Computational Techniques for Solving Asset Pricing Models*

Andrew J. Culham
Department of Mathematics
Florida State University

June 10, 2005

Abstract

Frequently in financial economics, problems arise for which no closed form solution exists. In these cases one must resort to numerical techniques to approximate the solution. This paper will examine several of these numerical techniques and apply them to a model with a known, exact solution. The paper will begin with a brief introduction to the Lucas asset pricing model and conclude with a brief overview of agent-based economic models.

Contents

*This paper is being submitted to the Department of Mathematics for the Advanced Topics Exam in Financial Mathematics.

Academic Advisors: Dr. Alec Kercheval and Dr. Paul Beaumont

Exam Date: June 14, 2005

1 Introduction

Models used in financial economics are becoming increasingly complex. As the models get more sophisticated, the modelers move further away from a world in which analytical solutions exist and are easily found. Nonetheless, the solutions to these models are of major interest. As a result, economists have been forced to find alternative solution methods for these problems. In particular, the application of numerical analysis has become a common approach for gaining insight into these otherwise unsolvable models.

In this paper we will explore several different methods used to approximate the solutions to financial economics problems. We will use a very simple model so that the analytical solution remains tractable and we will be able to examine how closely the various techniques can approximate the true solution. Specifically, we will apply perturbation and projection methods to solve for the price-dividend ratio as a function of the growth rate of dividends in a model that has an exact form solution. *Perturbation methods* use the Implicit Function Theorem and Taylor's Theorem to expand the unknown price-dividend function about the deterministic steady state equilibrium. *Projection methods* approximate the solution as a linear combination of some basis functions. The problem then reduces to finding coefficients to make the fit as close as possible.

Finally, a section on the newly emerging field of *Agent-Based Computational Economics (ACE)* will be presented. This approach is drastically different from previous computational solution methods. Instead of imposing equilibrium conditions on a market, as is the norm, we allow agents to evolve and learn over time, trade with one another and build rules upon which to base trading decisions. The hypothesis (known as the *Marimon-Sargent Hypothesis*) is that these agents will eventually learn the equilibrium that previous models simply imposed up front.

The paper is organized as follows. Section 2 will describe the Lucas asset pricing model in some detail. The purpose of this section is to provide the reader unfamiliar with financial economics with the necessary background for the rest of the paper. Section 3 presents and solves the model of interest (henceforth referred to as the Burnside model). Section 4 will describe the perturbation method and proceed to approximate the solution to the Burnside model. Section 5 will provide a fairly general description of the projection method and then use a Galerkin approach to solve the Burnside model. Finally, section 6 will introduce ACE models. Included in this section will be a motivation for studying these models and a brief introduction to how to begin building them.

All computations throughout the paper were done using the free statistical package R.

2 The Lucas Asset Pricing Model

Consider an economy with N identical, infinitely lived agents and a single consumption good. Suppose, for concreteness, that the consumption good is apples. Suppose there are M trees that produce identical apples each period. The production of apples is stochastic and completely exogenous to the problem, that is, there is no way to influence a tree's production and it takes no resources from the economy to produce apples. The world these agent's live in is a pure exchange economy. This means that there is no money. In each period each agent will have a supply of apples. He may either eat them during that period (consumption) or he may use them to buy the right to some portion of a tree's production next period (stock). The apples produced by the tree should be thought of as a dividend paid by a stock. There is no savings in this world, the fruit will spoil at the end of each day and may not be consumed tomorrow. Thus, total consumption by the economy in any period may not exceed the sum of all apples produced by all trees.

The agents in the model have preferences. These preferences will be represented by a *utility function*. A utility function will take, as input, the agent's consumption and return a numerical value to indicate the level of happiness achieved from that amount of consumption. If an agent consumes nothing they die and thus will never choose this. In general, preferences can (and do) differ among agents. In our model, however, we will assume that all agents have the same utility function. In general, a utility function can be any function that is increasing and concave. Increasing implies that agents always prefer to consume more if they can. The concavity assumption implies that the increase in utility from consuming more decreases as consumption increases (the last bite is never as satisfying as the first). An easy way to think of this is to consider two people, A and B. Suppose A has \$10 and B has \$1000. Both would be happier with one more dollar, but the increase in happiness will be much greater for A than it would for B since he does not have much money to begin with. The assumption of increasing utility functions implies that, at equilibrium, all apples will be consumed every period (since agents can always increase utility by consuming more, and they will do so as long as they can).

We will now formalize the above description of our model economy. Assume for simplicity that there is only one tree (stock), that is $M = 1$, and that that there is a single, infinitely divisible share outstanding for this tree. The agent's consumption and holdings in the stock during period t will be denoted c_t and s_t, respectively. The agents all have the same utility function, $u \in C^2$, such that $u' > 0$ and $u'' < 0$. This utility function is the same in each period. Stock prices are competitively determined after dividends are declared and will be denoted p_t. In addition, we will add a risk free bond to this world. At first this seems to contradict the assumption made about no savings being permitted in this world. It turns out, however, that no agent will ever hold this bond. Its only purpose will be to discuss the risk free rate in a meaningful way. The price of the zero coupon bond at time t paying one with certainty at time $t+1$ will be denoted q_t, where $q_t \leq 1$. Then the risk free rate will be given by $R^f = 1/q_t$.

The agent has three choices to make at each time t: how much to consume today (c_t), how much to invest in the stock for tomorrow (s_{t+1}) and how much to invest in the bond for tomorrow (b_{t+1}). Before proceeding, it will be necessary to define some additional notation. We will denote the agent's choice set by $A_t = \{c_t, s_{t+1}, b_{t+1}\}$. Also, we let $z_t = \{s_t, b_t\}$ denote the individual states at time t and $Z_t = \{S_t, B_t, d_t\}$ denote the aggregate states at time t, where S_t and B_t represent the aggregate holdings of the stock and bond, respectively. Note that d_t should be thought of as the per capita dividend paid. We will define the agent's value function to be

$$v(z_t, Z_t) = \max_{A_t} E_0 \left\{ \sum_{t=0}^{\infty} \beta^t u(c_t) \right\}. \tag{2.1}$$

The agent's objective will be to maximize this function subject to a budget constraint. Since there is no savings and only one good, the budget constraint is simple. The agent's consumption and investment may not, in any period, exceed his gross return from last period, or

$$c_t + p_t s_{t+1} + q_t b_{t+1} \leq s_t(p_t + d_t) + b_t \quad \forall t, \tag{2.2}$$

where s_0 and b_0 are known. The agent is born with a certain amount of the stock and the bond and must plan the rest of his infinite life based on that. In addition to (2.2), we also impose the nonnegativity constraint on consumption, $c_t \geq 0$, for all t. It is not possible to consume a negative amount. It is now time to define what we will be looking for in this model.

Definition 2.1. *A rational expectations competitive equilibrium is a set of demand functions* $\{c^i(z_t, Z_t), s^i(z_t, Z_t), b^i(z_t, Z_t)\}_{i=1}^N$ *and a pricing function* $p(Z_t)$ *such that the following conditions hold.*

1. *These functions solve each agent's optimization problem (i.e. the first order conditions hold).*
2. *All markets clear:*

   (i) $B_t = \sum_{i=1}^{N} b_t^i = 0, \quad \forall t$

   (ii) $S_t = \sum_{i=1}^{N} s_t^i = 1, \quad \forall t$

   (iii) $\sum_{i=1}^{N} (c_t^i + p_t s_{t+1}^i + q_t b_{t+1}^i) = \sum_{i=1}^{N} (s_t^i (p_t + d_t) + b_t^i), \quad \forall t$

We will briefly discuss the three conditions in 2. The first condition makes sure that the bond market clears. Since all agents are identical this will imply that at equilibrium no agent will hold any bond. This is simple to see. If agent i has some holding of the bond, then all agents must hold the same amount. The only way this will sum to zero is if all agents hold zero. The second condition implies a similar result, but for the stock. In this case, the stock has outstanding exactly one share. Thus, at equilibrium, all agents must each hold exactly $1/N$ of it. The final condition states that all the budget constraints balance, that is, there is nothing left over. This condition, along with the first two, implies that $c_t^i = d_t$ for all i and t. At equilibrium, we know that $B_t = 0$ and $S_t = 1$. Thus, the aggregate state vector reduces to just d_t and we can write the pricing function as $p(d_t)$. This function will be the main focal point of this model. Note that all the demand functions have a superscript i so they can vary over agents. The pricing function, however, must be the same for all agents since it is the price the market sets.

To simplify notation, let us drop the subscript on the period t variables and represent period $t+1$ by a $'$. For example, $s_t = s$ and $s_{t+1} = s'$. To derive the agent's first order conditions, note that the problem can be written recursively (see [11], Ch. 9):

$$v(z, Z) = \max_{A} \left[u(c) + \beta E v(z', Z') \right], \tag{2.3}$$

subject to the budget constraint (2.2). Let λ_t be the Langrange multiplier for each t and denote $\lambda = \lambda_t$. Then we can write (2.3) as

$$v(z, Z) = \max_{\{c, s', b', \lambda\}} \left\{ u(c) + \beta E v(z', Z') + \lambda [s(p + d) + b - c - ps' - qb'] \right\}. \tag{2.4}$$

Since we have imposed the nonnegativity constraint on consumption, maximizing with respect to c involves complimentary slackness conditions. We have not imposed any bounds on s or b (that is, there is no short sale constraints) so s' and b' will be optimized as unconstrained variables. It is customary in these problems to impose $\lambda \geq 0$. This is since the multiplier can be interpreted as a shadow price. Thus, taking derivatives with respect to the choice variables and the Lagrange multiplier, we have the following first order conditions (and complementary slackness conditions, if applicable):

$$\begin{aligned} c: &\quad u'(c)-\lambda \le 0, \quad c \ge 0 \quad \text{and} \quad c(u'(c)-\lambda)=0 && (2.5)\\ s': &\quad \beta E[v_{s'}(z',Z')] = \lambda p && (2.6)\\ b': &\quad \beta E[v_{b'}(z',Z')] = \lambda && (2.7)\\ \lambda: &\quad s(p+d)+b-c-ps'-qb' \le 0, \quad \lambda \ge 0 \\ &\quad \text{and} \quad \lambda[s(p+d)+b-c-ps'-qb'] = 0 && (2.8) \end{aligned}$$

The first inequality in (2.5) will bind in all cases. If not, then the agent is choosing $c = 0$ which means he will die. No agent will ever choose this. Also, the first inequality in (2.8) will bind, for if it does not, then $\lambda = 0$ which would imply that $u'(c) = 0$, which contradicts our assumption about u. These four conditions are the first order conditions for the maximization problem. They are not immediately useful, however, since we do not know the derivatives of v. In order to find $v_{s'}$ and $v_{b'}$, we differentiate (2.4) with respect to s and b:[1]

$$\frac{\partial}{\partial s} v(z,Z) = \lambda(p+d), \tag{2.9}$$

$$\frac{\partial}{\partial b} v(z,Z) = \lambda. \tag{2.10}$$

Since these equations will hold for all t, we can update the time from t to $t+1$ to get

$$v_{s'}(z',Z') = \lambda'(p'+d'), \tag{2.11}$$

and

$$v_{b'}(z',Z') = \lambda' \tag{2.12}$$

Combining this with (2.5), (2.6) and (2.7), we get the Euler equations:

$$u'(c_t)p_t = \beta E_t[u'(c_{t+1})(p_{t+1}+d_{t+1})], \tag{2.13}$$

and

$$u'(c_t)q_t = \beta E_t[u'(c_{t+1})]. \tag{2.14}$$

The Euler equations have an intuitive economic interpretation. Consider the left hand side of (2.13). This represents the loss in utility to the investor if he chooses to purchase one more unit of the risky asset at the price p_t. On the other hand, the right hand side represents the gain in (expected, discounted) utility from the extra payoff tomorrow associated with having the additional unit of the stock. The agent will continue to buy or sell the asset until (2.13) holds with equality. To see why, consider the case that the gain in utility tomorrow exceeds the loss today. Then it makes sense to buy more of the asset, since the agent gains more than he loses by doing so. The second Euler equation, (2.14), does exactly the same thing as (2.13), except that it is determining the price of the risk free asset, with final price of one and no dividend payment. This equation is actually what we would use to compute the risk free rate. Since the bond is a discount bond, the gross rate of return will be $1/q_t$. Thus, we can define the risk free rate, R^f, to be

$$R^f = \frac{1}{E_t(m_{t+1})}, \tag{2.15}$$

[1] This is called the *envelope condition* or *Benveniste-Scheinkman condition*. It can be derived by computing the total derivatives of v and using the first order conditions (2.5)-(2.8).

where we have defined $m_{t+1} = \beta \frac{u'(c_{t+1})}{u'(c_t)}$. This m is called the *stochastic discount factor* and may be used to price any asset in the market. Consider any asset (risky or not) with payoff x_{t+1}. Then today's price of the asset is $p_t = E_t(m_{t+1}x_{t+1})$. See [4] for a full discussion of discount factors.

We saw above that in each time period there must be a zero net holding of the bond and all of the share of stock must be owned. Also, since all agents are identical this implies that each agent always has the same portfolio, namely, $(1/N)^{th}$ of the share of stock and zero bond. Thus, there is no trading in this model. Agents simply hold their portfolio and consume their dividends each period. It is important to note, however, that no trading does not imply that the price doesn't move. In fact, the price does move and will be our main focus. Once dividends are declared *all* agents will try to trade based on their new information. The problem is that since they are all identical, they will simply bid the market price to a new level where none of them wants to trade. The result is a movement in price without any trading volume.

2.1 An Example

We will conclude the section with an example that will be used throughout the paper. Suppose the agents have the the special *power form* of the utility function:

$$u(c_t) = \frac{c_t^{1-\gamma} - 1}{1-\gamma}, \tag{2.16}$$

where $\gamma > 0$ is a fixed parameter.[2] γ is a measure of the level of risk aversion for the agent. The larger γ is, the more risk averse the agent is. In the limiting case $\gamma \to 0$, the agent is risk neutral. The Euler equations for an agent with this utility function are:

$$p_t c_t^{-\gamma} = \beta E_t[c_{t+1}^{-\gamma}(p_{t+1} + d_{t+1})] \tag{2.17}$$

and

$$q_t c_t^{-\gamma} = \beta E_t[c_{t+1}^{-\gamma}]. \tag{2.18}$$

Thus, the stochastic discount factor is

$$m_{t+1} = \beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma}. \tag{2.19}$$

Finally, the risk free rate will be

$$R^f = \frac{1}{E_t(m_{t+1})} = \frac{1}{\beta} E_t \left(\frac{c_{t+1}}{c_t} \right)^{\gamma}. \tag{2.20}$$

3 The Burnside Model

In this section we will, through the use of a carefully chosen process for dividends, be able to specify an exact solution for the Lucas asset pricing model. Assume that there is a single representative

[2] This differs from the form commonly used in the literature. Usually, this utility function is written as $u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}$ and is then accompanied by a statement that the limit as $\gamma \to 1$ is $\ln c_t$. This, in fact, is wrong. The -1 is required in the numerator so that L'Hospital's rule applies. This "error" is made since we are usually only concerned with marginal utility and so the addition of a constant term is not relevant.

agent[3] whose utility function is power utility. That is, the agent wishes to maximize

$$E_0 \left\{ \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma} - 1}{1-\gamma} \right\}, \tag{3.1}$$

subject to

$$c_t + s_{t+1} p_t \leq (d_t + p_t) s_t. \tag{3.2}$$

Since there is only one agent in this model, he will hold the single share at equilibrium. This will imply that, at equilibrium, $c_t = d_t$ for all t. Using this fact and (2.17), we can write the first order conditions as

$$p_t = \beta E_t \left\{ \left(\frac{d_{t+1}}{d_t} \right)^{-\gamma} (p_{t+1} + d_{t+1}) \right\}. \tag{3.3}$$

Dividing both sides by d_t and defining the price-dividend ratio to be $v_t = \frac{p_t}{d_t}$, we get

$$v_t = \beta E_t \left\{ \left(\frac{d_{t+1}}{d_t} \right)^{1-\gamma} (v_{t+1} + 1). \right\} \tag{3.4}$$

We will suppose that dividends grow according to $d_t = e^{x_t} d_{t-1}$, where $x_t = (1-\rho)\mu + \rho x_{t-1} + \varepsilon_t$ with ε_t being i.i.d. $N(0, \sigma^2)$ and $|\rho| < 1$. Here ρ represents the persistence of the growth rate. For values of ρ close to one, the growth process will be highly persistent. This means that if the growth rate were high in a given period, it will likely be high in the next also. (As a special case, consider $\rho = 0$. This corresponds to the independent, identically distributed case.) Then (3.4) becomes

$$v_t = \beta E_t \left[e^{\theta x_{t+1}} (v_{t+1} + 1) \right], \tag{3.5}$$

where we have let $\theta = 1 - \gamma$. Burnside (1998) shows (after some algebra) that the solution, v_t, may be expressed as

$$v_t = \sum_{j=1}^{\infty} \beta^j \exp[a_j + b_j (x_t - \mu)], \tag{3.6}$$

where

$$a_j = \theta j \mu + \frac{1}{2} \theta^2 \frac{\sigma^2}{(1-\rho)^2} \left[j - 2 \frac{\rho}{1-\rho} (1 - \rho^j) + \rho^2 \frac{1-\rho^{2j}}{1-\rho^2} \right]$$

and

$$b_j = \theta \frac{\rho}{1-\rho} (1 - \rho^j).$$

A natural concern at this point is convergence of the above infinite series. The following proposition addresses this issue.

Proposition 3.1. *The series defined in (3.6) converges if and only if*

$$r \equiv \beta \exp \left[\theta \mu + \frac{1}{2} \theta^2 \frac{\sigma^2}{(1-\rho)^2} \right] < 1.$$

[3] This is a common simplifying assumption in the economics literature. The reason for doing so is that we have N identical investors so we need only look at one of them. This can be confusing too, however. It is not natural to think of an economy with only a single agent. To avoid confusion, always keep in mind that the agent could be thought of as many agents, all identical.

Proof. The proof is straightforward. Let $z_j = \beta^j \exp[a_j + b_j(x_t - \mu)]$. Then $v_t = \sum_{j=1}^{\infty} z_j$. Next, we will compute $\lim_{j\to\infty} \left|\frac{z_{j+1}}{z_j}\right|$. We need the following first:

$$\begin{aligned}
a_{j+1} - a_j &= \theta(j+1)\mu + \frac{1}{2}\theta^2 \frac{\sigma^2}{(1-\rho)^2}\left[(j+1) - \frac{2\rho}{1-\rho}(1-\rho^{j+1}) + \rho^2\frac{1-\rho^{2j+2}}{1-\rho^2}\right] \\
&\quad -\theta j\mu - \frac{1}{2}\theta^2 \frac{\sigma^2}{(1-\rho)^2}\left[j - \frac{2\rho}{1-\rho}(1-\rho^{j}) + \rho^2\frac{1-\rho^{j+1}}{1-\rho^2}\right] \\
&= \theta\mu + \frac{1}{2}\theta^2 \frac{\sigma^2}{(1-\rho)^2}\left[1 - \frac{2\rho}{1-\rho}(1-\rho^{j+1} - 1 + \rho^j) + \frac{\rho^2}{1-\rho^2}(1-\rho^{2j+2} - 1 + \rho^{2j}\right] \\
&= \theta\mu + \frac{1}{2}\theta^2 \frac{\sigma^2}{(1-\rho)^2}\left[1 - \frac{2\rho}{1-\rho}\rho^j(1-\rho) + \frac{\rho^2}{1-\rho^2}\rho^{2j}(1-\rho^2)\right] \\
&= \theta\mu + \frac{1}{2}\theta^2 \frac{\sigma^2}{(1-\rho)^2}\left[1 - 2\rho^{j+1} + \rho^{2j+2}\right]
\end{aligned}$$

Also, it is easy to see that $b_{j+1} - b_j = \theta\rho^{j+1}$. Thus we have

$$\begin{aligned}
\lim_{j\to\infty}\left|\frac{z_{j+1}}{z_j}\right| &= \lim_{j\to\infty} \beta \exp[(a_{j+1} - a_j) + (b_{j+1} - b_j)(x_t - \mu)] \\
&= \beta \exp\left\{\lim_{j\to\infty}\left[\theta\mu + \frac{1}{2}\theta^2\frac{\sigma^2}{(1-\rho)^2}\left(1 - 2\rho^{j+1} + \rho^{2j+2}\right) + \theta\rho^{j+1}\right]\right\} \\
&= \beta \exp\left[\theta\mu + \frac{1}{2}\theta^2\frac{\sigma^2}{(1-\rho)^2}\right]
\end{aligned}$$

where the last equality follows from the fact that $|\rho| < 1$. By the ratio test, if $r < 1$ the series converges and if $r > 1$ the series diverges. Consider the case where $r = 1$. Then the ratio test is inconclusive. However, if $r = 1$, then (after some algebra)

$$z_j = \exp\left\{\frac{1}{2}\theta^2\frac{\sigma^2}{(1-\rho)^2}\left[\rho^2\frac{1-\rho^{2j}}{1-\rho^2} - 2\frac{\rho}{1-\rho}(1-\rho^j)\right] + \theta\frac{\rho}{1-\rho}(1-\rho^j)(x_t - \mu)\right\}.$$

It is easy to check that $\lim_{j\to\infty} z_j \neq 0$ so the series will not converge when $r = 1$. This completes the proof. □

Proposition 3.1 is more important than it may seem at first glance. Our main usage of the Burnside model will be to check the accuracy of some methods of approximation. It will be important to make sure that the parameters we choose will lead to an exact solution that makes sense.

Table 3.1: The price dividend ratios for various cases and growth rates

	x_t				
Case	-0.2	-0.1	0	0.1	0.2
Benchmark	11.99	12.21	12.44	12.67	12.91
$\beta = 0.5$	3.40	3.46	3.53	3.59	3.66
$\beta = 0.99$	26.33	26.82	27.31	27.82	28.34
$\theta = -10$	3.83	4.34	4.91	5.57	6.30
$\theta = -5$	6.36	6.77	7.20	7.66	8.15
$\theta = 0$	19	19	19	19	19
$\theta = 0.5$	23.50	23.35	23.21	23.07	22.93
$\sigma = 0.001$	11.82	12.04	12.26	12.49	12.72
$\sigma = 0.1$	13.38	13.63	13.89	14.14	14.41
$\rho = 0$	12.53	12.53	12.53	12.53	12.53
$\rho = 0.5$	17.87	15.52	13.49	11.72	10.20
$\rho = 0.7$	28.83	21.08	15.48	11.42	8.47

Table 3.1 shows the price dividend ratios for many different cases. The benchmark values are (based on historical data) : $\mu = 0.0179$, $\sigma = 0.0348$, $\beta = 0.95$, $\theta = -1.5$ and $\rho = -0.139$. These values will be used throughout the paper. In each case all values but the one indicated are maintained at their benchmark levels. It is worthwhile to examine some of the individual cases presented in Table 3.1 in order to build some intuition about what is happening in this model. Typically, when we do comparative statics in financial economics we discuss price levels as opposed to price-dividend ratios. Therefore, in the discussions below it may be helpful to think of the dividend as fixed and instead of considering $v = \frac{p}{d}$, think of $p = d \cdot v$. The benchmark case is shown in Figure 3.1.

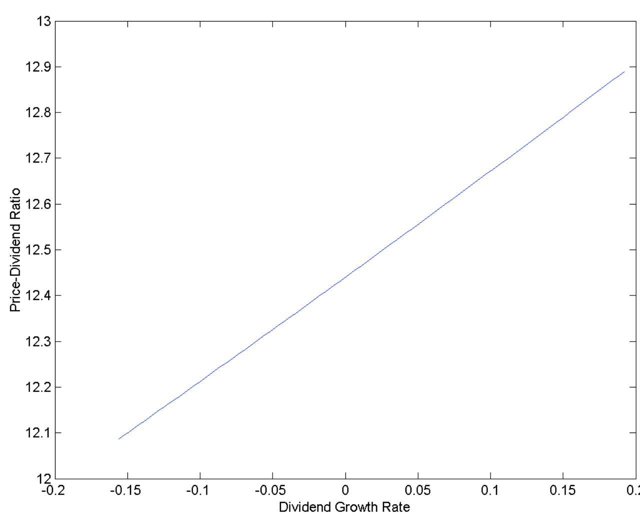


Figure 3.1: A plot of the exact solution in with all parameters at their benchmark levels.

First, consider the cases in which β is varied. If an agent has a particularly low β, then this means that he values money today much more than money tomorrow. Specifically, for an agent with $\beta = 0.8$, one dollar tomorrow (with certainty) is only worth 0.80 today. Since buying a stock essentially amounts to purchasing an infinite (albeit not certain) stream of payments, it only makes sense that the agent with the lower discount rate would not be willing to pay as much for it - he would rather have his money today than wait. A similar argument can be made for the agent with $\beta = 0.99$.

Second, we examine the various θ cases. Recall that the agent's utility function was $u(c) = \frac{c^{1-\gamma}-1}{1-\gamma}$ or $u(c) = \frac{c^{\theta}-1}{\theta}$. Thus, θ is the agent's risk aversion parameter, the lower the value of θ, the more risk averse is the agent.[4] Before proceeding we will need the following definition.

Definition 3.2. *The gross return from period t to $t+1$, denoted as R_{t+1}, is defined as*

$$R_{t+1} = \frac{p_{t+1} + d_{t+1}}{p_t}.$$

This can be thought of as the total payoff of the stock during period $t+1$ divided by what the agent paid for it one period earlier. We can also define the net return as $r_{t+1} = R_{t+1} - 1$.

It is not immediately obvious why varying θ gives the results it does. Suppose we fix the dividend in period t so that $d_t = d^*$. Then a high price-dividend ratio implies a high price and a low price-dividend ratio implies a low price. We now turn to the agent's preferences. The more risk averse an agent is, the higher return he will have to expect in order to invest in a risky asset. This will correspond to the case where p_t is lower. In other words, for a fixed dividend level, the most risk averse agent will pay the least for the stock and the least risk averse agent will pay the most. This explains why the price-dividend ratio is increasing in θ. The case when $\theta = -30$ is shown in Figure 3.2 (this is an unrealistically small value for θ but it illustrates the potential shape and slope of the graph).

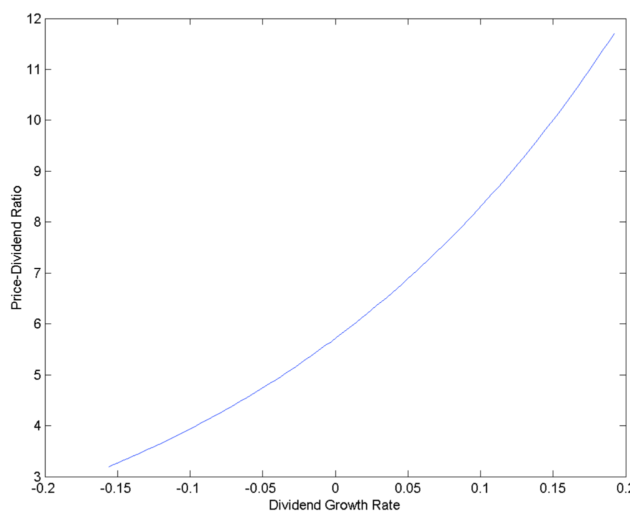


Figure 3.2: A plot of the exact solution for the case $\theta = -30$.

We would also like to explain why the price-divided ratio is increasing or decreasing in the growth rate for each θ case. Let us begin be rewriting (3.3) as

$$p_t = \beta E_t \left[\left(\frac{c_t}{c_{t+1}} \right)^{\gamma} (p_{t+1} + d_{t+1}) \right]. \tag{3.7}$$

[4]Actually, it makes more sense to talk about γ as the risk aversion parameter since a higher γ means a more risk averse agent.

It is slightly easier to visualize if we return to the previous γ notation, where $\theta = 1 - \gamma$. Now, suppose that today's growth rate is negative. Since all parameters except θ (γ) are maintained at their benchmark levels, we will have $\rho < 0$. This implies that tomorrow's growth rate will be positive which implies that tomorrow's consumption will be higher than today's. Thus, the ratio of consumption today versus tomorrow is less than 1. Now, suppose $\gamma < 1$ (which corresponds to $\theta > 0$), say $\gamma = \frac{1}{2}$. Taking a square root of something less than one results in a larger number. This leads to a larger price today and thus a larger price-dividend today, since d_t is known. As x_t increases, the ratio of consumption will decrease and thus, so will the price-dividend ratio. A similar argument can be made for the case $\gamma > 1$. In the case $\gamma = 1$, the problem becomes trivial since in (3.6) both the a_j and b_j will be 0. The solution then becomes $\sum_{j=1}^{\infty} \beta^j = \frac{\beta}{1-\beta}$.

Finally, we consider changes in the variable ρ. It turns out that in approximating the solution to this model, large values of ρ are the most troublesome. This is due to the fact that for large values of ρ, the price-dividend curve has a lot of curvature, whereas for lower values it is close to linear. The easiest case to explain is when $\rho = 0$. In this case we have $x_t = \mu + \varepsilon_t$. Thus dividend growth is i.i.d. over time. Therefore, the agent's valuation of the price-dividend ratio (and thus the price) should have nothing to do with the current dividend (nor the current x_t value). This is reflected in the constant value for all x_t. For $\rho < 0$ the price-dividend is increasing with respect to x_t. If x_t is negative today, then this will increase the mean dividend growth rate for tomorrow and thus we expect higher dividends. This causes the the p-d ratio to be smaller. On the other hand, if x_t is positive today then the mean tomorrow should be less and therefore dividends will be less, and the p-d ratio will be larger. A similar argument can be made for $\rho > 0$. A plot of the $\rho = 0.8$ case is given in Figure 3.3.

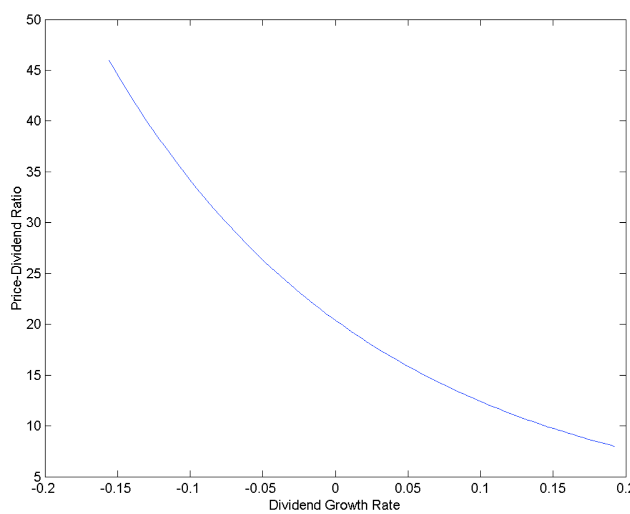


Figure 3.3: A plot of the exact solution for the case $\rho = 0.8$.

The plots given above are restricted to a small portion of the real line. As can be seen from (3.6), the plot of v on the whole real line will look like an exponential function. Extreme values of the growth rate are highly unlikely (since the shocks are normally distributed about 0) so we restrict our attention to x_t to five standard deviations about its long run mean.

4 The Perturbation Method

The first method of approximation that we will discuss is the method of perturbations. We follow the lead of Collard and Juillard (2001). Before we begin we will need the following preliminary result from advanced calculus.

Theorem 4.1 (Implicit Function Theorem)**.** *Suppose $H : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is C^k and there exists a point (x_0, y_0) such that $H(x_0, y_0) = 0$ and $H_y(x_0, y_0)$ is not singular. Then there exists a unique function $h : \mathbb{R}^n \to \mathbb{R}^m$ in C^k such that $y_0 = h(x_0)$ and $H(x, h(x)) = 0$ for x near x_0. Moreover, the derivatives of h may be computed by implicit differentiation of the identity $H(x, h(x)) = 0$.*

4.1 The Method

We begin with a fairly informal description of the method as in [7]. Below, we will apply this idea to the Burnside model and compare our approximations with the exact solution given in (3.6).

Suppose $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a differentiable, known function and that we wish to solve

$$f(x, \varepsilon) = 0 \tag{4.1}$$

for x, where ε is a parameter. Assume that for each ε, the above equation has at least one solution. We will search for a function, $x = x(\varepsilon)$ such that $f(x(\varepsilon), \varepsilon) = 0$. The Implicit Function Theorem tells us that such a function exists, is unique and is differentiable (as many times as f is), as long as $f_x \neq 0$.

The perturbation method is essentially nothing more than a Taylor series approximation, of whatever order we wish. Assume $x(0)$ is known. Suppose we wish to construct a linear approximation for $x(\varepsilon)$ about $\varepsilon = 0$:

$$x^L(\varepsilon) = x(0) + x'(0)\varepsilon. \tag{4.2}$$

We have assumed knowledge of $x(0)$ but not of $x'(0)$. To compute this we will apply implicit differentiation to (4.1). This gives

$$f_x(x(\varepsilon), \varepsilon)x'(\varepsilon) + f_\varepsilon(x(\varepsilon), \varepsilon) = 0. \tag{4.3}$$

Plugging in $\varepsilon = 0$ and solving for $x'(0)$ we get

$$x'(0) = -\frac{f_\varepsilon(x(0), 0)}{f_x(x(0), 0)}, \tag{4.4}$$

where all the terms on the right are known. Thus we have

$$x^L(\varepsilon) = x(0) - \frac{f_\varepsilon(x(0), 0)}{f_x(x(0), 0)}\varepsilon. \tag{4.5}$$

We could continue in this way. For example, let us define

$$x^Q(\varepsilon) = x(0) + x'(0)\varepsilon + \frac{1}{2}x''(0)\varepsilon^2, \tag{4.6}$$

to be the second order (quadratic) Taylor approximation to $x(\varepsilon)$. Then, applying implicit differentiation to (4.1) twice gives

$$f_x x'' + f_{xx}(x')^2 + 2f_{x\varepsilon}x' + f_{\varepsilon\varepsilon} = 0. \tag{4.7}$$

Plugging in $\varepsilon = 0$ and solving for $x''(0)$ we get

$$x''(0) = -\frac{f_{xx}(x(0),0)(x'(0))^2 + 2f_{x\varepsilon}(x(0),0)x'(0) + f_{\varepsilon\varepsilon}(x(0),0)}{f_x(x(0),0)}, \tag{4.8}$$

which completes the second order approximation. This process can be continued to any order desired without having to assume anything further about our knowledge of x, as long as f has the required derivatives.

To check how close the approximation is to the true solution, one can plug the Taylor approximation into (4.1) and see how close it is to zero. For example, suppose we wish to use a quadratic approximation to estimate $x(1)$. Then we could obtain the approximation as above and check the residual $r^Q \equiv f(x^Q(1), 1)$. If this is small, then we may decide that our quadratic approximation is sufficient. Otherwise we may have to try a higher order approximation. What constitutes "small" will depend on the problem.

4.2 Application to the Burnside Model

Let us write the Burnside model as

$$x_t = h(x_{t-1}, \varepsilon_t), \tag{4.9}$$

$$v_t = E_t[g(v_{t+1}, x_{t+1})], \tag{4.10}$$

where $g(v,x) = \beta \exp(\theta x)(1+v)$ and $h(x,\varepsilon) = (1-\rho)\bar{x} + \rho x + \varepsilon$. We are seeking a function $f(\,\cdot\,)$ such that $v_t = f(x_t)$. Then (4.10) becomes

$$f(x_t) = E_t[g(f(h(x_t, \varepsilon_{t+1})), h(x_t, \varepsilon_{t+1}))] = E_t[G(x_t, \varepsilon_{t+1})]. \tag{4.11}$$

We now proceed in a similar way as in the case presented above, however, we will not explicitly solve for the coefficients in the Taylor expansion, at least not initially. The reason for this alternative approach is that there are many functions and many variables in (4.11) and it is not immediately obvious how to tackle the problem in the preceding framework. Instead, we will proceed as follows. Let $H(x_t, \varepsilon_{t+1}) = f(x_t) - E_t[G(x_t, \varepsilon_{t+1})]$. Then (4.11) amounts to solving $H(x_t, \varepsilon_{t+1}) = 0$. Note that in this case, unlike before, we are not interested in solving for $x_t = x_t(\varepsilon_{t+1})$. Next, we will expand H in a Taylor series about the deterministic steady state. Thus, before proceeding, we will need to establish for which values x^* and v^* the economy is in a deterministic steady state. Using (4.9) and (4.10) we obtain

$$x^* = \rho x^* + (1-\rho)\bar{x} \iff x^* = \bar{x}, \tag{4.12}$$

and

$$v^* = g(v^*, x^*) \iff v^* = \frac{\beta \exp(\theta \bar{x})}{1 - \beta \exp(\theta \bar{x})}. \tag{4.13}$$

Suppose, initially, we are interested in a first order approximation. Then expanding H about $(x^*, 0)$ gives

$$\begin{aligned} H(x_t, \varepsilon_{t+1}) &\approx H(x^*, 0) + H_x(x^*, 0)\hat{x}_t + H_\varepsilon(x^*, 0)\varepsilon_{t+1} \\ &= [f(x^*) - E_t(G(x^*, 0))] + [f_x(x^*)\hat{x}_t - E_t(G_x(x^*, 0))\hat{x}_t] \\ &\qquad + [f_\varepsilon(x^*)\varepsilon_{t+1} - E_t(G_\varepsilon(x^*, 0))\varepsilon_{t+1}] \\ &= f(x^*) + f_x(x^*)\hat{x}_t - E_t[G(x^*, 0) + G_x(x^*, 0)\hat{x}_t - G_\varepsilon(x^*, 0)\varepsilon_{t+1}] \\ &= f_0 + f_1\hat{x}_t - E_t[G_{0,0} + G_{1,0}\hat{x}_t + G_{0,1}\varepsilon_{t+1}] \\ &= f_0 + f_1\hat{x}_t - E_t[G_{0,0} + G_{1,0}\hat{x}_t], \end{aligned}$$

where we have used the fact that $E_t[\varepsilon_{t+1}] = 0, \forall t$, by definition. Note that we have introduced following notation:

$$f_k = \frac{d^k}{dx^k} f(x)\bigg|_{x=x^*}, \quad G_{i,j} = \frac{\partial^{i+j}}{\partial x^i \partial \varepsilon^j} G(x,\varepsilon)\bigg|_{\substack{x=x^* \\ \varepsilon=0}} \quad \text{and} \quad \hat{x}_t = x - x^*.$$

Also, note that we are expanding around $\varepsilon = 0$ so $\hat{\varepsilon} = \varepsilon$. Equating coefficients gives

$$f_0 = G_{0,0} = g(v^*, x^*) = v^* \quad \text{and} \quad f_1 = G_{1,0}$$

All that remains is to compute the partial derivative of G with respect to x. This is not difficult since all the functions involved are known except for f. Using the chain rule we get

$$\begin{aligned} G_{1,0} &= f_1 \frac{\partial g(v,x)}{\partial v}\bigg|_{\substack{v=v^* \\ x=x^*}} \frac{\partial h(x,\varepsilon)}{\partial x}\bigg|_{\substack{x=x^* \\ \varepsilon=0}} + \frac{\partial g(v,x)}{\partial x}\bigg|_{\substack{v=v^* \\ x=x^*}} \frac{\partial h(x,\varepsilon)}{\partial x}\bigg|_{\substack{x=x^* \\ \varepsilon=0}} & (4.14) \\ &= f_1 \rho\beta \exp(\theta\bar{x}) + \rho\theta\beta\exp(\theta\bar{x})(1+f_0) & (4.15) \end{aligned}$$

Thus, we get $f_1 = \dfrac{\rho\theta\beta\exp(\theta\bar{x})(1+v^*)}{1-\rho\beta\exp(\theta\bar{x})}$. Therefore, our first order approximation for f is

$$f(x_t) = \frac{\beta\exp(\theta\bar{x})}{(1-\beta\exp(\theta\bar{x}))} + \frac{\rho\theta\beta\exp(\theta\bar{x})(1+v^*)}{1-\rho\beta\exp(\theta\bar{x})}\hat{x}_t \tag{4.16}$$

At first, this seems like a decent accomplishment. We were able to approximate our unknown solution function in terms of what we know. There is a fairly severe problem however - the approximation in (4.16) does not involve σ in any way. Recall that our dividend growth rate was assumed to evolve according to $x_{t+1} = (1-\rho)\bar{x} + \rho x_t + \varepsilon_{t+1}$, where $\varepsilon \sim N(0, \sigma^2)$. Thus, the only source of risk in the dividend process is in the shocks to the growth rate. Since (4.16) does not depend on σ, the solution will be the same regardless of the risk.

The above strategy can be extended to any order of expansion we wish. Using higher orders of approximation will allow higher moments of the distribution of ε to be taken into account. This should greatly improve the accuracy of the approximation. For example, the second order approximation will be computed as

$$f_0 + f_1\hat{x}_t + \frac{1}{2}f_2\hat{x}_t^2 = E_t\left[G_{0,0} + G_{1,0}\hat{x}_t + G_{0,1}\varepsilon_{t+1} + \frac{1}{2}G_{2,0}\hat{x}_t^2 + \frac{1}{2}G_{0,2}\varepsilon_{t+1}^2 + G_{1,1}\hat{x}_t\varepsilon_{t+1}\right]. \tag{4.17}$$

Given the distribution for ε we have $E_t[\varepsilon_{t+1}] = 0$ and $E_t[\varepsilon_{t+1}^2] = \sigma^2$. After taking expectations, the above may be simplified to

$$f_0 + f_1\hat{x}_t + \frac{1}{2}f_2\hat{x}_t^2 = G_{0,0} + G_{1,0}\hat{x}_t + \frac{1}{2}G_{2,0}\hat{x}_t^2 + \frac{\sigma^2}{2}G_{0,2}. \tag{4.18}$$

Equating coefficients gives

$$f_0 = G_{0,0} + \frac{\sigma^2}{2}G_{0,2}, \quad f_1 = G_{1,0} \quad \text{and} \quad f_2 = G_{2,0}.$$

Computing the partial derivatives will result in the following linear system:

$$\begin{aligned} f_0 &= \beta\exp(\theta\bar{x})\left[\left(1 + \frac{(\theta\sigma)^2}{2}\right)(1+f_0) + \frac{\sigma^2}{2}(2\theta f_1 + f_2)\right], & (4.19) \\ f_1 &= \rho\beta\exp(\theta\bar{x})(f_1 + \theta(1+f_0)), & (4.20) \\ f_2 &= \rho^2\beta\exp(\theta\bar{x})(\theta^2(1+f_0) + 2\theta f_1 + f_2). & (4.21) \end{aligned}$$

It is clear that these coefficients now depend on the variance of the shocks to the growth rate. This is a step in the right direction. In fact, we would like the approximation to take into account higher moments as well, such as skewness and kurtosis which are known to be important factors to consider in asset pricing models.

The following system of equations will give the n^{th} order approximation (see [5] for a derivation):

$$f_0 = \beta \exp(\theta \bar{x}) \sum_{j=0}^{n} \frac{1}{j!} \left[\theta^j + \sum_{l=0}^{j} \binom{j}{l} \theta^{j-l} f_l \right] \mu_j, \tag{4.22}$$

$$\vdots$$

$$\frac{1}{k!} f_k = \sum_{j=0}^{n-k} \frac{1}{(j+k)!} \binom{k}{j} \beta \exp(\theta \bar{x}) \rho^k \left[\theta^{j+k} + \sum_{l=0}^{j+k} \binom{j+k}{l} \theta^{j+k-l} f_l \right] \mu_j, \tag{4.23}$$

$$\vdots$$

$$f_n = \beta \exp(\theta \bar{x}) \rho^n \left[\theta^n + \sum_{l=0}^{n} \binom{n}{l} \theta^{n-l} f_l \right], \tag{4.24}$$

where $\mu_j = E_t(\hat{\varepsilon}_{t+1}^j)$.

4.3 Error Analysis

Since the model has a closed form solution, error analysis is quite easy to do. We will simply compute the true and approximate solutions to the model at several different points and evaluate the closeness of the two solutions. The following two formulas will be used in measuring this closeness:

$$E_1 = 100 \times \frac{1}{N} \sum_{t=1}^{N} \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad \text{and} \quad E_\infty = 100 \times \max \left\{ \left| \frac{y_t - \hat{y}_t}{y_t} \right| \right\},$$

where y_t is the true solution and $\hat{y}_t$ is the approximate solution. The N points were picked to be uniformly spaced in the interval $[\bar{x} - 5\sigma, \bar{x} + 5\sigma]$.

The benchmark values were chosen to be the same as those given on page 8. The series defined in (3.6) was truncated after 800 terms and $N = 20$ was chosen as the number of points at which to compute the error.

Table 4.1: Average and maximum relative (percentage) errors for the perturbation method

	Order 1		Order 2		Order 4	
Case	E_1	E_∞	E_1	E_∞	E_1	E_∞
Benchmark	1.44(0)	1.48(0)	1.79(-3)	3.94(-3)	1.63(-3)	3.51(-3)
$\beta = 0.5$	2.55(-1)	2.93(-1)	2.22(-3)	4.54(-3)	2.02(-3)	3.98(-3)
$\beta = 0.99$	2.94(0)	2.98(0)	1.83(-3)	4.26(-3)	1.68(-3)	3.88(-3)
$\theta = -10$	2.38(1)	2.53(1)	6.12(-1)	1.36(0)	5.49(-1)	1.40(0)
$\theta = -5$	9.31(0)	9.70(0)	7.09(-2)	1.60(-1)	6.42(-2)	1.53(-1)
$\theta = 0$	0	0	0	0	0	0
$\theta = 0.5$	2.87(-1)	2.90(-1)	6.45(-5)	1.39(-4)	5.87(-5)	1.21(-4)
$\sigma = 0.001$	1.19(-3)	1.22(-3)	4.18(-8)	8.60(-8)	3.80(-8)	7.29(-8)
$\sigma = 0.1$	1.18(1)	1.21(1)	4.72(-2)	1.10(-1)	4.35(-2)	1.06(-1)
$\rho = 0$	1.81(0)	1.81(0)	1.25(-3)	1.25(-3)	5.70(-7)	5.70(-7)
$\rho = 0.5$	7.49(0)	9.65(0)	2.02(-1)	4.88(-1)	1.25(-1)	2.58(-1)
$\rho = 0.7$	2.06(1)	3.17(1)	1.99(0)	5.19(0)	1.16(0)	2.43(0)
$\rho = 0.8$	4.33(1)	6.97(1)	9.92(0)	2.23(1)	6.67(0)	1.35(1)

The first order approximations are typically very poor. This is mainly due to the fact that, as previously discussed, the first order approximation ignores risk altogether. The exception here is the case $\sigma = 0.001$. In this case the volatility of the shocks is quite low and so ignoring it does not make a huge difference. Certainly as $\sigma \to 0$ we would expect approximations of all orders to converge to the true solution since all uncertainty would be removed and we would essentially be left with a bond. For most cases the accuracy improves quite nicely as the order is increased. The exception to this are the cases for larger ρ. In these cases the true solution has much more curvature in the region of interest than the benchmark and so we should expect to need higher order approximations. It is important to remember that the values in Table 4.1 are actually percentages. One could divide each by 100 to get two more decimals of accuracy. For comparison, all three orders are plotted along with the exact solution for the $\rho = 0.5$ case in Figure 4.1.

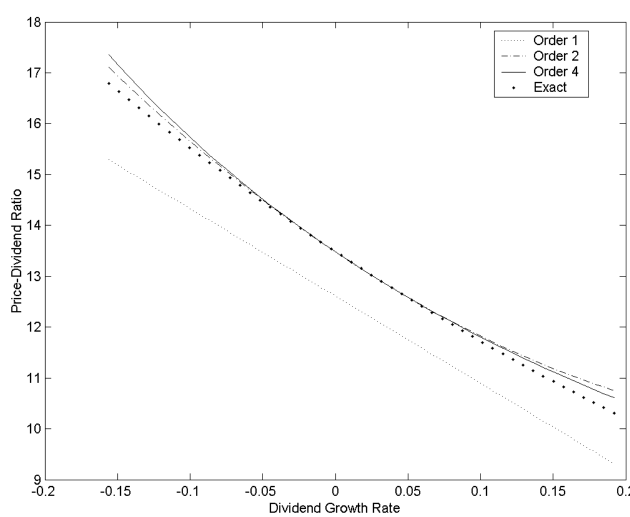


Figure 4.1: All three approximations plus the exact solution for the $\rho = 0.5$ case.

5 The Projection Method

5.1 The Method

The projection method differs from most numerical methods. Typically a numerical approximation to a problem will only offer a solution at a discrete number of points. To evaluate the solution at a point which is not in this set of discrete values one would need to use some interpolation rule. The way in which projection methods differ is that we are approximating a solution function which is valid at all points in the region of interest. We will begin with a fairly general description of the method, as in [7], then proceed to applying it to solve the Burnside model.

Suppose we wish to solve $\mathcal{N}(f) = 0$, where $\mathcal{N} : B_1 \to B_2$ is some continuous operator (linear or not) and B_1 and B_2 are inner product spaces of functions $f : D \subset \mathbb{R}^k \to \mathbb{R}^m$. The method will be comprised of several steps (as in [7]), each of which is outlined below.

1. Choose a basis in $\Phi = \{\varphi_i\}_{i=1}^{\infty}$ over B_1 and denote the norm as $\| \cdot \|$. Also, choose a basis $\Psi = \{\psi_i\}_{i=1}^{\infty}$ over B_2 and denote the inner product as $\langle \cdot , \cdot \rangle$. Typically these basis functions will increase in complexity as i increases. If the basis is polynomial functions then usually the i^{th} function will be the degree $i-1$ polynomial (with the first one being constant).

2. Choose how many basis functions from B_1 to use in the approximation. Define
$$\hat{f}(x) = \sum_{i=1}^{n} a_i \varphi_i(x),$$
and denote $\mathbf{a} = (a_1, a_2, \ldots, a_n)$. In general, it will not be possible to represent the actual solution as a finite linear combination of basis functions (that is, the only n that truly works is $n = \infty$). Also note that it is not usually going to be possible to pick the best value for n a priori. Instead, several approximations should be done until one is satisfied with the "closeness" of the approximation. This will be addressed in the next few steps.

3. In many cases it will be necessary to also construct an approximation, $\hat{\mathcal{N}}$, to the operator $\mathcal{N}$. This is required when it is not possible to numerically compute $\mathcal{N}(\hat{f})$. A typical example of this (and one that we will see later) is if $\mathcal{N}$ involves the integration of f. In this case one would need to use an appropriate quadrature rule and thus replace the integral operator $\mathcal{N}$ by a sum $\hat{\mathcal{N}}$.

4. Define the residual function
$$R(x; \mathbf{a}) = (\hat{\mathcal{N}}(\hat{f}(\,\cdot\,; \mathbf{a})))(x).$$
Compute $\|R(\,\cdot\,; \mathbf{a})\| = \langle R(\,\cdot\,; \mathbf{a}), R(\,\cdot\,; \mathbf{a})\rangle$ or choose l test functions in B_2, $p_i : D \to \mathbb{R}^m, i = 1, \ldots, l$, and for each guess of $\mathbf{a}$ compute $P_i(\,\cdot\,) = \langle R(\,\cdot\,; \mathbf{a}), p_i(\,\cdot\,)\rangle$. Examples of how this might be done will be given below.

5. Find $\mathbf{a} \in \mathbb{R}^n$ that either minimizes $\|R(\,\cdot\,; \mathbf{a})\|$ or solves $P(\mathbf{a}) = 0$. This will give us an approximate solution, $\hat{f}$, to the approximate problem, $\hat{\mathcal{N}}(\hat{f})$.

6. Finally, it needs to be verified that this $\hat{f}$ sufficiently approximates $\mathcal{N}(f)$. This can be done by computing the norm $\|\mathcal{N}(\hat{f})\|$ and/or the projections of $\mathcal{N}(\hat{f})$ against test functions not used already in finding $\hat{f}$ and seeing how close these values are to zero.

There are several ways in which one can actually solve for the vector $\mathbf{a}$ of coefficients. Most vary in degree of difficulty and accuracy. Some methods will be presented below. In all cases the overall goal is similar: make the approximation as close as possible. The methods differ in how they define what "close" is.

- One of the most familiar approaches to problems such as this, especially from an economics standpoint, is the *method of least squares*. This amounts to solving the following minimization problem:

$$\min_{\mathbf{a}} \langle R(x;\mathbf{a}), R(x;\mathbf{a})\rangle$$

- Perhaps the easiest and most intuitive method is the *method of collocation*. The idea behind the method is as follows: since our ultimate goal would be to have the residual be zero everywhere (which is equivalent to having the exact solution), let us make the residual zero at n places and hope that it doesn't deviate much from zero in between. A particularly easy implementation of this method would be to choose the n points to be uniformly spaced in the interval in question. This is called *uniform collocation*. The problem with this approach is that it is possible for a function to vary substantially from zero in between the nodes where it is chosen to be zero. A similar, yet much more reliable approach, is to choose the points to correspond to the Chebyshev nodes. For example if the domain in question were simply $[-1, 1]$, then we would need only find the roots of the n^{th} order Chebyshev polynomial and make those our collocation points. This method is called *Chebyshev collocation*. It can be shown that a smooth, well-behaved function which is zero at these Chebyshev nodes is close to zero for all $x \in [-1, 1]$ (see [7], Ch. 6 for details). For any other domain, $[a, b]$, it is straightforward to map $[a, b]$ into $[1, 1]$ and apply the same technique.

- The *subdomain method* involves splitting the domain, D, into n pieces (or subdomains) and ensuring that the approximation is (on average) good on each piece. Another way of stating this is that the residual function is, on average, zero on each subdomain, or

$$P_i(\mathbf{a}) = \langle R(x;\mathbf{a}), \mathbb{1}_{D_i}\rangle = 0, \quad \text{for each } i = 1, \ldots, n,$$

  where $D = D_1 \cup D_2 \cup \cdots \cup D_n$ and $\mathbb{1}_{D_i}$ is the indicator function of the set D_i.

- The *method of moments* insists that the residual be orthogonal to each of the first n monomials. That is,

$$P_i(\mathbf{a}) = \langle R(x;\mathbf{a}), x^{i-1}\rangle = 0 \quad \text{for each } i = 1, \ldots, n.$$

  This method turns out to be quite similar to the next.

- Finally, we will discuss the *Galerkin method*. In this case we insist that the residual be orthogonal to each of the basis functions we chose. That is,

$$P_i(\mathbf{a}) = \langle R(x;\mathbf{a}), \varphi_i(x)\rangle = 0 \quad \text{for each } i = 1, \ldots, n.$$

  In this case we need to make the assumption that $B_1 \subset B_2$ so that the inner products make sense. Notice that if we had chosen our basis to be $\{x^{i-1}\}_{i=1}^{\infty}$, then the method of moments and the Galerkin method coincide. This method works particularly well if the basis is chosen to be orthogonal. In that case, since R is defined in terms of the φ_i, the inner products become simplified due to the orthogonality of the basis functions (since the inner product of two distinct, orthogonal functions is zero). This is the method that will be applied to the Burnside problem below.

Thus far, we have discussed the projection method in a fairly general context and provided some ideas on how to impose the projection conditions. We have yet to discuss, however, how one would go about choosing a suitable basis. In general, any basis will work. However, there are some choices which will be obviously bad and others will be poor choices for more subtle reasons. In physics problems, it is typical to use trigonometric basis functions and apply Fourier analysis. This is a good choice for problems which are naturally periodic, ones involving waves for example. In economic applications, however, this is a very poor choice. Solutions to most economics problems are not periodic and to approximate a nonperiodic function with periodic basis functions will require many terms. This will cause the method to become too computationally intensive. A particularly easy choice of basis would be to simply use $\{1, x, x^2, ...\}$. While this will work, it will not be the most efficient choice. It turns out that the best choice is to use a family of orthogonal polynomials as your basis. As discussed briefly above, this will aid in reducing the computational complexity of the problem.

There are many families of orthogonal polynomials to choose from. Some common ones are Legendre, Chebyshev and Hermite. In this paper we will use Legendre polynomials as our basis functions. The reason for this is that they are orthogonal with respect to a weighting function of 1. This will eliminate some additional computational headaches later when we need to compute integrals since the weight function for some families of polynomials can be singular at the endpoints. The Legendre polynomials, P_i, can be defined recursively as

$$\begin{aligned} &P_0(x) = 1, \\ &P_1(x) = x, \\ &P_k(x) = ((2k-1)xP_{k-1} - (k-1)P_{k-2})/k. \end{aligned} \tag{5.1}$$

These polynomials are orthogonal with respect to the weight function $w(x) = 1$ on $[-1, 1]$.

5.2 Order n Galerkin Approach to the Burnside Model

Recall Equation 3.5:

$$v_t = \beta E_t \left[e^{\theta x_{t+1}} (v_{t+1} + 1) \right]. \tag{5.2}$$

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function defined on all of $\mathbb{R}$. Let $v_t = f(x_t)$ and write $x_t = x$. Then we have

$$f(x) = \beta E \left[e^{\theta x_{t+1}} (f(x_{t+1}) + 1) \middle| x \right]. \tag{5.3}$$

Note that we have $\mathcal{N}[f](x) = f(x) - \beta E \left[e^{\theta x_{t+1}} (f(x_{t+1}) + 1) \middle| x \right]$, where x_{t+1} is also a function of x. Also, in this case we will take $B_1 = \{g \in L^2 : g(x) = 0 \text{ if } x \notin [a, b]\}$ and $B_2 = L^2$. The best choice for the values of a and b will be computed later. For now, we will assume they are known and fixed. Let $\{P_i\}_{i=0}^n$ be the first $n+1$ Legendre polynomials on $[-1, 1]$ as defined in (5.1) and define

$$\varphi_i(x) = \begin{cases} P_i\left(2\frac{x-a}{b-a} - 1\right) & a \le x \le b \\ 0 & \text{otherwise} \end{cases}. \tag{5.4}$$

The φ_i are now orthogonal on $[a, b]$ with respect to the weighting function $w(x) = 1$ (in fact, they are orthogonal on any interval containing $[a, b]$). That is,

$$\langle \varphi_i(x), \varphi_j(x) \rangle = \int_a^b \varphi_i(x)\varphi_j(x)dx = 0, \quad \text{for all } i \ne j. \tag{5.5}$$

Define the approximate solution to (5.3) as

$$\hat{f}(x) = \sum_{i=0}^{n} a_i \varphi_i(x). \tag{5.6}$$

To simplify notation, we will write $E_x[\,\cdot\,]$ to mean $E[\,\cdot\,|x]$. Plugging (5.6) into (5.3) we get

$$\begin{aligned}
\sum_{i=0}^{n} a_i \varphi_i(x) &= \beta E_x \left[e^{\theta x_{t+1}} \left(\sum_{i=0}^{n} a_i \varphi_i(x_{t+1}) + 1 \right) \right] && (5.7) \\
&= \beta E_x \left[e^{\theta x_{t+1}} \sum_{i=0}^{n} a_i \varphi_i(x_{t+1}) + e^{\theta x_{t+1}} \right] && (5.8) \\
&= \beta E_x \left[e^{\theta x_{t+1}} \sum_{i=0}^{n} a_i \varphi_i(x_{t+1}) \right] + \beta E_x \left[e^{\theta x_{t+1}} \right] && (5.9) \\
&= \beta \sum_{i=0}^{n} a_i E_x \left[e^{\theta x_{t+1}} \varphi_i(x_{t+1}) \right] + \beta E_x \left[e^{\theta x_{t+1}} \right] && (5.10)
\end{aligned}$$

Collecting terms we have

$$\sum_{i=0}^{n} a_i \left(\varphi_i(x) - \beta E_x \left[e^{\theta x_{t+1}} \varphi_i(x_{t+1}) \right] \right) = \beta E_x \left[e^{\theta x_{t+1}} \right]. \tag{5.11}$$

In order to solve for the coefficients a_j, we let

$$R(x;\mathbf{a}) = \sum_{i=0}^{n} a_i \left(\varphi_i(x) - \beta E_x \left[e^{\theta x_{t+1}} \varphi_i(x_{t+1}) \right] \right) - \beta E_x \left[e^{\theta x_{t+1}} \right] \tag{5.12}$$

and require that

$$\langle R(x;\mathbf{a}), \varphi_k(x) \rangle = 0 \quad \text{for each } k = 0, 1, \ldots, n, \tag{5.13}$$

where the inner product is defined in (5.5).

Combining the orthogonality condition (5.13) with (5.11) we get

$$\int_a^b \varphi_k(x) \sum_{i=0}^{n} a_i \left(\varphi_i(x) - \beta E_x \left[e^{\theta x_{t+1}} \varphi_i(x_{t+1}) \right] \right) dx = \int_a^b \varphi_k(x) \beta E_x \left[e^{\theta x_{t+1}} \right] dx, \tag{5.14}$$

for each k. Interchanging the (finite) sum and integral we have

$$\sum_{i=0}^{n} a_i \int_a^b \varphi_k(x) \left(\varphi_i(x) - \beta E_x \left[e^{\theta x_{t+1}} \varphi_i(x_{t+1}) \right] \right) dx = \int_a^b \varphi_k(x) \beta E_x \left[e^{\theta x_{t+1}} \right] dx. \tag{5.15}$$

This can be written as a matrix system

$$\mathbf{Ma} = \mathbf{b},$$

where

$$M_{k,i} = \int_a^b \varphi_k(x) \left(\varphi_i(x) - \beta E_x \left[e^{\theta x_{t+1}} \varphi_i(x_{t+1}) \right] \right) dx \tag{5.16}$$

and

$$b_k = \int_a^b \varphi_k(x)\beta E_x\left[e^{\theta x_{t+1}}\right]dx. \tag{5.17}$$

It is now time to choose our a and b so that everything will make sense. In particular, we need to choose these limits in such a way that we will never need to evaluate a Legendre polynomial outside of them. First, we need to discuss how we will evaluate the expectations in (5.16) and (5.17).

Since x_{t+1} is normally distributed, it is natural to use Gauss-Hermite quadrature to evaluate the above expectations. The general rule is as follows. Suppose $Y \sim N(\mu,\sigma^2)$ and $h(\,\cdot\,)$ is an arbitrary function defined on $(-\infty,\infty)$. Then

$$E[h(Y)] = \frac{1}{\sqrt{\pi}}\sum_{i=1}^{m}\omega_i h(\sqrt{2}\sigma x_i + \mu) + \frac{m!}{2^m}\cdot\frac{h^{(2m)}(\xi)}{(2m)!},$$

for some $\xi \in (-\infty,\infty)$ and where ω_i and x_i are the Gauss-Hermite weights and nodes, respectively (see [7] for details). We will use $m = 10$ here. In this case, the largest node will be approximately 3.44 standard deviations from the mean.

Let $\alpha(x) = (1-\rho)\bar{x} + \rho x$. Then $x_{t+1} = (1-\rho)\bar{x} + \rho x + \varepsilon = \alpha(x) + \varepsilon$ and so $x_{t+1} \sim N(\alpha,\sigma^2)$. This implies that we can apply the above approximation to evaluate (5.16) with $h(y) = e^{\theta y}\varphi_i(y)$. To make things a little clearer, suppose for the moment that $\bar{x} = 0$. Then we need to choose a and b such that $\{\sqrt{2}\sigma x_i + \rho x : x \in [a,b]\} \subseteq [a,b]$ for any i. Since $|\rho| < 1$, this amounts to choosing b such that $3.44\sqrt{2}\sigma \leq (1-|\rho|)b$. Therefore, take $b = \frac{3.5\sqrt{2}\sigma}{1-|\rho|}$. Similarly, take $a = -\frac{3.5\sqrt{2}\sigma}{1-|\rho|}$. Thus, in our case (when $\bar{x} \neq 0$), we will have

$$a = \bar{x} - \frac{3.5\sqrt{2}\sigma}{1-|\rho|} \quad \text{and} \quad b = \bar{x} + \frac{3.5\sqrt{2}\sigma}{1-|\rho|}.$$

It is important to note that while this will hold for any ρ, for large values of $|\rho|$ this interval will become too large and most accuracy will be lost. This turns out, at least in the context of this paper, to not matter much. Typically in asset pricing problems ρ will be small and negative. Throughout this paper it will be assumed that $|\rho| \leq 0.8$.

The remaining integrals in (5.16) and (5.17) can be computed using a Gauss-Legendre quadrature. The number of nodes used in these approximations will vary from one to the next depending on the order. If $n < 10$, then 10 nodes are used. If $n \geq 10$, then $n+1$ will be used. Finally, the $n+1$ by $n+1$ linear system is solved for the vector $\mathbf{a}$ of coefficients and we will have our approximation for $v_t = f(x_t)$.

5.3 Error Analysis

In performing the error analysis, the following formulas were used:

$$E_1 = 100 \times \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \quad \text{and} \quad E_\infty = 100 \times \max\left\{\left|\frac{y_t - \hat{y}_t}{y_t}\right|\right\},$$

where y_t is the true solution and $\hat{y}_t$ is the approximated solution.[5] The code was then run for increasing orders until each of the above errors seemed to roughly converge. In most cases this was around 10^{-7}. As before, $N = 20$ was used for the error computations.

[5] The error was reported in this way since this is how it was done in Section 4, which was how it was done in [5].

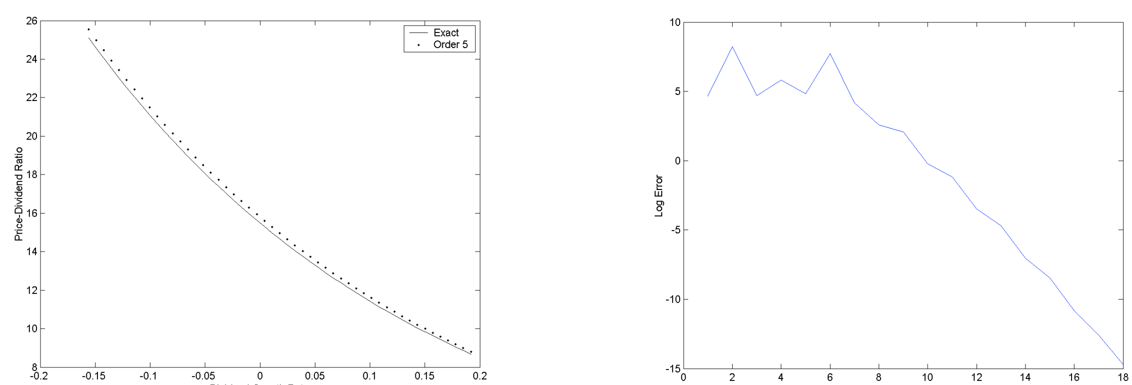
The error analysis shown on the next page is split up in the following way. Table 5.1 shows the errors of the approximation method for the benchmark case as well as cases when β and σ are varied. Table 5.2 shows the results when θ is varied. Finally, Table 5.3 shows cases when ρ is varied, up to and including $\rho = 0.8$.

It is important to note that the cases in which θ is large negative or ρ is positive result in a true solution that has much more curvature than in the other cases. This means that it will require many higher order polynomials to accurately approximate the function. This can be seen in the tables; it takes $n = 21$ in the $\rho = 0.8$ case to get the same accuracy as $n = 6$ in the benchmark.

For all the cases considered here, we were able to obtain decent error estimates, eventually. In the cases for large ρ, much higher order was needed. The method allows for this although at the cost of taking much longer to return the answer. For many cases, as with perturbation methods, the approximation required very little effort to get quite close. These are cases where the price-dividend ratio is constant or nearly linear in the region of interest.

Upon comparing the error results from the perturbation method and the projection method, one could come to the conclusion that projection is much better. This may or may not be true. In terms of ease of computation, projection is much better. It takes virtually no extra effort to write computer code that works for general n than it does to specify n up front and never be able to change it. Unfortunately, the same is not true for perturbation methods. For any given n, the system of equations (4.22)-(4.23) must be expanded and solved. Since the equations involve many partial derivatives, it would be very difficult to do this numerically for general n. It may be possible to have Maple or Mathematica compute the equations for general orders, but this was not attempted here. It is hard to say what would happen in terms of comparison if we were easily able to perform the perturbation method for arbitrary n.

Figure 5.1(a) shows the plot of the order 5 approximation and the exact solution in the $\rho = 0.7$ case. This is the highest order in which the two curves are distinguishable. For order 6 and above the approximation is good enough that the two curves appear to lie on top of one another. Finally, Figure 5.1(b) shows the log error for the $\rho = 0.8$ case. The fact that the curve is asymptotically linear with the order says that the error is decreasing exponentially.



(a) (b)

Figure 5.1: (a) A plot of the order 5 approximation and the exact solution for the $\rho = 0.7$ case. (b) A plot of $\log E_1$ for the $\rho = 0.8$ case.

n	Benchmark		$\beta = 0.5$		$\beta = 0.99$		$\sigma = 0.001$		$\sigma = 0.1$	
	E_1	E_∞	E_1	E_∞	E_1	E_∞	E_1	E_∞	E_1	E_∞
1	2.56(-1)	2.77(-1)	2.79(-2)	4.72(-2)	5.55(-1)	5.76(-1)	2.08(-4)	2.25(-4)	2.39(0)	2.57(0)
2	1.00(-4)	2.03(-4)	1.15(-4)	1.65(-4)	1.21(-4)	2.60(-4)	1.86(-7)	1.90(-7)	3.86(-3)	7.25(-3)
3	6.15(-6)	6.94(-6)	6.72(-7)	1.47(-6)	1.33(-5)	1.41(-5)	1.86(-7)	1.86(-7)	4.50(-4)	5.04(-4)
4	1.89(-7)	1.92(-7)	2.74(-8)	3.17(-8)	3.98(-7)	4.01(-7)			4.40(-7)	9.70(-7)
5	1.90(-7)	1.90(-7)	2.75(-8)	2.75(-8)	4.00(-7)	4.00(-7)			1.90(-7)	1.96)(-7)
6	1.90(-7)	1.90(-7)	2.75(-8)	2.75(-8)	4.00(-7)	4.00(-7)			2.22(-7)	2.22(-7)
7									2.22(-7)	2.22(-7)

Table 5.1: Average and maximum relative (percentage) errors for the benchmark, β and σ cases

n	$\theta = 0$		$\theta = 0.5$		$\theta = -5$		$\theta = -10$	
	E_1	E_∞	E_1	E_∞	E_1	E_∞	E_1	E_∞
1	2.80(-7)	2.80(-7)	5.21(-2)	5.43(-2)	1.70(0)	1.94(0)	4.86(0)	5.86(0)
2	2.80(-7)	2.80(-7)	3.56(-6)	7.19(-6)	4.49(-3)	9.07(-3)	4.50(-2)	8.42(-2)
3			4.74(-7)	4.83(-7)	4.32(-4)	5.32(-4)	4.69(-3)	6.33(-3)
4			3.38(-7)	3.38(-7)	9.22(-7)	1.96(-6)	3.74(-5)	7.46(-5)
5			3.38(-7)	3.38(-7)	8.33(-8)	9.62(-8)	1.76(-6)	2.76(-6)
6					1.26(-7)	1.26(-7)	1.25(-7)	1.45(-7)
7					1.26(-7)	1.26(-7)	1.13(-7)	1.14(-7)
8							1.13(-7)	1.13(-7)

Table 5.2: Average and maximum relative (percentage) errors for the θ cases

n	$\rho = 0$		$\rho = 0.5$		$\rho = 0.7$		$\rho = 0.8$	
	E_1	E_∞	E_1	E_∞	E_1	E_∞	E_1	E_∞
1	1.91(-7)	1.91(-7)	6.74(1)	6.95(1)	1.30(2)	1.32(2)	1.05(2)	1.06(2)
2	1.91(-7)	1.91(-7)	4.65(-1)	8.39(-1)	5.06(1)	1.30(2)	1.83(3)	3.84(3)
3			2.34(-1)	2.56(-1)	4.18(1)	4.37(1)	1.07(2)	1.10(2)
4			1.78(-3)	2.60(-3)	1.96(0)	4.18(0)	1.24(2)	3.40(2)
5			4.26(-4)	5.01(-4)	1.82(0)	2.00(0)	1.21(2)	1.28(2)
6			2.34(-6)	3.99(-6)	5.22(-2)	8.40(-2)	9.23(1)	2.25(2)
7			5.98(-7)	6.97(-7)	2.66(-2)	3.11(-2)	6.05(1)	6.60(1)
8			2.07(-7)	2.11(-7)	7.16(-4)	1.04(-3)	6.79(0)	1.33(1)
9			2.06(-7)	2.09(-7)	2.32(-4)	2.91(-4)	6.96(0)	7.91(0)
10			2.06(-7)	2.09(-7)	6.01(-6)	1.07(-5)	4.77(-1)	8.02(-1)
11					1.62(-6)	2.11(-6)	2.55(-1)	3.07(-1)
12					2.71(-7)	3.20(-7)	1.82(-2)	3.10(-2)
13					2.41(-7)	2.58(-7)	7.17(-3)	9.19(-3)
14					2.47(-7)	2.62(-7)	5.06(-4)	8.84(-4)
15					2.47(-7)	2.62(-7)	1.51(-4)	2.09(-4)
16							1.07(-5)	1.98(-5)
17							2.08(-6)	3.35(-6)
18							2.18(-7)	4.15(-7)
19							4.04(-7)	4.43(-7)
20							3.74(-7)	4.24(-7)
21							3.72(-7)	4.21(-7)

Table 5.3: Average and maximum relative (percentage) errors for the ρ cases

6 Agent-Based Models

The final section of this paper will introduce a modeling methodology that is very different from the one considered in earlier sections. Previously, we were forced into making assumptions about things such as investor homogeneity, equilibrium conditions, rational expectations, etc. This was necessary for many reasons, but most importantly it keeps our models tractable. As an example, consider the model described in Section 2 but instead of assuming all investors are identical, assume (more realistically) that there are N different investors. Each of these agents will have a different utility function (or at the very least a different level of risk aversion), different endowments and access to different information sets. We now find ourselves in a position where we need to keep track of many things. Instead of all investors holding an equal share of the stock and zero bond, there will now be diverse holdings across agents. Thus, in Section 2, there are nice solutions but the model is somewhat less "real" than we might like. Agent based modeling takes a sort of opposite approach. Instead of supposing equilibrium conditions, a dynamic model is created in which emergent equilibrium behavior is possible. This, unfortunately, comes at the price of no longer having the nice solutions we had before.

This section will be divided into two parts. First, we will discuss the motivation for wanting to take an alternative approach to modeling agents. Then, a brief and fairly nontechnical overview of the method used to accomplish this will be presented.

6.1 Motivation

To someone who has spent years studying economics, assumptions such as representative agents and perfect information become second nature. Economists think nothing of making assumptions such as these to simplify their problems and keep solutions tractable. These assumptions lead to results similar to those shown previously in this paper; an economy in which no one trades for example. For someone new to the field, most of these assumptions are probably quite troublesome, and justifiably so. If one looks around, we see nothing but heterogeneity. This is true in all aspects of life, but especially in how consumers behave. Given any two people in the economy, they will likely disagree about almost everything, from their discount factors to what they believe the price of apples will be tomorrow. It is for these reasons that a new field of economic thought is emerging - *Agent-Based Computational Economics (ACE).* ACE models attempt to capture the heterogeneity that is obviously present in the world. Leigh Tesfatsion, one of the leading researchers in ACE modeling, writes ([12], p. 4):

> "Economies are complex dynamic systems. Large numbers of micro agents engage repeatedly in local interactions, giving rise to global regularities such as employment and growth rates, income distributions, market institutions, and social conventions. These global regularities in turn feed back into the determination of local interactions. The result is an intricate system of interdependent feedback loops connecting micro behaviors, interaction patterns, and global regularities."

Typically in a traditional economic model agents do not interact directly. Instead, there is a third party "overseeing" all transactions in the economy. This is called a *Walrasian auctioneer.* Its purpose to to take all agents' demands and set prices such that the market clears perfectly. Thus, all agents and firms are price takers in this economy. This means that there cannot be any strategic behavior or learning. For example, a firm could not choose to set its price lower in order to increase volume. In this world all agents (and firms) are simply optimizing their utility (or revenue) as functions of the given prices and dividends. Also, in these traditional models (henceforth referred to as *Walrasian models*) it is also assumed that this auctioneer can perform its job completely absent of friction. That is, there are no problems with buyers finding sellers and there is never any excess demand or supply. It also does not cost anything to have this auctioneer do this job. This is quite unsatisfactory if we wish to have a model that accurately describes the real world in which we live. There are many questions left unanswered and many aspects of consumer behavior left unaddressed. Tesfatsion sums up the Walrasian model very nicely in the following ([12], p. 5):

> "Walrasian equilibrium is an elegant affirmative answer to a logically posed issue: can efficient allocations be supported through decentralized market prices? It does not address, and was not meant to address, how production, pricing, and trade actually take place in real-world economies through various procurement processes."

What processes would we like our agents and firms to go through? Consumers should identify what goods and services they wish to purchase, and at what price and volume. Suppliers should identify what goods and services they are willing to provide, and at what price and volume. Potential trade partners should be found and deals negotiated. Also, long term relationships should be possible between buyers and sellers. It is these processes that are traditionally replaced by the assumption of equilibrium conditions, a completely nontrivial assumption to make. Fisher provides a lengthy argument ([6], Ch. 1) that it is, in fact, not acceptable to simply assume that the economy is constantly in equilibrium. The only way one can justify studying equilibrium points alone is if two (nontrivial) conditions hold. First, the system must be stable, that is, it must converge to some equilibrium. Second, this convergence must be sufficiently fast. Otherwise, the time the economy spends in disequilibrium states is very relevant and must not be ignored. A common argument supporting the study of equilibrium points alone is that the economy could not stay at a point of disequilibrium. While this is true, it does not in turn imply that it will converge to an equilibrium either. Again, this raises the question of stability.

As a final case for a new model, we turn our attention to empirical evidence. In Section 2 we saw that in the homogeneous agent case, there is no trading, that is, volume is zero. It was still possible to get price movements, which is interesting in its own right, but in real markets trading takes place. In addition to this, several other empirical properties of real price data cannot be explained within the REE world. Arthur et. al. (1997) provides references for many of these. Specifically, volume and price volatility are large and both show autocorrelation. Stock returns also show small serial correlations. Some technical trading rules can produce statistically significant long-run profits, something that is only possible by luck under the REE assumptions. Finally, under the assumption of full rationality, investors lack incentives to gather information. We conclude this subsection with a quote from the aforementioned paper by Arthur et. al.:

> "By now, enough statistical evidence has accumulated to question efficient-market theories and to show that the traders' viewpoint cannot be entirely dismissed. As a result, the modern finance literature has been searching for alternative theories that can explain these market realities."

6.2 Methodology

At this point, we have provided an argument that the classical approach to these models is insufficient. The next question is, then, what do we do to fix this? The first idea would be to assume everything the same as we did in Section 2 but to allow investors to differ in their expectations. Suppose that for a payoff x_{t+1}, investor i will compute the price to be

$$p_t^i = E_t^i[m_{t+1}^i x_{t+1}], \tag{6.1}$$

where m^i is agent i's stochastic discount factor and E_t^i is agent i's expectation conditional on information available at time t. The information set could be anything that might influence the outcome of future dividends, including, but not limited to, economic indicators, rumors, historical price data, news, etc. The assumption of identical investors amounts to each agent using the information set in an identical way. In this case we are able to drop the superscripts and get $p_t = E_t[m_{t+1}x_{t+1}]$, as in Section 2. In fact, in this simplified case, agents can explicitly compute the price of the risky asset. By iterating the right hand side we can arrive at

$$p_t = E_t \sum_{j=0}^{\infty} m_{t,t+j} d_{t+j}, \tag{6.2}$$

where $m_{t,t+j} = \beta^j \frac{u'(c_{t+j})}{u'(c_t)}$ (see [4], p. 27). Note that this relies heavily on the fact that all investors are identical. This is called *deductive reasoning*; the agent was able to compute the correct answer.

Now, suppose that agents do not share expectations. Given some information set, "there may be many different, perfectly defensible statistical ways, based on different assumptions and different error criteria to use it to predict future dividends" ([1], p. 5). In this situation, agents do not know each other's expectations and there is no way for them to deductively figure it out. But in keeping with our original idea, to only change the agents' expectations, suppose that agent i tries to deductively compute his expectation of tomorrow's price. Then we would have

$$E_t^i[p_{t+1}] = \beta E_t^i \left\{ \sum_j \left[w_{j,t+1} E_t^j(x_{t+2}) \right] \right\}, \tag{6.3}$$

where $w_{j,t+1}$ is the weight, or confidence, placed in agent j's forecast. In the previous case, all agent's were identical and so j could be ignored. Now, however, the expectations of all agents in the market must be known and taken into account in order to deductively compute the expectation of tomorrow's price. But each of these expectations rely on all agents expectations for the next period which rely on all expectations for the period after that. In addition, all other agent's are basing their expectations on agent i's forecast. So each agent implicitly defines its own expectations as well. This can lead to instability. It is clear that this situation is hopeless. Note that the agents in this explanation are in no way limited in their intelligence or rationality. The problem is simply ill-posed under these conditions.

Instead of agents following deductive reasoning as above, in this heterogeneous world they will be assumed to follow *inductive reasoning*. To understand what this means imagine each investor acting as a little econometrician. Instead of being able to actually compute the solution, the agent will build a model to predict the price. He can estimate parameters based on past data, much like an econometrician would do. Each agent will have a different model, but each will be correctly formulated. By this I mean that the agents will rationally and intelligently build models, they will just differ from one another. In fact, each agent will have several of these models to use and will choose to use the one(s) that are most reliable from a historical standpoint. Also, every so often bad models are eliminated and new ones created. These points will be discussed for the remainder of this section. The following will be largely based on the Sante-Fe Artificial Stock Market (see [1]).

Suppose that at each time, the time series of current and past prices and dividends is summarized in an array of J market descriptors. These descriptors could be anything at all. For example, the fourth descriptor could be "price has risen for z consecutive periods". Each agent will possess a set of M predictors.[6] A predictor consists of two parts. First, there is a market condition that must be true in order to make the predictor active. A "1" indicates that the market state in that position has occurred, a "0" indicates that it did not occur and a "#" indicates that we are not concerned with that state. The second part of a predictor is the forecast formula, that is, what to forecast if the market condition is true. The forecast rule will be a 2 vector, (a, b), that forecasts $E(p_{t+1} + d_{t+1}) = a(p_t + d_t) + b$. It may be helpful at this point to consider a simple example.

Example 6.1. *Suppose, for simplicity, that* $J = 4$*. Then the market can be summarized by just 4 descriptors. These are*

1. *The price is greater than the 5 period moving average.*
2. *The price is greater than the 10 period moving average.*
3. *The current price-dividend ratio is greater than 10.*
4. *The current price-dividend ratio is greater than 15.*

Then the condition $(1\#\#\#)$ *will recognize any state in which the price exceeds the 5 period moving average, regardless of all other descriptors. For example, this condition would recognize the state* (1011) *or* (1100)*, but not* (0111)*. Similarly, the condition* $(\#0\#1)$ *will only recognize states in which the price is not greater than the 10 period moving average and the price dividend ratio is greater than 15.*

A predictor could look like $(1\#\#\#)/(0.9, 1)$*. This says that if the price now exceeds the 5 period MA, then forecast tomorrow's price plus dividend as 90% of today's plus 1.*

After the agent has his M predictors determined, he awaits the market conditions for each period. Once the prices and dividends are revealed, the agent will have some set of active predictors, that is, the ones that correspond to that state of the world. He then uses a linear forecast of the H most accurate active predictors to calculate an amount to bid or offer. After the markets clear, new prices and dividends are revealed and the rules are updated for accuracy. In this sense, the agent is constantly learning. If a particular rule does well it will be used more often and will continue to be used as long as it is accurate while rules that do poorly will rarely be used.

In addition to updating the accuracy of the rules, the agents are also able to generate new ones via a genetic algorithm. Every so often (250 periods, for example) the agent will discard its worst rules and try to formulate new ones. New rules are generated using mutation or crossover, with some predetermined probability. Mutation involves changing the bits of the predictor at random. Crossover involves taking two parent rules and creating a new "offspring" from them. See [9] for more on forming new rules.

We conclude by posing a natural question. If a model economy, in which a rational expectations equilibrium exists, is populated with intelligent agents, will they eventually discover this equilibrium? The *Marimon-Sargent hypothesis* says that the answer is yes. It has been shown, however,

[6] This assumes all investors are equally intelligent. More complex models may allow for investor i to have M_i predictors, thus creating further diversity among agents.

that one can construct situations in which this fails to be true. For example, Arthur et. al. mention that for fast enough learning rates agents will not learn the REE [1]. It remains an open problem to exactly specify under which conditions the hypothesis is true. It appears that in most cases that we are concerned about it will prove true. This is a very vague statement and needs to be made much more precise. This is one of the main problems I will be studying during my Ph.D. research.

References

[1] W. B. Arthur, J. H. Holland, B. LeBaron, R. Palmer, and P. Tyler, *Asset pricing under endogenous expectations in an artificial stock market*, in The Economy As an Evolving Complex System II, W. B. Arthur, S. N. Durlauf, and D. A. Lane, eds., Addison-Wesley, 1997, pp. 15–44.

[2] C. Burnside, *Solving asset pricing models with Gaussian shocks*, Journal of Economic Dynamics and Control, 22 (1998), pp. 329–340.

[3] A. C. Chiang, *Fundamental Methods of Mathematical Economics*, Princeton University Press, third ed., 1984.

[4] J. H. Cochrane, *Asset Pricing*, Princeton University Press, 2001.

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

[6] F. M. Fisher, *Disequilibrium Foundations of Equilibrium Economics*, Cambridge University Press, 1983.

[7] K. L. Judd, *Numerical Methods in Economics*, The MIT Press, 1998.

[8] B. LeBaron, *Agent-based computational finance: Suggested readings and early research*, Journal of Economic Dynamics and Control, 24 (2000), pp. 679–702.

[9] B. LeBaron, W. B. Arthur, and R. Palmer, *Time series of an artificial stock market*, Journal of Economic Dynamics and Control, 23 (1999), pp. 1487–1516.

[10] R. E. Lucas Jr., *Asset prices in an exhange economy*, Econometrica, 46 (1978), pp. 1429–1445.

[11] N. L. Stokey and R. E. Lucas Jr., *Recursive Methods in Economic Dynamics*, Harvard University Press, 1989.

[12] L. Tesfatsion, *Agent-based computational economics: A constructive approach to economic theory*, Handbook of Computational Economics, Volume 2: Agent-Based Computational Economics, (to appear).