Agricultural and Environmental Policy Models: Calibration, Estimation and Optimization

Richard E. Howitt

January 18, 2005

Contents

Chapter 1

Introduction

1.1 Introduction to Linear Models

1.1.1 Definition of a Model

Everyone uses models to think about complex problems. Usually our model is a simple weighting of past experiences that simplifies decisions. For example, after an initial learning period most people drive a car with a model that assumes a certain steering and braking action and only make radical changes from an established pattern when an unexpected emergency occurs. After the emergency most drivers return to their basic model. Why is the model of driving by exception normally optimal? The answer is that this driving model reduces the number of standard decisions that we have to think about and allows us to be more observant for the exceptional situation that requires a different action.

It is often thought that models are limited to algebraic representations and, as such, are hard to construct or interpret. This puts up an artificial barrier to mathematical models that often prevents an evolutionary approach to thinking about them. In reality, everyone uses models to think about complex events, as the process of constructing a model is part of the human process of thinking by analogy. For example, many people use astrology to guide their decisions, a curious but ancient model of relating the position of planets and stars to events in their lives. Skeptics point out that the ambiguity of most astrological forecasts makes quantitative measures hard to confirm. Perhaps they miss the point of astrology, which may not be to accurately predict events, but to give illusion of knowledge over unpredictable events. However, as economists we should be interested in astrology as a product for which the demand has been strong for several millennia.

For this book, the point is to see mathematical models as a practical extension of the graphical models with which we started our micro economic analysis. Mathematical models allow us to explore many more dimensions and interactions than graphical representations, but often we can usefully use simple graphical examples to clarify a mathematical problem. With their larger number of variables, mathematical models can be specified in a more realistic manner than graphical analysis but are still limited by data and computational requirements.

A model is by definition abstracted and simplified from reality and should be judged by its ability to deliver the required precision of information for the task in hand. It is easy in economics to judge a model on its mathematical elegance or originality, that is, as a work of art or artifact rather than a tool

1.1.2 Types of Models

Verbal Models

Thomas Kuhn has proposed that most scientific thought takes place within paradigms that gradually evolve. Given the evolutionary nature of science it is not surprising that most research takes place within a paradigm rather than trying to change paradigms. One of the older and best-known paradigms in economics is Smith's analogy of the price and market system to an "invisible hand." Simple verbal models such as this are very helpful in concisely defining the qualitative properties of a paradigm. The ability to give a simple verbal explanation of the model is probably a necessary condition for full understanding of a complex mathematical model. If you are unable to explain the essence of what you are modeling to your Grandmother, you probably don't really understand it.

Geometric (Graphical) Models

Geometric methods are the way that we are introduced to economic models and a method where our natural spatial instincts are more easily harnessed to show interrelationships between functions and equilibria. For most empirical applications, graphical models in two or three dimensions are simply too small to adequately represent the empirical relationships needed. Most graphical models can be represented by a system of two equations, whereas optimization models that we will encounter later in the book can have tens or hundreds of equations and variables. However, like verbal models, graphical models are very useful for conceptualizing mathematical relationships before extending them to multidimensional cases.

Mathematical Models

In economics, the term model has become synonymous with algebraic models since they have been the essential tools for empirical and theoretical work for the past five decades. For this book, a critical difference in model specification is between optimizing behavioral models and optimizing structural models.

Behavioral models yield equations that describe the outcome of optimizing behavior by the economic agent. Assuming that optimizing behavior of some sort has driven the observed actions allows us to deduce values of the parameters that underlie it. For example, observations on the different amounts of a commodity purchased as its price changes, allows the specification and estimation of the elasticity of demand.

An alternative approach to specification of this problem is to define structural equations for the consumer's utility function, represent the budget constraint and alternative products as a set of constraint equations, and explicitly solve the resulting utility optimization problem for the optimal purchase quantity under different prices. With a deterministic model and a full set of parameters, both these approaches would yield the same equilibrium. However this situation rarely occurs and each approach has its relative advantages.

Another distinction is between positive and normative models. Behavioral models are invariably positive models where the purpose is to model what economic agents actually do. In contrast, structural models are often normative and are designed to yield results that show what the optimal economic outcome should be. Inevitably normative models require some objective function that purports to represent a social objective function. This is very difficult to specify without strong value judgments.

The Development of Computational Economics

In the past, it has usually been the case that econometric models are positive and programming models are normative since they have an explicitly optimized objective function. This has led to an unfortunate methodological division among empirical modelers on the supply side of Agricultural and Development economics into practitioners of econometric and programming approaches. The difference in approach has been divided along the lines of normative and positive models in the past. With the development of calibration techniques for optimization models, programming approaches can now incorporate some positive data based parameters and thus build a

continuous connection from the pure data based econometrically estimated models through to the linearly constrained programming models. From one viewpoint econometric models are data intensive while calibrated nonlinear optimization models are more computationally intensive.

In recent years the sub-discipline of has emerged, driven initially by the empirical application of macro-economic models, but now used extensively for micro economic applications. The two leading texts in the area are "Applied Computational Economics and Finance" and "Numerical Methods in Economics" (Miranda and Fackler, 2002; Judd, 1999).

As we would expect, shifts in both the supply and demand have stimulated the emergence of this new economic field. The shift in the supply function of computation ability and cost has largely been driven by which states that "The number of transistors on a given chip doubles every eighteen months without any increase in cost." This remarkable trend, which was first proposed by Gordon Moore, a co-founder of Intel, is predicted to continue until at least 2010. Clearly we are in the middle of a dramatic reduction in the cost of computation. Along with the changes in hardware supply, there have been similar changes stimulated in the supply of software for computational economics.

The demand for computational economics is also shifting out due to the increasing in stochastic dynamic problems in applied economic analysis. Most realistic problems in stochastic dynamics and game theory are simply too complicated to solve analytically, and like most empirical sciences, economics will increasingly have to rely on numerical methods to test hypotheses and predict behavior.

Of more concern to those in this book is the fact that growth areas for applied economic analysis are in environmental, resource and development economics. Both these fields are characterized by the absence of large reliable data series and the need for disaggregate analysis. It's not that econometric approaches are unsuited to supply side analysis in these areas, its just that the data needed is usually absent.

These two shifts bode well for the growth in optimization models in the future. While there are many books on optimization modeling using linear and quadratic structural approaches, for example Paris (1991) and Hazell and Norton (1986), there is no published text on calibrating micro-economic models. This reader is a start on an introductory text for calibrating optimization models.

Chapters 1 – 4 are a brief introduction to the specification, solution and interpretation of linear structural models. The specification and solutions are defined in terms of linear algebra for reasons of compactness, clarity and

continuity with the remaining sections. Chapters 5 – 11 give an introduction to the development of nonlinear calibrated behavioral models.[1]

1.1.3 Uses for Models

Given an economic phenomenon there are three tasks that we may want to perform with economic models.

1. *Explaining observed actions* This is usually performed by structural analysis using positive econometric models. Given a structure in the form of a specific set of equations, the parameters that most closely explain the observed behavior are accepted as the most likely explanation.

2. *Predicting economic phenomena* As the Druids found out with Stonehenge, forecasting significant economic events is a source of power. Forecasting models are the ultimate outcome of the positivistic viewpoint where the structure is unimportant in itself and the accuracy of the out of sample forecast is the key determinant of model value. Econometric time series models are the best examples of pure forecasting models, although the ability to produce accurate out of sample forecasts should be used to assess the information value of all types of models.

3. *Policy models controlling or influencing certain economic outcomes* This process is generally referred to as policy evaluation since public economic policies are justified on the basis of improving some set of economic values. Both structural econometric and optimization models are used for policy evaluation, however due to the dearth of sample data and the wealth of physical structural data; policy models of agricultural production and resource use are often specified as optimization models.

1.1.4 Types of Policy Models

Econometric Models (Positive Degrees of Freedom)

Econometric structural models have been the standard approach to agricultural economic models for the past twenty years. Econometric models

[1] Note — the ARE252 reader only contains chapters 1–9.

of agricultural production offer a more flexible and theoretically consistent specification of the technology than programming models. In addition, econometric methods are able to test the relevance of given constraints and parameters given an adequate data set. The initial econometric research on production models was performed on aggregated data for multi-output/multi-input systems, or single commodities for more disaggregated specifications. However, despite several methodological developments econometric methods are rarely used for disaggregated empirical microeconomic policy models of agricultural production. This is usually because time series data is not generally available on a disaggregated basis and the assumptions needed for cross-section analysis are not usually acceptable to policy makers with regional constituencies. In short, flexible form econometric models have not fulfilled their empirical promise mostly due to data problems that do not appear to be improving.

Constrained Structural Optimization (Programming) Models

Optimization models have a long history of use in agricultural economic production analysis. There is a natural progression from the partial budget farm management analysis that comprised much of the early work in agricultural production to linear programming models based on activity analysis and linear production technology. Often linear specifications of agricultural production are sufficiently close to the actual technology to be an accurate representation. In other cases the linear decision rule implied by many Linear Programming (LP) models is optimal due to Leontief or Von Liebig technology in agriculture.

Despite the emphasis of methodological development for econometric models, programming models are still the dominant method for microanalysis of agricultural production and resource use. Their applications are widespread due to their ability to reproduce detailed constrained output decisions and their minimal data requirements. As noted above, econometric model applications on a microeconomic basis are hobbled by extensive data requirements.

LP models are also limited largely to normative applications as attempts to calibrate them to actual behavior by adding constraints or risk terms have not been very successful.

Calibrated Positive Programming (PMP) Models (Zero Degrees of Freedom)

Much of this book is focused around a method of calibrating programming models in a positive manner that has been a major focus of my research over many years (Howitt, 1995b). The approach uses the observed allocations of crops and livestock to derive nonlinear cost functions that calibrate the model without adding unrealistic constraints. The approach is called Positive Mathematical Programming (PMP). The focus of the book is on specifying, solving and interpreting several Positive and Normative Programming models used in Agricultural and Environmental Economics

Computable General Equilibrium (CGE) Models

CGE models have been used in macro-economic and sectoral applications for the past fifteen years, using a combination of fixed linear proportions from (SAMs) and calibrated parameters from exogenous elasticities of supply and demand. CGE models have much in common with PMP models in their data requirements and conceptual calibration approach. They will not be addressed directly in this book due to space constraints.

Ill-Posed Maximum Entropy (ME) Models (Negative Degrees of Freedom)

This class of models is emerging as extensions of both PMP and econometric approaches. Briefly this approach enables consistent reconstruction of detailed flexible form models of cost or production functions on a disaggregated basis. Often the more complex specifications require that the model contain more parameters than there are observations — hence the term Ill-Posed" problems. We will use two versions of these ME models in the last part of the book, but will not explore the entropy estimation in any depth. An early application of ME to micro production in agriculture is found in Paris and Howitt (1998).

1.1.5 Selecting Policy Models

Selection of the best model for the research task at hand is an art rather than a science. The model builder is constantly balancing the requirements of realism that complicate the model specification and solution against the practicality of the model in terms of its data and computational requirements. This trade-off is similar to the selection of the optimal photographic

models for a mail order catalog where the publisher has to make the subjective trade off between the beauty of the model and the degree of realism that the model will portray. The optimal customer response to the catalog will come from models who are eye-catching but with whom the customers can identify. In economic policy models, they must be simple enough so that the decision maker can identify with the model concept, but at the same time be tractable and able to reproduce the base year data.

There is no ideal model, just some that are more manageable and useful than others. In the words of G.E.P. Box, "All models are wrong, but some are useful."

The aim of this book is first: to give you the theoretical and empirical tools to make informed decisions on the best model specification for particular data and research situations. Second, having selected an emprical modeling method, to give you the ability to implement the model and interpret the results.

The process of econometric model building has three well-defined stages: Specification, Estimation, and Simulation. Programming model building methods have not formally separated these stages. The equivalent stages are Specification, Calibration and Policy Optimization. However the important process of calibrating the models is usually buried in the model specification stage, and often accomplished by the ad hoc method of adding increasingly unrealistic constraints to the model. One of the few programming model texts that even mentions model calibration is Hazell and Norton (1986), who briefly address calibration tests more than methods. This book will explicitly address these different stages of optimization model building and differs from the usual treatment by having a strong emphasis on model calibration.

Chapter 2

Specifying Linear Models

2.1 Constrained versus Unconstrained Models

Simple graphical models and nonlinear models in micro-economic texts are represented as unconstrained demand and supply functions that are optimized using calculus. A simple profit maximizing output is calculated given the following specification.

Given the general nonlinear production function $q = f(x_1, x_2)$, the price of the output q is p per unit output, and the cost per unit of input x_i is w_i.

If the objective is to maximize the profit Π subject to the production function $f(\cdot)$, the model is specified as:

$$\begin{gathered} \max \Pi = pq - x_1 w_1 - x_2 w_2 \\ \text{subject to } q = f(x_1, x_2) \end{gathered}$$

This equality constrained and differentiable problem can be expressed by the familiar Lagrangian function formulation, which by introducing the multiplier λ enables the equality constraint to be incorporated with the objective function. The resulting Lagrangian function can be optimized like an unconstrained function.

$$L = pq - \sum_{\forall i} x_i w_i - \lambda[q - f(x_1, x_2)] \tag{2.1}$$

Figure 2.1 represents the Lagrangian function, which can be maximized by the usual unconstrained approach of taking the partial derivatives $\frac{\partial L}{\partial x_i}$ and setting them equal to zero.

If the production function is defined as a linear relationship, defining a Leontief technology with fixed proportions of inputs per unit output, the

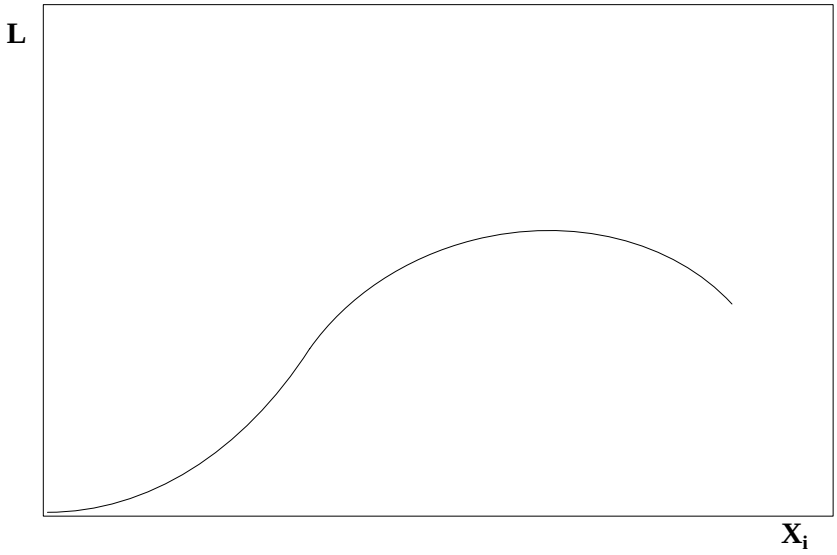


Figure 2.1: the Lagrangian function

problem becomes:

$$\begin{aligned} \max \Pi &= pq - x_1 w_1 - x_2 w_2 \\ \text{subject to} \quad q &= Min[a_1 x_1, a_2 x_2] \\ x_1 &\leq b_1 \\ x_2 &\leq b_2 \end{aligned}$$

This linear profit maximizing production problem can be solved as a Lagrangian, and it can also be rewritten using linear algebra. Note that there is only one constraint in this example.

$$\begin{aligned} &\max c'x \\ &\text{subject to} \quad Ax \leq b \end{aligned}$$

where $c' = [p, -w_1, -w_2]$ and $\tilde{a}_1 = [1, -a_1, -a_2]$, $\tilde{a}_2 = [0, 1, 0]$, $\tilde{a}_3 = [0, 0, 1]$, and $b' = [0, b_1, b_2]$. The vector of input and output activities is: $x' = [q, x_1, x_2]$. Try multiplying this out to check that it is the same problem as above.

2.2 Linear Programming

Linear Program: The equality constraint on the production function above is restrictive in that it implies that all the resources have to be exactly used up in the production processes. Given the nature of farm inputs such as land, labor, tractor time etc, inputs are available in certain quantities, but often they are not fully used up by the optimal production set in a given year. The relationship between the output levels q and the input levels x should be specified as inequality constraints for a more realistic and general specification. This inequality specification results in the Linear Program specification.

Given a set of m inequality constraints in n variables (x), we want to find the non-negative values of a vector x which satisfies the constraints and maximizes some objective function.

2.2.1 Yolo County Farm Model Example

In many places in this book we will use the following simple farm problem as a template. It is based loosely on my local agriculture in Yolo County, and we use the Yolo template to learn how linear programs are specified, solved and interpreted. We will solve it for both analytical and empirical programming exercises.

The farm has the possibility of producing four different crops: alfalfa, wheat, corn and tomatoes. Yields are fixed, so we can measure output by the number of acres of land allocated to each crop. The objective function is measured directly in net returns to a unit of land. This is usually measured in terms of "gross margins" with the variable costs subtracted from revenues for simplicity. Constraints on production are all inequalities and represent the maximum amounts of land, water, and labor available, and a contract marketing constraint on the maximum quantity of tomatoes that the farmer can sell in any year.

The resulting linear program can be written as follows:

Maximize the scalar product of net returns: $c'x$ subject to $Ax \leq b$, where A is the matrix of technical coefficients, b is the vector of input resources available and x is the vector of production or activity levels.[1]

The choice variables (measured in acres of land) are: x_1 = Alfalfa, x_2 = Wheat, x_3 = Corn, and x_4 = Tomato.

The Objective Function, $\Pi = 121x_1 + 160x_2 + 135x_3 + 825x_4$ (with the

[1] Remember that both c and x are $n \times 1$.

net revenue here measured in dollars) is maximized subject to $Ax \leq b$, e.g.:

$$\begin{array}{r} Land\,(acres) \\ Water\,(acre-feet) \\ Labor\,(hours) \\ Contract\,(tons) \end{array} \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 \\ 4.5 & 2.5 & 3.5 & 3.25 \\ 6.0 & 4.2 & 5.6 & 14.0 \\ 0.0 & 0.0 & 0.0 & 33.25 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \begin{array}{c} \leq \\ \leq \\ \leq \\ \leq \end{array} \begin{bmatrix} 600.0 \\ 1800.0 \\ 5000.0 \\ 6000.0 \end{bmatrix}$$

The optimal solution is $x' = [0.0, 419.549, 0.0, 180.451]$

Note that because there are only two binding constraints, there are only two non-zero activities in the optimal solution. We expect tomatoes to come into the profit maximizing solution since they have a high profit margin per unit of land.

On one acre, we can grow 33.25 tons, but can only sell 6000 tons to the processor. Therefore, the maximum acres of tomatoes is:

$$\frac{6000\,tons}{33.25\,tons/acre} = 180.451\,acres$$

The rest of the land ($600 - 180.451 = 419.549$) is used for the next most profitable crop, wheat. This optimal solution is not constrained by either water or labor (i.e., they are not binding).

2.2.2 The General Formulation of a Linear Program

	Columns →					
Row Name ↓	x_1	x_2	$\cdot$	$\cdot$	x_n	RHS
Objective Function	c_1	c_2	$\cdot$	$\cdot$	c_n	
Resource Constraints						
1	a_{11}	a_{12}	$\cdot$	$\cdot$	a_{1n}	$\leq$ b_1
2	a_{21}	a_{22}	$\cdot$	$\cdot$	a_{2n}	$\leq$ b_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\leq$ $\vdots$
m	a_{m1}	a_{m2}	$\cdot$	$\cdot$	a_{mn}	$\leq$ b_m

Alternatively,the previous problem can be written in a compact form:

$$\begin{aligned} &\max c'x \\ &\text{subject to} \quad Ax \leq b, \quad x \geq 0 \end{aligned}$$

2.3 Transformations

Economics is about transformations. In production the transformation is between inputs and outputs, and in consumption it is between consumed products and utility. Mathematical precision is essential when formulating economic models. We therefore need to think very precisely about the economic actions that we are trying to model, and the most commonly modeled action is a transformation. We will start with linear transformations since they are easier, but most micro theory is based on nonlinear transformations such as the decreasing utility that occurs when you eat too many donuts. All economic activities involve a transformation from input to output space or from product to utility space, such as eating donuts.

In our initial case of production, the economic transformation goes from an "m-dimensional" space of inputs (b), to "n-dimensional" space of outputs (x) and then to the scalar space of profit. In other words the production process being modeled takes a set of m inputs, say land, labor, and capital, and transforms them into n outputs, say corn, potatoes and milk, which are all sold for a common commodity, money. There are two transformations in this model of production. From inputs to outputs and from outputs to farm return. In addition we assume that the farmer is trying to maximize his return and will be constrained by some inputs.

These simple transformations characterize the way in which most of the world's population get their living. It is very important to be able to visualize the economic processes that underlie the linear algebra definitions, and be able to go back and forth between the algebraic definitions and the economic interpretations.

In the Yolo problem, the production transformation (mapping) is from land, water, labor, and contract constraints ($m = 4$) to alfalfa, wheat, corn, and tomatoes ($n = 4$) and the objective function transformation (mapping) is from 4-space to the 1-space of a single total farm return.

For example, the mapping in "Classroom space" is:

$$\begin{bmatrix} height & width & length \end{bmatrix} \begin{bmatrix} 0 \\ 6 \\ 12 \end{bmatrix} = \text{scalar}$$

The coordinates in 3-space locate a particular point on the floor that is 6 ft from one wall and 12 ft from another.

2.3.1 Other Examples of Production Transformations

Ford Car Production (Fixed Proportion):

$$\underbrace{\begin{bmatrix} Capital & Labor & Steel & Energy \end{bmatrix}}_{1\times 4} \underbrace{\begin{bmatrix} a_{CT} & a_{CE} \\ a_{LT} & a_{LE} \\ a_{ST} & a_{SE} \\ a_{ET} & a_{EE} \end{bmatrix}}_{4\times 2} = \underbrace{[Truck \quad Mustang]}_{1\times 2}$$

Candy Bar Production (Secret Recipe):

$$\underbrace{\begin{bmatrix} Sugar & Chocolate & Gum & CornSyrup \end{bmatrix}}_{1\times 4} \underbrace{\begin{bmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \end{bmatrix}}_{4\times 1} = \underbrace{[Milky\,Way\,Bar]}_{1\times 1}$$

2.4 Linear Algebra Definitions Used in Linear Programming

Linear Transformation: A *linear transformation* T from n-space to m-space is a correspondence on the space E^n which maps each vector x in E^n into a vector $T(x)$ in m-space, E^m. Transformations are performed by matrices or vectors as in the previous car or candy examples.

Note that scalar multiplications can be carried through the linear transformation. This means that scalar multiplications can be factored out of the transformation.

Example: Given the vectors x_1, x_2, in E^n, and scalars λ_1, λ_2, the transformation $T(\cdot)$ can be written as:

$$T(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 T(x_1) + \lambda_2 T(x_2)$$

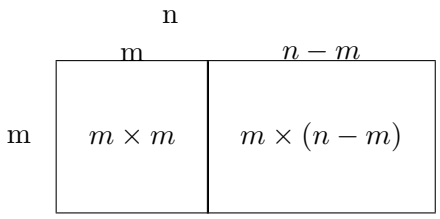
Linear Dependence: If a vector a_i in an $m \times n$ matrix A can be expressed as a linear combination of the other vectors then it is *linearly dependent* (Informal, intuitive definition).

Math definition: Given a set of $n\times 1$ vectors $a_1, \ldots, a_m$ in E^n space where $m < n$, the vectors a_i are *linearly dependent* if there exist $\lambda_i (i = 1, \ldots, m)$ such that $\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_m a_m = 0$ where not all the $\lambda_i = 0$.

It is sometimes easier to see the opposite case of linear independence. The vectors are *linearly independent* if the only set of values for λ_i for which the linear transformation can be made to equal zero is when all $\lambda_i = 0$.

Example of Linear Dependence: Set $\lambda_1 = 1$. Since the definition of linear dependence is that $\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_m a_m = 0$, we know $\lambda_2 a_2 + \ldots + \lambda_m a_m = -a_1$. Since $a_1 \neq 0$, thus some $\lambda_i (i = 2, \ldots, m)$ must also be non-zero.

Rank: The *rank* of a matrix is equal to the number of linearly independent vectors in the matrix.



Notes:

1. The number of linearly independent vectors cannot exceed m if $m < n$.

2. The number of linearly independent vectors cannot exceed smaller of the two dimensions, because rank is equal to the dimension of the largest invertible matrix. Remembering that matrices are only invertible if they are of equal dimensions (square)

3. The rank of A is denoted $r(A)$

2.5 Existence of Solutions

Solution: Given a system of constraints $Ax = b$, the vector $\tilde{x}$, is a *solution* to this system if $\tilde{x}$ satisfies the constraints. We want to find the unique set of that optimizes the objective function value.

Note:

1. That the values of a solution vector x are the "weights" in a linear transformation.

2. That since any set of values for x that satisfy the constraints is a solution, there is often a large number of potential feasible solutions, and the problem is to find the best one.

2.6 A Homogenous System

Homogeneous: A system is *homogenous* if all the values for the right hand side are zero ($Ax = 0$). In other words, b is defined as zero. Homogenous systems are often used as they are simpler to represent and manipulate. The trivial solution of $x = 0$ always exists.

Non-homogenous systems can always be converted to homogenous systems by matrix augmentation. We can convert $Ax = b$ to $\hat{A}\hat{x} = 0$, where $\hat{A} = Ab$ the "augmented matrix" and $\hat{x} = \begin{bmatrix} x \\ -1 \end{bmatrix}$, the augmented vector.

Example:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \qquad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

In the example above $Ax = b$, alternatively we can write the same equations in the form of $Ax - b = 0$. Note, we are assuming that the vector b is not linearly independent of A.

$$\begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

or, by redefining the matrices and vectors, we get the equivalent: $\hat{A}\hat{x} = [0]$.

2.6.1 Case 1: No Solution to the System Exists

No solutions exist to $Ax = b$ if the rank of A, $r(A)$, is less than the rank of the matrix $\hat{A}$, where $\hat{A}$ is defined as the augmented matrix, $[Ab]$. Here we compare the rank for the augmented and unaugmented matrices.

The rank condition where $r(\hat{A}) > r(A)$ means that b is linearly independent of A, and therefore there no solutions exist, except the trivial solution.

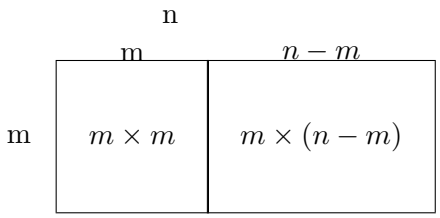
Note that b must be linearly independent of all the vectors in A if augmenting A by b increases the rank of $\hat{A}$ over the rank of A.

From the definition of a solution as a linear transformation, the linear independence of b from every vector in A means that no solutions can exist. In other words, the only set of weights (or allocations) of $\hat{x}$ that can make $\hat{A}\hat{x} = 0$ is the trivial solution when every value in $\hat{x} = 0$.

2.6.2 Case 2: The System has an Infinite Number of Solutions

For the system $Ax = b$, where A is an $m \times n$ matrix, and x is an $n \times 1$ vector. If $r(A) < n$ then there exist an infinite number of non-trivial solutions.

For simplicity, assume that the rank of A is $= m$, $(n > m)$ and arrange the linearly independent columns first. The partitioned matrix dimensions are shown:



Starting with the system $Ax = b$, where x is $n \times 1$, b is $m \times 1$, and A is $m \times n$, A and x can be partitioned as follows:

$$A \rightarrow [A_1 \vdots A_2] \qquad \text{and} \qquad x \rightarrow \begin{bmatrix} x_1 \\ \cdots \\ x_2 \end{bmatrix}$$

We can express the system $Ax = b$ as:

$$[A_1 \vdots A_2] \begin{bmatrix} x_1 \\ \cdots \\ x_2 \end{bmatrix} = b \quad \text{or (alternatively)} \quad A_1 x_1 + A_2 x_2 = b$$

But, by definition, if $r(A) = m$, and A_1 has m linearly independent vectors, then A_1 is "nonsingular" and A_1^{-1} exists. Rearranging and using the inverse yields:

$$x_1 = A_1^{-1}(b - A_2 x_2)$$

Multiplying it out yields:

$$\underset{\substack{solution \\ (unknowns)}}{x_1} = \underset{known}{A_1^{-1} b} - \underset{\substack{known\, or\, zero \\ (chosen)}}{A_1^{-1} A_2 x_2}$$

Note that the value of x_1 depends on x_2, which can have an infinite number of values. This common situation leaves us with an infinite number of feasible solutions to search over for the optimal feasible solution. The

solution approach is to make this intractable infinite problem tractable, by restricting our search to the finite number of feasible solutions that make up the basic solutions defined over the page.

2.6.3 Case 3: A Unique Solution Exists to the System

The system $Ax = b$ has a unique solution if these two conditions hold:

1. $r(A) = r(Ab)$
   That is, the RHS augmented matrix has the same rank as A; and

2. The matrix A is square and of full rank. That is, $A = m \times m$ and $r(A) = m$

2.6.4 Summary

Solutions to the set of equations $Ax = b$ depend on which of the following conditions hold:

$r(A) < r(Ab)$	No Solution
$r(A) < dim(x)$	Infinite Number of Solutions
$r(A) = dim(x)$ and A is of full rank	Unique Solution

2.7 Basic Solutions

Basic Solutions Given $Ax = b$, A is $m \times n$, *and* $r(A) = m$, a *basic solution* to the system is when $(n - m)$ predetermined values of x (eg, x_2) are set equal to zero.

Using the previous example, set the values in $x_2 = 0$. The basic solution is $x_1 = A_1^{-1}b$.

For convenience, I am now changing the set notation to the standard form for this book and define A_1 as B and A_2 as D. We can now write the system as $x_B = B^{-1}b$ where x_1 is denoted x_B and x_2 (set $= 0$) is denoted x_D.

Note that since B is $m \times m$ and non-singular, the inverse B^{-1} exists. Also note that since there are $(n - m)$ non-basis vectors, there is a large, but finite, number of alternative basic solutions.

2.7.1 A Notation Change

Basic Feasible Solution (BFS): A *basic feasible solution* x has non-negative values for the basic solution vectors and the x values in the basis

are all non-negative, i.e.:

Basic Solutions	$x = B^{-1}b$
Feasible	$x \geq 0$

Convex Sets: A set {X} is *convex* if for any points x_1 and $x_2 \in \{X\}$, the line segment $\overline{x_1\, x_2}$ is also $\in \{X\}$ (See Figure 2.2 for a convex set and Figure 2.3 for a non-convex set). Put another way, the set {X} is *convex* if there exist x_1,$x_2 \in \{X\}$, such that the linear combination $\lambda x_1 + (1-\lambda)x_2 \in \{X\}$, for $0 \leq \lambda \leq 1$. This linear combination can be anywhere on the line between x_1 and x_2. Note that $\lambda x_1 + (1-\lambda)x_2$ is, by definition, a convex linear combination.

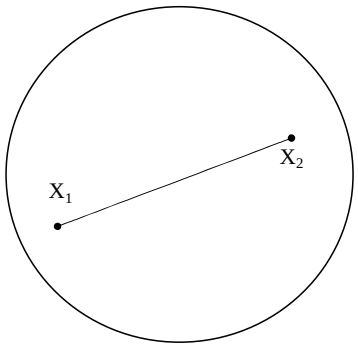


Figure 2.2: $\lambda = 0$ implies that we are at point x_2; $\lambda = 1/2$ implies a point half way between x_1 and x_2; $\lambda = 1$ implies point x_1

Extreme Point: The *extreme point* of a convex set is x if $x \in \{X\}$, but there do not exist any other $\bar{x}_1$, $\bar{x}_2 \in \{X\}$ such that $x = \lambda\bar{x}_1 + (1-\lambda)\bar{x}_2$ for $0 < \lambda < 1$(See Figure 2.4). What this says that an extreme point is a member of the set, but it cannot be expressed as a linear combination of any other points in the set. Imagine that you are, like Leonardo Di Caprio or Kate Windslow, standing on the bow of the Titanic— there's nowhere to go but down (i.e., off the ship)!

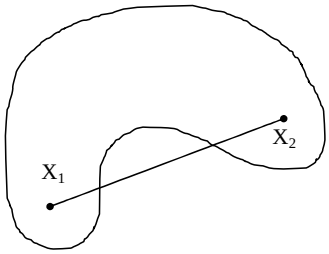


Figure 2.3: Sets with holes or dents are not convex

Notes:

1. In this definition λ is strictly $<$, not $\leq$.

2. Any point that satisfies the definition above is a single point, because it is in the set, but only at an extreme point.

3. Intuitively, an extreme point of a convex set is part of the convex set, but cannot be expressed as a linear combination of any other two points in the convex set.

In a linear system an infinite number of solutions often exist, the objective function is used to select the maximum or minimum value. However to reduce the number of values to search for optimal value we use the properties of the basic feasible solution to reduce the search problem from an infinite set to one over a finite set of possible optimal values.

The number of extreme points of a set of linear constraints is finite. Accordingly, if we search the set of basic feasible solutions for the optimal value of the objective function, we will have found the optimal for the whole set.

Note that if the constraints are nonlinear, the resulting convex set now

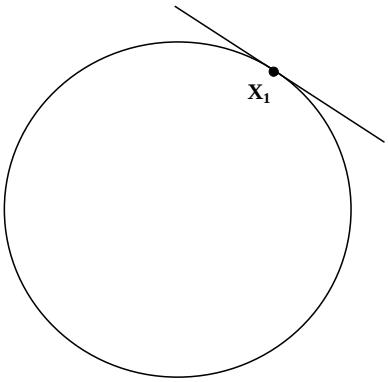


Figure 2.4: Extreme points are not part of any line inside the set.

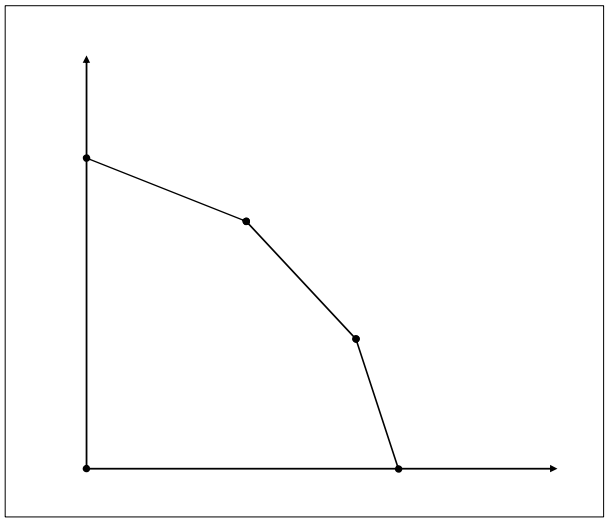


Figure 2.5: Basic feasible solutions are the non-negative extreme points.

has an infinite number of solutions again. To solve for the optimum in this case we have to use a different approach that is addressed later in the book.

2.8 Slack and Surplus Variables

Slack and surplus variables in a linear program are used to convert the inequality constraints into equality constraints, thus making the problem easier to write mathematically and helping the interpretation of the model. They are some times called "artificial variables." While you have to understand the interpretation of these variables, in actual empirical Linear Programming (LP) models they are usually put in automatically by the computer algorithm.

The two types of artificial variable correspond to the two types of inequality constraint. "Less than or equal to" ($\leq$) constraints are converted to equality constraints by slack variables, while "greater than or equal to" ($\geq$) constraints require surplus variables.

Assume that you don't know whether an inequality constraint is binding, but you want to express it mathematically as a binding constraint. In this case, you will have to take up the slack for an input if the constraint is "less than or equal to" or dispose of the surplus input if the constraint is "greater than or equal to."

Example of land as a Slack variable in the Yolo model:

	Alfalfa	Wheat	Corn	Tomato	Slack Land	
Land	1	1	1	1	1	= 600

Notes:

1. Objective function coefficient values for slack(or surplus) activities are zero.
2. Most LP programs put them in automatically (in GAMS).
3. The slack (or surplus) activities give us an initial basic feasible solution for the simplex method to use.
4. This initial basic feasible solution is:
   (a) Guaranteed basic due to the diagonal constraint matrix
   (b) Guaranteed feasible by the non-negativity condition on slack/surplus variables.
   (c) Won't add to the objective function value as slack and surplus activities have zero objective function values.

Surplus variables (for $\geq$ constraints) have a negative signed coefficient in the constraint because we want to reduce the surplus above the right hand side value, and thus reformulate the constraint as an equality.

2.9 Linear Program Objective Function Specification — Traditional Normative Approach

Economic Properties of Linear Program Objective Functions:

- Linearity of the objective function in the parameters. Max $c'x$ where the elements of the vector c are the linear gross margin values.
- Constant Returns to Scale, i.e., cost/unit production is constant.
- Constant output prices (price-taker); no regions, nations or large firms.

Some common examples of objectives are:

- Maximization of profit (Neoclassical Firm objective)
- Minimizing linear deviations from central planning targets
- Minimizing costs in a planned economy
- Minimizing risk of starving next season

The units in the objective function are usually defined by the price units, for example $ per ton. In the constraint matrix it is essential that there is consistency between the constraint units and objective function units.

2.9.1 Yolo Farm Example

No costs are specified in the Yolo problem. The objective function parameters are "Gross margin/acre." which are based on primary data and are equal to total revenue/acre – variable costs/acre.

This simple objective function specification works well until we need to consider capital investments, or changes in production technology as part of the problem to be solved. If changes in the amount of input used is part of the problem to be optimized then the net return per acre will change which requires a more complicated specification.

2.10 Specifying Linear Constraints

2.10.1 Types of Constraints

Linear constraints can usually be classified into three broad classes:

1. Resource Constraints
   In most linear models of production or distribution, resource constraints are intuitive. They usually take the form of a set of "m" summation constraints over the "n" activities. This form assumes that there are n possible production activities and m possible fixed resources used by the activities. The fixed resources are available in quantities $b_1 \ldots b_m$.

   The standard specification of this form is:

   $$\max c'x$$
   $$\text{subject to} \quad Ax \leq b, \quad x \geq 0$$

   The matrix A has individual elements, which are the input requirement coefficients for the production activities.

2. Bounds
   Where there are institutional limits on the activity levels, or because of bounds on the range of the linearity assumption, we may wish to bound the individual activity values. Bounds can be specified using a single row constraint for each bounded activity. The general "less than or equal to" form of constraint can be used in the following way: $Ix \leq b$ where the values for the b vector components b_i are the levels of the upper or lower bounds. Upper bounds have positive values for b_i . Lower bounds have negative values on both b_i and a corresponding -1 value in the identity matrix.

3. Linkage Constraints
   This type of constraint links two or more activities in a prespecified manner. The most common use of linkage constraints is to sum up total output or input use by activities. This operation is often needed where the total input use must be held in storage or purchased from another economic unit. Another common specification is an inventory equation that keeps track of commodity levels used, produced and on hand in the model. The units in the linkage constraint row usually determine the coefficients corresponding to each activity.

   Linkage constraints are best approached systematically:

(a) Decide on the best units for the constraint row.

(b) Write out the logic of the constraint in words. For example, a hay inventory row should be specified in tons. If the activities influencing the row are:

i. Hay grown (acres)
ii. Cattle to be fed (head)
iii. Hay sold (tons)
iv. Hay purchased (tons)

The logical basis for this constraint is — "Hay consumed by cows plus the hay sold equals hay grown plus hay purchased." The easiest way to think of linkage constraints is to define the "Flows in" and "Flows out" of the commodity that the constraint defines. Then decide if the problem requires that the flows in are greater than, less than, or equal to the flows out.

(c) Although linkage constraints are usually equalities, LP problems solve more easily if the constraints are specified as inequalities. The trick is to specify the signs of the coefficients so that the constraint is driven to hold as an equality by the objective function. For example if a constraint is set as a "greater than" inequality that requires a basic ration for animal production but allows a greater amount of food to be fed. An optimizing model will always constrain the ration to the basic level if the food input is costly or constrained.

Example 1: Hay allocation. The level of the four hay activities above, growing, feeding, selling, buying are defined by x_1, x_2, x_3 and x_4. The different impacts of the activities on the hay constraint are defined by the a_{ij} coefficients: a_{41}, a_{42}, a_{43}, a_{44}.

$$\begin{array}{lcccc} \text{Objective Row} & -c_1 & c_2 & c_3 & -c_4 \\ \text{Linkage Row} & -a_{41} & a_{42} & 1 & -1 \quad \leq 0 \end{array}$$

The row is measured in units of the variable input, tons. The linkage row above allows the possibility that more hay can be purchased or grown than is needed for feeding (x_2) or selling (x_3). However, since the extra units have a cost associated with them the objective function will be reduced by the "slack" activities and hence an optimizing model will not purchase more input than is needed. The linkage row does require that the units of hay required for feeding or selling, (x_2 + x_3) are summed and

that this sum is less than or equal to $x_1 + x_4$. In short, you can have too much, but you cannot have too little.

(d) If the problem solution is infeasible or unbounded, check the signs in the constraint.

(e) Check the operation of the linkage constraint by hand calculations on a representative constraint at the optimal solution values.

(f) Inventory stocks can be incorporated most simply by right hand side values (see Example 2). An alternative approach is to specify a separate column for the inventory stock activity, which may itself be constrained by upper or lower bounds (see Example 3). This approach is required if the stock is priced as a separate activity in the objective function.

Example 2: A Begining Inventory A stock of 50 units of x_4 is available at the start of the problem.

Objective Row	$-c_1$	c_2	c_3	$-c_4$	
Linkage Row	$-a_{41}$	a_{42}	1	-1	≤ 50

The resulting constraint in the optimal solution will have the form $-x_1 a_{41} + x_2 a_{42} + x_3 a_{43} - x_4 \leq 50$.

The interpretation is that the model can satisfy the requirements of activities x_2 and x_3 by using some of the unpriced 50 units on hand, or they can grow hay in activity x_1 for a cost c_1 or buy hay for a price c_4 through activity x_4.

Example 3: An Ending Inventory The problem starts with no initial stocks of hay, but is required to have 100 units in stock at the optimal solution. We now specify activity x_5 as the stock of hay. Note that the minimum stock constraint is defined as a double negative to simplify its specification in Gams.

Objective Row	$-c_1$	c_2	c_3	$-c_4$	c_5		
Linkage Row	$-a_{41}$	a_{42}	1	-1	1	$\leq$	0
Minimum Stock Row					-1	$\leq$	-100

2.11 Linear Transportation Problems

Linear optimization is particularly good at solving problems that minimize the cost of transporting a commodity from defined sources to destinations.

If there are n sources (numbered by i = 1 ... n) and m destinations (numbered by j = 1 ... m), then there are $n \times m$ possible ways to transport the commodity.

The activity (the amount shipped) is therefore defined as x_{ij} and has an associated cost of transport of c_{ij}.
The objective function is therefore:

$$\min z = \sum_{\forall i} \sum_{\forall j} c_{ij} x_{ij}$$

Demand at destinations: The transport problem is constrained by a set of minimum demand quantities at each destination. If the quantity demanded at destination j is defined as b_j, the demand constraint is:

$$\sum_{\forall i} x_{ij} \geq b_j$$

It says "the sum of the amounts that arrive from all sources must be greater or equal to the amount demanded at destination j."

Source "Supply" Constraints: The total amount shipped out of any supply source cannot exceed its capacity. Given a maximum capacity of a_i at source i, the supply constraint is:

$$\sum_{\forall j} x_{ij} \leq a_i$$

It says: "The amount shipped from source i to all destinations must be less than or equal to the amount available at source i."

The complete transportation problem for n sources ($i = 1 \ldots n$) and m destinations ($j = 1 \ldots m$) is:

$$\text{Minimize Total Cost } z = \sum_{\forall i} \sum_{\forall j} c_{ij} x_{ij}$$

$$\text{subject to: } \sum_{\forall j} x_{ij} \leq a_i \text{ and } \sum_{\forall i} x_{ij} \geq b_j$$

Demand Shortage Costs and Transportation Cost: Often problems can be written more realistically by redefining the demand constraints as not having to hold exactly, but to incur Shortage Costs if they are not met. To influence the optimal solution, the outage costs of not having enough product must exceed the transportation and supply costs.

Outage activities are included in the left-hand side of the demand constraints:

$$\sum_{\forall i} x_{ij} + out_j \geq b_j$$

Given an shortage cost of c_{out_j}, the transportation model objective function is now:

$$\sum_{\forall i} \sum_{\forall j} c_{ij} x_{ij} + \sum_{\forall j} c_{out_j} out_j$$

The model now finds the optimum pattern of transportation for the supplies on hand, and calculates the cost minimizing way to spread the shortage among destinations.

2.12 Readings

Williams "Model Building..." Ch 3, pp 20-47; Ch. 5, pp 63-82.
Hazell & Norton, Ch 2, pp 9-31; Ch 3, pp 32-53.

Chapter 3

Solving Linear Models

3.1 Solution Sets

A linear equality constraint defines a line in two space, a plane if it is in three space, and a hyperplane if the constraint is in n dimensions. It follows that a linear constraint (inequality) in "n space" divides the space into two half spaces. Therefore the set of values that satisfy several linear constraints must be common to (or contained in) the intersection of several half-spaces. Fortunately it turns out that the intersection of linear half-spaces is a convex set. Therefore the set of possible solutions, which can satisfy several linear constraints at the same time, is a convex set. This convex set is known as the feasible solution set for the linear inequality constraints, since any point in this set can satisfy all the constraints. We use the properties of convex sets to search over the large set of possible solutions in an efficient way for the optimal solution that maximizes some particular objective.

Example: The following set of inequality constraints that define the solution set:

$$\begin{array}{ccccc} x_1 & & & \leq & 5 \\ 3x_1 & + & 5x_2 & \leq & 30 \\ x_1 & \text{and} & x_2 & \geq & 0 \end{array}$$

We can show these inequality constraints graphically in x_1x_2 space depicted in Figure 3.1. The intersection of the two half-spaces is the feasible solution space. Note that the feasible solution set is a convex set with four extreme points at each corner of the set.

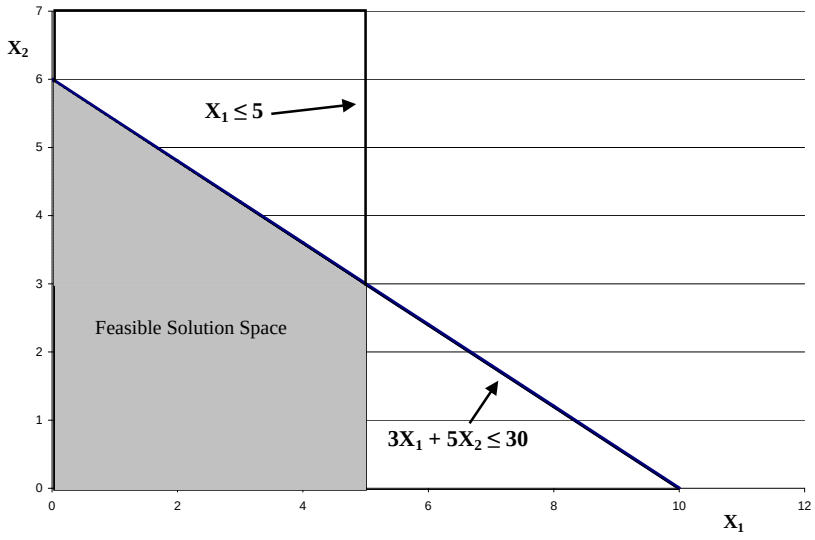


Figure 3.1: Feasible Solution Space

3.1.1 Extreme Points and Basic Solutions

Extreme Point: Given a system $Ax = b$ $(x \geq 0)$ where A is $m \times n$, and $r(A) = m$: if K is the convex set of all solutions to the system, that is, the set of possible $n \times 1$ vectors x that satisfy the system and if x is a basic solution to the system, then a vector x is an extreme point of K.

Basic Solution: Defined as a solution where all the basic variables are non-zero, all other non-basic variables are zero.

Corollary: If a feasible solution exists then there exist a finite number of potentially optimal solutions.

Given that we can be sure that our optimal solution is among the basic feasible solutions we can now concentrate the search for optimal solutions among the finite set of basic feasible solutions. Note that for the feasible solution set, the number of elements at each extreme point equals the number of binding constraints, and also equals the number of non-zero basis activities.

3.2 The Fundamental Theorem of Linear Programming

Given a linear programming problem in the usual matrix algebra form and A is $m \times n$, and $r(A) = m$:

$$\begin{aligned} &\max c'x \\ &\text{subject to} \quad Ax \leq b, \quad x \geq 0 \end{aligned}$$

The Theorem can be summarized by stating that:

1. If a feasible solution exists to the problem, a basic feasible solution to the problem also exists.

2. If an optimal feasible solution exists, then an optimal basic feasible solution to the problem also exists.

In other words, if there are multiple optimal solutions to an LP problem, then among all the optimal solutions at least one of them is an extreme point, and thus a basic solution. For a full mathematical proof see Fang and Puthenpura (1993).

Therefore all we have to do is check which of the basic feasible solutions maximizes the objective function to know that we have checked all the possible candidates for the optimal basic feasible solution.

3.3 The Simplex Algorithm

An algorithm is a set of systematic instructions to the computer that enables us to program it to perform a given task. The Simplex Algorithm is one of the oldest, but still the best algorithm for most problems of linear optimization. Its operation can be summarized as:

1. Change the basis of the problem, and hence the solution, by changing basis vectors.

2. Use the objective function value for a systematic choice of basis vectors that always improve the objective function.

Since the algorithm is driven by the effect of a change of basis on the objective function, to understand its operation we need to analyze the algebra

and economics of a change of basis for the following familiar LP problem:

$$\begin{gathered}\max \Pi = c'x \\ \text{subject to: } Ax = b, \\ x \geq 0,\ dim(A) = m \times n, r(A) = m\end{gathered}$$

Partition A into basis and non-basis matrices denoted respectively B and D. The resulting partition is $A \equiv [B \vdots D]$ where B is the $m \times m$ basis matrix and D is the $m \times (n-m)$ non-basis matrix.

When partitioned, x and c' become:

$$x = \begin{bmatrix} x_B \\ \cdots \\ x_D \end{bmatrix} \text{ and } c' = [c'_B \vdots c'_D]$$

Therefore, the LP problem after partition becomes:

$$\max[c'_B \vdots c'_D] \begin{bmatrix} x_B \\ \cdots \\ x_D \end{bmatrix} \text{ subject to } [B \vdots D] \begin{bmatrix} x_B \\ \cdots \\ x_D \end{bmatrix} \leq b$$

Multiplying out the partitioned matrices results in:

$$\begin{gathered}\max \Pi = c'_B x_B + c'_D x_D \\ \text{subject to: } Bx_B + Dx_D \leq b, \\ x_B \geq 0,\ x_D \geq 0\end{gathered}$$

An optimal basic feasible solution (BFS) has the elements of x_B, that are all non-zero (assuming non-degenerate solutions, a special case that is explained in later chapters), and all the x_D elements are zero. The constraints for a basic feasible solution become:

$$Bx_B + D[0] = b \text{ or, more concisely, } Bx_B = b$$

Since B^{-1} exists by definition of a basis ($m \times m$, with m linearly independent rows), a basic feasible solution to the system is:

$$x_B = B^{-1}b$$

However the point of this analysis is to find the effect of a change of the basic solution on the value of the objective function. Accordingly, the

partitioned x vector is used to write the objective function in terms of basis and non-basis variables.

The partitioned objective function is $\Pi = c'_B x_B + c'_D x_D$ and substituting

$$x_B = B^{-1}b$$

in the solution above:

$$\Pi = c'_B B^{-1} b + c'_D x_D$$

For basic solutions where $x_D = 0$, the objective function is $\Pi = c'_B B^{-1} b$.

What happens to the objective function value (z) when we consider introducing one of the x_D non-basis vectors, say x_j, into the basis? The analogy is the "In group" and "Out group" in high schools where teenage popularity and being cool is very important. One way in which one can judge, and be judged, as to what group you are in is with whom you eat lunch. Assume for mathematical reasons that the lunch bench has a finite dimension, (the rank of the basis matrix) and the number of people who can sit on the bench is limited. Since teenage popularity is fickle, it is quite likely that individuals (vectors) move in and out of the "In group" over time. The point is that if a new individual is popular enough to be admitted to the "In group," someone will have to leave the bench (the basis) and the other members of the group will have to rearrange their seating on the bench with the arrival of the new entrant.

To duplicate this process mathematically, first write out a solution to the partitioned constraints:

$$Bx_B + Dx_D = b \qquad \text{then subtract the non-basis values}$$

$$Bx_B = b - Dx_D \qquad \text{then pre-multiply by } B^{-1}$$

$$\text{and get: } x_B = B^{-1}b - B^{-1}Dx_D$$

Now we substitute this result into the objective function to see the effect of the introduction of a non-basis activity into the basis. This changes the objective function value. Since $\Pi = c'_B x_B + c'_D x_D$ substituting in for x_B from above yields:

$$\Pi = c'_B(B^{-1}b - B^{-1}Dx_D) + c'_D x_D$$

Collecting the terms multiplied by x_D we can now factor out x_D to get:

$$\Pi = c'_B B^{-1} b + (c'_D - c'_B B^{-1} D)x_D$$

$$\text{or } \ldots \Pi_2 = \Pi_{original} + (c_j - z_j)x_D$$

Where c_j is the vector of revenues from the new activity and z_j is the vector of the cost of the needed adjustments in the x_B basis values to accommodate the new x_j vector. Note that difference between the revenues and opportunity cost $c_j - z_j$ is termed the reduced cost and defined as r_j.

Note that $c'_D - c'_B B^{-1} D$ is a vector representing net values of x_D's (non-basis vectors). In the basic equation for Π they're multiplied by vector x_D of zeros. However if one of the x_D activities is set to non-zero value, that is, it is brought into the basis, the objective function will be changed by this amount.

The net change in objective function value from moving a new vector in is:

$\underbrace{c'_D}$	–	$\underbrace{c'_B B^{-1} D}$
Revenue contributed by a new activity from x_D		Change in the x_D values forced to satisfy the constraints defined as: y_j times the unit revenue lost

In other words ...

Marginal new revenue	–	Marginal opportunity cost

That is $c'_B B^{-1} D$ is the cost of moving the old basis values to fit a unit of the x_D vector. In other words, $c'_D - c'_B B^{-1} D$ is the marginal revenue minus the marginal opportunity cost of vector x_D, the incoming non-basic vector.

3.3.1 The Entering Activity Rule

If you want to maximize $c'x$, you select the change of basis that maximizes the difference between revenue and opportunity cost of the new vector. That is, for a maximization problem we select among those vectors that have:

$$c'_D > c'_B B^{-1} D \text{ or } c'_D - c'_B B^{-1} D > 0 \text{ or } MR - MC > 0$$

Using this criterion we only change the basis if we improve the objective function.

The Dual price or "shadow value" is defined as:

$$\lambda' \equiv c'_B B^{-1}$$

Note that λ is the "marginal" associated with the constraint rows in the Gams printout. Substituting the expression for λ above into the equation defines the term z_j where:

$$z = [c'_B B^{-1} D] = \lambda' D$$

The *reduced cost* vector is defined as:

$$r = c' - z' \equiv (c'_D - c'_B B^{-1} D)$$

For a maximization problem, the algorithm rule is:

- If there are any r_j greater than zero ($r_j > 0$), add the vector with the highest r_j.
- If all $r_j \leq 0$, you're at optimal solution.

In a minimization problem, the rule is to bring in the x vector with lowest r_j. If all $r_j \geq 0$, the minimization problem is at the optimum.

3.3.2 Intuition

- r_j is the net benefit of an activity entering the basis and is the "marginal" on the variables in the GAMS printouts.[1]
- c_j is the "benefit" of activity j (the incoming activity) to the objective function.
- z_j is the opportunity cost of moving the current basis values to accommodate incoming unit of x_j, and is equal to the lost revenues from current basis activities. The row vector z with elements z_j is defined as:

$$z = c'_B B^{-1} D = \lambda' D$$

c'_B are returns from x's in basis.

$B^{-1}D$ are the input release quantities for x's.

λ' are shadow values of resources (b)

Alternatively the same z_j value can be expressed in terms of the y_j vector. The new vector y_j is a vector that is a function of the jth non basis activity d_j, and y_j is defined as $B^{-1}d_j$:

[1] See the Rock Music example later in this chapter.

$$z = \begin{array}{rl} c'_B & B^{-1}d_j \\ \uparrow & \uparrow \\ \text{value (revenue) of a unit of each of those } x_i\text{'s} & y_j = \text{The reduction in the output of the current basis activities } x_i \text{ needed to release resources for 1 unit of the entering } x_j. \end{array}$$

$$\Pi = \text{ value of objective function}$$

$$\Pi = c'x = c'_B x_B + c'_B x_D$$

therefore can also be written as:

$$\Pi = c'_B B^{-1} b = c'_B x_B$$

3.4 An Outline of the Simplex Method

The Simplex algorithm optimizes using four critical pieces of information.

1. The Reduced costs of the non-basis activities, defined in the vector r:

$$[\underbrace{c_B B^{-1} D}_{cost} - \underbrace{c'_D}_{revenue}] = \underbrace{z}_{cost} - \underbrace{c}_{revenue} = r. \tag{3.1}$$

   Note that the signs of the r_j elements of r are reversed here — always check the LP package for the definition of r_j.

2. The value of the current Objective function

$$c_B B^{-1} b = c'_B x_B = \Pi \tag{3.2}$$

   Recall $\Pi = c'x = c'_B x_B + c'_D x_D$, but $c'_D x_D = 0$ for a basic solution, since x_D is defined to equal zero for a basic solution.

3. The values of a new parameter y_j, a vector that is calculated for all non-basis vectors that might enter the basis at the next iteration.

$$B^{-1}D = [y_1, y_2, \ldots, y_{n-m}] \tag{3.3}$$

   Note that $\mathrm{B}^{-1}\mathrm{D}$ is an $m \times (n-m)$ matrix, and that each vector y_j is an $m \times 1$ column vector of scalar resource requirements y_{ij}. The

scalar value y_{ij} is the amount of resource i used in the current basis, and needed by a unit of x_j if it enters the basis. That is, the amount that an activity level x_i in the basis must "move over" and transfer to the activity x_j to accommodate x_j entering the basis.

4. The value of the current basis variables x_B is given as:

$$B^{-1}b = x_B \tag{3.4}$$

An overview of the simplex method builds on the four matrix expressions derived above.

Step I First Iteration. Your GAMS program adds slack or surplus variables to convert all the inequality constraints to equalities. The slacks have two important characteristics:

1. Zero values in the objective function coefficient row (c).
2. A constraint matrix which is an identity matrix.

Thus, an initial basis of all slack and surplus variables is always:

1. Of full rank
2. Feasible
3. Zero valued objective function and, therefore, zero opportunity costs z_j for alternative activities.

Step II For each x_j not in the basis, calculate the reduced cost $r_j = c_j - z_j$, which is the net benefit of x_j entering the basis (For the first iteration $r_j = c_j$). For a maximization problem, select the x_j with maximum $r_j > 0$ to enter the basis. If any $r_j > 0$, the problem is not optimal. (For a minimization problem, select the minimum $r_j < 0$ to enter the basis).

Step III The y_j vector is used to select the activity that leaves the basis. y_j is an $m \times 1$ vector of values that show the technical rate of substitution between the basis activities x_{Bi} and the incoming vector x_k chosen in Step II. From equation 3.3 we see that $B^{-1}D$ yields $(n - m)$ y_j vectors, each with m elements. If we pick a non-basis activity and corresponding vector from the matrix D, say d_k, to enter the basis, the corresponding y_k vector will be:

$$y_k = B^{-1}d_k = B^{-1}a_k \tag{3.5}$$

Note that d_k is the k^{th} vector in A thus we can perform the above matrix operation. Using 3.5 we can also write a_k — the input requirements for the incoming vector — as a function of y_k:

$$\text{Since } B^{-1}a_k = y_k, \text{ pre-multiplying by } B \text{ yields } a_k = By_k \tag{3.6}$$

But since B is the basis submatrix of A,

$$B = [a_1, a_2, \ldots, a_m] \tag{3.7}$$

Thus from 3.7the potential incoming vector a_k, selected in Step II can be expressed as a linear combination of the basis vectors. Substituting 3.7 back into 3.6 yields:

$$a_k = a_1y_{1k} + a_2y_{2k} + \ldots + a_my_{mk} \tag{3.8}$$

Now we use the first algebraic trick. Multiply each element in 3.8 by an arbitrary nonzero scalar ϵ:

$$\epsilon a_k = a_1\epsilon y_{1k} + a_2\epsilon y_{2k} + \ldots + a_m\epsilon y_{mk} \tag{3.9}$$

Expanding the basic feasible solution $Bx_B = b$ using 3.7 for B results in:

$$a_1x_1 + a_2x_2 + \ldots + a_mx_m = b \tag{3.10}$$

We now introduce $-\epsilon a_k$ and ϵa_k into the feasible basis in equation 3.10. Note that if $\epsilon \equiv 0$, no basis change occurs.

$$a_1x_1 + a_2x_2 + \ldots + a_mx_m - \epsilon a_k + \epsilon a_k = b \tag{3.11}$$

Now, we want to drive a single activity to zero. First, multiply 3.9 by -1 and substitute the right hand side of 3.9 for $-\epsilon a_k$ in 3.11 to get:

$$a_1x_1 + a_2x_2 + \ldots + a_mx_m + (-a_1\epsilon y_{1k} - a_2\epsilon y_{2k} \ldots - a_m\epsilon y_{mk}) + \epsilon a_k = b \tag{3.12}$$

Factor out the $a_1, \ldots, a_m$ vectors to get:

$$a_1(x_1 - \epsilon y_{1k}) + a_2(x_2 - \epsilon y_{2k}) + \ldots + a_m(x_m - \epsilon y_{mk}) + \epsilon a_k = b \tag{3.13}$$

Note that as the value of the scalar ϵ is increased, the influence of a_k increases and $x_i - \epsilon y_{ik}$ will be driven to zero for some activity i. This activity will leave the basis.

Notes:

1. That by the definition of a basic solution, when an activity is driven to a zero value it leaves the basis.
2. An $m+1$-dimensional object in m-dimensional space is called a "simplex," hence the name of this algorithm.
3. In general, all the "old" basic activities change in quantity, but one activity is driven to zero and exits the basis.
4. The first term to be driven to zero will have the smallest x_i/y_{ik}, since $x_i - \epsilon y_{ik} = 0$ when $\epsilon = x_i/y_{ik}$.

From 3.13, we can see two outcomes of changes in the value of ϵ :

1. If $\epsilon = 0$ we get the old basis in 3.7
2. If ϵ is made large, the importance of the new vector a_k increases, but there is a danger that one of the new variable values $(x_i - \epsilon y_{ik})$ will be driven to a negative value, which would produce an infeasible solution.

Question: How do we select the value of ϵ that will drive one basis activity level exactly to zero, without driving any others to negative values (which would make those activities infeasible?)

Answer: Since

$$x_i - \epsilon y_{ik} = 0 \Rightarrow \epsilon = x_i/y_{ik} \tag{3.14}$$

If the exiting vector is selected as the first basis vector to have its coefficient driven to zero by the entry of a_k in the basis, we will have a new basis and ensure feasibility — remember this requires that x_i is non-negative.

The criterion is therefore (given $y_{ij} > 0$):

$$\min \left\{ \frac{x_{B_i}}{y_{ij}} \right\}_{y_{ij}>0}$$

Note that x_{B_i} and y_{ij} are scalars.

If there are no $y_{ij} > 0$, the problem is unbounded since the trade-offs between the inputs must be negative. Remember that an unbounded problem is one where the constraints are not sufficient to prevent the maximization of the objective function from driving some variables to very large values.

Why is this? If the resource requirement is negative, a situation that implies that $y_{ij} < 0$, then adding x_j to the basis will free up resources

necessary for x_i but may consume resources necessary for other activities. If $y_{ij} < 0$ for all i, then adding x_j to the basis frees up resources and doesn't consume any. This implies you can do it forever, and that there is such a thing as a free lunch.

Step IV Proceed with these iterations (by returning to Step II) changing the basis each time until all $r_j \leq 0$(for a maximizing problem). You now have the optimal solution $X_B = \mathrm{B}^{-1}\mathrm{b}$ and the optimal objective function $\Pi = c_B x_B$.

3.5 A Matrix Simplex Solution Example

As an illustration of matrix manipulations involved in solving for solutions of systems of linear equations, we develop the problem of the music production firm which can promote bands in four groups x_A (Alternative Rock), x_C (Country), x_G (Grunge Rock), and x_H (Hip Hop).

Assume that the firm has two stocks of input needed to produce a successful band, namely promotion AirPlay time (AP) and Recording Studio time (RS). A successful band will bring profits to the firm via CD sales. Both these assets are fixed in their maximum availability, max AP = 620, max RS = 180. In addition, the music firm manager knows how much of each input is required for each type of band. The technology required for the music business is therefore represented by $Ax \leq b$:

$$\begin{array}{cccccc} x_A & x_C & x_G & x_H & & \\ 25 & 32 & 18 & 28 & \leq & 620 \\ 12 & 14 & 17 & 10 & \leq & 180 \end{array}$$

To convert the set of inequality constraints into a set of equality equations we add two more activities for the slacks on AP and RS, respectively S_1 and S_2.

$$\begin{array}{cccccccc} x_A & x_C & x_G & x_H & S_1 & S_2 & & \\ 25 & 32 & 18 & 28 & 1 & 0 & = & 620 \\ 12 & 14 & 17 & 10 & 0 & 1 & = & 180 \end{array}$$

The manager also knows the marginal net revenue (gross margin) for each type of CD. Under current market conditions they are:

$$\begin{array}{rcccccc} & c_A & c_C & c_G & c_H & S_1 & S_2 \\ c' = & [3.5 & 4.2 & 5.6 & 4.8 & 0 & 0] \end{array}$$

The simplex method starts the search for the optimal solution with basic solution that we know will always be feasible but can be improved. The initial basic solution is composed of the slack variables for the binding constraints. In this example, the initial basic solution is composed of the vectors S_1 and S_2. Therefore the initial basis called B_1 is:

$$B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Since the inverse is the same matrix, the basic solution for the basis B_2 is:

$$x_{B_1} = B_1^{-1}b = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 620 \\ 180 \end{bmatrix} = \begin{bmatrix} 620 \\ 180 \end{bmatrix}$$

Since the gross margin from slack resources is zero the objective function is also zero. The value for the objective function is:

$$\Pi_1 = c'_{B_1} x_{B_1} = \begin{bmatrix} 0.0 & 0.0 \end{bmatrix} \begin{bmatrix} 620 \\ 180 \end{bmatrix} = 0.0$$

To select the next activity which will enter the basis we have to calculate the vector of r_j (or c_j - z_j) values for the four music activities that are currently in the non-basic set x_{D_1}. Since the formula for the vector of z_j values is $z = c'_{B_1} B_1^{-1} d$ and since c_{B_1} is composed of zero values, the opportunity cost of using slack inputs to produce CDs is zero. Thus the value of the vector of r_j's is equal to c_{D_1} and:

$$r_{j1} = \begin{bmatrix} 3.5 \\ 4.2 \\ 5.6 \\ 4.8 \end{bmatrix}$$

Since we wish to increase the objective function as fast as possible, we select the largest r_j value, which brings x_G (Grunge Rock) into the basis. To calculate the level at which we can bring in the Grunge rock band and which slack activity leaves the basis, we now calculate the y_j vector for x_G:

$$y_{x_G} = B_1^{-1} d_{x_G} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 18 \\ 17 \end{bmatrix} = \begin{bmatrix} 18 \\ 17 \end{bmatrix}$$

Applying the criteria for the exiting activity also sets the level at which the new activity comes into the basis:

$$\min \left\{ \frac{x_{B_i}}{y_{ij}} \right\}_{y_{ij}>0} = \min \left\{ \frac{620}{18}, \frac{180}{17} \right\} = \min \{34.44, 10.58\}$$

Accordingly, the Grunge rock comes in at a level of 10.58 units which drives the slack on Recording Studio time to zero and out of the basis. The new basis B_2 is composed of S_1 and x_G activities and is:

$$B_2 = \begin{bmatrix} 1 & 18 \\ 0 & 17 \end{bmatrix} \text{ and } B_2^{-1} = \begin{bmatrix} 1 & -\frac{18}{17} \\ 0 & \frac{1}{17} \end{bmatrix}$$

The new solution for the basic activities is:

$$x_{B_2} = B_2^{-1}b = \begin{bmatrix} 1 & -\frac{18}{17} \\ 0 & \frac{1}{17} \end{bmatrix} \begin{bmatrix} 620 \\ 180 \end{bmatrix} = \begin{bmatrix} 429.41 \\ 10.588 \end{bmatrix}$$

The new value for the objective function is:

$$\Pi_2 = c'_{B_2} x_{B_2} = \begin{bmatrix} 0.0 & 5.60 \end{bmatrix} \begin{bmatrix} 429.41 \\ 10.588 \end{bmatrix} = 59.29$$

This level of return is clearly better than the initial solution that did not use the resources at all, but is it the best use that we can make of the limited studio resources?

With the new basis there is a new set of opportunity costs for the resources. Studio space is fully used on the Grunge band under the current allocation, but AirPlay time still has a lot of slack. The new set of r_j values are $r = c'_{D_2} - z'_{D_2} = c'_{D_2} - c'_{B_2} B_2^{-1} D_2$

$$r = c'_{D_2} - \begin{bmatrix} 0.0 & 5.60 \end{bmatrix} \begin{bmatrix} 1 & -\frac{18}{17} \\ 0 & \frac{1}{17} \end{bmatrix} \begin{bmatrix} 25 & 32 & 28 \\ 12 & 14 & 10 \end{bmatrix} = \begin{bmatrix} -0.45 & -0.41 & 1.51 \end{bmatrix}$$

Using the maximum r_j rule, the only music activity for which the marginal contribution to the objective function exceeds that of the Grunge band is x_H (Hiphop). Therefore x_H comes into the basis. The new y_j values are:

$$y_{x_H} = B_2^{-1} d_{x_H} = \begin{bmatrix} 1 & -\frac{18}{17} \\ 0 & \frac{1}{17} \end{bmatrix} \begin{bmatrix} 28 \\ 10 \end{bmatrix} = \begin{bmatrix} 17.412 \\ 0.588 \end{bmatrix}$$

Applying the criteria for the exiting activity also sets the level at which the new activity comes into the basis.

$$\min \left\{ \frac{x_{Bi}}{y_{ij}} \right\}_{y_{ij}>0} = \min \left\{ \frac{429.41}{17.412}, \frac{10.58}{0.588} \right\} = \min \{24.66, 17.99\}$$

This says that the Grunge band should give up their studio time to the Hiphop band, and AirPlay time will still be slack. The next (third) basis has S_1 and x_H and is:

$$B_3 = \begin{bmatrix} 1 & 28 \\ 0 & 10 \end{bmatrix} \text{ and } B_3^{-1} = \begin{bmatrix} 1 & -2.8 \\ 0 & 0.1 \end{bmatrix}$$

The new solution for the basic activities is:

$$x_{B_3} = B_3^{-1}b = \begin{bmatrix} 1 & -2.8 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} 620 \\ 180 \end{bmatrix} = \begin{bmatrix} 116.0 \\ 18.0 \end{bmatrix}$$

The new value for the objective function is:

$$\Pi_3 = c'_{B_3} x_{B_3} = \begin{bmatrix} 0.0 & 4.80 \end{bmatrix} \begin{bmatrix} 116 \\ 18.0 \end{bmatrix} = 86.4$$

Clearly Hiphop is an improvement over the Grunge bands.

The r_j values for the new basis B_3 are as follows:

$$r = c'_{D_3} - z'_{D_3} = c'_{D_3} - c'_{B_3} B_3^{-1} D_3$$

$$r = c'_{D_3} - \begin{bmatrix} 0.0 & 4.80 \end{bmatrix} \begin{bmatrix} 1 & -2.8 \\ 0 & 0.1 \end{bmatrix} \begin{bmatrix} 25 & 32 & 18 \\ 12 & 14 & 17 \end{bmatrix} = \begin{bmatrix} -2.26 & -2.52 & -2.56 \end{bmatrix}$$

Since all the r_j values for the third basis are negative, this tells us that we are at the optimum solution with the Hiphop band using all the Recording Studio time and AirPlay time is in surplus. Note that the LP problem does not solve directly for the optimal Hiphop production since Grunge music made a larger gross margin per CD. We found that Hiphop produced the greatest value per unit of studio time only after we knew studio space was the limiting resource, and had a shadow value.

This problem is specified and solved in Gams on the class webpage. Check the format of the Gams set-up in matrix form. Note that every number on the Gams output has been, or can be, calculated in the matrix operations by hand. Remember that the "Duals" or "Marginals" on the resource constraints are calculated above in the r_j equation as $\lambda = c'_{B_3} B_3^{-1}$.

Check that the matrix calculations agree relatively closely with the Gams printout.

Chapter 4

The Dual Problem

4.1 Primal and Dual Objective Functions

The problem of minimizing the cost of inputs subject to constraints on a minimum output level is equivalent to the problem of maximizing profit subject to production technology and constraints on the total input available. Thus every optimization problem can be posed in its Primal or Dual form. For every Primal problem there exists a Dual problem which has the identical optimal solution. So far in this course we have only dealt with the Primal form of the problem since its intuitive explanation is easier. The standard form of the twin Primal and Dual problems is written as follows:

Primal	Dual
$\max c'x$	$\min \lambda' b$
s.t. $Ax \leq b$	s.t. $A'\lambda \geq c$
$x \geq 0$	$\lambda \geq 0$

where x is a $n \times 1$ vector of Primal variables and λ is an $m \times 1$ vector of Dual variables.

The Primal objective functions asks the maximum value of a firm's output while the Dual objective function asks the minimum acceptable price that one can pay for the firm's assets.

The Dual specification of a problem is particularly useful when:

1. When the Dual specification is simpler to solve than the Primal specification.

2. When you know production costs but not production technology.

4.2 The Economic Meaning of the Dual

The Dual variables λ_i are elements in the vector of imputed marginal values of the resources b_i, given that these resources are in short supply (or "binding") with respect to the objective function. Equivalent intuitive interpretations are the opportunity costs of not having the last unit of resource, how much you'd pay for one more unit of the resource b_i, and/or the marginal effect on the objective function of a small change in resource availability.

$$\lambda_i = \frac{\partial(obj)}{\partial b_i} \geq 0$$

If the constraint isn't binding, λ_i is always equal to zero by the Kuhn-Tucker complementary slackness conditions (the Kuhn-Tucker conditions can be found in most economic optimization texts or specifically Hazell and Norton (1986)Paris page 140).

4.2.1 Dual Objective function

The Dual objective function $\lambda' b$ is equal to the sum of the imputed values of the total resource stock of the firm. It is the sum of money that you would have to offer a firm owner for them to agree to a buy-out. For example, what would you pay for the Nike corporation ? Nike owns very few factories, but develops and markets a wide and profitable range of shoes, mainly on the strength of their trademark and sports celebrity advertising. Clearly, the valuation of the company has to be based on these contracts rather than the physical plant used in shoe manufacturing.

4.3 Dual Constraints

The Dual constraints $A'\lambda \geq c$ can be interpreted as defining the set of prices λ for the fixed resources or assets (b) of the firm that would yield at least an equivalent return to the owner as producing a vector of products (x) from these resources, which can be sold for prices (c) . The Dual constraint is a "greater than or equal to" because you can pay too much for a productive input, but you cannot pay too little. Market forces ensure that input prices reflect the value of the input in the final, saleable product.

These economic facts are reflected in the matrix algebra development. Post-multiplying the transpose of the technical input requirement matrix A by the Dual prices λ results in an $m \times 1$ vector of marginal opportunity

costs for each of the n potential production activities. For example, $A'\lambda \geq c$, where $A'\lambda$ = vector of marginal opportunity costs of production.

For a single production activity x_i, its opportunity cost of production for a vector of Dual prices λ is

$$a_i'\lambda = (\text{column } i \text{ of } A)'(\lambda \text{ vector}) = (a_{1i} \dots a_{mi}) \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$$

Where:

$a_i'\lambda$ is the imputed cost of the last unit of x_i produced or the cost of producing a unit of x_i if I have to pay λ for resources b.

$(a_{1i} \dots a_{mi})$ are the coefficients on resources required for one unit of x_i production.

$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$ is the marginal imputed value per unit of the resources used to produce x_i.

The constraint $A'\lambda \geq c$ for the production problem says that the Marginal Opportunity cost of producing a different vector of x must be greater than or equal to the marginal revenue (c) for each of the x actually in production. The profit-maximizing owner actually produces where the opportunity cost is equal to the firm value of $\lambda' b$.

4.3.1 Music Production Example

In Section 3.5, we found that the optimal solution resulted in HipHop. The resources used in this production can be calculated from the coefficients for $A'\lambda$ previously derived and measured against the marginal revenue on page 40.[1] Let's check this result again, this time from the Dual perspective.

HipHop production:

$$Cost_{HipHop} = Cost_{AirPlay} * Time_{AirPlay} + Cost_{RecordingStudio} * Time_{RecordingStudio}$$

$$= 28 * 0.0 + 10 * 0.48 = \$4.80$$

This is equal to the marginal revenue of \$4.80 for HipHop and therefore Marginal Opportunity Cost = Marginal Revenue.

[1] For the values of A, recall page 40. For the values of $\lambda = c'_{B_3} B_3^{-1}$, recall page 43.

Alternative production:

$$Cost_{Alternative} = Cost_{AirPlay} * Time_{AirPlay} + Cost_{RecordingStudio} * Time_{RecordingStudio}$$
$$= 25 * 0.0 + 12 * 0.48 = \$5.76$$

Since the marginal revenue from an Alternative CD is \$3.50, for production of Alternative CDs, the marginal opportunity cost exceeds the marginal revenue and Alternative CDs are not produced.

4.4 Showing the Primal/Dual Linkage

We want to show that the Primal optimality conditions imply that the Dual constraints must hold.

4.4.1 Primal Problem

$$\begin{aligned} &\max c'x \\ &\text{subject to: } Ax \leq b,\ x \geq 0,\ dim(A) = m \times n,\ r(A) = m \end{aligned} \tag{4.1}$$

Suppose there exists a basic feasible solution $x' = [x_B \vdots 0]$ with a corresponding partition of A into the basis matrix B and non-basis matrix D. It follows that:

$$x_B = B^{-1}b \tag{4.2}$$

The reduced cost vector r' equals:

$$r' = c'_D - c'_B B^{-1} D \tag{4.3}$$

If the x_B solution to 4.1 is optimal, then the scalar reduced cost for a maximization problem (a minimization problem would have the opposite sign) is:

$$r_j = c'_D - c'_B B^{-1} d_j \leq 0 \quad \forall j \tag{4.4}$$

Or — equivalently — by stacking the d_j vectors together to yield the matrix D, the right side of 4.4 becomes:

$$c'_B B^{-1} D \geq c'_D \tag{4.5}$$

Defining the $1 \times m$ vector of Dual variables

$$\lambda' \equiv c'_B B^{-1} \tag{4.6}$$

allows us to transform 4.5 into $\lambda' D \geq c_D$ at the optimum. This also means that:

The shadow values of resources you'd need to bring x_D into the basis, i.e., the opportunity costs of bringing activities into the basis, times the input requirement.	$\geq$	Marginal benefits of bringing x_D into the basis (since $c_j x_j$ is the benefit from x_j, $c_j x_j = 0$ when $x_j = 0$ or when x_j is a non-basic activity).

4.4.2 Dual Problem

The Dual version of the primal problem 4.1 is specified as:

$$\begin{aligned} &\min \lambda' b \\ &\text{subject to: } A'\lambda \geq c \text{ and } \lambda \geq 0 \end{aligned} \tag{4.7}$$

We want to show that the Dual solution vector λ defined in 4.6 is both:

Feasible: That is, it satisfies the constraints in 4.7.

Optimal: By showing that the optimal Dual objective function value is equal to the optimal Primal objective function value.

Feasibility of λ

$$\lambda' A = [\lambda' B \vdots \lambda' D] = [c'_B B^{-1} B \vdots c'_B B^{-1} D] = [c'_B \vdots c'_B B^{-1} D] \tag{4.8}$$

by substituting the inequality condition for optimality $c'_B B^{-1} D \geq c'_D$ in 4.5 we get:

$$[c'_B \vdots c'_B B^{-1} D] \geq [c'_B \vdots c'_D] = c' \tag{4.9}$$

combining 4.8 and 4.9:

$$\lambda' A \geq c' \text{ transposing yields } A'\lambda \geq c \tag{4.10}$$

Thus, by definition, λ is a feasible solution to the constraints $A'\lambda \geq c$.

Optimality of λ

Using the Dual objective function in 4.7 and our definition of λ in 4.6, By substituting the definition of $x_B = \mathrm{B}^{-1}\mathrm{b}$ from 4.2, we get:

$$\lambda' b = c'_B B^{-1} b = c'_B x_B \tag{4.11}$$

Therefore, the Dual Objective Function value = Optimal Primal Objective Function value and the Dual value λ is optimal by the Strong Duality Theorem.[2]

The Point: If the standard Primal LP problem has an optimal basic feasible solution with a basis B, then the vector $\lambda' \equiv c'_B B^{-1}$ is a feasible and optimal solution for the corresponding Dual problem.

4.5 Numerical Matrix Example — Yolo Model

The A matrix in Yolo is 4×4. The "*" denotes the basic solution activities and binding constraints for the optimal solution.

		Alfalfa	*Wheat	Corn	*Tomato
	*Land	1.0	1.0	1.0	1.0
A =	Water	4.5	2.5	3.5	3.25
	Labor	6.0	4.2	5.6	14.0
	*Contract	0.0	0.0	0.0	33.25

The optimal solution to Yolo has two binding constraints (Land and Contract), and two non-zero activities (Wheat and Tomato). We collapse the A matrix to the basis matrix B by removing the rows and columns that do not have * and do not constrain the optimal solution. That is, if the rows are not binding, their coefficients are not in the basis.

The optimal basis matrix is therefore 2×2 — ignoring the slack Labor and Water constraints. The optimal basis B is:

$$B = \begin{bmatrix} 1 & 1 \\ 0 & 33.25 \end{bmatrix} \qquad D = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

$$B^{-1} = \begin{bmatrix} 1 & -0.03007 \\ 0 & 0.03007 \end{bmatrix} \qquad b = \begin{bmatrix} 600 \\ 6000 \end{bmatrix} c_B = \begin{bmatrix} 160 \\ 825 \end{bmatrix}$$

Note that for B to be invertible the number of activities in the basis must equal the number of binding constraints.

Now that we know the optimal basis, let's find the various algebraic quantities at the optimum. Finding these quantities in GAMS will be left as an exercise.

[2]The Strong Duality Theorem can be informally summarized as $\lambda' b = c' x$ ... if and only if x and λ are optimal solutions to the Primal and Dual problems respectively.

- Optimal Primal Solution: equation 4.2 is $x_B = B^{-1}b$. Forget about x_D; they are all zeroes.

$$x_B = \begin{bmatrix} 1 & -0.03007 \\ 0 & 0.03007 \end{bmatrix} \begin{bmatrix} 600 \\ 6000 \end{bmatrix}$$

$$= \begin{bmatrix} (1)(600) + (-0.03007)(6000) \\ (0)(600) + (0.03007)(6000) \end{bmatrix} = \begin{bmatrix} 419.58 \\ 180.42 \end{bmatrix}$$

- Optimal Dual Solution (GAMS: Marginals on resources): equation 4.6 is $\lambda' = c'_B B^{-1}$. λ values are greater than zero for binding constraints and equal to zero for non-binding constraints.

$$\lambda' = c'_B B^{-1} = [160\ 825] \begin{bmatrix} 1 & -0.03007 \\ 0 & 0.03007 \end{bmatrix}$$

$$\lambda' = \{[(160)(1) + (825)(0)]\ [(160)(-0.03007) + (825)(0.03007)]\}$$

$$\lambda' = [160\ 20]$$

- Reduced Cost (GAMS: Marginals on activities): equation 4.4 can be stated as $r_j = c_j - z_j$ and $z'_j = c'_B B^{-1} D = \lambda' D$, where c_j are elements of the vector c_D. r_j values are less than zero for non-basic activities and equal to zero for activities in the basis.

$$z'_j = [160,\ 20] \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = [160\ 160]$$

$$c_D = [121, 135]$$

$$\therefore r' = [121\ 135] - [160\ 160] = [-39\ -25]$$

- Optimal Primal Objective Function: equation 4.1 is — in terms of B alone now — $\Pi = c'_B x_B$. Note that optimality implies the maximum value.

$$\Pi = [160\ 825] \begin{bmatrix} 419.549 \\ 180.451 \end{bmatrix} = [67,127.84 + 148,872.07] = 216,000$$

- Optimal Dual Objective Function: equation 4.7 is $\lambda' b$, which will also be minimized at the optimum.

$$\lambda' b = [160\ 20] \begin{bmatrix} 600 \\ 6000 \end{bmatrix} = [96,000 + 120,000] = 216,000$$

4.6 Parametric Analysis

Parametric analysis is the main method for obtaining policy results from optimization models. By changing the output prices or quantities of input available over a range of values we can generate empirical estimates of Derived Demand for inputs to the modeled process (§ 4.6.1) or the Supply Function from the modeled process (§ 4.6.2).

We may also be concerned with the Sensitivity of the model. Sensitivity tests the stability of conclusions from the model when we change a constraint or coefficient value. This is useful when we are uncertain of the exact value of a parameter and need to know whether knowing the value precisely is important.

4.6.1 Generating Derived Demand Functions for Inputs

The resource availability vector b is parameterized, that is, changed by small incremental values over a specified range. At each point the model is optimized and the value of the resource λ is plotted against the amount of the input b to form the derived demand function. Given the problem:

$$\begin{aligned} &\max c'x \\ &\text{subject to} \quad Ax \leq b, \quad x \geq 0 \end{aligned}$$

The optimal solution is $x = [x_B \ 0]$ if we assume the solution is not degenerate. Degeneracy occurs in the situation where constraints bind simultaneously, and there are more binding constraints than positive valued activities.

$$\therefore x_B = B^{-1}b \text{ and } \Delta x_B = B^{-1}\Delta b$$

Suppose Δx_B is small enough so there is no change in the basis, then we can define new x values ($\tilde{x}$) as:

$$\tilde{x} = [x_B + \Delta x_B \vdots 0]$$

This results in an update from the old objective function value to a new value:

$$\Pi = c'_B x_B \text{ (old)} \rightarrow \tilde{\Pi} = c'_B(x_B + \Delta x_B) \text{ (new)}$$

$$\Delta\Pi = \tilde{\Pi} - \Pi = c'_B \Delta x_B = c'_B B^{-1}\Delta b$$

$$\Delta\Pi = \lambda \Delta b$$

By selecting the b_i^{th} activity and solving for λ, we find the impact on the objective function of a marginal change in resource availability, given no change in the basis B. This is:

$$\frac{\Delta\Pi}{\Delta b_i} = \lambda_i$$

Note that from the equation $\lambda' = c'_B B^{-1}$ we see that for linear problems, the value of λ does not change with Δb unless the basis (B) changes, or the objective function coefficients c_B are changed by parameterization. To get the derived demand for an input, the output price is not changed.

There are two important results from this analysis:

1. For LP problems, the marginal change in the objective function for changes in resources ($\Delta\Pi/\Delta b$) is constant until Δb is large enough to cause a change in basis.

2. When the basis changes, the value of λ changes due to new values in c'_B and B. This gives rise to the stepwise response to parameterization shown in Figure 4.1.

.

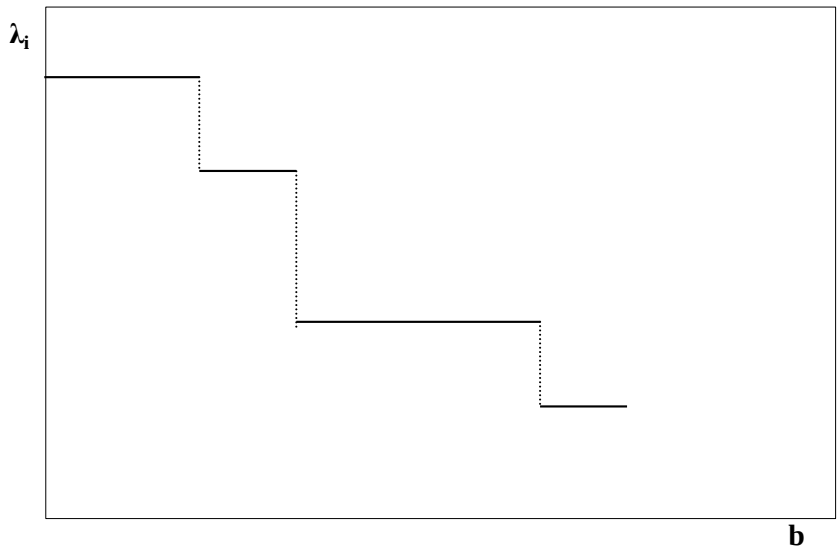


Figure 4.1: Derived Demand for an input found by parameterizing the RHS.

4.6.2 Generating Supply Functions for Outputs

Similarly to the demand functions, supply functions are obtained by parameterizing a single objective function coefficient (say c_i) by small incremental values over a specified range. At each point the model is optimized and the quantity of the output produced (x_i) is plotted against the price to form the supply function. Clearly, for LPs there will only be a change in the product x_i when there is a change of basis caused by the change in c_i. Accordingly, the reaction when x_i is zero and therefore a non-basis activity is going to be different from the situation when $x_i > 0$ and is in the basis.

Non-Basis activity supply parameterization

If $x_i = 0$, then from the optimality conditions we know that the reduced cost for this activity (assuming maximization) is $r_i = c_i - z_i < 0$.

Since $z_i = \lambda' d_i$ and $\lambda' = c'_B B^{-1}$, the value of z_i is unchanged by a change in c_i because it is not in the vector c_B. However, if the value of c_i is increased by less than r_i, then the reduced cost will still be negative (< 0) though reduced in numerical value, and x_i will still be a non-basis variable with a value of zero.

If the increase in c_i is greater than r_i, that is: $\Delta c_i > r_i$, then r_i will become positive and this will induce a change of basis which will bring the x_i activity into the basis with a positive value.

Basis activity supply parameterization

Consider parameterizing the returns from a basis variable x_k. The coefficient c_k is now part of the vector c_B. Since x_k is an optimal basic variable, we know that $r_k = c_k - z_k = 0$.

Again, $z_k = \lambda'$ d_k and $\lambda' = c'_B \mathrm{B}^{-1}$ Thus the value of z_k is changed by a change in c_k, and from the simplex criteria we know that:

- There is some value of c_k where the basis will change, since Δc_k will cause $\Delta\lambda$, and some activities will leave the basis.
- The new basis will have a larger value for x_k since the reduced cost with the higher c_k value in the basis value vector will increase the value of r_k.

Thus whether we start with a non-basis or basic activity, parameterizing the objective coefficient over a discrete range will result in a series of stepped increases in the quantity of x_k in the optimal solution. The resulting supply function will be an upward sloping supply function.

4.7 Complementary Slackness

4.7.1 Primal Complementary Slackness

The concept of Complementary Slackness (CS) applies to both the Primal and Dual problems, but is easier to conceptually understand in the Primal case. The formal proofs are developed for the Dual complementary slackness case.

Given the standard LP problem of:

$$\begin{aligned} &\max c'x \\ &\text{subject to} \quad Ax \leq b, \quad x \geq 0 \end{aligned}$$

The Primal complementary slackness theorem says that for the j^{th} constraint:

$$\text{If } b_i - a_i'x > 0 \text{ then } \lambda_i = 0$$

This is summarized as $(b_i - a_i'x)\lambda_i = 0$ or written in summation form as $(b_i - \Sigma_j a_{ij}x_j)\lambda_i = 0$.

This says that if the total use of the i^{th} input in all productive uses is less than the amount of input available, then the marginal scarcity value of additional units of input is zero.

Note the special case in which both $b_i - \Sigma_j a_{ij}x_j$ and λ_i are zero. This says that the constraint is binding, but an additional unit of b_i will not add to the objective function value.

An alternate definition says if $\lambda_i > 0$, then $(b_i - \Sigma_j a_{ij}x_j) = 0$, i.e., the shadow value cannot be positive if the constraint is not binding.

To get an intuitive idea of Primal complementary slackness, imagine that you are sunbathing on a large, sandy, uncrowded and hot beach. If you are offered additional sand for $10 you are unlikely to purchase it, as your sand constraint is not binding. However if you are offered a cold drink (you do not have any), you are probably willing to pay a price higher than the supermarket price to get a cold drink right there, right then.

4.7.2 Dual Complementary Slackness

The Standard Dual Problem is:

$$\begin{aligned} &\min \lambda'b \\ &\text{subject to} \quad A'\lambda \geq c, \quad \lambda > 0 \end{aligned}$$

Theorem

Let x and λ be feasible solutions to the standard Primal and Dual problems. A necessary and sufficient condition for them to be optimal is:

1. If $x_i > 0$ then $\lambda' a_i = c_i$.

2. If $\lambda' a_i > c_i$ then $x_i = 0$.

These conditions can be combined in matrix form as: $(c' - \lambda' A)x = 0$. This is a very important equality.

The "Free Lunch" Theorem

Note x_i is never < 0 by definition of feasibility, and c_i is never $> \lambda' a_i$ by Dual constraint that $\lambda' A \geq c$. The intuition for the direction of this constraint is that under economic (and optimizing) assumptions $\lambda' A < c$ cannot exist; if productive resources were priced below their immediate productive value one could make instantaneous capital gains (A "Free Lunch").

Proof

Sufficiency: The logic of sufficiency can be summarized as "If the conditions hold then the solution must be optimal".

Assume that $(c' - \lambda' A)x = 0$ holds. Rewrite it as:

$$\lambda' A x = c' x$$

Since $Ax = b$ for any basic feasible solution, the sufficient conditions can be written as $\lambda' b = c' x$. By the Duality theorem λ and x must be optimal if $\lambda' b = c' x$.

Necessity: The logic of necessary conditions can be summarized as: "If the problem is optimal, then the conditions hold".

Assume optimality.
If λ^* and x^* are optimal solutions, then from $(c' - \lambda' A)x = 0$ we get:

$$\lambda^{*'} b = c' x^*$$

Dropping the * notation and substituting for b gives $\lambda' A x = c' x$ and therefore $(c' - \lambda' A)x = 0$.

4.7.3 Demonstrating Complementary Slackness

Case I: Show that: $x_i > 0 \Rightarrow c_i' - \lambda' a_i = 0$

In the production problem, we know that if something is produced ($x_i > 0$), then its marginal revenue equals its marginal opportunity cost at the optimum. This means that $c_i' - \lambda' a_i$=0.

We select the x_B (Basis) partition of x, since only $x_i > 0$ are in the basis with an optimal solution and all their r_i values are zero. Note that the r_j for each x_i is the incremental benefit of bringing x_i into the basis. When x_i is already in the basis, r_j is equal to zero. This reflects the common sense notion that you are not willing to pay more for an activity that you are already enjoying (but may be willing to pay to get it back if it is taken from you).

For the basis vectors B, the vector of $r_i = (c_B' - c_B') = 0$. Rewriting this gives $r_i = (c_B' - c_B' B^{-1} B)$. (Remember, $r_j = (c_D' - c_B' B^{-1} D)$ when x_j is not in the basis). It follows that:

$$c_B' - \lambda' B = 0 \text{ since } r_i = 0 \text{ for basis activities and } \therefore c_B' = \lambda' B$$

Note that since the non-basis x_D are zero by definition, $c' - \lambda' A$ becomes $c' - \lambda' B$.

Case II: Show that: $x_i = 0 \Rightarrow c_i' - \lambda' a_i < 0$

In a production problem, if a product is not produced at the optimum, its net revenue must be less than its opportunity cost. If $x = 0$, it implies that it is not in the basis. Therefore, the partition of $c' - \lambda' A$ for non-basis activities (x_D) becomes $c_D' - \lambda' D$.

Recall that $r_j = c_j - z_j = c_D' - c_B' B^{-1} D = c_D' - \lambda' D$ so the vector of $r_j = c_D' - \lambda' D$. Since all $r_j < 0$ for the non-basic x_j values found at the optimal solution, we find that (note the strict inequalities):

$$c_D' - \lambda' D < 0 \Rightarrow \lambda' D > c_D' \text{ or } c_j' - \lambda' d_j < 0$$

Corollary: If λ and x are optimal Dual and Primal solutions, this implies that $(c' - \lambda' A)x = 0$ for all values.

Example: Auto Dealer Hype

"Trust me, this car is selling below my invoice cost."[3]

Here, we are optimizing from the buyer's perspective. The buyer wants to minimize the price for a car with certain attributes. c_j is the dealer's invoice cost. Since cars are traditionally sold as a vector of attributes and options on a base unit, a "loaded" car will be composed of a base unit, air conditioning, FM stereo, "dealer's prep," etc. These attributes of the car are represented by the $m \times 1$ vector a_i. The dealer's problem is to assign a set of prices λ' to the a_i vector such that $\lambda' a_i - c_i \geq 0$.

Remember that prices on a car lot are always negotiable. The buyer wants to minimize the cost of a car with a selection from the $m \times 1$ set of attributes. The buyer wants to set their λ values to minimize $\lambda' b$. The dealer wants to convince you that — at the asking λ price vector — the complementary slackness theorem does not hold and you are getting a bargain " below dealer cost".

The complementary slackness theorem says that if the dealer truly has set prices λ below his invoice, then $\lambda' a_i - c_i < 0$ and the dealer will set $x_i = 0$. This is more often the case as the advertised below-invoice car always seems to have "just sold" when you get to the dealer.

4.7.4 Duality and Reduced Cost

Since λ is the vector of opportunity costs on the binding resources, and r_j elements form a vector of revenue minus activity opportunity costs, they must be connected to each other.

$$\lambda = c'_B B^{-1}$$

$$r = c'_D - c'_B B^{-1} D = c'_D - \lambda' D$$

Each individual element in the vector r, is $r_j = c_j - \lambda d$ where:

$$\lambda' d_j = [\lambda_1 \dots \lambda_m] \begin{bmatrix} \mathrm{d}_{ji} \\ \vdots \\ \mathrm{d}_{jm} \end{bmatrix}$$

Therefore $\lambda' \mathrm{d}_j$ is equal to the shadow value of inputs required to produce a unit of activity x_j that is not currently in the basis. For an optimal solution, we would expect that all values in $c'_D - \lambda' D \leq 0$.

[3] If you believe this — stop reading this book and try to sell it for more than the retail price.

4.8 Exercises

1. Check the matrix derivations in Section 4.5 against the GAMS computer printout for the Yolo model.

Chapter 5

PMP Calibration of Optimization Models

Even with a constraint structure and parameters that are theoretically correct, it is highly unlikely that a linear model will calibrate closely to the base year data. This is inherent in the structure of models that are, by definition, simplified abstractions of the real systems. In the process of abstracting and simplifying a real system the model loses information and needs to be verified against actual behavior. Just as in econometric modeling there are two phases of model development, namely estimation and prediction, in optimization modeling there are the two phases of model calibration and simulation.

Fundamentally, the calibration process is one of using a hypothesized function and data on input and output levels in the base year to derive specific model parameter values that "close" the model. By closing the model, we mean that the calibration parameters lead to the objective function being optimized for the base year conditions at the observed base year values. Like econometrics, calibration methods assume that actions by individuals are motivated by their optimization of some set of criteria. It follows that if we can derive the parameters, that when optimized, lead to the observed actions, we have derived the parameters most likely to have been used by the decision maker.

Given that most optimization models are specified to optimize profits to economic decision makers, they can in theory be calibrated from the demand (price) or supply (cost) sides. For many years builders of large quadratic programming models have calibrated model prices by deriving demand parameters. However demand side calibration can only help when the model

is on a scale large enough so that changes in the levels of output change product prices. In addition, because optimization models are usually specified with several regions supplying a single market demand, a small number of market demand parameters cannot calibrate the cropping patterns over several regions.

5.1 Calibrating on the Model Supply Side

This section contains a short overview and critique of the traditional methods of calibrating the supply side of optimization models using linear constraints. The shortcomings of the constraint calibration methods lead to a discussion of ways to derive nonlinear supply functions that are based on observed behavior by decision makers, but calibrate the model in general. These methods are termed "Positive Mathematical Programming" (PMP) since they are based on positive inferences from the base year data, rather than normative assumptions (Howitt, 1995b).

5.1.1 A Review of Supply Side Constraint Calibration

Programming models should calibrate against a base year or an average over several years. Policy analysis based on normative models that shows a wide divergence between base period model outcomes and actual production patterns is generally unacceptable. However calibration by adding linear constraints is also unsatisfactory. Models that are tightly constrained can only produce the subset of normative results that the calibration constraints dictate. The policy conclusions are thus bounded by a set of constraints that are expedient for the base year, but are often inappropriate under policy changes. This problem is exacerbated when the model is built on a regional basis with very few empirical constraints, but a wide diversity of crop production. For example, the Yolo model presented in the previous chapter is highly simplified with only four cropping activities, but still requires unrealistic constraints on the amount of labor and water that can be employed to produce all four crops in the optimal solution. While labor requirements do vary, they are subject to a labor supply which is rarely constrained. The supply function will increase labor available in a given quarter at an increased cost of overtime or operations performed by custom operators. To suggest that a more profitable crop in some policy scenario will never be grown beyond a certain limit because of labor constraints is to radically depart from the actual empirical solution. In addition to the labor constraint, the water constraint may also be artificially constraining if the farmers have access

to groundwater or river pumping. More complex linear production models require more complicated constraint structures to reproduce the observed cropping pattern. In many traditional optimization models the proportions of crops are restricted by "rotational" constraints or "flexibility" constraints. These constraints determine the optimal solution not only for the base year, for which they are appropriate, but also for policy runs that attempt to predict the outcome of changed prices, costs or resource availability. The solution of the model under policy runs is therefore significantly restricted by the base year solution constraints.

This section is a brief overview of some of the past calibration methods in mathematical programming models. For a discussion of other calibration methods see Hazell and Norton (1986) or Bauer and Kasnacoglu (1990). It is worth noting that no single linear constraint calibration method has proved sufficiently satisfactory to dominate the mathematical programming literature.

Previous researchers such as Day (1961) have attempted to provide added realism by imposing upper and lower bounds to production levels as constraints. McCarl (1982) advocated a decomposition methodology to reconcile sectoral equilibria and farm level plans. Both approaches require additional micro level data and result in calibration constraints influencing policy response.

Meister et al. (1978), in their national quadratic programming model, specify 103 producing regions and aggregate the results to 10 market regions. Despite this structure, they note the problem of overspecialization and suggest the use of rotational constraints to curtail the overspecialization. However, it is comparatively rare that agronomic practices are fixed at the margin, more commonly they reflect net revenue maximizing tradeoffs between yields, costs of production, and rotational externalities between crops. In this latter case, the rotations are themselves a function of relative resource scarcity, output prices, and input costs.

Hazell and Norton (1986) suggest six tests to validate a sectoral model. First, a capacity test checks whether the model constraint set allows the base year production. Second, a marginal cost test ensures that the marginal costs of production, including the implicit opportunity costs of fixed inputs, are equal to the output price. Third, they suggest a comparison of the Dual value on land with actual rental values. Three additional comparisons of input use, production level, and product price are also advocated. Hazell and Norton show that the percentage of absolute deviation for production and acreage over five sectoral models ranges from 7 percent to 14 percent deviation. The constraint structures needed for this validation are not de-

fined.

In contrast, the PMP approach achieves exact calibration in acreage, production and price (Howitt, 1995b). The PMP approach was applied to the Turkish Agricultural Sectoral Model (TASM) which is one of the models listed by Hazel and Norton. The resulting PMP version of TASM calibrated exactly with the base year and showed consistency in the parameters over the seven years used for calibration (Bauer and Kasnacoglu, 1990). A PMP calibrated model was developed to analyze the effects of large inter-sectoral water reallocations in California (USBR Staff, 1997). The model termed the Central Valley Production Model (CVPM) was tested by out of sample predictions of regional crop acreage changes during a recent drought period. The CVPM predictions were close with three contract crops (Sugar beet, Tomatoes, and Subtropical orchard) having a 14-23 % error. The remaining nine crops had prediction errors below 7%. Regional crop acreage was predicted for eleven regions. For all of the regions the crop acreage predictions had errors below six percent.

The calibration problem for farm level, regional, and sectoral LP models can be mathematically defined by the common situation in which the number of binding constraints in the optimal solution (m) are less than the number of non-zero activities (n) observed in the base solution. If the modeler is fortunate enough to have empirical data to specify, *a priori*, a realistic constraint set that reproduces the optimal base year solution, then additional model calibration may be redundant. The PMP approach is developed for the majority of empirical model builders who, for lack of empirical justification, data availability, or cost, find that the empirical constraint set does not reproduce the base year result. The LP solution is an extreme point of the binding constraints. In contrast, the PMP approach views the optimal farm production as a boundary point which is a combination of binding constraints and first order conditions.

Relevant constraints should be based on either economic logic or the technical environment under which the agricultural production is operating. Constraints should generally represent allocatable input quantities, actual rotational limits and relevant policy constraints. When the basis matrix of valid empirical constraints has a rank less than the number of observed base year activities, the resulting optimal solution will suffer from overspecialization bias of production activities compared to the base year.

A root cause of these problems is that linear programming was originally used as a normative farm planning method where full knowledge of the production technology is assumed. Under these conditions, any production technology can be represented as a linear Leontief specification, subject

to resource and stepwise constraints. For aggregate policy models, this normative approach over-simplifies the production and cost technology due to inadequate knowledge. In most cases, the only regional production data available is an average or "representative" figure for crop yields and inputs. This common data situation means that the analyst using linear production technology in programming models is attempting to estimate marginal behavioral reactions to policy changes, based on average data observations. The average conditions can be assumed to be equal to the marginal conditions only where the policy range is small enough to admit linear technology over the whole range.

Two broad approaches have been used to reduce the specialization errors in optimizing models. The demand-based methods have used a range of methods to add risk or endogenize prices. These have reduced the calibration problem, but substantial calibration problems remain in many models (Just, 1993).

A common alternative approach is to constrain the crop supply activities by rotational (or flexibility) constraints or step functions over multiple activities (Meister et al., 1978). In regional and sectoral models of farm production the number of empirically justifiable constraints are comparatively few. Land area and soil type are clearly constraints, as is water in some irrigated regions. Crop contracts and quotas, breeding stock, and perennial crops are others. However, it is rare that some other traditional programming constraints such as labor, machinery, or crop rotations are truly restricting to short-run marginal production decisions. These inputs are limiting, but only in the sense that once the normal availability is exceeded, the cost per unit output increases due to overtime, increased probability of machinery failure or disease. In this situation the analyst has a choice. If the assumption of linear production (cost) technology is retained, the observed output levels infer that additional binding constraints on the optimal solution should be specified. Fixed proportion rotational constraints are a common example of this approach.

An alternative explanation of the situation, where there are more crop activities than constraints, is that the profit function is nonlinear in land for most crops, and the observed crop allocations are a result of a mix of unconstrained and constrained optima. The equilibrium conditions for this case are satisfied if some, or all, of the cropping activities have decreasing returns to land as the crop acreage is increased. The most common reasons for a decreasing returns per acre are declining yields due to heterogeneous land quality, risk aversion, or increasing costs due to restricted management or machinery capacity.

5.2 Positive Mathematical Programming

The positive mathematical programming (PMP) approach is being adopted quite rapidly for agricultural sector models. In the introduction to their book on "Agricultural Sector Modelling and Policy Information Systems," Heckelei et al. state that they received :

> a rich supply of about 60 proposals from 16 countries reflecting the breadth of work directed to agricultural sector modeling and policy information systems. Because the sample of proposals is indicative of current emphasis in research, it is worth mentioning that, from a methodological point of view almost 25% of the supply might be called "econometric partial' analysis, another 25% "programming models" (half of which relying on PMP) whereas the remaining half of the proposals covered a multitude of quantitative methods ...

In European agricultural economics, the PMP calibration method has become a widely accepted standard method for agricultural economic optimization models. A bibliography of 40 published models that have utilized PMP is in an appendix at the end of the book.

5.2.1 Behavioral Calibration Theory

The process of calibrating models to observed outcomes is an integral part of constructing physical and engineering models but is rarely formally analyzed for optimization models in agricultural economics. In this section we show that observed behavioral reactions yield a basis for calibrating models in a formal manner that is consistent with microeconomic theory. Analogously to econometrics, the calibration approach draws a distinction between the two modeling phases of calibration (estimation) and policy simulation (prediction).

On a regional level, the information on the product output levels and farm land allocations is usually more accurate than the estimates of marginal crop production costs. This is particularly true when micro data on land class variability, technology, and risk feature in the farmers' decisions, but are absent in the aggregate cost data available to the model builder. Accordingly, the PMP approach uses the observed acreage allocations and outputs to infer marginal cost conditions for each regional crop allocation observed. This inference is based on parameters that are known to be accurately observed and the usual maximizing and concavity assumptions on the profit function.

The Nonlinear Calibration Proposition: If the model does not calibrate to observed production activities with the set of linear constraints that can be empirically justified, a necessary condition for profit maximization at the observed values is that the objective function is nonlinear in at least some of the activities.

Many regional models have some nonlinear terms in the objective function reflecting endogenous price formation or risk specifications. Although it is well known that the addition of nonlinear terms improves the diversity of the optimal solution, there is usually an insufficient number of independent nonlinear terms to accurately calibrate the model.

The Calibration Dimension Proposition: The ability to calibrate the model with complete accuracy depends on the number of nonlinear terms that can be independently calibrated.

The ability to adjust some nonlinear parameters in the objective function, typically the risk aversion coefficient, can improve model calibration. However, if there are insufficient independent nonlinear terms the model cannot be made to calibrate precisely. In technical terms, the number of instruments the modeler has available to calibrate the model may not span the set of activities that need to be calibrated.

Proving the Propositions[1]

Proposition 1: Given an agent who maximizes a multi-output profit subject to linear constraints on some inputs or outputs. If the number of nonzero nondegenerate production activity levels observed (k) exceeds the number of binding constraints (m), then a necessary and sufficient condition for profit maximization at the observed levels is that the profit function be nonlinear (in output) in some of the (k) production activities.

Proof: Define the profit function in general as a function of input allocation x, $f(x)$.

$$\begin{aligned} & \max f(\bar{x}) \\ \text{subject to} \quad & \bar{A}\bar{x} \leq b \\ & \bar{x} = n \times 1,\ \bar{A} = m \times n,\ m < n \end{aligned} \tag{5.1}$$

At the observed optimal solution (nondegenerate in Primal and Dual specifications) there are k non-zero values of $\bar{x}$. Drop the zero values of $\bar{x}$ and define the $m \times m$ basic partition of A as the $(m \times m)$ optimal solution basis matrix B and the remaining partition of A as N $(m \times k - m)$. Partitioning the $k \times 1$ vector x into the $m \times 1$ vector x_B and $(k - m) \times 1$ vector x_N, Equation 5.1 is rewritten as:

$$\max f(x) \qquad \text{subject to } [B \vdots N] \begin{bmatrix} x_B \\ x_N \end{bmatrix} = b \tag{5.2}$$

or

$$\max f(x_B, x_N) \qquad \text{subject to } Bx_B + Nx_N = b \tag{5.3}$$

Given the constraint set in Equation 5.3, x_B can be written:

$$x_B = B^{-1}b - B^{-1}Nx_N \tag{5.4}$$

Since 5.4 are binding constraints, substituting Equation 5.4 into the objective function, Equation 5.3, gives:

$$\max f(B^{-1}b - B^{-1}Nx_N, x_N) \tag{5.5}$$

Taking the gradient of Equation 5.5 with respect to x_N yields the reduced gradient (r_{x_N}):

$$r_{x_N} = \nabla f_{x_N} - \nabla f_{x_B} B^{-1} N \tag{5.6}$$

[1]From Howitt (1995b)

A zero-reduced gradient is a necessary condition for optimality (Leunberger, 1984). Without loss of generality, we define the basic part of the objective function as linear with coefficients c_B which yields the optimality condition:

$$r_{x_N} = \nabla f_{x_N} - c'_B B^{-1} N = 0 \tag{5.7}$$

The objective function associated with the independent (x_N) variables has either zero coefficients, linear coefficients, or a nonlinear specification. If $f(x_N)$ had zero coefficients, x_N would have to be zero at the optimum given the positive opportunity cost of resources. If $f(x_N)$ was linear, say c_N then Equation 5.7 would be the reduced cost of the activity. A zero reduced cost of a nonbasic activity implies degeneracy when coupled with a zero activity level x_N. Since $x_N > 0$ at the optimum, $f(x_N)$ cannot be linear and hence must be nonlinear for Equation 5.7 to hold.

Proposition 2: A necessary condition for the exact calibration of a $k \times 1$ vector x is that the objective function associated with the $(k-m)\times 1$ vector of independent variables x_N contain at least $(k-m)$ linearly independent instruments that change the first derivatives of $f(x_N)$.

Proof: By Proposition 1, $f(x_N)$ is nonlinear in x_N. Each element of the gradient $\nabla f(x_N)$ has a component that is a function of x_N, and probably also a constant term. The optimality conditions in Equation 5.7 are modified by subtracting the constant components in the gradient (k) from both sides to give:

$$\nabla \bar{f}_{x_N} = c^* \tag{5.8}$$

where:

$$\nabla \bar{f}_{x_N} = \nabla f_{x_N} - \bar{k}' \text{ and } c^* = c'_B B^{-1} N - \bar{k}'$$

The $1 \times (k-m)$ vector $\nabla \bar{f}_{x_N}$ can be written as the product of x_N and a $(k-m) \times (k-m)$ matrix F, where the i^{th} column of F has elements

$$\frac{\partial f(x_N)}{\partial x_i} \frac{1}{x_i}$$

as in Equation 5.4. Using this decomposition:

$$\nabla \bar{f}_{x_N} = x'_N F \tag{5.9}$$

The necessary reduced gradient condition (Equation 5.8) can now be rewritten as:

$$x'_N F = c^* \tag{5.10}$$

Calibration of an optimization model requires that the observed solution vector $\tilde{x}$ results from the optimal solution of the calibrated model. From Equation 5.4, the independent values $\tilde{x}_N$ determine the dependent values $\tilde{x}_B$. Since from Equation 5.8, c^* is a vector of fixed parameters, the necessary condition (Equation 5.10) can only hold at $\tilde{x}_i$ if the values of F^{-1} can be calibrated to map c^* into $\tilde{x}_N$. Thus the matrix of calibrating gradients F^{-1} must span $\tilde{x}$ such that:

$$\tilde{x}'_N = c^* F^{-1} \tag{5.11}$$

It follows that the rank of F must be $(k - m)$ and there have to be $(k - m)$ linearly-independent instruments which change the values of F to exactly calibrate $\tilde{x}$.

Example: Let x_n be a 2×1 vector, $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, and

$$f(x_N) = \alpha' x_N - x'_N Q x_N \tag{5.12}$$

where

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$$

and Q is symmetric. Writing 5.7 as:

$$[\alpha_1 - 2x_1 q_{11} - 2x_2 q_{12},\ \alpha_2 - 2x_2 q_{22} - 2x_1 q_{21}] - c'_B B^{-1} N = 0 \tag{5.13}$$

defining the $1 \times (k - m)$ row vector c^* as in Equation 5.8 results in:

$$[2x_1 q_{11} + 2x_2 q_{12},\ 2x_1 q_{21} + 2x_2 q_{22}] = c^* \tag{5.14}$$

By definition, the left-hand side of Equation 5.14 can be written as the product of x'_N and a matrix F where:

$$F = \begin{bmatrix} 2q_{11} & 2q_{21} \\ 2q_{12} & 2q_{22} \end{bmatrix} \tag{5.15}$$

Therefore the optimality condition that the reduced gradient equals zero requires that $x_N F = c^*$. If particular values of x_N, say $\tilde{x}_N$, are required by changing the coefficients of F, then $\tilde{x}_N = c^* F^{-1}$.

Note from Equation 5.8 that $-c^*$ is the difference between the constant linear term in the objective function $\bar{k}$ and the opportunity cost of the resources. Thus $-c^*$ is equal to the vector of PMP Dual values λ_2. Solving for the parameters of F, given c^* and $\tilde{x}_N$, is computationally identical to solving for the vector of δ_i parameters which requires the necessary condition that F is linearly independent and of rank $(k - m)$.

Corollary: The number of calibration terms in the objective function must be equal to or greater than the number of independent variables to be calibrated.

Decoupling Proof[2]

Constraint Decoupling:
Constraint decoupling is achieved by the perturbation of the calibration constraints. Where the binding and slack resource constraints under values $\tilde{x}$ are separated into groups I and II, it is shown to preserve the Primal and Dual values.

Problem P1:

$$\begin{aligned} \max \quad & f(x) && (5.16)\\ \text{subject to} \quad & Ax = b && \text{(I)}\\ & \hat{A}x < \hat{b} && \text{(II)}\\ & Ix = \tilde{x} && \text{(III)} \end{aligned}$$

$$x = k \times 1,\ A = m \times k,\ \hat{A} = (l-m) \times k$$
$$\tilde{x} = k \times 1, k > m,\ \ b = m \times 1,\ \hat{b} = (l-m) \times 1$$

$\tilde{x}$ is a $k \times 1$ vector of activities that are observed to be nonzero in the base-year data; $k > m$ implies that there are more nonzero activities to calibrate than the number of binding resource constraints (I).

We assume that $f(x)$ is monotonically increasing in x with first and second derivatives at all points and that Problem P1 is not Primal or Dual degenerate.

Proposition 3: There exists a $k \times 1$ vector of perturbations $\epsilon(\epsilon > 0)$ of the values $\tilde{x}$ such that:

- Constraint set I in Equation 5.16 is decoupled from constraint set III, in the sense that the Dual values associated with constraint set I do not depend on constraint set III;
- The number of binding constraints in constraint set III is reduced so that the problem is no longer degenerate; and
- Binding constraint set I remains unchanged.

Proof: Define the perturbed problem with the calibration constraints defined as upper bounds without loss of generality.

[2] Also from Howitt (1995b)

Problem P2:

$$\begin{aligned} \max \quad & f(x) \\ \text{subject to} \quad & Ax = b \quad \text{(I)} \\ & \hat{A}x < \hat{b} \quad \text{(II)} \\ & Ix \leq \tilde{x} + \epsilon \quad \text{(III)} \end{aligned} \tag{5.17}$$

Any row of the nonbinding resource constraints, $\hat{A}x < \hat{b}$ in Problem P1 can be written as:

$$\sum_{j=1}^{k} |\hat{a}_{ij} x_j| < \hat{b}_i \qquad i = 1, \ldots, (1-m) \tag{5.18}$$

Select the constraint $i = 1, \ldots, (1-m)$ such that:

$$b_i - \sum_{j=1}^{k} \hat{a}_{ij} \tilde{x}_j$$

is minimized. If $\epsilon_j > 0,\ j = 1, \ldots, k$ are selected such that

$$\sum_{j=1}^{k} |\hat{a}_{ij} \epsilon_j| < \left[b_i - \sum_{j=1}^{k} \hat{a}_{ij} \tilde{x}_j \right] \tag{5.19}$$

By rearranging Equation 5.19, an inequality holds for the constraint when $x = \tilde{x} + \epsilon$, but x cannot exceed $\tilde{x} + \epsilon$ from constraint set III; therefore, those constraints in $Ax < b$ that are inactive under the values $\tilde{x}$ will remain inactive after the perturbation to $\tilde{x} + \epsilon$. The invariance of the binding resource constraints for (I) under the perturbation ϵ can be shown using the reduced gradient approach (Leunberger, 1984). Using Equation 5.19 we can write Problem P2 using only constraint sets I and III as:

$$\begin{aligned} \max \quad & f(x) \\ \text{subject to} \quad & Ax = b \\ & Ix \leq \tilde{x} + \epsilon \end{aligned} \tag{5.20}$$

where $A(m \times k)$ and $I = k \times k$. Invoking the nondegeneracy assumption for A and starting with the solution for Problem P1 (i.e., $\tilde{x}$), the constraints can be partitioned as:

$$\begin{bmatrix} B & N \\ I_1 & \\ & I_2 \end{bmatrix} \begin{bmatrix} x_B \\ x_N \end{bmatrix} \begin{matrix} = \\ \leq \\ \leq \end{matrix} \begin{bmatrix} b \\ \tilde{x}_B + \epsilon_B \\ \tilde{x}_N + \epsilon_N \end{bmatrix} \tag{5.21}$$

For brevity, the partition of A has been made so that the $(k-m)$ activities associated with N have the highest value of marginal products for the constraining resources. The reduced gradient for changes in $\tilde{x}_N$ is therefore:

$$r_{x_N} = \nabla f_{\tilde{x}_N} - \nabla f_{\tilde{x}_B} B^{-1} N \tag{5.22}$$

Since $f(\cdot)$ is monotonically increasing in x_N and x_B, the resource constraints will continue to be binding since the optimization criterion will maximize those activities with a nonnegative reduced gradient until the reduced gradient is zero or the upper-bound calibration constraint $\tilde{x}_N + \epsilon$ is encountered. Since $m < n$, the model overspecializes in the more profitable crops when subject only to constraint sets I and II. Under the specification in Problem P2, the most profitable activities will not have a zero-reduced gradient before being constrained by the calibration set II at values of $\tilde{x}_N + \epsilon$. Thus, the binding constraint set I remains binding under the ϵ perturbation.

The resource vector for the resource constrained crop activities (x_B) is now:

$$b - N(\tilde{x}_N + \epsilon) \tag{5.23}$$

and from Equation 5.21:

$$x_b = B^{-1}[b - N(\tilde{x}_N + \epsilon)]$$

Since B is of full rank m, exactly m values of x_B are determined by the binding resource constraints, which depend on the input requirements for the subset of calibrated crop acre values $\tilde{x}_N + \epsilon$.

The slackness in the m calibration constraints associated with the m resource constrained output levels x_B, follows from the monotonicity of the production function in the rational stage of production. Since the production function is monotonic, the input requirement functions are also monotonic, and expansion of the output level of the subset of crop acreage to $\tilde{x}_N + \epsilon$ will have a nonpositive effect on the resource vector remaining for the vector of crop acreages constrained by the right-hand side, x_B. That is:

$$b - N(\tilde{x}_N + \epsilon_N) \leq b - N\tilde{x}_N \text{ for } \epsilon_N > 0 \tag{5.24}$$

But since the input requirement functions for the x_B subset are also monotonic, Equations 5.24 and 5.21 imply that:

$$x_B \leq \tilde{x}_B \text{ or } x_B < \tilde{x}_B + \epsilon_B \text{ for } \epsilon_B > 0 \tag{5.25}$$

From Equation 5.25, it follows that the m perturbed upper-bound calibration constraints associated with x_B will be slack at the optimum solution.

Given Equations 5.19 and 5.25, the constraints at the optimal solution to the perturbed Problem P2 are:

$$\begin{bmatrix} B & N \\ \hat{A}_1 & \hat{A}_2 \\ I_1 & \\ & I_2 \end{bmatrix} \begin{bmatrix} x_B \\ \tilde{x}_N + \epsilon_N \end{bmatrix} \begin{matrix} = \\ < \\ < \\ = \end{matrix} \begin{bmatrix} b \\ \hat{b} \\ \tilde{x}_B + \epsilon_B \\ \tilde{x}_N + \epsilon_N \end{bmatrix} \tag{5.26}$$

Thus, there are k binding constraints, $b(m \times 1)$ and $x_B + \epsilon_N[(k-m) \times 1]$. The Dual constraints to this solution are:

$$\begin{bmatrix} B' & 0 \\ N' & I_2 \end{bmatrix} \begin{bmatrix} \lambda_1^* \\ \lambda_2^* \end{bmatrix} = \begin{bmatrix} \nabla_{x_B} f(x^*) \\ \nabla_{x_N} f(x^*) \end{bmatrix} \tag{5.27}$$

using the partitioned inverse,

$$\begin{bmatrix} \lambda_1^* \\ \lambda_2^* \end{bmatrix} = \begin{bmatrix} P & 0 \\ Q & I \end{bmatrix} \begin{bmatrix} \nabla_{x_B} f(x^*) \\ \nabla_{x_N} f(x^*) \end{bmatrix} \tag{5.28}$$

where $P = B'^{-1}$ and $Q = -N'B'^{-1}$.

Thus, the ϵ perturbation on the upper-bound constraint set II decouples the Dual values of constraint set I from constraint set II. This ensures that k constraints are binding and the partitioning of A into B and N is the unique outcome of the optimal solution to problem P2 in the first stage of PMP.

End of Proofs

5.2.2 A Cost-Based Approach to PMP Calibration.

This section demonstrates the PMP calibration process using nonlinear costs and constant yields to calibrate the model. The derivation is shown in its simplest form. Once you have the concept, the more complex development of changing yields will be clearer.

A key concept in PMP calibration based on work by Fiacco & McCormack , is that every linear constraint in an optimization problem can also be modeled by a nonlinear cost function with appropriately chosen coefficients.

Single Crop Cost-Based PMP Calibration

A single linear crop production activity is measured by the acres x allocated to the crop. The yield is assumed constant. The data available to the modeler is:

Marginal revenue/acre is assumed constant at	\$500/acre
Average Cost is	\$300/acre
Observed acres allocated to the crop	50 acres

In the first step, a measure of the value of the residual cost needed to calibrate the crop acreage to 50 (by setting marginal revenues equal to marginal cost at that acreage) is obtained from a constrained linear program. See Figure 5.1.

$$\begin{aligned} & \max \Pi(x) = 500x - 300x \\ \text{subject to} \quad & x \leq 100 \end{aligned} \tag{5.29}$$

The calibration proceeds in five steps:

Step I From the nonlinear calibration proposition we know that either (or both) the cost or production function is nonlinear if we need calibration constraints. In this case we define the total cost function to be quadratic in acres (x). There are very many possible nonlinear forms, but this is the simplest general form.

$$TC = \alpha x + \frac{1}{2}\gamma x^2$$

Step II Under unconstrained optimization crop acreage expands until the marginal cost equals marginal revenue. Therefore $MC = MR$ at $x = 50$.

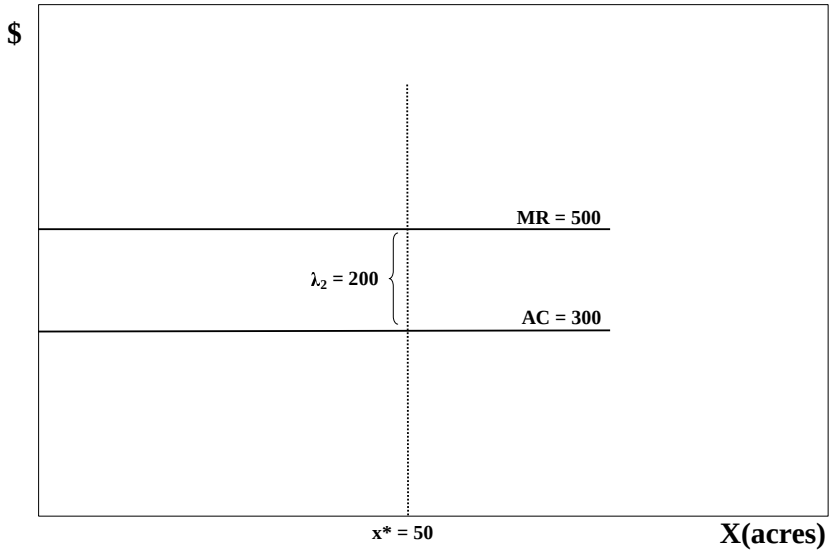


Figure 5.1: LP Constrained by calibration constraints

Step III It follows that the value λ_2 in the linear model is the difference at the constrained calibration value and is equal to $MR - AC$. But (from Step II) we know that $MR = MC$, and therefore $\lambda_2 = MC - AC$ (since $MR = MC$ at $x = 50$). Given the hypothesized total cost function TC:

$$MC = \alpha + \gamma x \qquad AC = \alpha + \frac{1}{2}\gamma x$$

$$MC - AC = \alpha + \gamma x - (\alpha + \frac{1}{2}\gamma x)$$

$$\therefore \lambda_2 = MC - AC = \frac{1}{2}\gamma x$$

and the cost slope coefficient is calculated as:

$$\gamma = 2\lambda / x^* = (2 \times 200)/50 = 8$$

Step IV We can now calculate the value of the cost function intercept α using the AC information in the basic data set:

$$300 = \alpha + (\frac{1}{2} \times 8 \times 50) \Rightarrow \alpha = 300 - 200 = 100$$

Step V Using the values for α and γ, the unconstrained quadratic cost problem is:

$$\max \Pi = 500x - \alpha x - \frac{1}{2}\gamma x^2 = 500x - 100x - \frac{1}{2}8x^2$$

$$\frac{\partial \Pi}{\partial x} = 500 - 100 - 8x$$

Setting $\partial\Pi/\partial x = 0$ (which implies MR = MC) results in:

$$8x = 400 \Rightarrow x^* = 50$$

Notes
1. The unconstrained model calibrates exactly in x and also in Π.
2. $MC = MR$ at $x = 50$.
3. $AC = 300$ at $x = 50$.
4. The cost function has been "tilted".
5. Two types of information are used: x^* and AC.
6. The observed x^* quantities need to be mapped into value space (λ_2) by the calibration constrained LP before it can be used.
7. The model now reflects the preferences of the decision maker.
8. The model is unconstrained by calibration constraints for policy analysis.

5.2.3 An Analytic Derivation of Calibration Functions

This section will show that the PMP non-linear calibration approach can be applied to any non-degenerate linear problem. The derivation of the general result proceeds in three steps. The first step shows that the Dual value on the calibration constraint for the calibrated activity set x_k is equal to the reduced cost of the activity x_i in the un-calibrated base problem. The second step shows that if the correct non-linear penalty function is added to the objective function, the resulting nonlinear problem satisfies the necessary conditions for optimality at the required value of x_i. Finally, it is shown that the correct penalty function has a gradient at the required value of x_i equal to the negative of the calibration Dual. The general linear programming optimization problem can be compactly written as:

$$\begin{aligned} &\max_x \; c'x \\ &\text{subject to } Ax \leq b, \quad x \geq 0 \end{aligned} \tag{5.30}$$

Where c is an $p \times 1$ vector of net returns per unit activity; the matrix A and right hand side b are, respectively, the usual technical constraints and right hand sides. The dimension of A is $m \times p$, $(m < p)$. The basis dimension of A is m. If the number of observed activities is $n(n \leq p)$, where $n = k + m$, then — in addition to the x_m basis activities — there are an additional k activities (x_k) that are observed and need to be calibrated into the optimal model solution. For simplicity, define the LP problem as only subject to one set of upper bound calibration constraints:

$$\begin{gathered} \max_x \; c'x \\ \text{subject to } Ax \leq b,\; Ix \leq \tilde{x} + \epsilon \text{ and } x \geq 0 \end{gathered} \tag{5.31}$$

where ϵ is added to the calibration constraints to prevent degeneracy.[3]

The optimal basic solution to this problem will have a mix of n binding resource and calibration constraints. The A matrix can be partitioned into an $m \times m$ basis matrix B that corresponds to the m-least-profitable "marginal" activities (x_m), and an associated $m \times k$ matrix N for the k calibrated activities x_k. Dropping out the $p - n$ zero activities, the optimal basic solution to equation 5.31 can be written as:

$$\begin{gathered} \max_x \; c'_m x_m + c'_k x_k \\ \text{subject to } \hat{A}x = \hat{b} \text{ — which is partitioned as:} \\ \begin{bmatrix} B & N \\ 0 & I \end{bmatrix} \begin{bmatrix} x_m \\ x_k \end{bmatrix} = \begin{bmatrix} b \\ \tilde{x}_k + \epsilon \end{bmatrix} \end{gathered} \tag{5.32}$$

The optimal Dual constraints for this problem are $\hat{A}'\lambda = c$ which are partitioned as:

$$\begin{bmatrix} B' & 0 \\ N' & I \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} c_m \\ c_k \end{bmatrix} \tag{5.33}$$

Equation 5.33 can be solved for the values of λ_1 and λ_2 by inverting the partitioned constraint matrix, using the partitioned form of the inverse, to yield:

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} P & 0 \\ Q & I \end{bmatrix} \begin{bmatrix} c_m \\ c_k \end{bmatrix} \tag{5.34}$$

$$\text{Where } P = B'^{-1} \text{ and } Q = -N'B'^{-1}$$

[3] See the proof of proposition 3 on pages 71-73.

From equation 5.34 we see that the k x 1 vector of Dual values for the binding calibration constraints (λ_2) has the value:

$$\lambda_2 = c_k - N' B'^{-1} c_m \tag{5.35}$$

The right hand side of equation Equation 5.35 is the difference between the gross margin of the calibrating activity c_k and the equivalent gross margin that can be obtained from the less profitable marginal cropping activities c_m. In other words λ_2 is the marginal opportunity cost of restricting the calibrated activities x_k by the amount needed to bring the marginal x_m activities into the expanded basis. This cost of restricting the more profitable activities x_k in the basis is similar to the familiar reduced cost term.

Note first that when land is the *numeraire*,[4] the corresponding coefficients in the N and B matrices are one.

Second, the sign on λ_2 is positive for GAMS printouts as a marginal increase in the right hand side upper bound on the more profitable activities will increase the objective function value.

The Dual values associated with the set of binding calibration constraints (λ_2) are independent of the resource and technology constraint Dual values (λ_1), since the constraint decoupling proposition (page 71-73) shows that the values for λ_1 are not changed by the addition of the calibration constraints.

If an increasing nonlinear cost function is added to the objective function for the x_k activities that need to be calibrated, the marginal and average costs of producing x_k will differ. The net return to land from x_k now decreases as the acreage is increased. The net returns to land from x_k reach an internal equilibrium solution at the point where they are equal to the opportunity cost of land set by the marginal crops x_m. This condition is the microeconomic "equimarginal" principle of optimal input allocation across products.

If the calibration constraints are removed and a nonlinear cost for x_k is added, equation 5.31 becomes:

$$\begin{aligned} &\max_x \; c'_m x_m + c'_k x_k - f(x_k) \\ &\text{subject to } [B\ N] \begin{bmatrix} x_m \\ x_k \end{bmatrix} = b \end{aligned} \tag{5.36}$$

The reduced gradient (the nonlinear equivalent of the reduced cost) for activities x_k is derived by rewriting the set of binding constraints in Equa-

[4] Leontieff production needs a common unit of measurement.In agricultural models land is usually chosen

tion 5.36 so that x_m is a function of x_k:

$$x_m = B^{-1}b - B^{-1}Nx_k \tag{5.37}$$

Substituting the expression for x_m back into the objective function in equation 5.36 defines the problem as an unconstrained function of x_k.

$$Max\ c'_m(B^{-1}b - B^{-1}Nx_k) + c'_k x_k - f(x_k) \tag{5.38}$$

The unconstrained gradient of this nonlinear problem in x_k is defined as the reduced gradient. Taking the derivative of equation 5.38 and transposing yields:

$$c_k - N'B'^{-1}c_m - \nabla f'(x_k) \tag{5.39}$$

Where $\nabla f'(x_k)$ is the gradient of $f(x_k)$ and is a row vector by definition.

Leunberger (1984) shows that a zero valued reduced gradient is a necessary condition for the optimum of a nonlinear problem. The calibrated equation 5.36 will optimize with a zero reduced gradient at the values $\tilde{x}_k$ when $c_k - N'B'^{-1}c_m = \nabla f'(x_k)$ or substituting into equation 5.35, when $\nabla f'(x_k) = \lambda_2$.

The PMP Proposition: If the parameters of $f(x)$ are calibrated such that at the value $\tilde{x}_k$, $\nabla f'(\tilde{x}_k) = \lambda_2$, then the model will be optimal exactly at the calibrating acres.

To reiterate, equation 5.35 shows that λ_2 is equal to the first two terms of equation 5.39. It follows that the reduced gradient of the resulting nonlinear problem will equal zero at $\tilde{x}_k$. As this is a necessary condition for the optimum, the problem in equation 5.36 will calibrate at the values $\tilde{x}_k$ without calibration constraints.

Equation 5.38 shows that optimal solution to the calibrated problem responds to changes in the linear gross margin (c), the right hand side values (b)," or the constraint coefficients (B or N).

The economic interpretation of the calibration cost function $f(x_k)$ is as follows. From equation 5.35, we see that λ_2 is equal to the difference between the gross margins per unit land for x_k and x_m. The gross margins are calculated from the observed average variable costs, that is $c_i = (p_i y_i - AC_i)$. It follows that:

$$\lambda_2 = (p_k y_k - AC_k) - (p_m y_m - AC_m) \tag{5.40}$$

but since the gross margin for the marginal crop x_m is equal to the opportunity cost of land, and since the land coefficients in N and B equal

one, using equation 5.34 and 5.35 we can rewrite 5.40 as:

$$\lambda_2 = (p_k y_k - AC_k) - \lambda_1 \text{ or } \lambda_2 + \lambda_1 = (p_k y_k - AC_k) \tag{5.41}$$

but at the optimal allocation of land, all crops must have a marginal net return equal to the opportunity cost of land (by the equimarginal principle), therefore at the optimal solution to the nonlinear problem defined by equation 5.36:

$$\lambda_1 = p_k y_k - MC_k = p_k y_k - AC_k - (MC_k - AC_k) \tag{5.42}$$

substituting 5.41 into 5.42 yields:

$$\lambda_2 = MC_k - AC_k \tag{5.43}$$

To summarize, this section has shown that linear and nonlinear optimization problems can be exactly calibrated by the addition of a specific number of nonlinear terms. We have used a general nonlinear specification to show that since the calibrated equation 5.31 yields the necessary conditions 5.35. If the nonlinear equation 5.36 has a nonlinear cost function f(x_k) that satisfies equations 5.35, 5.39, and 5.43, the resulting nonlinear problem will calibrate exactly in the Primal and Dual values of the original problem, but without any inequality calibration constraints.

In the next section we show how the calibration procedure can be simply implemented using a quadratic cost function in a two-stage process that is initiated with a calibrated linear program.

5.3 An Empirical Calibration Method

The previous section showed that if the correct nonlinear parameters are calculated for the $(k-m)$ unconstrained (independent) activities, the model will exactly calibrate to the base year values x without additional constraints. The problem addressed in this section is to show how the calibrating parameters can be simply and automatically calculated using the minimal data set for a base year LP.

Given that nonlinear terms in the supply side of the profit function are needed to calibrate a production model, the task is to define the simplest specification that is consistent with the technological basis of agriculture, microeconomic theory and the data base available to the modeler.

A highly probable source of nonlinearity in the profit function is due to heterogeneous land quality. This will cause the marginal cost per unit of

output to increase as the proportion of a crop in a specific area is increased. This phenomenon, first formalized by Ricardo, is widely noted by farmers, agronomists, and soil scientists, but often omitted from quantitative production models (Peach, 1993).

Defining yields per acre as constant and marginal cost as increasing in land allocation is a considerable simplification of the complete production process. Given the applied goal of this "positive" modeling method, the calibration criteria used is not whether the simple production specification is true, but rather, does it capture the essential behavioral response of farmers, and can it be made to work with the restricted data bases and model structures available.

Since land is assumed to be the only input activity and the Leontief numeraire, the coefficient is equal to one.

In many LP problems such as the Yolo example, the objective function coefficients c_i represent the Gross margin per acre faced by the farmer. The c_i coefficient is composed of the product of the average yield times the price per unit output, with the average variable costs deducted. Since we are specifying the yield as constant but the marginal cost as increasing with acreage in this model, we have to separate the cost and revenue components of the gross margin.

The calibrated optimization problem equivalent to equation 5.36 and with land as the restricting factor becomes:

$$\begin{array}{c} Max \ \sum_i p_i y_i x_i - (\alpha_i + 0.5\gamma_i x_i)x_i \\ \text{subject to } Ax \leq b, \quad x \geq 0 \end{array} \tag{5.44}$$

where $a_{i1} = 1$ and in this case A is a $1 \times n$ vector, x_i is the acreage of land allocated to crop i, y_i is the yield per acre, α_i and γ_i are (respectively) the intercept and slope of the total land cost function. In addition, ω_j is the average cost per acre of land.

The PMP calibration approach uses three Stages. In the first Stage a constrained LP model is used to generate the Dual values for both the resource and calibration constraints, λ_1 and λ_2 respectively. In the second Stage, the calibrating constraint Dual values (λ_2) are used along with the data-based average cost function to derive the unique calibrating cost function parameters (α_i and γ_i). In the third Stage, the cost parameters are used with base-year data to specify the PMP model in equation 5.44. The resulting model calibrates exactly to the base-year solution and original constraint structure.

The procedure is illustrated using a very simple problem which has a single land constraint (5 acres) and two crops (wheat and oats). The following parameters are used:

	Wheat (w)	Oats (o)
Crop prices	P_w = \$2.98/bu.	P_o = \$2.20/bu.
Variable cost/acre	ω_w = \$129.62	ω_o = \$109.98
Average yield/acre	$\bar{y}_w$ = 69 bu.	$\bar{y}_o$ = 65.9 bu.
Gross Margin/acre	c_w = 76.0	c_o = 35.0

The observed acreage allocation in the base year is 3 acres of wheat and 2 acres of oats.

Figure 5.2 shows the initial problem in a diagrammatic form for two activities, with one resource constraint and two upper bound calibration constraints. Note that at the optimum, the calibration constraint will be binding for wheat, the activity with the higher average gross margin, while the resource constraint will restrict the acreage of oats.

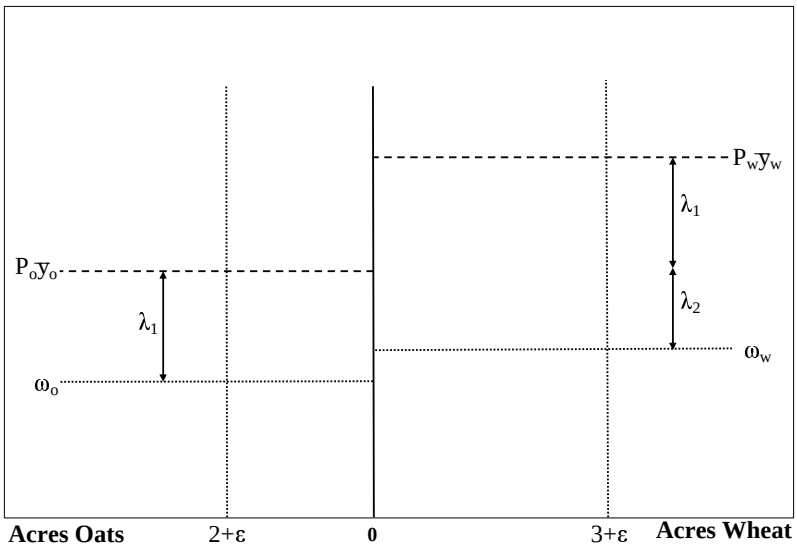


Figure 5.2: Two LP Activities

The problem in Figure 5.2 is:

$$\max(2.98 \times 69 - 130)x_w + (2.20 \times 65.9 - 110)x_o \qquad (5.45)$$

subject to Constraints 1 – 3:

$$\begin{array}{lcl} 1) & x_w + x_o & \leq 5 \\ 2) & x_w & \leq 3.01 \\ 3) & x_o & \leq 2.01 \end{array}$$

Note the addition of the ϵ perturbation term (≈ 0.01) on the right hand side of the calibration constraints. The average gross margin from wheat is \$76/acre and oats \$35/acre. The optimal solution to the Stage 1 problem is when the wheat calibration constraint is binding at a value of 3.01 and Constraint 1 is binding when the oat acreage equals 1.99. The oat calibration constraint is slack.

Two equations are solved for the two unknown yield parameters (α and γ). Using the quadratic total land cost function specified in 5.44 and the first order conditions in equation 5.39, the first equation for the marginal cost coefficient sets λ_2 equal to the difference between marginal and average cost based on equation 5.43 and derives the calibrated value for γ.

$$\begin{array}{c} f'(\tilde{x}_k) - \bar{f}(\tilde{x}_k) = \lambda_{2k} \\ \alpha + \gamma_k \tilde{x}_k - \alpha - 0.5\gamma_k \tilde{x}_k = \lambda_{2k} \\ \gamma_k = \frac{2\lambda_{2k}}{\tilde{x}_k} \end{array} \tag{5.46}$$

Equation 5.46 uses the value of the Dual on the LP calibration constraint (λ_2) which is shown in figure 5.1 to be the difference between the average cost (AC) of the crop and the marginal cost (MC).

The second equation is the average cost for crop i_i:

$$\begin{array}{c} \omega_i = \alpha_i + 0.5\gamma_i x_i \\ \therefore \alpha_i = \omega_i - 0.5\gamma_i x_i \\ \therefore \alpha_i = \omega_i - \lambda_2 \end{array} \tag{5.47}$$

The derivation of the two types of Dual value λ_1 and λ_2, can be shown for the general case (Howitt, 1995b). The A matrix in 5.31 is partitioned by the optimal solution into an $m \times m$ matrix B associated with the marginal variables x_m, an $m \times 1$ subset of x with inactive calibration constraints. These activities set the opportunity costs for the $m \times 1$ set of binding resource constraints. The second partition of A is into an $m \times k$ matrix N associated with a $k \times 1$ partition of x, x_N of non-zero activities constrained by the calibration constraints. The equation for λ_1 is the usual LP form of:

$$\lambda_1 = c'_m B^{-1} \tag{5.48}$$

The elements of vector x_m are the acreage produced in the crop group limited by the general constraints, and λ_1 are the Dual values associated with the set of $m \times 1$ binding general constraints. Equation 5.48 states that the value of marginal product of the constraining resources is a function of the revenues from the constrained crops. Remembering the reduced gradient specification in equation 5.39, the "independent" crops (x_k) do not influence the Dual value of the resource by the decoupling proposition. This is consistent with the principle of opportunity cost in which the marginal net return from a unit increase in the constrained resource determines its opportunity cost. Since generally the more profitable crops x_k are constrained by the calibration constraints, the less profitable crop group x_m are those that could use the increased resources and hence set the opportunity cost.

Equation 5.49 determines the Dual values on the upper bound calibration constraints on the crops.

$$\begin{gathered} \lambda_2 = -N'B'^{-1}c_m + Ic_k \\ \text{and substituting 5.48 gives:} \\ \lambda_2 = Ic_k - N'\lambda_1 \end{gathered} \tag{5.49}$$

The Dual values for the binding calibration constraints are equal to the difference between the marginal revenues for the calibrated crops (x_k) and the marginal opportunity cost of resources used in production of the constrained marginal crops (x_m). Since the Stage I problem in Figure 5.2 has a linear objective function, the first term in 5.49 is the crop average value product of land in activities x_k. The second term in 5.49 is the marginal opportunity cost of land from equation 5.48. In this PMP specification, the difference between the average and marginal cost of land is attributed to changing land quality. Thus the PMP Dual value (λ_2) is a hedonic measure of the difference between the average and marginal cost of land, for the calibrated crops. Analogously to revealed preference, PMP can be thought of as revealed efficiency based on observed land allocations.

Equation 5.49 substantiates the Dual values shown in Figure 5.2, where the Duals for the calibration constraint set (λ_2) in the Stage I problem are equal to the divergence between the LP average cost per acre and the marginal opportunity cost per acre.

The Dual value on land (λ_1) is \$35 and on the calibration constraint (λ_2) is [41 0]. Using equation 5.45, the λ_2 value for wheat, and the base-year data, the cost function slope for wheat is calculated as:

$$\gamma_w = \frac{2 \times 41}{3} = 27.333 \tag{5.50}$$

γ_w is now substituted into equation 5.47 to calculate the cost intercept α_w.

$$\alpha_w = 129.62 - (0.5 \times 27.333 \times 3) = 88.62 \tag{5.51}$$

Using the cost function parameters, the Stage II Primal PMP problem becomes (see Figure 5.3):

$$\begin{aligned} &\max[(2.98 \times 69)x_w + (2.20 \times 65.9)x_o \\ &\quad - (88.62 + 0.5 \times 27.333x_w)x_w - 109.98x_o] \\ &\text{subject to } x_w + x_o \leq 5 \end{aligned} \tag{5.52}$$

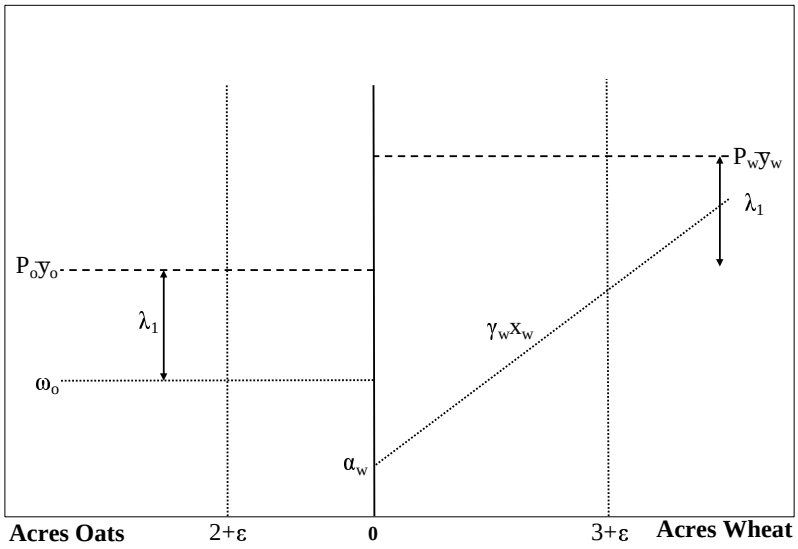


Figure 5.3: One PMP Activity and One LP Activity

A quick empirical check of the calibration to the base values is performed by calculating the VMP per acre of wheat at 300 acres. If it is close to the VMP (VAP) of oats and converging, the model will calibrate without the additional calibration constraints.

The marginal cost per acre of wheat is:

$$\begin{aligned} MC_{(w=3)} &= 88.62 + 27.333 \times 3 &&&&= 170.619 \\ VMP_{(w=3)} &= 2.98 \times 69 - 170.619 &&= 205.62 - 170.619 &&= 35.001 \\ VMP_o &= 2.20 \times 65.9 - 110 &&= 144.98 - 110 &&= 34.98 \end{aligned}$$

The VMP for wheat at 3 acres of $35.01 is marginally above the VMP for oats ($34.98). Thus, the unconstrained PMP model will calibrate within the rounding error of this example.

5.3.1 Calibration Using Supply Elasticity Estimates

Since the PMP procedure solves for the marginal cost function it also solves for the range of supply elasticities based on the marginal costs. However, as can be seen from equation 5.46, the marginal cost parameter γ_k depends on the empirical parameters of λ_2 and $\tilde{x}$. The resulting supply elasticity is not bounded and thus can assume values for a short calibration period that may be inconsistent with estimates based on a larger representative sample of crop response. The elasticity of supply is the essential measure of how the calibrated PMP model responds to policy changes. Accordingly, and consistent with the philosophy of using the best information available, the modeler must check the equilibrium elasticities implied by the calibrated cost functions, and if reliable parameters are available, use the prior information on elasticities to calibrate the model.

The PMP marginal cost slope is calibrated against prior econometric estimates, but also reflects the conditions that are present in the base year model conditions. Modelers should be aware that using an elasticity based on prior econometric estimates to calculate the adjustment value does not ensure a positive net return. Net returns over variable costs should be checked after the adjustment factor is calculated.

The supply elasticity based on prior econometric estimates is defined as:

$$\eta_s = \frac{\partial q}{\partial p}\frac{P}{Q}$$

Using the assumption of a constant per acre yield and the usual marginal cost supply function specification, the elasticity can be rewritten in terms of crop land allocations as:

$$\eta_s = \frac{\partial x}{\partial mc}\frac{P}{x^*} \quad \text{or} \quad \frac{\eta_s x^*}{P} = \frac{\partial x}{\partial mc} \tag{5.53}$$

Since the nonlinear total cost is $\alpha x + 0.5\gamma x^2$ and the supply function (marginal cost function) is $\alpha + \gamma x$, the change in marginal cost with output (γ) is:

$$\frac{\partial mc}{\partial x} = \gamma \Rightarrow \frac{1}{\gamma} = \frac{\partial x}{\partial mc} = \frac{\eta_s x^*}{P} \tag{5.54}$$

$$\gamma = \frac{P}{\eta_s x^*}$$

The supply elasticity implied by the unrestricted PMP calibration 5.54 is calculated from:

$$\eta_s = \frac{P}{x^* \gamma} \tag{5.55}$$

In the Standard PMP model elasticities are available for all crops and can be used to calibrate the PMP functions as follows. First, the γ coefficients are calculated using equation 5.54 and prior estimates of supply elasticities. Since we have set the slope of the marginal cost function, we have to calculate the marginal cost function intercept parameter α using λ_2 and the average cost:

$$\alpha_k = AC_k + \lambda_{2k} - \gamma_k x_k^*$$

The resulting PMP model will calibrate in inputs and outputs, and will also have elasticities of supply that calibrate to the base data.

5.3.2 Calibrating Marginal Crops

A valid objection to the simple PMP specification in 5.52 is that we assume an increasing cost of production/acre for the more profitable unconstrained crops x_k, but the marginal crops x_m that are constrained by resources are assumed to have constant production costs per acre.

Calibrating the marginal crops (x_m) with increasing cost functions requires additional empirical information. The independent variables, as x_k are termed, use both the constrained resource opportunity cost (λ_1) and their own calibration Dual (λ_2) to solve for the yield function parameters implied by the observed crop allocations (see Figure 5.3). However, the marginal crops (x_m) have no binding calibration constraint, and thus cannot empirically differentiate marginal and average cost at the observed calibration acreage, using the minimal LP data set specified.

Clearly some additional data on the marginal cost function for this group of crops is needed. For cost function calibration, the best additional data comes from prior estimates of elasticities of supply. Since we are now changing the opportunity cost of the restricting resources by changing the costs of the marginal crops, we will have to adjust all the PMP λ_2 values. We use a prior elasticity of supply to calculate the adjustment factor.

Defining the adjustment factor $Adj = MC - AC$ at $\tilde{x}_m$. It is the additional cost that we need to add to the LP average cost to obtain a nonlinear cost function. The Adj value should be thought of as the PMP term for

the marginal activities, but since it increases the marginal opportunity of binding resources, it will also change all the other non-marginal PMP values.

$$Adj = \frac{1}{2}\gamma x^* = \frac{P}{2\eta_s} \tag{5.56}$$

Defining the slope from equation 5.54 and the prior elasticity, yields the second term in 5.56.

Now we redefine the PMP values for the non-marginal crops as:

$$\hat{\lambda}_{2i} = \lambda_{2i} + Adj$$

We can now calculate the PMP cost function values of α and γ using the adjusted values and the average costs from the data set.

Returning to the example in equation 5.52 and Figure 5.2, the Stage 1 calibrated problem is run exactly as before. One of the important pieces of information from the optimal solution of the Stage 1 problem is which activities are in the x_k and x_m groups. The modeler is unlikely to know this beforehand.

In the example, let us assume that the a priori information on the elasticity of supply for oats is that it is 2.25. Using equation 5.56 for the adjustment term (Adj_m), the adjustment term for Oats is:

$$Adj_o = \hat{\lambda}_{2o} = 2.20 \times 65.9/2 \times 2.25 = 28.996 \tag{5.57}$$

Note that this adjustment factor is per acre, so instead of the price per unit product used in the normal elasticity formula, we have to use total revenue ($price \times yield$) per acre, hence (2.20×65.9) which is the price of oats times the yield per acre for oats.

This Adj value now plays the role of λ_2 for the marginal crops (Oats in this example). The residual Dual value on land set by the oat cropping activity is reduced accordingly by 28.996 from 35 and becomes 6.004.

The nature of marginal crops would lead one to expect that they would have a highly elastic supply. This is a factor to be considered carefully, as from equation 5.56, it can be seen that there is no bound on the value of Adj. This means that if small elasticities are used, they can result in large Adj values, that in turn, can lead to negative shadow values and resulting calibration problems.

The PMP Dual on Wheat (λ_{2w}) must also be increased by this same amount to ensure the first order conditions hold. The new value for $\hat{\lambda}_{2w}$ is:

$$\hat{\lambda}_{2w} = 41.0 + 28.996 = 69.996 \tag{5.58}$$

The calculations for the cost coefficients in 5.50 and 5.51 are now applied to all activities, both marginal (x_m) and independent (x_k). Note that the adjusted $\hat{\lambda}_2$ values are used for the independent activities and the *Adj* value based on the prior data is used for the marginal crops.

The PMP problem given the information on marginal yields for the oat crop is re-defined using the new $\hat{\lambda}_2$ values for both Wheat and Oats in equations 5.50 and 5.51 and now becomes (also see Figure 5.4):

$$\begin{aligned} \max \quad & (2.98 \times 69) - (59.624 + (0.5 \times 46.664)x_w)x_w \\ & + (2.20 \times 65.9 - (80.984 + (0.5 \times 28.996)x_o))x_o \\ \text{subject to} \quad & x_w + x_o \leq 5 \end{aligned} \tag{5.59}$$

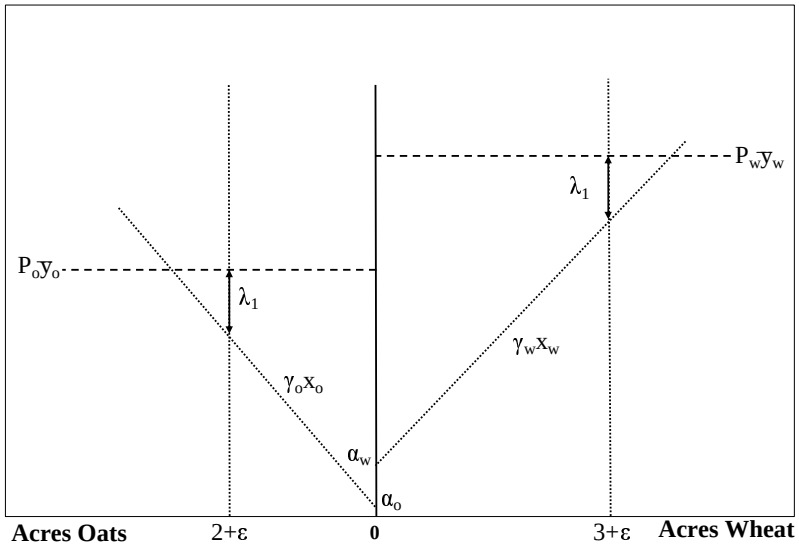


Figure 5.4: PMP Calibration on All Crops.

The calibration acreage can be checked by calculating the VMP for each crop at the calibration acreage of $\tilde{x}_w = 3$ and $\tilde{x}_o = 2$. This is:

$$\begin{aligned} VMP_w|_{\tilde{x}_{(w=3)}} &= 2.98 \times 69 - (59.624 + 46.664 \times 3) = 6.004 \\ VMP_o|_{\tilde{x}_{(o=2)}} &= 2.20 \times 65.9 - (80.984 + 28.996 \times 2) = 6.004 \end{aligned} \tag{5.60}$$

Since the VMP's are equal to each other and also equal to the new opportunity cost of land, the PMP model with the new cost functions will calibrate arbitrarily close to the base year acreage.

The resulting model will calibrate acreage allocation and input use, and the objective function value precisely. However, the Dual value on land will be lower reflecting the additional, and presumably more accurate, data on the marginal cost of the marginal crops obtained through the elasticities of supply.

5.4 Calibrating Production and Intermediate Activities

5.4.1 A Crop and Livestock Farm Model

The agricultural models that have been used as examples for calibration have been exclusively simple cropping models where the product is produced and sold directly. However, many agricultural enterprises or regional aggregations produce a mixture of crops for direct sale, and livestock that are also sold. In addition, crops such as pasture and fodder crops are grown to be fed directly to livestock. A farm manager has the option of selling, purchasing, or producing intermediate feed products. We define the activities of purchasing and selling products as intermediate activities, as they do not directly involve the productive resources of the farm. Farmers are small agents in the commodity markets from which they buy and sell, and thus are probably price takers facing fixed market prices. If follows that intermediate activities should be modeled as linear constant cost or price activities, unless there is good empirical evidence to the contrary. A second extension to the basic crop model is that the calibration unit will differ for different crop and livestock activities. In the basic crop model we used a unit of land as the calibration unit, however for a more general model the calibration unit varies with the activity. The model builder has to specify a meaningful unit for the activity such as an acre of land, a cow, a sheep , or for more complex livestock operations, an animal unit year index can simplify the specification. The general model specification starts with defining the activities into two subsets of production processes, and transfer processes. Revenues and costs are defined for both production processes and transfer processes, but for some production processes the revenues may be defined as zero since they are transferred the to appropriate transfer constraint for reallocation to other productive processes. A typical example is

the production of hay that uses farm inputs, but is not sold directly, but is fed to an animal production process. In this case, hay would be assigned its share of production costs, but its marginal revenue would be defined as zero. However a hay transfer constraint shows the required hay for livestock production, and enables the manager to avoid purchasing it by producing it on the farm. Thus the production of hay would have an opportunity cost marginal revenue equal to the price of hay. Transfer activities also have different revenues and costs. An activity that purchased hay would involve a cost, but not revenue, likewise the sale of hay will result in revenue, but no direct costs. The section 2.10 in chapter 2 shows how to define the linear linkage constraints that are used to transfer between activities. An example of a hay transfer constraint is used to show the linear specification. The constraints are defined for each transfer activity such as working capital, hay, grain, or labor when there is a mix of family and hired labor used on the farm. Only the productive activities are calibrated in the model. Thus the set of calibration activities is only defined over the production subset.

$$\max_x \; mr'x - ac'x \tag{5.61}$$

$$\text{subject to } Ax_p \leq b_p, \tag{5.62}$$

$$Ix_p \leq \tilde{x_p} \tag{5.63}$$

$$Tx = b_t \quad x_p \quad \& \quad x_t \geq 0 \tag{5.64}$$

Note that the resource and calibration constraints are only applied to the productive activities as before, but the matrix of transfer constraint coefficients links both the production activities (x_p) and the transfer activities (x_t). The Midwest Farm Gams template program illustrates how the production and transfer activities interact and are calibrated. The activities are growing corn,wheat, soybeans, hay, and fattening cattle. Clearly the calibration units differ. The four crop growing activities are calibrated to the land used, while the cattle fattening activity is calibrated to the number of cattle fattened. There are two inventory constraints for hay and corn. Hay is not sold and is fed to the cattle, but hay can be purchased at the going market price. Corn is more complex as it can be bought, grown, sold or fed to cattle. The amount of hay and corn that each steer needs to fatten is defined as a fixed input requirement, since we are still assuming a linear production technology. In the Midwest model we assume that land is the only resource constraint, although a constraint on the number of cattle that can be housed should also be included. Calibration is defined by prior estimates of supply elasticities and the observed cropping and livestock ac-

tivity levels in the base year data set. Since the productive activities are calibrated, the purchase and sale levels for hay and corn will also calibrate to the observed base levels. Policy parameterization of the model will result in the production activities changing with respect to their elasticities, and the levels in the transfer activities responding accordingly. To see this try loading the Midwest Gams template and increasing the marginal revenue for cattle parametrically. Note the changes in cattle, hay grown, and corn grown and sold.

5.5 Policy Modeling with PMP

The purpose of most programming models is to analyze the impact of quantitative policy scenarios which take the form of changes in prices, technology, or constraints on the system. The policy response of the model can be characterized by its response to sensitivity analysis and changes in constraints.

Advantages of the PMP specification are not only the automatic calibrating feature, but also its ability to respond smoothly to policy scenarios. Paris (1993) shows that the input demand functions and output supply functions obtained by parameterizing a PMP problem satisfy the Hicksian conditions for the competitive firm. In addition, the input demand and supply functions are continuous and differentiable with respect to prices, costs, and right hand side quantities. (At the point of a change in basis the supply and demand functions are not differentiable.) The continuity of input demand and output supply functions is in contrast to LP or stepwise problems. In linear problems the Dual values, and sometimes the optimal solution, are unchanged by parameterization until there is a discrete change in basis, when they jump discontinuously to a new level.

The ability to represent policies by constraint structures is important. The PMP formulation has the property that the nonlinear calibration can take place at any level of aggregation. That is, one can nest an LP subcomponent within the quadratic objective function and obtain the optimum solution to the full problem. An example of this is used in technology selection where a specification that causes discrete choices may be appropriate. Suppose a given regional commodity can be produced by a combination of five alternative linear technologies, whose aggregate output has a common supply function. A PMP model can calibrate the supply function while a nested LP problem selects the optimal set of linear technology levels that make up the aggregate supply (Hatchett et al., 1991).

Since the intersection of the convex sets of constraints for the main prob-

lem and the convex nested sub-problem is itself convex, then the optimal solution to the nested LP sub-problem will be unchanged when the main problem is calibrated by replacing the calibration constraints with quadratic PMP cost functions. The calibrating functions can thus be introduced at any level of the linear model. In some cases, the available data on base year values will dictate the calibration level. Ideally, the level of calibration would be determined by the properties of the production functions, as in the example of linear irrigation technology selection. The basic PMP approach does not replace all linear cost functions with equivalent quadratic specifications, but only replaces those that data or econometric estimates suggest are best modeled as nonlinear.

If the modeler has prior information on the nature of yield externalities and rotational effects between crops, they can be explicitly incorporated by specifying cross crop yield interaction coefficients in equations 5.42 and 5.43. The PMP yield slope coefficient matrix (Γ) is positive definite, $k \times k$, and of rank k. Without the cross crop effects the matrix is diagonal.

Resource using activities such as fodder crops consumed on the farm may be specified with zero valued objective function coefficients. Where an activity is not resource using, but merely acts as a transfer between other activities, there is no empirical basis or need to modify the objective function coefficients. An application of PMP to an example of joint crop and livestock production is developed in a later chapter.

Chapter 6

Nonlinear Duality, Prices, and Risk

6.1 Duality in Nonlinear Models

6.1.1 Deriving the Dual for Nonlinear Problems

The Standard PMP Problem defined in chapter 5 with nonlinear cost functions for all crops is concisely stated as:

$$\begin{array}{c} \max_x \; mr'x - \alpha'x - \frac{1}{2}x'\Gamma x \\ \text{subject to } Ax \leq b \end{array} \tag{6.1}$$

Given a nonlinear Primal problem, the Dual problem can be derived using the following steps:

1. Set up the problem as a Primal Lagrangian. (Here a maximization).

2. Apply the first Kuhn Tucker condition, $\frac{\partial L}{\partial x} \leq 0$, which yields the constraints for the Dual problem.

3. Apply the second Kuhn Tucker condition, $\left(\frac{\partial L}{\partial x}\right) x = 0$. Rearrange the equation to obtain $\lambda' Ax$ on the left hand side. Substitute the expression for $\lambda' Ax$ back into the Primal objective function. Multiply out and simplify to obtain the Dual objective function.

   The logic is that by taking the Primal first order conditions and objective function, and by substitution, expressing them in terms of prices and costs, we obtain the equivalent Dual problem.

Applying this procedure to equation 6.1, we get the following results:

1. Form a Lagrangian for the problem:

$$\begin{aligned} L &= mr'x - \alpha'x - \tfrac{1}{2}x'\Gamma x + \lambda'(b - Ax) \qquad (6.2)\\ \frac{\partial L}{\partial x} &= mr' - \alpha' - x'\Gamma - \lambda'A \end{aligned}$$

2. Apply Kuhn Tucker (KT) Conditions:

$$\frac{\partial L}{\partial x} \leq 0 \quad \left(\frac{\partial L}{\partial x}\right) x = 0$$

$$\begin{aligned} \frac{\partial L}{\partial x} = \quad & mr' - \alpha' - x'\Gamma - \lambda'A \quad \leq 0 \qquad (6.3)\\ \frac{\partial L}{\partial x} = \quad & mr - \alpha - \Gamma x - A'\lambda \quad \leq 0 \ldots \text{ 6.3 transposed} \\ & A'\lambda \qquad\qquad\quad \geq mr - \alpha - \Gamma x \qquad (6.4) \end{aligned}$$

Note that equation 6.4 states necessary conditions in terms of prices, variable costs and imputed costs and thus becomes the Dual problem constraint set.

3. Now we derive the Dual objective function. Using the second KT condition $\frac{\partial L}{\partial x}x = 0$ yields the following condition:

$$\begin{aligned} &(mr' - \alpha' - x'\Gamma - \lambda'A)x = 0 \\ &\therefore mr'x - \alpha'x - x'\Gamma x - \lambda'Ax = 0 \qquad (6.5) \end{aligned}$$

Since, the constraint holds exactly at the optimum, we know:

$$mr'x - \alpha'x - x'\Gamma x = \lambda'Ax \qquad (6.6)$$

Now substitute this into the Lagrangian (6.2) to get:

$$L = mr'x - \alpha'x - \frac{1}{2}x'\Gamma x - (mr'x - \alpha'x - x'\Gamma x) + \lambda'b \qquad (6.7)$$

cancelling terms and subtracting the quadratic terms gives:

$$L = \lambda'b + \frac{1}{2}x'\Gamma x \qquad (6.8)$$

The Dual Problem to the PMP production equation 6.1 is:

$$\begin{aligned} \min \quad & \lambda'b + \tfrac{1}{2}x'\Gamma x \qquad (6.9)\\ \text{subject to} \quad & A'\lambda \geq mr' - \alpha' - x'\Gamma, \quad \lambda \geq 0 \end{aligned}$$

6.1.2 The Economic Interpretation Nonlinear Dual Problems

The Dual Objective Function

Note that the Dual objective function has both Primal and Dual variables in it. The part with dual variables is familiar.

$$\lambda' b \tag{6.10}$$

Equation 6.10 has the same interpretation as in LP, i.e., the opportunity cost of firm's resources b. What about the $\frac{1}{2}x'\Gamma x$ term?

The producer's surplus from producing a quantity x of some crop is the area above the marginal cost function and below the marginal revenue for x, or the total revenue from producing x minus the total cost. Ignoring the shadow values of constraining resources for simplicity, producer's surplus is:

$$\text{Producer surplus} = mr'x - \alpha' x - \frac{1}{2}x'\Gamma x \tag{6.11}$$

but at the unconstrained optimum output marginal cost is equal to the marginal revenue (mr), therefore at the optimum we can substitute $mr' = \alpha' + x'\Gamma$. Using this in 6.11 yields:

$$\text{Producer surplus} = (\alpha' + x'\Gamma)x - \alpha' x - \frac{1}{2}x'\Gamma x = \frac{1}{2}x'\Gamma x \tag{6.12}$$

Therefore the PMP Dual objective function in 6.9 minimizes the sum of imputed resource value and producer's surplus

The Dual Constraints

$$A'\lambda \geq mr' - \alpha' - x'\Gamma \tag{6.13}$$

Where $A'\lambda$ is an $n \times 1$ vector of marginal opportunity costs of production of the vector of outputs x. $[A'\lambda + \alpha'] + [x'\Gamma]$ is [Marginal opportunity cost]+ [Marginal cash cost], and mr' is an $n \times 1$ vector of marginal revenue.

Therefore the PMP Dual Constraint says: "The sum of marginal costs of production must be equal to or greater than the marginal revenue for all outputs x." This is the "free lunch" theorem again. That is, if someone offers to sell you a Rolex for $19.95, there is probably something wrong with the deal.

Nonlinear Dual Values for Resources

For the binding constraints, substitute the basis matrix B (invertible) into equation equation 6.13, for A (non-invertible since not square). This gives a formula for λ on the binding constraints. Note that, from complementary slackness, $\lambda = 0$ for the slack constraints.

$$B'\lambda = mr' - \alpha' - x'\Gamma \tag{6.14}$$

Transposing 6.14 and inverting B yields:

$$\lambda' = B^{-1}mr - B^{-1}\alpha - B^{-1}\Gamma x \tag{6.15}$$

From 6.15 we see that in nonlinear programming problems, λ is a continuous function of x, unlike LP where the Dual vector is $\lambda = mr'_B B^{-1}$

6.1.3 Parameterizing Nonlinear Problems

Since the Dual values for the Primal Linear programming problem are:

$$\lambda = c_B B^{-1} \tag{6.16}$$

This implies that a particular Dual value λ_i of b_i is constant until the basis B^{-1} changes (Note that c_B is a vector of linear net revenues).

The quadratic Primal PMP problem, however, has a Dual of:

$$\lambda' = B^{-1}mr - B^{-1}\alpha - B^{-1}\Gamma x \tag{6.17}$$

$$\text{or} \quad \lambda' = B^{-1}(mr - \alpha) - B^{-1}\Gamma x \tag{6.18}$$

where α and Γ are the intercepts and slopes of the cost functions for x_i, and the vector mr is defined here as the constant marginal revenue/unit from producing x.

The Dual λ_i is now a linear function of x, therefore as b_i changes and x changes, the Dual will change. However, the "intercept" $B^{-1}(mr - \alpha)$ and the "slope" $B^{-1}\Gamma$ of the Dual function will change when the basis changes.

Condensing the notation by defining $\mu \equiv B^{-1}(mr - \alpha)$ and $\xi \equiv B^{-1}\Gamma$, the Dual for the monopoly problem becomes:

$$\lambda = \mu + \xi x \tag{6.19}$$

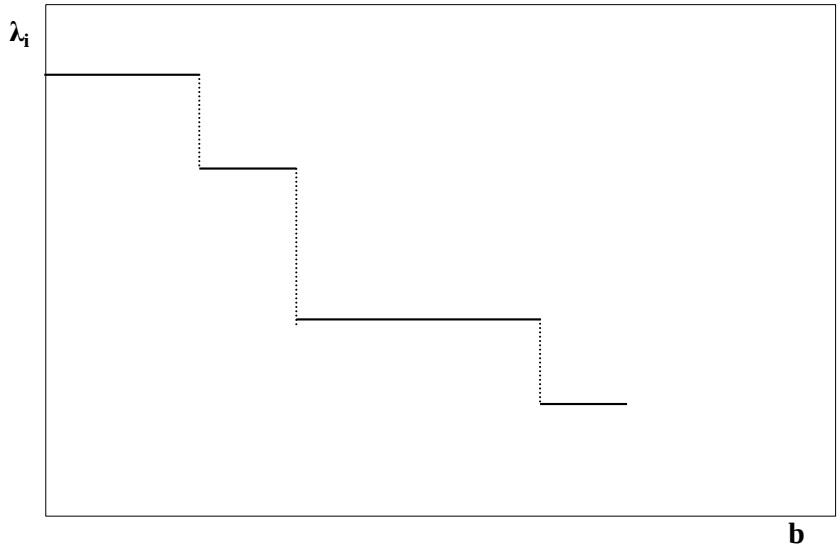


Figure 6.1: Parameterization of b_i

6.1.4 An Empirical Example

The following simple Quadratic problem illustrates the nonlinear dual (also see Figure 6.2):

$$\begin{aligned} \max \quad & \phi' x - \tfrac{1}{2} x' D x \\ \text{subject to} \quad & Ax \leq b, \quad x \geq 0 \end{aligned}$$

$$\text{where} \phi' = [8\ 6\ 4], \quad D = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 0 \\ 2 & 0 & 2 \end{bmatrix}, \quad A = [1\ 1\ 2]$$

and b is parameterized over the values $0 \rightarrow 4$. As one would expect from the equations 6.18 and 6.19, the quadratic Dual is a continuous linear function of b_j within a basis, since x = B^{-1}b, and from 6.19 we see that lambda is a continuous function of x. With a change of basis, the slope of the linear function changes discretely. Equation 6.19 shows that both the intercept and slope of the Dual value function changes with a change in the basis B.

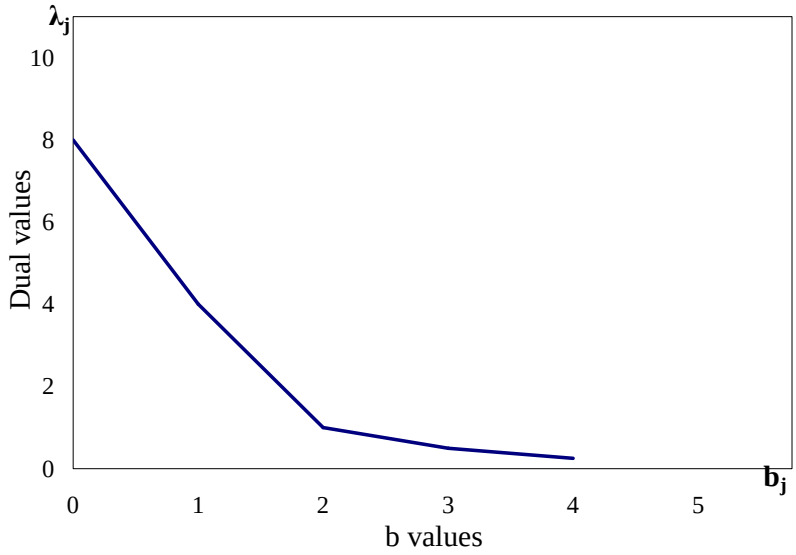


Figure 6.2: QP Parameterization - Plot of the Dual value and RHS

6.2 Incorporating Endogenous Supply and Demand Prices

Linear Programs assume constant prices and costs. In reality marginal costs are rarely constant and output prices are only constant for individual firms. Analysis which is performed on a regional, national, or commodity basis should have prices that change with changes in the solution.

6.2.1 Case I: Changes in Output Price Only

Assuming that we are given, or have estimated the parameters of a linear demand function, we can relate the quantity of output demanded q with its price p as $q = a + Sp$ where p and q are vectors of prices and quantities; the parameter a is a vector of positive slope intercepts; and S is a negative-definite matrix of demand slopes and cross demand effects.

Note: A matrix S is *negative definite* if the scalar product $k'Sk < 0$ for all non-zero values in a conformable vector k.

Since we are modeling the aggregate outcome of individual farmer behavior, we are interested in the opposite effect — i.e., how output levels q affect the price received. This assumption implies that farmers are so small

in their individual output that they are price takers. In addition, for most agricultural crops, the farmer has to commit to purchasing the inputs before it is clear what the price will be at harvest. Therefore we invert the demand function to get it in a price dependent form.

$$p = -S^{-1}a + S^{-1}q$$

$$\text{or} \ldots\ p = \phi + Dq \text{ where} \phi \equiv -S^{-1}a \text{ and} D \equiv S^{-1}$$

Assuming a constant yield per acre for the moment, we can replace the output quantity (q) by the number of acres allocated to a crop (x) and substitute the resulting expression for price into the objective function. The price endogenous objective function differs with the assumptions on the objectives of the decision maker. The two main specifications are (i) the objectives of a monopolist and (ii) the objectives of a perfect competitor:

A Monopolist Objective Function

Assume a monopolist faces a set of linear, price-dependent demand functions for the vector of outputs x and where D is a symmetric negative-definite matrix results in.

$$\text{Demand System} \quad p = \phi + Dx \tag{6.20}$$

If the monopolist has a constant marginal cost of production vector c, the net revenue objective function will be:

$$\max J = p'x - \omega'x \tag{6.21}$$

Substituting in 6.20 for p gives:

$$\begin{aligned} \max J &= (\phi + Dx)'x - \omega'x && (6.22)\\ \text{or} \max J &= \phi'x + x'Dx - \omega'x && (6.23)\\ \text{subject to} \quad & Ax \leq b, \quad x \geq 0 \end{aligned}$$

1. Unconstrained Equilibrium
   Set the derivative $\frac{\partial J}{\partial x} = 0$ and transpose it. In this problem there are no binding resources (the constraints are all slack), therefore:

$$\begin{aligned} \text{Marginal Revenue} &= \text{Marginal Cost} \\ \phi + 2Dx &= \omega && (6.24)\\ \text{Vector of Marginal Revenues} &= \text{Vector of Marginal Costs} \end{aligned}$$

Also, monopoly rent = total revenue - total cost. Therefore:

$$\begin{aligned} \text{Monopoly rent} &= p'x - \omega'x \\ &= (\phi + Dx)'x - (\phi + 2Dx)'x \\ \text{Monopoly rent} &= -x'Dx \end{aligned}$$

We know from 6.20 that $p = \phi + Dx$ and $\omega = \phi + 2Dx$ from 6.24 in addition, we know that a monopolist produces where $MR = MC$. Also note that $x'Dx$ is a positive scalar value since D is a negative definite matrix.

2. Constrained Monopoly Equilibrium
The monopolist is now constrained by a vector of fixed inputs b and the linear technology matrix A. The Lagrangian now becomes:

$$\max J = \phi'x + x'Dx - \omega'x + \lambda'(b - Ax) \tag{6.25}$$

The first order optimum conditions are:

$$\frac{\partial J'}{\partial x} = \phi + 2Dx - \omega - A'\lambda \stackrel{set}{=} 0 \tag{6.26}$$

$$\phi + 2Dx - \omega = A'\lambda \tag{6.27}$$

That is, the difference between the marginal revenue and marginal cost of an output is the sum of shadow values of the inputs used to produce it. This shows that even monopolists exhaust the rents from a fixed input.

The Perfect Competition Objective Function

1. Unconstrained Perfect Competition

The unconstrained perfectly competitive equilibrium condition is defined as where:

$$\begin{aligned} \text{Price} &= \text{Marginal Cost} \\ \phi + Dx &= \omega \end{aligned} \tag{6.28}$$

The perfectly competitive objective function is obtained by integrating the optimal marginal condition from 6.28, namely:

$$\int (\phi + Dx - \omega)\,\mathrm{d}x = \phi'x + \frac{1}{2}x'Dx - \omega'x$$

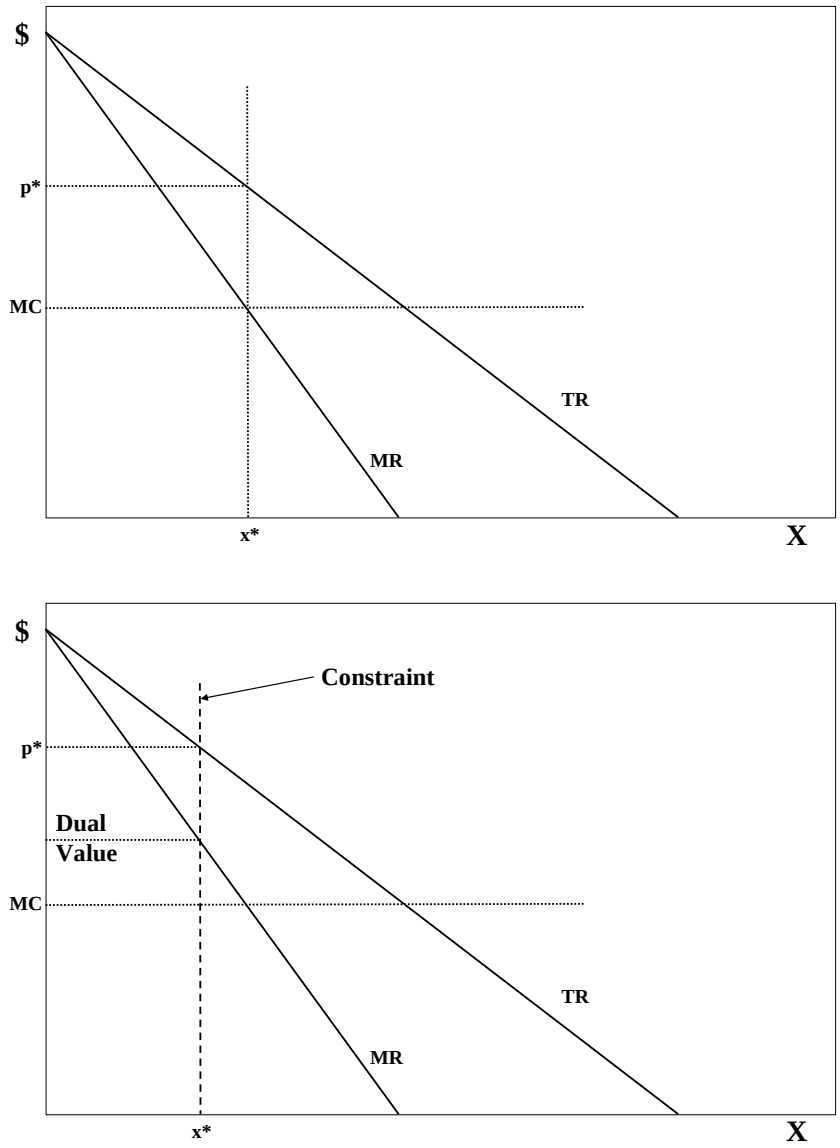


Figure 6.3: Monopoly Producer: Unconstrained (top) & Constrained (bottom)

Accordingly, we specify a different objective function that satisfies the marginal conditions for unconstrained perfect competition.

$$\max z = (\phi + \frac{1}{2}Dx)'x - \omega'x \tag{6.29}$$

A good question is: Why is the one half in the objective function multiplying the slope parameter? The answer is that a Perfectly Competitive market is defined by its marginal conditions, so to correctly define the objective function, we have to start with the marginal conditions and derive the objective function. Essentially we have to ask, what objective function would an optimizing decision maker have had to have, to result in the perfectly competitive first order conditions? We therefore start with the marginal conditions and integrate them to obtain the objective function.

2. Constrained Perfect Competition

$$\begin{aligned} \max z \quad &= \phi'x + \tfrac{1}{2}x'Dx - \omega'x \\ \text{subject to} \quad & Ax \leq b, \quad x \geq 0 \end{aligned} \tag{6.30}$$

The perfectly competitive objective function also maximizes the sum of consumer's surplus and producer's quasi-rent (producer surplus) Initially, assume an unconstrained solution for simplicity.

$$\begin{aligned} z = \quad & (\phi + \tfrac{1}{2}Dx)'x - \omega'x \\ = \quad & (\phi + Dx)'x - \omega'x - \tfrac{1}{2}x'Dx \quad (\text{add and subtract}\frac{1}{2}x'Dx) \\ = \quad & p'x - \omega'x - \tfrac{1}{2}x'Dx \quad (\text{substitute } p \text{ from 6.28}) \\ = \quad & (p - \omega)'x - \tfrac{1}{2}x'Dx \end{aligned} \tag{6.31}$$

Since Price minus Variable Cost is defined as producer's surplus or "Quasi Rent", the left-hand-side term $(P-\omega)'x$ is equal to producer's surplus. Since the marginal cost is defined as a constant value "ω", the producer's surplus only occurs because of the constraint on the amount of product sold.

What about the right-hand-side term, $\frac{1}{2}x'Dx$?

$$\begin{aligned} -\frac{1}{2}x'Dx &= \tfrac{1}{2}x'(-Dx) && (6.32)\\ &= \tfrac{1}{2}x'(\phi - \phi - Dx) && \text{(add and subtract}\phi\text{)}\\ &= \tfrac{1}{2}x'(\phi - p^*) && \text{(substitute } p^* \text{ from 6.28)}\\ -\frac{1}{2}x'Dx &= \text{consumer's surplus} && \text{(see Figure 6.4)} \end{aligned}$$

Thus equation 6.28 satisfies competitive marginal conditions and maximizes aggregate net social benefits, which are usually defined as the sum of producer and consumer surplus.

SUMMARY

Unconstrained Perfect Competition	Constrained Perfect Competition
Producer Surplus	Producer Surplus
(p.s.) $= 0$	(p.s.) $= (p_i - \omega_i)x_i$
Consumer Surplus	Consumer Surplus
(c.s.) $= \frac{1}{2}(\phi_i - p_i)x_i$	(c.s.) $= \frac{1}{2}(\phi_i - p_i)x_i$

6.2.2 Case II: Aggregate Perfect Competition — Endogenous Prices and Costs

In this specification the marginal costs are no longer constant and there is a linear supply function (Marginal Cost) as well as endogenous demand prices.

Given the price dependent demand function and a linear marginal cost (supply) function below:

$$\begin{aligned} \text{Price} &= \phi + Dx && D \text{ is negative definite}\\ \text{Marginal Cost} &= \alpha + \Gamma x && \Gamma \text{ is positive definite.} \end{aligned} \qquad (6.33)$$

We integrate equation 6.33 and obtain the unconstrained perfectly competitive problem with supply functions as:

$$\max J = \phi'x + 1/2x'Dx - \alpha'x - \tfrac{1}{2}x'\Gamma x \qquad (6.34)$$

$$\frac{\partial J}{\partial x} = \phi' + x'D - \alpha' - x'\Gamma \overset{set}{=} 0 \qquad (6.35)$$

As expected for perfect competition, price = marginal cost.

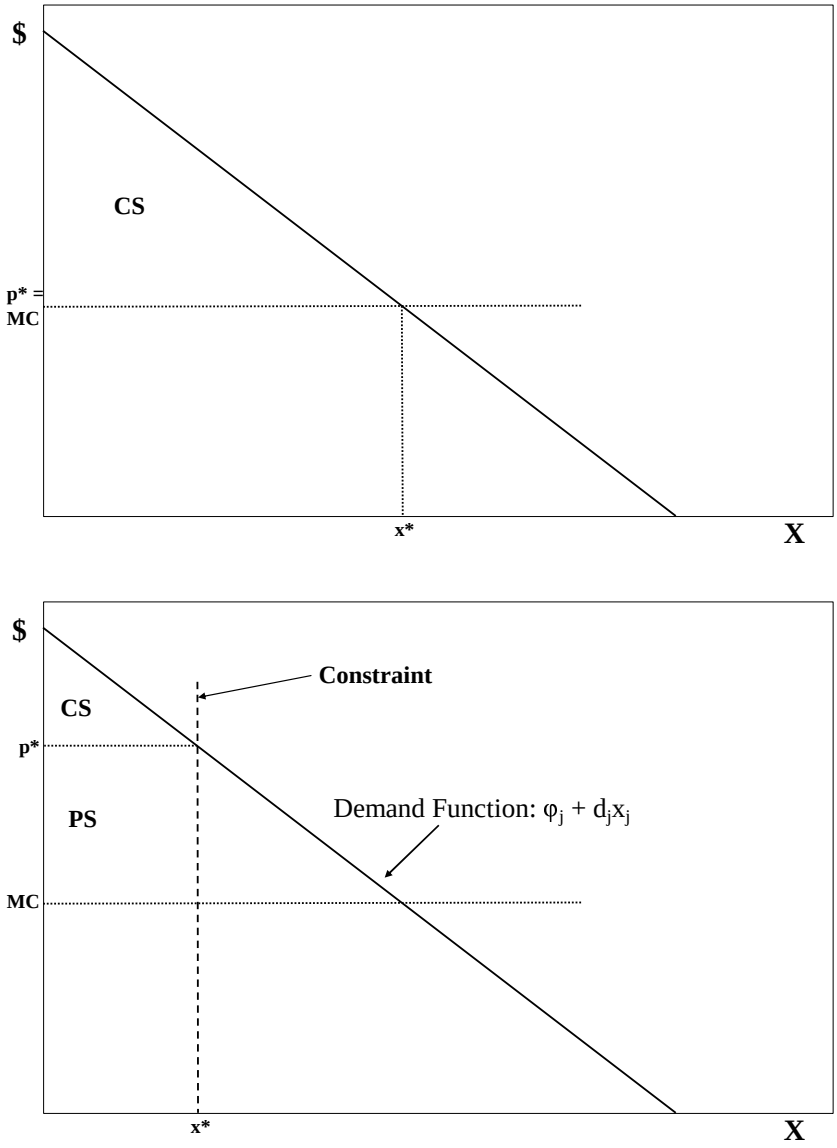


Figure 6.4: Perfectly Competitive Aggregate Producers of a Single Constant Cost Product

The Interpretation of the Objective Function

Trick #1. Add and subtract $1/2x'Dx$ and $1/2x'\Gamma x$ to 6.33 to yield:

$$J = \phi'x + x'Dx - \alpha'x - x'\Gamma x - 1/2x'Dx + 1/2x'\Gamma x \tag{6.36}$$
$$J = (\phi + Dx)'x - (\alpha + \Gamma x)'x - \tfrac{1}{2}x'Dx + \tfrac{1}{2}x'\Gamma x \tag{6.37}$$

But from 6.35, we see that the first two terms cancel out (Price = Marginal cost) at the optimum. Therefore at the optimum the objective function is:

$$J = -1/2x'Dx + 1/2x'\Gamma x \tag{6.38}$$

From 6.32, we see that the first term ($\frac{1}{2}x'Dx$) is equal to consumer's surplus. The second term is changed by trick #2: factor out x and add/subtract α to yield:

$$\begin{aligned} \frac{1}{2}x'\Gamma x &= \tfrac{1}{2}(\alpha + \Gamma x - \alpha)'x \\ &= \tfrac{1}{2}(MC - \alpha)x \end{aligned} \tag{6.39}$$

Since price equals marginal cost, $\frac{1}{2}x'\Gamma x$ is one-half the area above the marginal cost intercept and below the price line as shown in Figure 6.5. Note that this term and logic is identical to the original nonlinear PMP Dual derivation at the beginning of the chapter.

6.2.3 Case III: Inter-regional Trade Models

Empirical trade models can be solved using this same objective function expanded to several regions. The effect of transport costs and tariffs can be added to the supply functions to solve for changes in trade policies or conditions. The seminal paper in this area is Takayama and Judge (1964).

There are several different ways of setting up the inter-regional trade problem, but the simplest method to show is to extend the quantity dependent supply-demand concept above to J regions that are linked by trade which results in a trading cost of c_{ij} per unit commodity traded from region i to region j. The resulting problem solves a wide range of trade problems.

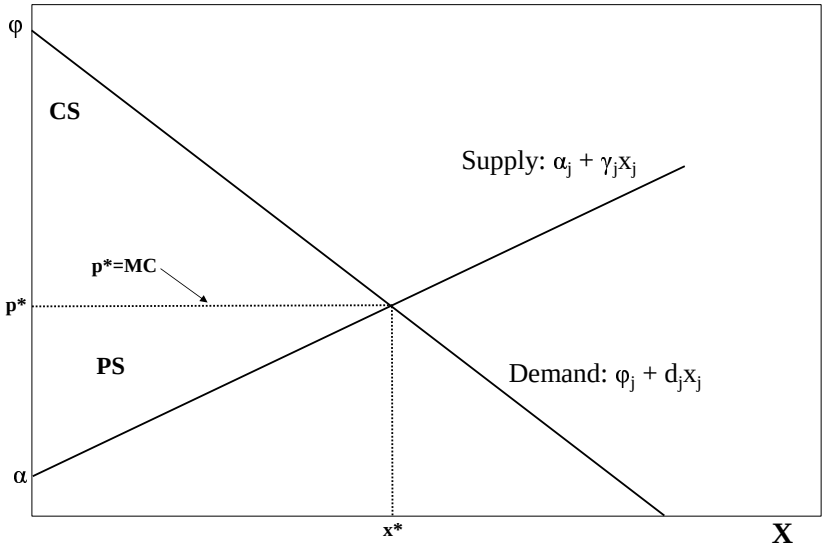


Figure 6.5: Perfect Competition — Endogenous Price and Cost

Price Dependent Optimal Inter-regional Trade Specification

$$maxF(\cdot) = \sum_{j=1}^{J}(\phi_j + \frac{1}{2}d_j\ xd_j)\ xd_j \quad - \sum_{i=1}^{I}(\alpha_i + \tfrac{1}{2}\gamma_i\ xs_i)\ xs_i \tag{6.40}$$

$$- \sum_i \sum_j c_{ij}x_{ij} \tag{6.41}$$

$$\text{subject to } xs_i = \sum_j x_{ij} \quad xd_j = \sum_i x_{ij}, \text{ and } x_{ij} \geq 0 \tag{6.42}$$

$$\tag{6.43}$$

Note that the only constraints are those needed to aggregate the individual demands xd_j and supplies xs_i for the aggregate regional demand and supply functions.

The objective function maximizes the definite integral under the demand and supply function for each region in terms of the post trade quantities demanded and supplied in each region. The costs of trading between regions is deducted to yield the net social benefit of trade shown in the two region diagram. The adding up constraints ensure that the quantities demanded and supplied in each region balance.

The specification of the regional trade problem is an excellent illustration of the efficiency and beauty of Dual specifications. The Primal specification above solves optimally, but it is more complicated than needed. Since the decision variable is the quantity of product traded between regions, the cost of trading is explicitly defined in the objective function, and the aggregate quantities are generated by the summing up constraints.

An alternative to using price dependent supplies and demands, is to formulate the Dual of the price dependent problem that can be termed the quantity dependent form, since it uses the standard quantity dependent demands and supplies. The quantity dependent specification solves the problem with two simple equations that use the standard quantity dependent demand and supply functions, and instead of solving for $i \times j$ quantities traded, the price based model solves for the $i + j$ set of equilibrium prices.

The two equations are the producer and consumer surplus objective function, and the first order price condition for trade. Returning to the original price dependent demand 6.41 we have:

$$xd_j = a_j + s_j \; pd_j$$

We also define a similar quantity dependent supply function:

$$xs_i = b_i + g_i \; ps_i$$

The Quantity Dependent Interregional Trade Problem

$$\max F(\cdot) = \sum_{j=1}^{J}(a_j + \frac{1}{2}s_j \; pd_j)pd_j - \sum_{i=1}^{I}(b_i + \frac{1}{2}g_i \; ps_i)ps_i$$

$$\text{subject to } pd_j - ps_i - c_{ij} \leq 0$$

Note that in the quantity dependent formulation, the constraints are the Kuhn-Tucker conditions for each possible trade. They contain the aggregate demand and supply prices, but also the trade specific transport costs (c_{ij}).

The quantities traded are generated as the Dual values to the $(i \times j)$ interregional pricing constraints. This is an example of the complementary slackness principle working for us. From the complementary slackness principle we know that if the trade price constraint is slack, that is the supply and transport cost exceed the demand price, then the quantity traded will be zero. The corollary is that the Dual value, when the trade price constraint is binding, is the quantity traded. Note that all the information from the

optimal solution of the Primal problem is also obtained from solving the simpler Dual problem.

An empirical example of the GAMS code for these problems is found in the Gams Templates on the class webpage. By running both problems it can be seen that the results are identical.

6.2.4 Calibrating Demands Using Elasticity Estimates

Often the modeler is faced with the need to derive parameters for demands when the only data available is the equilibrium price and quantity in the base year for the model, and an estimate of the elasticity from an previous econometric study. Similar to the PMP development in the previous chapter, there is a very simple derivation that enables one to calibrate the slope and intercept coefficients using the elasticity. Assume that a demand can be specified as linear in the price dependent form used above.

$$pd = \phi + dx$$

Recall that demand elasticity is defined as:

$$\eta = \frac{\partial q}{\partial p}\frac{p}{q}$$

If you have a base value for $q = 2.9, p = 173$, and $\eta = -0.6$ the values can be substituted in to the above formula to solve for the slope of the demand function:

$$-0.6 = \frac{\partial q}{\partial p}\frac{173}{2.9}$$

The slope of the price dependent linear demand function that results is:

$$\frac{\partial p}{\partial q} = -99.42$$

Substituting this value back into the original price dependent demand equation results in an intercept value of 461.3.

The resulting calibrated demand curve that will yield a price of $173 at a quantity of 2.9 and a point elasticity of -0.6 has the form:

$$pd = 461.3 - 99.42q$$

The concept of calibrating models against prior econometric elasticity estimates is well illustrated in the Central Valley Production Model (CVPM) that has been developed by S. Hatchett in the consulting firm CH2M -

Hill. The CVP model is currently used to analyze the economic impacts of large reallocations of water between agricultural, environmental and urban interests in California. The model use both demand, supply and substitution elasticities to reflect the economic impacts of changes in water availability and cost in California.

6.3 Incorporating Risk and Uncertainty

In our problem specifications we have implicitly assumed that the parameters in the problem are known and constant. For example, the objective function $c'x$ assumes a deterministic vector c of prices or costs. We can make this more realistic in two ways:

1. Assume the elements of c are not known with certainty, but they are stochastic with known distributions.

$$c \sim N(\bar{c}, \Sigma_c)$$

   where Σ_c is an $n \times n$ variance/covariance matrix of revenues.

2. The decision maker's objective function values both the expected return and its variance.

Question:
If the vector of net revenues c is distributed $c \sim N(\bar{c}, \Sigma_c)$, what is the distribution of the objective function $c'x$, where x is a non-stochastic vector?

Answer:
Since x is a deterministic linear operator;

Expected Value of $c'x = E(c'x) = \bar{c}'x$.

Variance of $c'x = E[c'x - E(c'x)]^2$

$$\begin{aligned}
Var(c'x) = &\quad E\{[c'x - E(c'x)]'[c'x - E(c'x)]\} && \text{Inner product} \\
= &\quad E\{x'[c - E(c)][c - E(c)]'x\} && \text{Inner product of an outer product} \\
= &\quad x'E\{[c - E(c)][c - E(c)]'\}x && \text{Take expectation of stochastic terms} \\
= &\quad x'\Sigma_c x && \text{by covariance matrix definition}
\end{aligned}$$

Therefore, if $c \sim N(\bar{c}, \Sigma_c)$ then $c'x \sim N(\bar{c}'x, x'\Sigma_c x)$ where x is deterministic.

If the decision maker is risk neutral, the objective function can be specified by substituting $\bar{c}'$ for c'. More frequently, the decision maker is risk

averse. If the degree of aversion to income variance can be measured by a parameter ρ, then the problem becomes:

$$\begin{array}{lccc} \max z = & \bar{c}'x & & -\rho\ x'\Sigma_c x \\ & \uparrow & & \uparrow \\ & \text{expected revenue vector} & & \text{variance of revenue} \\ \text{subject to} & Ax \leq b, \quad x \geq 0 & & \end{array}$$

6.3.1 The Effect of Uncertainty and Risk Aversion

If there is a nonlinear cost of risk, it will have several different effects on the optimal solution of the problem.

- The optimal solution will show more diversification to offset risk.
- Since the problem is no longer linear in x, some x_j may have interior optima and will not be restricted by binding constraints. In this case, there will be more x_j activities than constraints.
- Note that ρ is a scalar, since the variance of a vector product $(c'x)$ is a scalar. ρ measures the cost of risk to the decision maker. The variance of the return from the portfolio $x'\Sigma_c x$ and the expected return $\bar{c}'x$ are both changed by changing the proportions of x_i in the portfolio.

The point is demonstrated by focusing on a single x_i, and denoting the sum vector of derivatives that result from the covariance by the short hand expression $\frac{\partial(var)}{\partial x_i}$.

$$\frac{\partial z}{\partial x_i} = \frac{\partial(\bar{c}'x)}{\partial x_i} - \rho\frac{\partial(var)}{\partial x_i} \stackrel{set}{=} 0$$

$$\rho\frac{\partial(var)}{\partial x_i} = \frac{\partial(\bar{c}'x)}{\partial x_i}$$

$$\therefore \rho = \frac{\partial(\bar{c}'x)}{\partial x_i}\frac{\partial x_i}{\partial(var)} = \frac{\partial(\bar{c}'x)}{\partial(var)} = \frac{1}{\lambda}$$

ρ is equal to the marginal rate of tradeoff between expected income and variance at the optimum.

6.3.2 Measuring Risk Aversion- E/V Analysis

The risk aversion parameter can be calculated for an individual by solving the following problem, termed an E/V problem since it minimizes the variance of revenues, subject to a pre-specified expected revenue:

$$\begin{aligned} \min v = \quad & x'\Sigma_c x \\ \text{subject to} \quad & \bar{c}'x \geq e^* \end{aligned}$$

Where: $x'\Sigma_c x$ = variance of revenue, $\bar{c}'x$ = expected revenue and e^* = the value chosen for the minimum expected revenue.

Expressing this problem as a Lagrangian:

$$\begin{aligned} L &= Var + \lambda(e^* - \bar{c}'x) \\ \frac{\partial L}{\partial x_i} &= \frac{\partial (var)}{\partial x_i} - \lambda \frac{\partial (\bar{c}'x)}{\partial x_i} \overset{set}{=} 0 \\ \therefore \lambda &= \frac{\partial (Var)}{\partial (\bar{c}'x)} = \frac{1}{\rho} \end{aligned}$$

In other words, the risk aversion parameter $\rho = \frac{1}{\lambda}$, where λ = opportunity cost of constraint when you minimize the variance of the objective function. By parameterizing the model over a range of values for e^* , we can generate "E/V" frontiers that show those output combinations that give the best combination of expected revenue (E) and variance of revenue (V). See Figure 6.6.

The E/V frontier is generated in two steps:

1. Min $var(c'x)$ s.t. $\bar{c}'x \geq e^*$ for a range of e^* values.

2. Plot results.

A simple example of the Gams code for the E/V problem is found in the Gams templates on the class webpage.

Note: There is a linear approximation to the mean/variance objective function called MOTAD (see Hazell and Norton (1986, 86-90)). With the growth in nonlinear algorithms and computing power, this linear approximation is rarely needed.

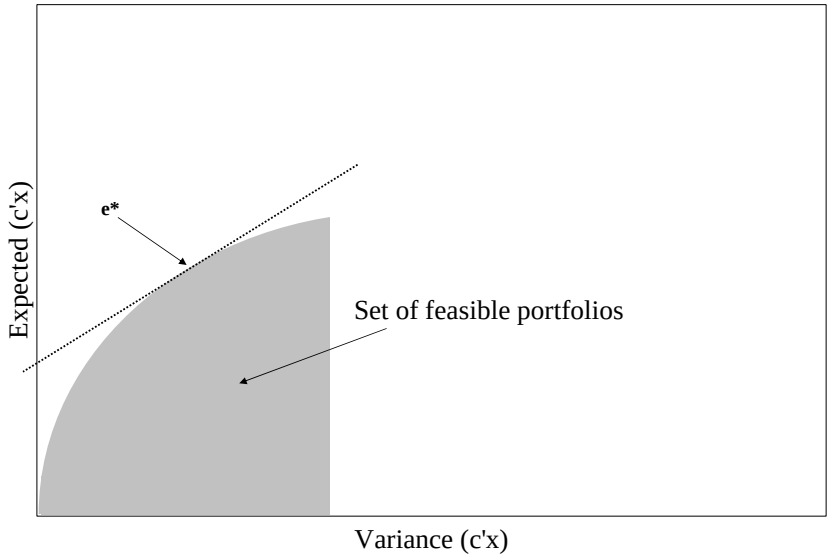


Figure 6.6: E/V Frontier

6.3.3 Uncertainty in the Constraints: Chance Constrained Programming

The constraints of a problem sometimes are not known with certainty. Often the quantity available of input resources (b_i) is uncertain, for farming this may be reflected by the distribution of growing season length, rainfall or seasonal labor availability. The a_{ij} technological coefficients may also be stochastic, but this more complex case is set aside for the moment.

Case: A Single Right Hand Side Value b_i is Normally Distributed

Two assumptions are required in this specification.
First: Some or all of right-hand-side b_i values are stochastic. Their distribution is known.

Second: The problem decision maker has specified that the problem solution must have the uncertain constraint satisfied for a known proportion $(1 - \alpha)$ of the time. That is the probability that the constraint is satisfied is specified at a particular level.

Probability Review

Any normal random variable can be transformed to a variable that has a Standard Normal Distribution, $\sim N(0,1)$, whose cumulative probabilities are calculated and tabulated in the z distribution, i.e.:

$$b_i \sim N(\bar{b}_i, \sigma^2_{b_i}) \Rightarrow \frac{b_i - \bar{b}_i}{\sigma_{b_i}} \sim N(0,1)$$

.

Also the probability that a z-distributed random variable exceeds a specified value is equal to one minus the cumulative probability at that value.

Z Table Review

Z tables are usually set up so that:

- They only tabulate half the distribution, so you have to add or subtract $\frac{1}{2}$ from the cumulative probability.
- They give the cumulative probability in the "tail" of the distribution.

If the variable z_i is distributed:

$$z_i \sim N(0,1) \tag{6.44}$$

That is, z_i is distributed with a standard normal distribution, and $\text{Prob}\{z_i \geq Q\} \geq \alpha$ holds only if $Q \leq k_\alpha$ and where the cumulative probability (α) is the area under the curve (density function) in the tail beyond k_α. See Figures 6.7 and 6.8

The Problem

Given the problem:

$$\begin{aligned} &\max c'x \\ &\text{subject to } Ax \leq b \end{aligned} \tag{6.45}$$

the ith row in the constraints is written using summation notation for simplicity:

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i \tag{6.46}$$

But if b_i is stochastic and we want this constraint to hold with probability (1- α), where α is defined as the critical region, it is rewritten as:

$$\text{Prob} \left\{ b_i \geq \sum_{j=1}^{n} a_{ij} x_j \right\} \geq 1 - \alpha \tag{6.47}$$

This is the probability that the constraint is slack, or just holds. Note that the right hand side of the expression in the brackets $a_{ij}x_j$ is deterministic but changes with changes in x.

Step 1 If we require the probability that a normally distributed random variable will be equal to or bigger than the specified value (1- α), we need to calculate a value where cumulative probability is α, our specified level. It is quick and convenient to convert the distribution to an equivalent standard normal and look up the probabilities in a table.

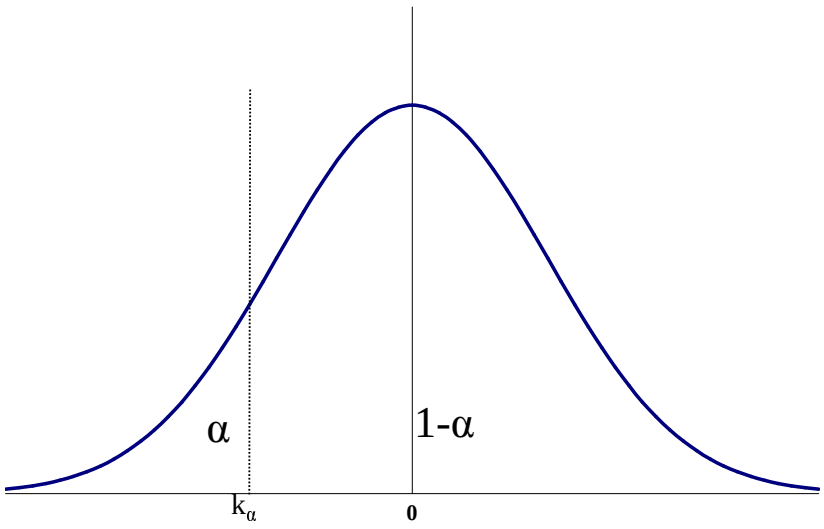


Figure 6.7: Z distributions and critical regions

Step 2 Convert the normally distributed b_i to an equivalent standard normal distribution, as above. b_i is "standardized" to:

$$z_i = \frac{b_i - \bar{b}_i}{\sigma_{b_i}} \tag{6.48}$$

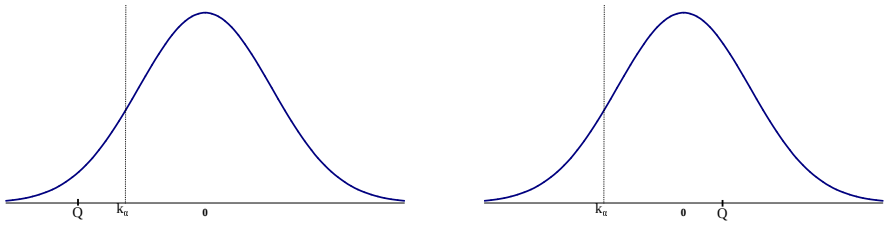


Figure 6.8: Left: Low Q Value where Prob$\{z_i \geq Q\} > (1-\alpha)$ and Right: High Q Value where Prob$\{z_i \geq Q\} < (1-\alpha)$.

Since we want to put this standardized value into constraint 6.46 on the left hand side,we have to perform the same b_i standardization transformation on the other side of the constraint ($\sum_{j=1}^{n} a_{ij}x_j$). Applying the transformation to both sides, constraint 6.46 is rewritten for the standard normal distribution as equation 6.47. Note that the left-hand term in the brackets is a random variable, while the right-hand term is deterministic.

$$\text{Prob}\left(\frac{b_i - \bar{b}_i}{\sigma_{bi}} \geq \frac{\sum_{j=1}^{n} a_{ij}x_j - \bar{b}_i}{\sigma_{bi}}\right) \geq (1-\alpha) \tag{6.49}$$

Notes

- $\frac{b_i - \bar{b}_i}{\sigma_{b_i}} \sim N(0,1)$. That is, it has a standard normal distribution.
- $\frac{\sum_{j=1}^{n} a_{ij}x_j - \bar{b}_i}{\sigma_{bi}}$ is composed of fixed and known parameter values and thus is equivalent to the z value "Q" used above.
- From Step 1 and equation 6.47, we see that Prob $\{z_i \geq Q_i\} \geq (1-\alpha)$, holds only if $Q_i \leq k_\alpha$. Therefore, equation 6.49 only holds if:

$$\frac{\sum_{j=1}^{n} a_{ij}x_j - \bar{b}_i}{\sigma_{bi}} \leq k_\alpha \tag{6.50}$$

Step 3 Multiplying out (6.50), we see that (6.49) only holds if:

$$\sum_{j=1}^{n} a_{ij}x_j - \bar{b}_i \leq k_\alpha \sigma_{bi} \tag{6.51}$$

rewritten as

$$\sum_{j=1}^{n} a_{ij}x_j \leq \bar{b}_i + k_\alpha \sigma_{bi} \tag{6.52}$$

An intuitive explanation is to think of $k_\alpha \sigma_{bi}$ as a risk adjustment factor that changes the probability of the constraint binding. For example, if $k_\alpha \sigma_{bi} = 0$, the constraint would bind 50 percent of the time.

Notes

- The left hand side is the familiar Ax constraint.
- The right hand side is the original b_i value adjusted by a term that is linear and can be calculated from the Z tables and the distribution of b_i.
- The adjusted constraint holds with a probability α.

A Numerical Example

In this example we work backwards, namely we are given a right hand side value, and want to know what the probability is that it is satisfied. Suppose b_i is distributed $b_i \sim N(42, 4)$, does a resource requirement $\Sigma a_{ij} x_j = 40.25$ hold with probability 0.8?

Step 1 Find k_α, using the Z tables. The tables for $\alpha = 0.2$ give us a value of 0.85, which we deduce from symmetry is -.85. see the Z table review if this is not clear. [1] $\therefore k_{0.2} = -.85$

Step 2 Find Q using values from the distribution of b_i Q =

$$\frac{\sum_{j=1}^{n} a_{ij} x_j - \bar{b}_i}{\sigma_{bi}} = \frac{40.25 - 42}{2} = -0.875$$

Step 3 Comparing k_α and Q, we see that $Q < k_\alpha$, i.e., $-.875 < -.85 \therefore$ we know that the constraint value of 40.25 will be satisfied slightly more than 80% of the time.

6.3.4 Uncertainty in the Technical Coefficients

This special case requires two assumptions about the technology embodied in the a_{ij} coefficients:
First: Individual a_{ij} have a known mean and variance
Second: Two or more a_{ij} values have a non-zero covariance.
Given these assumptions, we can use the same probability concept as with the stochastic b_i values. However, in this case, the derivation results in nonlinear (quadratic) constraints in x.

[1] Note the 0.8 and 0.85 values in this problem are completely different parameters.

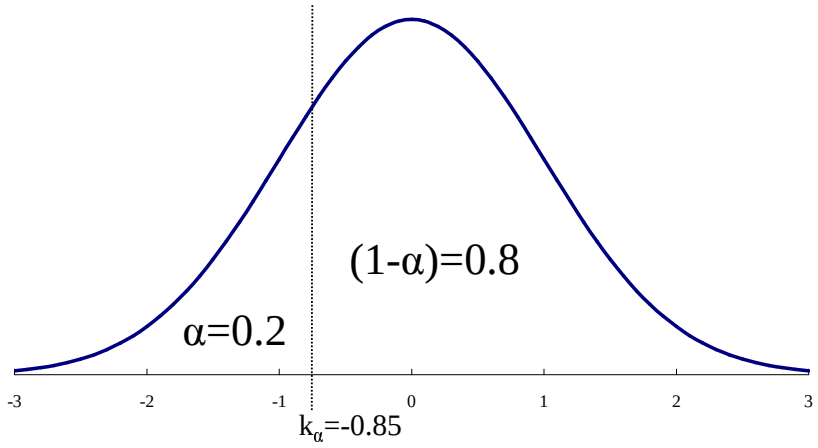


Figure 6.9: A Numerical Example

For details see Hazell and Norton (1986, 107-110). For a more general model specification that combines the mean/variance objective function and chance constraints on the input supply, see Paris (1979).

Chapter 7

Calibrating Models with Nonlinear Yield and Production Functions

Despite using a nonlinear calibration procedure, the models used so far are based on a production technology that is linear and does not allow any substitution among inputs. This section relaxes this stringent and unrealistic assumption and intro duces three ways in which more general production specifications can be incorporated in calibrated models.Howitt (1995a) We develop three cases. In the first two cases,land use is assumed to remain linear, but the relationship between particular inputs is known a priori. In the first case we want to incorporate econometric estimates of yield functions, and in the second case we know the isoquant function that shows the possible tradeoffs between water use per acre and the capital and management costs that result in a constant yield. Finally, the third approach is to define a CES production function for all the production inputs. Given that the elasticity of substitution is known, and the base input allocations, input and output prices are also known CES functions can be calibrated analytically. We present a simple example, and a more complex nested example that enables the modeler to specify different elasticities of substitution between different sets of inputs.

Agricultural models that are used for policy analysis are often required to be disaggregated by region, commodity and input use. The level of disaggregation depends on the policy, but for analysis of the interaction between agricultural price supports and environmental outcomes, the model requirements frequently exceed the capacity of the data base for direct estimation.

In this case, the modeler has to use formal or informal calibration methods to match the model outcome to the available data base. In microeconomic modeling the process of calibrating models is widely practiced, but rarely formally discussed. In contrast, calibration methods for macroeconomic models have stimulated an emerging literature. Hoover (1995) provides a survey and analysis of the contending viewpoints. Gregory and Smith (1993) conclude that "Studies which use calibration methods in macroeconomics are now too numerous to list, and it is safe to say that the approach is beginning to predominate in the quantitative application of macroeconomic models." In an earlier paper these same authors (Gregory and Smith, 1990) define calibration as involving the choice of free parameters in a model by matching certain moments of simulated models to those of the data.

The ability formally to model input substitution makes the model particularly suitable for the analysis of agricultural input policies where substitution is an important avenue of adjustment for farmers.

Regional modelers often face the added difficulty of a severely restricted data set which requires a compromise between the specification complexity of the model and the degree of disaggregation. The trade-off required to model the preferred specification with less than optimal data usually determines the economic modeling methodology used. Using this approach, we can calibrate nonlinear CES production functions in agricultural models using a minimum data set that usually restricts the modeler to a linear program.

In the following section the calibration approach to model specification is outlined. This calibration approach has some characteristics of both econometric and programming models in that it has a more flexible production specification than linear or quadratic programming (LP, QP) models, but the free parameters in the model are based on observed farmer behavior subject to resource and policy constraints.

7.1 Calibrating with Known Yield Functions

In some cases a nonlinear yield function is known and needs to be incorporated into the model calibration. For example in models where the level of feed determines livestock growth, or the intensity of supplementary water use determines crop yield, yield functions estimated from field or experimental data can be simply incorporated into the calibrated model, which then optimizes both the output and input levels based on their market and opportunity costs and the value of the output. Using a very simple case of supplementary irrigation for corn production in the midwest we suppose

that the yield of corn (crop i) per acre (yld_i)is a quadratic function of the quantity of supplementary irrigation water per acre (w_i).

$$yld_i = a_i + b_i w_i + 0.5 c_i w_i^2 \tag{7.1}$$

It is probably simplest to specify the yield equation above as a separate equation in Gams, and make the resulting yield an endogenous variable. The major change needed to calibrate a model with endogenous yields is that since the yield, price, and input quantity used for each crop with a yield function are defined, the base case value marginal product (VMP) of the input must be calculated before the calibration stage LP is defined. If the crop specific VMP is not calculated before the initial calibration stage, and used as the crop specific input "price" in both the LP and PMP stage, the model will not calibrate. If the base year VMP is used as the crop specific input price, the PMP model will calibrate in output and input use, and yield to the base quantities. In addition, when parameterized, the model will adjust the optimal production on all margins changing both the optimal input use and the resulting yield of the crops produced. Defining the nonlinear yield function generally as $yld_i = f(x_{ij})$ the endogenous yield problem is defined as:

$$\max \Pi = p_i yld_i x_{li} - vmp_{ij} x_{ij} \tag{7.2}$$

$$\text{subject to } Ax_l \leq b, \tag{7.3}$$

$$Ix_l \leq \tilde{x}_l \tag{7.4}$$

$$yld_i = f(x_{ij}), \quad and \quad x_{ij} \geq 0 \tag{7.5}$$

A Gams Program Template An example of a simple PMP yield calibration Gams program is shown in the Gams templates under Republic-1. The model is of the Republic river region in Nebraska, and has been simplified to only grow two crops, alfalfa and corn. To illustrate the procedure, I have only defined a single quadratic yield function for corn. Alfalfa is assumed to have the standard fixed yield and water requirements. Parameterizing the cost of water shows that both the yield and water use change in corn production.

7.2 Calibrating with Known Isoquants

In some cases the model builder may find that the best way to represent productive alternatives is not to have the yield vary with input combinations,

but to have a constant yield from different combinations of inputs. A CES isoquant for two inputs x_1 and x_2 and an elasticity of substitution of σ , where the parameter γ is defined as:$\sigma = 1/(1+\gamma)$ is defined in equation 7.6 as:

$$a\left(b_1 x_1^{\gamma} + (1-b_1)x_2^{\gamma}\right)^{\frac{1}{\gamma}} = 1 \tag{7.6}$$

The CES isoquant approach was first developed by Hatchet who designed and built the calibrated production model that is the central method in USBR.. Hatchet estimated CES functions in two inputs that show the ability to trade off water application per acre against the capital and management costs of field level application of the water. Essentially this enables the modeler to specify the objective as a function of a fixed yield, but not be confined to the fixed proportion assumption that underlies all the linear Leontieff input requirement matrices that we have used so far. The isoquant approach reconciles a fixed yield with an infinite combination of inputs that can result in the yield. It is also consistent with the underlying production technology. Of course, the isoquant can be a function of several variables, although the CES form used here has to assume that each input has the same elasticity of substitution. The isoquant approach also assumes that it is not profitable for the farmer to change input proportions so that crop yields increase or decrease. This is not the case in many situations, for example, the deliberate reduction of irrigation water below the rate that optimizes physical yields is termed "deficit irrigation" and is commonly practiced under drought conditions when the opportunity cost of water is much higher than average years. The endogenous adjustment of land allocations, inputs and yields requires a full production function and is addressed in the next section. The empirical specification of the isoquant model is:

$$\max \Pi = \sum_i [p_i yld_i xl_i - \sum_j \omega_j x_{ij}] \tag{7.7}$$

$$\text{subject to } \sum_i x_{ij} \leq b_j \quad \forall i \tag{7.8}$$

$$xl_i \leq \tilde{xl}_i \quad \forall i \tag{7.9}$$

$$\alpha_i(\sum_j \beta_{ij} x_{ij}^{\gamma_i})^{\frac{1}{\gamma_i}} = 1 \quad \forall i \tag{7.10}$$

$$\text{and} \quad x_{ij} \geq 0 \tag{7.11}$$

A Gams Program Template An empirical Gams template of the

isoquant approach applied to the simple Republic river model can be found in the Gams templates under Republic-2.gms.

7.3 Calibrating CES Production Functions

Analysis of a wider response to agricultural policy requires the introduction of more flexible production functions. The PMP and CGE calibration approaches can be combined to calibrate agricultural production models consistently and simply. In this example we will use the simplest crop- production data set possible, although this approach can be easily applied to mixed or pure livestock production.

The CES calibration procedure uses a three-stage approach. A constrained linear program is specified for the first stage. In the second stage, the regional production and cost parameters that calibrate the nonlinear CES model to the base-year data are derived from the numerical results of the linear program. The resource and policy constraints that reflect the empirical data are also included in the calibration process. The third-stage model is specified with a nonlinear objective function that incorporates the nonlinear production functions and PMP land costs. The CES model also has resource and policy constraints. However, the calibration constraints used in the first stage are absent.

7.3.1 The Analytical Derivation of the CES Parameters

A CES production function with one output, three inputs and constant returns to scale is defined as:

$$y = \alpha(\beta_1 x_1^{\gamma} + \beta_2 x_2^{\gamma} + \beta_3 x_3^{\gamma})^{\frac{1}{\gamma}} \tag{7.12}$$

where $\gamma = \frac{\sigma-1}{\sigma}$, $\sum \beta_i = 1$ and $\sigma =$ prior value elasticity of substitution. Taking the derivative of 7.12 with respect to x_1, we obtain:

$$\frac{\partial y}{\partial x_1} = \gamma\beta_1 x_1^{(\gamma-1)} \frac{1}{\gamma} \alpha(\beta_1 x_1^{\gamma} + \beta_2 x_2^{\gamma} + \beta_3 x_3^{\gamma})^{\frac{1}{\gamma}-1} \tag{7.13}$$

since $\gamma - 1 = \frac{1}{-\sigma}, \frac{1}{\gamma} - 1 = \frac{1}{\sigma-1}$.
Simplifying and substituting, 7.13 can be rewritten as:

$$\frac{\partial y}{\partial x_1} = \beta_1 x_1^{\frac{-1}{\sigma}} \alpha(\beta_1 x_1^{\gamma} + \beta_2 x_2^{\gamma} + \beta_3 x_3^{\gamma})^{\frac{1}{\sigma-1}} \tag{7.14}$$

Equating $\rho\frac{\partial y}{\partial x_1} = \omega_1$ and $\rho\frac{\partial y}{\partial x_2} = \omega_2$ results in:

$$\frac{\omega_1}{\omega_2} = \frac{\beta_1 x_1^{\frac{-1}{\sigma}}}{\beta_2 x_2^{\frac{-1}{\sigma}}} \tag{7.15a}$$

$$\frac{\omega_1}{\omega_3} = \frac{\beta_1 x_1^{\frac{-1}{\sigma}}}{\beta_3 x_3^{\frac{-1}{\sigma}}} \tag{7.15b}$$

From 7.15a we obtain:

$$\beta_2 = \beta_1 \frac{\omega_2}{\omega_1} \left(\frac{x_1}{x_2}\right)^{\frac{-1}{\sigma}} \tag{7.16}$$

Likewise from Equation 7.15b:

$$\beta_3 = \beta_1 \frac{\omega_3}{\omega_1} \left(\frac{x_1}{x_3}\right)^{\frac{-1}{\sigma}} \tag{7.17}$$

But from the constant returns to scale assumption:

$$\beta_3 = 1 - \beta_1 - \beta_2 \tag{7.18}$$

Substituting 7.16 and 7.17 into 7.18, we obtain:

$$\beta_1 \frac{\omega_3}{\omega_1} \left(\frac{x_1}{x_3}\right)^{\frac{-1}{\sigma}} = 1 - \beta_1 - \beta_1 \frac{\omega_2}{\omega_1} \left(\frac{x_1}{x_2}\right)^{\frac{-1}{\sigma}} \tag{7.19}$$

Dividing through by β_1 and rearranging yields:

$$\frac{1}{\beta_1} = 1 + \frac{\omega_2}{\omega_1} \left(\frac{x_1}{x_2}\right)^{\frac{-1}{\sigma}} + \frac{\omega_3}{\omega_1} \left(\frac{x_1}{x_3}\right)^{\frac{-1}{\sigma}} \tag{7.20}$$

Solving 7.20 for β_1 and substituting into Equation 7.16 solves for β_2. Substituting the values into equation 7.17 solves for β_3. The numerical value for the total production, y, in Equation 7.12 is known from the observed acreage $\bar{x}_1$ and the average yield $\bar{y}$. Using the known values for $\beta_1 \ldots \beta_3$ and Equation 7.12, we can solve for α as follows:

$$\alpha = \frac{\bar{y}\bar{x}_1}{(\beta_1 x_1^{\gamma} + \beta_2 x_2^{\gamma} + \beta_3 x_3^{\gamma})^{\frac{1}{\gamma}}} \tag{7.21}$$

The minimal data set needed to specify an LP model are the input allocations and prices, the expected yield, price and any resource or policy constraints. If the elasticity of substitution value and the constant returns to scale assumption are added to this basic data set, the scale and share parameters of the CES production function can be recursively solved for any number of inputs using Equations 7.16, 7.17, 7.18, and 7.21.

7.3.2 The Empirical Application of CES Calibration

The data set, which can be termed the minimum LP data set, is a single cross- section observation of regional production over i crops. Observations include product prices p_i, acreage allocation $\bar{x}_{i1}$ crop input use x_{ij} cost per unit input w_j, and average yields $\bar{y}_i$ Allocable resource limits or policy constraints are defined as b_j, the right-hand side values of inequality constrains on the production activities. Regional subscripts have been omitted for simplicity. The first stage LP model is defined in Equations 7.22a to 7.22c. The generation of the dual values for the two types of constraint in the model is an essential step in the derivation of adjusted factor costs that will allow the more complex CES specification to be calibrated from the simple data base.

$$\begin{aligned} \max \quad & \textstyle\sum_i p_i \bar{y}_i x_i - \sum_j \omega_j a_{ij} x_i && \text{(7.22a)} \\ \text{s.t.} \quad & Ax \leq b && \text{(7.22b)} \\ & Ix \leq \bar{x} + \epsilon && \text{(7.22c)} \end{aligned}$$

The two constraint sets 7.22b and 7.22c yield two sets of dual values. The vector λ_1 consists of the resource shadow value Duals associated with constraint set 7.22b. The vector of elements λ_2 are the PMP Duals from the calibration constraint set (7.22c).

These two sets of dual values are used to calculate the equilibrium opportunity cost of land and other fixed but allocable inputs. These values are then used in the derivation of the production function coefficients.

CGE models are by definition and convention based on Walras' law for factor allocation, which defines the set of prices that equate excess supply and demand (Dervis et al., 1982). For partial-equilibrium models, the fixed resource endowment and local adjustment costs result in resource factors having scarcity costs that may not be fully reflected in the nominal resource or rental prices. While CGE calibration methods can use market prices and quantities to define the share equations and production function parameters, partial-equilibrium agricultural models have to augment the nominal prices by the resource and crop-specific shadow values generated in the first LP stage of the calibration.

Equation 7.23 shows a three-input CES production function for a single crop, i:

$$y_i = \alpha_i(\beta_{i1} x_{i1}^{\gamma_i} + \beta_{i2} x_{i2}^{\gamma_i} + \beta_{i3} x_{i3}^{\gamma_i})^{\frac{1}{\gamma_i}} \qquad (7.23)$$

where $\gamma_i = \frac{\sigma_i - 1}{\sigma_i}$, $\beta_{i3} = 1 - \beta_{i1} - \beta_{i2}$ and $\sigma_i =$ is a prior estimate of the elasticity of substitution.

The production function is specified as having constant returns to scale for a given quality of land, since use of the two sets of dual values and the nominal factor prices exactly allocates the total value of production among the different inputs. If the modeler needs to specify groups of inputs with differing elasticities of substitution, perhaps zero for some inputs, the nested approach suggested by Sato (1967)is shown in section 1.4. The Cobb-Douglas production function or restricted quadratic specifications can be used instead of the CES.

The definition of model calibration in the introduction, and over a decade of empirical practice with calibrating CGE models, has established the precedent of using robustly estimated parameters from other studies for calibration. Elasticity parameters are often used as they represent underlying preferences or technologies and, as such, are less likely to vary over specific model applications.

Given the data, Equation 7.23 with j inputs has j unknown parameters to calibrate. Namely, $(j-1)$ share parameters β_{ij} and one scale parameter, α_i. As shown in section 7.3.1 the $(j-1)$ unknown share parameters are expressed in terms of the factor cost and input shares. The first-order conditions for input allocation equate the value marginal product to the nominal input cost plus any shadow costs for constrained resources. Algebraic manipulation of the first-order conditions yields the recursive Equations 7.24a – 7.24c that are solved for the crop and regional-specific share coefficients as shown above.

$$\frac{1}{\beta_1} = \frac{\bar{\omega}_2}{\bar{\omega}_1}\left(\frac{x_1}{x_2}\right)^{\frac{-1}{\sigma}} + \frac{\bar{\omega}_3}{\bar{\omega}_1}\left(\frac{x_1}{x_3}\right)^{\frac{-1}{\sigma}} \tag{7.24a}$$

$$\beta_2 = \beta_1\frac{\bar{\omega}_2}{\bar{\omega}_1}\left(\frac{x_1}{x_2}\right)^{\frac{-1}{\sigma}} \tag{7.24b}$$

$$\beta_3 = 1 - \beta_1 - \beta_2 \tag{7.24c}$$

where $\bar{\omega}_j$ equals factor plus opportunity cost and σ equals elasticity of substitution.

Share equations for variable factor inputs whose supply functions are assumed elastic are calibrated similarly to those in CGE model production functions. An important difference between partial equilibrium and CGE models is in the specification of the resource share equations. In regional partial-equilibrium models the physical limits on the availability of these resources has to be reflected in the allocations. In most partial-equilibrium models these fixed resources will have a market price, but it is likely that the

physical limits will also result in a dual value for the resource. Accordingly, the share equations for allocable resource inputs other than land have the resource shadow cost, measured by the dual for constrained resource group λ_1 , added to the market price of the input to yield $\bar{\omega}_j$. Owing to changes in quality, the cost of land inputs is derived by adding the market price, shadow value ($\lambda 1_l$) and the marginal crop-specific PMP cost, ($\lambda 2_i$) to yield the land factor cost $\bar{\omega}_{li}$. This crop-specific cost of land reflects both the scarcity value of land and the quality differences in land allocated to different crops.

The adjusted factor costs $\bar{\omega}_j$ exactly exhaust the total revenues for each cropping activity and are used in Equations 7.24a – 7.24c to calibrate the share coefficients.

The crop and regional scale coefficient α in Equation 7.23 is calibrated by substituting the values of β, σ, y, and x back into Equation 7.23, as shown in Equation 7.21.

Since the marginal implicit cost of changing crop acreage is included in the share equations via the parameter $\bar{\omega}_{li}$, the cost function must also be explicitly represented in the objective function. Using the standard PMP approach, we specify the implicit cost function for each crop in Equation 7.26a as quadratic in the acreage allocated to the crop. Note that the previously used notation for the PMP coefficients γ has been changed to ψ to avoid confusion in the specification of the CES function.

$$\text{Implicit cost} = \psi_i x_{il}^2 \tag{7.25a}$$

$$\lambda_{i2} = 2\psi_i x_{il} \tag{7.25b}$$

$$\text{therefore } \psi_i = \frac{\lambda_{i2}}{2x_{il}} \tag{7.25c}$$

Defining the quadratic cost function in Equation 7.26a as the implicit cost of increasing regional crop acreage, the marginal implicit cost is calibrated using the crop-specific PMP dual value. Equation 7.25b shows how λ_{2i} from Equation 7.25a is used to calibrate the implicit cost function coefficient ψ_i in Equation 7.25c.

Using the coefficients calibrated above, a general CES representation of the agricultural resource production problem is:

$$\max \sum_i p_i y_i - \sum_j \omega_{ij} x_{ij} - \sum_i \psi_i x_{il}^2 \tag{7.26a}$$

$$\text{subject to } y_i = \alpha_i (\sum_j \beta_{ij} x_{ij}^{\gamma})^{\frac{1}{\gamma}} \tag{7.26b}$$

$$Ax \leq b \tag{7.26c}$$

The model in Equation 7.26a differs from that in the first stage, Equation 7.22a, in three significant ways. First, the production technology is more general and has the empirical elasticity of substitution incorporated in it. This means that the model in 7.26a solves for the optimal input proportions in conjunction with the land allocation, but not in fixed proportions to it as in the Leontief specification in model 7.22a.

Second, the objective function has the additional implicit cost function specified for each land allocation. The basis of this cost is in the heterogeneity of land, other inputs, and the fixed nature of some farm inputs such as family labor and major machinery units.

Third, the set of calibration constraints (7.22c) are omitted from the CES model in 7.26a. The CES model still calibrates with the base-year inputs and outputs since the dual values from model 7.22a are incorporated in the first-order condition used to calibrate the production and cost coefficients. Thus the CES model calibrates exactly to the base-year data without any arbitrary or empirically insupportable constraints.

To summarize, this section has shown how a minimal data set for a constrained LP model can be used to generate a more general self-calibrating CES model.

A Gams Program Template An empirical Gams template of the CES calibration approach applied to the Yolo county model can be found in the Gams templates under Yolo-CES.gms.

7.4 Using a Nested CES Production Function

The CES calibration used in the previous section has the implicit restriction that all inputs have to have the same elasticity of substitution. This may be unacceptable when the number of inputs is large and varied. For example, labor and machinery may be highly substitutable for some operations, for example the size of a tractor versus the number of operators. However land and irrigation water may have a much lower elasticity of substitution. This empirical situation has been accommodated for many years by dividing the inputs into subsets that have similar elasticities of substitution. This division requires that first the subsets are calibrated, then another CES function is calibrated to reflect the elasticity of substitution between the two subsets. This function is called a nested CES production function. Many studies use computable general equilibrium modeling to simulate the effects of change in input prices on input use and output. However, CGE models, due to their complexity, are usually specified with far fewer inputs

and agricultural production activities than are usually required for regional agricultural and environmental policy. The nested CES approach enables a partial equilibrium model to be calibrated in sufficient detail to accurately reflect the physical differences between regions, but at the same time be entirely consistent with the more general CGE approach. In fact, there is a natural symbiosis between detailed calibrated production models and more general CGE models.

7.4.1 The Nested CES Production Function

The regional total output from each cropping activity q_{iwr} (indices are dropped for brevity) is defined by a nested CES production function with seven categories of inputs in this example:

$$Q = C\left\{\beta_F\left[C_F\left(\sum_{j=1}^{2} b_j x_j^{\gamma_F}\right)^{\frac{1}{\gamma_F}}\right]^{\gamma} + \beta_V\left[C_V\left(\sum_{j=3}^{7} b_j x_j^{\gamma_V}\right)^{\frac{1}{\gamma_V}}\right]^{\gamma}\right\}^{\frac{1}{\gamma}} \tag{7.27}$$

The function consists of two nests. The first nest, expressed in the first set of brackets, includes the first two categories of inputs, the allocatable inputs of land and water. The second nest, expressed in the second set of brackets, is for the remaining five variable inputs. Each nest is in itself a CES function. A nested CES function is more flexible than a regular CES function in that more than one elasticity of substitution coefficient between inputs can be modeled. In agricultural crop production, the ability to substitute inputs varies significantly (Debertin et al., 1990; Hertel et al., 1989; Rendleman, 1993; Ray, 1982). For example, the ability to substitute cultivation for herbicide weed control is greater than the substitution potential between water and nitrogen. In the nested CES formulation, the nests can be thought of as hierarchies. Equation 7.27 has the higher nest parameters on the outside. The scalar C is the top-nest scale parameter, and β_F and β_V are the top-nest share parameters for allocatable and variable inputs, respectively. Moving to the lower nests, C_F and C_V are scale parameters for allocatable and variable input nests, respectively. The quantity of input j allocated to each cropping activity is indicated by x_j where the j values of 1 and 2 correspond to allocatable inputs of land and water, and the remaining values of j (from 3 to 7) correspond to variable inputs. The parameter β_j is the share parameter of the j^{th} input. In addition to index j, input quantity x is indexed over i, w, and r, which are dropped here for brevity, The coefficient $\gamma = (s-1)/s$, where s is the top-nest elasticity of substitution coefficient. Finally, $\gamma_F = (s_F - 1)/s_F$ and $\gamma_V = (s_V - 1)/s_V$, where s_F and s_V are the

elasticity of substitution between the allocatable inputs and the elasticity of substitution between the variable inputs, respectively. For the dryland cropping activities, the allocatable input nest has only the land input as its argument.

7.4.2 An Application- Measuring Energy Cost Effects

Farmers will react to higher energy costs in several ways. First, they are expected to substitute low-energy inputs (such as land, capital, and labor) for energy-intensive inputs (such as fertilizer, chemicals, and on-farm direct energy use). Second, they will reduce the acreage of energy-intensive crops in favor of less energy-intensive crops. Third, farmers will likely shift acreage from irrigated to dryland cultivation. Fourth, depending on the magnitude of the energy price increases and the profitability of the farm enterprises, in some regions the acreage of low-energy-using crops may also be reduced. This can occur when output price effects dominate input price effects. Furthermore, farmers' responsiveness to energy price changes will differ depending on the regional differences in profitability conditions and energy use.

All of these adjustments will take place in a highly interactive environment in which the changes in input use affect crop yields and, combined with the cropping pattern and acreage changes, also affect the total output of each crop produced and ultimately the market price of the crops. The changes in crop prices in turn influence input use, cropping choice, and planting decisions. To account for these interactions, the economic model employed in the analysis needs to be national in scope, embody regional crop production characteristics, and have the ability to explicitly account for farmers' input substitution behavior.

The nested approach used to analyze farmer reactions to energy prices (Konyar and Howitt (2000)) is very similar to models used by Howitt (1995a; 1995b), and by Edwards et al. (1996). Each cropping activity is defined by a nested constant elasticity of substitution (NCES) production function with seven inputs as the arguments, allowing the model to endogenously determine the quantities of inputs used in each activity. The smallest decision-making unit is a region. There are 12 regions spanning the 48 contiguous states in the USA. Regions are modeled as aggregate farm units producing all the major crops in their respective areas under dry and irrigated conditions. The model incorporates aggregate (domestic and export) demand equations for each crop which endogenously determine crop prices. The model simulates a conditional near-term sectoral equilibrium in a compara-

tive static setting. Farmer behavior is predicted for given policy shocks in terms of acres allocated to specific crops in each region (with or without irrigation), the amount of each input used, and the impact on crop prices. More specifically, the model is defined by the following objective function and constraints.

The Objective Function The objective function represents the aggregate consumer (domestic and foreign) and producer welfare for all regions and activities:

$$\begin{aligned}\Pi = & \sum_i \left[\alpha_i \sum_w \sum_r q_{iwr} + 0.5\delta_i \left(\sum_w \sum_r q_{iwr} \right)^2 \right] - \sum_i \sum_w \sum_r \tau_{iwr} q_{iwr} \\ & - \sum_r \left[v_{r1} \sum_i \sum_w x_{iwr1} + 0.5\omega_{r1} \left(\sum_i \sum_w x_{iwr1} \right)^2 \right] \\ & - \sum_i \sum_w \sum_r \eta_{iwr1} x_{iwr1} - \sum_i \sum_w \sum_r \sum_j \left(\rho_{iwrj} x_{iwrj} + \phi_{iwrj} x_{iwrj}^2 \right)\end{aligned} \tag{7.28}$$

This formulation ensures a competitive market equilibrium solution. The first expression in brackets measures the area under the crop-specific linear market quantity-dependent demand equations, where α_i is the intercept and δ_i is the slope of the quantity-dependent demand equation for crop i. The variable q_{iwr} described in Equation 7.28, represents the output of crop i, produced under cultivation condition w (1 = dry and 2 = irrigated), in region r. The coefficient τ_{iwr} in the second expression accounts for the marketing and transportation costs of output produced. The third expression allows the land rents to be endogenous at the regional level, where v_{r1} and ω_{r1} are the intercept and slope of regional linear land supply equations, respectively. The coefficient η_{iwr1} accounts for the difference between the regional average land rent and the crop activity-specific land rents in that region. The last term in the objective function is the PMP cost function, where ρ and ϕ are the coefficients, and the variable x_{iwrj} is the amount of input j ($j = 1, \ldots, 7$, with 1 = land input) used in the cropping activity i, w, r. This function is quadratic in the land input and linear in the others. For the non-land inputs, ϕ_{iwrj} is zero and ρ_{iwrj} is the linear per acre cost c_{iwrj} for each input.

The Model Constraints In the following discussion, indices on x are reintroduced: i (crop), w (irrigation condition), r (region), and j (input).

The regional irrigation water constraint limits the total irrigation water used by all irrigated crops ($w = 2$) in a region to the actual total irrigation water ($j = 2$) used in the region in the base year, $\bar{X}_{2r2}$. This constraint is specified as:

$$\sum_i X_{i2r2} \leq \sum_i \bar{X}_{i2r2} \tag{7.29}$$

Likewise, the regional irrigated land constraint restricts the total land allocated to irrigated cultivation to the total actual base year irrigated acreage:

$$\sum_i X_{i2r1} \leq \sum_i \bar{X}_{i2r1} \tag{7.30}$$

Constraint 7.30 is redundant in the base run. Since it is not needed in the calibration step, the right-hand-side value of this constraint is kept slightly above its actual base year value so as not to bind. It is included under policy shocks so that the irrigated land in regions will not exceed the base year levels.

In some arid regions, irrigated crop production occurs on land that has insufficient rainfall for dryland production. In these regions, dry cultivation is not an economically viable option for most of the crops grown under irrigated conditions. In the model, we include constraints to limit the availability of total dryland in these regions to be no greater than the actual base year dryland acreage.

7.5 Microeconomic Properties of Calibrated Production Models

In generalizing the production specification to the CES class of functions, calibrated production models show properties consistent with microeconomic theory that are not exhibited in LP or input/output models. The ability for unconstrained calibration has been addressed in the previous section.

With the specification of a nonlinear profit function in land in PMP models, the standard Hicksian microeconomic properties can be derived. By specifying the primal-dual model formulation, and making the usual assumption that the matrix of implicit cost coefficients Ψ is positive definite, it can be shown (Paris, 1993, Chp 11) that the slopes of the supply and demand functions derived from the calibrated model are respectively positive and negative, as in Equations 7.31a and 7.31b. The Hicks symmetry

conditions shown in Equation 7.31c also hold for the CES model.

$$\frac{\partial y}{\partial p} = \text{PSD} \tag{7.31a}$$

$$\frac{\partial x}{\partial \omega} = \text{NSD} \tag{7.31b}$$

$$\frac{\partial y}{\partial \omega} = -\frac{\partial x}{\partial p} \tag{7.31c}$$

The response of the model output to changes in price, or input use to changes in cost. The importance of this property is that politically acceptable agricultural and resource policies are usually constrained to relatively small changes in costs or policy constraints. The continuous functions in calibrated production models can reflect these small policy changes and simulate their economic and physical impact on a regional scale.

The approach presented in this chapter calibrates more flexible production functions than linear programs, but uses almost the same minimal data base. Calibrated production models can be viewed as a compromise between the ridgidity of linear programming and data requirements of econometric estimation. The properties of calibrated models are shown to meet many of the requirements for modeling regional agricultural policies, while the data requirements are satisfied by the minimal data sets usually available on a regional basis.

While potential difficulties in the nonlinear solution of a many-dimensional nonlinear calibrated production specification cannot be blithely ignored, initial empirical results indicate that these models are quite tractable. Given the common agricultural policy requirement for modeling regional economic and environmental consequences, the properties of the models seem to justify the additional complexity.

Chapter 8

Nonlinear Optimization Methods

8.1 Mathematics For Nonlinear Optimization

8.1.1 Concave Functions and Convex Sets

Concave Function

A function, $f(x)$, defined on a convex set Ω is strictly *concave* if for every $x_1, x_2 \in \Omega$ and every λ, $(0 \leq \lambda \leq 1)$:

$$f(\lambda x_1 + (1-\lambda)x_2) > \lambda f(x_1) + (1-\lambda) f(x_2) \tag{8.1}$$

The *tangent line* of a scalar-valued function, $f(x)$, evaluated at x_0, has a value at x_1 given by (also see Figure 8.1):

$$f(x_0) + f'(x_0)(x_1 - x_0)$$

It follows that the *tangent plane* for a vector-valued function, $f(x)$, at a point in n-space (say x_0) can be expressed as $f(x_0) + \nabla f(x_0)(x_1 - x_0)$ for $x_1 \in \mathbb{R}^n$ (see Figure 8.2).

If $f(x)$ is a concave function on the convex set Ω, then the set $\Omega = \{x : x \in \Omega, f(x) \geq c\}$ is convex for every real number c. This is a method for defining a nonlinear constraint set.

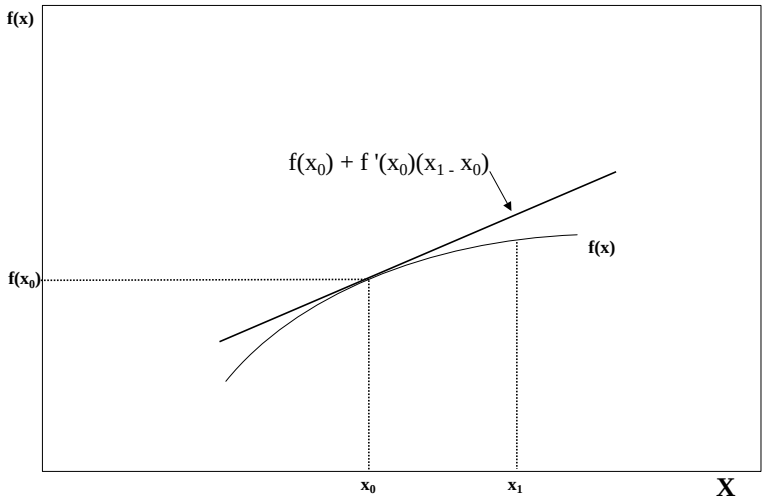


Figure 8.1: Tangent on a Concave Function

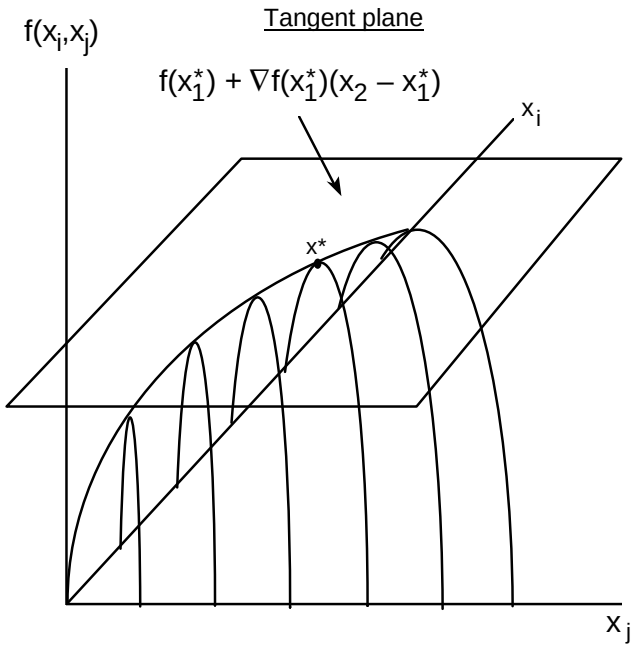


Figure 8.2: Tangent plane to a Concave Function ($\in \mathbb{R}^3$)

8.1.2 Taylor Series Expansion

A second-order *Taylor series expansion* of the scalar function $f(x)$ around the point x_0 is:

$$f(x) = f(x_0) + \frac{f'(x_0)}{1!}(x_1 - x_0) + \frac{f"(x_0)}{2!}(x_1 - x_0)^2 + r$$

Where r is a remainder term. This expansion is often approximated as:

$$f(x) \cong f(x_1) + f'(x_1)(x_2 - x_1) + \frac{1}{2}f"(x_1)(x_2 - x_1)^2$$

Where $f(x_0)$ is the initial value, $f'(x_0)(x_1 - x_0)$ is the tangent line times the difference in x values, and $\frac{1}{2}f"(x_1)(x_2 - x_1)^2$ is the 2^{nd} order term.

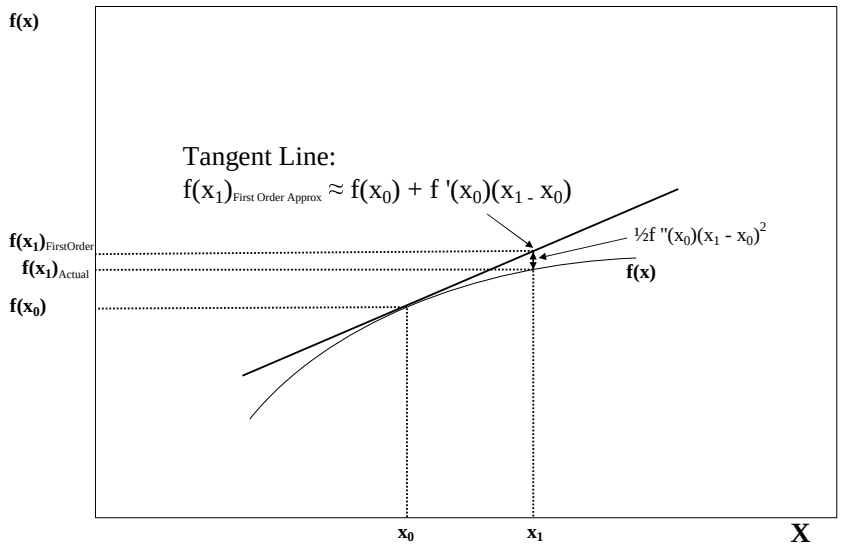


Figure 8.3: Taylor series expansion of $f(x)$ at x_0

8.1.3 Matrix Derivatives

Linear form

If $L(x) = c'x$ then $\frac{\partial L(x)}{\partial x} = c'$.

Quadratic form

If $Q(x) = x'Ax$ then $\frac{\partial Q(x)}{\partial x} = 2x'A$.

In both cases the derivative of a scalar value (resulting from the inner product of a row and column vector) with respect to a column vector is a row vector.

8.1.4 The Gradient Vector ($\nabla f(x)$)

A second convention is that the derivative of a scalar with respect to a column (row) vector is a row (column) vector. Thus, if the scalar y is a differentiable function of the column vector x, the vector of partial derivatives $\frac{\partial y}{\partial x_i}$ is a row vector called the *gradient vector*. For example, if an objective function has a scalar value which is a nonlinear function of n variables, i.e.:

$$y = f(x) = f(x_1, x_2, \ldots, x_n)$$

then the vector of first-order partial derivatives, the gradient vector, is the row vector:

$$\nabla f(x) \equiv \frac{\partial f(x)}{\partial x} = \left[\frac{\partial f(\cdot)}{\partial x_1}, \frac{\partial f(\cdot)}{\partial x_2}, \cdots, \frac{\partial f(\cdot)}{\partial x_n}\right]$$

8.1.5 Inner Products

Because we use inner products a lot we will adopt a new, clearer, notation. The Inner product of two vectors ($a'b$ = scalar) may now be denoted in an alternative form as $\langle a,\, b\rangle$ = scalar. Thus the familiar objective function $c'x = z$ can be written $\langle c,\, x\rangle = z$.

8.1.6 Hessian and Jacobian Matrices

Hessian matrix

The derivative of the gradient vector with respect to the $n \times 1$ column vector x is the $n \times n$ *Hessian matrix:*

$$\frac{\partial^2 f(x)}{\partial x^2} = \frac{\partial}{\partial x}\left(\frac{\partial f(x)}{\partial x}\right) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

For example, the Hessian matrix of the quadratic form $x'Ax$ is $2A$.

Jacobian matrix

Similarly the derivative of the column vector of m functions, $g_i = g_i(x)$, with respect to x (where each function depends on the $n \times 1$ column vector x) is the $m \times n$ *Jacobian matrix*:

$$\frac{\partial g(x)}{\partial x} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n} \end{bmatrix}$$

8.1.7 Taylor series expansion of a Vector Function

We can now use gradients, Hessians, and the scalar Taylor series to approximate functions of vectors.

Suppose $f(x)$ is a function of the $n \times 1$ vector x.

Given some vector of initial values x_0, we can expand around this vector to approximate the functional value of some other vector x_1.

$$f(x_1) = f(x_0) + \langle \nabla f(x_0), (x_1 - x_0) \rangle + \frac{1}{2}(x_1 - x_0)' H_{x_0}(x_1 - x_0) + r_0$$

Where:

- $\nabla f(x_0)$ is the gradient of $f(x)$ at x_0.
- H_{x_0} is the Hessian of $f(x)$ at x_0.
- $(x_1 - x_0)$ is an $n \times 1$ vector of the differences between x_1 and x_0.
- r_0 is the remainder term of the expansion that could be reduced by going to higher terms.

8.1.8 Definite Quadratic Forms

Some quadratic forms have the property $x'Ax > 0 \quad \forall \quad x$ except $x = 0$; some are negative for all x except $x = 0$.

Positive Definite Quadratic Form

The quadratic form $x'Ax$ is said to be *positive definite* if it is positive (> 0) for every x except $x = 0$.

Positive Semidefinite Quadratic Form

The quadratic form $x'Ax$ is said to be *positive semidefinite* if it is non-negative (≥ 0) for every vector x where $x \neq 0$, and there exist points for which $x'Ax = 0$.

Negative definite and negative semidefinite forms are defined by interchanging "negative" and "positive" in the above definitions. If $x'Ax$ is positive definite (semidefinite), then $x'(-A)x$ is negative definite (semidefinite).

8.2 An Introduction to Nonlinear Optimization

8.2.1 Some Non Linear Programming (NLP) Definitions

The Standard NLP Problem

$$\begin{aligned} \min \quad & f(x) \text{ where } x = n \times 1 \text{ vector} \\ \text{subject to} \quad & x \in \Omega \end{aligned}$$

and where Ω denotes the feasible solution set. $\Omega \in \mathbb{R}^n$ or a subset of $\mathbb{R}^n$. Note that the objective function and the solution set are no longer defined by linear functions.

Local Minima

A point $x^* \in \Omega$ is a local minimum of $f(x)$ on Ω if there is a small distance, ϵ, such that $f(x) > f(x^*) \quad \forall \quad x \in \Omega$ within the distance ϵ of x^*.

Verbally: "The objective function increases in all directions, therefore we are at the optimum point for a minimization problem."

Global Minima

A point $x^* \in \Omega$ is a global minimum of $f(x)$ on Ω if $f(x) > f(x^*) \quad \forall \quad x \in \Omega$.

The aim is to set criteria for a computer program to perform a systematic search over a mathematical surface. As in finding your way to a location, good directions will give you a sequence to follow. In each sequence or step you need to know the direction d to proceed, and how far α to go in that direction.

Feasible Directions

Along any given single direction d, the objective function $f(x)$ is a function of the distance moved in a direction. Note any direction in $\mathbb{R}^n$ is an n-dimensional vector.

For $x \in \Omega$, d is a *feasible direction* at x_0 if there exists a scalar $\alpha > 0$, such that:

$$(x_0 + \alpha d) \in \Omega \quad \forall \quad \alpha(0 \leq \alpha \leq \bar{\alpha})$$

i.e. A particle can move in direction d for a distance $\alpha \leq \bar{\alpha}$ without leaving the feasible set Ω.

Notes:

- A direction in "n-space" ($\mathbb{R}^n$) is an n-dimensional vector d.
- Along any given single direction d, the objective function $f(x)$ is a function of the distance moved in that direction.

Example in two-space: In R^2 where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, then d (the directional vector) can be $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 4 \\ 2 \end{bmatrix}$, or $\begin{bmatrix} 34 \\ 17 \end{bmatrix}$.

So, the vector d gives direction, not distance. A direction depends on the ratio of values in the vector not the values themselves.

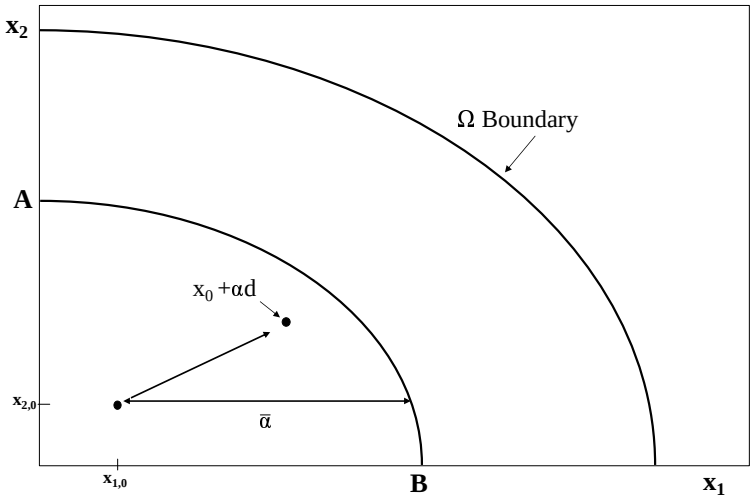


Figure 8.4: Feasible Directions

The feasible directions depend on

- $\bar{\alpha}$ = the step size limit. Note that the feasible directions d at a point x_0 are a function of the step size limit and vice versa.
- Where starting point x_0 is situated in Ω.
- In Figure 8.4, the arc from point A to point B describes the set of feasible directions at x_0 for step size $\bar{\alpha}$.

8.2.2 Nonlinear First Order Conditions

Local Minimum Point — Constrained Problem

Given the objective function f(x) and the constraint set Ω. If x^* is a local minimum point of f(x) over Ω, then for any feasible direction from x^*, the necessary condition is (for all feasible directions d $\in \mathbb{R}^n \cap \Omega$):

$$\langle \nabla f(x^*),\, d \rangle \geq 0 \tag{8.2}$$

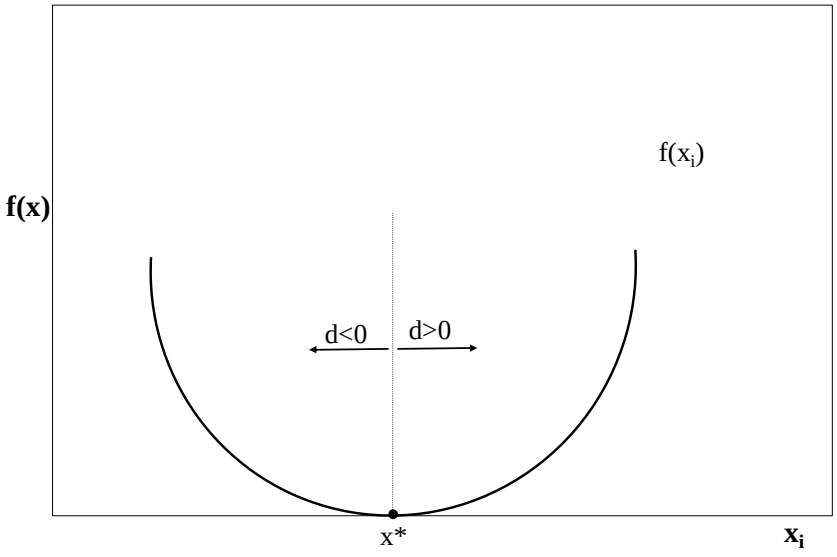


Figure 8.5: Minimizing a Scalar Function

Equation 8.2 can be explained from figure 8.5 as follows:
Moving left from x^*, $d < 0$ and $\nabla f(x) = \frac{\partial f(x)}{\partial x} < 0$ so $\nabla f(x) \cdot d > 0$.
Moving right from x^*, $d > 0$ and $\nabla f(x) = \frac{\partial f(x)}{\partial x} > 0$ so $\nabla f(x) \cdot d > 0$.

This works since d is defined as the direction you're moving from the point x^*, that is, each element is a distance $(x_i - x_i^*)$ as in the Taylor series expansion.

8.2.3 Proof of First Order Condition

1. Let x^* equal a local minimum.

2. Pick another point $x(\alpha)$ at an arbitrary distance α in direction d ($\alpha > 0$). $x(\alpha) = x^* + \alpha d$. Therefore the new objective value is $f(x^* + \alpha d)$.

3. Apply Taylor series expansion, that is truncated its first order, to $f(x(\alpha))$ around $f(x^*)$.

$$\begin{aligned} f(x(\alpha)) &= f(x^* + \alpha d) \\ &\approx f(x^*) + \langle \nabla f(x^*),\, (x(\alpha) - x^*) \rangle \\ &\approx f(x^*) + \langle \nabla f(x^*),\, \alpha d \rangle \end{aligned} \tag{8.3}$$

   Note that the second-order terms and the remainder term in the Taylor series expansion have been truncated, making this an approximation.

4. If x^* is a minimum point, by definition:

$$f(x^*) - f(x(\alpha)) \leq 0 \tag{8.4}$$

5. Substituting the expansion for $f(x(\alpha))$ defined in 8.3 into 8.4, we obtain:

$$f(x^*) - f(x^*) - \langle \nabla f(x^*),\, \alpha d \rangle \leq 0 \tag{8.5}$$

$$\therefore -\langle \nabla f(x^*), \alpha d \rangle \leq 0 \tag{8.6}$$

6. Factoring out the scalar α and multiplying by - 1 we get:

$$\alpha \langle \nabla f(x^*),\, d \rangle \geq 0 \tag{8.7}$$

Since $\alpha > 0$, x^* being a minimum implies:

$$\langle \nabla f(x^*),\, d \rangle \geq 0 \tag{8.8}$$

The Unconstrained Problem

If the problem is unconstrained, this implies that x^* is an interior point and therefore (for some small enough $\alpha > 0$) all directions d are feasible.

For unconstrained problems, the feasible-direction vector d can have any sign or direction, thus, the first order condition $\langle \nabla f(x^*),\, d \rangle \geq 0$ for all d contained in $\mathbb{R}^n$ implies that $\nabla f(x^*) = 0$ because:

$$\langle \nabla f(x^*),\, d \rangle = [f_1',\, f_2',\, \ldots f_n'] \begin{bmatrix} \pm \text{ anything} \\ \pm \text{ anything} \\ \vdots \\ \pm \text{ anything} \end{bmatrix} \geq 0 \text{ for all possible directions } d$$

This means $[f_1',\, f_2',\, \ldots f_n']$ must equal $[0,\, 0, \ldots,\, 0]$.

Unconstrained Optimization Example

$$\min f(x_1, x_2) = x_1^2 - x_1 x_2 + x_2^2 - 3x_2 \tag{8.9}$$

There are no constraints; therefore the feasible region is the whole of two-space on the real line, or (in symbolic terms) $\Omega = R^2$.

$$\text{FOC } \nabla f(\cdot)' = 0 \Rightarrow \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \end{bmatrix} \overset{set}{=} \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Solving the two gradient equations:

1. $2x_1 - x_2 = 0 \Rightarrow 2x_1 = x_2$
2. $-x_1 + 2x_2 = 3 \Rightarrow -x_1 + 2(2x_1) = 3 \Rightarrow -x_1 + 4x_1 = 3$

Therefore $x_1 = 1$ and $x_2 = 2$. Try substituting this into 8.9 and check other values.

Constrained Optimization Example

$$\min f(x_1,\, x_2) = x - x_1 + x_2 + x_1 x_2$$
$$\text{subject to } x_1 \geq 0 \text{ and } x_2 \geq 0$$

Problem B has a global minimum at

$$\begin{matrix} x_1 = \left[\frac{1}{2}\right] \\ x_2 = \left[0\right] \end{matrix}$$

Do the constrained first order conditions hold at this point? Check FOC

$$\nabla f(x^*) = [2x_1 - 1 + x_2,\ x_1 + 1]$$

Substituting the numerical $x_1 = \frac{1}{2}$ and $x_2 = 0$ for x_1 and x_2 into the FOC equation yields the gradient values at $[x_1 = \frac{1}{2},\ x_2 = 0]$ of $[0,\ \frac{3}{2}]$.

Now we pick a small step size, say $\alpha = 0.1$. Therefore at x^*, the feasible directions given the constraints and the initial values are:

$$d = \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} \text{anywhere} \\ \text{positive values only} \end{bmatrix}$$

The condition $\langle \nabla f(x^*),\ d \rangle > 0$ implies:

$$\left[0,\ \frac{3}{2}\right] \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} > 0 \text{ or } \frac{3}{2}\Delta x_2 > 0$$

While Δx_1 can be positive or negative, Δx_2 can only have positive values, therefore the constrained First Order Conditions in 8.7 hold.

8.3 Steepest Descent Algorithms

8.3.1 Steepest Descent Direction

The gradient vector of a function indicates the direction of movement which results in the greatest change in the function value.[1]

Example: Given $f(x) = a'x + x'Bx$, where:

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad a = \begin{bmatrix} 10 \\ 8 \end{bmatrix} \quad B = \begin{bmatrix} -2 & 0 \\ 0 & -1 \end{bmatrix}$$

We know that $f(x_1, x_2) = 10x_1 + 8x_2 - 2x_1^2 - x_2^2$.

Given the initial values of $x_0 = \left[\begin{smallmatrix} 1 \\ 3 \end{smallmatrix}\right]$, we can find that $f(x_0) = 23$. The gradient at x_0 can also be calculated:

$$\nabla f(x_0) = \left[\frac{\partial f(x)}{\partial x_1},\ \frac{\partial f(x)}{\partial x_2}\right] = [10 - 4x_1,\ 8 - 2x_2] = [6,\ 2]$$

[1] For a proof, see "directional derivative" in Greenberg (1978, 156-7).

Therefore, the gradient at x_0 indicates a direction in the ratio of 3:1 in $(x_1,\ x_2)$-space.

Numerical Example: Suppose we have an additional 2 units to allocate between x_1 and x_2 at x_0. We can compare three different allocation strategies in order to maximize our objective function.

Strategy I: Put all units (i.e., 2) on the most profitable activity. This is x_1, since it has the largest marginal product.

$$\tilde{x}_1 = 1 + 2 = 3 \quad \tilde{x}_1 = 3 \quad \Rightarrow f(\tilde{x}) = 30 + 24 - 18 - 9 = 27$$

Strategy II: Even split between x_1 and x_2, i.e., one unit is added to each.

$$\bar{x}_1 = 2 \text{ and } \bar{x}_2 = 4 \quad \Rightarrow f(\bar{x}) = 20 + 32 - 8 - 16 = 28$$

So this result is an improvement over Strategy I.

Strategy III: Use the gradient ratio to set the allocation. The gradient at x_0 is $\nabla f(x_0) = [6,\ 2]$. The ratio is 3:1. Thus of the two extra units, 1.5 goes to x_1 and 0.5 goes to x_2. The new values of x are:

$$\hat{x}_1 = x_1 + 1.5 = 2.5 \quad \hat{x}_2 = x_2 + 0.5 = 3.5$$

Which yields an objective value of:

$$f(\hat{x}) = 25 + 28 - 12.5 - 12.25 = 28.25$$

This strategy is best since it is an improvement over Strategy II.

8.3.2 An Outline of the Gradient Algorithm

Step 1: Pick an initial starting point x_0.

Step 2: Check if $\nabla f(x_0) = 0$ if so, stop since we are at a critical point.

Step 3: If $\nabla f(x_0) \neq 0$ i.e., < 0 for a minimization problem, we move to another point:

$$x_1 = x_0 + \alpha_1 d_1$$

Where d_1 = direction = $\nabla f(x_0)$ and α_1 = step size.

The objective function improvement condition $f(x_0 + \alpha_1 d_1) < f(x_0)$ holds if and only if $\langle \nabla f(x_0),\ d \rangle < 0$ for a minimization problem. Notes:

- This improvement condition is analogous to the negative r_j criterion for entering activities in a minimizing LP problem (See chapter 3).
- The objective function improvement condition has the opposite sign to the first-order optimality condition in equation 8.2 on page 144. This is because the optimum is defined as the point where no improvement of the objective function is possible.

Step 4: Return to Step 2.

A two-dimensional example See Figure 8.6.

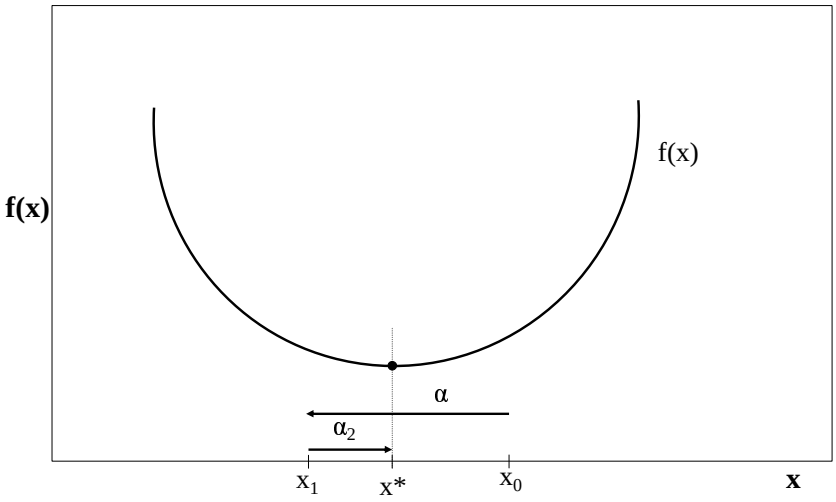


Figure 8.6: Example of steps converging on the minimum.

1. Objective: Select x to minimize $f(x)$.
2. Start at x_0. Slope $= \nabla f(x_0) > 0$, so choose $d < 0$ to satisfy the objective function improvement condition, $\langle \nabla f(x_0), d \rangle < 0$.
3. With α_1 step size, we arrive at x_1. At $x_1 \nabla f(x_1) < 0$, so we choose $d > 0$ to satisfy $\langle \nabla f(x_1), d \rangle < 0$.
4. With α_2 step size, we arrive at x^*. Note that I faked the selection of α_1 and α_2 so that the problem converged in two steps.

5. At $x^*, \nabla f(x^*) = 0$ and $\langle \nabla f(x^*), d \rangle \geq 0$. Therefore, we are at the minimum.

8.3.3 Practical Problems with gradient Algorithms

Starting points

Starting points must be feasible. One method of choosing starting points is to use a linear approximation to find points in the feasible set and use them as the starting points. Note that GAMS gives initial conditions that are feasible.

Step size

The step size must not take you out of feasible set at each iteration. See Figure 8.7 for examples.

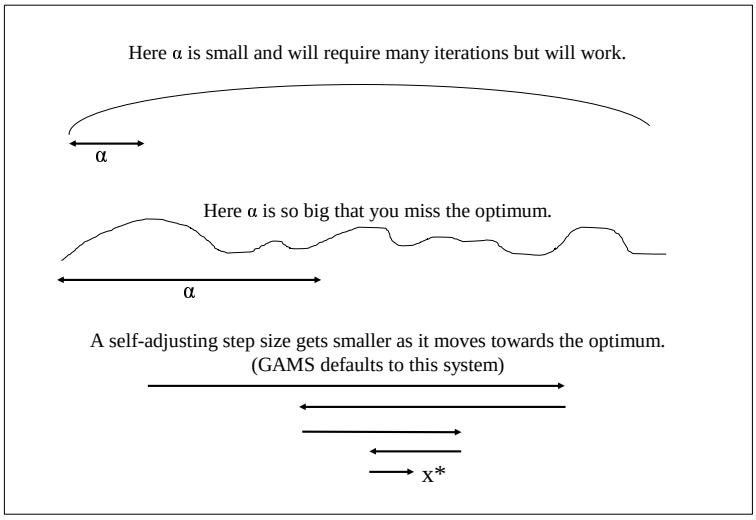


Figure 8.7: Step size examples

Scaling

Scaling the values of the data to balance the Hessian matrix of the objective function is a very important operation for all nonlinear solver routines. The essential aspect is to scale the data values and the corresponding coefficients so that the eigenvalues of the Hessian are within three orders of magnitude

of each other. Technically, the *condition value* for any matrix is the ration of the largest and smallest eigenvalue.

Example: A very simple unconstrained quadratic problem can be defined as:

$$Z = [\alpha_1\ \alpha_2] \begin{bmatrix} x \\ y \end{bmatrix} - [x\ y] \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

Well-scaled outcome: If $\alpha = [10.585, 2.717]$, then:

$$\Gamma = \begin{bmatrix} 0.3786 & 0.00578 \\ 0.00578 & 0.02193 \end{bmatrix}$$

The eigenvalues for Γ are 0.3787 and 0.0218, so the condition number is 17.37. This is a well-scaled result. The surface to be searched by the algorithm appears in Figure 8.8. Note that the surface gradients with respect to x and y are quite similar and the calculation of a gradient will have a similar rounding error in each direction.

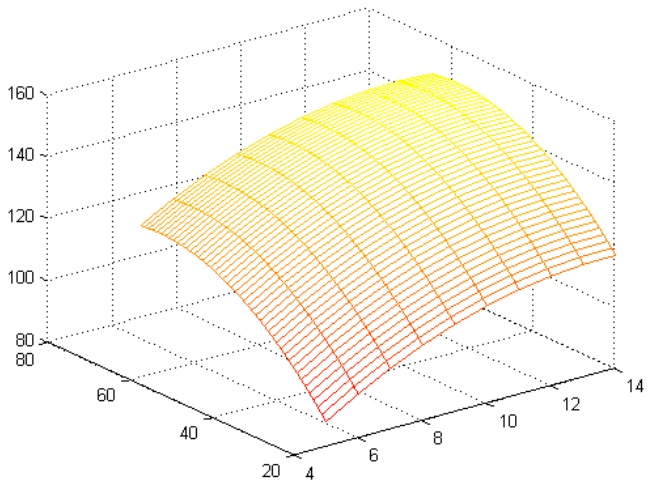


Figure 8.8: Well-scaled outcome. The surface is similar with respect to both x and y axes.

Poorly-scaled outcome: If $\alpha = [10.585, 0.2717]$, then:

$$\Gamma = \begin{bmatrix} 0.3786 & 0.00578 \\ 0.00578 & 0.000219 \end{bmatrix}$$

The eigenvalues for this matrix are 0.3787, 0.0001 and the condition number is 3787. Note that scaling y by a factor of 10 has increased the condition number 218 times. The surface to be searched by the algorithm appears is in Figure 8.9. Note that the gradients are very different and the y-axis values are multiplied by 10. The same rounding errors and step size are applied in both directions, but have very different effects on the change in the objective Z value. Thus the gradient will try and converge in one dimension but not the other.

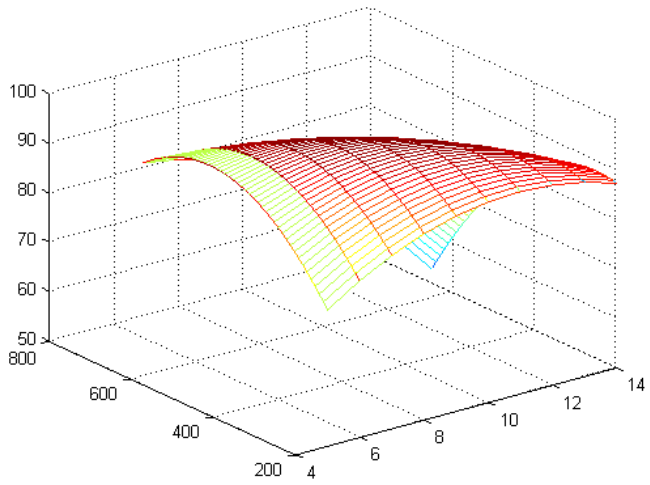


Figure 8.9: Poorly-scaled outcome. An algorithm will fall off the narrow objection function surface.

This "poorly scaled" condition creates a "hill" on the objective function surface that is very narrow, and the algorithm may "fall off" the surface. Given that the rounding error in calculating the numerical gradients is the same for the very large and small values the proportional error will be magnified by large differences in scaling.

8.4 Reduced gradients

For constrained nonlinear optimization problems it is very common for the number of nonzero activities in the optimal solution to be greater than the number of binding constraints. Thus, there may be m-binding constraints and k (k $\geq$ m) activities in the optimum solution, as in the case where we

could use the PMP approach. However, all k activities enter the constraint set despite the m-dimensional basis. The reduced gradient is the nonlinear analog to the reduced cost in LP, and incorporates the linkages between activities due to the constraints. The net effect of a marginal change must therefore consider the direct gradient and the effect on other gradients, due to the linkage of the constraints.

Given the constrained nonlinear problem:

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & Ax \geq b,\ x \geq 0,\ \text{and } q \times 1 \end{aligned}$$

At any point, the $q \times 1$ vector x can be partitioned into three sets:

x_B an $m \times 1$ set of basis (dependent) variables.
x_N a $k \times 1$ set of independent variables.
x_0 a $(q - m - k) \times 1$ set of zero valued variables.

Likewise, the matrix A can be partitioned into three matrices B (an $m \times m$ basis matrix), N (an $m \times k$ matrix of technical coefficients for the independent variables), and D (an $m \times (q - k - m)$ matrix for the zero valued variables). Without loss of generality, we can drop the zero-valued variables, (x_0), and the non-binding constraints from the problem. The problem becomes:

$$\begin{aligned} \min \quad & f(x_B, x_N) \\ \text{subject to} \quad & Bx_B + Nx_N = b \end{aligned} \tag{8.10}$$

Using the constraint, the dependent variables can be written as a function of the independent variables:

$$x_B = B^{-1}b - B^{-1}Nx_N$$

Substituting this expression back into 8.10 produces a new objective function: $\min f(B^{-1}b - B^{-1}Nx_N, x_N)$ with two useful characteristics:

1. The effect of the binding constraints are incorporated in the objective function.

2. The whole problem is expressed as a function of only the independent variables, x_N.

The resulting reduced gradient for this problem is:

$$r_{x_N} = \nabla f_{x_N}(\cdot) - \nabla f_{x_B}(\cdot)B^{-1}N$$

The reduced gradient captures the net effect of a marginal change in an independent variable.

8.4.1 Necessary Condition

It can be shown (Luenberger **get citation**) that a *necessary condition* for a linearly-constrained nonlinear optimization is that all the reduced gradients are zero.

Note that the Gams/MINOS algorithm uses reduced gradients for this purpose in nonlinear problems and prints the "rg" value on the right-hand side of the activities.

8.5 Newton's Method

Newton's Method uses the Hessian as well as the gradient to search over the set of all the critical stationary points, i.e., where $\nabla f(x) = 0$. Since we now consider a sequence of algorithm steps, we use the more general subscript notation of k, k+1, k+2, ... to represent the steps.

8.5.1 Derivation

Starting at some point x_k, we wish to move to a new point x_{k+1}, which has the property that it is a critical point, i.e.:

$$\nabla f(x_{k+1})' = 0 \tag{8.11}$$

In this method, we start with the gradient and expand around it. Using a Taylor series expansion of the gradient of x_k, we get:

$$\nabla f(x_{k+1})' \approx \nabla f(x_k)' + H_{x_k}(x_{k+1} - x_k) \stackrel{set}{=} 0 \tag{8.12}$$

where H_{x_k} is the Hessian of $f(x)$ at x_k. $\nabla f(x_{k+1})' = 0$ is called the "stationary point condition" and allows us to rewrite equation 8.12 as:

$$H_{x_k}(x_{k+1} - x_k) = -\nabla f(x_k)' \tag{8.13}$$

Assuming that H_{x_k} is nonsingular and invertible, we can multiply 8.13 by $H_{x_k}^{-1}$ and move x_k to the right-hand side to yield:

$$x_{k+1} - x_k = -H_{x_k}^{-1}\nabla f(x_k)' \Rightarrow \therefore x_{k+1} = x_k - H_{x_k}^{-1}\nabla f(x_k)' \tag{8.14}$$

Note that with a quadratic objective function, a stationary point x_{k+1} is reached in one step! An example of Newton's (1642 - 1727) mind at work, which is still topical after three hundred years. Beautiful!

Example: Newton's Method applied to a quadratic problem.

1. Start with $f(x) = a'x + \frac{1}{2}x'Bx$ where the Hessian matrix B is nonsingular and positive definite at x_0.

2. Take the gradient $\nabla f(x_0)' = (a+Bx_0)'$ where $H_{x_0} = B$ — independent of the choice of x_0.

3. Apply Newton's Method:

$$x_1 = x_0 - B^{-1}(a + Bx_0) = -B^{-1}a$$

   Also independent of choice of x_0.

4. Check if x_1 is a stationary point by substituting the Newton value for x_1 into $\nabla f(x_1)' = a + Bx_1$. This gets us:

$$\nabla f(x_1)' = a + B(-B^{-1}a) = a - a = 0$$

5. Therefore, we have reached a stationary point with only one estimation.

Note: For non-quadratic functions, the Newton is an approximation to the function and therefore will take several steps. However, if the Hessian is well-conditioned, Newton's method will converge rapidly.

8.5.2 Caveats on Newton's Method

- The Newton approach finds stationary points, it does not guarantee that they're global or even local optima.

- The Hessian is often hard to invert.

8.5.3 Criteria for Non-Linear Algorithms

There are several criteria used to choose which non-linear algorithm suits your situation. Here is how Newton differs from a gradient approach:

Convergence rate: Newton is faster than the steepest-decent method.

Computational difficulty: Newton requires generating, then inverting the Hessian.

Stability: Newton requires a well-conditioned Hessian.

8.5.4 Newton Step size

Used when the objective function is not quadratic in form. In this case the Taylor series expansion that forms the basis of the Newton derivation is only an approximation of the function:

$$x_{k+1} = x_k - \alpha H_{x_k}^{-1} \nabla f(x_k)' \text{ for a minimization problem.}$$

Note: If H_{x_k} is an identity matrix, then this is equivalent to the steepest-descent method and reduces to:

$$x_{k+1} = x_k - \alpha \nabla f(x_k)'$$

8.5.5 Movement Towards a Minimum

The necessary conditions for movement towards a minimum with the Newton method are clarified by a change to condensed notation. Define:

$$\begin{array}{rll} d_k & = -H_{x_k}^{-1} \nabla f(x_k)' & \text{direction} \\ M_k & = H_{x_k}^{-1} & \text{algebraic simplification} \\ g_k & = \nabla f(x_k)' & \text{algebraic simplification} \end{array}$$

For a minimum: We require the "downhill" condition in the new notation to be $\langle d_k,\, g_k \rangle < 0$ i.e., $d_k' g_k < 0$. Using the notation defined above, this is equivalent to:

$$(-M_k g_k)' g_k < 0 \;\text{ or }\; g_k' M_k g_k > 0$$

This is most easily fulfilled if $M_k = H_{x_k}^{-1}$ is a positive-definite matrix.

8.5.6 Desired Conditions for H_{x_k}

- Non-singular
- Positive definite
- Well-conditioned (eigenvalues within 10^3 of each other)

8.5.7 How to make an Ill-conditioned Hessian Well-conditioned

- Greenstadt's Method — similar to ridge-regression.
- Scaling — a much better approach that aims to change the units of measurement associated with the Hessian so they are close to each other (i.e. within 10^3 of each other).[2]

[2]For a more extensive discussion on scaling see Gill and Wright (1982, 346-354).

The GAMS solver *Conopt2* now has an effective scaling routine that updates the scaling factors as the solver progresses. This scaling option should not be confused with the standard GAMS scaling option that scales the GAMS problem before the solver is called. The *Conopt2* scaling system has to be called from an appropriate "Options" file in the GAMS program. For an example of a *Conopt2* options file see the Gams templates.

Chapter 9

Calibration with Maximum Entropy

9.1 Measuring Information

The PMP cost function in Chapter 5 has its quadratic coefficients solved analytically by solving two equations in two unknowns. However, this analytical solution requires that the quadratic cost matrix is specified as strictly diagonal. There is a significant practical problem with the diagonal specification in that is assumes that there are no "cross effects" between the amount of land allocated to crops, apart form the effect on the total land constraint. In economic parlance this assumption requires that there are no substitution or complementary cost effects between crops grown in the same district or farm. Clearly the almost universal existence of rotations in crop production implies that farmers are well aware of the interdependencies among crops, and use them to stabilize or increase profits. Clearly, the assumption of a diagonal cost matrix is unrealistic. But to calibrate a full matrix of coefficients requires solving an "ill-posed" problem in which we are trying to estimate more parameter values than we have data points, an estimation with negative degrees of freedom. As we saw in chapter 2, we cannot solve ill-posed problems by inverting matrices. Fortunately there is an alternative method that can obtain parameter estimates for ill-posed problems using information theory and the principle of maximum entropy.

Claude Shannon was a giant in information theory. In 1948, he published a paper that proposed a mathematical way of measuring information, and started the information revolution. Shannon noted that information is inextricably linked with the probability of an event about which the signal tells

us. Clearly a signal that tells us that an extremely unlikely event, such as a bad earthquake in Davis, has gone from a very low probability of happening to a very high probability has a very high information content. Note, we are not even considering whether the event has any particular value associated with it. Likewise a signal that tells us that a very likely (high probability) event will happen has a low information content.

Shannon proposed several axioms that a measure of information must have, and showed that the only measure of information content that satisfied the axioms is if the information content of a signal is:

$$\ln(p) \text{ where } p \text{ is the prior probability of the event happening} \tag{9.1}$$

Shannon extended the definition from single probabilities to discrete distributions and defined the expected information content of a prior distribution $\Sigma_i p_i$ as the entropy of a distribution:

$$H = -\Sigma_i p_i \ln p_i \tag{9.2}$$

It follows that a distribution that has a uniform distribution, in which each event is equally likely, has the highest entropy and the lowest information content. Conversely, a distribution that puts a weight of one on a single outcome and zero on the rest, has an entropy of zero. Remember that, counter intuitively, a distribution that shows that a given event will occur with probability one, has the highest information content (lowest entropy).

If we specify a set of discrete values over the feasible range of a parameter value termed the support set or space then multiplying each support value by its associated probability yields and expected parameter estimate, and there is an entropy value for the distribution associated with the parameter value. However, as one would expect with an ill-posed problem, there is an infinite set of probabilities that will yield any given parameter value. We have to use the entropy criteria to choose a unique distribution from among the infinite set of feasible distributions.

Edwin Jaynes, another information pioneer, showed that the distribution with the highest entropy can occur in the most number of ways. The concept of "multiplicity" is similar to frequency or maximum likelihood in conventional estimators. Essentially the distribution with the maximum entropy is the best estimator. Maximizing entropy also has another fine property, in that the entropy function in equation 9.2 has a unique solution at the maximum.

Thus by defining a support space for our parameter and solving for the maximum entropy distribution that has an expected value consistent with

the data, we can get a unique solution to the ill-posed problem, and estimate more parameters than we have observations!

9.1.1 The "Loaded" Dice Example

Here we see how we intuitively calculate six parameters (probabilities) from a single data observation. Jaynes in his 1963 Brandeis lectures on maximum entropy used the example of a game using a six-sided dice with the usual values from 1 – 6. Suppose that you know the average value of a large number of independent rolls of the dice. For a given mean value there are an infinite number of combinations of the six probabilities that could have generated the mean value. The problem is ill-posed because we are given one data point, the mean value, and we have to estimate six probabilities. The only structural constraint that we have is that the probabilities have to add to one.

Think about a game in which your opponent produces a dice and suggests that it is rolled twenty times, and if the average score is above 4 you win, and if the average score is below 3 they win. With "fair dice" (in which we assume that the probability of each side coming up is even), this is a fair game as one would expect that the average of a series of rolls would be 3.5. If you notice that the rolls of the dice come up consistently with 1, 2, or 3, you will alter your initial assumption that the dice are fair, and assume that your opponent is cheating with dice "loaded" to favor the low scores. You have just performed an ill-posed estimation of the probabilities.

Using the principle of multiplicity, the distribution that maximizes the entropy is the most likely to be probabilities underlying the dice. If we define the score values as x_i, the dice problem, defined and coded by W. Britz, is specified as:

$$\begin{aligned} \max \quad & H = -\Sigma_i p_i \ln p_i \\ \text{subject to} \quad & \text{Avg. Score} = \Sigma_i x_i p_i \\ & \Sigma_i p_i = 1, \quad p_i \geq 0 \end{aligned} \tag{9.3}$$

The GAMS solution to the dice problem is posted with the GAMS templates on the webpage. The Maximum Entropy (ME) probability results for a range of average scores from 1.5 to 5.5 is plotted in Figure 9.1 and the entropy value for different average scores is plotted in Figure 9.2.

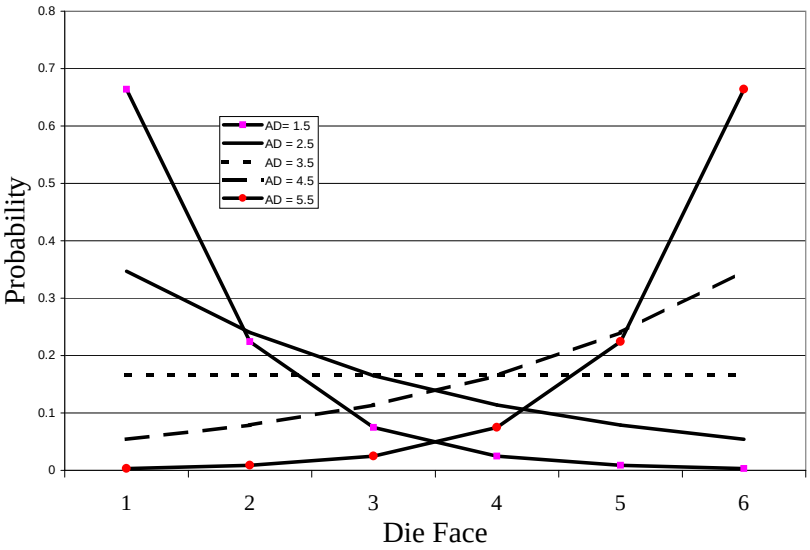


Figure 9.1: Max Entropy Dice Probabilities

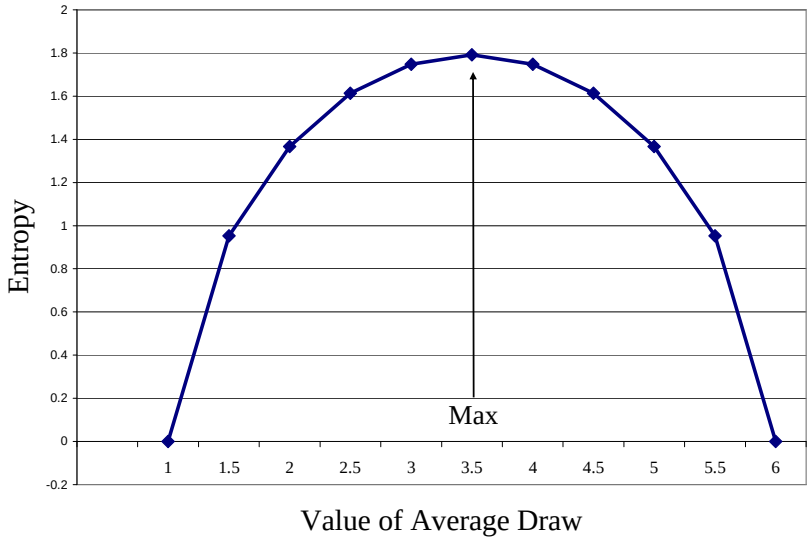


Figure 9.2: Entropy Values. Note the unique maximum value.

9.1.2 A Simple Example of Maximum Entropy Parameter Estimation

Assume that we want to estimate two parameters in the simple quadratic cost function:

$$TC = ax + \frac{1}{2}bx^2 \tag{9.4}$$

We only have one observation in which we see that the marginal cost is 60 when the output $x = 10$. The data relationship that we have to satisfy is:

$$60 = a + 10b \tag{9.5}$$

There are an infinite number of parameter values for a and b that satisfy this relationship. Suppose we consider five discrete values for a support space. If we rule out negative costs, the lower support space is bounded at zero and the upper support space can be defined by the coefficient value that would explain all of the cost when the other coefficient is zero. Using this as a basis for the support values, five evenly distributed support values would be:

$$za_i = [0,\ 8,\ 16,\ 32,\ 40] \quad zb_i = [0,\ 1,\ 2,\ 3,\ 4] \tag{9.6}$$

A feasible set of probabilities that would solve equation 9.7

$$\begin{aligned} MC_j \quad & = \Sigma_i z a_i p a_i + (\Sigma_i z b_i p b_i) x_j \\ \text{subject to} \quad & \Sigma_i p a_i = 1, \quad \Sigma_i p b_i = 1, \quad p a_i, p b_i \geq 0 \end{aligned} \tag{9.7}$$

is $pa_i = [0,\ 0,\ 0,\ 0.25,\ 0.75]$ and $pb_i = [0,\ 0.15,\ 0.5,\ 0.35,\ 0]$. The distributions for this solution are plotted in histograms in Figure 9.3

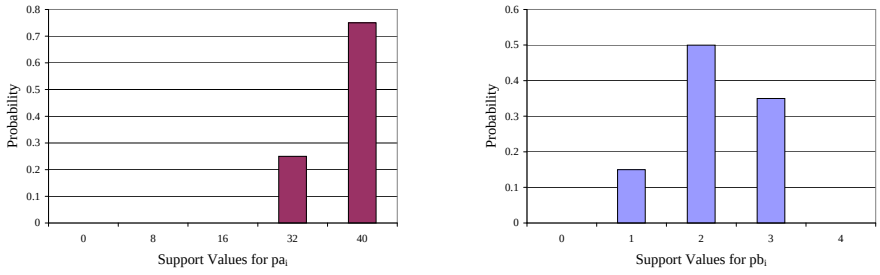


Figure 9.3: Parameter Probability Plots — Consistent, but *ad hoc* Values

9.1.3 The Maximum Entropy Solution

The maximum entropy problem that solves for the two distributions that are most likely to have generated a marginal cost of 60 for an output of 10 is:

$$\begin{aligned} \max \quad & -\Sigma_i pa_i \ln pa_i - \Sigma_i pb_i \ln pb_i \\ \text{subject to} \quad & MC_j = \Sigma_i z a_i pa_i + (\Sigma_i z b_i pb_i) x_j \\ & \Sigma_i pa_i = 1, \quad \Sigma_i pb_i = 1, \quad pa_i, pb_i \geq 0 \end{aligned} \tag{9.8}$$

The Maximum Entropy (ME) solution to this problem is plotted on the histograms in Figure 9.4. Clearly the ME solution satisfies the data constraint without having to use such specialized, and unlikely, distributions as the *ad hoc* solution above.

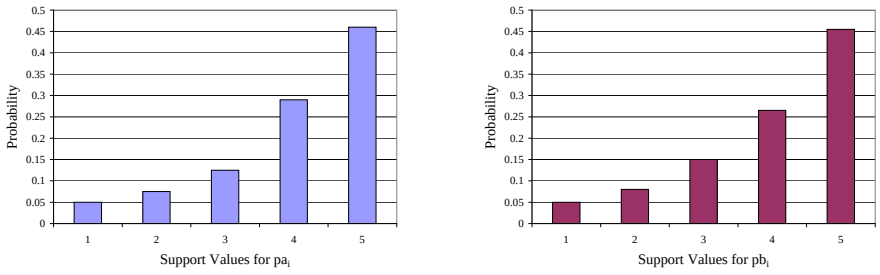


Figure 9.4: Consistent Maximum Entropy Distributions

The expected parameters that result from these two calculations are:

	Ad Hoc Values	Maximum Entropy
E(a)	38.0	30.217
E(b)	2.2	2.978

9.2 Maximum Entropy PMP Modeling

9.2.1 The Basic ME-PMP Model

So far the cost matrix for PMP models has been restricted to a diagonal specification that implicitly assumes that the cost of growing a particular acreage of a crop is completely disconnected from the other crop acreages. In a typical agricultural production situation crops compete for limited land resources of different quality. In addition, crop rotations are based on substitute or complementary relations between crops in the rotation. Under this more realistic situation the marginal cost of growing a quantity of crop depends on its own level of production, and also the level of production of all the other crops. The PMP total cost function for n crops is:

$$TC = \alpha' x + 0.5 x' \Gamma x \tag{9.9}$$

where Γ is an $n \times n$ matrix and the marginal cost for a given crop k is a function of all the x values.

$$MC_{x_k} = \alpha_k x_k + \Sigma_i \gamma_{ki} x_i \tag{9.10}$$

Equation 9.10 shows the fundamental problem of ME- PMP. For a problem with n crops in the optimal solution we have n marginal costs on the left hand side, but a total of $n + \frac{n^2+1}{2}$ unknown parameters in n equations. This total has n α parameters, and $\frac{n^2+1}{2}$ unknown γ parameters. This last set is based on the symmetry requirement of the cost matrix. Clearly we cannot derive this number of parameters analytically, or estimate them normally as we have negative degrees of freedom. The approach that we use is to solve a maximum entropy reconstruction problem that generates the α and γ parameters that satisfy the first order conditions of equation 9.10.

9.2.2 Defining the Support Space Values — z-values

To ensure feasible solutions that have a unique optimum the PMP matrix Γ has to have feasible support values that satisfy the first order conditions. Since the z-values are "priors" on the parameters that we want to reconstruct they can have a strong influence on the resulting parameters. We need to be aware that the defined z-values must be:

1. Feasible for the data constraints that we are going to impose on the support space.

2. Neutral in their effect on the estimate unless we have information that we want to incorporate in the priors

Feasible Supports

In the simple parameter and dice example the z-values of the dice are clear, and for the cost parameters they were glossed over. When defining a system for generating z-values for different models and crop components in the models we need a more formal approach. Feasible z-values are generated by defining the z-values (say five) as being the product of a single centering parameter value and five z-weight parameters:

$$z_{val}(j,p) = z_{wgt}(p) * c_{val}(j)$$

The centering $c_{val}(j)$ is an empirical value the modeler calculates will be feasible for the data set imposed on it. For example, if we are reconstructing a PMP cost function matrix as a function of the acreage planted of a crop, then a feasible value for the marginal cost coefficient is the average cost divided by the base acres.

Non-Informative z-values: The $z_{wgt}(p)$ parameters can be thought of as spreading the $c_{val}(j)$ value across a feasible range for estimation. For a diagonal cost parameter that is strictly positive, the range may go from 0, 0.5, 1.0, 1.5, 2.0. For an off-diagonal parameter the $c_{val}(j)$ weight will be reduced, possibly to 0.25 the diagonal weight and centered on 0, with weights of -1.5, -0.75, 0.0, 0.75, 1.5. This system of generating z-values is designed to automatically generate a set of feasible but noninformative prior z-values. Informative z values are used in a later section

9.2.3 Adding Curvature Conditions by Cholesky Decompositions:

To ensure that the resulting matrix PMP model converges to a stable solution the necessary second order conditions require that the Hessian of the cost function is negative definitive. Since the Hessian is the Γ matrix, this requires that Γ is positive definite. Diewert shows that a necessary condition for a matrix to be positive definite is that the diagonal elements of its Cholesky decomposition matrix are positive.

What is a Cholesky decomposition? Judd (1999) says that we can think of a Cholesky decomposition as being the "square root" of a matrix. The C matrix is a lower triangular matrix L that when post-multiplied by its

transpose yields the original matrix. If $\Gamma = LL'$ then L is the Cholesky decomposition.

If Γ is a 3×3 matrix ...

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{bmatrix}$$

The Cholesky decomposition ...

$$L = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix}$$

$$\Gamma = LL' = \begin{bmatrix} l_{11}l_{11} & l_{11}l_{21} & l_{11}l_{31} \\ l_{21}l_{11} & l_{21}l_{21} + l_{22}l_{22} & l_{21}l_{31} + l_{22}l_{32} \\ l_{31}l_{11} & l_{31}l_{21} + l_{32}l_{22} & l_{31}l_{31} + l_{32}l_{32} + l_{33}l_{33} \end{bmatrix}$$

This latter expression results in two sets of equations for the diagonal and off-diagonal elements of Gamma:

$$\gamma_{jj} = \sum_{k=1}^{j} l_{jk}^2 \quad \gamma_{ij} = \sum_{k=1}^{j} l_{ik}l_{jk}$$

by adding these equations as constraints on the entropy problem and constraining the diagonal Cholesky terms to be greater than zero (here $>$ 0.001), we can impose curvature conditions on the Gamma matrix that ensures that the simulated problem will satisfy the second order conditions.

9.2.4 Implementing ME-PMP on the Yolo County Model

The procedure that implements a ME estimate of the PMP cost matrix for the Yolo model is found in the Gams Template section on the website. The procedure is exactly the same as the basic and standard PMP approaches for the first stage LP model with calibration constraints. The major difference is in the next stage, which for basic and standard PMP, involves the analytic calculation of the α and γ parameters. As previously stated, for the ME-PMP method we solve a maximum entropy reconstruction of the α and γ parameters. The standard PMP method is used to generate the marginal costs for the left hand side of the entropy constraints. First, we use an elasticity on the marginal crop to calculate the "Adj" value, that is the

difference between average and marginal costs for the marginal crop. Second, the Adj value is used to increase all the λ_2 values, to adjust them for the lower opportunity cost of land. The new λ_2 values are then used to calculate the marginal cost for each crop, defined as Lam, where $Lam = \lambda_2 + AC$.

Defining the z weights and z-values: The z weights are used to impose prior restrictions or properties on the resulting parameters. For example, since the α values are cost function intercepts, they can be positive or negative. In this first example of Maximum Entropy PMP models we combine the curvature constraints and the prior support values for the PMP cost matrix Γ by defining Γ in terms of the product of its lower triangular Cholesky matrices, namely LL'. By defining the supports for the diagonal values to be positive we ensure that the resulting Γ matrix curvature restrictions are satisfied. Note that the LL' Cholesky decomposition of Γ has the restriction that the resulting Γ matrix is positive definite. For the more general case of a positive semi definite restriction Paris and Howitt (1998) use the decomposition LDL'. In the Yolo case the zweights go from - 4.5 to 4.5 and are centered on zero. The triangular Cholesky decomposition L must have positive diagonal elements, so the diagonals have a zweight set that are strictly positive and range from 0.00-1 to 2.0. In contrast, the off diagonal elements have to be able to represent the cost effects of complementary crops that would lower the cost of a given crop, or substitute crops that compete for resources and increase the costs of a given crop. It follows that the zweights for the off-diagonal elements of L can be positive or negative, but are unlikely to have as much influence on marginal costs as the direct diagonal coefficients, and are therefore set at values that range from -1.5 to 1.5. The z values are generated by multiplying the z weights by Lam, the PMP marginal cost value from the previous stage, namely the sum of the average cost and the λ_2 value for the relevant crop.

The ME Parameter Reconstruction Model Following the pattern of equation 9.8 the full PMP matrix for the Yolo county model is reconstructed by solving the following maximum entropy problem. Assuming that the z values for the alpha and L, and i x k triangular matrix, have been defined above, the ME problem for a production unit with i crops is:

$$\begin{aligned} \max \quad & -\Sigma_i \Sigma_p pa_{pi} \ln pa_{pi} - \Sigma_i \Sigma_k \Sigma_p \quad pl_{pik} \ln pl_{pik} \qquad (9.11) \\ \text{subject to} \quad & Lam_i = \Sigma_p za_{pi} pa_{pi} + \Sigma_i [\Sigma_k (\Sigma_p zl_{pik} pl_{pik}) * (\Sigma_p zl_{pki} pl_{pki})] x_i \\ & \Sigma_p pa_{pi} = 1, \quad \Sigma_p pl_{pik} = 1, \quad pa_{pi}, pl_{pik} \geq 0 \end{aligned}$$

The optimal values for pa_{pi} and pl_{pik} are used to generate the α and Γ parameters:

$$\begin{aligned} \alpha_i &= \Sigma_p \quad za_{pi} pa_{pi} \\ \gamma_{ik} &= (\Sigma_p \quad zl_{pik} pl_{pik}) * (\Sigma_p \quad zl_{pki} pl_{pki}) \end{aligned} \tag{9.12}$$

The resulting α vector and Γ matrix are now used to solve the standard nonlinear optimization PMP problem:

$$\begin{aligned} \max \quad & \Sigma_i p_i * yield_i * x_i - \Sigma_i \quad \alpha_i x_i - 0.5 \Sigma_i \Sigma_k \quad z_{ik} x_i x_k \\ \text{subject to} \quad & \Sigma_i a_{ij} x_i \leq b_j \\ & x_i \geq 0 \end{aligned} \tag{9.13}$$

This PMP-ME model will calibrate exactly in inputs and outputs, but the objective function will differ from the Linear program solution because of adjustment of the resource opportunity cost based on the marginal crop elasticity, as discussed in the previous chapter.

9.3 Using Prior Elasticities- The Heckelei-Britz ME-PMP Approach

Informative Prior z-values: A problem with the basic PMP-ME approach is that the curvature restrictions and the prior supports for the Γ matrix are inextricable. Essentially this prevents the direct use of priors on Γ such as estimates of the elasticity of supply. In a more general specification, Heckelei and Britz (2000) show how the z-values on the cost slope parameters are defined to be informative, in that their centering value is defined by the model parameters and a prior elasticity value. The elasticity value is entered in the same manner as in the PMP elasticity calibration, and forms a stronger basis for the defined z-values. Note this elasticity is based on "scalar" reasoning, in the calculation is based on single coefficients and ignores the effect of the off diagonal coefficients and resource constraints, however, as can be seen from the empirical elasticity test, they do give the model an operating range of elasticities.

More in this section later:

9.4 Obtaining Consistent Estimates of Resource Shadow Values

The basic PMP model defined in sections .. is defined as a primal model where the shadow values for the constraining resources are determined by the calibration constrained linear model, although the model that we maintain is closer to the true situation is the nonlinear PMP model.

9.4.1 The Symmetric Positive Equilibrium Model:

Paris and Howitt (2001) address the problem of inconsistent shadow values by specifying a symmetric model SPEP that simultaneously calibrates the primal and dual constraints of the problem. Using the notation in section ... the symmetric problem is defined as ..

More in this section later:

9.4.2 The Heckelei and Wolff Critique and Solution:

Heckelei and Wolff (2003) point out that the use of the dual shadow values estimated in the initial linear model is inconsistent with the nonlinear cost or production functions used in the PMP calibration specification...

More in this section later:

9.5 Reconstructing Production Function Models

While both ME and standard PMP models calibrate correctly and have consistent output supply and input demand elasticities, other than land, the production technology is still Leontieff. In this section we show how we can use the initial calibrated LP stage to calculate the crop specific land opportunity costs, and then use the ME approach to reconstruct the parameters of a production function. In order to have the multi-output crops calibrate with constant linear output and input prices we require that the production functions show decreasing returns to scale. There are several methods of analytically calibrating CES and Cobb Douglas production functions to observed cost shares (See Howitt (J Ag Econ 1995) Useful as these methods are, they are invariably specified as homogenous of degree one, with constant returns to scale (CRS). The CRS nature of the production function requires that the crops in the model have to be calibrated by PMP nonlinear cost functions. This raises a fundamental theoretical inconsistency in this type of "ad hoc" model, in that the production and cost specifications are not consistent. A consistent approach is to calibrate a more general class of production functions, namely the Quadratic production function and its associated Trans-log and Generalized Leontieff forms. Essentially, we are going to change our basic calibration assumptions from a model where we assume constant per acre yields and increasing costs, to one where the increasing costs are caused by the decreasing marginal productivity of inputs with expanded scale, and input prices are constant. The fundamental quadratic production function for each crop is defined below in equation 9.14. Note that the input quantities x_{ji} are now defined as the total input quantity for the farm , region, or other modeling unit. Inputs are **not** measured in per acre terms as in all the other models. This is because the production function for a given crop regards land as another input whose marginal conditions define its allocation. The model is now able to trade off on both the intensive and extensive margins, namely the extensive margin of how much land is used in producing crop j, and the intensive margin of the quantity of other inputs to allocate to crop j. The crop yield per acre is now an endogenous variable in contrast to the usual PMP model where yields are fixed parameters. Defining the subscript "k" as an alias of subscript "i", the production function for the jth crop is:

$$y_j = \Sigma_i \alpha_{ji} x_{ji} - 0.5 \Sigma_i \Sigma_k \quad z_{jik} x_{ji} x_{jk} \tag{9.14}$$

The production function based optimization model that uses equation 9.14 for j crops each with i inputs is defined as:

$$\begin{array}{lll} \max\limits_{x} & \Pi = & \Sigma_j(p_j[\Sigma_i\alpha_{ji}x_{ji} - 0.5\Sigma_i\Sigma_k \quad z_{jik}x_{ji}x_{jk}] - \Sigma_i\omega_i x_{ji}) \\ \text{subject to} & & \Sigma_j x_{ji} \leq b_i \\ & & x_{ji} \geq 0 \end{array} \tag{9.15}$$

The first order conditions for this model, when set equal to zero are:

$$\frac{\delta\Pi}{\delta x_{ji}} = p_j(\alpha_{ji} - \Sigma_k z_{jik}x_{jk}) - \omega_i - \lambda_i \quad = 0 \tag{9.16}$$

The first order constraints for the production function model can be explained as equating the marginal physical product of each input in each crop with the ratio of the total marginal cost of a unit of input (input price plus opportunity cost) to output price.

$$\frac{\omega_i + \lambda_i}{p_j} \quad = \quad \alpha_{ji} - \Sigma_k z_{jik}x_{jk} \tag{9.17}$$

Note that equations 9.16 and 9.17 satisfy the conditions that equate the value marginal product of an input used in all crops to its marginal cost. In addition to the usual first order conditions above, we also want the production model to calibrate to the average product or yield per acre, essentially ensuring that the marginal conditions integrate correctly to the total crop product observed in the base year. Given that crop and animal yields are a data source that farmers can remember with accuracy, it is important to use this information to calibrate the model. Thus the ME reconstruction of the production function is restricted by both marginal and total product constraints. The total product constraint is:

$$yield_j * x_{j,land} \quad = \quad \Sigma_i(\alpha_{ji}x_{ji} - 0.5\Sigma_k z_{jik}x_{jk}x_{ji}) \tag{9.18}$$

To ensure that the resulting optimization model satisfies the second order conditions for a unique optimum, we have to impose symmetry and positive definite restrictions on the matrix of the quadratic production function. We use the same Cholesky decomposition approach as used in the PMP-ME model. Accordingly, the Zeta matrix in the quadratic production function is estimated by the Cholesky matrices L, where $Z = LL'$. The resulting maximum entropy reconstruction problem is:

$$\max \quad -\Sigma_j\Sigma_i\Sigma_p pa_{pji} \ln pa_{pji} - \Sigma_j\Sigma_i\Sigma_k\Sigma_p \quad pl_{pjik} \ln pl_{pjik}$$

subject to

$$\frac{\omega_i + \lambda_i}{p_j} = \Sigma_p za_{pji} pa_{pji} - \Sigma_i[\Sigma_k(\Sigma_p zl_{pjik} pl_{pjik}) * (\Sigma_p zl_{pjki} pl_{pjki})]x_{ji} \quad (9.19)$$

$$y_j x_{j,land} = \Sigma_i[(\sigma_p za_{pji} pa_{pji}) - 0.5 * \Sigma_k[(\Sigma_p zl_{pjik} pl_{pjik}) * (\Sigma_p zl_{pjki} pl_{pjki})]x_{jk}]x_{ji}$$

$$\Sigma_p pa_{pji} = 1, \quad \Sigma_p pl_{pjik} = 1, \quad pa_{pji}, pl_{pjik} \geq 0$$

The ME production model optimizes the values for pa_{pji} and pl_{pjik}. The optimal values for pa_{pji} and pl_{pjik} are used to generate the α and Z parameters:

$$\begin{aligned} \alpha_{ji} &= \Sigma_p \quad za_{pji} pa_{pji} \\ z_{jik} &= (\Sigma_p \quad zl_{pjik} pl_{pjik}) * (\Sigma_p \quad zl_{pjki} pl_{pjki}) \end{aligned} \quad (9.20)$$

The resulting α vector and Z matrix are now used to solve the nonlinear production function optimization problem:

$$\begin{aligned} \max \quad & \Sigma_j[p_j * (\Sigma_i[\alpha_{ji} - 0.5 * \Sigma_k z_{jik} x_{jk}]x_{ji}) - \Sigma_i x_{ji}\omega_i] \\ \text{subject to} \quad & \Sigma_j x_{ji} \leq b_i \\ & x_{ji} \geq 0 \end{aligned} \quad (9.21)$$

This quadratic production function problem calibrates in terms of input quantities, output quantities, yields, and the Adj adjusted shadow value for land and other fixed inputs. As with the PMP-ME model the objective function does not calibrate to the initial LP value due to the use of prior supply elasticity values on the marginal land using crop. To reiterate, the production function model in equation 9.21 above responds to changes in output or input prices by adjusting at the intensive and extensive margins. Thus a change in the quantity of water available can no only change the proportions of crops grown, but also the amount of fertilizer applied and the resulting crop yield.

9.5.1 Calculating Comparative Static Parameters

The quadratic production function model has convenient properties for calculating policy parameters. Defining equation 9.21in matrix terms for a single crop j, produced by an $i \times 1$ vector of inputs x_j we get:

$$\Pi_j = p(\alpha' x_j - \frac{1}{2} x_j' Z x_j) - \omega' x_j \tag{9.22}$$

Note that the Hessian of the unconstrained profit function above is simply:

$$\frac{\partial^2 \Pi}{\partial x_{ij}^2} = Z$$

Calculating the Derived Demands for Inputs: For simplicity, we will use the unconstrained profit function for a single crop in equation 9.22:

$$\begin{aligned} \tfrac{\partial \Pi_j}{\partial x_j} &= p_j'(\alpha - Z x_j) - \omega = 0 \\ -Z x_j &= \tfrac{\omega}{p_j} - \alpha \\ x_j^* &= Z^{-1}\alpha - Z^{-1}\tfrac{\omega}{p_j} \end{aligned}$$

Define the matrix of demand slopes as: $G_j = Z^{-1}\frac{1}{p}$ and the vector of intercepts $a_j = Z^{-1}\alpha$. Using these definitions, the vector of optimal input demand quantities for crop j is a function of the crop price p_j and the vector of input prices ω and is:

$$x_j^* = a_j + G_j \omega \tag{9.23}$$

From the above equations, it is clear that if we can invert the Hessian of the profit function (Z^{-1}) we can calculate the derived demand for each input for each crop as a linear function of input and output price.
Note

First: That x_j^* is an $i \times 1$ vector of the optimal inputs for crop j, and a_j is an $i \times 1$ vector of intercept terms and G_j is an $i \times i$ matrix of derived demand slopes.

Second: That the demand for a given input i used in crop j is a function of the prices of the other inputs as well as its own price.

Third: The elasticity of demand is based only on the own price effect, thus we want GAMS to use only the i^{th} diagonal elements of DS_j. The elasticity of the input demand for x_i in crop j is:

$$\eta_{ij} = G_{j,i,i} \frac{\omega_i}{x_i^*} \tag{9.24}$$

Note that this elasticity is based on a single crop. For the usual multi-output case, we weight the individual crop contribution by their relative resource use to arrive at a weighted elasticity for the resource.

Calculating Supply Functions and Elasticities: Since production is a function of optimal input allocation and we now have the input demands as a function of input and output price, we can derive the output supply function by substituting the optimized input derived demands into the production function and simplify in terms of the output price. Going back to the derived demand and production function formulæ:

$$x^* = Z^{-1}\alpha - Z^{-1}\frac{\omega}{p_j} \text{ and } y_j^* = \alpha' x^* - \frac{1}{2} x^* Z x^* \tag{9.25}$$

Defining $r_i = \frac{\omega_i}{p_j}$ and using the vectors α and r, we get:

$$y^* = \alpha' Z^{-1}(\alpha - r) - \frac{1}{2}(\alpha - r)' Z^{-1} Z Z^{-1}(\alpha - r) \tag{9.26}$$

Multiplying out and collecting α and r terms yields:

$$y^* = \frac{1}{2} \quad \alpha' Z^{-1}\alpha - \frac{1}{2} \quad r' Z^{-1} r$$

$$y^* = \hat{\alpha} - \frac{1}{2}\left(\frac{\omega_1}{p_j} \cdots \frac{\omega_I}{p_j}\right)' Z^{-1} \left(\frac{\omega_1}{p_j} \cdots \frac{\omega_I}{p_j}\right)$$

Factoring out p_j results in:

$$y^* = \hat{\alpha} - \frac{1}{2p^2} \quad \omega' Z^{-1} \omega \tag{9.27}$$

Substituting the expression for $\frac{\partial y^*}{\partial p_j}$ into the supply elasticity formula, and separating out ω and p, we get the supply elasticity expression:

$$\begin{aligned}
\frac{\partial y^*}{\partial p_j} &= \frac{1}{p^3} \quad \omega' Z^{-1} \omega \\
\eta_{sj} &= \frac{\partial y^*}{\partial p_j} \frac{p_j}{y} = \frac{1}{p^3} \quad \omega' Z^{-1} \omega \frac{p_j}{y} \\
\eta_{sj} &= \frac{1}{p^2 y_j} \quad \omega' Z^{-1} \omega
\end{aligned}$$

Calculating Elasticities of Input Substitution: There are many elasticities of substitution with different advantages and disadvantages. To demonstrate that we can obtain crop and input specific elasticities of substitution we use the classic Hicks elasticity of substitution defined for the two input case by Chambers (1988, p 31) as:

$$y^* = f(x_1 \dots x_I)$$

$$\sigma_{1,2} = \frac{-f_1 f_2 (x_1 f_1 + x_2 f_2)}{x_1 x_2 (f_{11} f_2^2 - 2 f_{12} f_1 f_2 + f_{22} f_1^2)}$$

where f_i and f_{ij} are derivatives.

Further Reading See Golan et al. (1996); Mittelhammer et al. (2000); Paris and Howitt (1998)

9.5.2 Using Alternative Functional forms for the Production Function

Throughout this analysis we have used the quadratic production function. Two other functional forms that are widely used are the Generalized Leontief and Translog production functions. As would be expected, the ME empirical reconstruction methods outlined in this chapter apply equally well to these production functions. For illustration, I show the production surface for all three production specifications for the Yolo model simplified to the two inputs of land and water.

This section showed that we can reconstruct fully flexible production models from minimal data sets, and that there is a continuum of models between the basic LP and fully flexible econometric specifications, that are able to reflect complex production properties with minimal data sets.

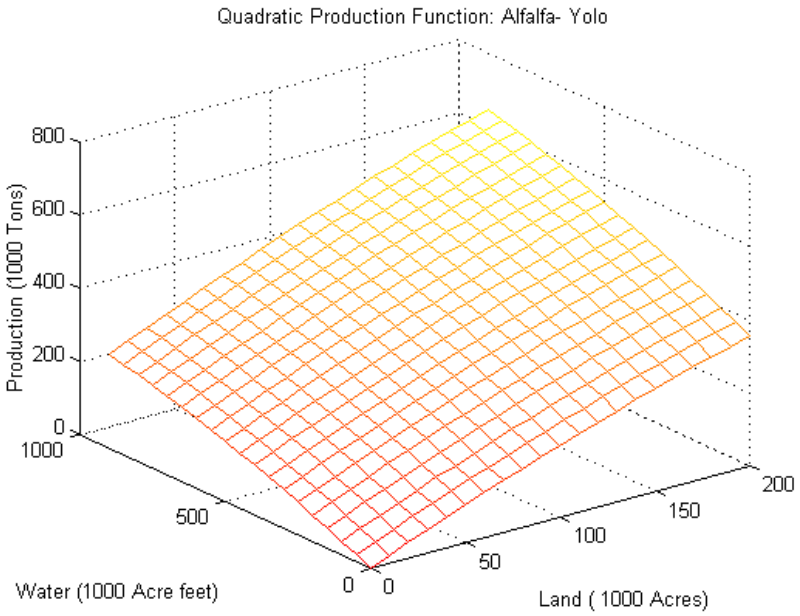


Figure 9.5:

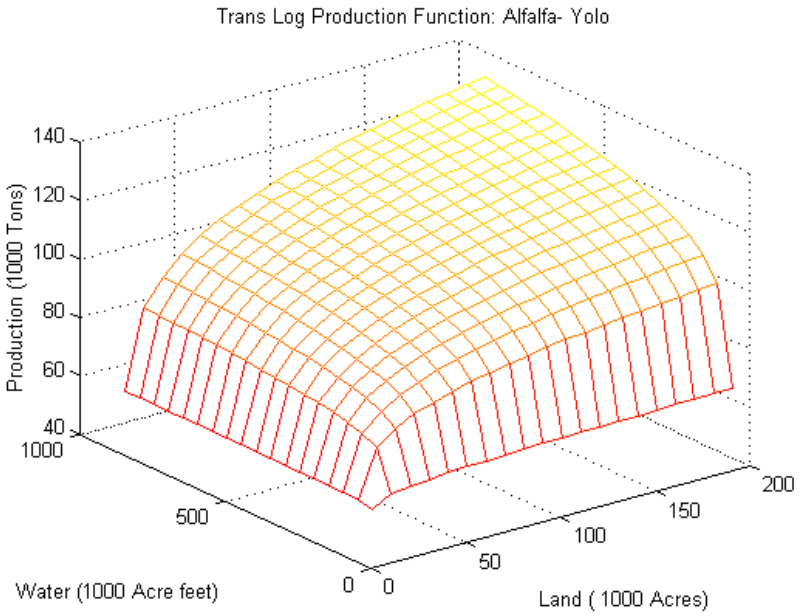


Figure 9.6:

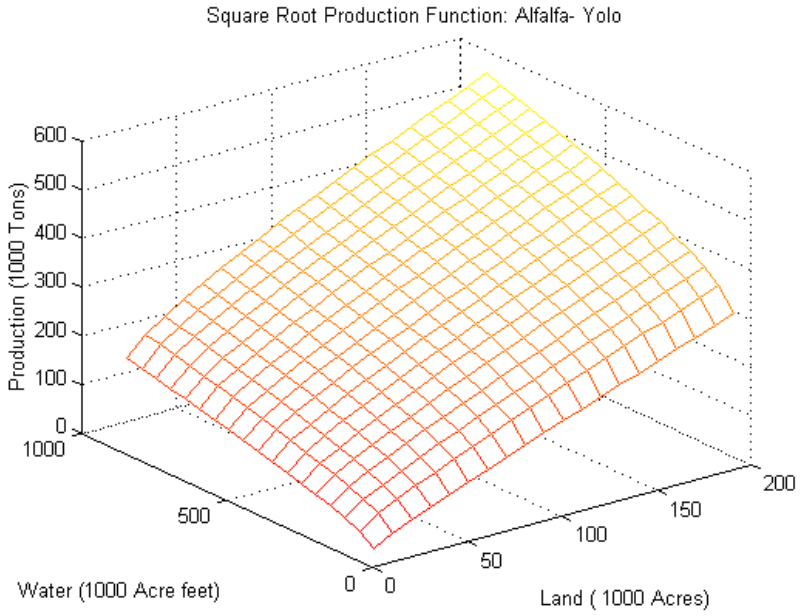

Square Root Production Function: Alfalfa- Yolo
Production (1000 Tons)
600
500
400
300
200
100
0
1000
500
0
Water (1000 Acre feet)
0
50
100
150
200
Land (1000 Acres)

Figure 9.7:

Chapter 10

Empirical Dynamic Policy Models

Using Polynomial Approximations to Solve Stochastic Dynamic Programming Problems: or A "Betty Crocker" Approach to SDP.[1] **Ed: Check dating on this wp.**

10.1 Introduction

Despite its tremendous methodological appeal in solving inter-temporal economic problems, Dynamic Programming (DP) has not been as widely used for the empirical analysis of natural resource problems as was initially expected. In the early sixties, Burt and Allison (1963) predicted that Dynamic Programming would shortly become a standard economic tool for agricultural and natural resource problems. Forty years later, however, there are still relatively few economic resource management studies published that use Deterministic or Stochastic Dynamic Programming (SDP) as an analytical tool. One reason may be that the computational methodologies for implementing these types of programming problems have not been widely accessible. SDP programs have mostly lain in the domain of Fortran specialists who have invested considerable time in crafting custom-written solution algorithms.

In this chapter we would like to demonstrate a methodology for solving deterministic DP or SDP problems in standard optimization software such as GAMS, with an efficient and robust computational algorithm that draws

[1] adopted from Howitt et al. (2002b).

upon approximation theory. The central philosophy of the approach is to minimize the computation stages that require discrete point evaluations. Instead, we use computational economic approximation methods Judd (1999) to fit continuous value functions,[2] and thus maximize the use of standard nonlinear solver routines, which have seen substantial advances over the past two decades. For example, we use the value iteration method to obtain the value function that solves the Bellman equation instead of the more popular policy iteration method, which is adopted mainly as an acceleration method (Judd, 1999; Bertsekas, 1976; Provencher and Burt, 1994). Miranda and Fackler advocate a hybrid approach that uses a continuous representation of the state variable while defining the control variables over a set of discrete values. They do this to avoid re-computation of the basis functions for the polynomial approximation to the value function and to maintain a linear solution procedure within the vector/matrix computing environment of MATLAB and GAUSS. In contrast, we specify both the state and control variables as continuous (but bounded) and use a non-linear solver in GAMS, which allows us to rapidly recompute the polynomial approximation terms at each iteration. The value iteration method converges rapidly and, in our case, enables us to specify both the state and the control variables as continuous, which allows us to examine a wider range of policies for natural resource problems.

While the individual components of the approach are not original in themselves,[3] the integration of these methods using a standard software package into a seemingly robust and general solution approach has value. We raise the analogy to Betty Crocker cake mixes, since predictable standard cakes mixes make a culinary contribution, and we think that the current state of SDP solution algorithms is similar to requiring that everyone bake a cake from scratch.

We feel that there is a significant contribution in making empirical economists aware of techniques that are available to solve this wide-ranging and important class of inter-temporal economic models. We are now able to claim that a large class of dynamic economic problems can be empirically solved by standard nonlinear optimization programs. The class of problems includes deterministic optimal control problems with finite or infinite horizons, and stochastic dynamic programming problems over finite or infinite horizons. In all cases, our approach is subject to the standard microeco-

[2] In contrast, Williams and Wright (1991) use a continuous approximation of the derivative of the value function.

[3] Provencher (1994) and Provencher and Bishop (1997) have used the same polynomial approximation to solve a stochastic resource problem.

nomic assumptions of:

- Value functions that are continuous in the controls and state variables.
- Value functions that are concave in the controls and state variables.
- A decision-maker who optimizes the sum of expected discounted values.

The remainder of the chapter is organized in the following way. In the next section, we describe the general value-iteration approach to solving an infinite-horizon DP problem, and then describe the particular polynomial approximation technique that we use. In the following section, we will show how well this approximation technique performs in comparison to a standard polynomial approximation, within the context of a well-known macro-economic growth model. We then extend the approximation methodology to solve finite time, stochastic dynamic problems. In the last section, we develop an empirical application of SDP to a simple natural resource management problem.

10.2 The Value Iteration Approach

In this approach, we seek a numerical approximation to the infinite horizon value function that maximizes the value of the problem resulting from decisions carried out in the future.

For a generic objective function $f(x_t, u_t)$ and an equation of motion for the state variable $x_{t+1} = g(x_t, u_t)$, we can write down the Bellman equation as:

$$V(x_t) = \max_u \{f(x_t, u_t) + \beta V(x_{t+1}) | x_{t+1} = g(x_t, u_t)\} \tag{10.1}$$

... and proceed to solve it forward, once we have an adequate representation of the value function, such that the above relationship will hold for each stage (time period) along the optimal path. In other words, we seek a polynomial approximation to an unknown function, such that the sum of the maximized intermediate function and the discounted "carry-over" value, evaluated by the function, maximize the total value function. This is essentially the idea of the mapping relationship given by $V^{s+1} = TV^s$, where T is a mapping that generates the sequence of approximations for $s = 0, 1, \ldots, S$, which converges to a stable value such that $V = TV$ holds. There are two fundamental propositions that underlie such a procedure, namely:

1. That any function can be approximated by a polynomial of sufficient order.

2. That such a function can be found within a finite number of iterations.

The proof for the second proposition is given by the contraction mapping theorem (applied in n stages by Stokey and Lucas with Prescott, 1989). [**Ed: need citation**]

The Bellman relationship in 10.1 can be written in a more compact form as:

$$V(x_t) = \max_{x_{t+1}} \left\{ f(x_t, g^{-1}(x_t, x_{t+1})) + \beta V(x_{t+1}) \right\} \tag{10.2}$$

... where we use the inverse of the equation of motion to express the control variable as a function of current and future values of the state variable.

The particular functional form that we've chosen for the polynomial approximation to the infinite-horizon is a Chebychev Polynomial, which belongs to a family of orthogonal polynomials described by Judd (1999) and implemented by Provencher and Bishop (1997) and Miranda and Fackler. We can think of the polynomial terms as analogous to a basis that spans the approximation space for the value function. The approximation takes the form:

$$V(x) = \sum_i a_i \phi_i M(x)$$

... where a_i is the coefficient of the i^{th} polynomial term $\phi_i M(x)$, which is defined over the interval given by the mapping $\hat{x} = M(x)$, which is [-1,1] in the case of the Chebychev polynomial. The terms of the Chebychev polynomial are sinusoidal in nature, and are given (for the n^{th} term) by the relationship $\phi_n(\hat{x})$, which is more easily enumerated by the following numerical recursion relationship for n polynomial terms:

$$\begin{aligned}
\phi_1(\hat{x}) &= 1 \\
\phi_2(\hat{x}) &= \hat{x} \\
\phi_3(\hat{x}) &= 2\hat{x}\phi_2(\hat{x}) - \phi_1(\hat{x}) \\
\vdots &= \vdots \\
\phi_n(\hat{x}) &= 2\hat{x}\phi_{n-1}(\hat{x}) - \phi_{n-2}(\hat{x})
\end{aligned}$$

Figure 10.1 shows five terms of the Chebychev polynomial terms over its domain where the orthogonal nature of Chebychev polynomials is quite clear. Notice that the function is defined on the [-1,1] interval.

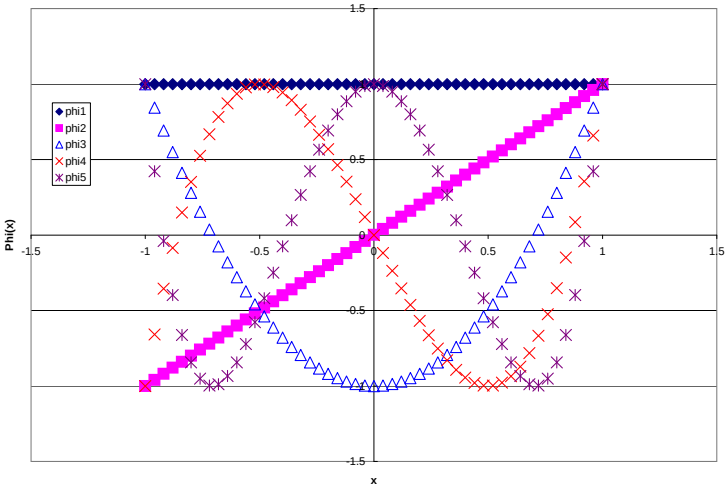


Figure 10.1: Graph of the Chebychev Polynomial Terms over its Domain

Following the discussion in Judd, the algorithm for obtaining the approximation to the value function $V(x)$ $(x \in [a, b])$ incorporates Chebychev regression, interpolation and scaling into one efficient procedure

1. The nodes at which the value function approximation will be evaluated are given by:

$$\hat{x}_k = \cos\left(\frac{2k-1}{2m}\pi\right) \in [-1,1] \text{for} k = 1, \ldots, m \text{where} m \geq n+1$$

2. Solve the Bellman Equation for each of the m interpolation nodes and save the maximized values given by:

$$V^{(j)}(x_k) = \max_{x_k^+} \left\{ f(x_k, g^{-1}(x_k, x_k^+)) + \beta V^{(j-1)}(x_k^+) \right\}$$

where x_k^+ is the value of the state variable in the next period resulting from the maximization at node k, and $V^{(j-1)}$ is the approximation of the value function from the previous iteration (initialized at iteration 0 by an initial guess). The mapping of $\hat{x}$ onto x $(\hat{x} \mapsto x)$ is given by

the relationship

$$\hat{x}_k = \left(\frac{2(x-a)}{b-a} \right) - 1$$

3. Update the polynomial coefficient values (for iteration j) by the regression:

$$a_i^{(j)} = \frac{\sum_{k=1}^{m} V^{(j)}(x_k)\phi_i(\hat{x}_k)}{\sum_{k=1}^{m} \phi_i(\hat{x}_k)\phi_i(\hat{x}_k)}$$

and obtain the updated value function

$$V^{(j)}(x) = \sum_i a_i^{(j)} \phi_i \left(\frac{2x - (a+b)}{b-a} \right)$$

for use in the next iteration with the Bellman equation.

This procedure is iterated until the polynomial coefficients are deemed sufficiently close for numerical convergence. In other words, the error sum of squares $\left\| a^{(j)} - a^{(j-1)} \right\|^2$ is evaluated for each iteration and is compared to a tolerance (ϵ) until the condition $\left\| a^{(j)} - a^{(j-1)} \right\|^2 < \epsilon$ holds, at which time the iterations terminate.

10.3 A Classical Growth Model Example

For our numerical example, we choose the neo-classical optimal growth model of Brock and Mirman (1972) **[Ed: need citation]**, which postulates fixed labor supply, Cobb-Douglas technology and logarithmic utility. The problem of the social planner is given by the infinite-horizon problem:

$$\max_{\{c_t, k_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \ln(c_t)$$
$$\text{subject to} c_t + k_{t+1} \leq A k_t^{\alpha}$$

which can be equivalently written as:

$$\max_{\{k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \ln(A k_t^{\alpha} - k_{t+1})$$

and restated in the recursive Bellman form as:

$$v(k_t) = \max_{k_{t+1}} \{\ln(Ak_t^{\alpha} - k_{t+1}) + \beta v(k_{t+1})\}$$

and solved for each period t.

Sargent (1987) shows that the infinite horizon value function is given (in the limit) as:

$$v(k) = \frac{\ln(A(1-\alpha\beta)) + \left(\frac{\alpha\beta}{1-\alpha\beta}\right)\ln(A\alpha\beta)}{1-\beta} + \frac{\alpha}{1-\alpha\beta}\ln(k)$$

which yields — when inserted into the Bellman Equation — the analytical steady state value for capital:

$$\tilde{k} = \left(\frac{1}{\alpha\beta A}\right)^{\frac{1}{\alpha-1}}$$

When we approximate the "true" value function with the Chebychev Polynomial given by successive contraction mapping according to: $Tv^{(j)}(k_t) = \max_{k_{t+1}} \left\{\ln(Ak_t^{\alpha} - k_{t+1}) + \beta v^{(j-1)}(k_{t+1})\right\}$, we obtain the results shown in figures 10.2 -10.4 :

Figure 10.2 shows how the Chebychev polynomial approximation converges on the "true" value function as the iterations increase. For comparison, we can see in Figure 10.3 how the Chebychev orthogonal polynomial approximation performs relative to a simple cubic polynomial approximation $\left(V(x) = \alpha_0 x + \alpha_1 x^2 + \alpha_3 x^3\right)$, by considering Figure 10.3 **Ed: check figure refs!**. It is clear that the Chebychev converges, while the cubic polynomial has a substantial deviation from the true value function.

The Chebychev function converged in 553 iterations (given a tolerance of 0.0001) compared to 751 for the cubic approximation. This demonstrates the superiority of the Chebychev algorithm, in both efficiency and accuracy of the approximation. As shown in Figure 10.4, the convergence is actually exponential.

10.4 Solving SDP Problems

The continuous value function approximation outlined in the previous section can be used to solve the larger class of more interesting stochastic finite horizon problems. In the previous section we assumed that the underlying value function is concave and continuous. In this section we add that

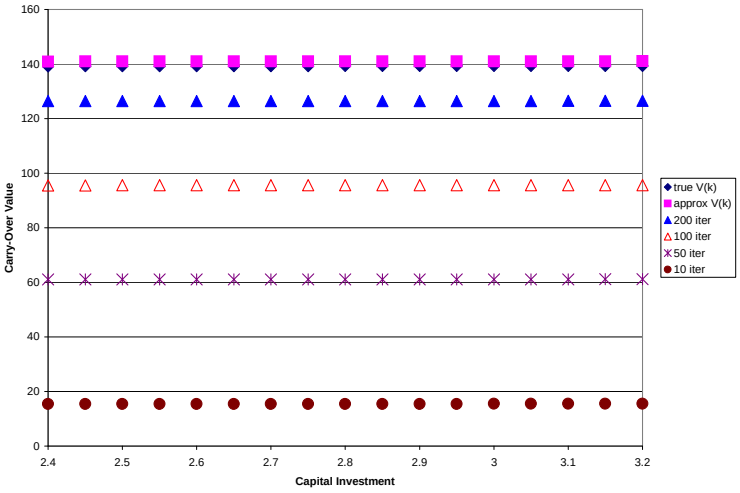


Figure 10.2: Comparison of the True Value Function and the Chebychev Approximation at Various Iteration Stages

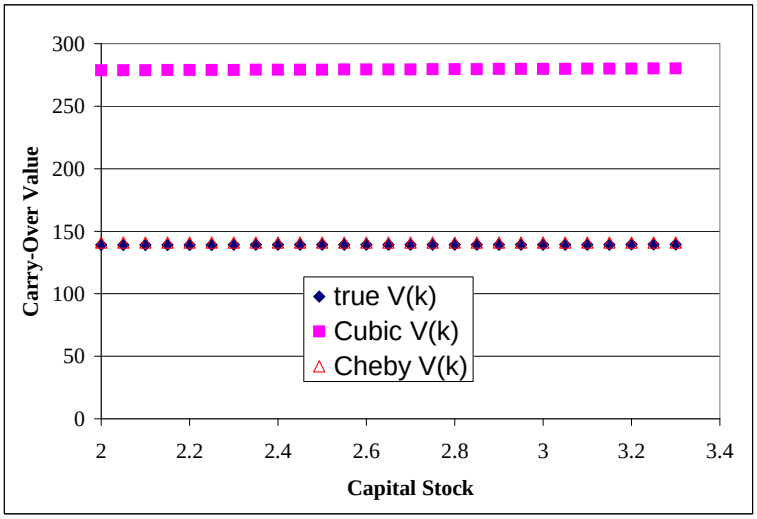


Figure 10.3: Comparison of the Chebychev and Cubic Approximations to the true Value Function

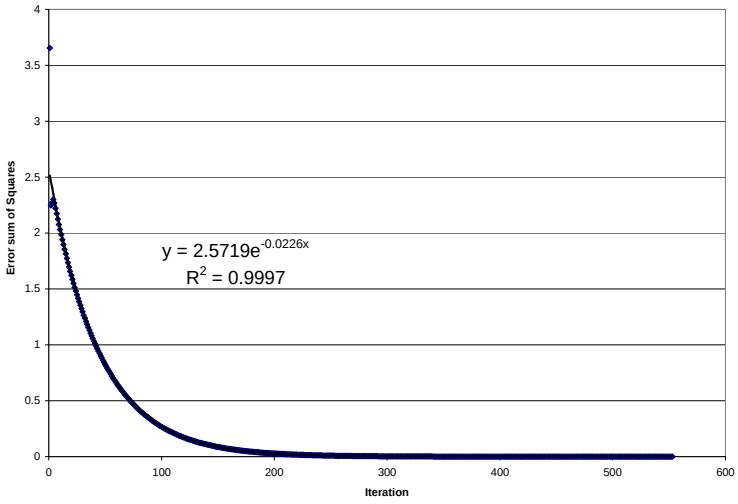


Figure 10.4: Convergence of the Chebychev Algorithm

the decision maker is interested in the expected outcomes of the stochastic decision problems. For ease of exposition, we shall only address a single state problem where the stochastic part of the problem is in the equation of motion.

10.4.1 The General SDP Formulation

Let X denote a vector of state variables, u a vector of control variables, and $\tilde{e}$ a vector of random events that influence the state variables, the objective function or both. Note that we slightly change matrix and vector notation here. The distribution of the stochastic vector is known, *a priori*. A general formulation of a finite SDP is:

$$\max_{u} V(u, X, \tilde{e}) \tag{10.3}$$

$$\begin{aligned} \text{subject to} \quad & X_{t+1} = g(X_t, u_t, \tilde{e} && \text{(10.4a)} \\ & u_t \in \Psi(X_t, \tilde{e}_t), \qquad X \in \Omega && \text{(10.4b)} \\ & \tilde{e} \text{is observable} \end{aligned}$$

The decision-maker's objective function is to find the optimal set of controls $\{u_1^*, \ldots, u_T^*\}$ that maximizes the objective 10.3 under a set of constraints. The objective function may be defined as the discounted sum of profits or utilities.

The equations of motion define the dynamic evolution of the resource. At each date, the level of a state variable is a function of the state variable level at the previous date, the control and the realized stochastic variables, Equation 10.4a. The problem is bounded by feasibility constraints. The set Ψ represents the feasibility constraints for controls given the level of the state variables and the stochastic variables. It may correspond to a set of physical or technical constraints. The set Ω characterizes the state variable feasibility constraints.

This general formulation of SDP encompasses most of the usual problems in the field of natural resource management. The field of natural resource management can be differentiated from conventional microeconomic theory by the degree of spatial or temporal market failure. Temporal market failure is rooted in the stochastic equations of motion for the relevant resource. Uncertainty in the evolution of an economic resource can be inherent in the biological dynamics or exogenously caused by weather, prices, or institutions. While SDP solutions are currently the only empirical approach to

solving resource problems characterized by discrete time stochastic dynamic equations of motion, the dominant use of SDP to date has been concentrated in engineering studies of optimal normative intertemporal water allocation.

We consider the simplest representation of a resource network based on a single state representation in a complex network system.[4] The dynamics of the system are given by:

$$X_{t+1} = X_t + \tilde{e}_{1t} - u_t \tag{10.5}$$

The change in natural resource stock must balance the stochastic change ($\tilde{e}_{1t}$) and the resource use (u_t). The index t in 10.5 denotes time period. The final demand for resource services is satisfied by resource flows from X_t, namely u_t.

We define the following timing of information and controls. First, the decision-maker observes the realization of the exogenous stochastic stock change variable $\tilde{e}_{1t}$ and hence X_t. Second, the decision-maker chooses the control u_t, the level of resource extraction or harvest.

The intermediate value of flow resources is defined by the inverse demand function: $P(u)$. The net surplus, $W(u)$, derived from resource consumption is denoted by:

$$W(u) = \int_0^q P(u)\,\mathrm{d}u \tag{10.6}$$

The net surplus of resource consumption is a concave increasing function of q.

The decision-maker wishes to maximize their utility subject to the equation of motion for the natural resource stock and the feasibility constraints. The resulting Bellman equation is:

$$\max_u V_t = \{W_t(u_t) + \beta E_{e_1}[V_{t+1}(X_{t+1})]\} \tag{10.7}$$

$$\begin{aligned} \text{subject to} \quad X_{t+1} &= X_t + \tilde{e}_{1t} - u_t && \text{(10.8a)} \\ X_{t+1} &\geq \underline{X} && \text{(10.8b)} \\ X_{t+1} &\leq \overline{X} && \text{(10.8c)} \\ u_t &\geq 0 && \text{(10.8d)} \end{aligned}$$

[4]For multi-state systems, the single state is decoupled from the network by approximating the network by two elements: a natural resource, with a stochastic change over time. $\tilde{e}_{1t}$ and stock X_t at each date t, and the resource flows from the rest of the system characterized by a stochastic variable $\tilde{e}_{2t}$. For an empirical example see Howitt et al. (2002a).

The stochastic control problem consists of choosing a sequence of decision rules for allocating resource flows that maximize the objective function (10.7) subject to (10.8a)-(10.8d). At each date, the current net surplus depends on the water releases. Consequently the objective function is the expectation of the current net surplus. All model parameters and functions are the same for all decision stages, implying that the problem is stationary. If the planning horizon is infinite, the optimal decision vector in state space for any decision stage is the same for all stages. The value of the system at any stage is likewise the same for all stages and is finite, even though the planning horizon is infinite, because the returns at all stages are discounted. The stochastic dynamic recursive equation that defines the optimal natural resource management problem is:

$$V_t(X_t, \tilde{e}_t) = \max_u \left\{ W_t(u_t) + \beta \left[\int V_{t+1}(X_{t+1}, e)\, \mathrm{d}\Phi_1 \right] \right\} \tag{10.9}$$

$$\text{subject to} \quad X_{t+1} = X_t + \tilde{e}_{1t} - u_t \tag{10.10a}$$

$$X_{t+1} \geq \underline{X} \tag{10.10b}$$

$$X_{t+1} \leq \overline{X} \tag{10.10c}$$

$$u_t \geq 0 \tag{10.10d}$$

where $V(\cdot)$ is the value function, u must be feasible and Φ_1 is the probability density function (pdf) for the random variable $\tilde{e}_1$.

Instead of the traditional methods of evaluating the value function for discrete points in the probability, control and state spaces, we make two approximations, the first on the value function and the second on the information accumulation by the decision-maker. We approximate the expected value function by assuming that it is a continuous function in state space. In addition, we assume that the decision-maker at any time regards the stochastic control problem as a closed loop problem based on the most recent information. The information contained in the observed level of the state variable is updated each time a stochastic realization is observed. Essentially, we assume that the state variable value function for the finite stochastic problem is adequately modeled for the next period by a function of the current state variable over an infinite horizon. Implicitly, this approximation assumes that the ability of the decision-maker to observe the updated level of the state variable in the future does not change the current optimal control given the current state of the system. These two assumptions enable us to solve the resulting SDP problems rapidly and reliably in a two-stage procedure.

Stage one, shown in Figure 10.5, [**Ed: if we make the boxes into a figure, it takes a page on its own. Let's discuss**], solves for the infinite time expected value function as a continuous function of the state variable. We use the value iteration method to solve for the second recursive term (the state variable value function) in equation 10.9 for a selected set of values of the state variable[5]. The value iteration method consists of assigning an arbitrary initial value for the value function, and then recursively solving the maximization problem until the decision rule converges to a constant function that is invariant over time. In our approach we also fit a continuous polynomial approximation to the value function over the set of selected state variable values at each iteration. It follows that the convergence of the value function in steady state also requires that the polynomial approximation converges.

Bellman has suggested that the curse of dimensionality could partially be overcome by approximating the value function as a polynomial on a relevant interval, (Bellman, 1961). This approach has been useful for researchers solving discrete choice dynamic programming problems, where the numerical challenges of multi-dimensional integration and approximation of the value function at alternative choice nodes have to be tackled simultaneously (Keane and Wolpin, 1994).[6] A second advantage of polynomial approximation is that it avoids the discrete approximations needed for a finite state Markov process. However, the major operational advantage of having a continuous function for the expected value function is that the forward solution becomes a simple nonlinear optimization problem (NLP). Assuming that the objective function (equation 10.9) is concave in the control variable, the NLP problem is sequentially solved in continuous, but constrained, state and control space. Clearly, the expected value function has to be approximated over a set of discrete points in state space. We use the Chebychev approximation of the value function to do this.

In stage two, shown in Figure 10.6, we start with the steady state values for the Chebychev parameters scaled over the relevant interval in state space from part one. The forward solution is initialized with the initial value for the state variable and any observed perturbations during the first time period. The first period optimal control results from the solution of a single period nonlinear optimization problem. The solution trades off the current short term returns against changes in the expected long term stock values

[5]We note that the derivative of the state value function at any time is the costate variable value.

[6]Although, in their case they are approximating and interpolating the maximized value of the value function over state-space, rather than the function itself.

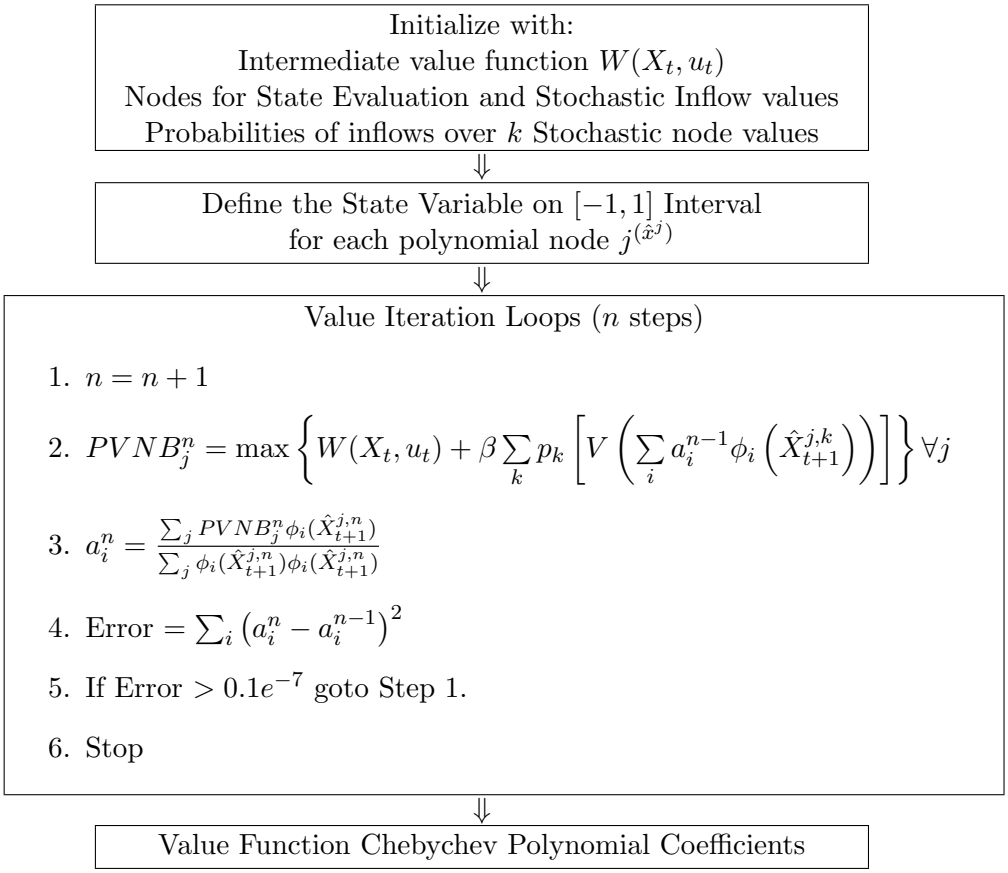


Figure 10.5: Solving for the Expected Value Function

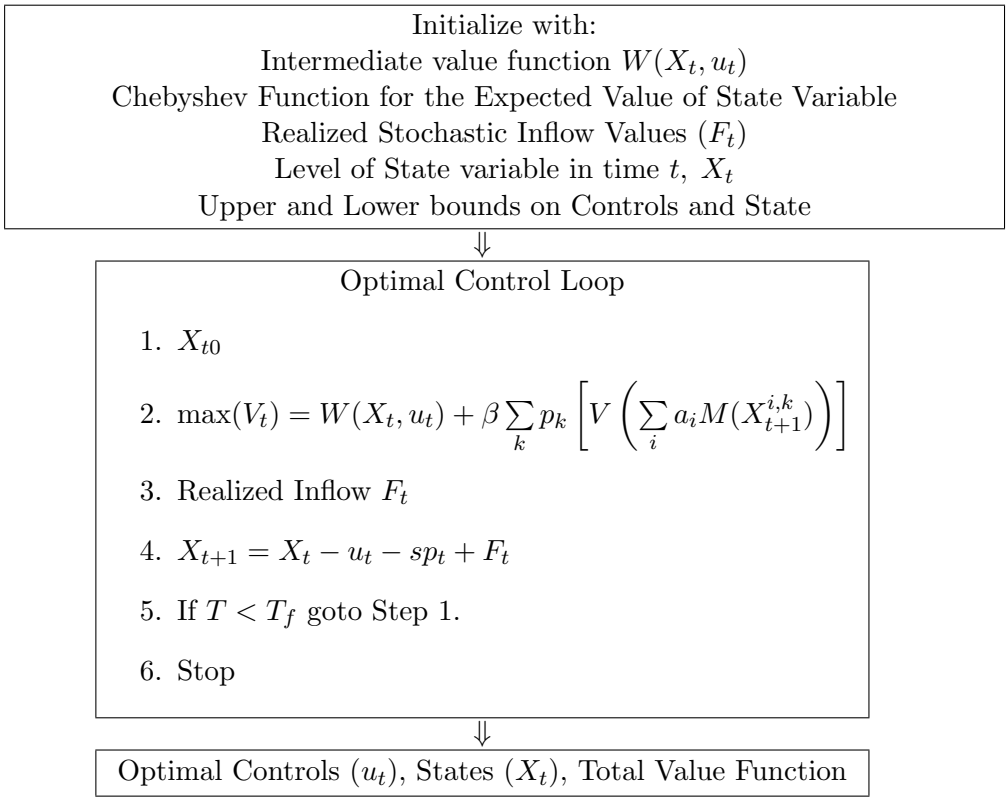


Figure 10.6: Solving for Optimal Stochastic Controls

implied by the control action and equations of motion. At the end of each time period, we assume that the decision maker observes the actual stochastic realization for that period, and updates the expected state value by the realized value. This updated state value provides the initial conditions for the NLP optimal control solution in the second period.

Given the continuous nature of the problem we can take advantage of modern NLP algorithms and standard packages. The problem is solved using Gams Conopt2, but any NLP procedure can solve the problem.[7]

[7]The program in Gams for this problem can be **obtained from the authors**. It is easily modified to solve other single state discrete time SDP problems.

10.5 An Empirical Application to Reservoir Management On the North Platte River

We illustrate the empirical solution of an SDP problem using a simplified representation of water management on the North Platte River in Nebraska. For ease of presentation we define the problem in its simplest terms and limit our analysis to the inter-year management problem, a more complex specification is used in Howitt et al. (2002a). The problem is to determine, from year to year, the optimal water releases and the optimal carryovers.

The North Platte Storage Project (NPSP) can be modeled as a single aggregated reservoir and a single release to the North Platte irrigators.[8] Accordingly, X_t is the stock of water in NPSP reservoirs, $\tilde{F}_{1t}$ are the stochastic levels of inflow to the reservoir, w_t are the water releases from the reservoir that produce water supply and hydropower benefits. The variation of the reservoir storage plus the stochastic inflow must be equal to the water release w_t, and the spills from the reservoir, sp_t. The spills balance the system in times of high flows, but have no economic value in the model.

We assume that yearly inflows $\tilde{F}_{1t}$ are independently and identically distributed (IID) with a log-normal distribution, i.e.:

$$\tilde{F}_{1t} \overset{iid}{\sim} LN(\mu, \sigma^2) \tag{10.11}$$

The aggregate inverse demand was generated using irrigation district data from Supalla (1999). A simple quadratic form is fitted to the resulting data yielding:

$$P(w) = 148.15 - 129.58w + 28.859w^2 \tag{10.12}$$

where w is the quantity of water in millions of acre-feet (MAF) and $P(\cdot)$ is the associated marginal value in millions of dollars per MAF.

The resulting net benefit function for water consumption is:

$$W(w) = 148.15w - 64.79w^2 + 9.6196w^3 \tag{10.13}$$

The net benefit function is increasing and concave with water consumption (as seen in Figure 10.7).

[8]The following eleven North Platte districts are aggregated to form the NPSP group: Pathfinder, Gering-Ft Laramie, Gering, Farmers, Northport, Central, Brown's Creek, Beerline, Chimney Rock, Lingle & Hill, and Goshen.

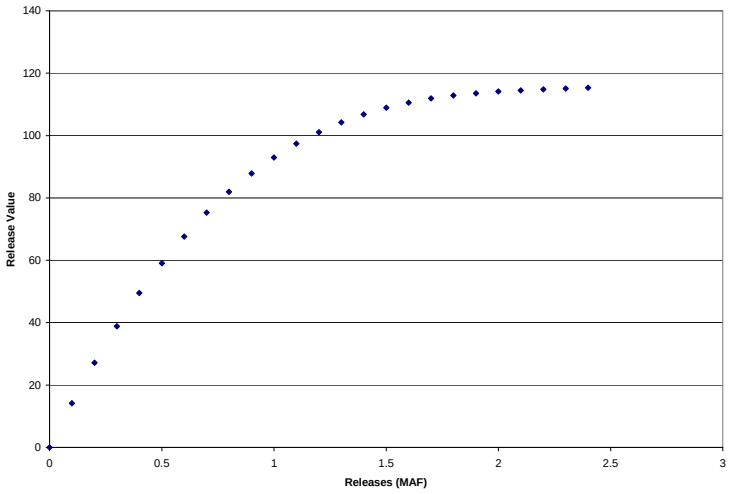


Figure 10.7: North Platte Water Release Benefit Function

The SDP Specification

The maximum storage capacity in the NPSP dams each year is 1.9 million acre-feet (MAF). We assume a minimum storage constraint equal to 0.7 MAF. Data on actual inflows, releases and storage are available for 1960 to 1992. A log-normal distribution was fitted to the observations and used to generate a set of eight discrete probabilities for the associated inflow quantities. The decision-maker is assumed to maximize the sum of the expected net present value of the water releases over this time period. The maximization is subject to the equation of motion for the reservoir stock and the feasibility constraints. The stochastic optimization program is:

$$\max_{w} U_t = \{W_t(w_t) + \beta E_{e_1}\left[V_{t+1}(X_{t+1})\right]\} \tag{10.14}$$

$$\begin{aligned} \text{subject to:} \quad X_{t+1} &= X_t + \tilde{F}_t - w_t - sp_t && \text{(10.15a)} \\ X_{t+1} &\geq 0.7 && \text{(10.15b)} \\ X_{t+1} &\leq 1.9 && \text{(10.15c)} \\ w_t &\geq 0 && \text{(10.15d)} \end{aligned}$$

The discount rate β is set at 0.943, corresponding to an interest rate of 6%.

10.5.1 Solving the model

The state variable (reservoir storage) is discretized in seven points from 0.7 MAF to 1.9 MAF. We consider a seven degree Chebychev polynomial approximation of the value function:

$$V_C(X) = \sum_{i=0}^{5} a_i T_i(\hat{X}) \text{ where } \hat{X} = M(X) \tag{10.16}$$

Where Chebychev polynomial coefficients a_i $(i = 1, \ldots, 7)$ are iteratively computed using the Chebychev regression algorithm, and $M(X)$ is a mapping of X onto the [-1,1] interval (Judd, 1999). The program converged in 240 iterations and took four minutes and forty seconds on a two gigahertz desktop machine. This result shows the substantial advantage of using a Chebychev approximation to approximate the state variable value function.

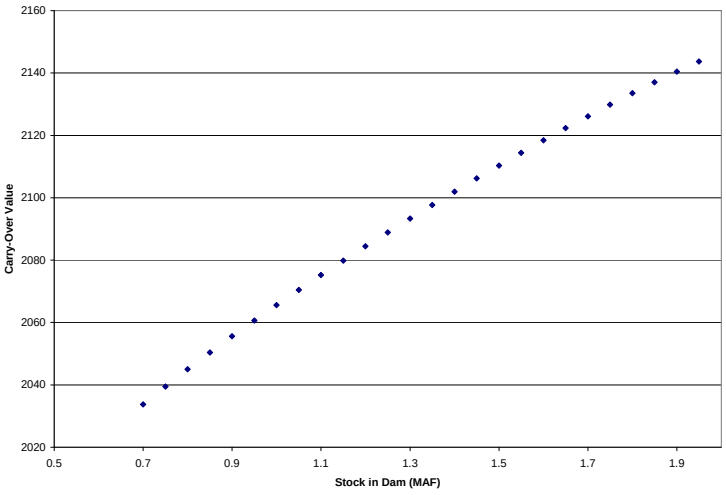


Figure 10.8: North Platte Storage Value Function

The criterion for convergence is that the sum of squared errors between polynomial coefficients at two consecutive iterations, must be smaller than $1.00E^{-7}$ **Ed: clarify scientific notation**: Table 10.1 shows the Chebychev polynomial coefficients **Ed: for what, over what?**.

1	2	3	4	5	6	7
2090.3675	53.15513	- 3.1195	0.17897	-0.1456	-0.01425	0.006614

Table 10.1: Chebychev polynomial coefficients

To evaluate the quality of fit of our SDP, we simulate the optimal predicted releases and storage for NPSP reservoirs over the historic time period, using the actual realized inflows to set the initial conditions for each year's optimization. As noted by Bertsekas (1976), the solution requires the solution of t sequential optimization problems by nonlinear optimization techniques.

Figure 10.9 presents the SDP policy simulations versus the observed ones for water release (the control) and storage (the state). We stress again that these predicted water storage and release levels result from the solution of a series of single year NLP problems. The information used by the model in any given year is only that which is available to the decision-maker, namely the current storage level, the current runoff from rain and snowfall, and the probability distribution of stochastic runoff in future years. The optimization at each stage balances the marginal value of current releases against the expected value of water stored and carried over for use in future years.

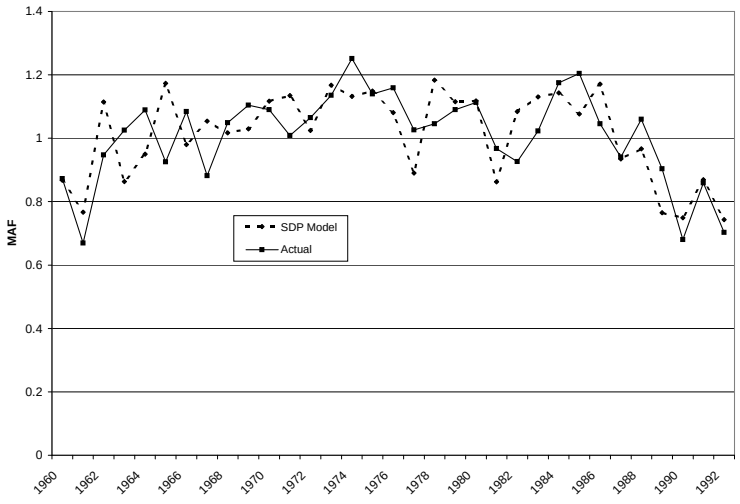


Figure 10.9: Simulation of North Platte Water Releases

Figure 10.9 shows that the annually optimized results for the SDP track

the actual releases fairly closely. It is worth noting that none of the release quantities are constrained by their upper or lower bounds, 1.5 and 0.1 MAF respectively. The resulting optimal releases are a result of trading off the marginal expected present value of storage against the immediate marginal returns from releasing more water.

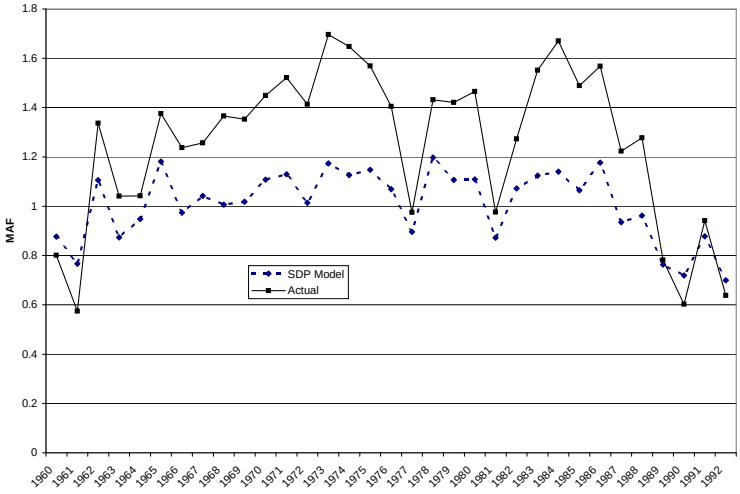


Figure 10.10: Simulation of North Platte Water Storage

The optimal storage results from the SDP solution in Figure 10.10 follow the same pattern as the actual storage quantities. However, the actual operators of the reservoirs were consistently more conservative, in that they stored larger amounts of water than the SDP solution. This is to be expected as the SDP is currently specified as risk neutral. However, it is more likely that public decision makers will be risk averse. Calibration of intertemporal preferences that improves the historical fit of the model using a recursive utility objective function can be found in Howitt et al. (2002a). Again, the optimal storage quantities in Figure 10.10 were obtained by a forward solution with no terminal constraints or penalty functions on the state variable. Neither of the bounds on the state variable bind during the time period simulated.

10.5.2 Using SDP as a policy tool

In this last section, we want to show how the SDP model can be easily used as a policy tool. Once the value iteration has solved for the Chebychev value

function coefficients, the forward solution of the SDP can be run swiftly and cheaply for a wide range of policy parameters. One cannot however change the stochastic properties of the inflows, the discount rate or the intermediate value function without resolving for the Chebychev coefficients. An additional advantage of this polynomial approximation approach for policy analysis is that both the controls (releases) and states (storage) are continuous functions of policy variables such as changes in storage for flood control reasons, or changes in the effective inflow due to environmental river flow regulations. The continuous functions are able to simulate subtle changes in policy that may prove difficult for the traditional discrete control sets.

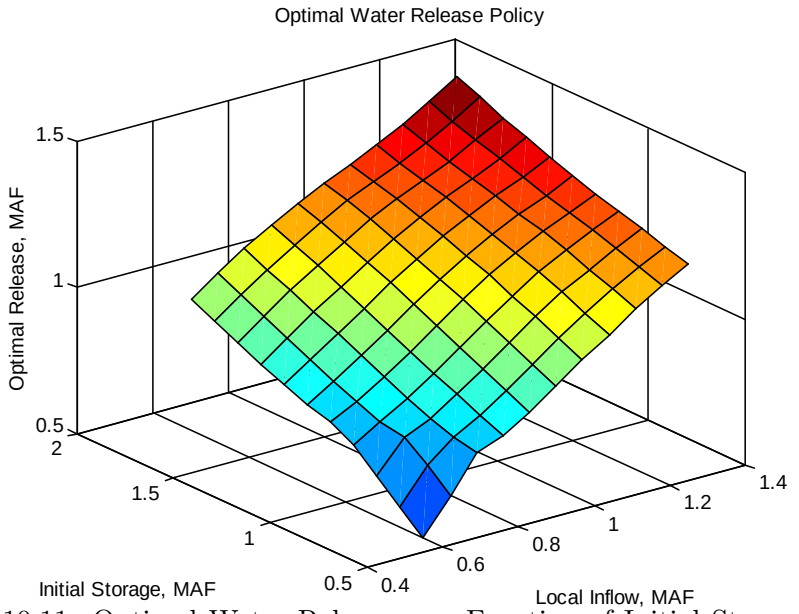


Figure 10.11: Optimal Water Release as a Function of Initial Storage and Realized Inflow

Figure 10.11 gives the optimal water release as a function of the two types of information that the decision maker has at the beginning of the year, namely the initial water storage and the local inflow realization. It is worth noting that due to the NLP formulation it took only 20 seconds to run 121 NLP solutions and generate the points on the surface of Figure 10.11.

In this risk-neutral SDP model, the level of initial storage in the reservoir has a similar effect on the optimal release as the level of inflow. Since the probability of spills from the reservoir is not a function of the storage. The shape of the surface shows the marginal effect on releases, and thus immediate revenues of changes in the initial storage quantity or realized

inflow.

10.6 Intertemporal Calibration and Estimation

The empirical problem in the previous section solves and simulates efficiently, but optimizes a simple net present value objective function that does not reproduce the observed dynamic behavior very closely as shown in figure ... A positive approach to dynamic modeling would lead us to hypothesize that there are additional parameters, or parameter values that may calibrate the dynamic results more closely to the observed decisions. The net present value objective function is separable over time, in that it only considers the trade off between actions in the current and future periods. It seems inappropriate to restrict the type of dynamic objective function to those forms that are not changed by the time profile of production or consumption. Thus we consider a more general recursive utility objective function specification which is "path dependent". That is to say, the current marginal rate of substitution between periods is a function of all past decisions. The general recursive specification ...

Continue with the development of the Howitt et al dynamic estimation paper

10.7 Conclusions

This chapter has been deliberately technical and computational. Our aim is to show precisely how easy it is to obtain empirical solutions to a wide range of dynamic problems that face researchers and policy analysts in agricultural and resource economics. We have aimed the chapter at a natural resources researcher who is mathematically literate, but has not invested in becoming proficient in the wide range of ingenious custom-tailored Fortran programs that have been used to solve SDP problems over the past forty years. **We hope that the GAMS code available from the author's web site will not be used as a "Black Box"**. Despite this misgiving, we want to restate that the only requirements to run the program are:

- A continuous intermediate value function in the controls.
- A set of discrete quantities that span the values of the stochastic component.
- A matching set of discrete probabilities for the stochastic values.

- An equation of motion and bounds on states and controls.

In addition, we want to emphasize the point of the emerging field of computational economics, namely, that like most "hard" sciences, there are classes of problems in economics whose solutions can only be meaningfully expressed by computational solutions. Stochastic dynamic programming is one such class of problems.

We realize that this chapter has not addressed many of the harder discrete choice, non-concave, and multi-state specifications that SDP researchers have addressed in the past, but given its introductory nature and space limitations, we hope that this chapter will stimulate additional work of this nature.

Ed: need to "nocite" the other refs to get them in the biblio...

Bibliography

Anderson, J. R. (1974). Sparse Data, Climatic Variability and Yield Uncertainty in Response Analysis. *American Journal of Agricultural Economics*, 55:77–82.

Antle, J. M. (1983). Sequential decision making in production models. *American Journal of Agricultural Economics*, 65(2):282–290.

Bauer, S. and Kasnacoglu, H. (1990). Nonlinear Programming Models for Sector Policy Analysis. *Economic Modelling*, pages 275–290.

Bellman, R. (1961). *Adaptive Control Processes: A Guided Tour*. Princeton University Press.

Bertsekas, D. (1976). *Dynamic Programming and Stochastic Control*. Academic Press, New York, NY (USA).

Box, G. E. P. (1979). *Robustness in Statistics*, chapter Robustness in the Strategy Of Scientific Model Building. Academic Press, New York, NY (USA).

Brooke, A., Kendrick, D., and Meeraus, A. (1992). *GAMS: A User's Guide*. The Scientific Press, San Francisco, CA (USA).

Burt, O. and Allison, J. (1963). Farm Management Decisions with Dynamic Programming. *Journal of Farm Economics*, 45(1):1–22.

Chambers, R. G. (1988). *Applied Production Analysis: A Dual Approach*. Cambridge University Press.

Day, R. H. (1961). *Agricultural Supply Functions*, chapter Recursive Programming and the Production of Supply. Iowa State University Press.

Debertin, D., Pagoulatos, A., and Aoun, A. (1990). Impacts of Technological Change on Factor Substitution Between Energy and Other Inputs with U.S. Agriculture, 1950-79. *Energy Economics*, 12:2–10.

Dervis, K., de Melo, J., and Robinson, S. (1982). *General Equilibrium Models for Development Policy*. Cambridge University Press, Cambridge.

DWR Staff (1974-96). State water project annual report of operations. Technical report, California Department of Water Resources.

Edwards, B., Howitt, R., and Flaim, S. (1996). Fuel, Crop, and Water Substitution in Irrigated Agriculture. *Resource and Energy Economics*, 18:311–31.

Fang, S. and Puthenpura, S. (1993). *Linear Optimization and Extensions: Theory and Algorithms*. Prentice Hall, Englewood Cliffs, NJ (USA).

Gill, P. E. and Wright, M. H. (1982). *Practical Optimization*. Academic Press.

Golan, A., Judge, G. G., and Miller, D. (1996). *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. John Wiley & Sons.

Greenberg, M. D. (1978). *Foundations of Applied Mathematics*. Prentice Hall.

Gregory, A. W. and Smith, G. W. (1990). Calibration as Estimation. *Econometric Reviews*, 9(1):57–89.

Gregory, A. W. and Smith, G. W. (1993). *Handbook of Statistics, Econometrics*, volume 11, chapter Statistical Aspects of Calibration in Macroeconomics, pages 707–719. North Holland, Amsterdam (NL).

Hatchett, S. A., Horner, G. L., and Howitt, R. E. (1991). *The Economics and Management of Water and Drainage in Agriculture*, chapter A Regional Mathematical Programming Model to Assess Drainage Control Policies, pages 465–489. Kluwer, Boston, MA (USA).

Hazell, P. B. R. and Norton, R. D. (1986). *Mathematical Programming for Economic Analysis in Agriculture*. MacMillan Co, New York, NY (USA).

Heckelei, T., Witzke, H., and Henrichsmeyer, W. (2001). *Agricultural Sector Modelling and Policy Information Systems*. Wissenschaftsverlag Vauk, Kiel, KG (Germany).

Hertel, T., Ball, V., Huand, K., and Tsigas, M. (1989). Computing Farm-level General Equilibrium Demand Elasticities for Agricultural Commodities. Bulletin 988, Agricultural Experimental Station, Purdue University, West Lafayette, IN (USA).

Hoover, K. D. (1995). Facts and Artifacts: Calibration and the Empirical Assessment of Real-Business-Cycle Models. *Oxford Economic Papers*, 47.

Horner, G. L., Corman, J., Howitt, R. E., Carter, C. A., and MacGregor, R. J. (1992). The canadian regional agricultural model: Structure, operation and development. Technical Report 1/92, Agriculture Canada, Ottawa, ON (CAN).

House, R. M. Usmp regional agricultural model. National Economics Division Report, address =.

Howitt, R., Reynaud, A., Msangi, S., and Knapp, K. (2002a). Calibrated Stochastic Dynamic Models For Resource Management. Working Paper, Department of Agricultural & Resource Economics, University of California, Davis.

Howitt, R., Reynaud, A., Msangi, S., and Knapp, K. (2002b). Using Polynomial Approximations to Solve Stochastic Dynamic Programming Problems: or A "Betty Crocker" Approach to SDP. Working Paper, Department of Agricultural & Resource Economics, University of California, Davis.

Howitt, R. E. (1995a). A Calibration Method for Agricultural Economic Production Models. *Journal of Agricultural Economics*, 46(2):147–159.

Howitt, R. E. (1995b). Positive Mathematical Programming. *American Journal of Agricultural Economics*, 77:329–342.

Judd, K. (1999). *Numerical Methods in Economics*. The MIT Press, Cambridge, MA (USA).

Just, R. (1993). Discovering Production and Supply Replationships: Present Status and Future Opportunities. *Review of Market and Agricultural Economics*, 61:11–40.

Keane, M. and Wolpin, K. (1994). The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence. *The Review of Economics and Statistics*, 76:648–672.

Konyar, K. and Howitt, R. E. (2000). The Cost of The Kyoto Protocol to U.S. Crop Production: Measuring Crop Price, Regional Acreage, Welfare and Input Substitution Effects. *American Journal of Agricultural Economics*, 25(2):347–367.

Leunberger, D. (1984). *Linear and Nonlinear Programming.* Addison-Wesley, Reading, MA (USA).

McCarl, B. A. (1982). Cropping Activities in Agricultural Sector Models: A Methodological Proposal. *American Journal of Agricultural Economics*, 64:768–771.

Meister, A. D., Chen, C. C., and Heady, E. O. (1978). *Quadratic Programming Models Applied to Agricultural Policies.* Iowa State University Press.

Miranda, M. and Fackler, P. *Computing in Economics and Finance '99*, chapter Hybrid methods for Continuous State Dynamic Programming.

Miranda, M. and Fackler, P. (2002). *Applied Computational Economics and Finance.* The MIT Press, Cambridge, MA (USA).

Mittelhammer, R. C., Judge, G. G., and Miller, D. J. (2000). *Econometric Foundations.* Cambridge University Press.

Paris, Q. (1979). Revenue and Cost Uncertainty, generalized Mean-Variance, and Linear Complementarity Problem. *American Journal of Agricultural Economics*, 61(2):268– 275.

Paris, Q. (1991). *An Economic Interpretation of Linear Programming.* Iowa State University Press, Ames, IA (USA).

Paris, Q. (1993). Lecture Notes in Mathematical Programming. Department of Agricultural Economics, University of California, Davis, California.

Paris, Q. and Howitt, R. E. (1998). An Analysis of Ill-Posed Production problems using Maximum Entropy. *American Journal of Agricultural Economics*, 80:124–138.

Peach, T. (1993). *Interpreting Ricardo.* Cambridge University Press, Cambridge (UK).

Provencher, B. (1994). Structural Estimation of the Stochastic Dynamic Decision Problems of Resource Users: An Application to the Timber Harvest

Decision. *Journal of Environmental Economics and Management*, 29:321–338.

Provencher, B. and Bishop, R. (1997). An Estimable Dynamic Model of recreation Behavior with an Application to Great Lakes Angling. *Journal of Environmental Economics and Management*, 33(2):107–127.

Provencher, B. and Burt, O. (1994). Approximating the Optimal Groundwater Pumping Policy in a Multiaquifer Stochastic Conjunctive Use Setting. *Water Resources Research*, 30:833–843.

Ray, S. (1982). A Translog Cost Function Analysis of U.S. Agriculture. *American Journal of Agricultural Economics*, 64:490–98.

Rendleman, M. (1993). Estimation of Aggregate U.S. Demands for Fertilizer, Pesticides and Other Inputs. Technical Bulletin 1813, Resources and Technology Division, USDA/ERS, Washington DC (USA).

Sargent, T. (1987). *Dynamic Macroeconomic Theory*. Harvard University Press, Cambridge, MA (USA).

Sato, K. (1967). A Two-Level Constant Elasticity of Substitution Production Function. *Review of Economic Studies*, 34:201–218.

Supalla, R. (1999). Personal communication.

Takayama, T. and Judge, G. G. (1964). Equilibrium among Spatially Separated Markets: A Reformulation. *Econometrica*, 32(4):510–524.

USBR Staff (1997). Central Valley Project Improvement Act. Draft Programmatic Environmental Impact Statement 8, US Bureau of Reclamation.

Williams, J. C. and Wright, B. (1991). *Storage and Commodity Markets*. Cambridge University Press.