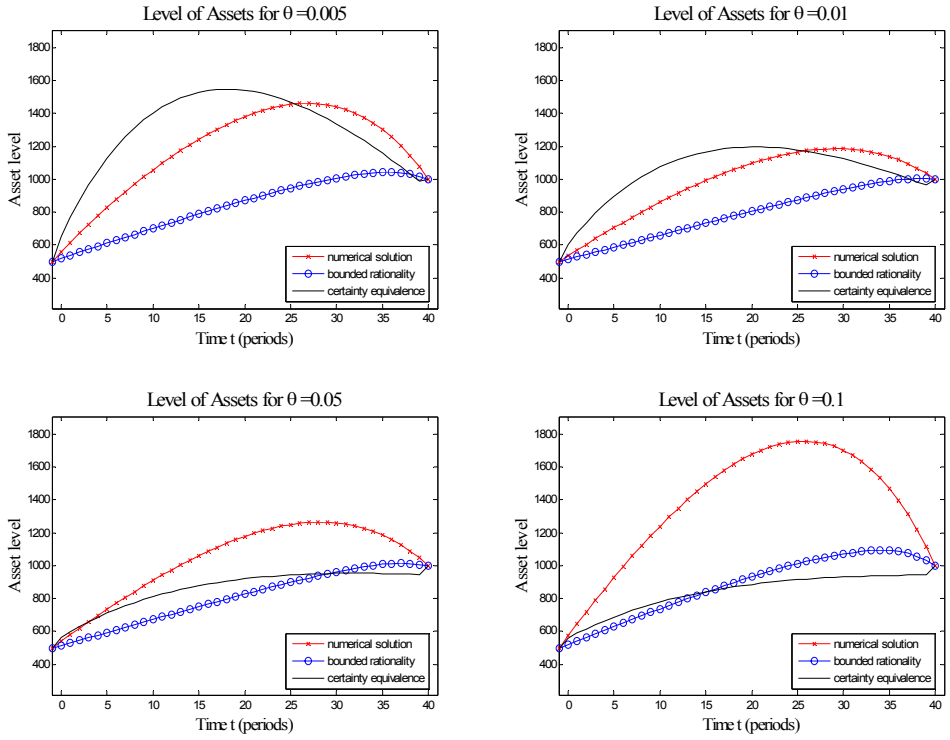

Figure 22. Asset paths for $\sigma_y = 10$, $r_0 = 0.06$ and $\sigma_r = 0.0025$.

In conclusion, in the case of rule 3, the rule employed by the rationally bounded consumer for the accumulation of assets is overshadowing the precautionary motives embedded in the functional specification of the model.

III.4. Final Remarks

The level of precautionary saving under bounded rationality depends quite heavily on the behavior assumptions. While in many of the simulations presented in this chapter the level of precautionary saving chosen on average by the rationally bounded consumer is below that resulting from a rational expectations model, there a few parameterizations of the model, under rule 2, for which the rationally bounded consumer saves more.

The simulations also show that for low coefficients of risk aversion, variation in income uncertainty does not affect much the level of saving. If one adds to this observation the possibility that self selection exists (individuals with high risk aversion choose occupations with low income uncertainty), it is easy to see why some empirical studies would find relatively low levels of precautionary saving.

Another interesting result is that under rule 3, where the rationally bounded consumer follows some form of financial planning, there is not much difference for asset paths across various levels of risk aversion and income uncertainty. This result is consistent with the observation made by Lusardi (1997) that the saving rates do not change much across occupations.

Most of the studies looking to asses the importance of precautionary saving, or the impact of income uncertainty on precautionary saving, have assumed that interest rate uncertainty does not play an important role in the decision making process. For the model discussed in this chapter, the assumption of a constant interest rate would result in an asset path that is constant regardless of the realizations for the income process. By introducing uncertainty in the interest rate process, that is no longer the case. The dynamic of the asset path is especially influenced by the realization of the interest rate

process for lower levels of risk aversion. Therefore, the empirical literature should also consider the impact of interest rate uncertainty when studying the importance of precautionary motives on the level of saving.

While the results presented in this chapter point to an important role for the bounded rationality in the decision making process, it would be difficult to test the model's validity in a standard empirical setting. The problem is that the results depend heavily on the rules adopted as well as on the parameterization of the model and it would be difficult to distinguish between the effects of the general assumptions corresponding to bounded rationality and those specific to a particular rule. Therefore, a more appropriate framework for testing the validity of the model would be an experimental setting. In such a framework, one can potentially "calibrate" the model by identifying the level of risk aversion and the level of patience for each subject. Once these parameters are determined it becomes easier to test hypotheses regarding the decision making process. There have been several studies in the field of experimental economics investigating consumption behavior under uncertainty (Hey and Dardanoni (1988), Ballinger et al. (2003) and Carbone and Hey (2004)) that concluded that actual behavior differs significantly from what is considered optimal. While these studies provide some insights in the decision making process, they do not test for any particular alternative to the optimal behavior corresponding to an unboundedly rational individual. Therefore a future area of research is the design of an experimental framework that could test the hypotheses regarding the decision making process advanced in this chapter.

Appendices

Appendix A. Technical notes to chapter 2

Appendix A1. Definitions for Scenarios, Equivalence Classes and Associated Probabilities

Suppose the world that can be described at each point in time by the vector of state variables x_t, and let u_t denote the control variable while ξ_t is the forcing variable. Suppose ξ_t is a random variable, with the underlying probability space[71] (Ω, Σ, P). ξ_t is defined as $\xi_t : \Omega \to R$ where Ω is countable and finite. If the horizon has $T+1$ time periods and $\xi_t(\omega)$ is a realization of ξ_t for the event $\omega \in \Omega$ in time period t, then the sequence

$$\xi^s(\omega) = \left(\xi_0^s(\omega), \xi_1^s(\omega), \ldots, \xi_T^s(\omega)\right)$$

is called a scenario[72]. From now on, for notation simplification, I will refer to a scenario s simply by ξ^s or by the index s and, in vector form, by $\xi^s = \left(\xi_0^s, \xi_1^s, \ldots, \xi_T^s\right)$.

Let $S(\omega)$ denote the set of all scenarios. Given that Ω is finite, the set $S(\omega)$ is also finite. Therefore, one can define an event tree $\{N, A\}$ characterized by the set of nodes N and the set of arcs A. In this representation, the nodes of the tree are decision points and the arcs are realizations of the forcing variables. The arcs join nodes from

[71] Ω is the sample space, Σ is the sigma field and P is the probability measure.

[72] Other definitions of scenarios can be found in Helgason and Wallace (1991a, 1991b) and Rosa and Ruszczynski (1994).

consecutive levels such that a node n_t^i at level t is linked to N_{t+1} nodes $n_{t+1}^k, k = 1,...,N_{t+1}$ at level $t+1$.

The set of nodes N can be divided into subsets corresponding to each level (period). Suppose that at time t there are N_t nodes. The arcs reaching the nodes $n_t^i, \quad i = 1,\ldots,N_t$ belong each to several scenarios $\xi^q(\omega), \ q = 1,...,L_t$ where L_t represents the number of leaves stemming from a node at level t. The bundle of scenarios that go through one node plays a very important role in the decomposition as well as in the aggregation process. The term equivalence class has been used in the literature to describe the set of scenarios going through a particular node.

By definition, the equivalence class $\left\{s^t\right\}_i, i = 1,\ldots,N_t$ is the set of all scenarios having the first $t+1$ coordinates, $\xi_0,\ldots,\xi_t$ common. This means that for two scenarios $\xi^j = \left(\xi_0^j, \xi_1^j, \ldots, \xi_{t-1}^j, \xi_t^j, ..., \xi_T^j\right)$ and $\xi^k = \left(\xi_0^k, \xi_1^k, \ldots, \xi_{t-1}^k, \xi_t^k, ...\xi_T^k\right)$ that belong to the equivalence class $\left\{s^t\right\}_i, i = 1,\ldots,N_t$ the first $t+1$ elements are common, that is, $\xi_l^j = \xi_l^k \quad for \quad l = 0,...,t$. Formally,

$$\left\{s^t\right\}_i = \left\{\xi^k \,|\, \xi_l^k = \xi_l^i \quad for \quad l = 0,...,t\right\}$$

As mentioned in the above description of the event tree, at time t there are N_t nodes. Then, the number of distinct equivalence classes $\left\{s^t\right\}_i$ is also N_t, that is, $i = 1,\ldots,N_t$. Every node $n_t^i, \quad i = 1,\ldots,N_t$ is associated with an equivalence class $\left\{s^t\right\}_i$. The number of elements of the set $\left\{s^t\right\}_i$ is given by the number of leaves stemming from node i, level (stage) t.

Since scenarios are viewed in terms of a stochastic vector ξ with stochastic components $\xi_0^s, \xi_1^s, \ldots, \xi_T^s$, it is natural to attach probabilities to each scenario. I denote the probability of a particular realization of a scenario, s, with

$$p(s) = prob(\xi^s).$$

These probabilities are non-negative numbers and sum to one. Formally, $p(s) > 0$ and $\sum_{s \in S} p(s) = 1$. I assume that for each scenario ξ^s the stochastic components $\xi_0^s, \xi_1^s, \ldots, \xi_T^s$ are independent. Then

$$p(s) = prob\left(\xi^s(\omega)\right) = \prod_{t=0}^{T} prob\left(\xi_t^s(\omega)\right) \tag{A.1.1}$$

Further on, I define the probabilities associated with a scenario conditional upon belonging to a certain equivalence class $\left\{s^t\right\}_i$ at time t:

$$p\left(s \middle| s \in \left\{s^t\right\}_i\right) = prob\left(\xi^s \middle| \xi^s \in \left\{s^t\right\}_i\right) = \frac{p(s)}{p\left(\left\{s^t\right\}_i\right)},$$

where $p\left(\left\{s^t\right\}_i\right)$ is the probability mass of all scenarios belonging to the class $\left\{s^t\right\}_i$. Under the assumptions outlined above, $p\left(\left\{s^t\right\}_i\right) = \prod_{\tau=0}^{t} prob\left(\xi_\tau^s(\omega)\right)$. Therefore, the conditional probability is easily computed as

$$p\left(s \middle| \left\{s^t\right\}^i\right) = prob\left(\xi^s \middle| \xi^s \in \left\{s^t\right\}_i\right) = \prod_{\tau=t+1}^{T} prob\left(\xi_\tau^s(\omega)\right)$$

The transition from the state at time t to that at time $t+1$ is governed by the control variable u_t but is also dependent on the realization of the forcing variable, that is, on a particular scenario s.

Appendix A2. Description of the Scenario Aggregation Theory

The idea is to show how a solution can be obtained by using special decomposition methods that exploit the structure of the problem by splitting it into manageable pieces and coordinate their solution.

Let us assume for a moment that the original problem can be decomposed into subproblems, each corresponding to a scenario. Then the subproblems can be described as:

$$\min_{u_t \in U_t \subseteq R^{m_u}} \sum_{t=1}^{T} F_t\left(x_t^s, u_t^s\right), \qquad s \in S \tag{A.2.1}$$

where u_t^s and x_t^s are the control and the state variable respectively, conditional on the realization of scenario s while S is a finite, relatively small set of scenarios.

Formally, by definition, a policy is a function or a mapping $U : S \to R^m$ assigning to each scenario $s \in S$ a sequence of controls $U(s) = \left(u_0^s, u_1^s, \ldots, u_t^s, \ldots, u_T^s\right)$, where u_t^s denotes the decision to be made at time t if the scenario happens to be s. Similarly, the state variable at each stage is associated with a particular scenario s. I use the notation x_t^s to show the link between the state variable and scenario s at time t. One can think of the mappings $U : S \to R^m$ as a set of time linked mappings $U_t : S \to R^{m_t}$ with $m = \sum_{t=1}^{T} m_t$.

The policy function has to satisfy certain constraints if two different scenarios s and s' are indistinguishable at time t on information available about them at time t. Then $u_t^s = u_t^{s'}$, that is, a policy can not require different actions at time t relative to scenarios s and s' if there is no way to tell at time t which of the two scenarios will be

followed. This constraint is referred to as the non-anticipativity constraint. One way to model this constraint is to introduce an information structure by bundling scenarios into equivalence classes[73] as defined above. In this way, the scenario set S is partitioned at each time t into a finite number of disjoint sets, $\{s^t\}_i$. Let the collection of all scenario equivalence classes at time t be denoted by B_t, where $\mathrm{B}_t = \bigcup_i \{s^t\}_i$. In most cases partition B_{t+1} is a refinement of partition B_t, that is, every equivalence class $\{s^t\}_i \in \mathrm{B}_t$ is a union of some equivalence classes $\{s^{t+1}\}_j \in \mathrm{B}_{t+1}$. Formally, $\{s^t\}_i = \bigcup_{j=1\ldots m_i} \{s^{t+1}\}_j$. Looking back to the event tree representation discussed in the previous section, m_i represents the number of nodes n_{t+1}^j at level $t+1$ that are linked to the same node n_t^i.

A policy is defined as implementable if it satisfies the non-anticipativity constraint, that is, $u_t(\omega)$ must be the same for all scenarios that have common past and present[74]. In other words, a policy is implementable if for all $t = 0,\ldots,T$ the t^{th} element is common to all scenarios in the same class $\{s^t\}_i$, i.e. if $u_t(\xi^i) = u_t(\xi^k)$ whenever $\{s^t\}_i = \{s^t\}_k$.

Let Σ be the space of all mappings $U : S \to R^n$ with components $U_t : S \to R^{n_t}$. Then the subspace

[73] Some authors, such as Rockaffeler and Wets (1991), use the term scenario bundle.

[74] For certain problems the non-anticipativity constraint can also be defined in terms of the state variable, that is, $x_t(\omega)$ must be the same for all scenarios that have common past and present.

$$\Pi = \left\{ U \in \Sigma \,|\, U_t \text{ is constant on each class } \left\{ s^t \right\}_i \in B_t, \text{ for } t = 1,...,T \right\}$$

identifies the policies that meet the non-anticipativity constraint.

A policy is admissible if it always satisfies the constraints imposed by the definition of the problem. It is clear that not all admissible policies are also implementable. By definition, a contingent policy is the solution, u^s, to a scenario subproblem. It is obvious that a contingent policy is always admissible but not necessarily implementable. Therefore, the goal is to find a policy that is both admissible and implementable. Such a policy is referred to as a feasible policy.

One way to create a feasible policy from a set on contingent policies is to assign weights (or probabilities) to each scenario and then blend the contingent policies according to these weights. Specifically, if the probabilities associated with each scenario are defined as in (A.2.1), one calculates for every period t and for every equivalence class $\left\{ s^t \right\}_i \in \mathrm{B}_t$ the new policy $\underline{u}_t$ by computing the expected value:

$$\underline{u}_t \left(\left\{ s^t \right\}_i \right) = \sum_{s' \in \left\{ s^t \right\}_i} p\left(s' \middle| \left\{ s^t \right\}_i \right) u_t(s') \tag{A.2.2}$$

Then one defines the new policy for all scenarios s that belong to the equivalence class $\left\{ s^t \right\}_i \in \mathrm{B}_t$ as:

$$\hat{u}_t^s = \underline{u}_t \left(\left\{ s^t \right\}_i \right) \text{ for all } s \in \left\{ s^t \right\}_i \tag{A.2.3}$$

Based on its definition, $\hat{u}_t^s$ is implementable. The operator $J : U \to \hat{U}$ defined by (A.2.2) and (A.2.3) is called the aggregation operator.

Let us rewrite equation (2.4.1) as:

$$\min_{u_t \in U_t \subseteq R^{m_u}} F^s\left(x^s, u^s\right), \qquad s \in S \tag{A.2.4}$$

by defining the functional $F^s\left(x^s, u^s\right) = \sum_{t=1}^{T} F_t\left(x_t(s), u_t(s)\right)$.

Then the overall problem can be reformulated as:

$$\min \sum_{s \in S} p^s F^s\left(x^s, u^s\right) \text{ over all } U \in \Sigma \bigcap \Pi \tag{A.2.5}$$

Let us assume for a moment that $\hat{u}^s$ is an implementable policy obtained as in (A.2.3) from contingent policies u^s and $\overline{u^s}$ is the optimal policy for the particular scenario s of the problem described by (A.2.5). Let $\hat{U}$ and $\overline{U}$ be the collections of policies $\hat{u}^s$ and $\overline{u^s}$ respectively. One can easily see that $\overline{U}$ represents the optimal policy for the problem described by (A.2.5). The question that the scenario aggregation methodology answers is how to obtain the optimal solution $\overline{U}$ from a collection of implementable policies $\hat{U}$.

Appendix A3. Solution to a Scenario Subproblem

In order to take advantage of the fact that scenario aggregation does not require the computation of an exact solution for each scenario, I transform the Lagrangian (2.6.8) by replacing the utility function with a first order Taylor series expansion around the solution obtained in the previous iteration. Hence:

$$e^{-\theta c_t^s} = e^{-\theta c_t^{s(k-1)}}\left[1 - \theta\left(c_t^s - c_t^{s(k-1)}\right)\right]$$

From the transition equation, consumption can be expressed as:

$$c_t^s = \left(1 + r\right) A_{t-1}^s + y_t^s - A_t^s$$

Then $e^{-\theta c_t^s} = e^{-\theta c_t^{s(k-1)}} \left\{1-\theta\left[(1+r)\left(A_{t-1}^s - A_{t-1}^{s(k-1)}\right) - \left(A_t^s - A_t^{s(k-1)}\right)\right]\right\}$. For iteration (k) and scenario s the Lagrangian becomes:

$$\min \sum_{t=0}^{T} \beta^t \left\{ \frac{e^{-\theta C_t^{s(k-1)}}}{\theta}\left[1-\theta(1+r)\left(A_{t-1}^s - A_{t-1}^{s(k-1)}\right) + \theta\left(A_t^s - A_t^{s(k-1)}\right)\right] + W_t^s\left[(1+r)A_{t-1}^s + y_t^s - A_t^s\right] + \right.$$
$$\left. + \frac{1}{2}\rho\left[(1+r)\left(A_{t-1}^s - \underline{A}_{t-1}^{(k-1)}\right) - \left(A_t^s - \underline{A}_t^{(k-1)}\right)\right]^2 \right\}$$

Then, the first order condition with respect to A_t^s is given by:

$$\beta^t\left\{e^{-\theta c_t^{s(k-1)}} - W_t^{s(k)} - \rho\left[(1+r)\left(A_{t-1}^s - \underline{A}_{t-1}^{(k-1)}\right) - \left(A_t^s - \underline{A}_t^{(k-1)}\right)\right]\right\} +$$
$$\beta^{t+1}\left\{-(1+r)e^{-\theta c_{t+1}^{s(k-1)}} + (1+r)W_{t+1}^{s(k)} + \rho(1+r)\left[(1+r)\left(A_t^s - \underline{A}_t^{(k-1)}\right) - \left(A_{t+1}^s - \underline{A}_{t+1}^{(k1)}\right)\right]\right\} = 0$$

Rearranging the terms leads to:

$$\frac{1}{\rho}\left[e^{-\theta c_t^{s(k-1)}} - (1+r)\beta e^{-\theta c_{t+1}^{s(k-1)}} - W_t^{s(k)} + (1+r)\beta W_{t+1}^{s(k)}\right] +$$
$$+(1+r)\underline{A}_{t-1}^{(k-1)} - \underline{A}_t^{(k-1)} - (1+r)^2\beta\underline{A}_t^{(k-1)} + (1+r)\beta\underline{A}_{t+1}^{(k-1)} = \qquad \text{(A.3.1)}$$
$$= (1+r)A_{t-1}^s - A_t^s - (1+r)^2\beta A_t^s + (1+r)\beta A_{t+1}^s$$

Let

$$\Gamma_t^{s(k)} = \frac{1}{\rho}\left[e^{-\theta c_t^{s(k-1)}} - (1+r)\beta e^{-\theta c_{t+1}^{s(k-1)}} - W_t^{s(k)} + (1+r)\beta W_{t+1}^{s(k)}\right] \qquad \text{(A.3.2)}$$

Then the first order condition with respect to A_t^s can be written as:

$$\Gamma_t^{s(k)} + (1+r)\underline{A}_{t-1}^{(k-1)} - \left[1+(1+r)^2\beta\right]\underline{A}_t^{(k-1)} + (1+r)\beta\underline{A}_{t+1}^{(k-1)} =$$
$$= (1+r)A_{t-1}^s - \left[1+(1+r)^2\beta\right]A_t^s + (1+r)\beta A_{t+1}^s$$

For $t = T-1$ the first order condition becomes:

$$\Gamma_{T-1}^{s(k)} + (1+r)\underline{A}_{T-2}^{(k-1)} - \left[1+(1+r)^2\beta\right]\underline{A}_{T-1}^{(k-1)} + (1+r)\beta\underline{A}_T^{(k-1)} =$$
$$= (1+r)A_{T-2}^s - \left[1+(1+r)^2\beta\right]A_{T-1}^s + (1+r)\beta A_T \qquad \text{(A.3.3)}$$

Noting that $\underline{A}_T^{(k-1)} = \underline{A}_T = A_T$ equation (A.3.3) can be written as:

$$\Gamma_{T-1}^{s(k)} + (1+r)\,\underline{A}_{T-2}^{(k-1)} - \left[1+(1+r)^2\beta\right]\underline{A}_{T-1}^{(k-1)} = (1+r)\,A_{T-2}^{s} - \left[1+(1+r)^2\beta\right]A_{T-1}^{s}$$

Similarly, for $t=0$ one obtains:

$$\begin{aligned}&\Gamma_{0}^{s(k)} + (1+r)\,\underline{A}_{-1}^{(k-1)} - \left[1+(1+r)^2\beta\right]\underline{A}_{0}^{(k-1)} + (1+r)\beta\,\underline{A}_{1}^{(k-1)} = \\ &(1+r)\,A_{-1}^{s} - \left[1+(1+r)^2\beta\right]A_{0}^{s} + (1+r)\beta A_{1}^{s}\end{aligned} \tag{A.3.4}$$

Again, noting that A_{-1} is given, $\underline{A}_{-1}^{(k-1)} = A_{-1}^{s}$ so equation (A.3.4) becomes:

$$\Gamma_{0}^{s(k)} - \left[1+(1+r)^2\beta\right]\underline{A}_{0}^{(k-1)} + (1+r)\beta\,\underline{A}_{1}^{(k-1)} = -\left[1+(1+r)^2\beta\right]A_{0}^{s} + (1+r)\beta A_{1}^{s}$$

Rewriting the system of equations in matrix form, leads to:

$$\begin{bmatrix} -\left[1+(1+r)^2\beta\right] & (1+r)\beta & 0 & 0 & \dots & 0 & 0 \\ (1+r) & -\left[1+(1+r)^2\beta\right] & (1+r)\beta & 0 & \dots & 0 & 0 \\ 0 & (1+r) & -\left[1+(1+r)^2\beta\right] & (1+r)\beta & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & (1+r) & -\left[1+(1+r)^2\beta\right] \end{bmatrix} \begin{bmatrix} A_0^s \\ A_1^s \\ A_2^s \\ \vdots \\ A_{T-1}^s \end{bmatrix} =$$

$$= \begin{bmatrix} \Gamma_0^{s(k)} - \left[1+(1+r)^2\beta\right]\underline{A}_0^{(k-1)} + (1+r)\beta\,\underline{A}_1^{(k-1)} \\ \Gamma_1^{s(k)} + (1+r)\,\underline{A}_0^{(k-1)} - \left[1+(1+r)^2\beta\right]\underline{A}_1^{(k-1)} + (1+r)\beta\,\underline{A}_2^{(k-1)} \\ \vdots \\ \Gamma_{T-1}^{s(k)} + (1+r)\,\underline{A}_{T-2}^{(k-1)} - \left[1+(1+r)^2\beta\right]\underline{A}_{T-1}^{(k-1)} \end{bmatrix}$$

Appendix B. Technical notes to chapter 3

Appendix B1. Analytical Solution for a Scenario with Deterministic Interest Rate

Consider the problem described by (3.3.1) - (3.3.4). Solving the period-by-period budget constraint (3.3.2) for c_t, $t = T-1$ and $t = T$, and substituting back into the utility function, the period $T-1$ optimization problem is given by:

$$\max_{A_{T-1}} \left\{ -\frac{\exp\left\{-\theta\left[\left(1+r_{T-1}\right)A_{T-2} + y_{T-1} - A_{T-1}\right]\right\}}{\theta} - E\left(\beta \frac{\exp\left\{-\theta\left[\left(1+r_T\right)A_{T-1} + y_T - A_T\right]\right\}}{\theta} \middle| I_{T-1} \right) \right\} \tag{B.1.1}$$

subject to

$$A_{T-1} \geq -b \tag{B.1.2}$$

Taking derivatives with respect to A_{T-1}, the Euler equation for (B.1.1) is given by:

$$\exp\left[-\theta\left(1+r_{T-1}\right)A_{T-2} - \theta y_{T-1} + \theta A_{T-1}\right] = \max\left\{ \begin{array}{c} \exp\left[-\theta\left(1+r_{T-1}\right)A_{T-2} - \theta y_{T-1} - \theta b\right], \\ \beta\left(1+r_T\right)E\left\{\exp\left[-\theta\left(1+r_T\right)A_{T-1} - \theta y_T + \theta A_T\right] \middle| I_{T-1}\right\} \end{array} \right\} \tag{B.1.3}$$

Note that $y_T = y_{T-1} + \xi_T$ while $E\left[\exp\left(-\theta\xi_T\right) | I_{T-1}\right] = \exp\left(\frac{\theta^2\sigma_y^2}{2}\right)$ and hence solving (B.1.3) for the optimal wealth level at the beginning of period $T-1$ yields:

$$A_{T-1}^* = \max\left\{-b, \frac{\left[\left(1+r_{T-1}\right)A_{T-2} + \Gamma_T^* + A_T\right]}{\left(2+r_T\right)}\right\}. \tag{B.1.4}$$

where $\Gamma_T^* = \Gamma + \left\{\log\left[\beta\left(1+r_T\right)\right]\right\}/\theta$, and $\Gamma = \theta\sigma_y^2/2$.

Going now to period $T-2$, the optimization problem is given by

$$\max_{A_{T-2}}\left\{-\frac{\exp\left\{-\theta\left[\left(1+r_{T-2}\right)A_{T-3}+y_{T-2}-A_{T-2}\right]\right\}}{\theta}-\beta E\left(\frac{\exp\left\{-\theta\left[\left(1+r_{T-1}\right)A_{T-2}+y_{T-1}-A_{T-1}^{*}\right]\right\}}{\theta}+\right.\right.$$
$$\left.\left.+\beta\frac{\exp\left[-\theta\left(\left(1+r_{T}\right)A_{T-1}^{*}+y_{T}-A_{T}\right)\right]}{\theta}\middle|I_{T-2}\right)\right\} \tag{B.1.5}$$

subject to

$$A_{T-2}\geq -b \tag{B.1.6}$$

Taking derivatives with respect to A_{T-2}, and noting that

$$E\left[\exp\left(-\theta A_{T-1}^{*}\right)\middle|I_{T-2}\right]=\exp\left(-\theta A_{T-1}^{*}\right),$$

the Euler equation for (B.1.5) is given by:

$$\begin{aligned}&\exp\left[-\theta\left(1+r_{T-2}\right)A_{T-3}-\theta y_{T-2}+\theta A_{T-2}\right]\\&=\max\left\{\begin{array}{c}\exp\left[-\theta\left(1+r_{T-2}\right)A_{T-3}-\theta y_{T-2}-\theta b\right],\\ \beta\left(1+r_{T-1}\right)\exp\left[-\theta\left(1+r_{T-1}\right)A_{T-2}+\theta A_{T-1}^{*}\right]E\left[\exp\left(-\theta y_{T-1}\right)\middle|I_{T-2}\right]\end{array}\right\}\end{aligned} \tag{B.1.7}$$

Since $y_{T-1}=y_{T-2}+\xi_{T-1}$, (B.1.7) can be rewritten as:

$$\begin{aligned}&\exp\left[-\theta\left(1+r_{T-2}\right)A_{T-3}-\theta y_{T-2}+\theta A_{T-2}\right]\\&=\max\left\{\begin{array}{c}\exp\left[-\theta\left(1+r_{T-2}\right)A_{T-3}-\theta y_{T-2}-\theta b\right],\\ \beta\left(1+r_{T-1}\right)\exp\left[-\theta\left(1+r_{T-1}\right)A_{T-2}-\theta y_{T-2}+\theta A_{T-1}^{*}\right]E\left[\exp\left(-\theta\xi_{T-1}\right)\middle|I_{T-2}\right]\end{array}\right\}\end{aligned}$$

Assuming that liquidity constraint is not binding, solving (B.1.7) for A_{T-2} yields:

$$\left(1+r_{T-2}\right)A_{T-3}-A_{T-2}=-\frac{\ln\left[\beta\left(1+r_{T-1}\right)\right]}{\theta}+\left(1+r_{T-1}\right)A_{T-2}-A_{T-1}^{*}-\frac{\theta\sigma_{y}^{2}}{2} \tag{B.1.8}$$

Using the notation from above, equation (B.1.8) can be written as:

$$\Gamma_{T-1}^{*}=-A_{T-1}^{*}+\left(2+r_{T-1}\right)A_{T-2}-\left(1+r_{T-2}\right)A_{T-3} \tag{B.1.9}$$

Similarly, for period t, the equivalent of equation (B.1.9) is given by:

$$\Gamma_{t+1}^{*} = -A_{t+1}^{*} + \left(2 + r_{t+1}\right) A_t - \left(1 + r_t\right) A_{t-1} \tag{B.1.10}$$

It is clear that the optimal wealth level at the beginning of period t does not depend on labor income received at the beginning of the period. This result is not general, but is rather specific to the life-cycle model with a negative exponential utility function and labor income following an arithmetic random walk process.

Solving for the beginning-of-period wealth levels from $t = 0$ to $t = T - 1$ means solving the system of linear equations:

$$\mathrm{D}\begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ \vdots \\ A_{T-3} \\ A_{T-2} \\ A_{T-1} \end{pmatrix} = \begin{pmatrix} \left(1 + r_0\right) A_{-1} + \Gamma_1^{*} \\ \Gamma_2^{*} \\ \Gamma_3^{*} \\ \vdots \\ \Gamma_{T-2}^{*} \\ \Gamma_{T-1}^{*} \\ A_T + \Gamma_T^{*} \end{pmatrix} \tag{B.1.11}$$

where **D** is a tridiagonal coefficient matrix,

$$\mathrm{D} = \begin{pmatrix} 2 + r_1 & -1 & 0 & \cdots & 0 & 0 & 0 \\ -\left(1 + r_1\right) & 2 + r_2 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\left(1 + r_{T-2}\right) & 2 + r_{T-1} & -1 \\ 0 & 0 & 0 & \cdots & 0 & -\left(1 + r_{T-1}\right) & 2 + r_T \end{pmatrix} \tag{B.1.12}$$

Once the values for wealth levels are computed, the consumption levels follow. The solution presented in this section is in fact the solution for a scenario obtained by discretizing the distribution of the forcing variable for the interest rate. Since an analytical solution can be obtained when income follows an arithmetic random walk and interest rate is deterministic, it is no longer necessary to discretize both forcing variables, but only the interest rate. This approach reduces considerably the computational burden.

For different labor income processes, a dual discretization is necessary, that is, for both forcing variables.

Appendix B2. Details on the Assumptions in Rule 1

In period t the consumer wants to solve the optimization problem given by:

$$\max_{\{c_\tau\}_{\tau=t}^{T}} E\left[\sum_{\tau=t}^{T} -\beta^{\tau-t}\left(\frac{1}{\theta}\right)\exp\left(-\theta c_\tau\right) | I_t\right] \tag{B.2.1}$$

subject to

$$\begin{aligned} A_\tau &= \left(1+r_\tau\right)A_{\tau-1} + y_\tau - c_\tau, \qquad \tau = t, t+1, ..., T, \\ &\text{with } A_{t-1}, A_T \text{ given} \quad t = 0,1,...,T-1, \end{aligned} \tag{B.2.2}$$

$$\begin{aligned} y_\tau &= y_{\tau-1} + \xi_\tau, \qquad \tau = t+1, ..., T, \\ &\text{with } \mathrm{y}_t \text{ given} \quad \mathrm{t} = 0,1,...,T-1, \end{aligned} \tag{B.2.3}$$

$$\begin{aligned} r_\tau &= r_{\tau-1} + \upsilon_\tau, \qquad \tau = t+1, ..., T, \\ &\text{with } \mathrm{r}_t \text{ given} \quad \mathrm{t} = 0,1,...,T-1, \end{aligned} \tag{B.2.4}$$

The assumption is that the forcing variable υ_τ has three possible realizations, $\{\upsilon_a, \upsilon_b, \upsilon_c\}$. The set of its realizations determines the event tree and consequently the set of scenarios. For T_h periods the number of all scenarios is 3^{T_h}. The consumer considers all the possible scenarios from period t to period $t+T_h$. From there on it assumes that for every leaf the scenario will be determined by υ_τ taking its unconditional mean, that is, zero. For example, if the short optimizing horizon is given by $T_h = 4$ and the sequence of realizations for υ_τ up to period $t+4$, for a particular scenario, is $\{\upsilon_a, \upsilon_c, \upsilon_b, \upsilon_c\}$, the assumption made by consumer is that for this particular scenario the realizations of υ_τ for the rest of the periods will be 0, that is, the whole scenario is $\{\upsilon_a, \upsilon_c, \upsilon_b, \upsilon_c, 0, 0, ..., 0\}$.

This process is repeated as the consumer advances to period $t+1$ and goes again through the optimization procedure. The number of scenarios considered remains the same unless $T-t<T_h$, which is to say that there are fewer than T_h periods left until the terminal period.

Appendix B3. Details on the Assumptions in Rule 2

In period t the consumer wants to solve the optimization problem given by:

$$\max_{\{c_\tau\}_{\tau=t}^{T}} E\left[\sum_{\tau=t}^{T} -\beta^{\tau-t}\left(\frac{1}{\theta}\right)\exp\left(-\theta c_\tau\right) | I_t\right] \tag{B.3.1}$$

subject to

$$\begin{aligned} A_\tau &= \left(1+r_\tau\right)A_{\tau-1}+y_\tau-c_\tau, \qquad \tau=t,t+1,...,T, \\ &\text{with } A_{t-1}, A_T \text{ given} \quad t=0,1,...,T-1, \end{aligned} \tag{B.3.2}$$

$$\begin{aligned} y_\tau &= y_{\tau-1}+\xi_\tau, \quad \tau=t+1,...,T, \\ &\text{with } \mathrm{y}_t \text{ given} \quad \mathrm{t}=0,1,...,T-1, \end{aligned} \tag{B.3.3}$$

$$\begin{aligned} r_\tau &= r_{\tau-1}+\upsilon_\tau, \quad \tau=t+1,...,T, \\ &\text{with } \mathrm{r}_t \text{ given} \quad \mathrm{t}=0,1,...,T-1, \end{aligned} \tag{B.3.4}$$

The assumption is that the forcing variable υ_τ has three possible realizations, $\{\upsilon_a,\upsilon_b,\upsilon_c\}$. The set of its realizations determines the event tree and consequently the set of scenarios. For T_h periods the number of all scenarios is 3^{T_h}. The consumer considers all the possible scenarios from period t to period $t+T_h$. From there on it assumes that for every leaf only three more scenarios emerge, with υ_τ taking only one of the three values $\{\upsilon_a,\upsilon_b,\upsilon_c\}$ every period until the end of the horizon. For example, if the short optimizing horizon is given by $T_h=4$ and the sequence of realizations for υ_τ up to

period $t+4$, for a particular scenario, is $\{\upsilon_a,\upsilon_c,\upsilon_b,\upsilon_c\}$, the assumption made by consumer is that only three more scenarios will stem from the leaf corresponding to scenario $\{\upsilon_a,\upsilon_c,\upsilon_b,\upsilon_c\}$. These three scenarios are given by $\{\upsilon_a,\upsilon_c,\upsilon_b,\upsilon_c,\upsilon_a,\upsilon_a,...,\upsilon_a\}$, $\{\upsilon_a,\upsilon_c,\upsilon_b,\upsilon_c,\upsilon_b,\upsilon_b,...,\upsilon_b\}$ and $\{\upsilon_a,\upsilon_c,\upsilon_b,\upsilon_c,\upsilon_c,\upsilon_c,...,\upsilon_c\}$. Effectively, the total number of scenarios considered is 3^{T_h+1} as opposed to 3^{T-t} which would represent the total number of scenarios for the horizon from period t to period T.

This whole process is repeated as the consumer advances to period $t+1$ and goes again through the optimization procedure. The number of scenarios considered remains the same unless $T-t<T_h$, which is to say that there are fewer than T_h periods left until the terminal period.

Bibliography

Allen, T. W., Carroll, C. D. (2001). Individual Learning About Consumption. *Macroeconomic dynamics,* 5, 255-271.

Anderson, G., and Moore, G. (1986). An Efficient Procedure for Solving Nonlinear Perfect Foresight Models. Working Paper, January 1986.

Atkinson, K.E. (1976). A Survey of Numerical Methods for the Solution of Fredholm Integral Equations of the Second Kind. Society for Industrial and Applied Mathematics, Philadelphia.

Baker, C. T. H. (1977). The Numerical Treatment of Integral Equations. Clarendon Press, Oxford.

Ballinger, T. P., Palumbo, M. G., and Wilcox, N. T. (2003). Precautionary Saving and Social Learning across Generations: An Experiment. *The Economic Journal*, 113, 920-947.

Banks, J., Blundell, R., and Brugiavini, A. (2001). Risk Pooling, Precautionary Saving and Consumption Growth. *Review of Economic Studies*, 68, 757-779.

Baxter M., Crucini, M., and Rouwenhorst, K. G. (1990). Solving the Stochastic Growth Model by a Discrete-State-Space, Euler Equation Approach. *Journal of Business and Economic Statistics*, 8, 19-21.

Binder, M., Pesaran, M. H., Samiei, S. H. (2000). Solution of Nonlinear Rational Expectations Models with Applications to Finite-Horizon Life-Cycle Models of Consumption. *Computational Economics,* 15, 25-57.

Birge, J. R. (1985). Decomposition and Partitioning Methods for Multistage Stochastic Linear Programs. *Operations Research,* 33, 989-1007.

Bitros, G. C., and Kelejian, H. H. (1976). A Stochastic Control Approach to Factor Demand. *International Economic Review*, 17, 701-717.

Burnside, C. (1993). Consistency of a Method of Moments Estimator Based on Numerical Solutions to Asset Pricing Models. *Econometric Theory*, 9, 602-632.

Burnside, C. (1999). Discrete State-Space Methods for the Study of Dynamics Economies. *Computational methods for the study of dynamic economies*, ed. by R. Marimon and A. Scott. Oxford University Press, Oxford and New York, 95-113.

Carbone, E., and Hey, J. D. (2004). The Effect of Unemployment on Consumption: An Experimental Analysis. *The Economic Journal*, 114, 660-683.

Carroll, C. D. (1994). How Does Future Income Affect Current Consumption? *Quarterly Journal of Economics*, 109, 111-147.

Carroll, C. D., and Samwick, A. (1997). The Nature of Precautionary Wealth. *Journal of Monetary Economics*, 40, 41-71.

Cecchetti, S. G., Lam, P. -S., and Mark, N. C. (1993). The Equity Premium and the Risk-Free Rate: Matching the Moments. *Journal of Monetary Economics*, 32, 21-45.

Christiano, L. J. (1990a). Solving the Stochastic Growth Model by Linear-Quadratic Approximation and by Value-Function Iteration. *Journal of Business and Economic Statistics*, 8, 23-26.

Christiano, L. J. (1990b). Linear-Quadratic Approximation and Value-function Iteration: A Comparison. *Journal of Business and Economic Statistics*, 8, 99-113.

Christiano, L. J. and Fischer, J. D. M. (2000). Algorithms for Solving Dynamic Models with Occasionally Binding Constraints. *Journal of Economic Dynamics and Control*, 24, 1179-1232.

Chow, G. C. (1973). Effect of Uncertainty on Optimal Control Policies. *International Economic Review*, 14, 632-645.

Chow, G. C. (1976). The Control of Nonlinear Econometric Systems with Unknown Parameters. *Econometrica*, 44, 685-695.

Coleman, W.J., II (1990). Solving the Stochastic Growth Model by Policy-Function Iteration. *Journal of Business and Economic Statistics*, 8, 27-29.

Collard, F. and Juillard, M. (2001). Accuracy of Stochastic Perturbation Methods: The Case of Asset Pricing Models. *Journal of Economic Dynamics and Control*, 25, 979-99.

Conlisk, J. (1996). Why Bounded Rationality? *Journal of Economic Literature,* 34, 669-700.

Dardanoni, V. (1991). Precautionary Savings under Income Uncertainty: A Cross-Sectional Analysis. *Applied Economics*, 23, 153-160.

Deaton, A. and Laroque, G. (1992). On the Behavior of Commodity Prices. *Review of Economic Studies*, 59, 1-23.

Den Haan, W. J. and Marcet, A. (1990). Solving the Stochastic Growth Model by Parameterizing Expectations. *Journal of Business and Economic Statistics*, 8, 31-34.

Den Haan, W. J., and Marcet, A. (1994). Accuracy in Simulations. *Review of Economic Studies*, 61, 3-17.

Dotsey, M. and Mao, C. S. (1992). How Well Do Linear Approximation Methods Work? The Production Tax Case. *Journal of Monetary Economics,* 29, 25-58.

Dynan, K. (1993). How Prudent Are Consumers? *Journal of Political Economy*, 101, 1104-1113.

Fair, R. C. (2003). Optimal Control and Stochastic Simulation of Large Nonlinear Models with Rational Expectations. *Journal of Economic Dynamics and Control*, 21, 245-256.

Fair, R. C., and Taylor, J.B. (1983). Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models. *Econometrica*, 51, 1169-1185.

Fuhrer, J. C., and Bleakley, C. H. (1996). Computationally Efficient Solution and Maximum Likelihood Estimation of Nonlinear Rational Expectations Models. Mimeo, Federal Reserve Bank of Boston.

Gagnon, J. E. (1990). Solving the Stochastic Growth Model by Deterministic Extended Path. *Journal of Business and Economic Statistics*, 8, 35-36.

Gali, J., Lopez-Salido, J. D., Valles, J. (2004). Rule-of-Thumb Consumers and the Design of Interest Rate Rules. *NBER Working Paper Series*. Working Paper 10392.

Guariglia, A. (2001). Saving Behaviour and Earnings Uncertainty: Evidence from the British Household Panel Survey. *Journal of Population Economics*, 14, 619-634.

Guiso, L., Jappelli, T., Terlizzese, D. (1992). Earnings Uncertainty and Precautionary Saving. *Journal of Monetary Economics*, 30, 307-337.

Helgason, T., and Wallace, S. W. (1991a). Approximate Scenario Solutions in the Progressive Hedging Algorithm. *Annals of Operation Research,* 31, 437-444.

Helgason, T., and Wallace, S. W. (1991b). Structural Properties of the Progressive Hedging Algorithm. *Annals of Operation Research,* 31, 445-456.

Hey, J. D., and Dardanoni, V. (1988). Optimal Consumption under Uncertainty: An Experimental Investigation. *The Economic Journal*, 98, 105-116.

Ingram, B. F. (1990). Equilibrium Modeling of Asset Prices: Rationality versus Rules of Thumb. *Journal of Business and Economic Statistics,* 8, 115–125.

Judd, K.L. (1992). Projection Methods for Solving Aggregate Growth Models. *Journal of Economic Theory*, 58, 410-452.

Judd, K. L. (1998). *Numerical Methods in Economics*. MIT Press, Cambridge, MA.

Klenow, P.J. (1991). Externalities and Business Cycles. Ph.D. thesis, Department of Economics, Stanford University.

Krusell, P., Smith A. A. Jr. (1996). Rules of Thumb in Macroeconomic Equilibrium - A Quantitative Analysis. *Journal of Economic Dynamics and Control*, 20, 527-558.

Kydland, F., and Prescott, E. (1982). Time to Build and Aggregate Fluctuations. *Econometrica*, 50, 1345-1370.

Lettau, M., and Uhlig, H. (1999). Rules of Thumb versus Dynamic Programming. *The American Economic Review*, 89, 141-172.

Lusardi, A. (1997). Precautionary Saving and Subjective Earnings Variance. *Economic Letters*, 57, 319-326.

Marcet, A. (1988). Solving Nonlinear Stochastic Growth Models by Parameterizing Expectations. Carnegie-Mellon University manuscript.

Marcet, A. (1994). Simulation Analysis of Dynamic Stochastic Models: Application to Theory and Estimation. *Advances in Econometrics, Sixth World Congress*, Vol. II, ed. by C. Sims. Cambridge University Press, Cambridge U.K., 91-118.

Marcet, A., and Lorenzoni, G. (1999). The Parameterized Expectations Approach; Some Practical Issues. *Computational methods for the study of dynamic economies*, ed. by R. Marimon and A. Scott. Oxford University Press, Oxford and New York, 143-171.

Marcet, A. and Marshall, D. A. (1994a). Convergence of Approximate Model Solutions to Rational Expectations Equilibria using the Method of Parameterized Expectations. Working Paper No. 73, Department of Finance, Kellogg Graduate School of Management, Northwestern University.

Marcet, A. and Marshall, D. A. (1994b). Solving Non-linear Rational Expectations Models by Parameterized Expectations: Convergence to Stationary Solutions. Federal Reserve Bank of Chicago, Working Paper 94-20.

McGrattan, E. R. (1990). Solving the Stochastic Growth Model by Linear-Quadratic Approximation. *Journal of Business and Economic Statistics*, 8, 41-44.

McGrattan, E. R. (1996). Solving the Stochastic Growth Model with a Finite-Element Method. *Journal of Economic Dynamics and Control*, 20, 19-42.

McGrattan, E. R. (1999). Application of Weighted Residual Methods to Dynamic Economic Models. *Computational methods for the study of dynamic economies*, ed. by R. Marimon and A. Scott. Oxford University Press, Oxford and New York, 114-142.

Mehra, R. and Prescott, E. C. (1985). The Equity Premium: A Puzzle. *Journal of Monetary Economics*, 15, 145-161.

Merrigan, P., Normandin, M. (1996). Precautionary Saving Motives: An Assessment from UK Time Series of Cross-Sections. *Economic Journal*, 106, 1193-1208.

Miles, D. (1997). A Household Level Study of the Determinants of Income and Consumption. *Economic Journal*, 107, 1-25.

Miranda, M.J., and Helmberger, P.G. (1988). The Effects of Commodity Price Stabilization Programs. *American Economic Review*, 78, 46-58.

Miranda, M. J., and Rui, X. (1997). Maximum Likelihood Estimation of the Nonlinear Rational Expectations Asset Pricing Model. *Journal of Economic Dynamics and Control,* 21, 1493-1510.

Mulvey, J. M., and Ruszczynski, A. (1992). A diagonal quadratic approximation method for large scale linear programs. *Operations Research Letters*, 12, 205-215.

Novales, A. et al. (1999). Solving Nonlinear Rational Expectations Models by Eigenvalue-Eigenvector Decomposition. *Computational methods for the study of dynamic economies*, ed. by R. Marimon and A. Scott. Oxford University Press, Oxford and New York, 62-92.

Prucha, I. R., and Nadiri, M. I. (1984). Formulation and Estimation of Dynamic Factor Demand Equations under Non-Static Expectations: A Finite Horizon Model, *NBER Working Paper Series*. Revised technical working paper no. 26.

Prucha, I. R., and Nadiri, M. I. (1986). A Comparison of Alternative Methods for the Estimation of Dynamic Factor Demand Models under Non-Static Expectations. *Journal of Econometrics* 33, 187-211.

Prucha, I. R., and Nadiri, M. I. (1991). On the Specification of Accelerator Coefficients in Dynamic Factor Demand Models. *Economic Letters* 35, 123-129.

Reiter, M. (2000). Estimating the Accuracy of Numerical Solutions to Dynamic Optimization Problems. Mimeo.

Rockafellar, R. T., and Wets, R. J. -B. (1991). Scenarios and Policy Aggregation in Optimization under Uncertainty. *Mathematics of Operations Research*, 16, 1-23.

Rosa, C., and Ruszczynski, A. (1994). On Augmented Lagrangian Decomposition Methods for Multistage Stochastic Programming. *International Institute for Applied Analysis*, Working Paper WP-94-125.

Rust, J. (1996). Numerical Dynamic Programming in Economics. *Handbook of Computational Economics*, Vol. I, ed. by H. Amman, D. Kendrick and J. Rust. Amsterdam: North-Holland, 619-729.

Rust, J. (1997). A Comparison of Policy Iteration Methods for Solving Continuous-state, Infinite-horizon Markovian Decision Problems Using Random, Quasi-Random, and Deterministic Discretizations. Manuscript, Yale University.

Ruszczynski, A. (1986). A Regularized Decomposition Method for Minimizing a Sum of Polyhedral Functions. *Mathematical Programming,* 35, 309-333.

Ruszczynski, A. (1989). An Augmented Lagrangian Decomposition Method for Block Diagonal Linear Programming Problems. *Operations Research Letters*, 8, 287-294.

Ruszczynski, A. (1993). Parallel Decomposition of Multistage Stochastic Programs. *Mathematical Programming*, 58, 201-228.

Santos, M. S. (2000). Accuracy of Numerical Solutions Using the Euler Equation Residuals. *Econometrica*, 68, 1377-1402.

Santos, M. S. and Vigo, J. (1998). Analysis of a Numerical Dynamic Programming Algorithm Applied to Economic Models. *Econometrica*, 66, 409-426.

Sargent, T. J. (1993). Bounded Rationality in Macroeconomics. *Oxford University Press*.

Schmitt-Grohé, S. and Uribe, M. (2004). Solving Dynamic General Equilibrium Models Using a Second-Order Approximation to the Policy Function. *Journal of Economic Dynamics and Contro*l, 28, 755-775.

Skinner, J. (1988). Risky Income, Life Cycle Consumption, and Precautionary Savings. *Journal of Monetary Economics,* 22, 237-255.

Tauchen, G. (1990). Solving the Stochastic Growth Model by Using Quadrature Methods and Value-Function Iterations. *Journal of Business and Economic Statistics*, 8, 49-51.

Tauchen, G. and Hussey, R. (1991). Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models. *Econometrica*, 59, 371-396.

Taylor, J. B. and Uhlig, H. (1990). Solving Nonlinear Stochastic Growth Models: A Comparison of Alternative Solution Methods. *Journal of Business and Economic Statistics*, 8, 1-17.

Uhlig, H. (1999). A Toolkit for Analyzing Nonlinear Dynamic Stochastic Models Easily. *Computational methods for the study of dynamic economies*, ed. by R. Marimon and A. Scott. Oxford University Press, Oxford and New York, *30-61*.

Van Slyke, R., and Wets, R. J. -B. (1969). L-Shaped Linear Programs with Applications to Optimal Control and Stochastic Programming. *SIAM Journal on Applied Mathematics*, 17, 638-663.

Wets, R. J. -B. (1988). Large Scale Linear Programming. *Numerical methods in stochastic programming,* ed. by Yu Ermoliev and R. J. -B. Wets. Springer Verlag, Berlin, 65-94.

Wright, B.D. and Willam, J.C. (1982a). The Economic Role of Commodity Storage. *Economic Journal*, 92, 596-614.

Wright, B.D. and Willam, J.C. (1982b). The Roles of Public and Private Storage in Managing Oil Import Disruptions. *Bell Journal of Economics*, 13, 341-353.

Wright, B.D. and Willam, J.C. (1984). The Welfare Effects of the Introduction of Storage. *Quarterly Journal of Economics*, 99, 169-182.