Applications of

Mathematics in Economics

Warren Page
EDITOR

MAA
MATHEMATICAL ASSOCIATION OF AMERICA

MAA Notes #82

Applications of Mathematics in Economics

Library of Congress Control Number 2013942925

Print ISBN 978-0-88385-192-0

Electronic ISBN 978-1-61444-317-9

Printed in the United States of America

Current Printing (last digit):
10 9 8 7 6 5 4 3 2 1

Applications of Mathematics in Economics

Edited by

Warren Page

Professor Emeritus

New York City College of Technology

City University of New York

Published and Distributed by

The Mathematical Association of America

The MAA Notes Series, started in 1982, addresses a broad range of topics and themes of interest to all who are involved with undergraduate mathematics. The volumes in this series are readable, informative, and useful, and help the mathematical community keep up with developments of importance to mathematics.

MAA Notes

14. Mathematical Writing, by *Donald E. Knuth, Tracy Larrabee, and Paul M. Roberts.*
16. Using Writing to Teach Mathematics, *Andrew Sterrett,* Editor.
17. Priming the Calculus Pump: Innovations and Resources, Committee on Calculus Reform and the First Two Years, a subcomittee of the Committee on the Undergraduate Program in Mathematics, *Thomas W. Tucker,* Editor.
18. Models for Undergraduate Research in Mathematics, *Lester Senechal,* Editor.
19. Visualization in Teaching and Learning Mathematics, Committee on Computers in Mathematics Education, *Steve Cunningham and Walter S. Zimmermann,* Editors.
20. The Laboratory Approach to Teaching Calculus, *L. Carl Leinbach et al.,* Editors.
21. Perspectives on Contemporary Statistics, *David C. Hoaglin and David S. Moore,* Editors.
22. Heeding the Call for Change: Suggestions for Curricular Action, *Lynn A. Steen,* Editor.
24. Symbolic Computation in Undergraduate Mathematics Education, *Zaven A. Karian,* Editor.
25. The Concept of Function: Aspects of Epistemology and Pedagogy, *Guershon Harel and Ed Dubinsky,* Editors.
26. Statistics for the Twenty-First Century, *Florence and Sheldon Gordon,* Editors.
27. Resources for Calculus Collection, Volume 1: Learning by Discovery: A Lab Manual for Calculus, *Anita E. Solow,* Editor.
28. Resources for Calculus Collection, Volume 2: Calculus Problems for a New Century, *Robert Fraga,* Editor.
29. Resources for Calculus Collection, Volume 3: Applications of Calculus, *Philip Straffin,* Editor.
30. Resources for Calculus Collection, Volume 4: Problems for Student Investigation, *Michael B. Jackson and John R. Ramsay,* Editors.
31. Resources for Calculus Collection, Volume 5: Readings for Calculus, *Underwood Dudley,* Editor.
32. Essays in Humanistic Mathematics, *Alvin White,* Editor.
33. Research Issues in Undergraduate Mathematics Learning: Preliminary Analyses and Results, *James J. Kaput and Ed Dubinsky,* Editors.
34. In Evesò Circles, *Joby Milo Anthony,* Editor.
35. Youòre the Professor, What Next? Ideas and Resources for Preparing College Teachers, The Committee on Preparation for College Teaching, *Bettye Anne Case,* Editor.
36. Preparing for a New Calculus: Conference Proceedings, *Anita E. Solow,* Editor.
37. A Practical Guide to Cooperative Learning in Collegiate Mathematics, *Nancy L. Hagelgans, Barbara E. Reynolds, SDS, Keith Schwingendorf, Draga Vidakovic, Ed Dubinsky, Mazen Shahin, G. Joseph Wimbish, Jr.*
38. Models That Work: Case Studies in Effective Undergraduate Mathematics Programs, *Alan C. Tucker,* Editor.
39. Calculus: The Dynamics of Change, CUPM Subcommittee on Calculus Reform and the First Two Years, *A. Wayne Roberts,* Editor.
40. Vita Mathematica: Historical Research and Integration with Teaching, *Ronald Calinger,* Editor.
41. Geometry Turned On: Dynamic Software in Learning, Teaching, and Research, *James R. King and Doris Schattschneider,* Editors.
42. Resources for Teaching Linear Algebra, *David Carlson, Charles R. Johnson, David C. Lay, A. Duane Porter, Ann E. Watkins, William Watkins,* Editors.

43. Student Assessment in Calculus: A Report of the NSF Working Group on Assessment in Calculus, *Alan Schoenfeld,* Editor.
44. Readings in Cooperative Learning for Undergraduate Mathematics, *Ed Dubinsky, David Mathews, and Barbara E. Reynolds,* Editors.
45. Confronting the Core Curriculum: Considering Change in the Undergraduate Mathematics Major, *John A. Dossey,* Editor.
46. Women in Mathematics: Scaling the Heights, *Deborah Nolan,* Editor.
47. Exemplary Programs in Introductory College Mathematics: Innovative Programs Using Technology, *Susan Lenker,* Editor.
48. Writing in the Teaching and Learning of Mathematics, *John Meier and Thomas Rishel.*
49. Assessment Practices in Undergraduate Mathematics, *Bonnie Gold,* Editor.
50. Revolutions in Differential Equations: Exploring ODEs with Modern Technology, *Michael J. Kallaher,* Editor.
51. Using History to Teach Mathematics: An International Perspective, *Victor J. Katz,* Editor.
52. Teaching Statistics: Resources for Undergraduate Instructors, *Thomas L. Moore,* Editor.
53. Geometry at Work: Papers in Applied Geometry, *Catherine A. Gorini,* Editor.
54. Teaching First: A Guide for New Mathematicians, *Thomas W. Rishel*.
55. Cooperative Learning in Undergraduate Mathematics: Issues That Matter and Strategies That Work, *Elizabeth C. Rogers, Barbara E. Reynolds, Neil A. Davidson, and Anthony D. Thomas,* Editors.
56. Changing Calculus: A Report on Evaluation Efforts and National Impact from 1988 to 1998, *Susan L. Ganter*.
57. Learning to Teach and Teaching to Learn Mathematics: Resources for Professional Development, *Matthew Delong and Dale Winter.*
58. Fractals, Graphics, and Mathematics Education, Benoit Mandelbrot and Michael Frame, Editors.
59. Linear Algebra Gems: Assets for Undergraduate Mathematics, *David Carlson, Charles R. Johnson, David C. Lay, and A. Duane Porter,* Editors.
60. Innovations in Teaching Abstract Algebra, *Allen C. Hibbard and Ellen J. Maycock,* Editors.
61. Changing Core Mathematics, *Chris Arney and Donald Small,* Editors.
62. Achieving Quantitative Literacy: An Urgent Challenge for Higher Education, *Lynn Arthur Steen.*
64. Leading the Mathematical Sciences Department: A Resource for Chairs, *Tina H. Straley, Marcia P. Sward, and Jon W. Scott,* Editors.
65. Innovations in Teaching Statistics, *Joan B. Garfield,* Editor.
66. Mathematics in Service to the Community: Concepts and models for service-learning in the mathematical sciences, *Charles R. Hadlock,* Editor.
67. Innovative Approaches to Undergraduate Mathematics Courses Beyond Calculus, *Richard J. Maher,* Editor.
68. From Calculus to Computers: Using the last 200 years of mathematics history in the classroom, *Amy Shell-Gellasch and Dick Jardine,* Editors.
69. A Fresh Start for Collegiate Mathematics: Rethinking the Courses below Calculus, *Nancy Baxter Hastings,* Editor.
70. Current Practices in Quantitative Literacy, *Rick Gillman*, Editor.
71. War Stories from Applied Math: Undergraduate Consultancy Projects, *Robert Fraga*, Editor.
72. Hands On History: A Resource for Teaching Mathematics, *Amy Shell-Gellasch*, Editor.
73. Making the Connection: Research and Teaching in Undergraduate Mathematics Education, *Marilyn P. Carlson and Chris Rasmussen*, Editors.
74. Resources for Teaching Discrete Mathematics: Classroom Projects, History Modules, and Articles, *Brian Hopkins*, Editor.
75. The Moore Method: A Pathway to Learner-Centered Instruction, *Charles A. Coppin, W. Ted Mahavier, E. Lee May, and G. Edgar Parker*.
76. The Beauty of Fractals: Six Different Views, *Denny Gulick and Jon Scott*, Editors.
77. Mathematical Time Capsules: Historical Modules for the Mathematics Classroom, *Dick Jardine and Amy Shell-Gellasch*, Editors.
78. Recent Developments on Introducing a Historical Dimension in Mathematics Education, *Victor J. Katz and Costas Tzanakis*, Editors.
79. Teaching Mathematics with Classroom Voting: With and Without Clickers, *Kelly Cline and Holly Zullo*, Editors.
80. Resources for Preparing Middle School Mathematics Teachers, *Cheryl Beaver, Laurie Burton, Maria Fung, and Klay Kruczek*, Editors.
81. Undergraduate Mathematics for the Life Sciences: Models, Processes, and Directions, *Glenn Ledder, Jenna P. Carpenter, and Timothy D. Comar*, Editors.
82. Applications of Mathematics in Economics, *Warren Page*, Editor.

MAA Service Center
P.O. Box 91112
Washington, DC 20090-1112
1-800-331-1MAA FAX: 1-301-206-9789

Contents

Preface

An MAA NOTES volume that illustrates the uses of mathematics in economics would be a valued resource for mathematics faculty who want to enrich their undergraduate mathematics courses and better accommodate the needs of students interested in economics.

Applications of Mathematics in Economics presents an overview of the (qualitative and graphical) methods and perspectives of economists. Its objectives are not intended to teach economics, but rather to give mathematicians a sense of what mathematics is used at the undergraduate level in various parts of economics, and to provide students with the opportunities to apply their mathematics in relevant economics contexts.

The volume's applications span a broad range of mathematical topics and levels of sophistication. Each article consists of self-contained, stand-alone, expository sections whose problems illustrate what mathematics is used, and how, in that subdiscipline of economics. The problems are intended to be richer and more informative about economics than the economics exercises in most mathematics texts. Since each section is self-contained, instructors can readily use the economics background and worked-out solutions to tailor (simplify or embellish) a section's problems to their students' needs. Overall, the volume's 47 sections contain more than 100 multipart problems. Thus, instructors have ample material to select for classroom uses, homework assignments, and enrichment activities.

Although there are books available on mathematical economics, I know of no work that provides such an overview of economic subdisciplines or brings together so many different mathematics applications to such varied economics topics. I believe that mathematics instructors will be much more likely to use this volume's self-contained, stand-alone applications than to search through books on mathematical economics to find or develop their own such applications.

My hope is that in the near future, there will be an online MAA user forum where mathematics and economics faculty can collaborate, introduce new cross-disciplinary applications, and share teaching experiences. Such a forum would begin to address a second stage of issues that this volume does not attempt to address, such as how best to introduce these and other applications into mathematics courses, how to adjust the mathematics curriculum and approach to better serve economics students, and how to develop interdisciplinary programs.

Acknowlegements

The editor wishes to thank the following mathematicians and economists for their reviews and additional contributions.

David Allen, David Bressoud, James Case, William Dawes, David Garman, Raymond Greenwell, Don Hooley, Roger Horn, John Kenelly, Tanya Leise, Mary Lesser, Jane Lopus, Michael Murray, Roger Nelson, Chris Nevinson, Lynn Paringer, Kenneth Ross, Cecil Rousseau, Robert Rycroft, Karen Saxe, Anne & John Selden, Kathleen Shannon, Bruce Stephensen, Philip Straffin, Robert Tardiff, Alan Tucker, Frank Wang.

Notes on the Sections

Microeconomics (Mary H. Lesser and Warren Page)

Microeconomics is the study of a market's economy. The market for a good or service is said to be in *equilibrium* at a price when the quantity demanded equals the quantity supplied. This precalculus-based article provides an introductory overview of key notions in microeconomics. It shows how consumer behavior and firms' profit-optimizing decisions give rise to a market's demand and supply curves.

Section 2 (Supply and Demand) Algebraic and graphical representations of demand and supply functions are used to examine properties related to a market's equilibrium. The problems illustrate, perhaps surprisingly, what happens at equilibrium when a per-item tax is imposed on the supplier, and when the same tax is levied on the buyer. The notion of elasticity is used to explain why the fraction of the tax borne by the buyer is the same in each scenario, as is the fraction of the tax borne by the supplier.

Section 3 (Production Costs and Profit) and Section 4 (Consumer Behavior and Budget Constraints) show that supply and demand curves are outcomes of optimization processes. In Problem 3.3, for example, a firm's approximate supply curve is determined based on its labor cost, fixed and variable production costs, and profit. Section 4 considers consumers with M dollars to spend on a combination of x units of a good X and y units of a good Y. The most satisfying combination is determined by maximizing their utility $U(x, y) = xy$ subject to the budget constraint $xp_x + yp_y = M$, where p_x and p_y are the respective unit prices of X and Y. In Problem 4.3, the consumer's maximized utility is used to produce an approximate demand curve that, with the approximate supply curve of Problem 3.3, determines the market conditions at equilibrium for good X. Problem 4.4, an added option, requires the partial derivatives of $U(x, y) = \frac{3}{4}xy^4$.

Scenarios Involving Marginal Analysis (Julie Glass, Lynn Paringer, Jane Lopus)

The term "marginal" is used in economics to connote the change in a function due to a one-unit increase in its independent variable. For example, marginal cost means the change in total cost from producing one more unit, and marginal revenue is the change in total revenue from selling one more unit. Since a function's derivative at a value approximates the "marginal" at that value, economists use marginal analysis in situations where exact data is unavailable and where approximations suffice for making decisions. In this article, which requires single-variable calculus, the individual(s) in each scenario use marginal decision-making to solve problems in their economic undertakings.

Section 2 (Economic Industrial Espionage) Ingrid creatively uses analytic and graphical properties of marginal cost to sketch the graphs and determine the equations of a firm's variable cost function, total cost function, and average variable cost function. For another firm, she uses marginal cost to approximate the change in total cost, and then estimates the change in that firm's profit based on its expected marginal revenue.

Section 3 (Monopolistic Competitor) Physician Phyllis uses her demand function (number q of patients who pay p dollars for annual coverage) and total cost function to determine how many patients at what price will maximize her profit. Soon after, Phyllis uses her marginal revenue at maximum profit to decide if she should participate in a government program that pays her \$ 700 for each government patient she accepts. After establishing that she should participate, Phyllis needs to determine how many private and government patients to cover in order to calculate her new maximize profit.

Section 4 (Firm's Start Up) Sam and Stuart use a sampling method to estimate a demand function for their newly patented game. From their data, they create a marginal-cost function to estimate their variable-cost function and at what quantity their average variable cost is minimized. To better appreciate how demand changes with price, they use the price elasticity of demand to determine the effect on demand if they raise the price 3% per unit. Finally, Sam and Stuart explore the merits of increasing their fixed costs (advertising, rent) versus their anticipated corresponding increase in demand.

Intermediate Macroeconomics Theory (Michael K. Salemi)

This overarching survey shows what undergraduate mathematics is used in intermediate macroeconomics. The Assignments show the mathematics used in particular contexts, and the Advanced Projects illustrate additional mathematical skills that advanced students would employ The problems use algebra (mainly) through calculus, most of which can be done by students who have no knowledge of economics. But this doesn't mean that the problems are all easy.

Assignment 2.1 (Growth Accounting) concerns the construction and interpretation of macroeconomic data. Algebraic and graphical properties of exponential and logarithmic functions are used to compute and interpret compound growth rates for time-series data. The Advanced Project uses sigma notation and partial differentiation, in the method of least squares, to derive the linear regression equation for Gross Domestic Product data.

Assignment 2.2 (Production Functions and Labor Demand) introduces the Cobb-Douglas production function $Y = AK^{1-\alpha}L^{\alpha}(A > 0$ and $0 < \alpha < 1)$ that models a firm's output Y based on its capital cost K and labor cost L. Single-variable calculus is required to explore issues such as what quantity of labor at given wages that a profit-maximizing firm would employ. The Advanced Project repeats the Assignment for the production function $Y = A[\alpha L^{-\beta} + (1-\alpha)K^{-\beta}]^{-1/\beta}$, where $0 < \alpha < 1$ and $-1 < \beta \neq 0$.

Assignment 2.3 (Solow Growth Model) considers the long-run growth trajectory of a developed economy. Algebra only is needed to investigate a system of seven equations and four parameters that describes an economy. After expressing the state of the economy by a single equation, the growth rates are determined for capital, labor, output, and consumption. The Advanced Project uses a spreadsheet to exploit the recursive nature of the model. For given initial parameter values, one examines how the capital-labor relation, consumption per worker, and output per worker evolve from arbitrary initial conditions to their equilibrium values.

Assignment 2.4 (Short-run Model) examines the short-term deviations of an economy's variables (X) from their steady-state (long-run) values (X_S). The short-term model consists of two sectors: the IS sector determines balance between investment and savings, and the LM sector determines balance between the money supply and money demand. The IS sector's equations are combined into a single equation (called the IS schedule) that expresses output Y as a decreasing linear function of interest rate r. The LM equation $M = PLY^{e}r^{-f}$ $(L, e, f > 0)$ needs to be linearized, and $M - M_s = \frac{\partial M}{\partial P}(P - P_S) + \frac{\partial M}{\partial Y}(Y - Y_S) + \frac{\partial M}{\partial r}(r - r_S)$ recast as an equation (termed the LM schedule) that expresses r as an increasing linear function of Y. Issues considered include by what amount the interest rate must rise to offset the increase in money demand caused by a one percent increase in price level. The Advanced Project involves solving the IS and LM schedules for the model's equilibrium solution (Y, r), and predicting the changes in Y and r that result from changes in other variables. Everything, other than linearizing the LM equation, can be done using algebra. Presenting the LM's equation in its linearized form would render algebra sufficient for all parts of the Assignment and Advanced Project.

Assignment 2.5 (Business Cycle Model) focuses on monetary policy. The model, a system of three equations in six variables and three parameters, examines what interest rate r_t the Federal Reserve should choose in order to maintain the initial long-term equilibrium value of output Y_t or inflation π_t (or both) in the face of a change in spending or in inflation. In the Advanced Project, the Federal Reserve is assumed to follow a weighted policy $r_t = \theta\pi_t + \phi Y_t$. The Advanced Project also describes how to set up a spreadsheet and explore, for combinations of θ and ϕ, the long-run values of Y_t and π_t due to changes in spending or inflation. Everything in the Assignment and Advanced Project can be done using algebra.

Assignment 2.6 (Monetary Policy game) compares the effects on the economy's output and inflation when the Federal Reserve conducts monetary policy by using discretion (tailored to each situation) and when it is governed by a policy rule (applicable to every situation). Each scenario is treated as a two-person game between the private sector and Federal Reserve, whose objective function describes its level of preference for output and inflation. The Advanced Project repeats the Assignment's analysis for a new objective function. Everything can be done using algebra; where calculus is used to optimize quadratic functions, this can be done algebraically by determining a parabola's vertex. Although the interpretation and analysis of a few calculations would be beyond those with no knowledge of economics, explanations are included in the Solutions section.

Closed Linear Systems (Warren Page and Alan Parks)

An economy governed by linear equations is said to be *linear*. A *closed* economy is one that is self-contained in some sense – as, for example, when production equals consumption, or expenditures equal income. This article describes three closed, linear economies and the manner by which each attains equilibrium. Matrix operations and systems of linear equations in matrix form suffice to understand the text and handle the problems other than Problems 2.7–2.9, whose guided exercises lead to proofs of the theorems on which the text is based. These exercises require knowledge of the Monotone Convergence theorem (a monotone series converges if and only if it is bounded) and the meaning of a convergent infinite series.

Section 2 (Production Adjustment and Price Adjustment Models) The production adjustment model describes how production adjusts to satisfy consumption. The price adjustment model describes how price adjusts to balance income and the cost of labor. The models are combined to establish the *equilibrium principle*: the total spent on consumption equals the total spent on labor.

Section 3 (Normalized Leontief Model) The production adjustment and price adjustment models are incorporated into a single model by treating labor as a good whose total consumption and price are its usage and cost. The model's units are adjusted so that each good's total one-unit production is consumed. The investigation also leads to properties of probability vectors and Markov matrices.

Mathematics in Behavioral Economics (Michael Murray)

Section 2 (Fairness) uses a two-player game to illustrate how attitudes about fairness influence economic behavior, and to challenge standard economic theory's assumption that individuals care only about their benefits. Both players know the game's rules and that they will play only once with their unknown partner. They will divide a fixed amount of money if the second player accepts the fraction of that money offered by the first player. Algebra only is needed to discover what offer the second player would accept, and what the first player would offer knowing what the second player would accept.

Section 3 (Probabilities) Standard economic theory assumes rational behavior and the incorporation of new information in accord with the laws of mathematical probability. However, cognitive scientists have demonstrated that our brains' natural responses to new information are at odds with Bayes' rule. This section discusses systematic ways that people's probability assessments violate the mathematical rules of probability and their possible economic consequences.

Section 4 (Decisions Under Uncertainty) contrasts two models for explaining choices made under uncertainty. Expected utility theory used in standard economics assumes individuals make choices that will maximize the mathematical expectation of their utility functions (which represent preferences among outcomes). Since people's choices are often influenced by their risk-averse perception of immediate circumstances, rather than by maximizing expected well being, behavioral economists describe behavior by prospect theory, which uses a value function whose domain is the losses and gains of outcomes facing an individual. Everything in this section can be explored using algebra and the notion of mathematical expectation $E(z) = (1 - p)z_1 + pz_2$.

Section 5 (Inter-temporal Discounting) Two methods are compared of formalizing how people discount future costs and benefits relative to immediate costs and benefits. Economists using exponential discounting, based on a constant rate of discounting, assume inter-temporal consistence: decision makers will choose the same alternative for the future that they originally chose. Behavioral economists use hyperbolic discounting, which encompasses exponential discounting by including a parameter that characterizes people's observed tendency to be more present biased and cost averse. The problems involving hyperbolic discounting examine whether people will make or regret making the same choices for the future that they made earlier.

The introduction to exponential discounting shows that $dW(t)/dt = rW(t)$ yields $W(t) = W_0 e^{-rt}$, and uses $\int_0^\infty f(t)e^{-rt}\,dt$ to represent the present discounted value of lifetime benefit of $f(t)$. Nevertheless, everything else in Section 5 other than Problem 5.3, which requires simple differentiation, can be done algebraically using exponential functions and finite geometric series.

Econometrics (Ray Jean B. Goodman)

This article considers aspects of linear regression and its applications in econometrics. Section 2 (Linear Regression) uses the method of least squares to determine estimators b_0 and b_1 of β_0 and β_1 in a simple two-variable model. Problem 2.1 requires sigma notation and differentiation to minimize $L = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$ with respect to b_0 and b_1. Problem 2.2 asks to represents the system of n equations $e_i = y_i - b_0 - b_1 x_i$ in matrix form $\mathbf{e} = \mathbf{y} - X\mathbf{b}$, where $\mathbf{e}$, $\mathbf{y}$, $\mathbf{b}$ are column vectors and X is an $n \times 2$ matrix. Then $L = \mathbf{e^t e}$, and setting the vector derivative $\partial L/\partial \mathbf{b}$ equal to the column vector of zeros yields $X^t\mathbf{y} = X^t X\mathbf{b}$. Thus, $\mathbf{b} = (XX^t)^{-1}X^t\mathbf{y}$ for conditions on X under which matrix $X^t X$ is nonsingular.

Section 3 (Multiple Linear Regression) considers linear regression for k independent variables. Problem 3.1 asks to represent the system of n equations $e_i = y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \cdots - b_k x_{ik}$ in matrix form and show that $L = \mathbf{y^t y} - 2\mathbf{b^t}X^t\mathbf{y} + \mathbf{b^t}X^t X\mathbf{b}$.

Section 4 (Elasticity and Functional Forms) The elasticity of a function $y = f(x)$ may be interpreted as the % change in y due to a 1% increase in x. Problem 4.1 requires calculus to show that elasticity can be approximated by $xf'(x)/f(x)$ and expressed as $d(ln\, f(x))/d(ln\, x)$. The section's other problems use only logarithmic and semilogarithmic transformations of functions and linear regression on their transformed data. They investigate relevant economic questions such as the elasticity of output with respect to labor, the compound growth rates of the M1 and M2 money supply, and at what rate the labor force has grown.

Section 5 (Applications in an Econometrics Course) This section of worked-out problems illustrates graphical and statistical applications in an econometrics course, including analyses related to assessing statistical significance of a regression equation's estimators and the equation as a whole.

The Portfolio Problem (Kevin J. Hastings)

Financial economics, in large part, studies the interrelationships of financial variables such as the fair market price of an asset based on its risk, and the balance of a portfolio's risk relative to its expected return.

Section 2 (Average Rate of Return and Risk) uses the daily prices of four stocks to introduce the article's needed concepts. The sample mean and variance of the ith stock's prices are respective estimates for the stock's theoretic expected return (mean μ_i) and risk (variance σ_i^2). If w_i is the fraction of money invested in the ith asset, the portfolio's expected return μ and risk σ^2 are w_i-weighted sums of its individual assets' risk-return measures. The problems consider the maximum rate of return for a three-stock portfolio, and compare the minimized variances of a two-stock portfolio when the stocks' daily returns are independent and when their dependence is characterized by a given correlation coefficient. The need to minimize σ^2, expressed as a quadratic function of w_1, can be done using calculus or by algebraically determining the vertex of a parabola.

Section 3 (Solving the Optimization Problem) The objective is to maximizes a kind of portfolio net value $\mu - a\sigma^2$, where the constant $a \geq 0$ characterizes risk aversion. For a portfolio of four assets, Problem 3.1 uses the Lagrangian to maximize $f(w_1, w_2, w_3, w_4) = \sum_{i=1}^4 w_i \mu_i - a\sum_{i=1}^4 w_i^2 \sigma_i^2$ subject to the constraint $g(w_1, w_2, w_3, w_4) = w_1 + w_2 + w_3 + w_4 - 1 = 0$. For the solution's optimally determined weights, $\sum_{i=1}^4 w_i \mu_i$ is called the *portfolio's risk-averse expected return*. The risk-averse expected returns for the other portfolio problems can be obtained by substituting the given data into the expressions for the weights obtained in Problem 3.1's solution.

Section 4 (The Portfolio Separation Theorem) Problem 3.1's solution is used to establish the Portfolio Separation theorem: "For a portfolio with a risk-free asset, the ratios wi/wj of the optimally determined weights are independent of the investor's risk aversion." Problem 3.1's solution can be used to determine if the theorem holds for a portfolio of three stocks and a risk-free asset (e.g., a bond). However, in Problem 4.3, the Lagrangian with two multipliers is needed to determine if the theorem holds for a portfolio of four stocks and a risk-free asset.

Topics in Modern Finance (Frank Wang)

Section 2 (Statistics Background) The annual closing prices of IBM and HP stocks are used to introduce the needed statistical concepts: the mean, variance, covariance, and correlation coefficient. The problems require only numerical computation.

Section 3 (Modern Portfolio Theory) An asset's returns are used to characterize its risk (measured by the variance) and expected return (measured by the mean). The risk and expected return of a portfolio of assets are weighted sums of its individual assets' risk-return measures. Section 3.1 explores a two-asset portfolio and begins to hint at the rational for diversification. The problems, which require minimizing quadratic functions of x, can be done using calculus or by algebraically determining the vertex of a parabola. Thus, Section 2 may be combined with Section 3.1 as a richer exploration for students without calculus. Section 3.2 requires partial derivatives and Lagrange's method to explore a three-asset portfolio, and Section 3.4 requires matrix algebra for dealing with multi-asset portfolios.

Section 4 (The Capital Asset Pricing Model) is used to determine the expected return of an asset based on how risky it is. In Problem 4.1, algebra suffices to derive the Capital Market Line equation that expresses a portfolio's expected return $\mu = r + m\sigma$, where r is the rate of a risk-free asset (e.g., bond) and m is the Market's (expected return – r) per unit of Market risk. Problem 4.2 requires single-variable calculus to derive the Security Market Line equation that expresses a stock's expected return as a positive linear function of β, a measure of the stock's price covariate with the market.

Section 5 (The Black-Scholes Formula) Knowledge of mathematical probability is required to determine the exact value of a risky asset called an "option" (a contract for the right to buy or sell a stock at a fixed price within a specified time interval). In the problems, computer-simulated trajectories of a stock's prices over the year's 254 trading days are used to estimate the stock's future price. The current value of a stock's option obtained from simulation is compared with that from the Black-Scholes formula. The Black-Scholes formula also is used to estimate the implied volatility of a stock based on its option's market price.

Section 6 (Afterwords) examines some of assumptions underlying standard financial models. The section's problem guides the reader to calculate the z-score of the daily change of the Down Jones Industrial Average to illustrate why the normal distribution is inadequate to characterize the behavior of the market. And it shows the assumption that decision makers act rationally needs to be modified in accordance with recent findings by psychologists and behavioral economists. (See, for example, Sections 3 and 5 in Michael Murrays article Mathematics of Behavioral Economics.)

Maximizing Profit with Production Constraints (Jennifer Wilson)

Mathematical economics is the subdiscipline of economics that includes the formulation and derivation of mathematical methods for analyzing and solving problems in economics. Aspects of this are illustrated in Wilson's article that, using multivariable calculus and matrices, discusses mathematical principles and methods for optimizing functions with constraints.

Section 2 (Production Functions) Partial derivatives are used to explore production functions and economically interpret some of their key properties.

Section 3 (Unconstrained Optimization) A function's matrix of second derivatives is used to investigate the relationship between the function's concavity, the existence of a unique extremum, and the second-derivative test.

Section 4 (Constrained Optimization) uses the method of Lagrange multipliers to optimize functions with one or two constraints, and shows how to modify the second-derivative test for such problems. It also discusses the economic interpretation of the Lagrange multiplier at optimal production as the approximate amount of production increase for each additional unit in the total budget.

Section 5 (Optimization with Inequality Constraints) discusses two theorems that give necessary and sufficient conditions for optimization. They are illustrated in an example whose function of three variables is subject to three inequality constraints.

At first glance, this article may appear intimidating because of the densely displayed symbols and equations. This is due in part to messy computations associated with the functions in economics, and because all derivations and examples are worked out for the reader. The sections, therefore, are straightforward to read. And their problems are accessible to students who have the aforementioned background. Although the analysis is carried out in generality, the problems can be made simpler and more user-friendly by substituting specific values for parameters. Computer-algebra software also can be used to handle some of the more involved calculations. Using such technology, instructors can generate a variety of related problems at different levels of sophistication, and students can explore more realistic optimization problems in economics.

Problems and Subject Areas

Problem s.n denotes the nth problem of Sections, and APs.n denotes the Advanced Project associated with Asssignment s.n, in the referenced articles. Problems that can be explored in more than one area of mathematics are listed in each subject area.

Precalculus

Microeconomics 2.1, 2.2, 2.3, 2.4, 2.5, 3.1, 3.2, 3.3, 3.4, 4.1, 4.2, 4.3
Intermediate Macroeconomics Theory 2.1, 2.3, AP2.3, 2.5, AP2.5, 2.6, AP2.6
Behavioral Economics 2.1, 2.2, 2.3, 2.4, 4.1, 4.2, 4.3, 4.4, 4.5, 5.1, 5.2, 5.4, 5.5, 5.6
The Portfolio Problem 2.1, 2.2, 2.3, 3.2, 3.3, 3.4, 3.5, 4.1, 4.2
Topics in Modern Finance 2.1, 2.2, 3.1, 3.2, 3.3, 4.1

Single-variable Calculus

Microeconomics 4.1, 4.2, 4.3
Intermediate Macroeconomics Theory 2.2, AP2.2, 2.6, AP2.6
Econometrics 4.1
The Portfolio Problem 2.3
Topics in Modern Finance 3.1, 3.2, 3.3, 4.2

Multivariable Calculus

Microeconomics 4.4
Intermediate Macroeconomics Theory A2.1, 2.4, AP2.4
Econometrics 2.1
The Portfolio Problem 3.1, 4.3
Topics in Modern Finance 3.4
Maximizing Profit with Production Constraints 2.1, 2.2, 2.3, 3.1, 3.2, 3.3, 3.4, 4.1, 4.2, 4.3, 5.1, 5.2

Linear Algebra

Closed Linear Systems 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.1, 3.2
Econometrics 2.2, 3.1
Topics in Modern Finance 3.5

Probability

Behavioral Economics 3.1, 3.2, 4.1, 4.2, 4.3, 4.4, 4.5

Statistics

Econometrics 4.2, 4.3, 4.4, 5.1, 5.2, 5.3
The Portfolio Problem 2.1, 2.2, 2.3
Topics in Modern Finance 2.1, 2.2, 3.1, 3.2, 3.3, 5.1, 6.1

1

Microeconomics

Mary H. Lesser
Charles M. Snipes School of Business, Lenoir-Rhyne University, Chapel Hill

Warren Page
Professor Emeritus, New York City College of Technology, CUNY

1 Introduction

Microeconomics is the study of decisions made by individual economic units. It is often called “price theory” because the emphasis is on how prices bring decisions made by individuals (consumers or producers) into balance. Marginal analysis is key in microeconomics and decisions are “made at the margin,” meaning the decision often is reduced to “should one more unit be consumed, or produced?” When the benefit from one more unit is less than its cost then the answer is “no” and an optimum has been achieved.

Most standard textbooks at the intermediate level approach microeconomic analysis using geometry and algebra, with optimizations often shown as tangency points between two curves. Calculus can also be employed as the “marginal” is closely approximated by the first derivative of a function. As a simple example, firms are assumed to choose to produce a level of output that maximizes total profit. Total profit is defined as total revenue minus total cost. The optimal output is where marginal revenue (the change in total revenue for a one-unit change in output) is equal to marginal cost (the change in total cost for a one-unit change in output). This is really just a statement that total profit is maximized at the level of output where marginal profit (defined as marginal revenue less marginal cost) is zero.

Section 2 introduces aspects of supply and demand, including a perhaps unexpected result concerning payment of excise taxes. Section 3 considers production costs and profit maximization. It introduces marginal analyses and the optimization processes that culminate in the supply curve discussed earlier. In Section 4, we examine consumer behavior and budget constraints when consumers demand more than one good. This illustrates how the demand curve develops. The problems in this paper can all be done using algebra, except for Problem 4.4 which requires calculus. Section 5 contains the solutions to the problems in the previous sections.

2 Supply and Demand

When most people hear “economics,” they immediately think supply and demand. “Demand,” which refers to an inverse relationship between quantity purchased and the price of a good or service, represents the buyers’ side of a market. Since people generally buy more at lower prices and less at higher prices, the quantity demanded is a decreasing function of price, assuming other factors (such as income, prices of other goods, and expectations) are held constant. "Supply" is typically considered in terms of producer (seller) behavior and refers to a (usually) positive relationship between the quantity of a good produced for sale and price of the good or service. The quantity supplied is an increasing function of the good's or service's price.

Supply is essentially a "max-min" relationship: at a given price, there is a maximum amount that a seller would be willing to produce and offer to the market. However, given any quantity, the associated price would be the minimum the seller would be willing to accept in order to produce and offer that quantity on the market. Demand is a "max-max" relationship: at a given price, the associated quantity is the maximum that would be purchased, and at a given quantity the associated price is the maximum that buyers would be willing to pay.

Supply and demand depend on price. When plotting a demand curve or a supply curve, the practice in economics is to put price on the vertical axis, even though the quantity supplied or demanded is the dependent variable. This reversal of the axes is generally attributed to English economist Alfred Marshall [1] and is a great delight to economics majors who can claim that economists have been "doing it backwards" for over a century.

The quantity supplied $S(p)$ is a strictly increasing function of price p, whereas the quantity demanded $D(p)$ is a strictly decreasing function of price p. Equilibrium occurs when $S(p^*) = D(p^*)$ for price p^*, called the *equilibrium price*; the corresponding quantity q^* is called the *equilibrium quantity*. Economists use the vertical axis to represent price p and the horizontal axis to represent quantity q. Thus, the point (q^*, p^*) where the curves $q = S(p)$ and $q = D(p)$ intersect is referred to as the *equilibrium point*.

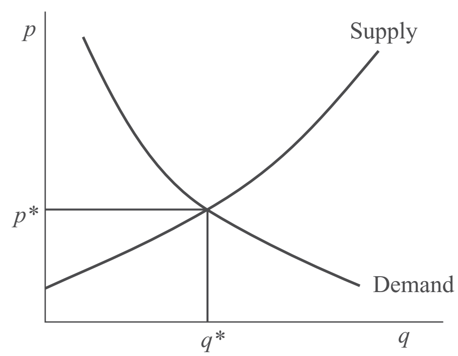


Figure 1

Problem 2.1 The supply and demand for a good at price p are

$$S(p) = 3p - 50 \quad \text{and} \quad D(p) = -2p + 100.$$

(a) How many units would be purchased at \$40 per unit, and how many at \$20 per unit? In each case, would the supplier and consumer both be satisfied? On the same axes, with the vertical axis representing price, show the curves $q = S(p)$ and $q = D(p)$, and the equilibrium point (q^*, p^*).

(b) Suppose $S(p) = mp - 50$ for $m > 0$ and $D(p) = -2p + 100$. Explain how the equilibrium point $(\widehat{q}, \widehat{p})$ is related to (q^*, p^*) when $m > 3$, and when $m < 3$? Confirm your answers graphically.

(c) Suppose $S(p) = 3p - 50$ and $D(p) = -mp + 100$ for $m > 0$, and equilibrium occurs for $\widehat{q}$ units at price $\widehat{p}$. Explain how $\widehat{q}$ and $\widehat{p}$ are related *to* q^* and p^* when $m > 2$, and when $m < 2$. Confirm your answers graphically.

(d) Economists say that a supply curve (demand curve) is "steeper" if it appears more vertical. How must m change for the supply curve $q = mp - 50$ to be steeper, and how must m change for the demand curve $q = -mp + 100$ to be steeper? Confirm your answers by expressing p as a function of q.

Problem 2.2 Let $S = 3p - 50$ and $D = -2p + 100$.

(a) First explain why D can be expressed as a decreasing function of S. Then determine the function $F(S)$ such that $D = F(S)$.

(b) Display the graph of $D = F(S)$, where D is represented on the vertical axis. Then show graphically how to determine the point at which equilibrium occurs. Confirm your result algebraically.

(c) Suppose $D = f(p)$ and $S = g(p)$. Explain why D is a decreasing function of S. What function of S is D?

Problem 2.3 If the supply and demand for a good at price p are

$$S(p) = 3p - 50 \quad \text{and} \quad D(p) = -2p + 100,$$

at equilibrium $q^* = 40$ units are sold at $p^* = \$30$ per unit. Now suppose that a \$5 excise tax is imposed on that good. Then at equilibrium, the per unit price p_c the consumer pays and the per unit price p_s the supplier receives are not necessarily equal.

(a) Suppose the supplier pays the \$5 per item tax for each unit produced; the quantity $D(p)$ demanded by consumers is unaffected. Since the supplier receives \$5 less per item produced, the quantity it will supply is $S(p - 5)$. Show how the supply curve $q = S(p - 5)$ is related to the supply curve $q = S(p)$. Then determine the equilibrium point $(\widehat{q}, \widehat{p})$ for $S(p - 5)$ and $D(p)$. Explain why $p_c = \$33$ and $p_s = \$28$. What fraction of the tax is levied on the consumer and what fraction is levied on the supplier?

(b) Suppose the consumer pays the \$5 tax for each unit bought; the amount $S(p)$ supplied is unaffected. Since the price to consumers has increased \$5, the demand for the good is $D(p + 5)$. Determine the equilibrium point for $S(p)$ and $D(p + 5)$, and use it to show why $p_c = \$33$ and $p_s = \$28$.

(c) On axes with p as the vertical, display the points (q^*, p^*), $(\widehat{q}, p_c)$, and $(\widehat{q}, p_s)$. Show that $\Delta p_c/|\Delta p_s| = 3/2$. What fraction of the \$5 tax is levied on the buyer and what fraction is levied on the supplier? (Note: the absolute value of Δp_s is used because the producer will receive a lower price after the tax. It is, however, the proportions that matter.)

Problem 2.4 Let the supply and demand for a good at price p be $S(p) = m_s p + b_s$ and $D(p) = -m_d p + b_d$, where m_s and m_d are positive.

(a) Determine the quantity q^* and price p^* at which equilibrium occurs.

(b) Suppose an excise tax of \$$T$ per item is imposed on the supplier. Show that at equilibrium,

$$p_c = p^* + (m_s T)/(m_s + m_d) \quad \text{and} \quad p_s = p^* - (m_d T)/(m_s + m_d)$$

for $\widehat{q} = q^* - (m_d m_s T)/(m_s + m_d)$ units.

(c) Suppose the buyer pays the \$$T$ per item tax. Show that at equilibrium, p_c and p_s are the same as in part (b). How many units are sold?

(d) Show that

$$\Delta p_c/|\Delta p_s| = (p_c - p^*)/(p_s - p^*) = m_s/m_d.$$

When is $\Delta p_c > |\Delta p_s|$, and when is $\Delta p_c < |\Delta p_s|$?

(e) Sketch the graph of p_c on the vertical versus m_s/m_d on the horizontal axis. What do p_c and p_s approach as m_s/m_d increases?

Problem 2.5 The elasticity of $y = f(x)$ at a point (x^*, y^*) is defined as the percentage change in y divided by the percentage change in x,

$$\varepsilon = \frac{\Delta y/y^*}{\Delta x/x^*},$$

which is the percentage change in y corresponding to a 1% change in x. Show that at the equilibrium point (q^*, p^*) of $S(p) = m_s p + b_s$ and $D(p) = -m_d + b_d$, the elasticities of supply and demand are $\varepsilon_s = m_s \cdot \frac{p^*}{q^*}$ and $\varepsilon_d = m_d \cdot \frac{p^*}{q^*}$. Thus, $\Delta p_c/|\Delta p_s| = \varepsilon_s/\varepsilon_d$.

The important application of this analysis relates to the development of tax policy. Economists make a distinction between the incidence of a tax and the tax burden. As an example, stores are required to collect sales tax but the burden of the tax is obviously on the consumer.

In the case of taxes put on producers, the policy implication depends on whether the producer can "shift" the tax onto the consumer, and if so, to what extent. The answer depends on the relative elasticities of supply and demand.

Imagine a consumer whose demand for a product was "perfectly inelastic," meaning that the quantity demanded would not change at all in response to any price change. The demand curve would be shown graphically as a vertical line and the elasticity would be 0. If a tax were imposed on the producers of that product, it would be fully shifted onto consumers. Thus, policymakers should know that in taxing producers the burden is really going to fall on consumers. There is also reverse tax shifting, which occurs when the burden is shifted onto producers. As would be expected, this occurs when supply is relatively inelastic compared to demand.

In the analysis of tax burdens, the goal is to assess the impact on economic welfare, called the tax incidence. Tax incidence can be measured using a pass-through ratio defined as $\varepsilon_s/(\varepsilon_s - \varepsilon_d)$.

3 Production Costs and Profit

While "supply and demand" are thought of as the basics of economics, they really represent the outcomes of optimization processes that rely on marginal analysis (what most economists would consider one of the fundamental principles of economic thinking). The term "marginal" is used to define the slope of curves that graphically describe relationships in economics. It always refers to the change in the dependent variable that results when there is a unit change in the independent variable. "Thinking at the margin" is the basis of economic decision making. A consumer might ask questions like, "Do I want one more slice of pizza?" or "Should I work another hour?" whereas a producer might ask "Should I hire one more worker?" or "Should I produce one more unit of the good?"

In most microeconomics texts, the first use of the marginal is the analysis of consumer choice. Economists define "utility" as the benefit a consumer receives from consuming goods and services. Students are usually presented with data similar to that in Table 1, where q is the number of units of a good, and U (measured in utils) is the utility (amount of benefit or value). MU is the *marginal utility* (the additional utility from one unit change in q), defined as $\Delta U/\Delta q$. For example, as q changes from 0 to 1, the utility U changes from 0 to 10, and so the MU of the first unit is 10 (utils).[1]

q	U	MU
0	0	
1	10	10
2	17	7
3	22	5
4	26	4

Table 1

Marginal utility decreases because most consumers will report that at some point, the extra (marginal) utility will decrease (for instance, a third slice of pizza has less marginal utility than the second, etc.). As more of something is consumed, utility should (eventually) increase at a decreasing rate. Economists consider this behavior to be so pervasive that it is termed the "Law of Diminishing Marginal Utility."

Although consumers are assumed to be able to use nonnegative real numbers to measure utility, we are using integers to simplify matters. Since utility is a continuous function of q, there is a value of q at which the slope will begin to decrease. Maximum total utility would occur at the value of q where marginal utility is zero, what economists call the "saturation" point. Consumption of additional units of q would result in negative MU or a decrease in U.

[1] If a total utility function were provided, marginal utility could be calculated at a particular value of q. In this treatment MU is calculated between two values and, in the graphical representation, the value of MU would be placed between the values of q; that is, one would plot the MU of 10 vertically with a horizontal coordinate of .5.

Problem 3.1 (The Law of Diminishing Marginal Utility) Refer to Table 1.

(a) Show why U and MU are functions of q. Indicate the domain and range of each function. Suppose a new table is to have double the values of q in Table 1. Use Table 1 to determine the new table's values of U and MU, or explain why you cannot enter some value.

(b) On the same axes, show the graphs of the functions $U(q)$ and $(MU)(q)$. What does the value of MU at q represent on U's graph? Use the words "increasing" and "decreasing" to complete the following: U is ___ at a(n) ___ rate, and MU is ___ at a(n) ___ rate.

(c) Show the graph of MU if the values of U are increased by k > 0, and if the values of U are multiplied by k. How is each of these graphs related to the graph of MU in (b)?

(d) Explain why, in general, neither U nor MU can be a linear function of q as q becomes large.

(e) Explain why q is a function of U, denoted $q(U)$. Then show the graphs of $U(q)$ and $q(U)$ on XY-axes, where X represents domain values and Y represents range values. Prove that the line $Y = X$ is the perpendicular bisector of each pair of corresponding points (q, U) and (U, q). [Graphs with this property are said to be symmetric with respect to the line $Y = X$.]

Production costs

Assume that a firm's production of its product involves only a fixed cost $K = \$500$ for its capital (which represents its building and its machinery) and a variable amount of labor, L. The firm's variable costs are the product of L and the payment $w = \$200$ per unit of L. Table 2 lists the firm's total fixed costs (TFC), total variable costs (TVC), and total costs (TC), together with its corresponding average costs (the per-unit costs).

q	L	MP_L	TFC	TVC	TC	AFC	AVC	ATC	MC
0	0	–	500	0	500	–	–	–	
5	1	5	500	200	700	100	40	140	40.00
13	2	8	500	400	900	38.46	30.77	69.23	25.00
18	3	5	500	600	1,100	27.78	33.33	61.11	40.00
22	4	4	500	800	1,300	22.73	36.36	59.09	50.00
25	5	3	500	1,000	1,500	20.00	40.00	60.00	66.67
27	6	2	500	1,200	1,700	18.52	44.44	62.96	100.00

Table 2. $K = \$500$, $w = \$200$

The table's third column shows the productivity of labor or marginal product of labor, $MP_L = \Delta q/\Delta L$ defined as the change in output that results from a change in the quantity L of labor. For example, when L changes from 0 to 1, the output q changes from 0 to 5, so the marginal product of the first unit of labor is $MP_L = 5$ units of output.[2]

As the firm adds more and more labor to its fixed capital (in order to produce more output), it eventually runs into another economic law, the "Law of Diminishing Returns." There will be a value of L at which MP_L will begin to decrease. As MP_L decreases, it will eventually cause the average total cost $ATC = TC/q$ of producing one unit of output to increase. Therefore, at the quantity of labor where MP_L starts to decrease, the marginal cost MC starts to increase (since TC's constant change of \$200 is divided by decreasing changes in q) and will result in increasing per unit costs.

Problem 3.2

(a) Average Fixed Cost (AFC) decreases as output increases (called "spreading the overhead"). The table's six data pairs (q, AFC) belong to the graph of a function defined for positive q. Use this function to estimate the number of units q at which $AFC = \$25$.

[2] See earlier footnote on marginal utility; the same comments apply to marginal product and to marginal cost.

(b) Show why MC equals marginal TVC. How are the graphs of TC and TVC related?

(c) Prove that MC increases if and only if MP_L decreases.

(d) Use the graph of q as a function of L to show that AVC increases if and only if MP_L decreases.

(e) In Table 2, observe that AVC decreases when $MC < AVC$, and AVC increases when $MC > AVC$. Explain why this makes sense. Then algebraically show why by calculating the average cost of $n+1$ items if the average cost of n items is d, and an additional item costs kd ($k > 0$).

(f) The firm's cost curves are graphed on a set of axes with cost on the vertical and quantity of output on the horizontal. Typically, a cost-curve graph includes only AVC, ATC, and MC, with the marginal cost curve crossing the AVC and ATC curves at the minimum point of each curve. Use (c), (d), and (e) to sketch a firm's cost-curve graph.

Profit maximization

A firm makes decisions with the goal of making the largest possible profit, which is the difference between total revenue (TR) and total cost (TC). To maximize profits, the firm must choose the "best" or optimal level of output. This requires some assumptions about how many firms are in the industry, what market shares they have, and whether the product being produced is homogeneous or differentiated in the eyes of the consumer. That is, does the consumer view the output of each firm as identical or perfect substitutes for each other, or does the consumer perceive differences that result in a preference for the output of one firm over any other? The firm has to consider the demand for its product in this context.

Suppose that the firm is a monopolist, the only producer of a good or service in its particular market. In this case, the market demand curve is a decreasing function of price, and to sell more of its product the monopolist must reduce the price. Above a certain high price, people will not buy any of the good or service. Total revenue is the product of price (p) and quantity (q), $TR = qp$. Marginal revenue, $MR = \Delta(TR)/\Delta q$, is the additional revenue from selling one more unit of the product at price p. For any firm facing a downward-sloping demand curve (a demand curve with less than perfect elasticity), $MR < p$.

The opposite of monopolist would be a firm in "pure competition,"[3] which means being in a market structure where many small firms produce the same product and charge the same price. Such a firm can sell as much as it can produce as long as it charges the same price p as all of its rivals. The demand curve this firm faces is a horizontal line set at the level of the "going" or market price, and $MR = p$. (With reference to Section 2, this demand curve is perfectly elastic.)

Table 3 shows the data for a firm in pure competition at price $p = \$55$ per unit.

q	L	MP_L	TFC	TVC	TC	MC	p	TR	MR
0	0	–	0	0	0		–	–	
5	1	5	0	200	200	40.00	55	275	55
13	2	8	0	400	400	25.00	55	715	55
18	3	5	0	600	600	40.00	55	990	55
22	4	4	0	800	800	50.00	55	1,210	55
25	5	3	0	1,000	1,000	66.67	55	1,375	55
27	6	2	0	1,200	1,200	100.00	55	1,486	55

Table 3

A firm wishing to maximize total profit should keep producing as long as the additional revenue exceeds the additional cost. The optimal or profit-maximizing quantity for a firm would be where marginal revenue equals marginal cost. Table 3 shows that the firm's optimum level of output is 22 units. The firm should not produce 25 units since the marginal

[3] Some economics texts describe "perfect competition" as "pure competition" with perfect information. That is, everyone in the market knows all the relevant information about product availability, etc. This assumption seemed much more extreme before the Internet.

cost would be greater than the marginal revenue. In cases where $MR = MC$ at two different quantities, the optimization should choose the quantity at which $MR = MC$ and MC is rising (since the goal is to maximize total profit).

The analysis of profit maximization is the basis for the supply curve—the quantity that the firm will produce and sell at each price. If the market price increased to \$70 in Table 5, the firm (and all the other firms) seeking to maximize profits will choose to produce 25 units of output. Given the firm's cost structure, changes in price change marginal revenue, and therefore change the optimal level of output. In short, the levels of output shown on a supply curve are the firms' optimal levels of output at each particular price. Indeed, the marginal cost curve is the firm's supply curve and the market supply curve is found by horizontally summing the firms' supply curves; that is, adding up the quantity that each firm will produce at any price. The sum at each price is the quantity that will be supplied to the market at each price.

Problem 3.3 (Supply Curve) Suppose the firm in Table 3 must produce a minimum of 10 units.

(a) On axes with the table's values of q on the horizontal, plot the points (q, p) where p is the minimum price for which the firm produces q units.

(b) The polygonal line segments that connect successive points on the plot in (a) are an approximate representation of the firm's supply curve. Use the approximate supply curve to estimate how many units the firm would produce at $p = \$34$ per unit.

Since the marginal cost curve is the firm's supply curve, the market supply curve is found by horizontally summing the firms' supply curves; that is, adding up the quantity that each firm will produce at any price. The sum at each price is the quantity that will be supplied to the market at each price.

Problem 3.4

(a) Show why $MR < p$ for a monopolist firm, and why $MR = p$ for a firm in pure competition.

(b) In Table 3, show why $MR \geq MC$ if and only if $MP_L \geq 200/p$.

(c) On the same axes with q on the horizontal, use line segments to join Table 3's successive values of TR, and Table 3's successive values of TC. Use MR and MC to determine at which q the firm's profit is maximum. At which q is $MR - MC$ greatest, and at which q is $MR - MC$ least?

(d) For what fixed cost K would $TC < TR$ for all q? Explain why $MR > MC$ at q does or does not mean that the firm is profitable for q.

(e) Suppose a firm's increasing number of workers $\{L_i\}$ produce an increasing quantity $\{q_i\}$ of units, where the firm's production involves only its fixed expense K, wage w paid to each worker, and market price p per unit produced.

(1) Show that $MR \geq MC$ if and only if $MP_L \geq w/p$.

(2) What relation among q_i, L_i, K, w, p is equivalent to $TR \geq TC$? What must be the output for L_i workers in order for the firm to have a profit?

4 Consumer Behavior and Budget Constraints

Section 2 considered a consumer choosing how much to consume of only one good. However, consumers purchase many goods. Since all goods and services have prices, and consumers have a specified amount of money to spend, consumers must maximize their utility subject to a "budget constraint." To determine the optimal combination of goods X and Y the consumer should choose, we can employ the constructs of budget line and indifference curves.[4]

It is assumed that people are rational and will seek to maximize their utility (benefit or satisfaction) subject to the constraint of the budget. Moreover, it is assumed that the amount of a good can be any nonnegative real number, and people can use nonnegative real numbers to assess how much more utility one good gives than another. Therefore,

[4] There are generally no limitations on the relationship between goods X and Y. The two goods could be substitutes (goods that are used instead of each other) or complements (goods that are used together) or unrelated in the sense of direct consumption of the two goods.

utility preferences can be represented by curves, called *indifference curves*, on a set of axes with quantity y of good Y on the vertical and quantity x of good X on the horizontal. (This is by tradition in economics, and does not assume that one variable is dependent and the other is independent.)

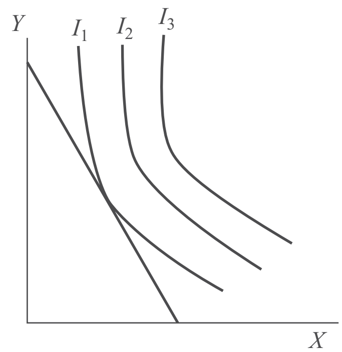


Figure 2

On an indifference curve, each plotted point represents a combination of the two goods that yield the same overall level of utility. It might be helpful to liken indifference curves to curves that describe the same altitude on a topographical map. Indifference curves that are closer to the origin represent lower levels of overall utility, whereas those farther away represent higher levels.

Economists make the following assumptions about indifference curves.

(1) Indifference curves can't intersect (if they did, the same combination of X and Y would simultaneously yield two different amounts of total utility).
(2) The curves are negatively sloped (in order to maintain the same overall level of total utility, a reduction in the quantity of one good must be compensated for by an increase in the amount of the other good).
(3) There is an infinitely dense field of such curves, meaning that for each $x > 0$ and $y > 0$, there is exactly one indifference curve that passes through the point (x, y).

Indifference curves are convex, as shown in Figure 2.[5] This is based on the premise that the less one has of a good the more valuable each unit of that good is (and vice versa). For example, suppose a consumer's combination (x, y) on an indifference curve is high up to the left and close to the vertical axis. Having a lot of Y and relatively little X, the person would be willing to sacrifice a relatively large amount of Y for a small amount of X, and so the curve is steep when x is very small. If the consumer continues to give up Y to get more X, then Y becomes scarcer (and X more abundant), so the consumer is willing to give up less Y for an increase of X. Thus, the curve becomes flatter as x becomes large.

To maximize utility, a consumer would like to determine the best possible point on the highest possible indifference curve. What limits this optimization process is the consumer's amount of money M that can be spent, and the respective prices P_x and P_y of goods X and Y. The line whose intercepts are M divided by the price of each good is called the consumer's *budget line*. Suppose, for example, that a consumer has \$10 to spend. If the price of X is \$2 per unit and the price of Y is \$1 per unit, the consumer could buy all possible combinations of X and Y for which $2x + y = 10$.

To determine the consumer's maximum utility, the consumer's budget line is overlaid on the field of indifference curves. Because the field is infinitely dense, the budget line will be tangent to one indifference curve. In the figure below, the point of tangency between the budget line and I represents the consumer's optimum combination of the

[5] There are exceptions to this but they represent goods with particular characteristics, such as the case of a good being a "neuter," that is, neither adding to nor detracting from the utility derived from the other good. Depending on which good is the neuter the indifference curves could be horizontal or vertical lines.

two goods; the combination that will provide the consumer with the highest possible level of utility given the budget constraint.

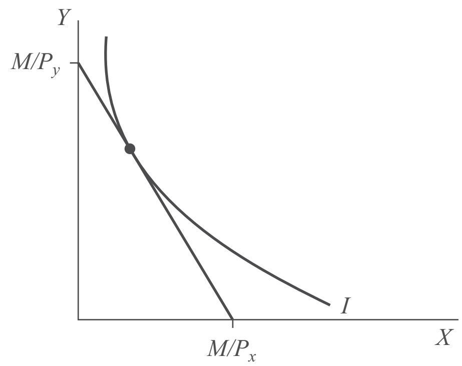


Figure 3

Problem 4.1 Suppose your utility is $U(x, y) = xy$ for x number of good X and y number of good Y, and your budget constraint is $2x + y = 10$.

(a) On the same xy-axes, sketch the graphs of $U(x, y) = 4, 8, 16$. For what value k is $U(x, y) = k$ tangent to your budget line and for which pair (x, y) does it occur?

(b) Show that $U(x, y) = x(10 - 2x)$. What is the maximum of $U(x, y)$ and for which pair (x, y) does it occur?

(c) If the number of each good X and Y must be an integer, what combination of X and Y would you choose to maximize your utility? On what indifference curve would your chosen (x, y) values lie?

Problem 4.2 (Changes in available budget and price) Suppose your utility is $U(x, y) = xy$.

(a) If your budget constraint is $2x + y = M$, what is your maximum utility? For what combination (x, y) does it occur?

(b) On the same axes, sketch the indifference curve tangent to your budget constraint for $M = 5$ and for $M = 10$. How are the graphs related?

(c) If your budget constraint is $ax + y = 10$, what is your maximum utility? For what combination (x, y) does it occur?

(d) On the same set of axes, sketch the indifference curve tangent to your budget constraint for $a = 2$ and $a = 5$. How are the graphs related?

(e) For $U(x, y) = Cxy$ and budget constraint $ax + by = M$, show that the maximum utility is

$$U = \frac{CM^2}{4ab} \quad \text{at} \quad (x, y) = \left(\frac{M}{2a}, \frac{M}{2b}\right).$$

Problem 4.3 (Equilibrium for good X) Suppose your utility is $U(x, y) = xy$ and your budget constraint is $xP_x + yP_y = M$.

(a) Show that a maximum utility the number of units demanded for good X is $x = 840/P_x$. How many units of X and Y would be demanded if $P_x = \$30$ per unit of X?

(b) Let P_x have the values MC in Table 3 (page 9) for the quantities $x = 13, 18, 22, 25, 27$. On the same axes, use the line segments joining consecutive points (x, P_x) to approximate X's supply curve, and the line segments joining consecutive points $(840/x, P_x)$ to approximate X's demand curve.

(c) What is the equilibrium point for X's demand and supply curves?

For the utility $U(x, y)$ of goods X and Y, the marginal utility of X, denoted MU_x, is the increase in $U(x, y)$ from the addition of one more unit of X (with the number y fixed), and the marginal utility of Y, denoted MU_y, is the increase in $U(x, y)$ from the addition of one more unit of Y (with the number x fixed). At the point where the budget line is tangent to the indifference curve, it can be shown that

$$MU_x/P_x = MU_y/P_y.$$

This condition, called the "equimarginal principle," says that total utility is maximized when the marginal utility per dollar for the last dollar is the same for each good purchased. (Since MU_x is the increase in utility from the addition of one more unit of X and the price of X is P_x dollars per unit, MU_x/P_x is the increase in utility from spending one more dollar on X. Likewise, MU_y/P_y is the increase in utility from spending one more dollar on Y.)

The rationale for the principle can also be explained as follows. If $MU_x/P_x > MU_y/P_y$, the consumer would buy more of X and less of Y. However, in buying more of X, the consumer experiences diminishing MU_x of X, and less utility per dollar spent on X. And buying less of Y means that MU_y for one unit of Y will be higher, as will the marginal utility per dollar spent on Y. So the tendency would be to start buying more Y until the inequality is reversed. This continues for each dollar spent until the choice point is driven to the point of equality. Therefore, for the given budget constraint, the consumer cannot achieve higher level of total utility than at the point where the marginal utility per dollar is the same for each good purchased.

We can use calculus to determine the maximum utility $U(x, y)$ subject to the budget constraint $ax + by = M$.

Problem 4.4 (Maximize utility subject to the budget constraint) You seek to purchase a combination of goods X and Y subject to your budget constraint $2x + y = 10$ and utility $U(x, y) = (3/4)xy^4$.

(a) Express $U(x, y)$ as a function of y, and use a sketch of the graph to show why $U(x, y)$ must have a maximum value for $y \in (0, 10)$.

(b) Determine the maximum of $U(x, y)$ and the values x^*, y^* at which it occurs.

(c) Is the Equimarginal Principle satisfied at (x^*, y^*)?

Producer behavior and the cost minimizing combination of inputs

The same approach described above can be used to analyze how a firm chooses the "best" combination of inputs. The constructs used are basically the same; in place of the indifference curves is a set of isoquants, each illustrating that a given level of output can be produced with different combinations of the two inputs, usually capital (K) and labor (L). The isoquant's slope is the ratio of the marginal products of the inputs.

In place of the budget line is an isocost line, which represents an amount of total spending on inputs (total cost) and whose slope is the ratio of the input prices. The optimization point is the tangency; at that point the slope of the isoquant, which is defined as the marginal product of labor divided by the marginal product of capital is equal to the slope of the isocost line, which is the price of labor divided by the price of capital. With some simple manipulation it can be shown that firms select a combination of resources at which the marginal product divided by the price is equal for all inputs purchased (note the similarity to the equimarginal principle discussed on page 15).

The combination of inputs chosen can be viewed in two ways: the firm reaches the highest isoquant it can with its given level of cost, or it produces a given level of output at the lowest possible cost.[6] As such, the analysis provides an intuitive illustration of duality theory.

5 Solutions

Problem 2.1

(a) Since $S(40) = 70$ and $D(40) = 20$, the consumer will buy 20 units at \$40 and the supplier will not be satisfied. Since $S(20) = 10$ and $D(20) = 80$, the supplier will sell all 10 units at \$20 and the consumer will not be satisfied. At equilibrium, $(q^*, p^*) = (40, 30)$, which means that 40 units will be exchanged at \$30 per unit.

[6] The analysis of the least-cost combination is generally attributed to Paul Samuelson [2].

(b) For $S = mp - 50$ and $D = -2p + 100$, equilibrium occurs at $\widehat{p} = 150/(m + 2)$ and $\widehat{q} = 100 - 2\widehat{p}$. If $m > 3$, then $\widehat{p} < 30$ and the point $(\widehat{q}, \widehat{p})$ on the demand curve lies below and to the right of the point (q^*, p^*). If $m < 3$, then $\widehat{p} > 30$ and the point $(\widehat{q}, \widehat{p})$ on the demand curve lies above and to the left of the point (q^*, p^*). As m increases (decreases) the equilibrium price decreases (increases), which means that the amount demanded increases (decreases).

(c) For $S = 3p - 50$ and $D = -mp + 100$, equilibrium occurs at $\widehat{p} = 150/(m + 3)$ and $\widehat{q} = 3\widehat{p} - 50$. If $m > 2$, then $\widehat{p} < 30$ and the point $(\widehat{q}, \widehat{p})$ on the supply curve lies below and to the left of the point (q^*, p^*). For $m < 2$, the point $(\widehat{q}, \widehat{p})$ on the supply curve lies above and to the right of (q^*, p^*). As m increases (decreases) the equilibrium price decreases (increases), which means that the amount supplied decreases (increases).

(d) The supply curve $S = mp - 50$ becomes steeper as m decreases, and the demand curve $D = -mp + 100$ becomes steeper as m decreases.

Problem 2.2

(a) As p increases, the amount supplied increases and the amount demanded decreases. Therefore, for values of $D = -2p + 100$ on the vertical axis and values of $S = 3p - 50$ on the horizontal axis, the graph of (S, D) is decreasing. Substituting $p = (S - 50)/3$ into the equation for D yields $D = -2(S - 100)/3$.

(b) The equilibrium point occurs where the line $D = S$ intersects the graph of $D = -2(S - 100)/3$. The solution is $S = 40 = q^*$ and $p^* = 30$.

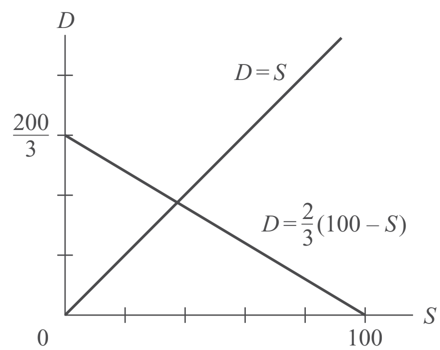


Figure 4

(c) $D = f(p)$ and $S = g(p)$ are one-to-one functions. Therefore $p = g^{-1}(S)$ and $D = f(g^{-1}(S))$. Suppose $S_1 < S_2$. Since $g(p)$ and $g^{-1}(S)$ are increasing functions and $f(p)$ is a decreasing function of p, it follows that $p_1 = g^{-1}(S_1) < p_2 = g^{-1}(S_2)$ and $D(S_1) = f(p_1) > f(p_2) = D(S_2)$.

Problem 2.3

(a) The supply curve $S(p - 5)$ is the curve $S(p)$ raised 5 units in the vertical direction. Equilibrium for $S(p - 5) = 3p - 65$ and $D(p) = -2p + 100$ occurs for $(\widehat{q}, \widehat{p}) = (34, 33)$. Since $p_c = \$33$, and the supplier receives 5 less, $p_s = \$28$.

(b) The equilibrium point for $S(p)$ and $D(p + 5)$ is $(\widehat{q}, \widehat{p}) = (34, 28)$.

(c) The consumer (p_c) pays 3/5 of the tax and the producer (p_s) pays 2/5 of the tax.

Problem 2.4

(a) For $S(p) = m_s p + b_s$ and $D(p) = -m_d p + b_d$ where m_s and m_d are positive, equilibrium occurs at $p^* = (b_d - b_s)/(m_d + m_s)$, $q^* = (m_s b_d + m_d b_s)/(m_d + m_s)$.

(b) For $S(p - T) = m_s p + b_s$ and $D(p) = -m_d p + b_d$ equilibrium occurs at $p = p^* + (m_s T)/(m_s + m_d)$. Since $p_c = p$ and $p_s = p - T$

$$p_c = p^* + (m_s T)/(m_s + m_d) \quad \text{and} \quad p_s = p^* - (m_d T)/(m_s + m_d).$$

Furthermore $\widehat{q} = -m_d(p^* + (m_s T)/(m_s + m_d)) = q^* - (m_d m_s T)/(m_s + m_d)$ units.

(c) Equilibrium for $S(p) = m_s p + b_s$ and $D(p + T) = -m_d p + b_d$ occurs at $p = p^* + (m_d T)/(m_s + m_d)$. In this case,

$$p_c = p + T = p^* + (m_s T)/(m_s + m_d) \quad \text{and} \quad p_s = p = p^* - (m_d T)/(m_s + m_d).$$

(d) $\Delta p_c/|\Delta p_s| = (p_c - p^*)/(p_s - p^*) = \{(m_s T)/(m_s + m_d)\}/|(m_d T)/(m_s + m_d)| = m_s/m_d$.

(e) Let $m = m_s/m_d$. Then $p_c = p^* + (mT)/(m + 1)$ increases to $p^* + T$ as m increases, and p_s decreases to p^*.

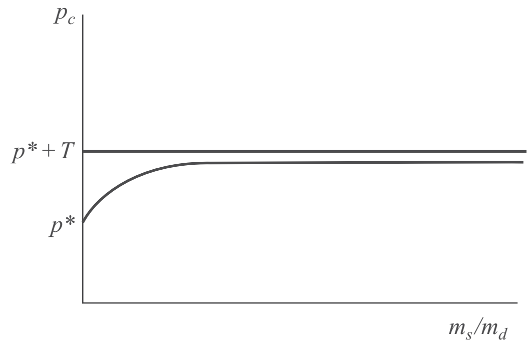


Figure 5

Problem 2.5

The elasticity of $y = f(x)$ at a point (x^*, y^*) can be written

$$\varepsilon = \frac{\Delta y/y^*}{\Delta x/x^*}.$$

For $S(p) = m_s p + b_s$ and $D(p) = -m_d p + b_d$, the elasticities at (q^*, p^*) of supply and demand are $\varepsilon_s = m_s \frac{p^*}{q^*}$ and $\varepsilon_d = m_d \frac{p^*}{q^*}$, where $m_s = \Delta S/\Delta p$ and $m_d = \Delta D/\Delta p$.

Problem 3.1

(a) U has domain $\{0, 1, 2, 3, 4\}$ and range $\{0, 10, 17, 22, 26\}$. And MU has domain $\{1, 2, 3, 4\}$ and range $\{10, 7, 5, 4\}$. The table whose values of q are doubled would have $U(2) = 17$ and $U(4) = 26$, and $MU(2) = 17$ and $MU(4) = 4$.

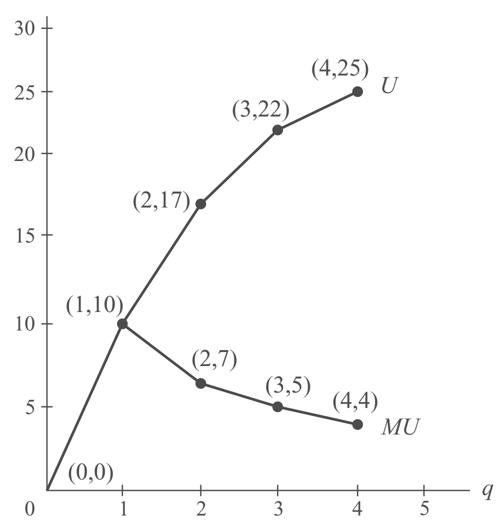


Figure 6

(b) U is increasing at a decreasing rate, and MU is decreasing at a decreasing rate.

(c) This is a consequence of the Law of Diminishing Marginal Utility.

Problem 3.2

(a) $AFC = 500/q$. *If* $AFC = \$25$, then $q = 20$.

(b) $MC = (\Delta TVC + \Delta TFC)/\Delta q$ and $\Delta TFC = 0$.

(c) $MC = \Delta TVC/\Delta q = (200\Delta L)/\Delta q = 200/MP_L$.

(d) $AVC = 200/(q/L)$ increases if and only if q/L decreases. Furthermore, the slopes in Figure 7 show that $q_{i+1}/L_{i+1} < q_i/L_i$ if and only if $MP_L(q_{i+1}) < MP_L(q_i)$.

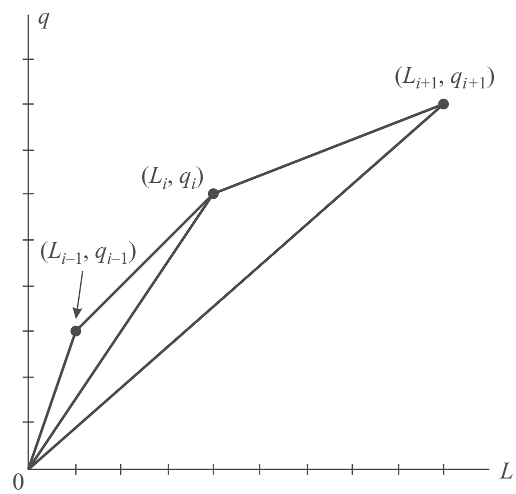


Figure 7

(e) The average cost of the $n + 1$ items is $d(n + k)/(n + 1)$.

(f) See Figure 8.

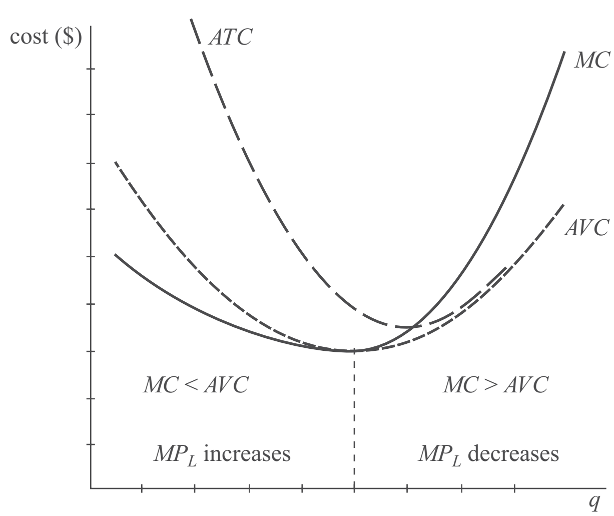


Figure 8

Problem 3.3

(a) The firm's approximate supply curve consists of the line segments joining successive points (q, MC) for $q > 13$.

(b) At $p = \$34$ per unit, the firm would produce $q = 16$ units.

Problem 3.4

(a) MR is the change in revenue pq for a one-unit increase in quantity purchased. If p_1 is the price for the $(q+1)$st unit, then $MR = \frac{p_1(q+1)-pq}{(q+1)-q}$. For a monopolist firm facing a decreasing demand function, $p_1 < p$ and $MR < p$. For a firm in perfect competition, $p_1 = p$ and $MR = p$.

(b) $MR = p$ and $MC - 200/p$.

(c) $TR - TC$ is greatest at $q = 22$, whereas $MR - MC$ is greatest at $q = 13$.

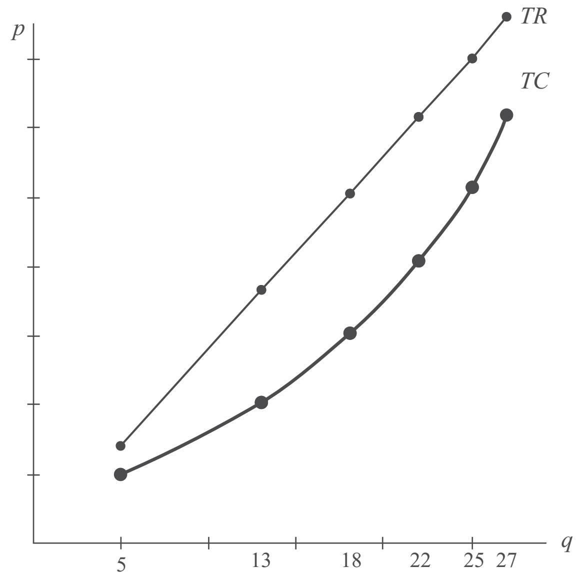


Figure 9

(d) $TC < TR$ for $K > \$410$. It is possible for $MR > MC$ for a quantity q at which $TR < TC$.

(e) $TR \geq TC$ at (q_i, L_i) if and only if $q_i p \geq K + L_i w$. In order for the firm to have a profit, $q_i > (K + L_i w)/p$.

Problem 4.1

(a) Setting $k/x = 10 - 2x$ yields $2x^2 - 10x + k = 0$. Using the quadratic formula, we obtain $x = (10 \pm \sqrt{100 - 8k})/4$. The curves $y = k/x$ and $y = 10 - 2x$ will be tangent if $k = 25/2$. This occurs at $(x, y) = (5/2, 5)$.

(b) The graph of $U = x(10 - 2x)$ is a parabola whose maximum is $25/2$ at $x = 5$. Thus, $y = 5$.

(c) Both (x, y) combinations (2, 6) and (3, 4) give you the same maximum utility $U = xy = 12$.

Problem 4.2

(a) The graph of $U = x(M - 2x)$ is a parabola whose maximum is $M^2/8$ at $(x, y) = (M/4, M/2)$.

(b) The graph of $U = x(10 - ax)$ is a parabola whose maximum is $25/a$ at $(x, y) = (5/a, 5)$.

Problem 4.3

(a) $U(x, y)$ has maximum value $(840)^2/P_x$ at $(x, y) = (840/P_x, 840)$. If $P_x = \$30$, then $x = 28$ units of X would be demanded.

(b) See Figure 10.

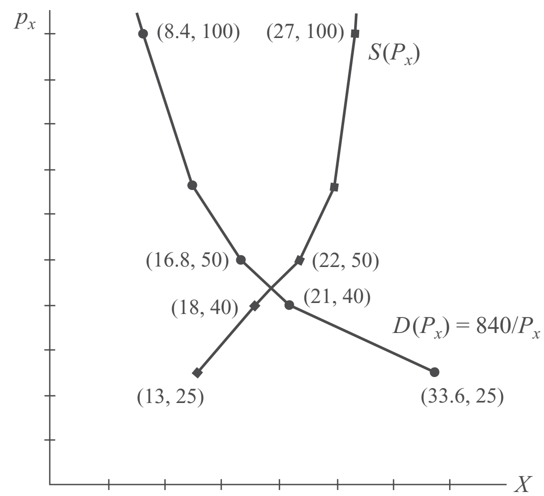


Figure 10

(c) The supply curves and demand curves intersect at the point $(p^*, q^*) = (20.19, 44.95)$. Thus, at equilibrium, 20.19 units of good X would be sold at \$44.95 per unit.

Problem 4.4

(a) Let $U(x, y) = (3/8)(10 - y)y^4$ be denoted $V(y)$. Since $V(0) = 0 = V(10)$, and $V(y)$ is positive for $y \in (0, 10)$ and negative for $y > 10$, it has a maximum in the interval (0, 10).

(b) Since $V'(y) = (3/8)(40y^3 - 5y^4)$ is zero at $y = 8$, and $y''(8) < 0$, the function $V(y) = U(x, y)$ has a maximum at $y^* = 8$ and $x^* = 1$.

(c) At $x = x^*$ and $y = y^*$, the Equimarginal Principle is satisfied:

$$MU_x/P_x = ((3/4)y^4)/2 = 3(8)^3 \quad \text{and} \quad MU_y/P_y = 3xy^3/1 = 3(8)^3$$

References

[1] Alfred Marshall, *Principles of Economics*, Macmillan and Co., Ltd., London 1920 (originally published 1890).

[2] Paul A. Samuelson, *Foundations of Economic Analysis*, Harvard University Press, Boston, MA, 1947, Chapter 4.

2

Scenarios Involving Marginal Analysis

Julie Glass
California State University, East Bay

Lynn Paringer and Jane Lopus
Department of Economics, California State University, East Bay

1 Introduction

Many economic decisions involve investigating the effects of small changes on various outcomes. Analyzing the effects of small or incremental changes allows us to investigate factors affecting cost, revenue, and profit in microeconomics. Economists use the term "marginal" to connote the change due to a one unit change in something. For example, to an economist, marginal cost (MC) means the change in total cost from producing one more unit of q, and marginal revenue (MR) is the change in total revenue from selling one more unit of q. Since the rate of change at a value is represented by the derivative, the word "marginal" in economics is approximated by the derivative. Thus, $MC = d(TC)/dq$. Knowing how a function changes can also be useful in economics situations where it may be difficult to obtain exact data, but where approximations will suffice for decision-making.

Fixed costs (e.g., lease payments on a rented building, fixed monthly fee for the business answer phone, installment payments for purchased equipment) are incurred regardless of whether any output is generated, and do not depend on the amount of output the firm produces. Variable costs vary with the output level of the firm. A cost function is derived from observation of the relationship between costs and level of output. A principal goal of the economist/mathematician is to model the costs of operation as closely as possible to what actually occurs in the production process and to tie the costs of production to the level of output being produced. Using the data points on output and cost, one utilizes some statistical technique (economists generally use regression with cost as the dependent variable and output as the independent variable) to identify the mathematical relationship between the level of the firm's output and the cost of producing that output.

In this paper, we present problems that apply calculus to marginal decision-making in microeconomics. The functions we consider are related to supply, demand, cost, revenue, and profit. Solutions to the problems are contained in Section 5.

2 Ingrid Estes Engages in Economic Industrial Espionage

Ingrid Estes is a garbage collector who has a side job of dumpster diving. Firms pay Ingrid to go through their competitors' garbage in order to determine their competitors' cost functions.

In general, the total cost (TC) of producing q units is the sum of the fixed costs (FC) and the variable costs (VC). The constant FC is not dependent on the value of q (or any other variable).

$$TC = FC + VC. \tag{1}$$

Marginal cost is the derivative of the total cost function with respect to q,

$$MC = \frac{d(TC)}{dq}. \tag{2}$$

Problem 2.1 (Finding Cost Functions)
On one of her latest dives, Ingrid came across a page with the graph of

$$\frac{d(MC)}{dq} = -12 + .3q \text{ with } q \text{ on the horizontal axis.} \tag{3}$$

Since the next page was torn, Ingrid could only make out

$$MC = 500 - \text{(everything after the minus sign was torn off).} \tag{4}$$

On the following water damaged page, the only thing Ingrid could read was "fixed cost equals $150,000." Being very resourceful, Ingrid figured out how to use the information above to determine the marginal cost function and the total cost function.

(a) Use (3), as Ingrid first did, to describe of the graph of MC. Then determine the marginal cost function and show its graph.
(b) Show how Ingrid could use MC to sketch the shape of TC's graph. Then explain how she determined that $TC = 150{,}000 + 500q - 6q^2 + .05q^3$.
(c) Assuming Ingrid's total cost function is correct, sketch a graph of the Average Variable Cost function, $AVC = \frac{VC}{q}$.
(d) Being curious, Ingrid sketched her graphs of AVC and MC on the same axes and observed that $AVC = MC$ at the value q where AVC is at a minimum. She also observed that AVC is increasing when $MC > AVC$, and AVC is decreasing when $MC < AVC$. For her submitted report, she confirmed her observations by proving that in general,

$$MC - AVC = q \cdot \frac{d(AVC)}{dq}. \tag{5}$$

Show Ingrid's algebraic work and a proof of (5) that she might have included in her report. Did she prove that AVC is equal to MC at q if and only if AVC has a minimum at q?

Remark: Since it costs more to produce more, total variable costs increase as more is produced. However, it is possible that these variable costs increase at a decreasing rate initially as workers gain efficiency in the production process. If variable costs increase at a decreasing rate, then the average variable cost (VC/q) would be decreasing over a range of output. However, at some point the variable costs are likely to increase at an increasing rate. This results from factors such as when one must pay overtime to workers beyond an 8 hour day or when equipment breakdowns increase as output expands or when one must pay an increasing incremental (marginal) cost to gain additional units of some inputs (e.g. hiring more highly specialized labor or purchasing more platinum or precious metal in manufacturing). Because the marginal cost is the slope of the total cost function (and the variable cost function), marginal cost may initially decrease, hit a minimum, and then increase, crossing the average variable cost function at its minimum and also crossing the average total cost function at its minimum.

Problem 2.2 (Estimating Marginal Costs and Total Costs)
Ingrid has a new contract with *Machines Incorporated (MI)*. It turns out that *Build Better (BB)* is thinking of purchasing some new equipment from *MI* and the two corporations are negotiating the price of that equipment. On a recent dive in *BB*'s dumpster, Ingrid discovered a torn sheet containing a partial table of TC's values, which she listed.

q	1300	1400	1600	1700
TC	732,400	735,200	742,000	748,400

According to *MI*, its new equipment will enable *BB* to increase production by 100 units per month and earn $90,000 in additional revenue per month. To better negotiate the price of its equipment, *MI* wants to estimate *BB*'s monthly profit (total revenue – total cost) when $q = 1400$ and when $q = 1500$.

(a) Ingrid used slope m_1 of the line joining (1300, 732400) and (1400, 735200), and the slope m_2 of the line joining (1400, 735200) and (1600, 742000), to estimate TC's change and *BB*'s monthly profit when q increased from 1400 to 1500, and when q increased from 1500 to 1600. What estimates did she get? What were her estimates for $TC(1500)$?

(b) Ingrid later found out that $TC = 150{,}000 + 1202q - .84q^2 + .0002q^3$. Being curious, she compared $MC(1400)$ with m_1, and $MC(1500)$ with m_2 and the slope m_3 of the line joining (1500, 738000) and (1600, 742000). Do likewise.

(c) Use $MC(1400)$ and $MC(1500)$ to estimate the corresponding changes in TC when q increases from 1400 to 1500, and from 1500 to 1600. Graphically compare your estimate with the exact changes in TC. Starting at which values $q \in \{1300, 1400, 1500, 1600\}$, will *BB* earn the largest monthly profit and the least monthly profit when production is increased 100 units?

(d) Ingrid realized that *BB*'s monthly profit might be greater than in (c) if *BB* chose a different value at which to increase production. So, to be sure, she determined the value q for which $\Delta TC = TC(q + 100) - TC(q)$ is minimum and profit is maximum. What did she discover is *BB*'s maximum monthly profit for increased production of 100 units?

3 Phyllis Physician is a Monopolistic Competitor

Phyllis Physician's demand function is $q = 15{,}000 - 10p$, where q is the number of patients in her practice who pay p dollars annually to be included in her practice. Her total cost function for q patients is $TC = 380{,}000 + 70q + .04q^2$.

Problem 3.1 (Maximizing Profit)
Phyllis wants to determine how many patients to cover in her practice in order to maximize her profit (π). [Note that economists use the symbol π to denote profit; it is not the number approximated by $22/7$.] In general, $\pi = TR - TC$ where $TR = pq$.

(a) How many private patients should she accept at what annual price p in order to maximize her profit?

(b) Phyllis is considering whether to participate in a government program that will pay her $700 for each government patient she includes in her practice. Use her marginal revenue at maximum profit to show why Phyllis should participate in the government program.

(c) To determine Phyllis' new maximum profit (assuming that her practice can include as many government patients as she wishes) use marginal revenue and marginal cost to calculate how many private patients and how many government patients she should cover.

4 Sam and Stuart Start Up

Sam and Stuart just received their first patent for a new hand-held game system, think *DS* meets *Twitter*. They have established their start-up firm in a warehouse and have not yet determined their fixed costs. Meanwhile, to determine the number q units demanded for their system at price p per unit, they put it on the market in two very similar areas at two different prices: \$720 in one area, and \$500 in the other. As long as they are able to support demand, and there is no shortage in either market, they will be able to use their experiment to establish their demand function. They discover that they sell $q = 1720$ units priced at $p = \$720$, and sell $q = 3480$ units prices at $p = \$500$.

Problem 4.1 (Revenue)

(a) Use their results to express p as a linear function of q. At what price and quantity will Sam and Stuart maximize their total revenue?

(b) Sam and Stuart sketched the graphs of TR and MR on the same axes (with q on the horizontal axis), and noticed that their maximum TR is the area of the triangle whose vertices are (0, 0), (0, 935), (3740, 0). This motivated them to prove that $TR(q)$ is the area under the MR curve from 0 to q. Show how they did this.

(c) Sam and Stuart were wondering: for $q \in [0, 3740]$, is the average TR per unit q the same as the average value of the TR function? What is your answer, and why?

(d) The price elasticity of demand, $e = \frac{\text{\% change in quantity demanded}}{\text{\% change in price}} = \frac{\Delta q}{q} / \frac{\Delta p}{p}$, is the percent change in the number of units demanded per 1% change in price per unit. The price elasticity of demand is said to be *elastic* when $|e| > 1$, *inelastic* when $|e| < 1$ and have *unit elasticity* when $|e| = 1$. Show that e can be approximated at a price p by

$$e = \frac{dq/q}{dp/p} = \frac{dq}{dp} \cdot \frac{p}{q}.$$

Using the fact that $TR = pq$, derive the equation $MR = p(1 + 1/e)$.

Problem 4.2 (Cost)

Sam and Stuart now want to estimate their cost function. They have fixed costs of \$40,000 in rent and overhead. Sam and Stuart believe that when $q = 1000$, marginal cost equals \$200 and that when $q = 5000$, marginal cost is \$680. They also believe that marginal cost can be approximated by a straight line.

(a) Using the information above, how did Sam and Stuart estimate the variable cost function?

(b) At what level of q is average cost a minimum?

Problem 4.3 (Maximizing Profit)

Sam and Stuart are sitting down to discuss profit based on their total cost function and total revenue function.

(a) Show why $MR = MC$ when their profit is maximum. How many units q at what price p yields the maximum profit? What are TC and TR when profit is at a maximum?

(b) To get a better feel for how demand changes as price changes, Sam and Stuart calculated the price elasticity of demand for the values of p and q at which profit is maximized. If they increase their price by 3% per unit, what would be the effect on demand for their product?

(c) Sam and Stuart are wondering if they should engage in an advertising campaign and move into nicer digs. Although the advertising and new rented space will raise their fixed cost to \$120,000, they think that they will have greater profit by selling more units at a lower price based upon the relationship $p = 960 - .1q$. Determine if Sam and Stuart should advertise and move into nicer quarters.

5 Solutions

Problem 2.1

(a) Use the graph of $\frac{d(MC)}{dq} = -12 + .3q$ to describe the graph of MC. Then determine the marginal cost function and show its graph.

$$\frac{d(MC)}{dq} = 0 \text{ at } q = 40, \text{ and } \frac{d(MC)}{dq} \text{ is negative for } p < 40 \text{ and positive for } p > 40.$$

Therefore, the graph of MC is decreasing when $p < 40$, has a horizontal tangent at $q = 40$, and is increasing for $p > 40$. Furthermore, the graph of $\frac{d(MC)}{dq}$ shows that the slope of the tangent to MC's graph at $40 - q$ is the negative of the tangent's slope at $40 + q$. So the graph of MC is symmetric about the vertical line $q = 40$. Therefore, for $q \geq 0$, the graph of MC is the portion of the parabola whose vertex at $q = 40$ is the graph's minimum point.

Since $\frac{d(MC)}{dq} = -12 + .3q$ is the derivative of the marginal cost function,

$$MC = \int \frac{d(MC)}{dq} = -12q + .15q^2 + c$$

for some constant c. Ingrid assumed that $c = \$500$ was the constant she saw on the previous page, and she concluded that $MC = 500 - 12q + .15q^2$.

The graph of MC is the portion of the parabola thorough (0, 500) that has vertex $(40, 260)$ as its minimum point.

(b) Show how Ingrid could use the graph of MC to describe the shape of TC's graph. Then explain how she determined TC.

The marginal cost function is the derivative of total cost function. Since MC is always positive, the graph of TC is always increasing. Furthermore, from the symmetry of MC's graph about $q = 40$, the slope of the tangent to TC's graph at $40 - q$ differs from the slope at $q = 40$ by the same amount as the slope at $40 + q$ differs from the slope at $q = 40$. Since TC must be a cubic, it must have general shape in Figure 1.

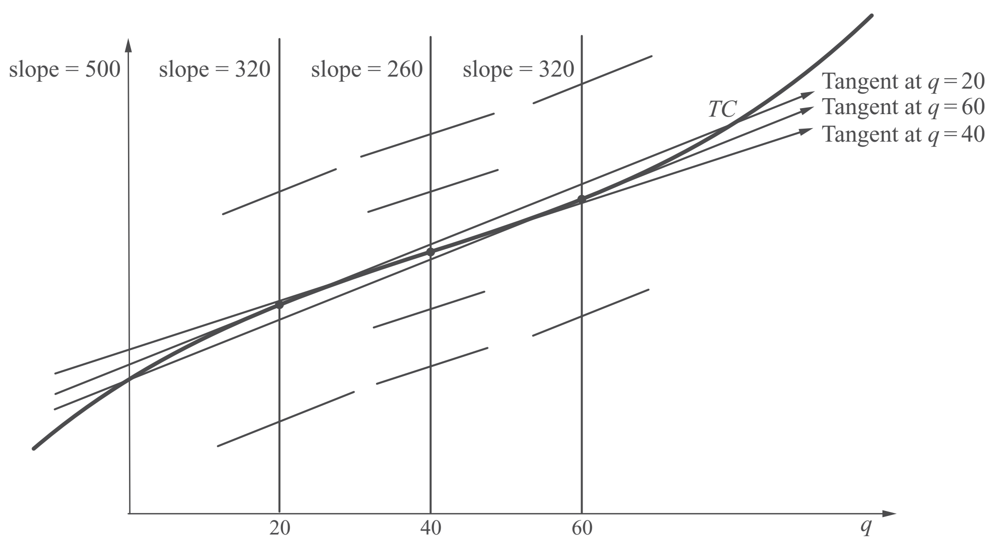


Figure 1. General shape of the graph of TC.

Ingrid integrated the marginal cost function to obtain

$$TC = \int MCdq = 500q - 6q^2 + .05q^3 + c.$$

She again reasoned that $c = \$150{,}000$ is the fixed cost she saw on the previous page, and concluded that

$$TC = \$150{,}000 + 500q - 6q^2 + .05q^3.$$

(c) Assuming that Ingrid's total cost function is correct, her average variable cost function is

$$AVC = \frac{VC}{q} = \frac{500q - 6q^2 + .05q^3}{q} = 500 - 6q + .05q^2.$$

Her sketches of AVC and MC on the same axes show that $AVC = MC$ at the value q where AVC is at a minimum.

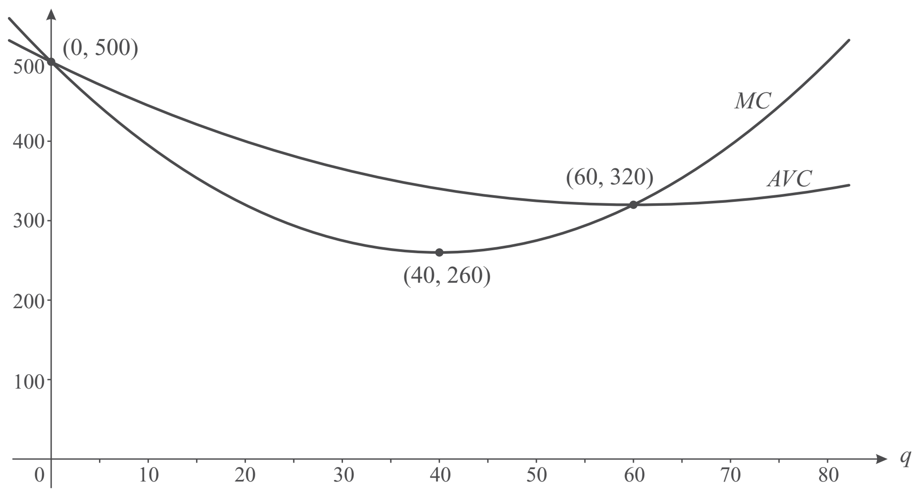


Figure 2. Graphs of AVC and MC.

(d) Ingrid proved that

$$MC - AVC = q \cdot \frac{d(AVC)}{dq}$$

by using the quotient rule to obtain

$$\frac{d(AVC)}{dq} = \frac{\frac{d(VC)}{dq}q - VC}{q^2} = \frac{\frac{d(VC)}{dq} - \frac{VC}{q}}{q} = \frac{MC - AVC}{q}.$$

Ingrid has not shown that AVC is equal to MC at q if and only if AVC has a minimum at q. She has only shown that if AVC has a minimum at q, then AVC is equal to MC. AVC having a minimum is a sufficient condition, but not a necessary one. Thus, AVC is increasing when $MC > AVC$, and AVC is decreasing when $MC < AVC$.

Problem 2.2

(a) Ingrid used the slopes

$$m_1 = \frac{TC(1400) - TC(1300)}{100} = 28 \quad \text{and} \quad m_2 = \frac{TC(1600) - TC(1400)}{200} = 34$$

to estimate that

$$\Delta\pi = 90{,}000 - \Delta TC = 90{,}000 - 2{,}800 = \$87{,}200 \text{ when } q \text{ increases from 1400 to 1500}$$

and

$$\Delta\pi = 90{,}000 - \Delta TC = 90{,}000 - 3{,}400 = \$86{,}600 \text{ when } q \text{ increases from 1500 to 1600.}$$

Her estimates for TC at 1500 are

$$TC(1500) = TC(1400) + 100m_1 = 738{,}000$$

and

$$TC(1500) = TC(1400) + 100m_2 = 738{,}600.$$

(b) The marginal cost is

$$MC = \frac{d(TC)}{dq} = \frac{d}{dq}\left(150{,}000 + 1202q - .84q^2 + .0002q^3\right) = 1202 - 1.68q + .0006q^2.$$

Thus, $MC(1400) = 26$ and $MC(1500) = 32$. By comparison, $m_3 = \frac{TC(1600)-TC(1500)}{100} = 40$.

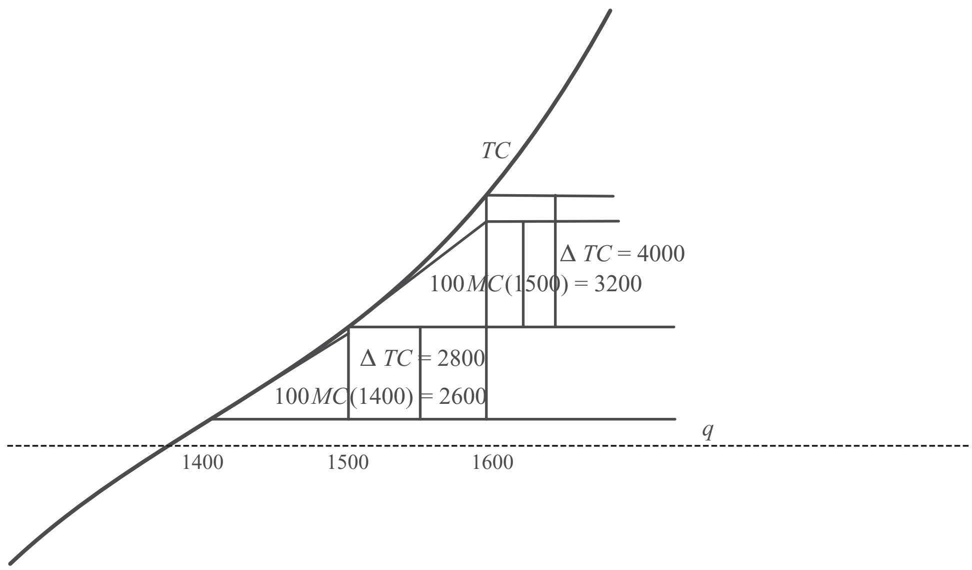


Figure 3. Marginal cost estimations of total costs.

(c) $\Delta TC = 100MC\,(1400) = 2600$ when q increases from 1400 to 1500, and $\Delta TC = 100MC\,(1500) = 3200$ when q increases from 1500 to 1600.

BB's maximum monthly profit is \$87,200 when $q = 1300$ and when $q = 1400$, and *BB*'s minimum monthly profit is \$83,600 when $q = 1600$.

(d) $\Delta TC = TC\,(q + 100) - TC\,(q) = .06q^2 - 162q + 112{,}000$. Solving $\frac{d(\Delta TC)}{dq} = .12q - 162 = 0$, we get $q = 1350$. Therefore, ΔTC has minimum value \$2650, and *BB* has maximum monthly profit \$87,350 when production increases from $q = 1350$ to 1450 units.

Problem 3.1

(a) How many private patients q should she accept at what annual price p in order to maximize her profit?

Profit, denoted by the symbol π, is total revenue minus total cost. Since total revenue equals pq, her profit function is

$$\pi = pq - TC = (1500q - .1q^2) - (380{,}000 + 70q + .04q^2) = 1430q - .14q^2 - 380{,}000.$$

Solving $\frac{d(\pi)}{dq} = 1430 - .28q = 0$ for the integer value closest to the critical value, we get $q = 5107$. Since $\frac{d^2(\pi)}{dq^2} < 0$, the profit is maximum at the critical value. At $q = 5107$, her maximum profit is \$3,271,607 and her total revenue is \$5,052,355. Substituting $q = 5107$ into the demand function, we get the price $p = \$989.30$ that is charged when her profit is maximum.

(b) Phyllis is considering whether to participate in a government program that will pay her \$700 for each government patient she includes in her practices. Use her marginal revenue at maximum profit to show why Phyllis should participate in the government program.

Her MR for private patients must be compared to the government sponsored price of \$700. Since her private revenue function is $TR = pq + 1500q - .1q^2$, her marginal revenue is $MR = 1500 - .2q$. At maximum profit,

$MR(2150) = 1500 - .2(5107) = \$478.60$. Since the government is offering her \$700, which is more than her private MR, she should participate.

(c) To determine Phyllis' new maximum profit (assuming that her practice can include as many patients as she wishes), Phyllis should take on patients as long as she makes a profit for each patient seen. So, she will add private patients until her $MC = 700$, which is the government guaranteed price. Since her $TC = 380{,}000 + 70q + .04q^2$, her $MC = 70 + .08q$. She will see patients until $70 + .08q = 700$, for a total of 7,875 patients.

In order to maximize her overall profit, she should take on private patients until her MR falls to \$700, which is for $q = 4000$ private patients at the price \$1,100 per patient. Then she should switch to adding $7875 - 4000 = 3875$ government patients. Phyllis's total revenue is the sum of her private revenue ($4000 \times \$1100$) and her government revenue ($3875 \times \$700$) for a total of \$7,112,500. Her total cost for 7,875 patients is \$3,411,875, so her maximum profit becomes \$3,700,625.

Problem 4.1

(a) Use their results to express p as a linear function of q. At what price and quantity will Sam and Stuart maximize their total revenue?

To determine the equation of the line $p = aq + b$ passing through the (q, p) points $(1720, 720)$ and $(3480, 500)$, they calculated the slope

$$b = \frac{\Delta p}{\Delta q} = \frac{p_2 - p_1}{q_2 - q_1} = \frac{500 - 720}{3480 - 1720} = \frac{-220}{1760} = -0.125$$

and used $p - 720 = -.125(q - 1720)$ to write $p = -0.125q + 935$. Therefore, their total revenue function is

$$TR = pq = (935 - .125q)q = 935q - .125q^2.$$

To maximize revenue, they set $MR = \frac{d(TR)}{dq} = 935 - .25q = 0$ and obtained the critical value $q = 3740$. Since $\frac{d^2(TR)}{dq^2} = -0.25$ is negative, $TR(3740) = \$1{,}748{,}450$ is the maximum total revenue. Substituting $q = 3740$ into their demand function yielded the price $p = -0.125(3740) + 935 = \$467.50$.

(b) The triangle whose vertices are $(0, 0)$, $(0, 935)$, $(3740, 0)$ has area $(935)(3740)/2 = 1{,}748{,}450 = TC(3740)$.

For a given value $q \in (0, 3740]$, the area under the MR curve from 0 to q is

$$\int_0^q (935 - .25t)\,dt = 935q - .125q^2 = TR(q).$$

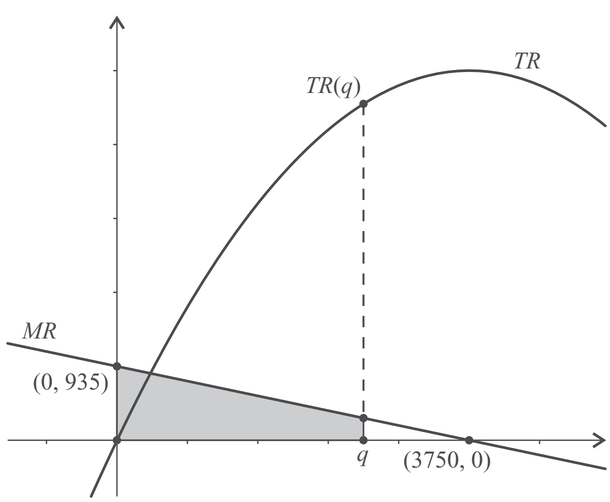


Figure 4. Graphs of TR and MR.

(c) For $q \in [0, 3740]$, the average TR per unit q is \$1,748,450/3,740 = \$467.50. The average value of the TR function is

$$\frac{1}{3740}\int_0^{3740}(935q - .125q^2)dq = \$1{,}165{,}633.$$

This total revenue is obtained approximately by selling $q = 1580$ units at $p = \$737.50$ per unit.

(d) The price elasticity of demand, $e = \frac{\text{\% change in quantity demanded}}{\text{\% change in price}} = \frac{\Delta q}{q} / \frac{\Delta p}{p}$, is the percent change in the number of units demanded per 1% change in price per unit. Since p and q were assumed to be differentiable functions of one another, e can be approximated at price p by

$$e = \frac{dq/q}{dp/p} = \frac{dq}{dp}\cdot\frac{p}{q}$$

because $e = \frac{\Delta q}{\Delta p}\cdot\frac{p}{q}$ and $\frac{\Delta q}{\Delta p}$ is approximated by $\frac{dq}{dp}$ for small Δp.

[It should be noted that if you estimate elasticity over a range (arc elasticity) there can be problems, including different answers depending on which values of p and q are used. So called "point elasticity" as calculated above using the derivative of p with respect to q avoids those problems.]

From $TR = pq$ we get

$$MR = \frac{d(TR)}{dq} = p + q\frac{dp}{dq} = p\,(1 + 1/e)\,.$$

Problem 4.2

(a) Using the information above, how did Sam and Stuart estimate the variable cost function?

$MC = a + bq$, where $b = (680 - 200)/(5000 - 1000) = .12$. Using $MC = 200$ when $q = 1000$ yields $a = 80$. So $MC = 80 + .12q$. Since MC is the derivative of VC, the variable cost is $80q + .06q^2$.

(b) At what level of q is average cost a minimum?

$AC = \frac{TC}{q} = 40{,}000q^{-1} + 80 + .06q$. Solving $\frac{d(AC)}{dq} = -40{,}000q^{-2} + .06 = 0$ for q, we get $q = 816.5$. Since $\frac{d^2(AC)}{dq} > 0$ at $q = 816.5$, the minimum average cost is $AC\,(816.5) = \$178$.

Problem 4.3

(a) Determine how many units q at what price p yield the maximum profit. What are TC and TR when profit is maximum?

Profit (π) is equal to total revenue minus total cost, $\pi = TR - TC$,

$$\frac{d(\pi)}{dq} = \frac{d(TR)}{dq} - \frac{d(TC)}{dq} = MR - MC.$$

Therefore, $MR = MC$ when profit is maximum. In our example, $TR = 935q - .125q^2$ and $TC = 40{,}000 + 80q + .06q^2$. Consequently, $\pi = -.185q^2 + 855q - 40{,}000$ and $\frac{d(\pi)}{dq} = 0$ for $q = 3211$. Since π's second derivative is negative, the maximum profit is \$947,583 at $q = 2311$. At maximum profit, $TR = \$1{,}492{,}906$ and $TC = \$545{,}323$. And the price, from our demand equation, is $p = 935 - .125(2311) \approx \$646$.

(b) To get a better feel for how demand changes as price changes, Sam and Stuart calculated the price elasticity of demand for the values of p and q at which profit is maximized. If they increase their price by 3% per unit, what would be the effect on demand for their product?

The elasticity of demand at the profit maximizing p and q is $e = -8\left(\frac{646}{2311}\right) = -2.24$. This suggests that q will decrease by 2.24% if p increases by 1%. So, if p is increased by 3% we would expect q to fall by 6.72%, from 2311 to 2156.

(c) Sam and Stuart are wondering if they should engage in an advertising campaign and move into nicer digs. Although the advertising and new rented space will raise their fixed cost to \$120,000, they think that they will have greater profit by selling more units at the lower price based upon the relationship $p = 960 - .1q$. Determine if Sam and Stuart should advertise and move into nicer quarters.

Under these circumstance their new *total revenue* function is $TR = 960q - .1q^2$ and their total cost function becomes $TC = 120{,}000 + 80q + .06q^2$. So the profit function is $\pi = TR - TC = -.16q^2 + 880q - 120{,}000$. Since $\frac{d(\pi)}{dq} = 0$ at $q = 2750$, and the second derivative is negative, the maximum profit is \$1,090,000 at $q = 2750$. This is higher than the maximum profit without advertising, so they should advertise. Sam and Stuart also used $q = 2750$ to determine the price $p = \$685$, and confirm that $TR = pq = \$1{,}883{,}750$.

3

Intermediate Macroeconomic Theory

Michael K. Salemi
Department of Economics, University of North Carolina, Chapel Hill

1 Introduction

The organizational chart for course work leading to an undergraduate major in economics has the shape of an inverted pyramid. All students begin by completing a one- or two-semester principles of economics course. Majors next complete three intermediate theory courses including intermediate microeconomics, intermediate macroeconomics, and statistics for economics. In many universities, all three intermediate theory courses are prerequisites for field courses so that students master a common set of tools, vocabulary, and problem solving techniques before specializing.

There is a good bit of variety in the course content of the intermediate macroeconomics theory (IMAC) course. The mainstream course teaches four broadly defined topics. The first topic concerns the construction and interpretation of data that describe the state of the economy. These data include current and constant dollar measures of total output, consumption, investment, government spending, exports and imports. They include measures of labor market outcomes such as employment, the unemployment rate, and the labor force. They also include measures of conditions in financial markets, notably interest rates and money stock measures. And they include price indices such as the consumer price index, the producer price index, and the GDP deflator. IMAC students learn how to collect these data and use them to make and support inferences about the state of the economy. They use mathematics to extract trends from the data, compute growth rates, and combine the data in valid and interesting ways.

The second topic is economic growth. The workhorse model of economic growth is the Solow model named for Nobel Laureate Robert Solow [11]. The model is dynamic and defines the exponential trend line to which an economy will converge over the long run. Students use mathematics to combine the model's constituent equations, to derive the difference equation that characterizes economic growth, to solve that difference equation for its steady state, to describe growth away from the steady state, and to solve a variety of problems.

The third topic is business cycle theory–using models to explain departures of the economy from its long run path. The typical business cycle model includes equations that predict the spending behavior of consumers, investors, the government sector, and foreigners who purchase domestically produced goods. Other equations describe decisions to divide wealth between money and interest-earning assets. Still others describe the demand for and the supply of labor, and equilibrium in the labor market. A business cycle model also includes equations that describe the demand for and supply of foreign exchange, and determine the equilibrium value of the exchange rate, the value of the US dollar in terms of a foreign currency such as the Euro.

Students use mathematics to derive the equations of the model from more primitive equations that characterize optimal economic behavior, and to combine the model equations to derive equations that describe equilibrium in different economic sectors. They use mathematics and graphing skills to explain how changes in exogenous variables affect the equations that characterize sectoral equilibria. Students use mathematics to combine the sectoral equations to obtain the model's solution or reduced form and to predict how a given change to an exogenous variable will affect each endogenous variable[1]. Finally, they use mathematics to design monetary and fiscal policies that will offset the effects of exogenous shocks[2] on output and inflation.

Models that explain business cycles are inherently dynamic because they are based on the assumption that prices adjust slowly and because they allow for temporary unemployment of labor and capital. Students use mathematics to derive equations that explain how the endogenous variables change both in the very short run when prices remain fixed, in the intermediate run when prices begin to adjust toward new long run values, and in the long run when prices have adjusted fully to the shocks that initiated the cycle.

The fourth topic considers issues in macroeconomics that are the subject of public attention. In the recent past, issues have included the question of whether monetary policy should be conducted under the discretion of policy makers or should be bound by a formal policy rule, analyses of the causes and consequences of government debt, the long-run relationship between interest rates and inflation rates, and the relationship between growth in the money supply and the inflation rate.

Issues covered in the IMAC course frequently derive from current research in macroeconomics. Thus, the IMAC course presents alternative hypotheses about household and firm behavior and investigates their implications. The alternative hypotheses frequently involve mathematical optimization and solution techniques for dynamic models. An excellent example is the rational expectations hypothesis that says, heuristically, that decision makers whose behavior is modeled know as much as the model builder. To study the implications of rational expectations requires students to understand and manipulate the formulas for conditional expected values.

This chapter presents applications, exercises, and projects that illustrate the kinds of mathematics students may encounter in the IMAC course. Section 2 sets out assignments that illustrate mathematics that students use in the IMAC course. Section 3 provides additional discussion and a list of standard texts for the IMAC course. Section 4 includes solutions or outlines of solution methods for the assignments in section 2.

2 The Mathematics of Intermediate Macroeconomics

This section illustrates the mathematics that students may encounter in the IMAC course. Depending on the course and the topic, students might use the mathematics in problem sets or encounter it in instructor lectures. Students who enroll in an honors section of the IMAC course or who are working toward a BS degree might work independently with much of the mathematics. Students in most BA programs will require instructor guidance to understand the mathematics.

The first subsection focuses on mathematics used to interpret macroeconomic data. The second covers the functions that economists use to describe production. The third uses difference equations to work with the Solow growth model. The fourth and fifth subsections are related. Subsection 4 asks students to derive linear approximations to nonlinear economic relationships and solve a system of simultaneous linear equations to make predictions about the effects of exogenous shocks on economic variables. Subsection 5 builds on subsection 4 by illustrating a dynamic analysis of business cycles. Finally, subsection 6 presents a game theory approach to monetary policy.

Each section contains three parts. The first part is an assignment that illustrates the mathematics used in conjunction with the economic topic. The second part is an advanced project that illustrates additional mathematical skills that advanced IMAC students would employ. The third part is a set of comments meant to guide the reader.

[1] The words exogenous and endogenous are widely used in economics to describe variables in a system of equations. An endogenous variable is a variable that is determined by the system of equations. That is, the endogenous variables of a system of equations must take on values such that all the equations hold. An exogenous variable is a variable that is determined outside the system of equations. When solving a system of equations, the exogenous variables are treated as constants.

[2] An exogenous shock is defined as a change in an exogenous variable. In a system of equations, an exogenous shock acts like a change in the constant term in one or more equations of the system that results in changes to the values of the endogenous variables that solve the system.

2.1 Growth Accounting

An early topic covered in the typical mainstream intermediate macroeconomic theory course is construction and interpretation of macroeconomic data.

Assignment Figure 1 displays the natural logarithm of real gross domestic product (RGDP) for the U.S. economy between 1947 and the first quarter of 2010.[3] The data are quarterly, meaning there are four observations per year.

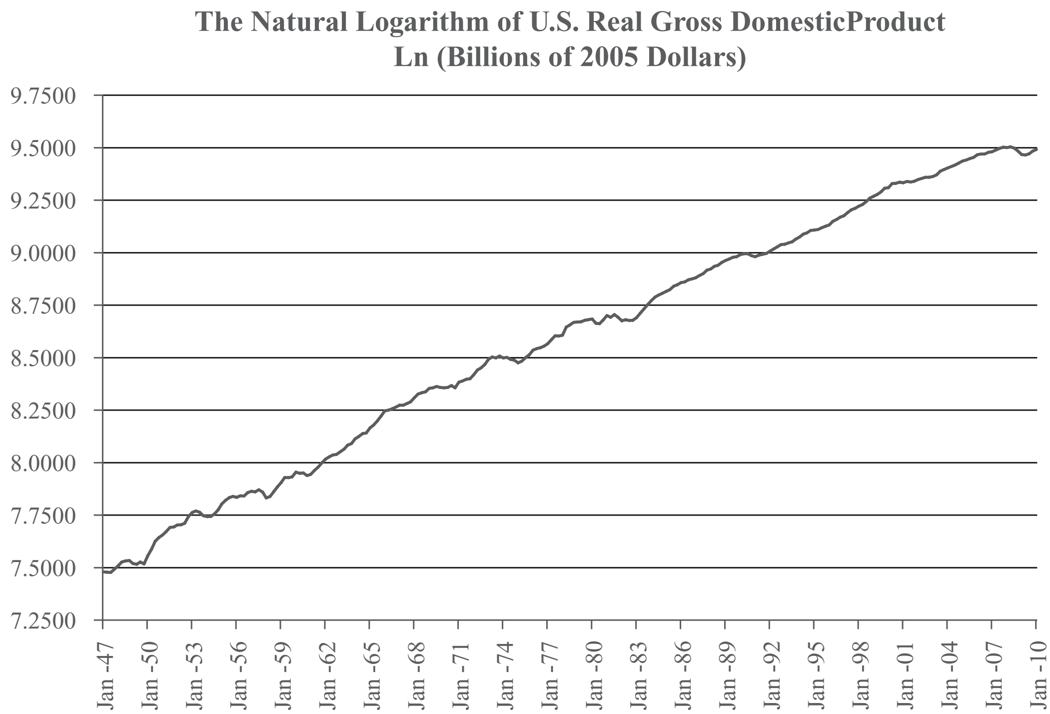


Figure 1

1. An exponential process is defined by the equation $Y_t = G\exp(gt)$, where Y_t is a function of time t, and G and g are fixed parameters. What function of t is $y_t = \ln(Y_t)$?
2. Does the graph of y_t resemble the displayed graph of the natural logarithm of Real Gross Domestic Product? Explain.
3. Since data for real GDP are quarterly (four observations per year), the units of t are quarters (three-month periods). How would you rewrite the exponential process for $Y_t = G\exp(gt)$ so that the units of time are years?
4. A compound growth process is defined by the equation $X_t = (1+h)X_{t-1}$, where the compound growth rate is h. Is an exponential process a compound growth process? Why or why not?
5. Use the information from the graph to estimate the annual compound growth rate of RGDP between the first and last dates of the sample. Explain how you estimated the annual growth rate.

Advanced Project Inspection of the graph presented in the assignment reveals that the data for $y_t = \ln(Y_t)$ do not lie exactly along a straight line. Suppose we allow for the departures of y_t from a straight line with the hypothesis that

[3] IMAC students would obtain these data by connecting to the internet and going to the website of the Federal Reserve Bank of St. Louis (http://www.stlouisfed.org/). Once there they would select "Research and Data" and then "Economic Data-Fred®". They would then download an excel spread sheet containing quarterly data for Real Gross Domestic Product (RGDP).

$y_t = H + gt + u_t$, where u_t is defined to be an error term. An interesting question is: what values of parameters H and g will yield the straight line that minimizes $S = \sum_{t=1}^{T}(u_t)^2$?

1. Show that $S = \sum(y_t)^2 - 2H\sum y_t - 2g\sum ty_t + 2gH\sum t + TH^2 + g^2\sum t^2$.
2. Differentiate S with respect to H and g and derive the conditions that H and g must satisfy if S is to be minimized.
3. Define the sample mean of $(x_1, x_2, \ldots, x_T)$ as $\overline{x} = \frac{1}{T}\sum_{t=1}^{T} x_t$. Show that the first condition you derived in part 2 implies that $H = \overline{y} - g\overline{t}$. What does this value for H imply about $\overline{u}$?
4. Show that the second condition you derived in part 2 implies that $g = \frac{\sum(y_t-\overline{y})(t-\overline{t})}{\sum(t-\overline{t})^2}$. Provide an interpretation for this result.

Comment One of the important competencies that economics majors should develop through the IMAC course is the ability to work with time series data like the series for RGDP used to produce the above graph. Part of that ability requires that students use the the arithmetic of logarithms to compute growth rates, a skill that is widely used in economics. Another part of that ability requires recognition that real GDP can be decomposed as the sum of an exponential-growth process and a residual ($y_t = \ln(Y_t) = H + gt + u_t$). Economists use different models to explain the growth-process component and the error component of RGDP and think of the error component as resulting from business cycle disturbances.

2.2 Production Functions and Labor Demand

The next assignment uses calculus to investigate the properties of a certain class of production functions and to derive a functional relationship between the profit-maximizing amount of labor employment and the variables that characterize the state of the economy.

Assignment Suppose a firm produces output by combining labor and capital. The amount of output obtained for the given input levels of K and L is $Y = AK^{1-\alpha}L^{\alpha}$, where $K > 0$ is the amount of capital employed (measured in constant-value dollars), $L > 0$ is the amount of labor employed (measured in years), $A > 0$ is a parameter that defines the level of technology used by the firm, and $\alpha \in (0, 1)$.

The capital stock of the firm is fixed at the beginning of the decision period. The firm's manager may decide to acquire more capital but newly acquired capital will not become available until a later production period. The only way the manager can change output in the current production period is by changing the amount of labor employed.

The manager operates in competitive markets and has no control over prices. The manager can sell as much output as he or she wants at a market price of P, and can hire as much labor as he or she wants at a wage rate of W. Both P and W are measured in current-year dollars.

1. Show that the production function is homogenous of degree one; that is, show that scaling every input by a factor $\lambda > 0$ changes output by λ.
2. For the given production function, derive the function that gives the marginal product of labor (the rate $\frac{dY}{dL}$ at which output changes as labor input changes).
3. Show that the marginal product of labor depends only the capital-labor ratio.
4. Show that the marginal product of labor is positive and declines as labor employment increases. [Economists call this result diminishing marginal productivity and argue that it is a necessary feature of an efficiently organized production process. The intution is that the firm manager will first assign the most productive activity to the first unit of labor employed, and only then assign less productive activities to subsequent employment units.]
5. The firm's profit is its revenue minus its costs for capital and labor. Derive the function that shows how the firm's profit varies with the amount of labor it employs. Why does it make sense that profit first rises as employment increases but ultimately falls as employment increases?

6. Assume that the firm cannot change the amount of capital it employs. Show that the firm maximizes its profit when it hires the amount of labor such that the marginal product of labor equals W/P, which is called the real wage. [This is an example of what economists call the cost-benefit principle. The benefit of additional employment is the additional output produced. The cost of the additional employment is the wages paid. Dividing the wage rate by the price level measures marginal cost in the same units as marginal benefit. Optimal employment occurs where marginal benefit equals marginal cost.]
7. Show that a profit-maximizing firm that cannot change the amount of capital it employs will employ the quantity of labor given by

$$L = K(\alpha A)^{\frac{1}{1-\alpha}} \left(\frac{W}{P}\right)^{\frac{-1}{1-\alpha}}.$$

Economists call this function a labor demand function. Use the labor demand function to predict how the quantity of labor demanded by the firm varies, other things unchanged, with the real wage ($\frac{W}{P}$), the technology parameter A, and capital employment parameter K. Explain the intuition behind your findings.

Advanced Project The production function given in the assignment is called the Cobb-Douglas function after mathematician Charles W. Cobb and economist Paul H. Douglas [3], who proposed its use to account for the apparent constancy of the ratio of the U.S. total wage bill to U.S. national product. The Cobb-Douglas function is one member of the class of functions that is homogenous of degree one. Another member of that class of functions is $Y = A[\alpha L^{-\beta} + (1-\alpha)K^{-\beta}]^{-1/\beta}$ ($\alpha \in (0, 1)$ and $-1 < \beta \neq 0$), where Y, L, and K are again defined as output, labor input and capital input. Repeat the above assignment using this new production function. For part 3, determine whether or not the marginal product of labor depends only on the capital-labor ratio or individually on the levels of capital and labor employment.

Comment The Cobb-Douglas production function is a workhorse in macroeconomics. The models that students encounter in the IMAC course typically assume that output and employment in the economy evolve through time as if output were produced by a representative firm using a Cobb-Douglas function. Important parts of the Solow growth model and the typical business cycle model use the Cobb-Douglas function. However, some models that students may encounter use the production function given in the advanced project.

2.3 The Solow Growth Model

Economists use the term "endogenous" to describe a variable that is explained by a system of economic equations. They use the term "exogenous" to describe a variable that is determined outside a system of economic equations. Economists use the term "structural" to describe an equation that is invariant to all changes in exogenous variables.

An economic model is a system of structural equations. The solution to the model is a set of m equations, where m is the number of endogenous variables each of which is expressed as a function of the exogenous variables. The set of solution equations (termed the reduced form of the model by economists) is typically obtained by linearizing the structural equations and then solving them simultaneously.

Advanced economics students study conditions under which there is a unique solution to a system of structural equations. In the IMAC course, students encounter only systems of equations for which unique solutions exist. The Solow model is one such system of equations.

The mainstream IMAC course uses the Solow growth model [11] to explain the long run growth paths followed by developed economies. The next assignment illustrates the mathematics of the Solow model.

Assignment Let Y_t, K_t, and N_t be real output, the capital stock, and the population of the work force (measured in worker years) at time $t = 1, 2, \ldots$ In the previous assignment, L was defined as labor employment which is different from N, the work force. Labor may either be unemployed ($L < N$) when the economy is in recession or overemployed ($L > N$) when firms mandate overtime. An assumption of the Solow model is that labor is fully employed which is why N rather than L is used.

Let $k_t = \frac{K_t}{N_t}$ be the capital-labor ratio of the economy and $y_t = \frac{Y_t}{N_t}$ be output per worker year at time t. Finally, let I_t, C_t, and S_t be investment, consumption, and saving in the economy at time t. The structural equations of the Solow model are:

$$Y_t = A_t K_t^{1-\alpha} N_t^{\alpha} \tag{1}$$
$$K_{t+1} = K_t + I_t - \delta K_t \tag{2}$$
$$N_{t+1} = (1+n)N_t \tag{3}$$
$$Y_t = C_t + I_t \tag{4}$$
$$S_t = sY_t \tag{5}$$
$$I_t = S_t \tag{6}$$
$$A_{t+1} = (1+g)A_t. \tag{7}$$

Equation (1) is the production function. Equation (2) explains how the capital stock evolves over time. It says that new investment adds to the capital stock with a one period lag, and that each period the capital stock depreciates at the rate δ. Equation (3) implies that the population of the work force grows at the constant rate n. Equation (4) is the GDP identity that says that output is either consumed or used to build capital. For simplicity, this version of the Solow model assumes that government spending, exports, and imports are all zero. Equation (5) says that national saving is a constant fraction s of national output. Equation (6) requires that investment equals the flow of national saving. Equation (7) says that the production function technology parameter grows at the rate g, allowing for steady improvement in economic productivity.

1. Show that the Solow model production function implies $y_t = A_t k_t^{1-\alpha}$. This result permits the state of the economy to be represented by a single variable, k_t.
2. Show that the Solow model implies that the capital-labor ratio evolves through time according to

$$k_{t+1} - k_t = \frac{sA_t k_t^{1-\alpha} - (\delta + n)k_t}{1+n}.$$

3. Suppose $g = 0$ so that the technology parameter of the production function remains constant through time. Find the value k_{BG} of k_t such that $k_{t+1} - k_t = 0$. Solow called k_{BG} a balanced growth equilibrium because capital, labor, and output all grow at the same constant rate when $k_t = k_{BG}$. At what rate do capital, labor, and output grow if $k_t = k_{BG}$?
4. Again suppose $g = 0$. Show that k_t converges to k_{BG} as $t \to \infty$.
5. Again suppose $g = 0$. Let $c_t = \frac{C_t}{N_t}$ be consumption per worker year. Use the equations of the Solow model to find an expression for the steady state value of c_t (that is, the value c_{BG} to which c_t converges as k_t converges to k_{BG}). Verify that steady state consumption is not monotone in s, the fraction of output that is saved. What features of the model account for this fact?

Advanced Project In the assignment, we assumed a constant technology. Suppose instead that $g > 0$, so that the technology improves at a constant rate.

1. Repeat parts 3, 4, and 5 of the assignment assuming that $g > 0$. It will be helpful to create a new measure of labor input, one that Solow referred to as "labor efficiency units". The new measure of labor is $E_t = (1+g)^t N_t$.
2. Assume that t is measured in years. Suppose that $\alpha = 0.7$, $\delta = 0.15$, $s = 0.05$, $n = 0.01$, and $g = 0.025$. Create an Excel spread sheet that shows how k_t, y_t, and c_t evolve from arbitrary positive initial values to their balanced growth equilibrium values.

Comment Solow originally used differential equations to set out his growth model. Some texts follow suit. Others use difference operators. The difference equation approach makes it more clear how to construct a spread sheet to simulate the model.

2.4 Short-run Model

A short-run macroeconomic model describes equilibrium in the overall economy in a period of time short enough to take the capital stock, the price level, and the wage rate to be constant. There are several different versions of short-run macro models. The assignment presents the traditional Keynesian version for a closed economy (IS-LM) that has appeared in IMAC textbooks for several decades. The model traces back to Keynes [7] as interpreted by Hicks [6]. In the assignment, students use a graphing strategy to solve the IS-LM model. In the advanced project, students derive analytical solutions for the model.

In order to work independently with short run macro models, students must be able to linearize a nonlinear equation and solve a system of simultaneous linear equations.

Assignment Consider a closed economy in which imports and exports are identically zero. Let X_S be the steady state (or long run) value of X as defined by a model like the Solow model. If X is real GDP, then X_S follows an exponential path through time. If, however, X is the capital-labor ratio, X_S will be constant. Let $\widehat{X} = \frac{X - X_S}{X_S}$ be the percentage difference between X and its long run value, X_S. For example, if P is defined as the economy's price level (say the consumer price index), $\widehat{P}$ is defined as the percentage difference between the price level and its steady state value. Thus $\widehat{P} = 0$, if and only if P equals P_S, and $\widehat{P} > 0$ ($\widehat{P} < 0$) if and only if $P > P_S$ ($P < P_S$). The purpose of a short run model is to explain the deviations of economic variables from their steady state values.

Because the short run model does not include dynamic relationships, the time subscript is understood but suppressed to make the notation simpler. The model is short-run in the sense that it predicts how its endogenous variables respond to changes in exogenous variables within one period (a quarter). Because the model is short-run, the inflation rate is constant and expected to remain constant in the near future.

The endogenous variables of the model are $\widehat{C}$, consumption, $\widehat{I}$, investment, $\widehat{Y}$, aggregate output, $\widehat{T}$, taxes, and $\widehat{r}$, the nominal rate of interest. The exogenous variables of the model are $\widehat{G}$, government spending, $\widehat{M}$, the money supply, $\widehat{P}$, the price level, and $\widehat{U}$, an exogenous changes to consumption such as might be caused by a change in consumer confidence[4].

The short-run model has two sectors—the sector that determines balance between investment and saving (IS) and the sector that determines balance between money demand and money supply (LM). Equilibrium occurs when both the IS and LM equations are satisfied. The equations of the IS sector are:

$$\widehat{Y} = \widehat{C} + \widehat{I} + \widehat{G} \tag{8}$$
$$\widehat{C} = \widehat{U} + b(\widehat{Y} - \widehat{T}) \quad (0 < b < 1) \tag{9}$$
$$\widehat{I} = c\widehat{Y} - d\widehat{r} \quad (0 < c < 1 \text{ and } d > 0) \tag{10}$$
$$\widehat{T} = t\widehat{Y} \quad (0 < t < 1). \tag{11}$$

There are four parameters in the equations of the IS sector: b is the rate at which consumption changes as income changes (called the marginal propensity to consume), c is the effect on investment of an increase in production, d is the decrease in investment that results from a one percent increase in the real rate of interest, and t is the rate at which taxes increase with income.

Equation (8) says that the sum of changes to consumption, investment, and government spending must equal changes to total economic output. Equation (9) says that changes in consumption result from changes in disposable (after-tax) income or the exogenous consumption shock. Equation (10) says that changes in investment are positively related to changes in income and negatively related to increases in the interest rate. Equation (11) says that changes in tax revenues are proportional to changes in total income (which given our system of national accounts must equal total output).

The LM sector has a single equation:

$$M = PLY^e r^{-f}, \tag{12}$$

[4] While it would make sense to consider other exogenous shocks such as shocks to investment or taxes, this is not done in order to keep the model simple.

where $L > 0$, $e > 0$, and $f > 0$ are parameters. The right-hand side is the demand for money and the left-hand side is the supply of money. The equation thus requires that the demand for money equal the supply of money. The demand for money increases with output because people use money to facilitate transactions and the volume of transactions increases with output. The demand for money decreases with the interest rate because holding wealth in the form of money means that that wealth cannot earn the higher interest rate available on other assets. In that sense, the interest rate is a cost of holding money. The demand for money increases proportionately with increases in the price level because individuals are presumed to care about the purchasing power of the money they hold, that is, the quantity of goods that the money will buy.

1. Show that the equations of the IS sector can be combined into the following equation, which economists term the IS schedule.
$$\widehat{Y} = \frac{\widehat{U} + \widehat{G}}{1 - b(1-t) - c} - \frac{d}{1 - b(1-t) - c}\widehat{r}. \tag{13}$$
2. Economists typically assume that $0 < 1 - b(1-t) - c < 1$. What is the implication of this assumption for the slope of the IS schedule, $\frac{d\widehat{Y}}{d\widehat{r}}$? What is the implication of this assumption for the effect on the intercept of the IS equation of $\Delta\widehat{U} > 0$?
3. Suppose there is a decrease in consumer confidence ($\Delta\widehat{U} < 0$) but that $\Delta\widehat{G} = \Delta\widehat{r} = 0$. Explain why it makes sense that $\Delta\widehat{Y} < \Delta\widehat{U} < 0$.
4. Suppose $\Delta\widehat{r} = .01$ and $\Delta\widehat{G} = \Delta\widehat{U} = 0$. Explain why $\Delta\widehat{Y} = \frac{-d}{1-b(1-t)-c}(.01)$ rather than $-d(.01)$.
5. The steady state of the money demand equation $M = PLY^e r^{-f}$ is $M_S = P_S L_S Y_S^e r_S^{-f}$. Linearize the money demand function around the steady state values Y_S, r_S and P_S. Show that the money demand equals money supply condition implies the following functional relationship among $\widehat{r}$, $\widehat{M}$, $\widehat{P}$, and $\widehat{Y}$ (termed the LM schedule by economists)
$$\widehat{r} = \frac{e\widehat{Y} - \widehat{M} + \widehat{P}}{f}.$$
6. Suppose $\Delta\widehat{M} = 0.01$ and $\Delta\widehat{Y} = \Delta\widehat{P} = 0$. What, according to the LM schedule, is the change in $\widehat{r}$ required to make money demand equal money supply? Why does it make sense that the interest rate must fall?
7. Suppose $\Delta\widehat{P} = 0.01$ and $\Delta\widehat{Y} = \Delta\widehat{M} = 0$. What, according to the LM schedule, is the change in $\widehat{r}$ required to make money demand equal money supply? Why does it make sense that the interest rate must rise?
8. Construct graphs of the IS and LM schedules, where $\widehat{r}$ is measured on the vertical axis and $\widehat{Y}$ is measured on the horizontal axis. Explain the effect of a change in $\widehat{U}$ on the IS schedule, and the effect of a change in $\widehat{M}$ on the LM schedule.
9. The equilibrium values of $\widehat{r}$ and $\widehat{Y}$ are the values at which the IS and LM schedules intersect. How does a change in $\widehat{U}$ change $\widehat{Y}$ and $\widehat{r}$? How does a change in $\widehat{M}$ change $\widehat{Y}$ and $\widehat{r}$?

Advanced Project

1. Solve the IS and LM schedules algebraically to obtain an analytic expression for $\widehat{Y}$ in terms of the exogneous variables $\widehat{G}$, $\widehat{U}$, $\widehat{P}$, and $\widehat{M}$. Use the solution equation to predict the changes in the equilibrium value of $\widehat{Y}$ that result from a change in (each of) $\widehat{U}$, $\widehat{G}$, $\widehat{P}$, and $\widehat{M}$.
2. Solve the IS and LM schedules to obtain an analytic expression for $\widehat{r}$ in terms of the exogenous variables. Use the solution to predict the changes in the equilibrium value of $\widehat{r}$ that result from a change in (each of) $\widehat{U}$, $\widehat{G}$, $\widehat{P}$, and $\widehat{M}$.
3. Verify that if output and the interest rate respond to exogenous variables in accordance with the analytic expressions derived in parts 1 and 2, then money supply will equal money demand after the shocks occur.

Comment Most IMAC courses use a geometric approach to solve the short-run macroeconomic model. The geometric approach amounts to deriving the IS and LM schedules, finding their intersection, and then showing how changes in the exogenous variables shift the two schedules and alter the point of intersection. Some IMAC courses produce analytic solutions (called reduced form equations) for some of the endogenous variables, as is done in the advanced project.

Most IMAC courses present an extension to the IS-LM model that allows for endogenous changes in imports, exports, the balance of payments and the foreign-currency price of the dollar. For the sake of simplicity and because it does not introduce new mathematical concepts, the extension is not included here.

It is important that students in the IMAC course are able to differentiate among the different concepts of "change" that are used in the course. In some contexts, "change" means a change in a variable over time. In others, "change" means the difference between a variable and its steady state value. In still other contexts, "change" means the change in an endogenous variable predicted by a model to occur if a certain exogenous variable changes.

2.5 Business Cycle Model

This subsection is closely related to the previous one. The typical IMAC course uses a short-run model, like that set out in the previous section, to construct a business cycle model in which prices (and wage rates) are no longer assumed to be exogenous. For a business cycle model, the horizon is longer and the price level, inflation rate, and expected future inflation rate are treated as endogenous variables. While it would be informative to derive the business cycle model explicitly from the short run model, it is not done here because the derivation is tedious and involves the same mathematical techniques illustrated previously. Instead, we directly set out equations of the business cycle model. In order to simplify notation, use of the hat symbol is dropped although variables continue to be defined as percentage differences from their long run values.

Assignment The business cycle model is a system of three equations in six variables: three endogenous variables, one variable controlled by the central bank, and two exogenous shocks. The endogenous variables are all percentage differences of current from steady state values. Y_t, π_t, and π_t^* are output, inflation, and expected inflation at time t. By expected inflation is meant the inflation rate that is expected at time t to occur between periods t and $t+1$. The exogenous variables D_t and S_t are shocks that occur at time t to desired spending and to inflation. D_t represents a change to spending that might result from a change in consumer confidence. S_t represents a change in inflation that might result from changes in the price of a key resource such as oil. Since D_t and S_t are exogenous, their time paths can take any shape at all. However, economists are often interested in studying the effects of one-time shocks ($D_1 = 0.01$ and $D_t = 0$ for $t \neq 1$).

The variable controlled by the central bank is the nominal rate of interest, r_t. The business cycle model abstracts from government spending and tax policy, and focuses on monetary policy. The model explicitly allows for changes in the interest rate designed to damp the business cycle but incorporates government spending and tax changes into the exogenous shock to desired spending.

The model has three parameters. Parameter $\sigma > 0$ is the rate at which inflation increases as output increases above its steady state value. Parameter $\rho > 0$ is the rate at which output decreases as the real rate of interest $(r_t - \pi_t^*)$ increases. The real rate of interest is the true cost of borrowing over the course of a loan contract because it adjusts for changes in prices that are expected to occur over the period of the loan. Parameter $0 < \gamma < 1$ is the rate at which the expected value of future inflation rises as the current inflation rate increases. Economists would use data to estimate parameters σ, ρ, and γ because they are expected to be different for different economies.

The equations of the model are:

$$\pi_{t+1} = \pi_t^* + \sigma Y_t + S_{t+1} \tag{14}$$

$$Y_t = D_t - \rho(r_t - \pi_t^*) \tag{15}$$

$$\pi_t^* = (1-\gamma)\pi_t + \gamma \pi_{t-1}^*. \tag{16}$$

Equation (14) is the Phillips schedule[5], which says that inflation increases when the expected rate of inflation increases, when output rises above its long run value, or when there are shocks to inflation (such as might be caused by shocks

[5] The Phillips schedule is named for economist A. W. Phillips, who noticed a negative correlation between unemployment growth and inflation. It has become common practice to call the equation that explains inflation a Phillips schedule even when the schedule does not have the exact form suggested by Phillips. Negative values of Y are assumed to be positively correlated with changes in unemployment.

to energy prices, farm prices, or unexpected increases in wages). Equation (15) is the aggregate demand schedule. It says that spending and output change relative to their long run values when there are shocks to aggregate spending or when there are changes in the real rate of interest.

Equation (16) explains how expectations about future inflation are formed. The equation says that inflation expectations adapt gradually to actual inflation. The parameter γ measures the speed of the adaptation. When γ is near 1, adaptation is very slow; when γ is near 0, adaptation is rapid.

The central bank (Federal Reserve) is assumed to control the nominal rate of interest.

1. Because the business cycle model is dynamic, the reduced form equations express the current values of the endogenous variables in terms of current and past exogenous variables and past endogenous variables. Show that the reduced form equation for Y_t is given by

$$Y_t = D_t - \rho r_t + \rho\pi^*_{t-1} + \rho(1-\gamma)\sigma Y_{t-1} + \rho(1-\gamma)S_t.$$

2. Suppose that the economy starts out at long run equilibrium $Y_0 = \pi_0 = \pi^*_0 = D_0 = S_0 = 0$, and that there is a one-time shock to desired spending ($D_1 > 0$ and $D_t = 0$ for $t = 2, 3, \ldots$; $S_t = 0$ for $t = 1, 2, \ldots$). Suppose that the Federal Reserve observes D_1 before it chooses the value for r_1. Find the value r_1^+ of r_1 that the Federal Reserve should choose if it wishes to keep output at its long run equilbrium value ($Y_1 = 0$)? Verify that if $r_1 = r_1^+$ and if $r_t = 0$ for $t > 1$, then $Y_t = 0$ for $t = 1, 2, \ldots$.
3. Show that the reduced form equation for π_t is given by

$$\pi_{t+1} = \sigma D_t - \sigma\rho r_t + (1+\sigma\rho)\pi^*_t + S_{t+1}.$$

4. Suppose that the economy starts out at long run equilibrium $Y_0 = \pi_0 = \pi^*_0 = D_0 = S_0 = 0$. Suppose, as in 2, that there is a one-time shock to desired spending and that the Federal Reserve observes D_1 before it chooses the value for r_1. Find the value r_1^{++} of r_1 that the Federal Reserve should choose if it wishes to keep inflation at its long run equilbrium value ($\pi_1 = 0$). Verify that if $r_1 = r_1^{++}$ and if $r_t = 0$ for $t > 1$, then $\pi_t = 0$ for $t = 1, 2, \ldots$.
5. Does the Federal Reserve face a tradeoff between keeping output and inflation at long run values when shocks to spending occur? Why or why not?
6. Suppose that the economy starts out at long run equilibrium $Y_0 = \pi_0 = \pi^*_0 = D_0 = S_0 = 0$ and that there is a one-time shock to inflation ($S_1 > 0$ and $S_t = 0$ for $t = 2, 3, \ldots$; $D_t = 0$ for $t = 1, 2, \ldots$), which the Federal Reserve observes before it chooses the value for r_0[6]. What value of r_0 should the Federal Reserve choose if it wishes to keep output at its long run equilbrium value? What value of r_0 should the Federal Reserve choose if it wishes to keep inflation at its long run equilbrium value? Does the Federal Reserve face a tradeoff between keeping output and inflation at long run values when inflation shocks occur? Why or why not?

Advanced Project In the assignment above, it was assumed that the Federal Reserve changed the rate of interest after observing shocks to aggregate demand and inflation. In this project, it is assumed that the Federal Reserve follows a policy rule that makes the rate of interest a function of output and inflation (presumably because it cannot observe shocks to spending and inflation directly). Assume that Federal Reserve policy has the form

$$r_t = \theta\pi_t + \phi Y_t, \tag{17}$$

where θ is the rate at which r changes in response to changes in π, and ϕ is the rate at which r changes in response to changes in Y.

1. Derive the reduced form equations for Y_t and π_t using the interest rate rule, equation (17), to eliminate values of r_t from the reduced form equations.
2. Suppose $\sigma = 0.25$, $\rho = 0.10$, and $\gamma = 0.50$. Using the reduced form equations you derived in part one and equation (16), create an Excel spread sheet that simulates the dynamic effects of shocks to spending and inflation on output and inflation.

[6] In reality, the Federal Reserve must base r_0 on a prediction of S_1. However, economists frequently study the effects of economic policy in the best-case scenario when the predictions of policy makers are correct.

3. Suppose the objective is to keep Y_t as close to zero as possible. What values for θ and ϕ appear to be appropriate? How do you know?
4. Suppose, alternatively, that the objective of monetary policy is to keep π_t as close to zero as possible. Would different values of θ and ϕ be appropriate? Explain.

2.6 Monetary Policy Game

One issue frequently treated in the IMAC course is the question of whether monetary policy should be conducted under discretion or be governed by a rule. Under a discretion regime, the Federal Reserve treats each situation as different and chooses the path of the nominal interest rate that is best for each situation. Under a rules regime, the Federal Reserve follows the same rule in every situation and makes that rule known to the general public. Many economists believe that a discretion regime creates a bias toward inflation. The Barro-Gordon model [2] explains the source of the bias. In order to simplify notation, I again drop use of the hat symbol although variables continue to be defined as percentage differences from their long run values.

Assignment Rewrite the Phillips schedule (14) presented in the business-cycle-model assignment as

$$Y_t = \frac{1}{\sigma}(\pi_{t+1} - \pi_t^*) - \frac{1}{\sigma}S_{t+1}. \tag{18}$$

The important implications of this equation are that output increases when the inflation rate is greater than expected ($\pi_{t+1} > \pi_t^*$) and that output decreases when there is a positive inflation shock ($S_{t+1} > 0$). The Barro-Gordon model is a version of this Phillips schedule for which time subscripts are supressed.

$$Y = \frac{1}{\sigma}(\pi - \pi^*) - \frac{1}{\sigma}S.$$

Barro and Gordon make several assumptions. First, they assume that the Federal Reserve can directly control the inflation rate (π). This assumption is reasonable because the Fed can, for example, raise the inflation rate by lowering the interest rate and raising aggregate demand. Second, Barro and Gordon assume that private citizens use available information to choose a value for π^* that accurately estimates π. Third, Barro and Gordon assume that the objectives of Federal Reserve policy are given by the objective function:

$$O = Y - \frac{\Psi}{2}\pi^2, \tag{19}$$

where the constant Ψ captures the Fed's relative dislike for inflation. Because the objective function is increasing in Y and decreasing in π^2, the objective function implies that the Federal Reserve likes output growth and dislikes departures of the inflation rate from its steady state. A large value for Ψ implies that the Fed cares more about stabilizing inflation and a small value for Ψ implies that the Fed cares more about stimulating output. If $\Psi = 1$, the Fed cares equally about stimulating output and stabilizing inflation.

Barro and Gordon treat monetary policy as a game played by two players: the Federal Reserve and the private sector. The private sector plays first by choosing π^*. The Federal Reserve plays second by choosing π. Barro and Gordon assume that private citizens know S when they make their choice of π^*, and that the Federal Reserve knows S and π^* when it makes its choice of π.

Barro and Gordon consider two policy regimes. In the rule regime, the Federal Reserve follows a policy rule of the form $\pi = \theta_1\pi^* + \theta_2 S$, where θ_1 and θ_2 are fixed parameters chosen by the Federal Reserve but known by the private sector. In the discretion regime, the Federal Reserve is free to set π to any value without the implication that it will choose the same π if it again finds itself confronting the same values of π^* and S. In the rule regime, the Federal Reserve takes into account the effect of its policy choice on π^*. In the discretion regime, the Federal Reserve ignores the effect of its policy choice on π^*.

First, suppose that the Federal Reserve follows a policy rule.

1. Explain why $\pi^* = \pi$ in this regime.
2. Given that $\pi^* = \pi$, what value of π will maximize the function O?

3. What is the implication for Y if the Federal Reserve follows a policy rule?
   Now, suppose that the Federal Reserve exercises discretion.
4. Determine the value of π that maximizes the Fed's objective function.
5. Explain why the optimal policy for the Federal Reserve in this case is to create inflation.
6. Suppose that over time private agents figure out what the Federal Reserve will do and respond by setting $\pi^* = \frac{1}{\sigma\Psi}$. What are the implications for output and inflation once private agents react in in this way?
7. Explain why policy discretion creates a bias toward inflation.

Advanced Project The Federal Reserve objective function assignment implies that the Federal Reserve likes output increases. Suppose alternatively that the Federal Reserve objective function is

$$O = -\frac{1}{2}Y^2 - \frac{\Psi}{2}\pi^2 \tag{20}$$

so that the Federal Reserve dislikes deviations of output and inflation from steady state values. Again, let Ψ measure the relative importance to the Fed of stabilizing inflation. Repeat the analysis of the assignment with this new objective function. Derive and interpret optimal Federal Reserve policy under rules and discretion. Does discretion create an inflation bias when the Fed's objective function is (20)?

Comment The Barro-Gordon analysis is the easiest way to introduce students to two important areas in macroeconomics: game theory models of policy and control theory models of policy. In an advanced macroeconomic theory course, students typically pursue both of these topics using more advanced texts and more advanced mathematics. The Barro-Gordon analysis also helps students appreciate why Kydland and Prescott [8] recommend that policy authorities follow a policy rule.

3 Discussion

The assignments and advanced projects set out in Section 2 illustrate the mathematical concepts and skills that students encounter in an intermediate macroeconomics course. Students enroll in intermediate macroeconomics after they complete Principles of Economics. While most students enroll in the IMAC course after they decide to major in Economics, some enroll in the course because it is a requirement for Business or International Studies. A few students will take the IMAC course as an elective.

Economics majors are often required to complete one course in calculus at a level below the calculus course designed for mathematics majors. In some cases, majors will have the alternative of taking a mathematics course designed for economics majors. Students pursuing a BS rather than a BA in economics will frequently be required to complete additional mathematics courses. Some economics departments require calculus as a pre-requisite for Intermediate Microeconomics, and Intermediate Microeconomics as a pre-requisite for the IMAC course.

There is a serious disconnect between the mathematical preparation necessary for students pursuing a terminal BA degree with a major in economics and students preparing for Ph.D. training in economics. Students preparing for Ph.D. training should complete courses in multivariate calculus, linear algebra, ordinary differential equations, and advanced calculus (real analysis) at a minimum and would also benefit from courses in discrete mathematics, topology, and functional analysis. The disconnect likely will remain because most undergraduate economics majors view economics as a way to prepare themselves for a career in business or in the financial services industry.

There are many available texts for the IMAC course[7]. Among the leading textbooks are Abel et al. [1], Froyen [4], Gordon [5], and Mankiw [9]. The assignments and advanced projects presented in this chapter more closely follow Mankiw but are more mathematical because they were created as a supplement to Mankiw's text.

Mathematics instructors who wish to contact economists about the use of mathematics in the teaching of undergraduate economics are perhaps best advised to direct their inquiries to the Committee on Economic Education (CEE) of the American Economic Association. The chair of the CEE is always well positioned to route an inquiry to an appropriate economists. The CEE maintains a web site at http://www.vanderbilt.edu/AEA/AEACEE/index.htm.

[7] See http://www.economicsnetwork.ac.uk/books/IntermediateMacroeconomics.htm.

4 Solutions

This section provides solutions to the assignments and suggestions for completing the advanced projects.

Assignment 2.1 (Growth Accounting)

1. $y_t \equiv \ln(Y_t) = \ln(G) + gt$
2. The points depicted in the graph lie near to but not on a straight line the slope of which is the value of g that is computed in part 5 of the assignment.
3. $Y_t = G \exp(\frac{g}{4}4t)$.
4. $Y_t = G \exp(gt) = G \exp(g(t-1) + g) = G \exp(g(t-1)) \exp(g) = \exp(g) Y_{t-1}$. But $\exp(g) \approx (1+g)$ provided that g is a small number. Therefore, an exponential process is a compound growth process.
5. $Y_{n+1} = (1+h)^n Y_1 \Longrightarrow h \approx \ln(1+h) = \frac{\ln(Y_{n+1}) - \ln(Y_1)}{n} = \frac{9.49-7.48}{252} = .0080$. The quarterly growth rate is 0.0080. The annual growth rate is 0.032.

Advanced Project 2.1

1. The result follows in a straight forward way after expanding $(y_t - H - gt)^2$.
2. $\frac{dS}{dH} = -2\sum y_t + 2g \sum t + 2TH = 0$
   $\frac{dS}{dg} = -2\sum (y_t t) + 2H \sum t + 2g \sum t^2 = 0$
3. $H = \overline{y} - g\overline{t}$ follows directly from the definition of the sample mean and $\frac{dS}{dH} = 0$. If H takes on the value that minimizes S, then $\overline{u} = 0$.
4. The numerator of g is the sample covariance between y and t. The denominator is the sample variance of t. For further explanation, please see pages 100–101 in *Topics in Modern Finance*.

Assignment 2.2 (Production Functions and Labor Demand)

1. $A(\lambda K)^{1-\alpha}(\lambda L)^\alpha = \lambda^{1-\alpha+\alpha} AK^{1-\alpha}L^\alpha = \lambda Y$
2. $\frac{dY}{dL} = AK^{1-\alpha}L^{\alpha-1}$.
3. $\frac{dY}{dL} = AK^{1-\alpha}L^{\alpha-1} = A(\frac{K}{L})^{1-\alpha}$.
4. $\frac{dY}{dL} > 0$ because A, K, L are positive and $\frac{d^2Y}{dL^2} = (\alpha - 1)AK^{1-\alpha}L^{\alpha-2} < 0$ because $\alpha \in (0, 1)$.
5. Let Ψ = Profit and R = Cost of a unit of capital. Then

$$\Psi = PY - WL - RK = PAK^{1-\alpha}L^\alpha - WL - RK$$
$$\frac{d\Psi}{dL} = \alpha PAK^{1-\alpha}L^{\alpha-1} - W = \alpha PA\left(\frac{K}{L}\right)^{1-\alpha} - W.$$

   For any W, there is an $L > 0$ small enough so that $\alpha PA(\frac{K}{L})^{1-\alpha} > W$ and $\frac{d\Psi}{dL} > 0$. But as shown in part 4, $\frac{d^2Y}{dL^2} < 0$. Therefore, one can always find L^* such that $\frac{d\Psi}{dL} < 0$ for $L > L^*$. Economists call $P\frac{dY}{dL}$ the value of the marginal product of labor, that is, the rate at which a firm's revenue increases as the firm employs more labor. If the value of the marginal product is greater than the wage, the firm raises profit by employing more labor. Eventually, because of diminishing returns, the marginal product of labor falls to a point where the value of the marginal product equals the wage rate. Additional employment beyond this point lowers profit.
6. $\frac{d\Psi}{dL} = \alpha PAK^{1-\alpha}L^{\alpha-1} - W = 0$ and $\alpha AK^{1-\alpha}L^{\alpha-1} = \frac{W}{P}$.
7. The labor demand function is obtained by algebraic manipulation of the the result obtained in part 6 ($\alpha AK^{1-\alpha}L^{\alpha-1} = \frac{W}{P}$). Since A, K, L are positive and $\alpha \in (0, 1)$, labor demand is increasing in A and K and decreasing in $\frac{W}{P}$. The intuition is that increases in A and K increase the marginal product of labor at any given level of labor employment and raise the value of L such that $\frac{d\Psi}{dL} = 0$. An increase in $\frac{W}{P}$ lowers the value of L such that $\frac{d\Psi}{dL} = 0$ because it requires an increase in the marginal product of labor.

Advanced Project 2.2 The advanced project is completed in a straightforward way by using the new function in place of the Cobb Doublas function.

Assignment 2.3 (The Solow Growth Model)

1. $y_t = \frac{Y_t}{N_t} = \frac{A_t K_t^{1-\alpha} N_t^{\alpha}}{N_t} = A_t K_t^{1-\alpha} N_t^{\alpha-1} = A_t k_t^{1-\alpha}$.
2. The proof requires several steps.

$$
\begin{aligned}
K_{t+1} - K_t &= I_t - \delta K_t \\
K_{t+1} - K_t &= sY_t - \delta K_t \\
\frac{K_{t+1}}{N_{t+1}} - \frac{K_t}{N_{t+1}} &= s\frac{Y_t}{N_{t+1}} - \delta\frac{K_t}{N_{t+1}} \\
k_{t+1} - \frac{k_t}{1+n} &= \frac{s}{1+n}y_t - \frac{\delta}{1+n}k_t \\
k_{t+1} - k_t &= \frac{s}{1+n}y_t - \frac{\delta}{1+n}k_t - \frac{1+n}{1+n}k_t + \frac{1}{1+n}k_t \\
k_{t+1} - k_t &= \frac{sy_t - (\delta+n)k_t}{1+n} \\
k_{t+1} - k_t &= \frac{sA_t k_t^{1-\alpha} - (\delta+n)k_t}{1+n}.
\end{aligned}
$$

3. $k_{t+1} - k_t = 0$ if and only if $sAk_t^{1-\alpha} = (\delta+n)k_t$, hence $k_t = (\frac{sA}{\delta+n})^{1/\alpha} = k_{BG}$. Here A is the constant value of A_t that occurs when $g = 0$. Since N_t grows at the rate n and $\frac{K_t}{N_t}$ is constant, K_t grows at the rate n in a balanced growth equilibrium. Since $y_t = Ak_{BG}^{1-\alpha}$ by part 1, it follows that Y_t also grows at rate n in a blanced growth equilibrium.
4. We have

$$
k_{t+1} - k_t = \frac{sAk_t^{1-\alpha} - (\delta+n)k_t}{1+n} = \frac{(\delta+n)}{1+n}k_t\left[\left(\frac{k_{BG}}{k_t}\right)^{\alpha} - 1\right].
$$

It follows that $k_{t+1} - k_t > 0$ if $k_t < k_{BG}$, and $k_{t+1} - k_t < 0$ if $k_t > k_{BG}$. Therefore, k_t converges to k_{BG} as $t \to \infty$.

5. $c_t = \frac{C_t}{N_t} = \frac{Y_t - I_t}{N_t} = \frac{Y_t - sYt}{N_t} = (1-s)y_t = (1-s)Ak_t^{1-\alpha}$. Therefore, $c_{BG} = (1-s)A(\frac{sA}{\delta+n})^{\frac{1-\alpha}{\alpha}} = (1-s)y_{BG}$, where y_{BG} is the value of y_t when $k_t = k_{BG}$. By inspection, s has two opposite effects on c_{BG}. A higher value for s raises k_{BG} and therefore y_{BG}. However, for a given y_{BG} a higher value of s raises saving and lowers consumption. Solow called the value for s that maximizes c_{BG} the golden-rule saving rate.

Advanced Project 2.3

1. The key to finding the balanced growth equilibrium when $g > 0$ is a change in variable. Define $E_t = (1+g)^t N_t$ to be labor efficiency units. Labor becomes more productive as time passes because the technology parameter in the production function grows at rate g. When $g > 0$, one can show by an argument exactly parallel to that given in parts 2 and 3 that $\frac{K_t}{E_t}$ converges to a fixed value as time passes. It follows that y_t and c_t grow at rate g in a balanced growth equilibrium.
2. The spread sheet program exploits the recursive nature of the Solow model. Given the parameter values and values for A_1 and k_1, the equations of the Solow model allow computation of y_1, c_1, and $k_2 - k_1$ in the first row of the spread sheet. In the second row of the spread sheet, one first computes A_2 and k_2 and then y_2, c_2, and $k_3 - k_2$. One can then use the “drop and drag” feature of a spread sheet to compute as many additional rows as desired.

Assignment 2.4 (Short-run Model)

1. The IS schedule is derived by substituting the equations for consumption and investment into the output accounting equation as follows.

$$\begin{aligned}
\widehat{Y} &= \widehat{C} + \widehat{I} + \widehat{G} = [b(1-t)+c]\widehat{Y} + \widehat{U} + \widehat{G} - d\widehat{r} \\
[1-b(1-t)-c]\widehat{Y} &= \widehat{U} + \widehat{G} - d\widehat{r} \\
\widehat{Y} &= \frac{\widehat{U} + \widehat{G} - d\widehat{r}}{1-b(1-t)-c}.
\end{aligned}$$

2. The parameter restriction $0 < 1 - b(1-t) - c < 1$ implies that $\frac{d\widehat{Y}}{d\widehat{r}} < 0$ so that, other factors unchanged, an increase in the rate of interest lowers the sum of consumption, investment, and government spending. The condition also implies that, other factors unchanged, $\Delta\widehat{U} > 0$ shifts the IS schedule toward higher levels of output.
3. $\Delta\widehat{Y} < \Delta\widehat{U} < 0$ because consumption and investment also decrease in response to $\Delta\widehat{U} < 0$. Economists call this result a "multiplier" effect.
4. If $\Delta\widehat{r} = .01$ and $\Delta\widehat{G} = \Delta\widehat{U} = 0$, then $\Delta\widehat{Y} = \frac{-d}{1-b(1-t)-c}(.01) < -d(.01)$ because the decline in output caused by an increase in interest rates in turn causes a decrease in consumption and an additional decrease in investment. The cummulative result of these decreases is larger than the direct effect of a one percent increase in the interest rate on investment and output.
5. To obtain the desired result, use the Taylor rule to linearize the money market equilibrium condition in the vicinity of the steady state as follows.

$$\begin{aligned}
&\text{Since } M = PLY^e r^{-f}, \\
&M - M_S \simeq LY_S^e r_S^{-f}(P - P_S) + P_S\left[eLY_S^{e-1} r_S^{-f}(Y - Y_S) - fLY_S^e r_S^{-f-1}(r - r_S)\right] \\
&\frac{M - M_S}{M_S} \simeq \frac{P_S}{M_S} LY_S^e r_S^{-f}\left[e\frac{Y - Y_S}{Y_S} - f\frac{r - r_S}{r_S}\right] \\
&\widehat{M} \simeq \widehat{P} + e\widehat{Y} - f\widehat{r} \\
&\widehat{r} \simeq \frac{e\widehat{Y} - \widehat{M} + \widehat{P}}{f},
\end{aligned}$$

   where the fourth step uses the fact that money demand equals money supply in the steady state.
6. If $\Delta\widehat{M} = 0.01$ and $\Delta\widehat{Y} = \Delta\widehat{P} = 0$, then $\Delta\widehat{r} = -\frac{0.01}{f}$ is the change in the interest rate required to increase money demand by one percent.
7. If $\Delta\widehat{P} = 0.01$ and $\Delta\widehat{Y} = \Delta\widehat{M} = 0$, then $\Delta\widehat{r} = \frac{0.01}{f}$. The interest rate must rise sufficiently to offset the increase in money demand caused by a one percent increase in the price level.
8. Of course, it does not really matter which variable is assigned to which axis. But it is traditional in IMAC courses to assign the interest rate to the vertical axis and output to the horizontal axis. The slope of the IS schedule is negative. The slope of the LM schedule is positive. The intersection of the two schedules gives the values for $\widehat{Y}$ and $\widehat{r}$ that jointly satisfy the requirements that investment equal saving and money demand equal money supply. An increase in $\widehat{U}$ causes a parallel shift in the IS schedule toward higher values of $\widehat{Y}$ and has no effect on the LM schedule. An increase in $\widehat{M}$ causes a parallel shift in the the LM schedule toward lower values of $\widehat{r}$ and has no effect on the IS schedule.
9. If $\Delta\widehat{U} > 0$, then $\Delta\widehat{Y} > 0$ and $\Delta\widehat{r} > 0$. If $\Delta\widehat{M} > 0$, then $\Delta\widehat{Y} > 0$ and $\Delta\widehat{r} < 0$.

Advanced Project 2.4

1. To obtain the reduced form equation for $\widehat{Y}$, use the LM schedule to substitute $\widehat{r}$ out of the IS schedule and then collect terms in $\widehat{Y}$. The reduced form equation for $\widehat{Y}$ is

$$\widehat{Y} = \frac{\widehat{U} + \widehat{G} + \frac{d}{f}\widehat{M} - \frac{d}{f}\widehat{P}}{1 - b(1-t) - c + \frac{de}{f}}.$$

   It shows how the equilibrium value of $\widehat{Y}$ varies with $\widehat{G}$, $\widehat{U}$, $\widehat{P}$, and $\widehat{M}$.

2. To obtain the reduced form equation for $\widehat{r}$, use the IS schedule to substitute $\widehat{Y}$ out of the LM schedule and then collect terms in $\widehat{r}$. The reduced form equation for $\widehat{r}$ is

$$\begin{aligned}\widehat{r} = \frac{e\widehat{Y} - \widehat{M} + \widehat{P}}{f} &= \left(\frac{e}{f}\right)\frac{\widehat{U} + \widehat{G} + \frac{d}{f}\widehat{M} - \frac{d}{f}\widehat{P}}{1 - b(1-t) - c + \frac{de}{f}} - \frac{\widehat{M} - \widehat{P}}{f}\\ &= \left(\frac{e}{f}\right)\frac{\widehat{U} + \widehat{G}}{1 - b(1-t) - c + \frac{de}{f}} - \frac{(1 - b(1-t) - c)}{f}\frac{\widehat{M} - \widehat{P}}{1 - b(1-t) - c + \frac{de}{f}}\end{aligned}$$

   It shows how the equilibrium value of $\widehat{r}$ varies with $\widehat{G}$, $\widehat{U}$, $\widehat{P}$, and $\widehat{M}$.

3. For the sake of specificity, let $\Delta\widehat{M} > 0$ and keep G, U, and P at their steady state values so that $\widehat{G}, \widehat{U}, \widehat{P} = 0$. Use the results from parts 1 and 2 to obtain expressions for $\Delta\widehat{Y}$ and $\Delta\widehat{r}$. Use the LM schedule to show that money demand changes by $\Delta\widehat{M}$.

Assignment 2.5 (Business Cycle Model)

1. The derivation requires several steps as follows.

$$\begin{aligned}Y_t &= D_t - \rho(r_t - \pi_t^*)\\ Y_t &= D_t - \rho r_t + \rho[(1-\gamma)\pi_t + \gamma\pi_{t-1}^*]\\ Y_t &= D_t - \rho r_t + \rho(1-\gamma)(\pi_{t-1}^* + \sigma Y_{t-1} + S_t) + \rho\gamma\pi_{t-1}^*\\ Y_t &= D_t - \rho r_t + [\rho(1-\gamma) + \rho\gamma]\pi_{t-1}^* + \rho(1-\gamma)\sigma Y_{t-1} + \rho(1-\gamma)S_t\\ Y_t &= D_t - \rho r_t + \rho\pi_{t-1}^* + \rho(1-\gamma)\sigma Y_{t-1} + \rho(1-\gamma)S_t.\end{aligned}$$

   The second step uses (16) to substitute π_t^* from the aggregate demand schedule. The third step uses (14) to substitute π_t from the second-step equation. The remaining steps involve algebraic simplification. Keep in mind that the final result gives Y_t as a function of exogenous variables r_t and S_{t-1} and past (or lagged) values of endogenous variables Y_{t-1} and π_{t-1}^*. Also keep in mind that r_t is exogenous because it is assumed to be chosen by the Federal Reserve.

2. Given the assumptions of the problem, $Y_1 = D_1 - \rho r_1 = 0$ if $r_1 = \frac{D_1}{\rho}$. The reduced form equation derived in part 1 implies $Y_2 = D_2 - \rho r_2 + \rho\pi_1^* + \rho(1-\gamma)\sigma Y_1 + \rho(1-\gamma)S_2 = \rho\pi_1^*$ by the assumptions of the model and the result that $Y_1 = 0$. But the Phillips schedule (14) and the adaptive expectations schedule (16) together imply that $\pi_1^* = 0$, so that $Y_2 = 0$. The same argument applied recursively shows that $Y_t = 0$, for $t = 3, 4, \ldots$. Thus, the business cycle model implies that if the Federal reserve fully offsets a one-time shock to aggregate demand in the period that the shock occurs, it need do nothing more in future periods in order to keep output at its steady state value.

3. The derivation uses (15) to substitute Y_t from the Phillips schedule (14) as follows.

$$\begin{aligned}\pi_{t+1} &= \pi_t^* + \sigma Y_t + S_{t+1}\\ &= \pi_t^* + \sigma[D_t - \rho(r_t - \pi_t^*)] + S_{t+1}\\ &= \sigma D_t - \sigma\rho r_t + (1 + \sigma\rho)\pi_t^* + S_{t+1}.\end{aligned}$$

4. Given the assumptions of the problem $\pi_1 = 0$, and $\pi_2 = \sigma D_1 - \sigma\rho r_1 = 0$ if $r_1 = \frac{D_1}{\rho}$. Thus, the same value for r_1 offsets the effect of $D_1 > 0$ on both output and inflation. By an argument parallel to that given in part 3, one can show that if the Federal Reserve fully offsets a one-time shock to aggregate demand in the period that the shock occurs, it need do nothing more in future periods in order to keep inflation at its steady state value.
5. The results from parts 3 and 4 together imply that the Federal Reserve does not face a tradeoff between keeping inflation and output at their long run values when shocks to demand occur. The same policy that stabilizes output, stabilizes inflation.
6. This part of the assignment assumes a one-time shock to inflation, $S_1 > 0$. Given the assumptions of the problem, $\pi_1 = S_1 - \sigma\rho r_0 = 0$ if $r_0 = \frac{S_1}{\sigma\rho}$. To offset the effect of an inflation shock in period 1, the Federal Reserve must raise interest rates by the right amount in period 0. But, again given the assumptions of the problem, $Y_0 = -\rho r_0 < 0$ if $r_0 = \frac{S_1}{\sigma\rho}$. The Federal Reserve cannot choose a value for r_0 such that $Y_0 = \pi_1 = 0$. Moreover, the effects of $S_1 > 0$ carry forward into the future both through the direct effect of S_1 on Y_1, and the indirect effects on output and inflation that work through changes in inflation expectations (π_k^*, for $k = 1, 2, \ldots$). Note that the problem is more difficult if the Federal Reserve cannot act at $t = 0$ to offset S_1. In that case, the economy will experience inflation in the first period, and the Federal Reserve will attempt to offset the lasting effects of S_1 on inflation and output. Again the Federal Reserve will face a tradeoff because no setting for the interest rate will achieve both goals of returning output and inflation to their steady state values.

Advanced Project 2.5

1. This task is straightforward although a little tedious. To obtain the reduced form for Y_t, substitute the interest rate rule (17) for r_t into the aggregate demand equation (15) and use the adaptive expectations equation (16) to substitute out π_t^*. Next collect terms in Y_t and π_t and use the Phillips schedule (18) to substitute for π_t. The reduced form equation for π_t is the same equation that was derived in part 3 of the assignment.
2. The key to completing the project is exploitation of the recursiveness of the model. Treat every row of the spread sheet as a time period. Create separate columns for the exogenous shocks, D_t and S_t. One can fill these columns with zeroes and chosen values in one row to simulate the effects of various one-time shocks, or design persistent shocks, or use a random number generator to allow the shocks to be chosen from a given distribution. Create columns for Y_t, π_t, and π_t^*. Set the initial values for $Y_1 = \pi_1 = \pi_1^* = 0$ in the first row of these three columns and then use the reduced form equations to compute Y_2 and π_2 and the adaptive expectations equation to compute π_2^*. It is sometimes convenient to define parameters and use them when computing Y_2, π_2, and π_2^*. Finally, use the drop and drag feature of the spread sheet to simulate the economy for as many periods as desired.
3. This part of the advanced project is an invitatiion to investigate the effect of different values of θ and ϕ on the time paths for Y_t. The experiment only makes sense if you hold the values (or the distribution) of the shocks constant as you vary the parameters of the Federal Reserve policy rule.
4. This part of the advanced project is an invitation to investigate the effect of different values of θ and ϕ on the time paths for π_t. The experiment only makes sense if you hold the values (or the distribution) of the shocks constant as you vary the parameters of the Federal Reserve policy rule.

Assignment 2.6 (Monetary Policy Game)

1. The private sector knows that the Federal Reserve will employ its rule, $\pi = \theta_1\pi^* + \theta_2 S$. So, it uses that rule to choose π^*. Because the private sector observes S before it chooses π^*, it can correctly predict the inflation rate chosen by the Federal Reserve with the result than $\pi^* = \theta_1\pi^* + \theta_2 S$ and $\pi^* = \frac{\theta_2}{1-\theta_1}S$.
2. If $\pi^* = \pi$, then $Y = -\frac{1}{\sigma}S$ so that Federal Reserve policy has no effect on Y. Therefore, inspection of the Federal Reserve objective function (19) implies that the optimal policy is $\pi = 0$.
3. If the Federal Reserve follows a policy rule that permits the private sector to accurately predict the Fed's choice of π, then $Y = -\frac{1}{\sigma}S$. Barro and Gordon assume that S is a zero mean random variable with the implication that Y is also a zero mean random variable. Thus, Barro and Gordon conclude that when the Federal Reserve follows an optimal rule, output will fluctuate randomly around its steady state value.

4. The objective function of the Federal Reserve is
$$O = Y - \frac{\Psi}{2}\pi^2 = \frac{1}{\sigma}(\pi - \pi^*) - \frac{1}{\sigma}S - \frac{\Psi}{2}\pi^2.$$
Furthermore, $\frac{dO}{d\pi} = \frac{1}{\sigma} - \Psi\pi = 0$ for $\pi = \frac{1}{\sigma\Psi}$ and $\frac{d^2O}{d\pi^2} < 0$. Note that the Federal Reserve assumes $\frac{d\pi^*}{d\pi} = 0$ in the policy discretion regime.
5. If the Federal Reserve ignores the effect of its choice of π on π^*, it believes that it can increase output by increasing inflation. Note that $\pi = \frac{1}{\sigma\Psi}$ is the optimal amount of inflation which is decreasing in Ψ.
6. If $\pi^* = \frac{1}{\sigma\Psi}$, then $Y = -\frac{1}{\sigma}S$. If the Federal Reserve continues to choose $\pi = \frac{1}{\sigma\Psi}$, then $O = -\frac{1}{\sigma}S - \frac{\Psi}{2}(\frac{1}{\sigma\Psi})^2 = -\frac{1}{\sigma}S - \frac{1}{2\Psi\sigma^2} < -\frac{1}{\sigma}S$.
7. Policy discretion creates a bias toward inflation because under discretion the Federal Reserve ignores the effects of its policy choices on private sector behavior. As the private sector figures out Federal Reserve policy, it adapts its own choices in a way that blunts the effect of policy on output.

Advanced Project 2.6 It should be clear that the answers to parts 1-3 of the assignment are the same as before. To solve part 4, it is necessary to use the new objective function
$$O = -\frac{1}{2}Y^2 - \frac{\Psi}{2}\pi^2 = -\frac{1}{2\sigma^2}(\pi - \pi^* - S)^2 - \frac{\Psi}{2}\pi^2.$$

It follows from $\frac{dO}{d\pi} = -\frac{1}{\sigma^2}(\pi - \pi^* - S) - \Psi\pi$ and $\frac{d^2O}{d\pi^2} < 0$ that the objective function is maximized at $\pi = \frac{\pi^*+S}{1+\sigma^2\Psi}$. Given the new objective function, the optimal choice of π in the policy discretion regime depends on S and π^*. If $\pi^* = \pi$, then $\pi^* = \pi = \frac{S}{\sigma^2\Psi} > 0$. So, in the discretion regime, there is an inflation bias even though the Federal Reserve prefers $Y = 0$ to $Y > 0$.

References

[1] Andrew B. Abel, Ben S. Bernanke, and Dean Croushore, *Macroeconomics*, Sixth Edition, Prentice Hall, Upper Saddle River, New Jersey, 2009.

[2] Robert J. Barro and David Gordon, "A Positive Theory of Monetary Policy in a Natural Rate Model," *Journal of Political Economy* 91 (1983) 589–610.

[3] Charles W. Cobb, Paul H. Douglas, "A Theory of Production," *American Economic Review* 18 (1928) 139–165.

[4] Richard J. Froyen, *Macroeconomics: Theories and Policies*, Ninth Edition, Prentice Hall, Upper Saddle River, New Jersey, 2009.

[5] Robert J. Gordon, *Macroeconomics*, 11th Edition, Prentice Hall, Upper Saddle River, New Jersey, 2009.

[6] J. R. Hicks, "Mr. Keynes and the Classics; A Suggested Interpretation," *Econometrica* 5 (1937) 147–59.

[7] John Maynard Keynes, *The General Theory of Employment, Interest and Money*, Harcourt, Brace and World, New York, 1964.

[8] Finn E. Kydland and Edward C. Prescott, "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy* 85 (1977) 473–92.

[9] N. Gregory Mankiw, *Macroeconomics*, Seventh Edition, Worth Publishers, New York, 2010.

[10] Phillips, A. W., "The Relationship betwen Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957," *Economica* (1958) 283–99.

[11] Robert M. Solow, "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics* 70 (1956) 65–94.

4

Closed Linear Economies

Warren Page
Professor Emeritus, New York City College of Technology, CUNY

Alan Parks
Lawrence University, Appleton

1 Introduction

A *linear economy* is governed by linear (matrix) equations. A *closed economy* is self-contained in some way – for instance, all goods produced are consumed, or expenditures equal income. Wassily Leontief's fundamental work [2], for which he received the Nobel Prize in Economics in 1973, explains how production interacts with consumption, labor, and prices. We present three models of a closed linear economy and show how each can attain equilibrium (state of stability). In Section 2, the Production Adjustment Model explains how production adjusts to satisfy demand, and the Price Adjustment Model finds prices that balance income and expenditures. Relating the models, we arrive at the *equilibrium principle*: total spent on consumption is equal to the total paid for labor. Section 3 treats the Normalized Leontief Model, a simplified version that combines the previous two models. This section is self-contained and may be dealt with independently of Section 2. Section 4 contains solutions to the problems.

To understand the models, students should be familiar with matrix operations and systems of linear equations in matrix form. To prove the theorems, in the Problems Section, students need to know the Monotone Convergence Theorem (a monotonic sequences converges if and only if it is bounded) and understand the meaning of a convergent infinite series.

Notation A matrix A is said to be *positive* (denoted $A > 0$) if each entry is positive. A is said to be *nonnegative* (denoted $A \geq 0$) if each entry is nonnegative and at least one entry is positive. Therefore, a nonnegative matrix cannot be the zero matrix, which is denoted O. For matrices of the same dimension, $A \geq B$ means $A - B \geq 0$. If A is a $n \times n$ matrix, $A^0 = I$ is the identity matrix. The vector with zero entries is denoted $\mathbf{0}$.

2 The Production Adjustment Model

Our economy consists of distinct goods $G_1, G_2, \ldots, G_n$ that are produced and consumed. The production of each good involves a certain amount of the other goods (used to create the good in question or consumed in the course of production). Furthermore, some of the goods are consumed after production. A *technology* matrix is a nonnegative $n \times n$ matrix $A = [a_{ij}]$, where a_{ij} is the number of units of G_j required in the production of one unit of G_i. So to produce u units of G_i, we need u a_{ij} units of G_j. The j-th column of A lists the amounts of good G_j used in the production of the other goods.

A *consumption vector* is a nonnegative vector $\mathbf{c} = (c_1\, c_2 \ldots c_\text{n})$, where c_i is the amount of G_i consumed (after production). In the Production Adjustment Model, we assume that the vector $\mathbf{c}$ is fixed. *A production schedule* is a nonnegative vector $\mathbf{x} = (x_1\ x_2 \ldots x_n)$, where x_i is the amount of good G_i produced. Given a production schedule $\mathbf{x}$, the demand on good G_j is the amount of G_j needed in the production of all the goods, along with the final consumption of G_j. This is the j-th entry of the vector $\mathbf{x}A + \mathbf{c}$.

Example 1 For $n = 3$, the right-hand side of the equation

$$(x_1 \quad x_2 \quad x_3)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} + (c_1 \quad c_2 \quad c_3) = \left(\sum_{i=1}^{3} x_i a_{i1} + c_1 \quad \sum_{i=1}^{3} x_i a_{i2} + c_2 \quad \sum_{i=1}^{3} x_i a_{i3} + c_3 \right)$$

shows the respective demands on G_1, G_2, G_3.

A production schedule $\mathbf{x}$ is said to be *feasible* if

$$\mathbf{x}A + \mathbf{c} \le \mathbf{x}$$

(the demand on each good is less than or equal to the amount of the good produced); it said to be *stable* if

$$\mathbf{x}A + \mathbf{c} = \mathbf{x}.$$

If we regard $\mathbf{x}$ as unknown, this system of linear equations can be solved by elimination. If there is a unique solution, we can observe if it is nonnegative and whether or not there exists a stable production schedule. There are technology matrices and consumption vectors for which there is no feasible production schedule because the modeled economies are unrealistic.

Example 2

(a) if $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $\mathbf{c} = (5\ \ 6)$, there is no feasible production schedule $\mathbf{x} = (x_1\ x_2)$ because $\mathbf{x}A + \mathbf{c} \le \mathbf{x}$ requires that $3x_2 + 5 \le 0$.

(b) For $A = \begin{pmatrix} \frac{5}{6} & 1 \\ \frac{1}{12} & \frac{1}{3} \end{pmatrix}$ and $\mathbf{c} = (\frac{1}{6}\ \frac{1}{3})$, the condition $\mathbf{x}A + \mathbf{c} \le \mathbf{x}$ is equivalent to requiring that $\mathbf{x} = (x_1 \quad x_2)$ satisfies both inequalities $x_2 \le 2x_1 - 2$ and $x_2 \ge \frac{3}{2}x_1 + \frac{1}{2}$. Therefore, $\mathbf{x} = (5\ \ x_2)$ is feasible for $12.5 \le x_2 \le 14$. Furthermore, $\mathbf{x} = (5\ \ 8)$ is the stable production schedule.

If there is a feasible production schedule, must there be a stable production schedule? Yes, as indicated by the following.

Theorem 1 Suppose A is a $n \times n$ technology matrix and $\mathbf{c}$ is a consumption vector. If there is a feasible production schedule, then there is a stable production schedule

$$\mathbf{y} = \sum_{n=0}^{\infty} \mathbf{c}A^n \text{ (the j-th entry of } \mathbf{y} \text{ is the sum of the j-th entries in the vectors } \mathbf{c}A^n)$$

and $\mathbf{y} \le \mathbf{x}$ for every feasible production schedule $\mathbf{x}$.

If $I - A$ has a nonnegative inverse and $\mathbf{c}$ is a consumption vector, then $\mathbf{x} = \mathbf{c}(I - A)^{-1}$ is a stable production schedule because $\mathbf{x}(I - A) = \mathbf{c}$ can be written $\mathbf{x}A + \mathbf{c} = \mathbf{x}$. In fact, this is the economy's only stable production schedule. We need $(I - A)^{-1}$ to be nonnegative in order to insure that $\mathbf{x} = \mathbf{c}(I - A)^{-1} \ge 0$.

Example 3 For A and $\mathbf{c}$ in Example 2(a), we get $(I - A)^{-1} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{3} \\ -\frac{1}{2} & 0 \end{pmatrix}$ and $\mathbf{x} = \mathbf{c}(I - A)^{-1} = \begin{pmatrix} -\frac{1}{2} & -\frac{5}{3} \end{pmatrix}$. Thus, $\mathbf{x}$ is not a feasible production schedule. For A and $\mathbf{c}$ in Example 2(b), we have $(I - A)^{-1} = \begin{pmatrix} 24 & 36 \\ 3 & 6 \end{pmatrix}$ and the stable production schedule $\mathbf{c}(I - A)^{-1} = (5\ \ 8)$.

When does $I - A$ have a nonnegative inverse? One answer is the following.

Theorem 2 Given a technology matrix A and a consumption vector $\mathbf{c}$, suppose that $\sum_{n=0}^{\infty} \mathbf{c}A^n$ converges to a vector $\mathbf{y} > \mathbf{0}$. Then $I - A$ has a nonnegative inverse, and $\mathbf{y} = \mathbf{c}(I - A)^{-1}$ is the unique stable production schedule.

A feasible production schedule $\mathbf{x}$ may not be stable because some producers make more of their goods than is required. Therefore, it is natural to suppose that they would decrease production from $\mathbf{x}$ to what was required to meet the demand $\mathbf{x}A + \mathbf{c}$. Although the resulting schedule is feasible, it may not be stable, and so the process of cutting back on production might continue. The following theorem illustrates how the economy might converge to the stable production schedule.

Theorem 3 Let A be an $n \times n$ technology and let $\mathbf{c}$ be a consumption vector. If $\mathbf{x}$ is a feasible production schedule, then each vector of the sequence

$$\mathbf{x_0} = \mathbf{x} \quad \text{and} \quad \mathbf{x_k} = \mathbf{x_{k-1}}A + \mathbf{c} \quad (k = 1, 2, \ldots)$$

is a feasible production schedule. As k increases, the vectors $\mathbf{x_k}$ converge to a stable production schedule.

The Price Adjustment Model As before, we have n distinct goods $G_1, G_2, \ldots, G_n$ that are produced and consumed. Each entry a_{ij} of our $n \times n$ nonnegative technology matrix A is the number of units of G_j used in the production of one unit of G_i. We assume that each good has a price: good G_j costs p_j dollars per unit. So the production of one unit of G_i involves spending $a_{ij}p_j$ on G_j. We wish to track the income and expenditures of the producers of each good.

Assume that the production of each good entails some amount of labor. Let $\mathbf{b} = (b_1 \;\; b_2 \ldots b_n)$ be the *labor cost vector*, where b_i is the cost of labor needed to produce one unit of G_i. The cost of producing one unit of G_i is $\sum_{j=1}^{n} a_{ij}p_j + b_i$.

In this model, the economy being closed assumes each unit that is produced is purchased. Thus, each unit of G_i brings p_i in income. We say that the price vector $\mathbf{p} = (p_1 \quad p_2 \ldots p_n)$ is *feasible* if each producer receives enough income to pay for each unit of production.

$$\sum_{j=1}^{n} a_{ij}p_j + b_i \le p_i \qquad \text{for each } i = 1, 2, \ldots, n.$$

The system of inequalities can be written as the matrix inequality

$$\mathbf{p}A^{\mathrm{T}} + \mathbf{b} \;\le \mathbf{p}.$$

A price vector $\mathbf{p}$ is *stable* if income equals expenditures for each producer. In other words,

$$\mathbf{p}A^{\mathrm{T}} + \mathbf{b} \;= \mathbf{p}.$$

If A is a square matrix, $I - A^{\mathrm{T}}$ has a nonnegative inverse if and only if $I - A$ has a nonnegative inverse because $(I - A^{\mathrm{T}})^{-1}$ is the transpose of $(I - A)^{-1}$. Therefore, the proofs of previous theorems hold when A, $\mathbf{c}$, and $\mathbf{x}$ are replaced by A^{T}, $\mathbf{b}$, and $\mathbf{p}$.

By analogy with Theorem 3, suppose A is an $n \times n$ technology matrix and $\mathbf{b}$ is a labor cost vector. If $\mathbf{p}$ is a feasible price vector, then each vector of the sequence

$$\mathbf{p_0} = \mathbf{p} \quad \text{and} \quad \mathbf{P_k} = \mathbf{P_{k-1}}A^{\mathrm{T}} + \mathbf{b} \quad (k = 1, 2, \ldots)$$

is a feasible price vector, and the vectors converge to the stable price vector $\mathbf{b}(I - A^{\mathrm{T}})^{-1}$.

The convergence of the sequence $\mathbf{P_k}$ can be explained by saying that if a producer's income is greater than expenditures, the price can be decreased while still satisfying demand. If each producer adjusts the price to cover costs including the cost of labor, the economy will move naturally to equilibrium.

Equilibrium

To combine the Production Adjustment Model and the Price Adjustment Model, let A be a technology matrix such that $I - A$ has a nonnegative inverse. Then for a given consumption vector and a given labor cost vector, we can show (Problem 2.7) that the total spent on consumption is equal to the total paid for labor.

Problem 2.1

(a) For each technology matrix A and consumption vector $\mathbf{c}$, determine the feasible productions schedules and the stable production schedule(s), or show that there is no stable production schedule.

$$A_1 = \begin{pmatrix} \frac{1}{2} & 1 \\ 0 & \frac{1}{3} \end{pmatrix}, \mathbf{c} = (1 \;\; 2) \quad A_2 = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \mathbf{c} = (0 \;\; 1)$$

$$A_3 = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix}, \mathbf{c} = (2 \;\; 1) \quad A_4 = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{pmatrix}, \mathbf{c} = (1 \;\; 0)$$

(b) For each matrix A in (a), compare the stable production schedule $\mathbf{x}$ with $\mathbf{c}(I - A)^{-1}$ if the inverse exists.

Problem 2.2

(a) Verify that $\mathbf{x_0} = (4 \;\; 1)$ is a feasible production schedule for A_4 and $\mathbf{c} = (1 \;\; 0)$ in Problem 2.1. Then (Theorem 3) determine the stable production schedule $\mathbf{x}$ to which the sequence $\mathbf{x_k} = \mathbf{x_{k-1}}A + \mathbf{c}$ $(k = 1, 2, \ldots)$ converges.

(b) Explain why $\sum_{n=0}^{\infty} \mathbf{c}A_4^n$ converges to a stable production schedule $\mathbf{y}$. Then calculate $\sum_{n=0}^{\infty} \mathbf{c}A_4^n = \mathbf{y}$ to see if $\mathbf{y} = \mathbf{x}$.

(c) Suppose $\mathbf{c}$ is a consumption vector for a technology matrix$A$ such that $I - A$ has a nonnegative inverse. If $\mathbf{y}$ and $\mathbf{x}$ are stable production functions, prove that $\mathbf{y} = \mathbf{x}$. Why does it follow that $\sum_{n=0}^{\infty} \mathbf{c}A^n$ converges to $\mathbf{c}(I - A)^{-1}$?

Problem 2.3 As a practical matter, we are interested in stable production schedules that are positive (all goods are produced). If A is a technology matrix such that $I - A$ has a nonnegative inverse and $\mathbf{c} \geq 0$ is a consumption vector, then $\mathbf{y} = \mathbf{c}(I - A)^{-1}$ is a stable production schedule. Prove that $\mathbf{y} > 0$ if either (i) $\mathbf{c} > 0$ or (ii) $\mathbf{c} \neq \mathbf{0}$ and $(I - A)^{-1} > 0$.

Problem 2.4 (a) For each technology matrix and consumption vector, determine the stable production schedule or show that none exists.

$$A_5 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 1 \end{pmatrix} \quad A_6 = \begin{pmatrix} \frac{1}{5} & 0 & 0 \\ 1 & \frac{3}{2} & 2 \\ 3 & 0 & \frac{2}{3} \end{pmatrix} \quad A_7 = \begin{pmatrix} \frac{1}{5} & 1 & 3 \\ 0 & \frac{1}{2} & 0 \\ 0 & 2 & \frac{1}{3} \end{pmatrix}$$

$$\mathbf{c} = (1 \quad 1 \quad 1) \qquad \mathbf{c} = (4 \quad 0 \quad 4) \qquad \mathbf{c} = (1 \quad 2 \quad 3)$$

(b) Which of the above technology matrices A has $(I - A)^{-1}$ with negative entries? Explain why this does or does not contradict Theorem 2.

Problem 2.5 Suppose $\mathbf{x} = (x_1 \;\; x_2 \ldots x_n)$ is a feasible production schedule for an $n \times n$ technology matrix A and consumption vector $\mathbf{c} = (c_1 \;\; c_2 \ldots c_n)$.

(a) Explain why no diagonal entry of A can be greater than 1.

(b) If A's diagonal has entry $a_{jj} = 1$, explain what this means for the economy's production and consumption. Confirm your explanation algebraically.

Problem 2.6 Let $\mathbf{b} = (4 \;\; 1 \;\; 8)$ be a labor cost vector for each technology matrix and consumption vector in Problem 2.4.

(a) Determine the stable price vector for each technology matrix or show why it does not exist.

(b) For each stable price vector $\mathbf{p}$ in (a), determine if the total spent on consumption $\mathbf{c} \cdot \mathbf{p^T}$ equals the total spent on labor $\mathbf{y} \cdot \mathbf{b^T}$.

(c) Suppose $\mathbf{c}$ is a consumption vector and $\mathbf{b}$ is a labor cost vector for a technology matrix A such that $I - A$ has a nonnegative inverse. If A has a feasible production schedule and a feasible price vector, then A's stable production schedule $\mathbf{y}$ and stable price vector $\mathbf{p}$ satisfy $\mathbf{c} \cdot \mathbf{p^T} = \mathbf{y} \cdot \mathbf{b^T}$.

3 The Normalized Leontief Model

The equilibrium principle of the previous section shows that labor and consumption can be incorporated into an economic model by treating labor as a good: its cost is the cost of this good, and its consumption is the use of labor. The Normalized Leontief model incorporates labor into production, and it adjusts units so that exactly one unit of each good is produced (including one unit of labor).

Assume that we have n goods $G_1, G_2, \ldots, G_n$ and that one unit of each good will be produced and consumed. This means that the economy is closed. Let A be the $n \times n$ matrix with nonnegative entries, whose entry a_{ij} is the number of units of good G_j used in the production of one unit of good G_i. (A is called a *technology matrix*) Since one unit of G_j is consumed,

$$\sum_{i=1}^{n} a_{ij} = 1 \text{ for each } j = 1, 2, \ldots, n.$$

Therefore, the entries of each column of A have sum one.

If p_i is the price of one unit of good G_i, then (since the one unit of G_i produced is consumed) p_i is also the income of the producer of G_i. Since the producer needs to buy the other goods to produce G_i, this producer's expenditure is $\sum_{j=1}^{n} a_{ij} p_j$. Let $\mathbf{p}$ be the $n \times 1$ vector whose entries $p_1, p_2, \ldots, p_n$ are the respective prices for goods $G_1, G_2, \ldots, G_n$. Then the producer's expenditure is the i-th entry of the n $\times$ 1 vector $A\mathbf{p}$. We say that the price vector $\mathbf{p}$ is *feasible* if

$$A\mathbf{p} \leq \mathbf{p} \text{ (income is at least as great as expenditure),}$$

and *stable* if

$$A\mathbf{p} = \mathbf{p} \text{ (income equals spending for each producer).}$$

Given a $n \times n$ technology matrix A, we ask if there is a stable price vector $\mathbf{p}$ with positive entries (that is, if there are real prices that support the production desired). Since the entries of each column of A have sum one, each column of A – I has sum zero. Therefore, the matrix A – I is singular, and $A\mathbf{p} = \mathbf{p}$ for a nonzero vector $\mathbf{p}$. This would be a stable price vector provided that $\mathbf{p}$ has nonnegative entries. But even if $\mathbf{p}$ had nonnegative entries, it doesn't tell us that the prices are positive.

A $n \times n$ matrix M with nonnegative entries is called a *Markov Matrix* if each column's entries have sum one. Since the rows of $M - I$ are linearly dependent, $M - I$ is singular and $M\mathbf{x} = \mathbf{x}$ for a nonzero vector $\mathbf{x}$. A vector $\mathbf{x}$ with nonnegative entries is called a *probability* vector if the sum of its entries is one. The following can be proved [1].

Theorem 4 If all the entries of M^k are positive for a positive integer k, then M^n converges to a matrix Q, where each of Q's columns is the same probability vector $\mathbf{q}$ whose entries are positive. Furthermore, $\mathbf{q}$ is the unique positive probability vector for which $M\mathbf{q} = \mathbf{q}$.

Problem 3.1

(a) Determine a stable price vector $\mathbf{p} = \binom{p_1}{p_2}$ for each technology matrix. For which matrix does the probability vector $\mathbf{q} = \frac{1}{p_1+p_2}\mathbf{p}$ have only positive entries?

$$A_1 = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix} \quad A_2 = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix} \quad A_3 = \begin{pmatrix} 0 & \frac{1}{4} \\ 1 & \frac{3}{4} \end{pmatrix} \quad A_4 = \begin{pmatrix} \alpha & \beta \\ 1-\alpha & 1-\beta \end{pmatrix} \quad \text{for } 0 < \alpha,\ \beta < 1$$

(b) Let Q be a matrix whose columns are the probability vector $\mathbf{q} = \binom{q_1}{q_2}$. Show that if $Q\mathbf{x} = \mathbf{x}$ for a probability vector $\mathbf{x} = \binom{x_1}{x_2}$, then $\mathbf{x} = \mathbf{q}$. Is this true if $x_1 + x_2 \neq 1$?

(c) As n increases, show that A^n converges to Q for each matrix A in (a). Does A_1's vector $\mathbf{q}$ contradict Theorem 4?

(d) Prove that if $\mathbf{p}$ is a stable price probability vector for a technology matrix A in (a), then $\mathbf{p} = \mathbf{q}$.

Problem 3.2

(a) For each technology matrix, determine the stable price probability vectors $\mathbf{q}$. For which matrix does $\mathbf{q}$ have all positive entries?

$$A_5 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{3} & \frac{2}{3} & \frac{3}{4} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{pmatrix} \quad A_6 = \begin{pmatrix} 1 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \quad A_7 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

(b) Let Q be the matrix each of whose columns is $\mathbf{q}$. As n increases, A^n converges to Q for which matrix A in (a)? Do your vectors $\mathbf{q}$ in (a) contradict Theorem 4?

(c) Let $\mathbf{e_j}$ be the column vector that has jth entry 1 and entry zero otherwise. Prove that if a technology matrix A has a diagonal element $a_{ij} = 1$, then A has a stable price vector $\mathbf{e_j}$, and A^n has column $\mathbf{e_j}$ for every integer $n > 1$.

Problem 3.3 Suppose $\mathbf{p_0}$ is a feasible price vector for a technology matrix A, and

$$\mathbf{p_k} = A\,\mathbf{p_{k-1}} \qquad \text{for } k = 1, 2, \ldots$$

(a) For each integer $k \geq 1$, show that $\mathbf{p_k} \leq \mathbf{p_{k-1}}$ and that $\mathbf{p_k}$ is a feasible price vector.

(b) Explain why the sequence $\mathbf{p_k}$ converges to a vector $\mathbf{p}$. Then show that $\mathbf{p} = \lim_{k\to\infty} \mathbf{p_k}$ is a stable price vector.

4 Solutions

Problem 2.1

(a) The feasible production schedules for $A_1 = \begin{pmatrix} \frac{1}{2} & 1 \\ 0 & \frac{1}{3} \end{pmatrix}$ and $\mathbf{c} = (1 \;\; 2)$ are the vectors $(x_1 \;\; x_2)$, where $x_2 \geq \frac{3}{2}x_1 + 3$ for $x_1 \geq 2$. The stable production schedule is $\mathbf{x} = (2 \;\; 6)$.

There is no feasible production schedule $(x_1 \quad x_2)$ for $A_2 = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ and $\mathbf{c} = (0 \;\; 1)$ because the conditions require that both $x_2 \leq 0$ and $x_2 \geq 2$.

The feasible production schedules for $A_3 = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix}$ and $\mathbf{c} = (2 \;\; 1)$ are the vectors $(x_1 \;\; x_2)$, where $\frac{1}{2}x_1 + 2 \leq x_2 \leq \frac{2}{3}x_1 - \frac{8}{3}$ for $x_1 \geq 28$. The stable production schedule is $\mathbf{x} = (28 \;\; 16)$.

The feasible production schedules for $A_4 = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{2} & 1 \end{pmatrix}$ and $\mathbf{c} = (1 \;\; 0)$ are the vectors $(x_1 \;\; x_1 - 2)$, where $x_2 \leq x_1 - 2$ for $x_1 \geq 2$. For $x_1 \geq 2$, every vector $\mathbf{x} = (x_1 \;\; x_1 - 2)$ is a stable production schedule.

(b) The matrix (I – A) does not have an inverse for A_2 or A_4. For A_1 and for A_3, the stable production schedule is $\mathbf{x} = \mathbf{c}(I - A)^{-1}$.

Problem 2.2

(a) The sequence

$$\mathbf{x_0} = (4 \;\; 1), \quad \mathbf{x_1} = \left(2 + \tfrac{3}{2} \;\; 1\right), \quad \mathbf{x_2} = \left(1 + \tfrac{3}{4} + \tfrac{3}{2} \;\; 1\right), \quad \mathbf{x_3} = \left(\tfrac{1}{2} + \tfrac{3}{8} + \tfrac{3}{4} + \tfrac{3}{2} \;\; 1\right), \ldots$$

converges to the stable production schedule $\mathbf{x} = (3 \;\; 1)$.

(b) By Theorem 1, the vector $\sum_{n=0}^{\infty} \mathbf{c}A_4^n = \mathbf{y}$ is a stable production schedule and $\mathbf{y} \leq \mathbf{x}$. Since $\mathbf{c}A_4^n = \left(\frac{1}{2^n} \;\; 0\right)$, we have $\mathbf{y} = \sum_{n=0}^{\infty} \mathbf{c}A_4^n = \left(\sum_{n=0}^{\infty} \frac{1}{2^n} \;\; 0\right) = (2 \;\; 0)$. Thus, $\mathbf{y} \neq \mathbf{x}$.

(c) If $\mathbf{x}$ and $\mathbf{y}$ are stable production schedules, $\mathbf{x} = \mathbf{c}(I - A)^{-1} = \mathbf{y}$.

Problem 2.3 Proof that $\mathbf{c}(I - A)^{-1} > 0$ if either (i) $\mathbf{c} > 0$ or (ii) $\mathbf{c} \neq \mathbf{0}$ and $(I - A)^{-1} > 0$.

(i) If $\mathbf{c} > \mathbf{0}$, then each entry of the vector $\mathbf{c}(I - A)^{-1}$ is positive because each column of $\mathbf{c}(I - A)^{-1}$ contains a positive entry.

(ii) Since $\mathbf{c}$ has a positive entry and each column of $(I - A)^{-1}$ has only positive entries, all entries of $\mathbf{c}(I - A)^{-1}$ are positive.

Problem 2.4

(a) There is no feasible production schedule for A_5 because $x_3 + 1 > x_3$. The stable production schedule for A_6 is $\mathbf{y} = (50 \ \ 0 \ \ 12)$ and the feasible production schedule for A_7 is $\mathbf{y} = \left(\frac{5}{4} \ \ 47 \ \ \frac{81}{8}\right)$.

$$A_5 = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 \\ 1 & \frac{1}{2} & 1 \end{pmatrix}, \quad A_6 = \begin{pmatrix} \frac{1}{5} & 0 & 0 \\ 1 & \frac{3}{2} & 2 \\ 3 & 0 & \frac{2}{3} \end{pmatrix}, \quad A_7 = \begin{pmatrix} \frac{1}{5} & 1 & 3 \\ 0 & \frac{1}{2} & 0 \\ 0 & 2 & \frac{1}{3} \end{pmatrix}$$

$$\mathbf{c} = (1 \quad 1 \quad 1) \qquad \mathbf{c} = (4 \quad 0 \quad 4) \qquad \mathbf{c} = (1 \quad 2 \quad 3)$$

(b) Although $(I - A_6)^{-1}$ has negative entries, this does not contradict Theorem 2 because $\sum_{n=0}^{\infty} \mathbf{c}A_6^n = \mathbf{y} = \big(50 \ 0 \ 12\big)$ is not positive.

Problem 2.5

(a) If A had a diagonal entry $a_{jj} > 1$, then more than one unit of good G_j would be required to produce one unit of G_j, which would not model a realistic economy.

(b) Suppose $a_{jj} = 1$. Since one unit of good G_j is used to produce one unit of G_j, zero amount of good G_j available to produce any other good G_i and additional good G_j for consumption after production. Therefore, $a_{ij}x_i = 0$ for each $i \neq j$ and $c_j = 0$. This explains why A_4 and $\mathbf{c} = (0 \ \ 1)$ in Problem 2.1, and A_5 and $\mathbf{c} = (1 \ \ 1 \ \ 1)$ in Problem 2. 4, have no feasible production schedule.

Problem 2.6

(a) For the labor cost vector $\mathbf{b} = (4 \ \ 1 \ \ 8)$, there is no feasible price vector for A_5 because $p_1 + p_2 + 8 > 0$, and there is no stable price vector for A_6 because $\mathbf{b}(I - A_6^T)^{-1} = (5 \ \ -288 \ \ 69)$. The stable price vector for A_7 is $\mathbf{p} = (75 \ \ 2 \ \ 18)$.

(b) For the closed economy modeled byA_7, $\mathbf{c}$, $\mathbf{b}$, and $\mathbf{p}$, the total spent on consumption $\mathbf{c} \cdot \mathbf{p}^\mathbf{T} = 1(75) + 2(2) + 3(18) = 133$ equals the total spent on labor $\mathbf{y} \cdot \mathbf{b}^\mathbf{T} = \frac{5}{4}(4) + 47(1) + \frac{81}{8}(8)$.

(c) Suppose $\mathbf{c}$ is a consumption vector and $\mathbf{b}$ is a labor cost vector for a technology matrix A such that $I - A$ has a nonnegative inverse. If A has a feasible production schedule and a feasible price vector, then A's stable production schedule $\mathbf{y}$ and stable price vector $\mathbf{p}$ satisfy

$$\mathbf{y}A + \mathbf{c} = \mathbf{y} \quad \text{and} \quad \mathbf{p}A^T + \mathbf{b} = \mathbf{p}.$$

Furthermore,

$$\mathbf{c} \cdot \mathbf{p}^\mathbf{T} = (\mathbf{y} - \mathbf{y}A) \cdot \mathbf{p}^\mathbf{T} = \mathbf{y} \cdot \mathbf{p}^\mathbf{T} - \mathbf{y}A \cdot \mathbf{p}^\mathbf{T} = \mathbf{y} \cdot \mathbf{p}^\mathbf{T} - \mathbf{y} \cdot (\mathbf{p}^\mathbf{T} - \mathbf{b}^\mathbf{T}) = \mathbf{y} \cdot \mathbf{b}^\mathbf{T}.$$

Problem 2.7 (Proof of Theorem 1) Suppose $\mathbf{x}$ is a feasible production schedule for a technology matrix A and consumption vector $\mathbf{c}$.

(a) Show that $\mathbf{c}A^n \le \mathbf{x}(A^n - A^{n+1})$ for each integer $n \ge 0$, and that for any positive integer N

$$\sum_{n=0}^{N} \mathbf{c}A^n \le \mathbf{x}(I - A^{N+1}) \le \mathbf{x}.$$

(b) Explain why the series $\sum_{n=0}^{\infty} \mathbf{c}A^n$ converges to a nonnegative vector $\mathbf{y} \le \mathbf{x}$.

(c) Show that $\mathbf{y}$ is a stable production schedule.

Problem 2.8 (Proof of Theorem 2) Suppose $\sum_{n=0}^{\infty} \mathbf{c}A^n = \mathbf{y} > \mathbf{0}$ for a technology matrix A and consumption vector $\mathbf{c}$. Then (as in Theorem 1) $\mathbf{y}$ is a stable production schedule, and $\mathbf{y} \le \mathbf{x}$ for every feasible production schedule $\mathbf{x}$.

(a) Proof (by contradiction) that $(I - A)^{-1}$ exists.
 (i) Show that if $I - A$ has no inverse, there exists a vector $\mathbf{z}$ with a positive entry such that $\mathbf{z}(I - A) = \mathbf{0}$.
 (ii) Explain why $\mathbf{y} > \alpha\mathbf{z}$ for a small positive constant α. Then show that $\mathbf{y} - \alpha\mathbf{z}$ is a stable production schedule, and that $\mathbf{y} \le \mathbf{y} - \alpha\mathbf{z}$ is a contradiction.

(b) Proof that $(I-A)^{-1} \geq 0$.

(i) Explain why there is a positive integer m such that $\sum_{k=0}^{m} \mathbf{c}A^k > 0$. Let $\mathbf{e}$ be the j-th row of βI for a positive constant β. Explain why β can be chosen so that $\mathbf{e} \leq \sum_{k=0}^{m} \mathbf{c}A^k$.

(ii) For fixed positive integer m and any positive integer n, show that

$$\sum_{i=0}^{n} eA^i \leq \sum_{i=0}^{n}\left(\sum_{k=0}^{m} \mathbf{c}A^k\right)A^i = \sum_{k=0}^{m}\left(\sum_{i=0}^{n} \mathbf{c}A^i\right)A^k \leq \sum_{k=0}^{m} \mathbf{y}A^k.$$

(iii) Explain why $\sum_{i=0}^{\infty} \mathbf{e}A^i = \mathbf{z}$ for some vector $\mathbf{z} \geq \mathbf{0}$. Then show that $\mathbf{z}A = \mathbf{z} - \mathbf{e}$, and why this proves that the j-th row of $(I-A)^{-1}$ has no negative entry. The same reasoning establishes that each row of $(I-A)^{-1}$ is nonnegative.

(c) Prove that the stable production schedule $\mathbf{y} = \mathbf{c}(I-A)^{-1}$ is unique.

Problem 2.9 (Proof of Theorem 3) Let $\mathbf{c}$ be a consumption vector for an $n \times n$ technology matrix A, and let $\mathbf{x}$ be a feasible production schedule.

(a) Show that $\mathbf{x_0} = \mathbf{x}$ and $\mathbf{x_k} = A\mathbf{x_{k+1}} + \mathbf{c}$ $(k = 1, 2, \ldots)$ is a decreasing sequence of feasible production schedules.

(b) Show that $\mathbf{x_n} = \mathbf{x}A^n + \sum_{k=0}^{n-1} \mathbf{c}A^k$ for each $n > 1$. Explain why $\mathbf{x_n}$ converges to a vector $\mathbf{z}$.

(c) Prove that $\mathbf{z} = \lim\limits_{n\to\infty} \mathbf{x}A^n + \sum_{k=0}^{\infty} \mathbf{c}A^k$ is a stable production schedule.

Problem 3.1

(a) Below are the stable price vectors $\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$ and corresponding probability vectors $\mathbf{q}$ for each technology matrix.

$$A_1 = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix} \qquad A_2 = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{2} \end{pmatrix} \qquad A_3 = \begin{pmatrix} 0 & \frac{1}{4} \\ 1 & \frac{3}{4} \end{pmatrix} \qquad A_4 = \begin{pmatrix} \alpha & \beta \\ 1-\alpha & 1-\beta \end{pmatrix}$$

$$\mathbf{p} = \begin{pmatrix} x \\ 0 \end{pmatrix}, \mathbf{q} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \mathbf{p} = \begin{pmatrix} x \\ \frac{3}{2}x \end{pmatrix}, \mathbf{q} = \begin{pmatrix} \frac{2}{5} \\ \frac{3}{5} \end{pmatrix} \quad \mathbf{p} = \begin{pmatrix} x \\ 4x \end{pmatrix}, \mathbf{q} = \begin{pmatrix} \frac{1}{5} \\ \frac{4}{5} \end{pmatrix} \quad \mathbf{p} = \begin{pmatrix} x \\ x \end{pmatrix}, \mathbf{q} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

(b) Let $Q = \begin{pmatrix} q_1 & q_1 \\ q_2 & q_2 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. If $Q\mathbf{x} = \mathbf{x}$, then $\begin{pmatrix} q_1(x_1+x_2) \\ q_2(x_1+x_2) \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Therefore, $\mathbf{x} = \mathbf{q}$ if $x_1 + x_2 = 1$. Note, however, that $Q\mathbf{x} = \mathbf{x}$ for $\mathbf{x} = 2\mathbf{q}$.

(c) As n increases, $A_1^n = \begin{pmatrix} 1 & \sum_{k=1}^{n} \frac{1}{2^n} \\ 0 & \frac{1}{2^n} \end{pmatrix}$ converges to $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$. This does not contradict Theorem 4 because A_1^n does not have all entries positive. Since A_2, A_3^2, and A_4 have all entries positive, Theorem 4 applies.

(d) If $A\mathbf{p} = \mathbf{p}$ for a probability vector $\mathbf{p}$, then $A^n\mathbf{p} = \mathbf{p}$ for $n = 2, 3, \ldots$ Since A^n converges to Q, and $Q\mathbf{p} = \mathbf{q}$, it follows that $\mathbf{q} = \mathbf{p}$.

Problem 3.2

(a) Below are the stable price vectors $\mathbf{p}$ and corresponding probability vectors $\mathbf{q}$ for each technology matrix.

$$A_5 = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{4} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{pmatrix} \qquad A_6 = \begin{pmatrix} 1 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \qquad A_7 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$$\mathbf{p} = \begin{pmatrix} x \\ \frac{5}{2}x \\ 2x \end{pmatrix}, \mathbf{q} = \begin{pmatrix} \frac{2}{11} \\ \frac{5}{11} \\ \frac{4}{11} \end{pmatrix} \qquad \mathbf{p} = \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix}, \mathbf{q} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad \mathbf{p} = \begin{pmatrix} x \\ 1-2x \\ x \end{pmatrix} = \mathbf{q}$$

Therefore, $\mathbf{q}$ is positive for A_5, and for A_7 if $x < \frac{1}{2}$.

(b) Since A_5^2 has all positive entries, A_5^n converges to Q. It can be verified that A_6^n converges to Q. However, $A_7^n = A_7$ if n is odd and $A_7^n = I$ if n is even. Therefore, A_7^n does not converge. The existence of positive stable probability vectors $\mathbf{q}$ for A_7 does not contradict Theorem 4.

Problem 3.3

(a) Since $\mathbf{p_0}$ is a feasible price vector, $\mathbf{p_1} = A\mathbf{p_0} \leq \mathbf{p_0}$, and $\mathbf{p_1}$ is feasible because $A\mathbf{p_1} \leq A\mathbf{p_0} = \mathbf{p_1}$. The same reasoning shows for each k = 2, 3, ... that $\mathbf{p_k} \leq \mathbf{p_{k-1}}$ and $\mathbf{p_k}$ is feasible.

(b) As k increases, the bounded monotone sequence $\mathbf{p_k}$ converges to a vector $\mathbf{p}$. To show that $\mathbf{p}$ is a stable price vector, let $\mathbf{p} = \lim_{k\to\infty} \mathbf{p_k}$. Since $\mathbf{p_k} = A^k\mathbf{p_0}$ for each k = 0, 1, 2, ...,

$$A\mathbf{p} = A(\lim_{k\to\infty} \mathbf{p_k}) = A \lim_{k\to\infty} (A^k\ \mathbf{p_0}) = \lim_{k\to\infty} (A^{k+1}\ \mathbf{p_0}) = \lim_{k\to\infty}\ \mathbf{p_{k+1}} = \mathbf{p}.$$

References

[1] J. Kemeny and J. Snell, *Finite Markov Chains*, Springer-Verlag, New York, 1976.

[2] Wassily Leontief, *The Structure of American Economy: 1919–1939*, Oxford University Press, 2nd edition, New York, 1951.

5

Mathematics in Behavioral Economics

Michael Murray
Department of Economics, Bates College, Lewiston, ME

1 Introduction[1]

Economists and psychologists who have explored the borderland between their two disciplines have created the field of behavioral economics. Economists have long understood that people don't always maximize their satisfaction or profits in the exact fashion that standard economic theory depicts. Behavioral economics has drawn economists' attention to numerous *systematic* deviations of human behavior that conflict with the predictions of standard economic theory[2]. These systematic deviations can often be accounted for by psychological insights into human behavior. Social psychology and cognitive psychology have contributed particularly rich insights to behavioral economics.

This chapter introduces four topic areas in which behavioral economics relies on mathematics: (i) how attitudes about fairness can affect economic behavior; (ii) how people assess probabilities in practice; (iii) how people behave in the face of uncertainty; and (iv) how people make choices in an inter-temporal setting. Section 2 introduces a model of fairness that requires algebraic skills and the notion of a line's slope. Section 3 discusses systematic ways that people's probability assessments violate the mathematical rules of probability, and provides examples of their possible economic consequences. Section 4 contrasts the "prospect theory" posed by behavioral economists for understanding choices under uncertainty with "expected utility theory" often used in standard economics to understand choices under uncertainty. Drawing the insights from each relies on understanding basic features of concave and convex functions, and the concept of "expected value." Section 5 contrasts the traditional way economists formalize people's tendency to discount future costs and benefits relative to immediate costs and benefits with how behavioral economists have formalized discounting. Behavioral economists better capture many people's observed tendency to be more present-biased than traditional formulations allow. Present-biased individuals rely on backward induction to inform their decision-making. Each section's discussion is accompanied by illustrative problems that might be given to undergraduate mathematics students. Section 6 contains solutions of the problems.

[1] I am grateful to Warren Page for his careful critical readings of earlier drafts that much improved the paper. An anonymous reviewer also provided quite valuable improvements.

[2] While non-standard, behavioral economics is not isolated on the fringes of modern economics. In 1978, Herbert Simon received the Nobel Prize in Economic Science for his rich, and often non-standard, work on decision theory. Daniel Kahneman received the Nobel Prize in Economic Science in 2002 for integrating insights from psychology into the study of decision-making under uncertainty. The American Economics Association has also acknowledged the richness of behavioral economists' work. In 1999, Andrei Schleifer was awarded the Association's John Bates Clark Medal (given every other year to the most promising US economist under 40) for his work infusing financial economics with behavioral insights. In 2001, Matthew Rabin received the Clark medal for his theoretical contributions to behavioral economics.

2 Fairness

Standard economic theory assumes that individuals care only about what happens to themselves. Behavioral economics has challenged that assumption with a wide variety of evidence. A particularly simple contradiction of the assumption arises when people play "the ultimatum game" [5]. In this two-player game, one player is given a sum of money, say \$10, with the instruction that she (or he) is to offer any portion of the sum (even none) to the second player. If the second player accepts the offered amount, the \$10 will be divided between the two players in the agreed upon fashion. However, if the second player refuses the offered amount, neither player will receive any money. In the prototypical formulation of this game, both players know the rules of the game, and that they will only play the game once with this particular unknown player. (For the standard economic thinker, the anonymity and one-shot character of the game eliminates inter-personal and strategic elements from the game.)

Standard economic theory predicts that the first player will offer as small a positive sum as possible and that the second player will accept this offer. The theory assumes that the first player knows the second thinks something is better than nothing (and will therefore accept any non-zero offer), so the first player will offer next to nothing so as to keep almost everything. This prediction is falsified in the vast majority of games played. Most offers of less than 20% of the available sum are turned down, and the first player often offers to split the sum equally with the second. The size of the game's initial sum has mattered little for the shares offered or for the shares turned down. See [6] for how the ultimatum game has been played by participants in a wide variety of societies.

Ernst Fehr and Klaus Schmidt [3] propose that the behavior observed in the ultimatum game is due to people being averse to inequalities. Other things equal, more is better than less. So taking as much as possible has some appeal for Player 1. But taking more than half would create an inequality between the players, and this troubles the first player. Player 2's decision is also made more complicated by inequality aversion. Accepting a small offered share of the sum means accepting an inequality, which is distasteful, but more is better than less, other things equal.

Fehr and Schmidt follow common economic practice; they represent people's preferences among outcomes by way of "utility functions." A utility function assigns to every outcome a numerical value that corresponds to the outcome's ranking in the person's preferences. If outcome A is preferred to outcome B, outcome A is assigned a higher utility value; outcomes an individual considers equally desirable are assigned the same utility value. Economists assume that people choose the available outcome that maximizes their utility functions. Fehr and Schmidt treat the difference between ultimatum game players' monetary gains as part of the game's outcomes. They incorporate an aversion to inequality into each person's utility function and then ask what utility-maximizing players of the ultimatum game would prefer to do, given their options.

To formalize these notions, assume X_1 is the amount that Player 1 receives from the game, X_2 is the amount that Player 2 receives from the game, and that each player prefers more for themself, given a fixed amount for the other player. Assume, also, that while each player dislikes inequality, each has greater dislike for an inequality that leaves her or him with less than the other player. Fehr and Schmidt assume that Player 1's utility function is

$$U_1 = X_1 - \beta_1 \max\{X_1 - X_2, 0\} - a_1 \max\{X_2 - X_1, 0\}$$

and Player 2's utility function is

$$U_2 = X_2 - \beta_2 \max\{X_2 - X_1, 0\} - \alpha_2 \max\{X_1 - X_2, 0\}.$$

The assumption of inequality aversion requires that $\beta_1, \beta_2, \alpha_1, \alpha_2$ are positive (so utility decreases with either kind of inequality). The assumption that players dislike inequities favorable to themself less than they dislike inequities favorable to the other player requires that $\alpha_1 > \beta_1$ and $\alpha_2 > \beta_2$. The assumption that each player prefers a larger amount for themself, given a fixed amount for the other person, requires that β_1 and β_2 are both less than one. This follows because when Player i is receiving more than the other player j, and X_i increases by one with no change in X_j the other person's X, then U_i increases by $(1 - \beta_1)$.

In the ultimatum game, in which players will divide amount G if Player 2 accepts the fraction s of G, offered by Player 1, then either $X_2 = sG$ and $X_1 = (1 - s)G$ because Player 2 accepts Player 1's offer, or $X_1 = X_2 = 0$ because Player 2 rejects Player 1's offer. In what follows, we'll see what offers Player 2 would accept from Player 1, and what offer Player 1 would make knowing what Player 2 would accept.

Problem 2.1

(a) What is Player 1's utility in terms of G and s if Player 2 accepts $sG < \frac{1}{2}G$?
(b) What is Player 1's utility in terms of G and s if Player 2 accepts $(1-s)G > \frac{1}{2}G$?
(c) Use the results of (a) and (b) to argue that Player 1 will always prefer that Player 2 accepts an offer of $sG < \frac{1}{2}G$ rather than $(1-s)G > \frac{1}{2}G$.

Problem 2.2 What utility level does each player receive if Player 2 declines Player 1's offered share?

Problem 2.3

(a) What level of utility does Player 2 get if Player 2 accepts an offered share s?
(b) Express in term of α_2 and β_2, the fraction s at which Player 2 would be indifferent between accepting an offer sG and rejecting that offer.
(c) Does a utility maximizing Person 2 need to know G to decide whether or not to turn down an offer?

Problem 2.4

(a) Given knowledge of what Player 2 will do in the ultimatum game, express in terms of α_1 and $\beta_1$the share of G a utility-maximizing Player 1 will offer to a utility-maximizing Player 2.
(b) Does the fraction s of G that a utility-maximizing Player 1 will offer to a utility-maximizing Player 2 depend on the value of G? If not, what does Player 1's offer depend on?

3 Probabilities

In standard economic theory, rational actors are assumed to use probabilistic information in accord with the laws of mathematical probability. Most importantly, rational actors incorporate new information in accord with Bayes's rule

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}.$$

Students of behavioral economics need to understand probability in order to appreciate how psychology can enrich standard economic models of behavior in the face of uncertainty. Behavioral economists offer considerable evidence that indicates people do not follow Bayes's rule when incorporating new information into their decision making. For example, students are generally startled to learn how a positive finding on an HIV test logically alters the probability that a randomly chosen college student is infected with the HIV virus. About one college student in a thousand in the United States is infected with the HIV virus. Suppose a randomly chosen college student was tested using a test that is correct 95% of the time, in the sense that 95% of infected students would test positive, and 95% of uninfected students would test negative. When asked what percent of randomly chosen student is infected with the HIV virus if the student tests positive, many students surmise 95%. Other students offer lower guesses, such as 75% or 85%. Almost none guess correctly: slightly less than 2%. Bayes's Rule yields

$$P(\text{infected} \mid \text{test positive}) = \frac{P(\text{infected})\ P(\text{test positive} \mid \text{infected})}{P(\text{test positive})}.$$

Since

$$P(\text{test positive}) = P(\text{infected})\ P(\text{positive} \mid \text{infected}) + P(\text{uninfected})\ P(\text{positive} \mid \text{uninfected}),$$

we get

$$P(\text{infected} \mid \text{test positive}) = \frac{(.001)(.95)}{(.001)(.95) + (.999)(.05)} = .018664.$$

Students, like many other people, focus too much on the test's properties and too little on the prior probability of being infected. Behavioral theorists refer to this propensity as the tendency to pay too little attention to "base rates." When students are asked to consider how many "false positives" there would be in a sample of 100,000 randomly chosen students (5% of 99,900, or 4,995) in comparison to the number of true positives in such a sample (.95*100, or 95), the "fewer than 2%" result begins to make intuitive sense.

Cognitive scientists have demonstrated that our brains' untutored responses to new information tend to be at odds with Bayes's rule because Bayes's rule is in tension with some human cognitive processes. Our mental processing of new information is not conducted independently of what we expect to see. For example, if we think A is true, we are apt to mistakenly view evidence that A is not true as evidence that A is true. As for assessing conditional probabilities, cognitive psychologists have found that both the strength of recently acquired information and the power of frequently "rehearsed" older memories matter for how people weight information when making conditional probabilistic assessments.

It is not surprising that people incorporate new probabilistic information in ways that are at variance with Bayes's rule when they are faced with new probabilistic information. More surprising are simpler errors in probabilistic reasoning that psychologists have found people tend to make. For example,

$$P(A \text{ and } B) \leq P(A)$$

seems intuitively accessible to most students. Nonetheless, many people fail to maintain this logic. For example [11], students were told that "Linda is 31 years old, single, outspoken and very bright; she majored in philosophy; as a student, she was deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations." When asked to rank the following statements by their probability of being correct, using 1 for the most probable and 8 for the least probable,

_____ a) Linda is a teacher in a primary school;
_____ b) Linda works in a bookstore and takes Yoga classes;
_____ c) Linda is an active feminist;
_____ d) Linda is a psychiatric social worker;
_____ e) Linda is a member of Women Against Rape;
_____ f) Linda is a bank teller;
_____ g) Linda is an insurance salesperson;
_____ h) Linda is a bank teller and is an active feminist,

a substantial majority of students rank "Linda is a bank teller and an active feminist" more likely than "Linda is a bank teller." They often do so because the latter seems "more representative" of who Linda is and hence, more likely to be true about her.

Another common probabilistic reasoning error occurs when people think samples are more representative than they really are of the probabilistic processes that give rise to the samples. If a baseball player with a batting average of .300 goes 0 for 4, people will think "the player is due for a hit"; that is, the sample data should tend to look like the population from which it is drawn. This same mode of thinking will lead people to think a coin is "unfairly" balanced when it shows 4 heads in a row; that is, the population should look like the sample even when the sample is small. In essence, such people believe in a false "law of small numbers" that is not really a law at all.

Behavioral economists have sought to develop models that capture the styles of people's probabilistic thinking. One assumption in such modeling is that people are "quasi-Bayesian" in their thinking; that is, they make some systematic errors, but otherwise correctly apply Bayes's rule. Mathew Rabin [9] models the false law of small numbers. Rabin assumes that people mistakenly assess samples as being taken from finite populations *without replacement*. Rabin's model implies the two common types of errors illustrated above. First, if people have a strong view about a population's distribution, they will expect small samples to look like the population. This leads people to expect tails to come up after a few heads appear in the toss of a coin picked up on the street. Second, people uncertain about a population's distribution will infer too much from a small sample of observations. For example, people who think in this way will too strongly assess as especially skilled stockbrokers who over a short period perform better than others.

Problem 3.1 Suppose 25% of cabs in New York City are green (G) and 75% are yellow (Y). Isabella (I) witnessed a hit-and-run accident on Fifth Avenue in which a cab injured a pedestrian. She testified in court that the cab she saw was green. The court performed reliability tests on Isabella's ability to identify the color of a cab in similar circumstances. The tests revealed that 20% of the time she mistakenly perceived a green cab when the cab was yellow and 10% of the time she mistakenly perceived a yellow cab when the cab was green. What is the probability that the cab involved in the hit-and-run accident is green?

(b) If there are 10,000 cabs in New York City, how many cabs would Isabella identify as green?

Problem 3.2 Suppose that half of stockbrokers are as likely to be wrong (W) as to be right (R) in any one decision $[P(W) = .5 = P(R)]$, and that the other half of stockbrokers are right sixty percent of the time on any one decision $[P(R = .6)$ and $P(W) = .4]$. For ease of computations, think of the stockbrokers as drawing randomly either from an urn $U(2R, 2W)$ containing two right and two wrong decisions or from an urn $U(3R, 2W)$ containing three right and two wrong decisions. Now consider a particular stockbroker making two right decisions in a row (denote this RR).

(a) Assume that this stockbroker's decisions are made independently from one another (in essence, made with replacement). Given the two right decisions that have been observed, what is the probability that this broker began with the urn $U(2R, 2W)$?
(b) Assume instead that the stockbroker randomly draws decisions without replacement from either urn. Given the two right decisions that have been observed, what is the probability that this broker began with the urn $U(2R, 2W)$?
(c) Suppose it is incorrect to think of a stockbroker's decisions as being made without replacement. How do (a) and (b) illustrate a consequence of the false law of small numbers?
(d) Returning again to assuming that decisions are made with replacement, what is the probability that this stockbroker's next decision will be right given the two right decisions that have already been observed?
(e) Returning instead to assuming that decisions are made without replacement, what is the probability that this stockbroker's next decision will be right given the two right decisions that have already been observed?
(f) Suppose it is incorrect to think of the decisions as being made without replacement. How do (d) and (e) illustrate a consequence of the false law of small numbers?

4 Decisions Under Uncertainty

Many economic decisions are made under conditions of uncertainty. Absent uncertainty, economists commonly assume that individual's choose the available outcome that they deem most preferred. Economists represent people's preferences among outcomes by way of "utility functions." A utility function assigns to every outcome a numerical value that corresponds to the outcome's ranking in the person's preferences. If outcome A is preferred to outcome B, outcome A is assigned a higher utility value; outcomes an individual considers equally desirable are assigned the same utility value. Individuals are assumed to make choices that maximize their utility functions.

When outcomes are uncertain, economists usually model people's choices by assuming that individual's choose the course of action that maximizes the mathematical expectation of the utilities they would receive as the uncertainty is resolved. This standard approach to behavior under uncertainty is called "expected utility theory." However, people often act in ways at odds with expected utility theorys' predictions. Behavioral economists argue that "prospect theory," invented by Kahneman and Tversky [7], better describes people's behavior. This section explores both expected utility theory and prospect theory.

Expected Utility Theory

In 1738, Daniel Bernoulli offered the expected utility approach to explain a paradox posed by his brother Nicholas, in 1713. (See [2] for a translation of Daniel's article.) In the paradox, which is now called the St. Petersburg Paradox, Nicholas asked how much one should be willing to pay to participate in the following gamble. The gambler tosses a coin repeatedly and is paid 2^n dollars, where n is the number of tosses required to obtain the first head. Although the expected value of this gamble is infinite,

$$E(G) = \sum_{n=1}^{\infty} 2^n P\left(1^{\text{st}} \text{ head occurs on } n^{\text{th}} \text{ toss}\right) = \sum_{n=1}^{\infty} 2^n \left(\frac{1}{2}\right)^n = \infty,$$

intuition suggests that people might not pay very much to undertake the gamble. Why would people forego an infinite expected payoff if charged a modest fee for taking the gamble? Daniel's answer was that the utility people obtain from wealth grows at a decreasing rate with increasing wealth; that is, there is *diminishing marginal utility of wealth*. (An extra \$100 now has less value to Bill Gates or Warren Buffett from when they were students.)

For a suitable utility function, the utility of wealth $u(2^n)$ decreases as the wealth 2^n increases so that the expected utility from this gamble,

$$E(u\,(G)) = \sum_{n=1}^{\infty} u(2^n)\left(\frac{1}{2}\right)^n,$$

is finite.

Daniel Bernoulli's expected utility theory received a tremendous boost in economists' esteem when John von Neumann and Oscar Morgenstern [12] provided a set of plausible assumptions under which a rational decision maker would make decisions in accordance with the expected utility model. A key prediction of their work was that an individual with a concave utility function would be "risk averse", that is, such an individual would prefer a sure thing worth X to a gamble with expected payoff X. For economics students at the elementary or intermediate level, it suffices to know that the utility function in Figure 1 is concave because every line segment connecting two points on the function's graph, such as A and B, will lie below the graph. (For a convex function, all such line segments would lie above the function's graph and the person would be "risk seeking", that is, the person would prefer a gamble with expected value X to a sure thing worth X.)

Demonstrating to students that expected utility models can capture both risk averse and risk seeking behavior is one of the early exercises in the undergraduate study of economic decision making under uncertainty. Most students need to be shown that the graph of a utility function can be used to depict the expected value and the expected utility of a gamble.

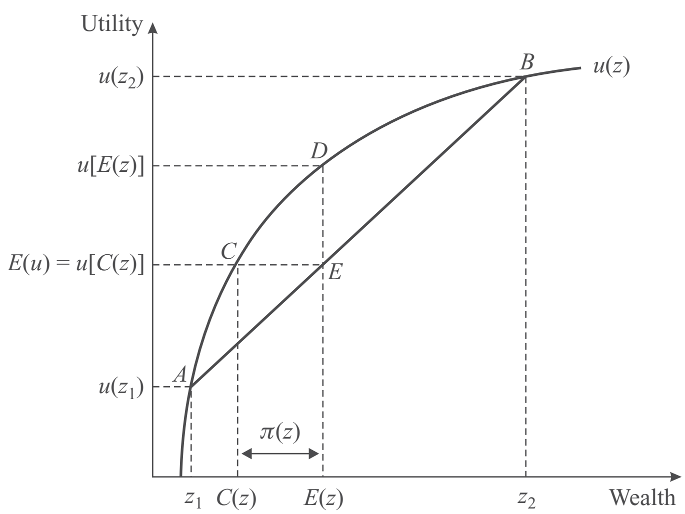


Figure 1. A concave utility function $u(z)$.

Consider a gamble with two possible mutually exclusive outcomes, the wealth of z_1 occurring with probability $(1 - p)$ and the wealth of z_2 occurring with probability p. The expected value of the wealth resulting from the gamble is

$$E(z) = (1 - p)z_1 + pz_2,$$

and the expected utility from the gamble is

$$E(u(z)) = (1 - p)u(z_1) + pu(z_2).$$

If one were assured of wealth equal to $E(z)$, with no gamble involved, the resulting utility from that wealth would be $(E(z))$. For each $p\varepsilon[0, 1]$, the point E with the coordinates $(E(z), E[u(z)])$ lies on the line segment AB. This follows from

$$\frac{E(u(z)) - u(z_1)}{E(z) - z_1} = \frac{u(z_2) - u(z_1)}{z_2 - z_1}$$

and $E(z) - z_1 = p(z_2 - z_1)$ for each $p\varepsilon[0, 1]$.

Figure 1 also shows that for a concave utility function, the gambler's expected utility $E(u(z))$ is less than the utility from a "sure" wealth equal to $E(z)$. Define $C(z)$ to be the level of wealth that yields the same expected utility as the gamble with expected value $E(z)$. Thus, a person with this utility function would be indifferent between a sure wealth of $C(z)$ and the expected value of the gamble between z_1 and z_2. Economists call $C(z)$ the "certainty equivalent" of the gamble. Economists use the term "risk averse" to describe a person who is indifferent between the lesser sure wealth $C(z)$ and the gamble with greater expected value $E(z)$. Economists also refer to $\pi(z) = E(z) - C(z)$ as the "risk premium" that a risk averse person would require in order to accept the shown gamble's expected wealth $E(z)$ over a sure thing.

A convex utility function describes a risk-seeking person who prefers a gamble over a sure equal expected value. Economists also have considered utility functions having regions of concavity and regions of convexity.

Problem 4.1 Suppose an expected utility maximizing individual's utility function for wealth is given by the square root of wealth.

(a) Which of two gambles would this person most prefer: Gamble A that offers a fifty percent chance of wealth equal to $1,440,000 and a fifty percent chance of wealth equal to $160,000, or Gamble B that offers a fifty percent chance of wealth equal to $640,000 and a fifty percent chance of wealth equal to $360,000?
(b) Would this person prefer the gamble chosen in (a) or a sure wealth of $700,000?
(c) Does the person's choice in (b) evidence risk aversion? Briefly explain.

Problem 4.2 Suppose $u(z) = \log(z)$ is an expected utility maximizing individual's utility function for wealth z.

(a) What is the utility associated with $100 wealth?
(b) What is the utility associated $10,000 wealth?
(c) What is the expected wealth from a gamble that would with equal probability result in $100 wealth or wealth $10,000?
(d) What is the certainty equivalent of this gamble?
(e) What is the risk premium associated with this gamble?

Problem 4.3 Demonstrate geometrically that if the utility function is convex, an expected utility maximizing individual will choose a gamble with expected wealth of z over a sure wealth of z. Such an individual is "risk seeking."

Prospect Theory

Behavioral economists argue that "prospect theory," invented by Kahneman and Tversky, better describes people's behavior. Expected utility theory is both elegant and tractable, but a large body of behavioral data indicates that people do not act in accord with expected utility theory. For example, when faced with a choice between two highway safety plans, one of which costs more than the other, people often differ in what they decide between the two plans depending on whether the outcomes from the two plans are contrasted in terms of lives saved or in terms of an equivalent number of lives lost. In general, such "framing" of choices often matters for the choices made – in conflict with the expected utility theory that assumes people are unaffected by how a problem is framed because they always see the consequences of their choices in terms of expected well-being (wealth or utility). However, in practice, people often react to how they frame immediate circumstances and options, without much regard to their overall well-being.

Kahnemann and Tversky, and others, have demonstrated many people's systematic tendency to react more negatively to an amount lost than they respond positively to an equal sized gain. Such "loss aversion" cannot be accounted for in expected utility theory where a "rational actor" always attends to the utility of a level of wealth – it doesn't matter whether that wealth arrives as a loss or gain. Differential treatment of losses and gains can arise only when people attend to the specific circumstances of a particular choice rather than to that choice's consequences for the individual's overall well-being. "Rational actors" don't frame choices in such a piecemeal fashion, and hence can't display loss aversion. Kahneman and Tversky built their "prospect theory" around framing and loss aversion.

Prospect theory specifies the outcomes facing an individual as involving losses or gains, where loss and gain is whatever the individual says they are in a given situation. For example, a homeowner might frame the decision to buy home insurance as "I lose \$100 with certainty by buying the insurance, or I risk a loss of \$10,000 from a fire with probability p, and no loss with probability $(1 - p)$." Prospect theory defines a "value function" v(x) over the domain of both losses and gains (positive and negative values of x). "Value" is close kin to "utility," but value applies to the losses and gains foreseen in a specific situation. This is in contrast to expected utility theory that focuses on the utility from total wealth or total utility viewed across the entirety of one's circumstances. For the homeowner thinking about home insurance, the consumer would focus on the utility from initial total wealth minus \$100 for the choice to insure, and on two possibilities for the choice to remain uninsured: (i) the utility from initial wealth less \$10,000 if there is a fire, and (ii) the utility from initial wealth if there is no fire.

Empirical evidence supports two features of value functions in prospect theory:

(i) the value function is concave in the gain domain and convex in the loss domain.

(ii) the value function is steeper approaching the zero point of no gain or loss from below (*i.e.*, from the loss domain) than it is approaching that point from above (*i.e.*, from the gain domain.)

The resulting function, concave in the gain domain and convex in the loss domain, is illustrated in Figure 2.

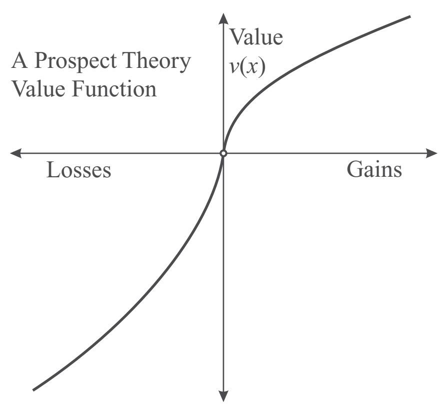


Figure 2. A prospect theory value function, v.

In expected utility theory, the possible outcomes are given weights equal to their probability of occurrence when forming the expected value of utility. Empirical evidence has shown that when weighting losses or gains, people do not use probabilities as weights for outcomes. That same evidence reveals that people assign to a highly likely outcome less weight than the outcome's probability of occurring, and assign to a very low probability outcome more weight than its probability of occurring. Prospect theory captures these phenomena by assuming people weight outcomes by a function $\pi(p)$ of the outcomes' probabilities p. In prospect theory, the function $\pi(p)$ has the following properties:

1. $\pi(p)$ is a continuous increasing function.
2. π (0) $= 0$ and (1) $= 1$.
3. $\pi(p) > p$ for p close to 0 (small probabilities are overweighted).
4. $\pi(p) < p$ for p close to 1 (large probabilities are underweighted).
5. $\pi\ (p) + \pi(1 - p) < 1$ for $0 < p < 1$.

The first, third, and fourth traits imply that $\pi(p) = p$ for some p. Empirically, this seems to be around $p = .5$, with $\pi(p)$ close to p from about .3 to .6. Because $\pi(p)$ can differ from p, the concavity or convexity of the value function doesn't always suffice to guarantee risk aversion or risk seeking, respectively; this feature of prospect theory is explored in Problem 4.5.

Consider an individual who chooses among n gambles, each having two possible outcomes, x_{1i} and x_{2i} ($i = 1, 2, \ldots, n$), whose respective probabilities of occurring are p_{1i} and $(1 - p_{1i})$. The numbers x_{1i} and x_{2i} are negative for losses and positive for gains. According to prospect theory, the individual will choose a gamble for which

$$f(x_{1i}, x_{2i}, p_{1i}) \equiv \pi(p_{1i})v(x_{1i}) + \pi(1 - p_{1i})v(x_{2i})$$

is largest, where $\pi(p)$ refers to a function of p (not the numerical constant pi). If multiple gambles share the largest value for $f(x_{1i}, x_{2i}, p_{1i})$, the individual will choose one from amongst those gambles. A sure thing, x_{1i} can be viewed as a gamble with $x_{2i} = 0$ and probability one for x_{1i} so that $f(x_{1i}, 0, 1) = v(x_{1i})$.

Prospect theory accounts for the framing of options and observed behavior in the face of risk. It also captures people's demonstrated tendency to overweight low probability events and underweight high probability events when making decisions.

Problem 4.4 An outcome x is positive for gain and negative for loss. Suppose the value function is $v = \sqrt{x}$ in the gain domain and $v = -2\sqrt{|x|}$ in the loss domain.

(a) What gain is needed to compensate for a loss of \$20?

(b) Consider a gamble in which one is equally likely to lose 4 or 36. If $\pi(.5) = .5$, would an $f(x_1, x_2, p_1)$-maximizing individual prefer the gamble to a sure loss of 20, the expected loss from the gamble? Is this risk-seeking behavior? Why or why not?

(c) Suppose the probability of losing 36 was .001 and the probability of losing 4 was .999. Also assume $\pi(.001) = .1$ and $\pi(.999) = .8$. Would the individual, according to prospect theory, prefer the gamble to a sure loss of 4.032, which is the expected loss from the gamble? Is this individual risk-seeking? Why or why not?

Problem 4.5 Buying home insurance entails a sure loss of the insurance premium. Forgoing home insurance has a very likely outcome of no loss at all and a very small probability of a large loss if the house burns down. Consider a consumer offered "actuarially fair" insurance (*i.e.*, insurance whose cost equals the expected loss from the covered risk, which is the probability of a loss times the magnitude of the loss).

(a) Use geometry and some algebra to explain why the convexity of the value function in the loss domain implies that an individual who maximizes the *expected* value of the value function would forgo the insurance; that is, would engage in risk-seeking behavior.

(b) According to prospect theory, individuals assess risky opportunities with known probabilities not by calculating the *expected value* of the value function associated with the risky outcomes, but by calculating $f(x_1, x_2, p_1)$. Explain how, despite the convexity of the value function in the loss domain, prospect theory can account for a homeowner avoiding risk by purchasing actuarially fair home insurance. In short, how can prospect theory account for the risk aversion that leads homeowners to buy actuarially fair insurance against the small chance of large losses from fire or such?

5 Inter-temporal Decision Making

Should a firm forgo some profits today to invest in machinery that will yield profits in the future? Should a consumer put money aside today in order to consume more after retirement? Economists have long studied inter-temporal allocation questions like these. The standard economic model recognizes that, other things the same, people like their pleasures sooner and their displeasures later. Economists say that firms and consumers "discount" future benefits and costs when comparing those benefits and costs to present opportunities. But the standard economics model requires that such discounting between any two adjacent periods be done at one same rate for the individual (the rate can vary across individuals). In 1937, Paul Samuelson introduced such discounting, called "exponential discounting," into economic modeling. As demonstrated below, exponential discounting implies a strong inter-temporal consistency for firms and consumers: barring any change in available opportunities, decision makers revisiting inter-temporal choices made at an earlier time will make the same choices for the future as they'd planned earlier. "Rational actors" never suffer regret.

Real people do suffer regret, sometimes wishing they had saved more or had not procrastinated. Behavioral economists attribute people's regretted behavior to a stronger preference for immediate benefits over future benefits, and a stronger aversion to present costs than to future costs, than is captured by the standard economic model. In this view, people are more "present biased" than the standard economic model assumes. In this section we explore both exponential discounting and an alternative model, called "hyperbolic discounting," that captures present bias. David Laibson [8] introduced the "hyperbolic" terminology and first applied the beta-delta model to inter-temporal choices of individuals in their everyday choices.

Exponential Discounting

If an investor has an initial wealth of W_0 at time $t = 0$, what will the investors wealth $W(t)$ be at time t if the money grows at a constant rate r? Mathematically, the assumption that the wealth grows at a constant rate r can be expressed as

$$\frac{dW(t)}{dt} = rW(t).$$

Separating variables and integrating both sides of $\frac{dw}{w} = r dt$, we obtain $(W(t)) = rt + C_0$. Since $C_0 = \ln W_0$ at $t = 0$, we obtain $W(t) = W_0 e^{rt}$.

The discounting question can now be framed as: "What amount of money V_{PV} invested at time $t = 0$ would accumulate to an amount V_t by time t? Economists call

$$V_{PV} = V_t e^{-rt}$$

the present discounted value of V_t realized at time t. The larger r, is the more future amounts of money are discounted. Economists refer to discounting of this sort as "exponential discounting." Standard economists write the present discounted value Π_{PV} of a firm's stream of profits $\pi(t)$ over time as

$$\prod_{PV} = \int_0^\infty \pi(t)e^{-rt} dt.$$

This is what investors would be willing to pay for the firm if the market rate of interest was r; the market value of the firm at time zero is Π_{PV}. In standard economic theory, firms make their inter-temporal decisions so as to maximize Π_{PV}.

Standard economics' strong assumption of inter-temporal consistency arises from its reliance on exponential discounting. Consider two alternative profit events: π_1, realized at time t_1, and π_2 realized at time $t_2 > t_1$. The present discounted values of these profits at time zero are

$$\prod_{1PV} = \pi_1 e^{-rt_1} \quad \text{and} \quad \prod_{2PV} = \pi_2 e^{-rt_2}.$$

In standard economic theory, the firm would choose the profit event for which the present discounted value is greater. Might the firm later change its mind about which alternative is better? No. Assessed in period T ε (0, t_1) the present discounted value of these alternatives, denoted Π_{1PVT} and Π_{2PVT}, respectively, would be

$$\prod_{1PVT} = \pi_1 e^{-r(t_1 - T)} = e^{rT} \prod_{1PV} \quad \text{and} \quad \prod_{2PVT} = \pi_2\, e^{-\,r(t_2 - T)} = e^{rT} \prod_{2PV}.$$

The initially preferred alternative remains the preferred alternative. The firm will never regret its earlier decisions unless the available alternatives or the interest rate changes. Exponential discounters are "time consistent."

In contrast to what economists say about firms, economists say that individuals do not seek to maximize the present dollar value of their wealth. Rather, economists say, people seek to maximize their overall well-being by choosing the available outcome they most prefer. Economists represent people's preferences among outcomes by way of "utility functions." A utility function assigns to every outcome a numerical value that corresponds to the outcome's ranking in the person's preferences. If outcome A is preferred to outcome B, outcome A is assigned a higher utility value; outcomes an individual considers equally desirable are assigned the same utility value. Individuals are assumed to make choices that maximize their utility function – that is, they choose their most preferred available option.

Because people are generally impatient, utility realized is appreciated more the sooner it occurs. Standard economic models, which capture people's impatience much like they do the firms, assume people maximize the present discounted value of lifetime utility. If one's immediate utility is $u(t)$ at time t, the individual's present discounted value of lifetime utility U_{PV}, is

$$U_{PV} = \int_0^\infty u(t)e^{-\rho t}dt,$$

where ρ is called the individual's "rate of time preference" or "individual discount rate." (One can assume that individuals don't care about what happens after they die, in which case $u(t)$ is zero after one dies, or one can assume that individuals do care about what happens after they die, in which case $u(t)$need not be zero after the individual dies.) Standard economic theory assumes that ρ is constant for any given individual, but can differ among individuals as a matter of taste.

Many economic models focus on discrete time periods within which decisions are made. Here too, standard economists rely on a constant discount rate. If depositing X dollars in the bank yields an interest payment of rX at the end of each year, then X dollars is worth $(1+r)X$ dollars after one year, $(1+r)^2X$ dollars after 2 years, and $(1+r)^tX$ dollars after t years. Thus, the present discounted value, V_{PV} of V_t dollars to be received in year t is

$$V_{PV} = (1+r)^{-t}V_t.$$

Hence, in discrete time, the present discounted value of a series of profits $\pi(t)$ realized over time becomes

$$V_{PV} = \sum_{t=0}^{\infty} \pi(t)(1+r)^{-t}.$$

Likewise, the present discounted value of lifetime utility for a consumer with personal discount rate ρ who lives forever is

$$U_{PV} = \sum_{t=0}^{\infty} u(t)(1+\delta)^{-t}.$$

Problem 5.1 Suppose you purchase a \$10,000 bond for 18 years that earns $r = 6.25\%$ each year.

(a) What is the present discounted value of your investment at the end of 18 years?
(b) What is the actual accumulated value of the investment at the end of 18 years?

Problem 5.2 A firm finishes its latest project and finds that it has \$1000 k in cash on hand that it will lose in taxes if it does not reinvest the money in another project. The firm's discount rate is $r = .50$. The firm can invest the \$1000 k in either: (a) a project that will return \$2025 k in three years or (b) a project that will return \$1000 k in two years and another \$675 k in three years. Which option is more profitable for the firm?

Problem 5.3 Assume that an individual's utility $u(t)$ at time t, depends on the individual's level of consumption $c(t)$ at time t, and the utility function is

$$u(t) = u(c(t)) = \frac{[c(t)]^{1-n}}{1-n} \text{ with } n \text{ a positive constant not equal to 1.}$$

Assume also that the present discounted value of consumption at time t with exponential discounting is $u(t)e^{-rt}$ with $r \geq 0$. (A reasonable real world value for r is around 2.)

(a) The marginal utility $\frac{du}{dc} = [c(t)]^{-n}$ is, to a first-order approximation, the change in $u(t)$ corresponding to a one unit increase in $c(t)$ at time t. The present value of such a change in utility at time t is $\lfloor c(t)\rfloor^{-n}e^{-rt}$. Show that P_f, the amount (approximately) of current consumption (at $t = 0$) that a consumer would be willing to forgo to get a one unit increment of consumption at future time $t = T$ is

$$P_f = e^{-rt}\left(\frac{c(0)}{c(T)}\right)^n.$$

(Hint: present value lost must match present value gained.)

(b) If individuals differ in their values of n, what does this say about their differing aversion to unequal consumption (for themselves) over time?

Hyperbolic Discounting

Exponential discounting and the accompanying inter-temporal consistency in decision making prove very tractable in economic models. Unfortunately, behavioral economists have shown people often do not display the inter-temporal consistency that follows from constant discount rates. People frequently regret past decisions, wishing they had shown more patience or less procrastination. (See, for example, [4] and [1].)

Psychologists and behavioral economists have that found the sharpest divergences of behavior from the predictions of exponential discounting occur in making tradeoffs between the present and the future, not in making tradeoffs between earlier and later times in the future. Evidence from scans of brain activity suggests that people making decisions about tradeoffs among future outcomes, with no present implications, mostly experience increased brain activity in the pre-frontal cortex. But when the tradeoffs include present considerations, evolutionarily older parts of the brain also become active. A rough interpretation of these findings is that when the costs or benefits are immediate, the older more primitive parts of our brain say "I want it NOW!" or "I don't like it; NOT NOW!", while the prefrontal cortex conducts our rational assessments of costs and benefits. This interpretation underlies formal models that have been widely used in behavioral economics. These models assume a specific form of time inconsistency referred to as "present bias"; the consequence of present bias is impatience in getting rewards and procrastination in embracing costs. The most common model that allows present bias is called (in an unfortunately mathematically misleading fashion) "hyperbolic discounting."

The hyperbolic discounting model (sometimes referred to as "the beta-delta model") uses the parameters $\beta \leq 1$ and $\delta \geq 0$ to describe the present value of utility from receiving $x(t)$ at time $t \geq 0$:

$$U_{PV} = u\left(x(0)\right) + \beta \sum_{t=1}^{\infty} u\left(x\left(t\right)\right)\left(1+\delta\right)^{-t}$$

(If β were to equal one, we would have the exponential discounting model.) When applying the hyperbolic discounting model to decision-making, $t = 0$ refers to the moment in which a decision is actually being made and $t > 0$ refers to future time periods. The beta-delta model with $\beta < 1$ predicts many individuals' behavior better than does the exponential discounting model. In this model, $u(x(t))$ is the immediate utility at time t from consuming $x(t)$ at time t.

As the term $(1 + \delta)^{-t}$ shows, the more remote in time a utility is, the less it is valued at time zero. This is true in the hyperbolic discounting model just as it is in the exponential discounting model. However, in the hyperbolic discounting model, there is also a further intensity to the "right now" that is related to the value of β. A cookie is never so tasty as when it begs to be eaten right now, and a push-up is never more unpleasant than when we are asked to do it right now. When $\beta < 1$, the satisfaction $u(x^*)$ associated with receiving x^* immediately at $t = 0$ is greater than the satisfaction $\beta u(x^*)(1 + \delta)^{-t}$ from receiving x^* at $t > 0$, even if $\delta = 0$, that is, even if the exponential discounting is absent. Since future utility when $\beta < 1$ is deemed less important than when $\beta = 1$, beta-delta discounters with $\beta < 1$ are said to be *present biased*.

For present-biased individuals, opportunities are always intensified when they become immediate, and this often paves the way for regret in the future. Suppose, for example, that at time $t = 0$ a present-biased individual chooses to receive x_b at time $t + 1$ instead of x_a at time t because $\beta u(x_b)(1 + \delta)^{-t-1} > \beta u(x_a)(1 + \delta)^{-t}$. Come time t, that person would prefer to receive x_a immediately rather than x_b in the following period if at the new present time $t = 0$ (which was time t in the past) $u(x_a) > \beta u(x_b)(1 + \delta)^{-1}$. In such a case, the individual would regret having committed to receiving x_b in the following period instead of receiving x_a in what has become the present.

Present-biased preferences introduce a complication absent in the time-consistent model. A distinction is needed between naïve and sophisticated present bias. Naïve present-biased people presume that in the future, they will behave in accord with their present judgments about the future. ("Oh, I won't do it today, but I will do it tomorrow," and because of their present bias, they might procrastinate again when tomorrow rolls around.) Sophisticated present-biased people recognize that they will be present-biased in the future. (So they might say, I'd better do it today because I can foresee that, given my present bias, I won't do it tomorrow, and doing it still later is a very bad idea.) Sophisticates act today in accord with their current preferences, but they take into account that in the future their choices will apply their hyperbolic preferences to a new present that is currently assessed as being in the future.

We can use the following problem due to Matthew Rabin and Ted O'Donoghue [10] to better understand the distinction between naifs and sophisticates, consider three people: a time-consistent person, a present-biased naif, and a present-biased sophisticate. Assume their preferences can be described by a $\beta - \delta$ model with discount rate $\delta = 0$, where $\beta = 1$ for the time-consistent person and $\beta = 0.5$ for the present biased people. (Setting $\delta = 0$ allows us to highlight present bias without the added complication of exponential discounting.) Each of the three people is offered a coupon good for attending a single movie that can be used in one of four weeks. In week one, a mediocre movie is playing. In week two, a good movie is playing. In week three a great movie is playing. And in week four, a truly amazing movie is playing. Using the coupon yields an immediate utility of 3 in the first week, 5 in the second week, 8 in the third week, or 13 in the final week. Assume the movies are all too expensive to attend without the coupon, so only in one week is a movie attended. Also assume that immediate utility is zero in weeks the movie is not attended. It follows that when considering in period t attending the movie in period $T \geq t : (i)\, u\,(x\,(t')) = 0$ for $t' \neq T$ and $u(x(T))$ is as given above for $T = 1, 2, 3$, or 4; (ii) the present discounted value of attending the movie in period t is $u(x(T))$ for $T = t$; and (iii) the present discounted value of attending the movie in period $T > t$ is $\beta u(x(T))(1+\delta)^{T-t}$, which reduces to $\beta u(x(T))$ since δ is assumed equal to zero.

The determination of what each person chooses today is influenced by what the person believes he or she would do in the future if they do not attend the movie today. Decisions to attend or not attend must be made each week, one after the other, based upon a surmise of what would happen in the future were one to not attend in the present period.

The time-consistent person, knowing that in each period the choice they will think best will be the same, confidently waits until week four when the utility from attending is highest.

Now consider the naïf. In week one, the naïve person believes (wrongly) that if he doesn't attend the movie in week one, he will act like a time-consistent person and wait until week four. In week one, the naïf prefers the discounted utility of week four's movie (13/2) to the utility of 3 from attending immediately, so the naïf does not go to the movie in week one. In week two, the naïf will think in the same fashion. Since one half of the fourth week's utility (13/2) exceeds the immediate utility of 5 from going in week two, the naïve person will not go the movie in week two. But when week three becomes the present, the immediate utility of 8 from the week three movie will exceed the discounted utility of attending the movie in week four (13/2). The naïve person will attend the movie in week three. Attending the movie in week three isn't what the naïf expected they'd do, but it is what they'd do.

Next consider the sophisticate. The sophisticate attends the mediocre movie in week one. A sophisticated present-biased individual uses backward induction to surmise in week one which week's movie she will attend if she doesn't attend in week one. The sophisticate surmises that if she arrives in week three not having attended a movie, she will attend in week three because the then immediate utility of 8 from attending in week three would be greater than the discounted utility from attending in week four (13/2). Week four's movie is off the table. The sophisticate also surmises that if she arrives in week two without having seen a movie, she will choose to attend in week two instead of waiting because she'd know the alternative to the immediate utility of 5 from attending in week two would be the discounted utility of attending in week three (8/2). Week three's movie is off the table. Having surmised that she will attend the movie in week two if she doesn't attend in week one, the sophisticate will choose the immediate utility of 3 from attending in week one over the discounted utility of attending in week 2 (5/2). Being naïve can sometimes be better than being sophisticated! (The problems show that the reverse can also be true.)

The movie example is somewhat contrived, but the vulnerability of sophisticates to difficulties that naifs might avoid arises in more serious circumstances, too. For example, having stopped smoking, a naïf might say, "If I just avoid smoking right now, I will do the right thing and continue to not smoke." That promise of a life without smoking might suffice to deter the naïf from smoking, at least until a more tempting moment arises when the naïf resumes smoking. Sadly, a sophisticate who foresees such a subsequent irresistibly tempting moment might say instead, "If I avoid smoking now, I'm just putting off an eventual slide back into smoking. I might as well light up now!" The naifs' self-deception works to the naifs' advantage in this case.

Problem 5.4 In period one, a beta-delta consumer with $\beta = .5$ and $\delta = .5$ two options: (i) attending a concert in period two that will yield 4.5 utils of immediate satisfaction in that period, or (ii) attending a play in period three which would yield nine utils of immediate satisfaction in that period. Assume immediate utility is zero in periods in which the person does not attend an entertainment.

(a) In period one, which event will the beta-delta consumer think more valuable (i) or (ii)?
(b) When period two arrives, will the consumer think differently than in (a)?

(c) For what value β (with $\delta = .5$) would the beta-delta consumer think the same in both period one and period two?

Problem 5.5 Suppose a naïf, a sophisticate, and a time-consistent person must each clean house on one of three nights: on Friday night when the cost of spending time cleaning is a loss of 2.5 utils, or on Saturday night when the cost is a loss of 4 utils, or on Sunday night when the cost is highest, say a loss of 6 utils. Assume the naïf and sophisticate have beta-delta preferences of $\beta = \frac{1}{2}$ and $\delta = 0$, and the time consistent person has $\beta = 1$ and $\delta = 0$. Assume immediate utility is zero when not cleaning.

(a) When will the time-consistent person clean house? Briefly explain.
(b) When will the naïf clean house? Briefly explain.
(c) When will the sophisticate clean house? Briefly explain.

Problem 5.6 Consider a present biased beta-delta sophisticate and present biased beta-delta naif who must clean house once, on Friday, Saturday, or Sunday, with consequent losses of immediate utility 2.5, 4, and 6, respectively. These people are offered a contract on Friday which must be either accepted or rejected on Friday: if the house isn't cleaned before Sunday, the cost of cleaning on Sunday will become 12 immediate utils lost, but the costs of cleaning Friday or Saturday would still be 2.5 and 4, respectively. Assume $\beta = .5$ and $\delta = 0$ for all. Assume immediate utility is zero when not cleaning.

(a) What will happen to a sophisticate who signs the contract Friday?
(b) Will sophisticates see any reason to sign or not sign the contract Friday?
(c) What will happen to a naïf who signs the contract Friday?
(c) Will the naïf see any reason to sign or not sign the contract Friday?
(d) Would a time-consistent person have any reason to sign the contract on Friday?

6 Solutions

Problem 2.1

(a) If Player 2 accepts an offer of $sG < \frac{1}{2}G$, Player 1 receives $X_1 = (1-s)\,G > \frac{1}{2}G$, and Player 1 experiences an advantageous inequality. Player 1's utility would therefore be

$$U_1 = (1-s)\,G - \beta_1\,(1-2s)\,G.$$

(b) If Player 2 accepts an offer of $(1-s)G > \frac{1}{2}G$, Player 1 receives $X_1 = sG < \frac{1}{2}G$ and Player 1 experiences a disadvantageous inequality. Player 1 therefore has utility

$$U_1 = sG - \alpha_1\,(1-2s)\,G.$$

(c) Since $a_1 > \beta_1$, it follows for $s < \frac{1}{2}$ that the utility of Player 1 in (a) is greater than the utility of Player 1 in (b). Thus, Player 1 prefers to offer the smaller of sG and $(1-s)G$ if that would be accepted.

Problem 2.2 If Player 2 declines the offer, both X_1 and X_2 equal zero, so $U_1 = 0 = U_2$.

Problem 2.3

(a) If Player 2 accepts share sG, Player 2's utility is

$$U_2 = sG - \beta_2 \max\left\{(2s-1)\,G, 0\right\} - a_2 \max\left\{(1-2s)\,G, 0\right\}.$$

(b) For $s < 1/2$, Player 2's utility is

$$U_2 = sG - \alpha_2(1-2s)G$$

because Player 2 is experiencing a disadvantageous inequality. Player 2 will be indifferent between accepting this inequality and rejecting the offer of sG if this utility equals that from refusing the offer, namely zero. Thus, Player 2 would be indifferent between accepting and rejecting the offers $sG < \frac{1}{2}G$ if $s - a_2(1-2s) = 0$, which occurs for

$$s_1 = \frac{\alpha_2}{(1+2\alpha_2)}.$$

Notice that as α_2 increases, s_1 increases and approaches (but does not equal) $1/2$—Player 2 will always accept an offered share of $\frac{1}{2}G$.

(c) G doesn't matter for Player 2's decision to accept or reject, it is the offered fraction s of G that matters. This aspect of the model is roughly consistent with a substantial number of observed cases.

Problem 2.4

(a) Player 1 will never offer more than $1/2G$ because a utility-maximizing Player 2 would accept some share less than $1/2G$. However, Player 1 will offer a share large enough to avoid rejection and the resulting zero utility. To examine how Player 1's utility varies for $sG \le \frac{1}{2}G$, we use Problem 1.1 (a) to write

$$U_1 = (1-s)\,G - \beta_1(1-2s)\,G,$$

which can be rearranged as:

$$U_1 = (1-\beta_1)\,G + (2\beta_1 - 1)\,Gs.$$

For $sG \le \frac{1}{2}G$, the graph of $U_1$is a line segment with slope $(2\beta_1 - 1)$. We can observe from the graph of U_1 that if $\beta_1 > 1/2$, Player 1 will offer $1/2G$ to Player 2 because U_1 increasing in that case. If $\beta_1 < 1/2$, then U_1 decreases with sG and Player 1 will offer Player 2 the smallest acceptable share, which is the smallest possible amount in excess of s_1G, unless the smallest possible amount in excess of s_1G were large enough and the prospect of an indifferent Player 2 rejecting s_1G small enough, in which case Player 1 would offer s_1G. If $\beta_1 = 1/2$, Player 1 will be indifferent among offers of any share $sG \in [s_1G, \frac{1}{2}G]$, with s_1G in the set only subject to the preceding qualifier.

(b) The fraction s of G a utility-maximizing Player 1 offers to a utility-maximizing Player 2 doesn't depend upon G; it depends on β_1 and a_2.

Problem 3.1

(a) According to Bayer's rules $P(G|IG) = \frac{P(G)P(IG|G)}{P(IG)}$. Given

$$P(G) = .25, \quad P(Y) = .75, \quad P(IG|G) = .90, \quad P(IG|Y) = .20,$$

$$P(IG) = P(G)P(IG|G) + P(Y)P(IG|Y) = (.25)(.90) + (.75)(.20) = .375.$$

Therefore,

$$P(G|IG) = \frac{(.25)(.90)}{.375} = .60.$$

(b) If there are 10,000 cabs in New York City, Isabella would correctly identify (2,500)(.90) = 2,250 green cabs. (And mistakenly identify (7,500)(.20) = 1,500 yellow cabs as green.)

Problem 3.2

(a) Given that $P(U(2R, W)) = \frac{1}{2} = P(U(3R, 2W))$, and that the first two decisions drawn from the urn are right (RR), Bayes's rule says

$$P(U(2R, 2W)|RR) = \frac{\left(\frac{1}{2}\right)P(RR|U(2R, 2W))}{\left(\frac{1}{2}\right)P(RR|U(2R, 2W) + \left(\frac{1}{2}\right)P(RR|(3R, 2W)}.$$

Since decisions are drawn from the urns with replacement,

$$P(RR|U(2R, 2W)) = \frac{1}{4} \quad \text{and} \quad P(RR|(U(3R, 2W)) = \frac{9}{25}.$$

Therefore,

$$P(U(2R,2W)|RR)) = \frac{\frac{1}{2}\left(\frac{1}{4}\right)}{\frac{1}{2}\left(\frac{1}{4}\right)+\frac{1}{2}\left(\frac{9}{25}\right)} = .41 \text{ and } P(U(3R,2W)|RR)) = .59.$$

(b) If sampling is drawn without replacement,

$$P(RR|U(2R,2W)) = \tfrac{1}{2}\left(\tfrac{1}{3}\right) = \tfrac{1}{6} = \text{and } P\big(RR|U(3R,2W)\big) = \left(\tfrac{2}{4}\right) = \tfrac{3}{10}.$$

Therefore, again using Bayes's rule,

$$P(U(2R,2W)|RR) = \frac{\left(\frac{1}{2}\right)\left(\frac{1}{6}\right)}{\left(\frac{1}{2}\right)\left(\frac{1}{6}\right)+\left(\frac{1}{2}\right)\left(\frac{3}{10}\right)} = .36 \text{ and } P(U(3R,2W)|RR) = .64.$$

(c) The person who mistakenly assumes sampling without replacement will overstate the probability that the population looks like the sample in hand, *i.e.*, the probability that the broker is a "good one" will be overstated.

(d) Returning to sampling with replacement,

$$\begin{aligned} P(RRR|RR) = {} & P\left(U\left(2R,2W\right)|RR\right) P(RRR|RR \text{ and } U(2R,2W)) \\ & + P(U\left(3R,2W\right)|RR)\, P(RRR|RR \text{ and } U(3R,2W)) \end{aligned}$$

thus

$$P(RRR|RR) = (.41)\left(\tfrac{1}{2}\right) + (.59)\left(\tfrac{3}{5}\right) = .56.$$

(Note that $P(RRR|RR) = P(R \text{ on 3rd draw} \mid RR)$.)

(e) Returning once more to sampling without replacement.

$$\begin{aligned} P(RRR|RR) = {} & P\left(U\left(2R,2W\right)|RR\right) P(RRR|RR \text{ and } U(2R,2W) \\ & + P(U\left(3R,2W\right)|RR)\, P(RRR|RR \text{ and } U(3R,2W)) \end{aligned}$$

Thus,

$$P(RRR|RR) = (.36)(0) + (.64)\left(\tfrac{1}{3}\right) = .21; \quad P(RRW|RR) = .79.$$

(f) The person who mistakenly assumes sampling without replacement will expect that the small sample will become more like the population than is likely, *i.e.*, the person will think it overly likely that the broker is "due to make a mistake."

Problem 4.1

(a) The expected utility from Gamble A is $.5\sqrt{1{,}440{,}000} + .5\sqrt{160{,}000} = 800$.
and the expected utility from Gamble B is $.5\sqrt{640{,}000} + .5\sqrt{360{,}000} = 700$.
This person prefers Gamble A.

(b) The expected utility from a sure thing is just the utility from the sure thing. Since $\sqrt{700{,}000 > 800}$, this person prefers the sure \$700,000 to Gamble A.

(c) A person is risk averse if he or she prefers the sure wealth of a gamble's expected wealth over the gamble. Thus, the choice in (b) evidences risk aversion. The expected wealth from Gamble A is $.5(\$1{,}440{,}000) + .5(\$160{,}000) =$ \$800,000. Since the person in (b) prefers a sure wealth \$700,000 over the expected value \$800,000 from Gamble A, that person would obviously also prefer a sure \$800,000 over the gamble.

Problem 4.2

(a) $U(100) = \log(100) = 2$.

(b) $U(10{,}000) = \log(10{,}000) = 4$.

(c) The expected utility of the gamble is $.5U(100) + .5U(10{,}000) = 3$.

(d) The sure thing that would give the individual the same utility as the gamble is $\log z = 3$, or $z = \$1000$. So $C(z) = \$1000$ is the certainty equivalent of this gamble.

(e) The expected outcome from the gamble is $E(z) = \$5050$. Therefore, the risk premium is $\pi(z) = E(z) - C(z) = \$4000$.

Problem 4.3 If the utility function in Figure 1 was convex, the line segment AB would lie above the utility function, and $E(u(z))$ would be greater than the utility from a "sure" wealth equal to $E(z)$. The individual would choose the gamble over the sure thing, indicating risk seeking behavior.

Problem 4.4

(a) The gain's prospect value $\sqrt{g}$ must have the same magnitude as the loss's prospect value $-2\sqrt{20}$. Thus, $\sqrt{g} = 2\sqrt{20}$ and $g = 80$.

(b) In assessing the gamble, the individual calculates

$$f(x_1, x_2, p_1) = -.5(2\sqrt{36}) - .5(2\sqrt{4}) = -8.$$

Because $-8 > -2\sqrt{20} = -8.94$, the value of the sure thing, the person chooses the gamble. The person prefers a risky gamble to a sure outcome equal to the gamble's expected outcome. This is risk-seeking behavior not just because the person chooses to gamble, but because they choose the gamble over a sure thing with equal expected value.

(c) In assessing the gamble, the individual calculates

$$f(x_1, x_2, p_1) = -.1\big(2\sqrt{36}\big) - .8\big(2\sqrt{4}\big) = -4.4.$$

Because $-4.4 < -2\sqrt{4.032}$, the value of the sure thing, the individual chooses the sure thing. This is risk-averse behavior not just because the person chooses not to gamble, but because they forego a gamble in favor of a sure thing with equal expected value.

Problem 4.5

(a) An insured person is sure to lose the actuarially fair price of the insurance (if insured, this is the outcome whether the house burns down or not). The actuarially fair price of such insurance is $-px_1$, which would exactly compensate the insurance company for the expected loss from fire. An individual acting in accord with prospect theory for whom insurance would cost $-px_1$ would assign a value $v(px_1)$ to being insured.

The outcomes facing a homeowner without insurance are a large loss $x_1 \ll 0$ if the house burns down, and a loss $x_2 = 0$ if there is no fire. For every probability $p > 0$ of a fire, the point $(x, y) = (px_1, pv(x_1))$ lies on the line segment connecting the origin $(0, v(0))$ to the point $(x_1, v(x_1))$. Because the value function is convex in the loss region (*i.e.*, in the lower left-hand quadrant in Figure 2), the line segment connecting $(0, v(0))$ to $(x_1, v(x_1))$ lies above the value function between $x = 0$ and $x = x_1$. Thus, for any $p\varepsilon(0, 1)$, going uninsured is preferable to insuring against a fire because the expected value of the value function calculated for gambling against a fire,

$$pv(v(x_1) + (1 - p)v(0)$$

is greater than $v(px_1)$, the value consequence of insuring. Consequently, mirroring the argument for risk aversion with concave utility functions, a consumer maximizing the expected value of the value function would choose to go uninsured because going without insurance yields higher expected value than does insuring. Thus, maximizing the *expected* value of the prospect value function entails risk seeking in the loss domain.

(b) Individuals who act in accord with prospect theory overweight low probability events. Consequently, for the probability p of a fire $\pi(p) > p$. Although

$$pv(x_1) + (1 - p)(0) > v(px_1),$$

as shown in (a), it might still be the possible that

$$\pi(p)v(x_1) + \pi(1-p)(0) < v(px_1),$$

in which case, the homeowner would choose to insure against a fire.

Problem 5.1

(a) The bond earns \$10,000(.0625) = \$625 each year for 18 years. Therefore, the present discounted value of its cumulative earnings is

$$PV = 10{,}000 + 640\sum_{t=1}^{18}\left[\frac{1}{1.0625^t}\right] = 10{,}000\left[\frac{1-\left(\frac{1}{1.0625}\right)^{18}}{.0625}\right] = \$16{,}801.42.$$

(b) The total accumulated value is 10,000 + 18(625) = \$21,250.00.

Problem 5.2 Since the discount rate is .50, the present discounted value of $\$W$ received t years later is $W(1+.5)^{-t}$. So, the first project is worth \$2025k/3.375 = \$600 k, and the second project is worth (\$1000/2.25) + (\$675/3.375) = \$644.44 k. Therefore, the second project is the more profitable one to undertake.

Problem 5.3

(a) The present discounted value of consumption at time t is $= \frac{c^{1-n}}{1-n}\, e^{-rt}$. The derivative of the PV with respect to consumption is $c^{-n}e^{-rt}$. To a first-order approximation, obtaining one unit of consumption at time T increases lifetime utility by $c^{-n}(T)e^{-rT}$. The loss in utility from giving up p_f units of consumption at time zero is, to a first-order approximation, $c^{-n}(0)e^0 p_f$. For what p_f are this gain and loss equal? They are equal when

$$p_f = \left(\frac{c\,(T)}{c\,(0)}\right)^{-n} e^{-rT} = \left(\frac{c\,(0)}{c(T)}\right)^{n} e^{-rT}.$$

The individual is willing to forego p_f units of consumption in period 0 to obtain one more unit of consumption in period T.

(b) When $c(0) = c(1)$, the value n does not matter for P_f. When $c(0) > c(T)$, a person for whom n is larger will pay more to redress the imbalance a bit than would someone with a smaller value of n. When $c(0) < c(T)$, a person for whom n is larger is more reluctant to increase the imbalance than is a person with a smaller n. (Economists call n a measure of inter-temporal consumption inequality aversion.)

Problem 5.4

(a) If the beta-delta consumer has $\beta = .5$ and $\delta = .5$, then *seen from period one*, the period-two concert's present discounted utility is $.5 \times (4.5/1.5) = 1.5$ and the period-three play's present discounted utility of 9 is $.5 \times (9/2.25) = 2$. The consumer will in period one think attending the play in the third period is better.

(b) Come the second period, the concert would give immediate satisfaction. *Looked upon in period two*, the present discounted utility of the period two concert would be 4.5 utils and the present discounted utility of the period three play assessed would be $.5 \times (9/1.5) = 3$. In period two, the beta-delta consumer would think attending the concert in the third period is better.

(c) $\frac{9\beta}{1.5} > 4.5$ or $\beta > .75$.

Problem 5.5

(a) The time-consistent person will minimize the lost utility by cleaning on Friday when the lost utility is least.

(b) The naïf cleans on Sunday. The naïf surmises on Friday that if he waits until Saturday he will clean on Saturday because the discounted loss of 2 utils from cleaning Saturday is less than the discounted cost of 3 utils from cleaning on Sunday. So on Friday, the naïf thinks Sunday is off the table. Hence, the naïf would decide not to clean on Friday because the immediate loss of 2.5 utils from cleaning Friday exceeds the discounted loss of

2 utils from cleaning on Saturday. But come Saturday, the naif's present bias kicks in and the immediate loss of 4 utils from cleaning on Saturday exceeds the discounted loss of 3 utils from cleaning on Sunday, so the naïf waits until Sunday. This example illustrates how life can unfold unhappily for naifs.

(c) The present-biased sophisticate, on the other hand, would anticipate that if she waits until Saturday to clean she will clean on Sunday. So Saturday is off the table. Since the immediate cost of 2.5 utils on Friday, is less than the discounted cost of 3 utils from cleaning on Sunday, the sophisticate will clean on Friday.

Problem 5.6

(a) The sophisticate who signs the contract cleans on Saturday. This sophisticate surmises on Friday that if she got to Saturday without having cleaned she'd clean on Saturday because Sunday's discounted cost of 6 would be more than Saturday's immediate cost of 4. Consequently, on Friday, the sophisticate would not clean because Friday's immediate cost of 2.5 exceeds the discounted (to Friday) cost of 2 from cleaning on Saturday.

(b) On Friday, cleaning on Saturday looks better than cleaning on Sunday. The sophisticate surmises correctly on Friday that with the contract, she will clean Saturday with discounted (to Friday) cost of 2 and that without the contract she will clean on Friday with cost of 2.5. The sophisticate will sign the contract clean on Saturday.

(c) A naïf who signs the contract will (correctly) anticipate cleaning on Saturday if the cleaning isn't done Friday. On Friday, cleaning Saturday looks better than cleaning on Friday, so the naïf would wait until Saturday. On Saturday, Sunday's discounted cost of 6 will be greater than Saturday's cost of 4. The naïf cleans on Saturday.

(d) A naïf sees no reason to sign or not sign the contract since the naïf (wrongly) believes on Friday that they will clean the house on Saturday either way.

(e) The time-consistent person will clean Friday with or without the contract, so the contract is irrelevant to the time consistent person.

References

[1] George W. Ainsle, "Impulse Control in Pigeons," *Journal of the Experimental Analysis of Behavior* 21 (1974) 485–489.

[2] D. Bernoulli, "Exposition of a New Theory on the Measurement of Risk," *Econometrica* 22 (1954) 23–36.

[3] Ernst Fehr and Klaus M. Schmidt, "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics* 3 (1999) 817–868.

[4] L. Green, A.F. Fry, and J. Myerson, J., "Discounting of Delayed Rewards: A Life Span Comparison," *Psychological Science* 5 (1994) 33–36.

[5] W. Güth, R. Schmittberger, and B. Schwarze, "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization* 3 (1982) 367–388.

[6] J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath, "In Search of Homo Economicus: Behavioral Experiments in 15 Small Scale Societies," *American Economic Review* 91 (2001) 73–78.

[7] Daniel Kahneman and Amos Tversky, "Prospect Theory: An Analysis of Decision Under Risk," Econometrica 47 (1979) 263–291.

[8] David Laibson, "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics* 112 (1997) 443–77.

[9] Matthew Rabin, "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics* 117 (2002) 775–816.

[10] Matthew Rabin and Ted O'Donoghue, "Doing It Now or Doing It Later," *American Economic Review* 89 (1999) 103–124.

[11] A. Tversky and D. Kahneman, "Judgments of and by Representativeness," in *Judgment Under Uncertainty: Heuristics and Biases*, D. Kahneman, P. Slovic & A. Tversky, eds., Cambridge University Press, Cambridge, UK, 1982.

[12] John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, 1944.

6

Econometrics

Rae Jean B. Goodman
Department of Economics, U.S. Naval Academy, Annapolis

1 Introduction

Ragnar Frisch, 1969 recipient of the Nobel Prize in Economic Sciences, stated that econometrics is "the unification of economic theory, statistics and mathematics." [2] Economic theory starts with a set of assumptions and produces a series of qualitative statements or hypotheses. Economists use data that are different from the data used in most mathematical statistics courses. These data are not usually generated by a controlled experiment. For example, economists are not able to change a tax rate to see how people respond, and then make policy recommendations. Instead they use past responses to changes in tax rates to estimate what the response will be in the current period. Because the economic situation is not precisely the same as in the past, and people may not respond in the same way, there is typically greater variance in economic data. There may also be measurement error due to data collection methods. So econometricians have developed statistical methods specifically to accommodate economic data. Econometrics provides the tools for estimating economic relationships, for testing economic theory, and for forecasting.

A typical undergraduate econometrics course starts with an explanation of the assumptions and the development of the simple least-squares regression model (the classical model). The course proceeds to multiple regression, hypothesis testing within the model, and discussions of selecting independent variables and the appropriate functional form. It continues with focus on the identification and implications of the violations of the classical assumptions: multicollinearity, autocorrelation, and heteroscedasticity. Depending on the length of the course, it might include such advanced topics as time-series models, limited dependent variable models, multiple equation models, and pooling cross-section and time series data (panel data) methods.

Students take an econometrics course at highly selective institutions. The econometrics course may or may not be required in the major but will be strongly recommended for those students who plan to continue in economics or an MBA program. To be successful in an econometrics course, a student should be comfortable with a semester of calculus and an elementary statistics course that includes hypothesis testing and simple, bivariate regression. If students have been exposed to matrix algebra, then an econometrics course can be taught using matrices. However, frequently at the undergraduate level, the course is taught using summation notation. Optimally, students would take econometrics following the intermediate theory courses or simultaneously with the last intermediate theory course, and prior to any field courses.

Given the heavy emphasis on statistics, and the desired economics preparation, mathematics students may reach too far if they attempt to solve econometrics problems. However, since mathematics students use regression for curve fitting, Sections 2–4 present applications from an econometrics course that may be profitably used in a mathematics course. The problems involve summation notation, differentiation of a function, matrix operations, and the solution of systems of linear equations.

The problems in Section 5 are intended only to illustrate some of the principles and statistical techniques used in an undergraduate econometrics course. Hopefully, this may enable teachers of statistics to appreciate and to emphasize what is needed in preparing students for work in economics.

The solutions, or outline of solutions, of problems are presented in Section 7.

2 The Linear Regression Model

In most econometrics courses, linear regression begins with finding the least squares estimators in a simple, two-variable model. The model describes the random response y_i for each value x_i of the independent variable x,

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,$$

where β_0 and β_1 are unknown parameters, and the random variable ε has expected value $E(\varepsilon_i) = 0$ and constant variance σ^2. Thus, for each value of x_i, the random response y_i is distributed with conditional mean $E(y_i|x_i) = \beta_0 + \beta_1 x_i$ and variance σ^2. The notation $E(y_i|x_i)$ indicates that once x_i is chosen, y_i's distribution has the mean $\beta_0 + \beta_1 x_i$. The line $E(y_i|x_i) = \beta_0 + \beta_1 x_i$ is called the *regression* line.

Suppose we have n observations

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (i = 1, \ldots, n).$$

If b_0 and b_1 are estimators of β_0 and β_1, then

$$\hat{y}_i = b_0 + b_1 x_i \quad (i = 1, \ldots, n)$$

are the respective estimates of y_i. The difference between y_i and its estimates

$$e_i = y_i - \hat{y}_i = y_i - b_0 + b_1 x_i \quad (i = 1, \ldots, n). \tag{1}$$

is called the i^{th} residual. The least squares estimators of β_0 and β_1 are the values of b_0 and b_1 that minimize the sum of the squared residuals

$$L = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - b_0 - b_1 x_i]^2. \tag{2}$$

Problem 2.1 Minimizing L with respect to b_0 and b_1

(a) Separately treating b_0 and b_1 as the variables, show that $\frac{\partial L}{d b_0} = 0$ and $\frac{\partial L}{\partial b_1} = 0$ yield the respective equations

$$\sum_{i=1}^{n} e_i = 0 \quad \text{and} \quad \sum_{i=1}^{n} x_i e_i = 0. \tag{3}$$

(b) In terms of $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, solve the equations in (3) to obtain $b_0 = \overline{y} - b_1\overline{x}$ and b_1.

(c) Show that b_0 and b_1 are the least squares estimators by proving that b_0 and b_1 minimize L.

Problem 2.2

(a) Verify that the systems of equations $e_i = y_i - b_0 + b_1 x_i$ is expressed in matrix form, or more compactly, as

$$\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}, \quad \text{or} \quad \mathbf{e} = \mathbf{y} - X\mathbf{b}.$$

(b) Show that $L = \mathbf{e}^{\mathbf{t}}\mathbf{e}$ and, that setting the vector of derivatives $\frac{\partial L}{\partial b} = \begin{pmatrix} \frac{\partial L}{\partial b_0} \\ \frac{\partial L}{\partial b_1} \end{pmatrix}$ to $\mathbf{0}$ yields

$$X^{\mathrm{t}}(\mathbf{y} - X\mathbf{b}) = \mathbf{0}.$$

(c) Show that the matrix $X^{\mathrm{t}}X$ has an inverse if some $x_i \neq x_j$, and then compare the values b_0 and b_1 of $\mathbf{b} = (X^{\mathrm{t}}X)^{-1}X^{\mathrm{t}}\,\mathbf{y}$ with their values obtained in Problem 2.1.

The equation $\hat{y}_i = b_0 + b_1 x_i$ is said to be an estimated regression equation. Using different sets of n observations will yield different values for b_0 and b_1, hence, different regression equations. This raises a number of questions. For example, how good an estimate is b_1 of the true coefficient β_1? And for a given value x_0, how good is the prediction $\hat{y}(x_0)$ of the true response $\beta_0 + \beta_1 x_0$? The answers to these questions depend on the statistical assumptions made about the x_i and ε_i. Section 5 illustrates how the estimated regression equation is used to make inference about β_0 and β_1.

3 Multiple Linear Regression

The extension of simple linear regression for two variables takes the form

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \quad (i = 1, \ldots, n),$$

where $x_1, x_2, \ldots, x_k$ are mutually independent variables, $\beta_0, \beta_1, \ldots, \beta_k$ are unknown parameters, and ε_i is a random variable with expected value $E(\varepsilon_i) = 0$ and constant variance σ^2. The only additional assumption that needs to be made for the multiple regression model is that there is not perfect correlation between the independent variables; that is, no x_k is a linear combination of other x's.

Problem 3.1 Suppose we have n independent observations

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \quad (i = 1, \ldots, n)$$

and $b_0, b_1, \ldots, b_k$ are respective estimators of $\beta_0, \beta_1, \ldots, \beta_k$. Then

$$\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \cdots + b_k x_{ik} \quad (i = 1, \ldots, n)$$

are the respective estimates of y_i. The system of equations

$$e_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \cdots - b_k x_{ik} \quad (i = 1, \ldots, n)$$

can be written

$$\begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ 1 & x_{n1} & x_{x2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{pmatrix}, \quad \text{or } \mathbf{e} = \mathbf{y} - X\mathbf{b}\,.$$

The least squares estimators of $\beta_0, \beta_1, \ldots, \beta_k$ are the values of $b_0, b_1, \ldots, b_k$ that minimize the sum of the squared differences $L = \mathbf{e}^{\mathbf{t}}\mathbf{e}$.

Show that

$$L = \mathbf{e}^{\mathbf{t}}\mathbf{e} = \mathbf{y}^{\mathbf{t}}\mathbf{y} - 2\mathbf{b}^{\mathbf{t}}X^{\mathrm{t}}\mathbf{y} + \mathbf{b}^{\mathbf{t}}X^{\mathrm{t}}X\mathbf{b}$$

and that setting $\frac{\partial L}{\partial \mathbf{b}}$ to zero yields $X^{\mathrm{t}}X\mathbf{b} = X^{\mathrm{t}}\mathbf{y}$.

It can be shown that the $k \times k$ matrix $X^{\mathrm{t}}X$ is invertible if $n > k$ and the columns of X are linearly independent.

4 Elasticity and Linear Functional Forms

Economists use the concept of elasticity in a wide range of applications. Elasticity is the percentage change in the dependent variable divided by the percentage change in the independent variable. If $y = f(x)$, the *elasticity* of y at x is

$$\varepsilon = \frac{\Delta y/y}{\Delta x/x} = \frac{\Delta y/\Delta x}{y/x} = \frac{\Delta y}{\Delta x}\frac{x}{y}$$

Thus, the elasticity of y at x can be interpreted as the percentage change in y due to a 1% increase in x. Economists say that $y = f(x)$ is *inelastic* at x if $|\varepsilon| < 1$, *elastic* at x if $|\varepsilon| > 1$, and has *unit elasticity* at x when $|\varepsilon| = 1$. Economists use elasticity to discuss the impact of the change of different variables on the function under discussion.

Problem 4.1

(a) Explain why the elasticity of $y = f(x)$ at x is approximated by $\varepsilon = \frac{xf'(x)}{f(x)}$ for small Δx.

(b) Explain why the function $y = kx^\alpha$ is called a *constant elasticity* function.

(c) The elasticity of $y = f(x)$ is often expressed in terms of natural logarithms. Use $Y = \ln f(x)$ and $X = \ln(x)$ to show that

$$\varepsilon = \frac{dY}{dX} = \frac{d(\ln y)}{d(\ln x)}$$

(d) The elasticity of $y = f(x_1, x_2, \ldots, x_k)$ at x_k is the percentage change in y divided by the percentage change in x_k. If $y = 2 + x_1 + 3x_2 - 4x_3$, what is the elasticity of y with respect to x_2 at $(x_1, x_2, x_2) = (3, 2, 1)$? If $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$, what is the elasticity of y with respect to x_2 at $(x_1, x_2, \ldots, x_k) = (x_1^*, x_2^*, \ldots, x_k^*)$?

A production function relates the amount of labor and capital necessary to produce different levels of output. By using a logarithmic transformation some nonlinear production functions can be changed into a linear form and can thus be estimated using linear regression techniques. Estimating production functions allows the economist to determine whether there are diminishing, constant, or increasing returns to labor and capital. The definition of diminishing returns is, for example, that as additional units of labor are added, the amount of additional output for each additional unit of labor decreases.

Problem 4.2 Consider the Cobb-Douglas production function $Q_t = \alpha L_t^\beta e^{u_t}$, where L_t is the amount of labor required to produce amount Q_t at a point in time t, and u_t represents the random error at time t caused by measurement error, different responses of people or firms at different points in time, etc. In Table 1 of the Appendix, the annual data for Real Gross Domestic Product (Q) and the labor force (L) are given for 1970–2009 period. Additional data may be collected from the St. Louis Federal Reserve data bank (http://research.stlouisfed.org/fred2/) for real gross domestic product and from the Bureau of Labor Statistics (www.bls.gov) for the labor force variable.

(a) Use linear regression to estimate the relationship $\ln Q = \alpha + \beta \ln L$.

(b) What is the elasticity of output with respect to labor?

(c) What does the coefficient β indicate about increasing or diminishing returns?

Economists use the semilog transformation for creating models involving rates of growth. To estimate the growth of the money supply or the labor force over a specific period of time, we might assume that the money supply or the labor force grew at a constant rate over that period. The growth of the money supply can be expressed as

$$M_t = \alpha(1+g)^t e^{u_t},$$

where α is a parameter, M_t is the size of the money supply in year t, and g is the compound growth rate of M_t. A compound rate of growth implies a linear relationship between $\ln M_t$ and t, not between M_t and t.

Problem 4.3 Has the money supply grown at a constant rate? Two measures of money supply used by economists are $M1$ (composed of currency in circulation, traveler's checks, and checkable deposits) and $M2$ (composed of $M1$ plus savings accounts, small time deposits, and money market accounts). Monthly data for $M1$ and $M2$ are in Table 2 of the Appendix. Additional $M1$ and $M2$ data may be collected from the St. Louis Federal Reserve data bank (http://research.stlouisfed.org/fred2/).

(a) Show that the equation $M_t = \alpha(1+g)^t e^{u_t}$ can be rewritten

$$M_t^* = \alpha^* + \beta^* t + u_t,$$

where $M_t^* = \ln M_t$, $\alpha^* = \ln \alpha$, and $\beta^* = \ln(1+g)$.

(b) Use linear regression to estimate the relation $M_t^* = \alpha^* + \beta^* t$ for both $M1$ and $M2$. Determine α and the compound growth rates of $M1$ and $M2$ over the entire period.

Problem 4.4 The annual labor force data are in Table 1 of the Appendix. Additional labor force data may be collected from the Bureau of Labor Statistics website (http://www.bls.gov/).

Use $L_t = \alpha(1+g)^t\, e^{u_t}$ and linear regression to estimate $\ln L_t = \ln \alpha + (1+g)t$ and at what rate the labor force has grown.

5 Examples of Applications in an Econometrics Course

5.1 Statistical Concepts and Graphical Analysis

In an introductory probability and statistics course, students generally learn to perform tests of hypotheses. Regression results permit several different types of hypotheses to be tested. These tests rest on assuming that the ε_i in a regression model are mutually independent, normally distributed random variables with zero mean and constant variances. The initial tests would be the significance of the regression as a whole, and subsequently tests of the individual regression coefficients would be performed. These tests use the F-distribution and the t-distribution. More advanced tests of hypotheses, such as testing whether a subset of independent variables contributes significantly to the results, also use the F-test.

Problem 5.1 Using annual data for the 1947–1978 period, Hong V. Nguyen [6] reported the following results from estimating a production function that included money.

$$\begin{aligned}\ln Y &= -3.136 + 0.875 \ln L + 0.515 \ln K + 0.240 \ln M1 \\ &\qquad (0.756) \quad (0.120) \qquad (0.059) \qquad (0.148) \\ \overline{R}^2 &= 0.998\end{aligned}$$

where Y is output, L is the labor force, K is the capital stock, and $M1$ is the money supply defined as $M1$. The estimated standard errors are reported below the coefficients.

(a) Using a 5% level of significance, determine which of the coefficients is statistically significantly different from zero.

(b) Given the results, should money be included in the production function?

(c) Using a 5% level of significance, determine if the regression equation is significant as a whole?

(d) Provide an economic interpretation of the coefficient on $\ln L$.

Problem 5.2 Fredland and Little [1] reported the following results from estimating a semi-log earnings function.

$$\begin{aligned}\ln W &= 8.1542 + 0.0626\, GC + 0.0115\, JT + 0.1169\, MVE + 0.1452\, CVE + \text{(other variables)} \\ &\qquad\qquad (.0061) \qquad (.0019) \qquad (.0614) \qquad (.0680) \\ \overline{R}^2 &= 0.30 \qquad F = 26.31 \qquad n = 714\end{aligned}$$

where W is the annual wage, GC measures education and is the highest grade completed, JT is job tenure, MVE is a bivariate variable for a user of military vocational education opportunities, CVE is a bivariate variable for a user of civilian vocational education opportunities, and the "other variables" are a series of socio-demographic and regional bivariate variables. (There are a total of 12 independent variables in the regression.) The estimated standard errors are in parentheses.

(a) Using a 5% level of significance, determine which of the coefficients is statistically significantly different from zero.

(b) Using a 5% level of significance, determine if the regression equation is significant as a whole. Explain.

(c) Is education an important determinant of the annual wage? Explain.

(d) Is vocational education an important determinant of the annual wage? Explain.

5.2 Graphical Analysis

We can use graphical tools to analyze whether or not a particular set of regression results conform to the some of the assumptions underlying the classical linear regression model. While there are formal statistical tests for the violations of each of the assumptions, scatter plots of the residuals can be used to provide clues about whether or not the assumptions are violated. Depending on the length and sophistication of the econometrics course, some of the formal statistical tests may not be included.

We can examine several different scatter plots of the residuals or error terms. The usual starting point is to examine a scatter plot of the residuals against each independent variable x_k or versus the predicted value $\hat{y}$. The independent variable is plotted on the horizontal axis and the residual on the vertical axis. If the points are clustered in a horizontal band around the zero line, this indicates that the assumption of a constant variance of the residuals or error term is not violated. If the width of the scatter plot increases or decreases as the value of x_k (or $\hat{y}$) increases, this suggests that the variances of the error terms are not constant. Therefore, the residuals indicate the presence of heteroscedasticity. (Heteroscedasticity is present if the variance of each error term, u_t, conditional on the x_k, is not constant.) If the scatter plot forms a "U" shape or a mountain shape, this usually indicates that the relationship between y and x_k is nonlinear. If a plot of the residuals versus time or observation, indicates a relationship (positive or negative cluster of points), then time might be included as a variable or the plot may be indicating that a variable is missing. A third residual plot is the residuals versus the residual from the previous observation or time period (the lagged residual). If there is any distinct relationship, the residual plot indicates that there is autocorrelation and this violates the assumption that the error terms are independent. A histogram of the residuals is one way of checking the assumption of a normal distribution of error terms. Encountering a violation of an assumption of the classical linear model does not mean that the analysis must be abandoned; one must use methods to correct for the violations. To see graphical representations of the assumptions and violations of the assumptions, look at Gujarati and Porter [3].

Generally, spreadsheet programs and statistical programs permit the collection of the residuals and frequently provide the plots discussed above using a specific command.

Problem 5.3 Collect data on real consumption expenditures C_t and real personal income Y_t at time t from the St. Louis Federal Reserve website.

(a) Estimate the regression equation $C_t = \alpha + \beta Y_t$.

(b) Use the residuals from this regression to test whether any of the assumptions of the classical linear regression model are violated? Explain

6 Final Comments

To expand on the data based questions, there are several online government data resources: (1) Federal Reserve of St. Louis: http://research.stlouisfed.org/fred2/ contains numerous macroeconomic time series data sets; (2) Board of Governors of Federal Reserve http://federalreserve.gov has data focused on variables related to monetary policy; (3) Census Bureau http://census.gov has data on population, housing, income, and business activity; and (4) U.S. Bureau of Economic Analysis http://bea.gov has international, national, regional and industry data. Two on-line data sources that charge for access to the data are Economagic (economagic.com) and Interuniversity Consortium for Political and Social Research (ICPSR.com). Your institution or Economics Department may have a site license for either of these data sources. Economagic has 400,000 data series and ICPSR has an extensive collection of downloadable data sets and related research articles. Econometrics courses use one or more of the following statistical programs to manipulate the data sets: SPSS, TSP, E-VIEWS, or STATA.

The commonly used texts in an econometrics course include Gujarati and Porter [3], Kennedy [4], Murray [5], Pindyck and Rubinfeld [7], Ramanathan [8], Stock and Watson [9], and Wooldridge [10]. The type of problems presented here would be found in any of these texts. Econometrics texts include data sets. Wooldridge has referenced articles related to each data set. These articles would provide additional questions like Problem 5.1.

7 Solutions

Problem 2.1

(a) For $L = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - b_0 - b_1 x_i]^2$.

$$\frac{dL}{db_0} = -2\sum_{i=1}^{n}[y_i - b_0 - b_1 x_i] = 0 \Rightarrow \frac{dL}{db_0} = \sum_{i=1}^{n} e_i = 0$$
$$\frac{dL}{db_1} = -2\sum_{i=1}^{n}[y_i - b_0 - b_1 x_i]x_i = 0 \Rightarrow \frac{dL}{db_0} = \sum_{i=1}^{n} x_i e_i = 0$$

(b) $\sum_{i=1}^{n} y_i = b_0\ n + b_1\left(\sum_{i=1}^{n} x_i\right) \Rightarrow \overline{y} = b_0 + b_1\overline{x}$

$$\sum_{i=1}^{n} x_i y_i = b_0\left(\sum_{i=1}^{n} x_i\right) + b_1\left(\sum_{i=1}^{n} x_i^2\right) \Rightarrow \sum_{i=1}^{n} x_i y_i = nb_0\overline{x} + b_1\sum_{i=1}^{n} x_i^2$$

Substituting $b_0 = \overline{y} - b_1\overline{x}$ in the second equation, we obtain

$$b_0 = \overline{y} - b_1\overline{x} \quad \text{and} \quad b_1 = \frac{n\overline{x}\overline{y} - \sum\limits_{i=1}^{n} x_i y_i}{n\overline{x}^2 - \sum\limits_{i=1}^{n} x_i^2}.$$

(c) The second partial derivatives of L are

$$\frac{\partial^2 L}{\partial b_0^2} = 2n > 0, \quad \frac{\partial^2 L}{\partial b_1^2} = 2\sum_{i=1}^{n} x_i^2 > 0, \quad \text{and} \quad \frac{\partial^2 L}{\partial b_0 \partial b_1} = 2n\overline{x}.$$

Furthermore,

$$\frac{\partial^2 L}{\partial b_0^2}\cdot\frac{\partial^2 L}{\partial b_1^2} - \left(\frac{\partial^2 L}{\partial b_0 \partial b_1}\right)^2 = 4n\left[\sum_{i=1}^{n} x_i^2 - n\,\overline{x}^2\right] = 4n\sum_{i=1}^{n}(x_i - \overline{x})^2 > 0.$$

Therefore, b_0 and b_1 minimize L.

Problem 2.2

(a) Since $\mathbf{e}$ is an $n \times 1$ column vector and $\mathbf{e^t}$ is a $1 \times n$ row vector, $L = \mathbf{e^t e} = e_1^2 + e_2^2 + \cdots + e_n^2$.

(b) From Problem 2.1,

$$\frac{\partial L}{\partial \mathbf{b}} = \begin{pmatrix} \dfrac{\partial L}{\partial b_0} \\ \dfrac{\partial L}{\partial b_1} \end{pmatrix} = \begin{pmatrix} -2\sum\limits_{i=1}^{n} e_i \\ -2\sum\limits_{i=1}^{n} x_i e_i \end{pmatrix} = -2X^t\mathbf{e} = -2X^t(\mathbf{y} - X\mathbf{b}).$$

(c) The matrix $X^t X = \begin{pmatrix} n & n\overline{x} \\ n\overline{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}$ has determinant $n\left(\sum_{i=1}^n x_i^2 - n\overline{x}^2\right) = n\sum_{i=1}^n (x_i - \overline{x})^2 > 0$, and inverse

$$(X^t X)^{-1} = \frac{1}{n^2\overline{x}^2 - n\sum\limits_{i=1}^n x_i^2} \begin{pmatrix} -\sum\limits_{i=1}^n x_i^2 & n\overline{x} \\ n\overline{x} & -n \end{pmatrix}.$$

Therefore,

$$\mathbf{b} = (X^{\mathrm{t}}X)^{-1}X^{\mathrm{t}}\,\mathbf{y} = \frac{1}{n\overline{x}^2 - \sum\limits_{i=1}^n x_i^2} \begin{pmatrix} \overline{x}\sum\limits_{i=1}^n x_i y_i - \overline{y}_i \sum\limits_{i=1}^n x_i^2 \\ n\overline{x}\,\overline{y} - \sum\limits_{i=1}^n x_i y_i \end{pmatrix} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}.$$

Problem 3.1 From $\mathbf{e} = \mathbf{y} - X\mathbf{b}$ **and** $\mathbf{e}^{\mathbf{t}} = (\mathbf{y} - X\mathbf{b})^{\mathbf{t}} = \mathbf{y}^{\mathbf{t}} - \mathbf{b}^{\mathbf{t}}X^{\mathrm{t}}$,

$$\begin{aligned} L &= \mathbf{e}^{\mathbf{t}}\mathbf{e} = (\mathbf{y}^{\mathbf{t}} - \mathbf{b}^{\mathbf{t}}X^{\mathrm{t}})(\mathbf{y} - X\mathbf{b}) = \mathbf{y}^{\mathbf{t}}\mathbf{y} - \mathbf{b}^{\mathbf{t}}X^{\mathrm{t}}\mathbf{y} - \mathbf{y}^{\mathbf{t}}X\mathbf{b} + \mathbf{b}^{\mathbf{t}}X^{\mathrm{t}}X\mathbf{b} \\ &= \mathbf{y}^{\mathbf{t}}\mathbf{y} - 2\mathbf{b}^{\mathbf{t}}X^{\mathrm{t}}\mathbf{y} + \mathbf{b}^{\mathbf{t}}X^{\mathrm{t}}X\mathbf{b} \end{aligned}$$

since $\mathbf{y}^{\mathbf{t}}X\mathbf{b}$ is a scaler and thus equal to its transpose. Furthermore,

$$\frac{\partial L}{\partial \mathbf{b}} = \begin{pmatrix} \dfrac{\partial L}{\partial b_0} \\ \dfrac{\partial L}{\partial b_1} \\ \vdots \\ \dfrac{\partial L}{\partial b_k} \end{pmatrix} = 0 - 2X^{\mathrm{t}}\mathbf{y} + 2X^{\mathrm{t}}X\mathbf{b}.$$

Problem 4.1

(a) When Δx is small, $\frac{\Delta y}{\Delta x} \approx f'(x)$.

(b) For $y = kx^\alpha$, we have $f'(x) = \alpha k x^{\alpha-1}$ and $\varepsilon = \frac{xf'(x)}{f(x)} = \frac{x(\alpha k x^{\alpha-1})}{kx^\alpha} = \alpha$.

(c) If $Y = \ln f(x)$ and $X = \ln x$, then $\frac{dY}{dX} = \frac{dY}{dx} \cdot \frac{dx}{dX} = \frac{f'(x)}{f(x)} \cdot x = \varepsilon$.

(d) The elasticity of $y = 2 + x_1 + 3x_2 - 4x_3$ with respect to x_2 at $(x_1, x_2, x_3) = (3, 2, 1)$ is $\varepsilon = \frac{\Delta y}{\Delta x_2} \cdot \frac{x_2}{y} = 3 \cdot \frac{2}{7} = \frac{6}{7}$. The elasticity of $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$ with respect to x_2 at $(x_1, x_2, \ldots, x_k) = (x_1^*, x_2^*, \ldots, x_k^*)$ is $\varepsilon = \frac{dy}{dx_2}\frac{x_2}{y} = \beta_2 \frac{x_2^*}{\beta_0 + \beta_1 x_1^* + \cdots + \beta_k x_k^*}$.

Problem 4.2 The data are in Table 1.

(a) The regression results are:

$\ln Q = -10.3143 + 1.6691 \ln L$
SE (0.4222) (0.0365)
t-stat -24.4291 45.6815
$R^2 = 0.982$ $F = 2086.80$ $n = 40$

where SE is the estimated standard error and t-stat is the calculated t-statistic.

(b) The elasticity of output with respect to labor is 1.6691, indicating a relatively elastic relationship.

(c) The coefficient of $\ln L$ in excess of one indicates increasing returns to labor for a given capital stock.

Problem 4.3 The money supply data are in Table 2.

The regression results for $M1$ are

$\ln M1 = 7.1610 + 0.002619\,t$
SE (0.0100) (0.0002)
t-stat 714.02 10.969
$R^2 = 0.632 \quad F = 120.33 \quad n = 72$

where SE is the estimated standard error and t-stat is the calculated t-statistic.

Therefore, $\ln(1+g) = 0.002619$ and $g = 0.0026$ or 0.26% average annual growth rate.

The regression results for $M2$ are

$\ln M2 = 8.6947 + 0.004968\,t$
SE (0.0029) (0.000069)
t-stat 2972.91 71.35
$R^2 = 0.9862 \quad F = 5091.18 \quad n = 72$

where SE is the estimated standard error and t-stat is the calculated t-statistic.

Thus, $\ln(1+g) = 0.004968$ and $g = 0.00498$ or 0.498%.

Problem 4.4 Using the data from Table 1, the regression results are

$\ln L = 11.1908 + 0.0177\,t$
SE (0.0115) (0.0005)
t-stat 976.23 36.29
$R^2 = 0.972 \quad F = 1316.78 \quad n = 40$

where SE is the estimated standard error and t-stat is the calculated t-statistic.

Therefore, $\ln(1+g) = 0.0177$ and $g = 0.0178$ or 1.78%.

Problem 5.1

(a) To test which coefficients are statistically significantly different from zero, economists use the t-distribution

$$t_{n-k-1,\alpha} = \frac{b_i - \beta_i}{\sigma_{b_i}},$$

where t has $n-k-1$ degrees of freedom (n = number of observation, k = number of independent variables) and $\alpha = .05$ is the level of significance.

$$H_0: \quad \beta_i = 0$$
$$H_A: \quad \beta_i \neq 0$$

Here $n = 32$ and $k = 3$, and the theoretical $t_{n-k-1,\alpha=0.05} = \pm 1.701$.

$\ln Y = -3.136 + 0.875 \ln L + 0.515 \ln K + 0.240 \ln M1$
SE (0.756) (0.120) (0.059) (0.148)
t-stat -4.148 7.292 8.729 1.622

where SE is the estimated standard error and t-stat is the calculated t-statistic.

Thus, all of the coefficients are statistically significantly different from zero except the coefficient of $\ln M1$.

(b) Should money be included in the production function given the results?

Because the coefficient of $\ln M1$ is not statistically significant from zero, money should not be included in the production function.

(c) Using a 5% level of significance, determine if the regression equation is significant as a whole?
To test if the regression equation significant as a whole, the F-distribution is used.

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$
$$H_A: H_0 \text{ is not true or not all } \beta\text{s are zero}$$

Because $\overline{R}^2 = 1 - (1 - R^2)\left[\frac{n-1}{n-k-1}\right] = 0.988$, the calculated F is

$$F_{(k)(n-k-1)} = \left[\frac{n-k-1}{k}\right]\left[\frac{R^2}{1-R^2}\right] = 5147.3$$

The theoretical value of F is 2.92 for a significance level of 0.05. Therefore, we reject the null hypothesis and conclude that the regression as a whole is significant.

(d) The coefficient of ln Lrepresents the elasticity between output and labor. Holding capital and money supply constant, a one percent increase in the labor force would increase output by 0.875 percent.

Problem 5.2

$$\ln W = 8.1542 + 0.0626\ GC + 0.0115\ JT + 0.1169\ MVE + 0.1452\ CVE + \text{(other variables)}$$

	SE	(.0061)	(.0019)	(.0614)	(.0680)
	t-stat	10.262	6.053	1.904	2.740

$\overline{R}^2 = 0.30$ $F = 26.31$ $n = 714$

where SE is the estimated standard error and t-stat is the calculated t-statistic.

(a) MVE is significantly different from zero at the 0.05 significance level, and GC, JT, CVE are significant at the 0.01 level.

(b) Because there are 12 variables, $n - k - 1 = 701$ and $\alpha = 0.05$, the theoretical value of $F = 1.752$. The calculated F is 26.31, we reject the hypothesis that all the coefficients equal zero. Thus, the regression equation is significant as a whole.

(c) Is education an important determinant of the annual wage? Explain. The coefficient of the education variable GC (the grade completed) is significantly different from zero. Therefore, GC it is an important determinant of the annual wage.

(d) Is vocational education an important determinant of the annual wage? Explain. The coefficients of both vocational education variables are significant. Thus, both military and civilian vocational education variables are important determinants of the annual wage.

Problem 5.3

(a) For the data in Table 3, the regression results are:

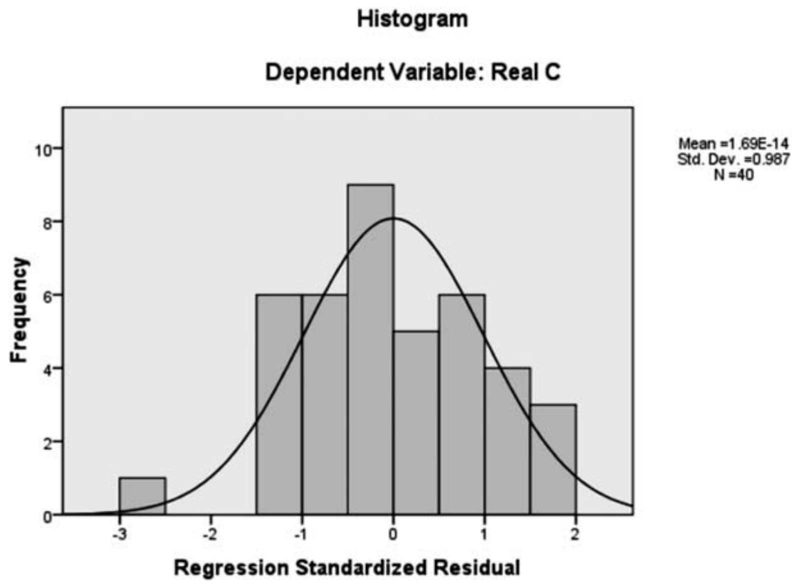
$$C_t = -170.913 + 0.954 Y_t$$

SE	(199.948)	(0.022)
t-stat	-0.855	43.950

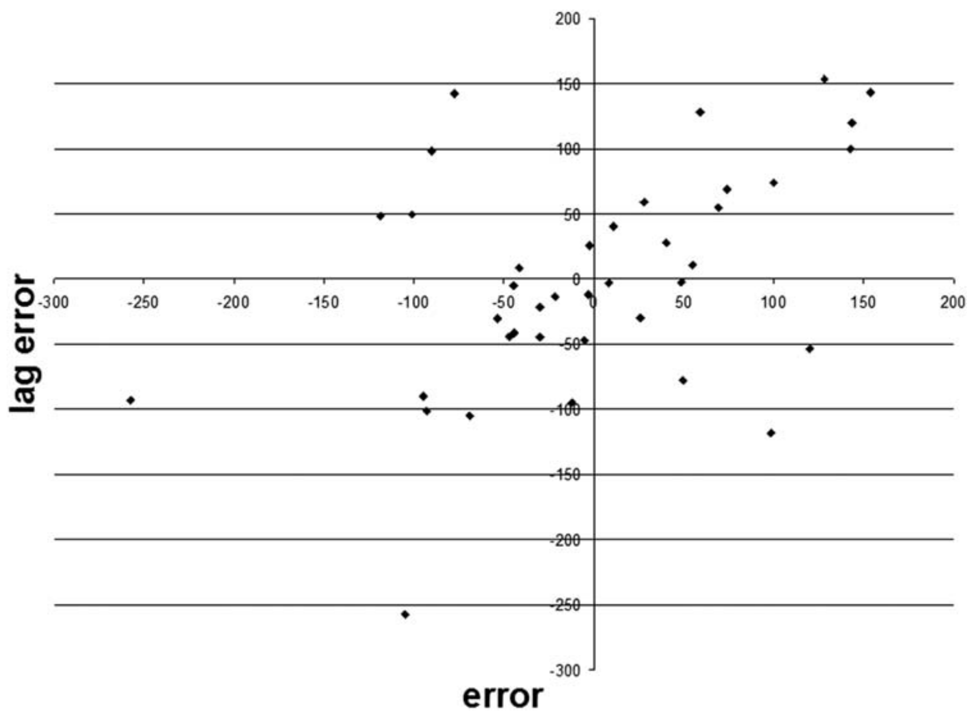
$R^2 = 0.981$ $F = 1931.572$ $n = 40$

where SE is the estimated standard error and t-stat is the calculated t-statistic.

(b) The histogram for the distribution of the error terms appears to be close to a normal distribution.



The graph of the error terms and the lagged error terms indicates a positive relationship. The correlation coefficient is 0.459, and is statistically significant. Thus, autocorrelation exists, and violates one of the assumptions of the least squares regression model.



To identify heteroscedasticity (Heteroscedasticity is present if the variance of each error term, u_t, conditional on the x_k, is not constant), we look at the graph of the error terms and the independent variable. The dispersion of the error term does not change over various levels of the dependent variable. We can conclude that there is no serious heteroscedasticity present.

References

[1] John Eric Fredland and Roger D. Little, "Long-Term Returns to Vocational Training: Evidence from Military Sources," *Journal of Human Resources* **15** (1980) 49–66.

[2] Radnar Frisch, "A Note on the Term 'Econometrics,"' *Econometrica* **4** (1936) 95.

[3] Damodar N. Gujarati and Dawn Porter, *Basic Econometrics*, McGraw-Hill, New York, 5th Edition, 2008.

[4] Peter Kennedy, *A Guide to Econometrics*, John Wiley & Sons, Hoboken, NJ, 6th Edition, 2008.

[5] Michael P. Murray, *Econometics: A Modern Introduction*, Addison-Wesley Publishers, Reading, MA, 2006.

[6] Hong V. Nguyen, "Money in the Aggregate Production Function," *Journal of Money, Credit and Banking* **18** (1986) 141–151.

[7] Robert S. Pindyck and Daniel L. Rubinfeld, *Econometric Models and Economic Forecasting*, McGraw-Hill, New York, 4th Edition, 2000.

[8] Ramu Ramanathan, *Introductory Econometrics with Applications*, Cengage-Southwestern, Mason, OH, 5th Edition, 2002.

[9] James H. Stock and Mark W.Watson, *Introduction to Econometrics*, Addison Wesley Publishers, Reading, MA, 2nd Edition, 2007.

[10] Jeffrey M. Wooldridge, *Introductory Econometrics: A Modern Approach*, Cengage-Southwestern, Mason, OH, 4th Edition, 2008.

Appendix

Year	Real GDP	Labor Force	Time, t	Year	Real GDP	Labor Force	Time, t
	(Billions)	(thousands)			(Billions)	(thousands)	
1970	4299.4	71006	1	1990	8081.8	109487	21
1971	4446.0	71335	2	1991	8055.6	108375	22
1972	4682.9	73798	3	1992	8326.4	108726	23
1973	4964.5	76912	4	1993	8563.2	110844	24
1974	4944.0	78389	5	1994	8900.5	114291	25
1975	4921.4	77069	6	1995	9129.4	117298	26
1976	5191.2	79502	7	1996	9471.1	119708	27
1977	5433.7	82593	8	1997	9881.8	122776	28
1978	5733.2	86826	9	1998	10304.0	125930	29
1979	5930.2	89932	10	1999	10812.1	128993	30
1980	5913.4	90528	11	2000	11268.8	131785	31
1981	6052.5	91289	12	2001	11404.6	131826	32
1982	5939.1	89677	13	2002	11606.9	130341	33
1983	6202.3	90280	14	2003	11914.2	129999	34
1984	6639.8	94530	15	2004	12358.5	131435	35
1985	6893.9	97511	16	2005	12735.5	133703	36
1986	7116.5	99474	17	2006	13046.1	136086	37
1987	7342.2	102088	18	2007	13362.8	137598	38
1988	7650.4	105345	19	2008	13442.6	136790	39
1989	7924.0	108014	20	2009	13083.7	130920	40

Table 1. Real GDP and Labor Force Annual Data 1970–2009

Year	Month	Time, t	Money Supply, M1	Money Supply, M2	Year	Month	Time, t	Money Supply, M1	Money Supply, M2
2004	Jan	1.0	1305.3	6066.2	2007	Jan	37.0	1374.0	7107.8
2004	Feb	2.0	1321.9	6107.5	2007	Feb	38.0	1367.1	7126.9
2004	Mar	3.0	1330.5	6147.4	2007	Mar	39.0	1369.1	7160.1
2004	April	4.0	1332.0	6185.6	2007	April	40.0	1379.2	7212.0
2004	May	5.0	1331.7	6254.8	2007	May	41.0	1380.8	7245.5
2004	June	6.0	1342.6	6267.3	2007	June	42.0	1364.6	7271.1
2004	July	7.0	1340.3	6272.4	2007	July	43.0	1368.6	7303.8
2004	Aug	8.0	1353.2	6297.5	2007	Aug	44.0	1371.8	7363.5
2004	Sept	9.0	1361.2	6332.2	2007	Sept	45.0	1372.4	7403.2
2004	Oct	10.0	1360.9	6356.9	2007	Oct	46.0	1379.0	7436.1
2004	Nov	11.0	1376.5	6392.2	2007	Nov	47.0	1375.5	7464.4
2004	Dec	12.0	1377.1	6407.1	2007	Dec	48.0	1375.8	7501.4
2005	Jan	13.0	1365.8	6405.5	2008	Jan	49.0	1381.1	7534.3
2005	Feb	14.0	1372.9	6423.1	2008	Feb	50.0	1387.0	7623.6
2005	Mar	15.0	1373.0	6438.9	2008	Mar	51.0	1389.7	7684.3
2005	April	16.0	1357.6	6444.0	2008	April	52.0	1392.1	7709.3
2005	May	17.0	1365.4	6461.1	2008	May	53.0	1391.5	7732.8
2005	June	18.0	1378.5	6493.9	2008	June	54.0	1398.1	7745.2
2005	July	19.0	1366.5	6521.7	2008	July	55.0	1415.1	7797.0
2005	Aug	20.0	1376.1	6553.9	2008	Aug	56.0	1400.0	7785.0
2005	Sept	21.0	1377.5	6589.1	2008	Sept	57.0	1459.5	7892.0
2005	Oct	22.0	1374.7	6620.1	2008	Oct	58.0	1472.7	8007.1
2005	Nov	23.0	1377.9	6647.0	2008	Nov	59.0	1518.1	8058.7
2005	Dec	24.0	1375.3	6671.6	2008	Dec	60.0	1594.7	8239.7
2006	Jan	25.0	1380.2	6711.8	2009	Jan	61.0	1573.8	8300.7
2006	Feb	26.0	1380.6	6739.3	2009	Feb	62.0	1562.0	8338.4
2006	Mar	27.0	1384.5	6755.2	2009	Mar	63.0	1564.2	8390.4
2006	April	28.0	1380.0	6780.3	2009	April	64.0	1592.7	8342.5
2006	May	29.0	1386.3	6798.0	2009	May	65.0	1593.0	8415.4
2006	June	30.0	1372.8	6832.8	2009	June	66.0	1641.0	8442.1
2006	July	31.0	1368.5	6868.4	2009	July	67.0	1649.9	8437.7
2006	Aug	32.0	1368.8	6898.3	2009	Aug	68.0	1648.3	8414.5
2006	Sept	33.0	1361.5	6928.2	2009	Sept	69.0	1660.9	8455.2
2006	Oct	34.0	1369.1	6987.3	2009	Oct	70.0	1674.6	8487.8
2006	Nov	35.0	1372.4	7023.5	2009	Nov	71.0	1687.6	8522.7
2006	Dec	36.0	1367.9	7071.7	2009	Dec	72.0	1696.4	8543.9

Table 2. M1 and M2 Monthly Data 2004–2009

Date	Real C	RDPI	Date	Real C	RDPI
2000-01-01	7501.3	8053.4	2005-01-01	8719.0	9189.6
2000-04-01	7571.8	8135.9	2005-04-01	8802.9	9253.0
2000-07-01	7645.9	8222.3	2005-07-01	8865.6	9308.0
2000-10-01	7713.5	8234.6	2005-10-01	8888.5	9358.7
2001-01-01	7744.3	8296.5	2006-01-01	8986.6	9533.8
2001-04-01	7773.5	8273.7	2006-04-01	9035.0	9617.3
2001-07-01	7807.7	8484.5	2006-07-01	9090.7	9662.5
2001-10-01	7930.0	8385.5	2006-10-01	9181.6	9788.8
2002-01-01	7957.3	8611.6	2007-01-01	9265.1	9830.2
2002-04-01	7997.8	8658.9	2007-04-01	9291.5	9842.7
2002-07-01	8052.0	8629.2	2007-07-01	9335.6	9883.9
2002-10-01	8080.6	8649.6	2007-10-01	9363.6	9886.2
2003-01-01	8122.3	8681.3	2008-01-01	9349.6	9826.8
2003-04-01	8197.8	8812.5	2008-04-01	9351.0	10059.0
2003-07-01	8312.1	8935.4	2008-07-01	9267.7	9838.3
2003-10-01	8358.0	8986.4	2008-10-01	9195.3	9920.4
2004-01-01	8437.6	9025.9	2009-01-01	9209.2	9926.4
2004-04-01	8483.2	9115.0	2009-04-01	9189.0	10077.5
2004-07-01	8555.8	9175.9	2009-07-01	9252.6	9984.4
2004-10-01	8654.2	9303.4	2009-10-01	9289.5	9985.5

Table 3. Quarterly Data for Real Consumption Expenditures and Real Personal Disposable Income

7

The Portfolio Problem

Kevin J. Hastings
Knox College, Galesburg

1 Introduction

"Buy low, sell high" may sound like wise advice, but it is certainly not easy to follow. If we open the daily paper to the stock quotations and see that currently a share of IBM stock sells for $103\frac{3}{4}$, down from $105\frac{1}{2}$ the day before, is the stock currently "low" and ready for an advance, or is it "high" and on the way down? Even when a stock is behaving in a relatively stable way over a time period, its daily volatility might make us nervous. We all want a decent return on our investments, but not at the cost of feeling insecure about their risks.

There are many kinds of information that might be used to predict the performance of stocks: general economic conditions, health of the industry the stock represents, productivity of the company as reflected in its annual report, and so on. Our purpose here is not to show how to play the stock market successfully, but rather to illustrate how, using stock quatations and a little bit of statistics, one can define and estimate the average rate of return on investment and the risk of investment. Based on these estimates, we will use the technique of Lagrange mulitpliers to put together an optimal portfolio of investments.

The problem of choosing an optimal combination of investments is known in the economic literature as the *portfolio problem*. The 1990 Nobel Prize in Economics was awarded to Harry Markowitz, William Sharpe, and Merton Miller for developing the theory of portfolio optimization introduced in this paper. For further reading, you might enjoy looking at Markowitz's original paper [2] and Sharpe's book [3] in the References.

2 Average Rate of Return and Risk

Suppose that P_t is the stock's price per share on day t. Then the daily rate of return on day t is

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}.$$

Evidently, we prefer high rates of return to low rates of return. However, on a given day, a stock's rate of return is unknown to us because we must wait for its price to be observed the next day. But then we have missed our investment opportunity to sell or buy more of the stock. The way to deal with this is to suppose that, at least over a short time period, there is a constant but unknown "average" or expected rate of return on the stock, denoted μ, about which R_t fluctuates daily. A reasonable way to estimate this theoretical expected rate of return is to follow the stock over as many days as possible and compute the average $\bar{R} = \frac{1}{n}\sum_{t=1}^{n} R_t$ of its daily rates of return R_t. As we will see, it will be useful to calculate the *variance* of the stock's daily rates of return:

$$S^2 = \frac{1}{n-1}\sum_{t=1}^{n}(R_t - \bar{R})^2.$$

Table 1 shows the prices for seventeen trading days of four companies trading on the New York Stock Exchange. The numbers in the rate of return columns are obtained by dividing the difference (current price minus previous price) by the previous price. The averages $\bar{R}_1$, $\bar{R}_2$, $\bar{R}_3$, and $\bar{R}_4$ of these observed rates of return are the respective estimates of the expected rates of return.

WalMart(1)	R_t	Goodyear(2)	R_t	Honda(3)	R_t	ComEd(4)	R_t
42		30.125		20.25		36.5	
41.375	−1.5%	31.375	4.1%	20.375	0.6%	36.75	0.7%
41.875	1.2%	32.125	2.4%	20.5	0.6%	36.5	−0.7%
41.625	−0.6%	33.25	3.5%	20.375	−0.6%	36.375	−0.3%
42.125	1.2%	33.25	0%	20	−1.8%	36.625	0.7%
43.125	2.4%	32.125	−3.4%	20.625	3.1%	37.375	2.0%
42.5	−1.4%	33.25	3.5%	20.625	0%	37.25	−0.3%
42.625	0.3%	33.125	−0.4%	21.125	2.4%	37.5	0.7%
43.375	1.8%	33.5	1.1%	21.375	1.2%	37.875	1.0%
42.75	−1.4%	34.875	4.1%	21.125	−1.2%	37.625	−0.7%
42.5	−0.6%	34	−2.5%	20.875	−1.2%	37.375	−0.7%
42.375	−0.3%	34.125	0.4%	20.625	−1.2%	37.125	−0.7%
43	1.5%	34.25	0.4%	21.25	3.0%	37.375	0.7%
43.375	0.9%	34.25	0%	21.5	1.2%	37.625	0.7%
42.375	−2.3%	33.875	−1.1%	21	−2.3%	36.375	−3.3%
42.25	−0.3%	33.5	−1.1%	21.375	1.8%	36.75	1.0%
43.375	2.7%	33.125	−1.1%	21.375	0%	37.125	1.0%
Mean $\bar{R}_1 = .212\%$		$\bar{R}_2 = .620\%$		$\bar{R}_3 = .352\%$		$\bar{R}_4 = .113\%$	
StdDev $S_1 = .148\%$		$S_2 = .232\%$		$S_3 = .170\%$		$S_4 = .122\%$	

Figure 1 shows graphs of the daily rates of return of Goodyear and Honda. The graphs show that there is considerable variability in the stocks' rates of return. Typically, risky investments that have high average rates of return also have high risks. If this was not the case, then investors would all flock to high-reward, low-risk investments, and market imbalances would occur.

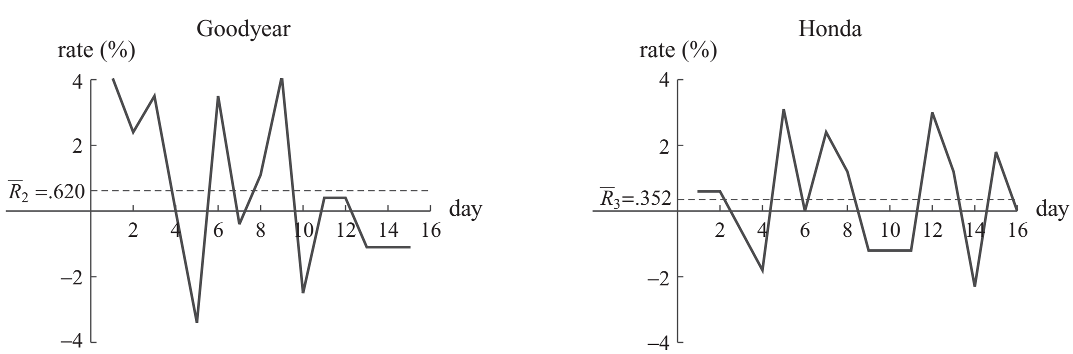


Figure 1

To increase safety, investors tend to diversify, that is, to spread their wealth among a collection or *portfolio* of assets. If the investor has total wealth W, then the decision to be made is what fraction w_i of that wealth is to be devoted to asset i for $i = 1, 2, 3, 4$. Since the total wealth invested in asset i is $w_i W$, and the expected rate of return per dollar invested is μ_i, the expected dollar return from our investment in asset i is $\mu_i w_i W$. Summing over all four assets, the total dollar return on the portfolio is expected to be $\sum_{i=1}^{4} \mu_i w_i W$. Since we invested W dollars, the portfolio's expected rate of return per dollar invested is

$$\mu = w_1\mu_1 + w_2\mu_2 + w_3\mu_3 + w_4\mu_4. \tag{1}$$

Estimating the theoretical means μ_i by their respective sample means $\bar{R}_i$ gives an explicit expression for the approximate average rate of return. If we did not care about riskiness, then to maximize the portfolio's expected rate of return μ we would find the largest individual rate of return, here μ_2 according to our estimates, and devote $w_2 = 1$ ($= 100\%$) of our wealth to that asset, leaving $w_i = 0$ ($= 0\%$) of our wealth to the other investments. The portfolio problem becomes much more interesting when we take riskiness into account.

Risk and Risk Aversion

Since a stock's rate of return is random, its daily rates of return cannot be predicted, and so there is riskiness in the investment. A theoretical measure of the riskiness of a stock is the *variance* σ^2, which is the expected squared difference between the stock's daily rate of return R and its expected rate of return μ. The squaring ensures that both positive and negative differences contribute positively to the variance. Note that the variance will be large when values of R are far from the mean μ and the variance will be small when the values of R are close to μ. In this way, the variance becomes a reasonable measure of risk. The *standard deviation* $\sigma = \sqrt{\sigma^2}$ also measures risk, and has the desirable quality of being expressible in the same units as the mean μ, but more often we will be working with the variance.

The theoretical variance σ^2 of a stock is estimated by the *sample variance*, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(R_t - \bar{R})^2$, which is the average squared difference of the sample's returns R_t from the sample mean $\bar{R}$. When there are n sample values, it would seem that we should divide the total by n, but technical reasons indicate that this method slightly underestimates σ^2. It is more common to divide by $n - 1$ instead. It is straightforward to calculate the sample variance S_i^2 for each of the stocks $i = 1, 2, 3, 4$ in Table 1. The square root $S_i = \sqrt{S_i^2}$ of the sample variance, called the *sample standard deviation*, which is in units of %, is displayed for each stock at the bottom of the table. This is also an estimated measure of risk, although it is less convenient computationally.

For the portfolio problem, we need to consider the portfolio's actual daily rate of return, which, by analogy with formula (1) is

$$R_p = w_1 R_1 + w_2 R_2 + w_3 R_3 + w_4 R_4.$$

If we assume that the random variables R_1, R_2, R_3, and R_4 are *stochastically independent*, meaning roughly that the value of one does not affect the values of the others, then the risk or variance σ^2 of the portfolio is calculated to be

$$\sigma^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_3^2\sigma_3^2 + w_4^2\sigma_4^2. \tag{2}$$

Thus, by estimating each stock's theoretical variance σ_i^2 by its sample variance S_i^2, we have an explicit expression for the estimated portfolio variance as a function of the portfolio weights. It is not too surprising that risks should be additive, though it requires an assumption. Interestingly, when the independence assumption is not satisfied, equation (2) contains additional terms that can be negative if one stock tends to go down as another goes up. These negative terms, arising from correlations among stocks, actually reduce porfolio risk as compared to the independent case. Intuitively, by combining two stocks that tend to move in opposite directions, the investor is protected against losses in one by gains in the other. This is the heart of diversification in investment problems.

Having introduced the portfolio's theoretical expected rate of return μ and measure of risk σ^2, we would like to determine the investment fractions w_i ($i = 1, 2, 3, 4$) that maximize μ and minimize σ^2. In general, these two goals are incompatible since the most profitable investments tend to be the riskiest (see Table 1).

One way to balance these objectives is to maximize a kind of portfolio net value $\mu - a\sigma^2$, where the constant $a \geq 0$ is a *risk aversion factor* that measures your reluctance to take risks. We will refer to this portfolio objective function as the *risk-averse expected return*, and the portfolio that maximizes it will be the *risk-averse optimal portfolio*. It is clear that a larger value of a means more worry about risk. For example, $a = 0$ means that you don't mind taking risks at all; your only goal is to maximize average rate of return. If $a = 10$, then an increase of one unit in risk would have to be offset by an increase in 10 units of rate of return in order for you to value the investment the same way.

If an investor with risk aversion a is presented with two portfolios having expected rates of return μ_1 and μ_2 and variances σ_1^2 and σ_2^2, the investor will be indifferent between the two portfolios if

$$\mu_1 - a\sigma_1^2 = \mu_2 - a\sigma_2^2.$$

Solving for a, we see that

$$a = \frac{\mu_1 - \mu_2}{\sigma_1^2 - \sigma_2^2}. \tag{3}$$

In assuming that an investor has risk aversion a, we are assuming that this quotient is the same for all pairs of investments between which the investor is indifferent. Equation (3) gives us a way to estimate an investor's risk aversion.

Problem 2.1 Suppose three stocks have mean rates of return $\mu_1 = 5\%$, $\mu_2 = 10\%$, $\mu_3 = 15\%$ and because the third stock is very speculative, you want no more than half as much wealth in it as in stock 1. What is the maximum portfolio rate of return that you can achieve? (Hint: Express the problem in terms of portfolio weights w_1 and w_2 only, and look at the region of points in the $w_1 - w_2$ plane that are legal under the problem conditions. Where in that region is the portfolio rate of return largest?)

Problem 2.2 What is your risk aversion constant if you are indifferent between the portfolio that has $\mu_1 = 5\%$ and $\sigma_1 = 1\%$, and the portfolio that has $\mu_2 = 8\%$ and $\sigma_2 = 2\%$?

Problem 2.3 When stock returns are dependent on one another, the variance of a sum is not just the sum of the variances, but also includes covariance terms. The covariance measures the degree of dependence between the two returns. For two stocks, the variance of $w_1 R_1 + w_2 R_2$ is

$$\sigma^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1 w_2 \mathrm{Cov}\,(R_1, R_2)\,.$$

Suppose $\sigma_1^2 = 2$, $\sigma_2^2 = 1$ and $\mathrm{Cov}\,(R_1, R_2) = -1$. What values of w_1 and w_2 minimize the variance of the portfolio? Are they different weights than those which minimize variance in the independent case? How much smaller is the minimum variance in the dependent case than the independent case?

3 Solving the Optimization Problem

Seeking to maximize the risk averse expected return $\mu - a\sigma^2$ reduces the portfolio problem to the constrained optimization problem

$$\begin{aligned}\text{Maximize } f\,(w_1, w_2, w_3, w_4) &= w_1\mu_1 + w_2\mu_2 + w_3\mu_3 + w_4\mu_4\\ &\quad -a\left(w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + w_3^2\sigma_3^2 + w_4^2\sigma_4^2\right)\end{aligned}$$

subject to the constraint $g\,(w_1, w_2, w_3, w_4) = w_1 + w_2 + w_3 + w_4 - 1 = 0$.

Problem 3.1 Use the partial derivatives $\frac{\partial L}{\partial w_i}$ $(i = 1, 2, 3, 4)$ and $\frac{\partial L}{\partial \lambda}$ of the Lagrangian

$$L(w_1, w_2, w_3, w_4) = f(w_1, w_2, w_3, w_4) - \lambda g\,(w_1, w_2, w_3, w_4)$$

to obtain a system of five equations in the unknowns λ, w_1, w_2, w_3, w_4. Show that the solution to the system's simultaneous equations is

$$w_2 = \frac{\sigma_1^2}{\sigma_2^2}w_1 + \frac{\mu_2 - \mu_1}{2a\sigma_2^2}, \quad w_3 = \frac{\sigma_1^2}{\sigma_3^2}w_1 + \frac{\mu_3 - \mu_1}{2a\sigma_3^2}, \quad w_4 = \frac{\sigma_1^2}{\sigma_4^2}w_1 + \frac{\mu_4 - \mu_1}{2a\sigma_4^2},$$

where

$$w_1 = \frac{1 - \frac{\mu_2 - \mu_1}{2a\sigma_2^2} - \frac{\mu_3 - \mu_1}{2a\sigma_3^2} - \frac{\mu_4 - \mu_1}{2a\sigma_4^2}}{1 + \frac{\sigma_1^2}{\sigma_2^2} + \frac{\sigma_1^2}{\sigma_3^2} + \frac{\sigma_1^2}{\sigma_4^2}}.$$

Problem 3.2 Let us see what happens if one of the assets has no risk (i.e. $\sigma = 0$).

(a) Suppose that there are three mutually independent assets, with $\mu_1 = 5\%$, $\sigma_1 = 0\%$, $\mu_2 = 8\%$, $\sigma_2 = \sqrt{2}\%$ (i.e. $\sigma_2^2 = 2$), $\mu_3 = 12\%$, $\sigma_3 = 2\%$. What is the risk-averse optimal portfolio for an investor whose risk aversion is $a = 2$? And for $a = 3$?

(b) How do these expected returns compare with the maximum risk-free expected return?

Problem 3.3 Suppose you have \$10,000 to invest and your risk aversion constant is $a = 12$.

(a) Use the data for the four assets in Table 1 to determine the weights w_1, w_2, w_3, w_4 that maximize the portfolio's risk-averse expected rate of return. What is the estimated return for this portfolio on a \$10,000 investment?

(b) How much of your \$10,000 are you investing in each stock, and what is the return on each investment? Do your answers agree with your return in (a)?

Problem 3.4 Repeat Problem 3.3 for risk aversion $a = 2$. (Note that some weights may be negative, particularly for the low-return, low-risk stock Commonwealth Edison. It is possible to hold negative wealth in an asset, by taking what is called in finance circles a *short position* in that asset. An investor can do this by contracting to acquire the proceeds of sale of the asset as if the investor actually owned it, then paying back at a later time by "buying" the asset back at current market price, thus reversing the order in the usual buy-sell cycle. The investor profits if the stock price goes down, so that the "purchase" price is less than the "selling" price.)

Problem 3.5

(a) Over the same seventeen day trading period, sixteen observations of rates of return on Exxon, Hormel, and Abbott Laboratories gave respective mean rates of return .085%, .134%, and .179%, and respective variances .011, .029, and .025. Thus, Abbott Labs appears to be strictly better than Hormel, in the sense of having a higher expected return, with lower risk. One might assume that an optimal risk-averse portfolio consisting of these stocks, together with, say Goodyear as a fourth stock, would not include Hormel at all. Using a risk aversion of $a = 4$, check that in actuality this is not the case.

(b) If you are curious and ambitious, write general expressions for the value function $\mu - a\sigma^2$ for a portfolio that uses all stocks, and for a second portfolio like the first, except that the weight formerly given to stock 2 (Hormel) is lumped together with the weight for stock 3 (Abbott Labs). Try to see what it is about the two values that makes the first better than the second, and think about what your study implies about diversification of investments.

4 The Portfolio Separation Theorem

In this section we will see that if in addition to risky investments like stocks ($\sigma^2 > 0$) we consider a risk-free investment opportunity like a bond or savings account ($\sigma^2 = 0$), then, under reasonable conditions, the weights for the risky investments will always be positive. (In fact, bonds do incur a risk of default, but we ignore that here.) In other words, in this more realistic context, the problem of negative weights for stocks does not arise. We will observe that as risk aversion increases, more of the wealth shifts from the more speculative stocks (at least in terms of our limited data), Honda and Goodyear, to the safer stocks, ComEd and WalMart.

To proceed, we use our results to prove a beautiful and important theorem of investment economics.

Portfolio Separation Theorem. If one possible investment in a portfolio is risk-free, then the ratios of the optimal weights of the other investments do not depend on the investor's risk aversion.

Problem 4.1 Suppose that asset 1 is risk-free, so that $\sigma_1^2 = 0$.

(a) Show that

$$w_2 = \frac{\mu_2 - \mu_1}{2a\sigma_2^2}, \quad w_3 = \frac{\mu_3 - \mu_1}{2a\sigma_3^2}, \quad w_4 = \frac{\mu_4 - \mu_1}{2a\sigma_4^2}$$

and for i, j in $\{2, 3, 4\}$,

$$\frac{w_i}{w_j} = \frac{(\mu_i - \mu_1)/\sigma_i^2}{(\mu_j - \mu_1)/\sigma_j^2}.$$

(b) The portion of the total wealth in risky assets that is devoted to each single risky asset does not depend on the risk aversion. For example, show for risky asset 2 that

$$\frac{w_2 W}{w_2 W + w_3 W + w_4 W} = \frac{\frac{\mu_2 - \mu_1}{\sigma_2^2}}{\frac{\mu_2 - \mu_1}{\sigma_2^2} + \frac{\mu_3 - \mu_1}{\sigma_3^2} + \frac{\mu_4 - \mu_1}{\sigma_4^2}}.$$

This means that any investor working with the same knowledge about rates of return and variance, and using the type of objective that we are using, will hold these investments in the same proportion to each other, regardless of the degree of risk aversion. These constant relative proportions describe what economists call the *market portfolio*. The role of the risk aversion a is only to determine what proportion of wealth is devoted to the risk-free asset, not the relative mixture of the risky assets.

(c) Show that if all risky assets have higher rates of return than the risk-free asset, then the weights of the risky assets w_2, w_3, and w_4 are all positive. (The weight w_1 of the risk-free asset may still be negative if our risk aversion is small. A negative w_1 in this case has the interpretation that we should borrow money to buy more of the risky assets.)

Problem 4.2 Suppose your investment opportunities are stocks in Exxon, Hormel, and Abbot Labs (see the data in Problem 3.5), and a bond with $\mu = .050\%$ and $\sigma^2 = 0$. What is your risk-averse optimal portfolio if your risk aversion is $a = 4$? If $a = 8$? If $a = 12$? (What does it mean if the weight you should invest in the bond is negative?) What is the market portfolio of the three stocks?

Problem 4.3 Consider a portfolio problem with five assets, the first risk-free, and with the extra constraint that assets 2 and 3 must have equal portions of the total wealth. Does a portfolio separation theorem still hold in this case? (Note. For this problem you need an extra Lagrange multiplier for the extra constraint.)

5 Solutions

Problem 2.1 Invest all in stock 2; 10%.

Problem 2.2 The risk aversion is $\frac{.08-.05}{.04-.01} = 1$.

Problem 2.3 The minimum variance in the dependent case occurs for $w_1 = 2/5$, $w_2 = 3/5$, and in the independent case it occurs for $w_1 = 1/3$, $w_2 = 2/3$. The variances come out to $2/3$ in the independent case and $1/5$ in the dependent case, so the reduction is $7/15$.

Problem 3.1 The partials of L with respect to the μ_i are of the form $\mu_i - 2a\sigma_i^2 w_i - \lambda$, and so $\lambda = \mu_1 - 2a\sigma_1^2 w_1$. Equating this expression for λ to the corresponding expressions $\mu_i - 2a\sigma_i^2 w_i$ for each of the other variables and solving for w_i easily gives the first group of equations. Setting $w_1 + w_2 + w_3 + w_4 = 1$ and solving for w_1 yields the second expression.

Problem 3.2 (a) For $a = 2$, $w_1 = 3/16$, $w_2 = 3/8$, $w_3 = 7/16$. With these weights, the expected return is $.05(3/16) + .08(3/8) + .12(7/16) = .091875$. For $a = 3$, $w_1 = 11/24$, $w_2 = 1/4$, $w_3 = 7/24$, and the expected return is $.05(11/24) + .08(1/4) + .12(7/24) = .0779167$. (b) To simply maximize the portfolio expected rate of return, ignoring risk, you would invest all wealth in asset 3 at 12% expected return. The effect of increasing a is to sacrifice expected return for the sake of reducing risk. As risk aversion grows, the balance of assets in the portfolio leans toward the risk-free asset and away from the risky assets.

Problem 3.3 (a) When $a = 12$, $w_1 = .21$, $w_2 = .40$, $w_3 = .36$, $w_4 = .03$. The expected rate of return is .422% or \$42.20 on the \$10,000 investment. (b) About \$2089 is going to stock 1 (Walmart), \$3999 to stock 2 (Goodyear), \$3597 to stock 3 (Honda), and \$314 to stock 4 (ComEd). The dollar returns are about \$4.43, \$24.80, \$12.66, and \$0.36 which up to rounding do total to the overall dollar return as they should.

Problem 3.4 (a) When $a = 2$, $w_1 = -.12$, $w_2 = 1.84$, $w_3 = 1.11$, $w_4 = -1.83$. The expected rate of return is about 1.3% or \$130 on the \$10,000 investment. (b) About $-$\$1227 is going to stock 1 (Walmart), \$18,389 to stock 2 (Goodyear), \$11,138 to stock 3 (Honda), and $-$\$18300 to stock 4 (ComEd). The dollar returns are about $-$\$2.60, \$114.01, \$39.21, and $-$\$20.68.

Problem 3.5 (a) The weights are about $-.45$, .04, .27, and 1.15; (b) For the second question, the issue of which is better turns on the size of $(w_2 + w_3)^2$ relative to $w_2^2 + w_3^2$.

Problem 4.1 Setting $\sigma_1^2 = 0$ in the expressions for the w_j in Problem 3.1 immediately gives $w_j = \frac{\mu_j - \mu_1}{2a\sigma_j^2}$, and taking the quotient of w_i and w_j allows the a factors to divide out. (b) The dollar value in asset j is $w_j W$, so the portion of total wealth in asset j is $w_j W/(w_2 W + w_3 W + w_4 W)$. The W factors divide out, yielding the desired expression. (c) By the form $w_j = \frac{\mu_j - \mu_1}{2a\sigma_j^2}$, as long as $\mu_j > \mu_1$ the weight will be positive.

Problem 4.2 For risk aversion $a = 4$, the bond weight $w_1 = -0.405$, meaning that we borrow bonds, and the stock weights are $w_2 = .398$, $w_3 = .362$, and $w_4 = .645$. For risk aversion $a = 8$, the bond weight is $w_1 = .298$, so that we no longer borrow, and the stock weights are $w_2 = .199$, $w_3 = .181$, and $w_4 = .322$. For risk aversion $a = 12$, we see the bond weight going up even higher to $w_1 = .532$, and the stock weights are reduced to $w_2 = .132$, $w_3 = .121$, and $w_4 = .215$.

Problem 4.3 By setting up the Lagrangian again with four variables w_1, w_2, w_4, and w_5, objective function

$$\begin{aligned} f(w_1, w_2, w_4, w_5) &= w_1\mu_1 + w_2\mu_2 + w_2\mu_3 + w_4\mu_4 + w_5\mu_5 \\ &\quad - a\left(w_2^2\sigma_2^2 + w_2^2\sigma_3^2 + w_4^2\sigma_4^2 + w_5^2\sigma_5^2\right) \end{aligned}$$

and constraint $g(w_1, w_2, w_4, w_5) = w_1 + 2w_2 + w_4 + w_5 - 1 = 0$, you can derive as in Problem 3.1 that w_4 and w_5 satisfy the same equations as in Problem 4.1, and w_2 satisfies an analogous equation with μ_2 replaced by the average of μ_2 and μ_3, and with σ_2^2 replaced by the average of σ_2^2 and σ_3^2. Thus, a separation theorem will hold.

References

[1] Robert V. Hogg, and E.A. Tanis, *Probability and Statistical Inference, 3rd ed.*, Macmillan, New York, 1988.

[2] Harry Markowitz, “Portfolio Selection,” *Journal of Finance* **7** (1952) 77–91.

[3] William Sharpe, *Portfolio Theory and Capital Markets*, McGraw-Hill, 1970.

[4] Paul A. Samuelson, *Foundations of Economic Analysis*, Harvard University Press, Cambridge, 1983.

8

Topics in Modern Finance

Frank Wang
LaGuardia Community College, The City University of New York

1 Introduction

Two most important concepts in investing are *risk* and *return*. Much of the economic study of these concepts falls within the sphere of modern finance, which has three major building blocks: Modern Portfolio Theory, the Capital Asset Pricing Model (CAPM, pronounced "cap-em"), and the Black-Scholes formula. In 1952, Harry Markowitz pioneered the Modern Portfolio Theory. His seminal paper [7] established the framework for selecting investments to reduce risk. Extending Markowitz's theory, William Sharpe devised the CAPM, that describes the relationship between risk and expected return [8]. Fisher Black and Myron Scholes, assisted by Robert Merton, found an analytic formula for valuing option contracts and assessing risk [1]. Their model is considered the most important single breakthrough in the 1970s. Essentially, these intellectual giants addressed three fundamental questions: What is risk? What is an asset worth? What is risk worth? For their contributions to financial economics, Markowitz, Sharpe, and Merton Miller were awarded the Nobel Prize for Economic Sciences in 1990. Scholes and Merton received the Nobel in 1997. (Black had died in 1995.) This paper guides the reader through a series of problems to discover how undergraduate mathematics are used in financial modeling.

Investors have known for long that risk has its reward, but the precise meaning of risk and reward had been elusive. Markowitz used statistical notion to quantify these concepts. He showed that, under certain conditions, an investor's portfolio can be reduced to balancing between the expected return, which is the weighted mean, and its risk, measured as its variance. Section 2 reviews statistical terminology crucial to this paper. Section 3 is devoted to the portfolio selection and involves calculus techniques. When Sharpe was a graduate student, he was asked by Markowitz to simplify the portfolio model. Surprisingly, he found that the expected return on an asset is a linear function of risk. This linear function, called the Security Market Line, will be derived in Section 4. In Section 5, stochastic processes and their application to option pricing are introduced. Our rudimentary treatment relies on computer simulations. Section 6 is an afterword that discusses criticisms of the orthodox theories, and their roles in major financial crises. It is useful to read the entire paper to gain an overview of modern finance, but each section is self-contained and can be used independently. For some tasks, *Maple*, *Mathematica*, and Excel are useful, and several problems have hints on key commands so that the reader can search for online help. Solutions of most problems are included in Section 7.

2 Statistical Background

Probability theory is the mathematical representation of uncertainty. Though probability as a mathematical axiomatic theory is well known, its interpretation is still the subject of debate. The economist John Maynard Keynes regarded probability as "intensity of belief," while Richard von Mises' interpretation of probability is "relative frequency." The two interpretations of probability are complementary: some parameters are obtained from subjective judgment, others from historical data. In financial analysis, we generally face different sources of risk. One kind of risk is associated with uncertain outcomes, given a known probability distribution; the other kind arises because our assumed distribution may itself be incorrect.

Since we do not know probability distributions of security returns, we use historical data to illustrate one of the many possible ways to estimate risk measures. Consider the closing prices of the IBM and HP stocks at the end of the year from 2002 to 2007, retrieved from `http://finance.yahoo.com`.

	IBM Stock		HP Stock	
Year	Dividend	Year-End Price	Dividend	Year-End Price
2002	—	$77.50	—	$17.36
2003	$0.63	92.68	$0.32	22.97
2004	0.70	98.58	0.32	20.97
2005	0.78	82.20	0.32	28.63
2006	1.10	97.15	0.32	41.19
2007	1.50	108.10	0.32	50.48

Table 1. IBM and HP closing prices from 2002 to 2007.

A stock's realized rate of return in any period t is defined as

$$R_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}},$$

where P_t and P_{t-1} are the respective closing stock prices at time t and $t-1$, and D_t is the dividend paid to the shareholders between time $t-1$ to t. The mean and variance of the returns of stock a are denoted

$$\bar{R}_a = \frac{1}{n}\sum_{t=1}^{n} R_{at}, \quad s_a^2 = \frac{1}{n-1}\sum_{t=1}^{n}(R_{at} - \bar{R}_a)^2.$$

The covariance between the returns of stocks a and b are

$$s_{ab} = \frac{1}{n-1}\sum_{t=1}^{n}(R_{at} - \bar{R}_a)(R_{bt} - \bar{R}_b).$$

It is a measure of the tendency of the stocks to rise or fall together. Note that $s_{aa} = s_a^2$. A related concept is the correlation coefficient

$$\rho_{ab} = \frac{s_{ab}}{s_a s_b},$$

whose value is between -1 and 1, inclusive. In the example in Figure 1, the correlation coefficient is $\rho_{ab} = 0.694$, which is related to the slope of the regression line, $\rho_{ab}\frac{s_b}{s_a} = 0.605$.

Problem 2.1

(a) Use the data given in Table 1 to verify that the sequence of annual returns on the IBM stock are $\{0.204, 0.0712, -0.158, 0.195, 0.128\}$. Then verify that the mean is $\bar{R} = 0.0880$, and the standard deviation is $s = 0.148$. All values are in units of per annum.

(b) Based on the data given in Table 1, find the mean and the variance of annual returns on HP stock in the given period.

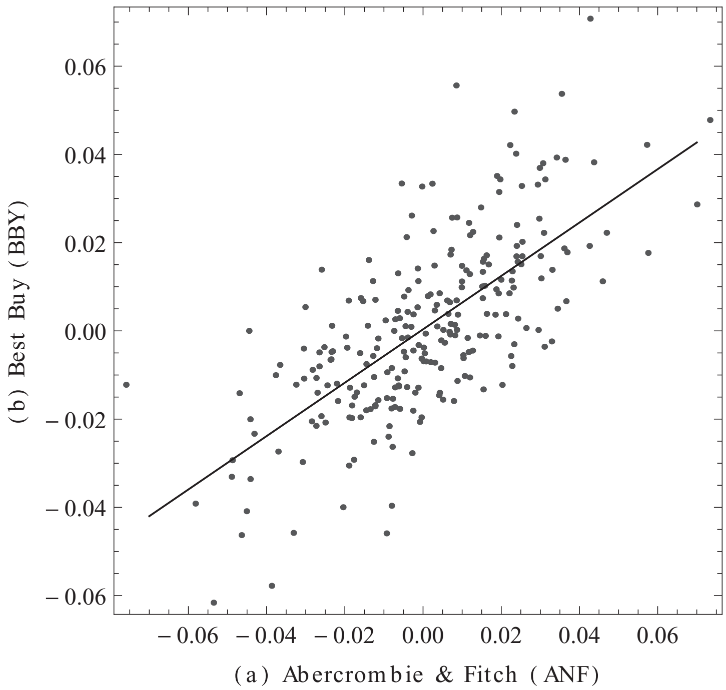


Figure 1. The daily returns of BBY and ANF, July 1, 2007–June 30, 2008.

(c) Verify that the covariance between the returns of the IBM and HP stocks is 0.00116. What is the correlation coefficient between these two stocks?

Problem 2.2 Select two of your favorite stocks on publicly traded companies. Visit `http://finance.yahoo.com`, and download monthly closing prices for three years to a worksheet.

(a) Find the realized rates of return, variances, and covariance.

(b) Make a scatter plot of these two securities, and find the slope of the regression line.

Hint: Excel's `average`, `var` and `covar` functions can be invoked for the obvious purpose.

3 Modern Portfolio Theory

Conventional wisdom tells us, "don't put all your eggs in one basket." In technical terms, this addresses the benefits of diversification. Instead of investing in just one stock, financial advisors generally advise customers to invest in a portfolio, which is a combination of stocks. A stock is characterized by its expected return, measured by its mean, and risk, measured by its variance. A portfolio's overall mean and variance are functions of the portfolio's *weight* in each stock. An *efficient portfolio* is one whose risk is the lowest for an expected return. In Section 3.1, we discuss a two-asset portfolio, so that students can apply single variable calculus to understand the rationale of diversification. In Section 3.2, we study a three-asset portfolio, which requires techniques of multivariable calculus. And in Section 3.3, we extend to n-asset portfolios using the language of linear algebra, which is convenient for computer implementation.

3.1 Two-Asset Portfolio

Consider a portfolio consisting of Stock A and Stock B, with respective expected returns (however decided upon) of μ_a and μ_b and variances σ_a^2 and σ_b^2. The correlation coefficient between these two stocks is ρ_{ab}. Let x be the weight of the portfolio invested in Stock A, and $1-x$ in Stock B. The portfolio's expected return is the weighted mean

$$\mu = x\mu_a + (1-x)\mu_b, \tag{1}$$

and its variance of this two-asset portfolio is

$$\sigma^2 = x^2\sigma_a^2 + (1-x)^2\sigma_b^2 + 2x(1-x)\rho_{ab}\sigma_a\sigma_b. \tag{2}$$

The portfolio is said to be attainable if $x \in [0, 1]$, meaning that an investor does not "hold" a negative amount of the stocks by borrowing from a broker. We remark that risk can refer to variance, or to standard deviation, which is the square root of variance. Variance is additive, but standard deviation is not. The standard deviation has the same unit as expected return (for example, 30% per annum). Note furthermore that the minimum of σ occurs at the same value of x as the minimum of σ^2, and the latter is easier to work with. In fact, one can use calculus or graphical reasoning to prove that a positive function $f(x)$ has an absolute minimum $f(x_0)$ at $x = x_0$ if and only if $f^2(x)$ has absolute minimum $f^2(x_0)$ at $x = x_0$.

Problem 3.1 One investor is interested in the stocks of Abercrombie & Fitch (NYSE: ANF) and Best Buy (NYSE: BBY). According to Standard & Poor's 500 Guide [9], the expected returns (per annum) on ANF and BBY are $\mu_a = 0.235$ and $\mu_b = 0.421$, respectively. Using historical rates of return, we found that the standard deviations (per annum) are $\sigma_a = 0.365$ and $\sigma_b = 0.436$, and the correlation coefficient is $\rho_{ab} = 0.694$. Determine the fraction x of the portfolio that should be invested in ANF if we wish to form the least-risky portfolio. What is the standard deviation of the least-risky portfolio, and how does it compare with the standard deviations of the individual stocks?

Problem 3.2

(a) Assume $\sigma_b > \sigma_a$. Express σ^2 as a quadratic function of x, and show that the vertex of the parabola occurs at

$$x_c = \frac{\sigma_b(\sigma_b - \rho_{ab}\sigma_a)}{\sigma_a^2 + \sigma_b^2 - 2\rho_{ab}\sigma_a\sigma_b}.$$

(b) Show that this parabola opens upward. Thus there is a global minimum.

(c) Show that the condition $-1 \le \rho_{ab} < \frac{\sigma_a}{\sigma_b}$ is required for $0 < x_c < 1$. Furthermore, explain why σ evaluated at such x_c is less than σ_a and σ_b.

(d) We can build a portfolio consisting of two portfolios. Consider the US large cap equity and EAFE (stock market index measuring markets outside the US) in Table 2. Find the portion you should invest in each so that the combined portfolio has minimum risk. What is your portfolio's expected return?

Expected return (μ_i)	Std. dev. of return (σ_i)	Asset class return correlations (ρ_{ij})		1	2	3	4
6.4%	4.7%	US bonds	1	1.00			
10.8	14.9	US large cap equity	2	0.32	1.00		
11.9	19.6	US small cap equity	3	0.06	0.76	1.00	
11.5	17.2	EAFE international equity	4	0.17	0.44	0.38	1.00

Table 2. From F. J. Fabozzi, H. M. Markowitz, *The Theory and Practice of Investment Management*, Wiley (2002), p. 51. See [3] also.

Problem 3.3 Choose two of your favorite stocks. Visit `http://finance.yahoo.com` and obtain their "one-year target estimate," which can be used for the expected returns. You can also download the daily closing prices of the past year to a spreadsheet, and calculate the variances and correlation coefficient. How would you invest in these two stocks so that the risk is minimal?

3.2 Three-Asset Portfolio

Consider three securities with expected returns μ_i, standard deviations σ_i, and correlations ρ_{ij}, for $i = 1, 2, 3$ and $j = 1, 2, 3$. Let x, y, and $1 - x - y$ be the weight of the portfolio invested in Securities 1, 2, and 3, respectively. The expected return μ and risk σ of such a portfolio is given by

$$\mu = x\mu_1 + y\mu_2 + (1-x-y)\mu_3, \tag{3}$$

and

$$\sigma^2 = x^2\sigma_1^2 + y^2\sigma_2^2 + (1-x-y)^2\sigma_3^2 + 2xy\rho_{12}\sigma_1\sigma_2 + 2y(1-x-y)\rho_{23}\sigma_2\sigma_3 + 2x(1-x-y)\rho_{13}\sigma_1\sigma_3. \quad (4)$$

We exclude negative values of weights, which gives the conditions

$$x \geq 0, \quad y \geq 0, \quad 1-x-y \geq 0. \quad (5)$$

The attainable set refers to the set of (x, y) that satisfy these conditions in (5). To find the minimum variance portfolio, we solve for x and y that satisfy $\frac{\partial\sigma^2}{\partial x} = 0$ and $\frac{\partial\sigma^2}{\partial y} = 0$. To find the minimum variance portfolio for a *given* mean (for an investor who tolerates greater risk for a higher expected return), we need to apply Lagrange's method, that is to find x, y, and λ such that $\nabla\sigma^2 = \lambda\nabla\mu$ (∇ is the gradient operator) for a specific value of μ in (3). Problem 3.4 illustrates these optimization methods.

Problem 3.4 Consider 3 stocks with these data: $\mu_1 = 0.10$, $\mu_2 = 0.15$, $\mu_3 = 0.20$; $\sigma_1 = 0.28$, $\sigma_2 = 0.24$, $\sigma_3 = 0.25$; $\rho_{12} = -0.10$, $\rho_{23} = 0.20$, $\rho_{13} = 0.25$. Here μ_i and σ_i are in units of per annum.

(a) To determine the minimum variance portfolio, substitute the numerical values into (4) to express σ^2 as a quadratic function of x and y. Then determine values x_c and y_c that minimize σ^2. Is (x_c, y_c) in the attainable set? How does the minimum value $\sigma_{\min}$ compare with the standard deviations of the individual stocks?

(b) To find the portfolio with the minimum σ^2 for a specific μ, use the Lagrangian $L = \sigma^2 - \lambda[\mu - x\mu_1 - y\mu_2 - (1-x-y)\mu_3]$, where λ is the yet undetermined Lagrange's multiplier. Solve for x, y and λ that satisfy $\frac{\partial L}{\partial x} = 0$, $\frac{\partial L}{\partial y} = 0$, and $\frac{\partial L}{\partial \lambda} = 0$. The solutions x_c and y_c should depend linearly on μ. Show that your solutions x_c and y_c yield the minimum variance $\sigma^2_{\min} = 9.85\mu^2 - 2.89\mu + 0.237$ at a specific μ.

(c) In the (σ, μ)-plane, sketch the minimum variance curve using $\sigma_{\min} = \sqrt{\sigma^2_{\min}}$ obtained in part (b). This curve is called the Markowitz bullet.

(d) On your graph in (c), mark at least 10 pairs of μ and σ calculated from different mixes of stocks 1, 2, and 3 in the attainable set. Observe that all these points are on or to the right of the Markowitz bullet. How do you interpret this result?

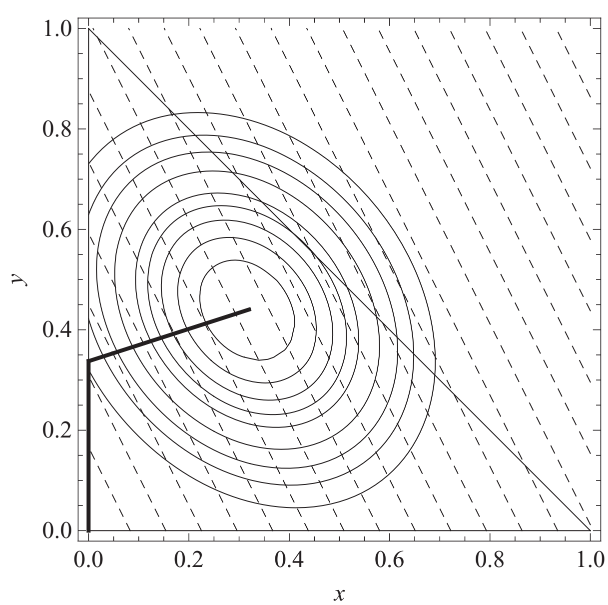


Figure 2. The optimization problem with three assets visualized in the (x, y)-plane.

Figure 2 illustrates the three-asset optimization in the (x, y)-plane. The boundary of the attainable set is the triangle with vertex $(0, 0)$, $(1, 0)$, and $(0, 1)$. As functions of two variables x and y, the graph of μ is a plane, and the graph of σ^2 is an elliptic paraboloid. In the (x, y)-plane, the level curves are lines and ellipses, called isomeans and isovariances, respectively. At a specific μ, represented by one isomean, the least-risky weight is the point (x_c, y_c) at which the isovariance is tangent to that isomean. If we trace these weights, it will be a *critical line* (recall that in Problem 3.4 (b), x_c and y_c are linear functions of μ). The thick line in Figure 2 represents the critical line starting out from the point that minimizes σ^2 (the critical point found in Problem 3.4 (a)) to the boundary of the attainable set, then follow the boundary (as a two-asset portfolio) to the single asset with the greatest return.

3.3 Multi-Asset Portfolio

As we add more assets to a portfolio, it is more convenient to adopt matrix notation. Let

$$\mathbf{m} = (\mu_1 \quad \mu_2 \quad \cdots \quad \mu_n)$$

be the expected return vector, and

$$C = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$$

be the covariance matrix, where $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$, and $\sigma_{ii} = \sigma_i^2$. Let

$$\mathbf{w} = (w_1 \quad w_2 \quad \cdots \quad w_n)$$

be the weight vector, and

$$\mathbf{u} = (1 \quad 1 \quad \cdots \quad 1)$$

be the row vector of unity. With the transposes of row vectors denoted by a superscript T, and the inverse matrix of C written as C^{-1}, the portfolio with the minimum risk at a given μ has the weight

$$\mathbf{w} = \frac{\begin{vmatrix} 1 & \mathbf{u}C^{-1}\mathbf{m}^T \\ \mu & \mathbf{m}C^{-1}\mathbf{m}^T \end{vmatrix} \mathbf{u}C^{-1} + \begin{vmatrix} \mathbf{u}C^{-1}\mathbf{u}^T & 1 \\ \mathbf{m}C^{-1}\mathbf{u}^T & \mu \end{vmatrix} \mathbf{m}C^{-1}}{\begin{vmatrix} \mathbf{u}C^{-1}\mathbf{u}^T & \mathbf{u}C^{-1}\mathbf{m}^T \\ \mathbf{m}C^{-1}\mathbf{u}^T & \mathbf{m}C^{-1}\mathbf{m}^T \end{vmatrix}}, \tag{6}$$

provided that the determinant in the denominator is nonzero. The weights depend linearly on μ, as in the three-asset case. The proof, which essentially extends the constrained optimization done in Problem 3.4, can be found in [2].

Problem 3.5

(a) With 3 securities, let $\mathbf{w} = (x \quad y \quad 1 - x - y)$. Show that $\mu = \mathbf{m}\mathbf{w}^T$ and $\sigma^2 = \mathbf{w}C\mathbf{w}^T$ give (3) and (4), respectively.

(b) For the data in Problem 3.4, use (6) to find the weight for the portfolio with minimum variance at a given μ. Then find this minimum variance by calculating $\sigma^2 = \mathbf{w}C\mathbf{w}^T$. The answer should be the same as that in Problem 3.4 (c).

(c) Consider a portfolio consisting of the 4 securities in Table 2 of Problem 3.1. For given μ, show that $\sigma^2_{\min} = 7.48\mu^2 - 0.978\mu + 0.0340$. You may use *Maple* or *Mathematica* for matrix operations involved.

4 The Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM), developed by William Sharpe, is based on Markowitz's portfolio theory discussed in Section 3. Consider a number of risky securities such as stocks. The ith stock has expected return characterized by its mean μ_i, and risk characterized by its standard deviation σ_i. A portfolio, which is a weighted combination of stocks, has an overall mean and standard deviation. By plotting the expected return against the standard deviation for every possible combination of stocks, a shaded area is formed in the (σ, μ)-plane, see Figure 3. Markowitz established a procedure to find the encompassing boundary of these points, called the Markowitz bullet.

Among portfolios, a rational investor will choose one with the highest return for a given standard deviation, and we define such a portfolio as an *efficient portfolio*. The set of efficient portfolios is the efficient frontier. Thus, the upper half of the Markowitz bullet is the efficient frontier of risky securities.

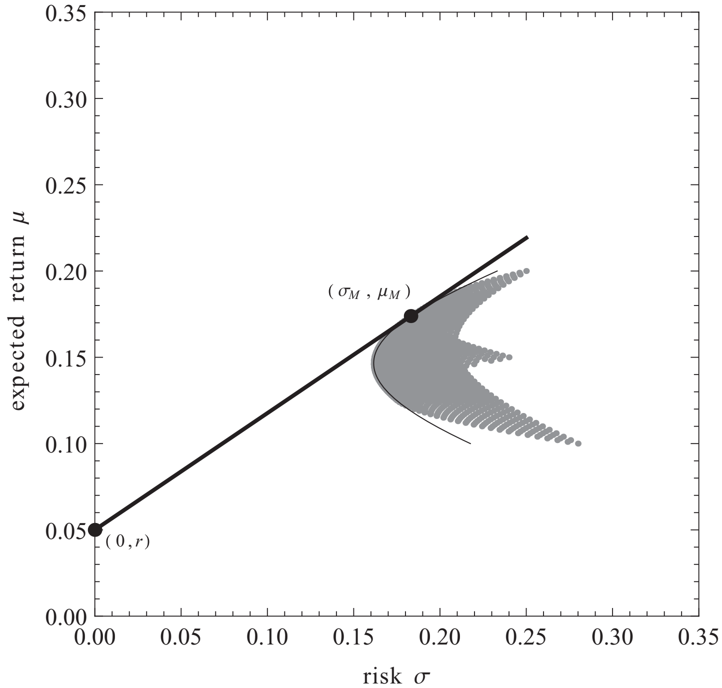


Figure 3. Attainable risky portfolios (shaded area) whose left boundary is the Markowitz bullet, and the Capital Market Line.

Sharpe added a risk-free asset (e.g., a bond) with a return denoted by r to the model; this asset is the point $(0, r)$ in Figure 3. The investor can choose to place x in the risk-free asset and the remainder $1 - x$ in a risky portfolio (one point in the shaded area in Figure 3). If the investor can lend or borrow at the risk-free rate r, then x can be any value. Sharpe discovered that after the risk-free asset is introduced, the efficient frontier changes to a line through the risk-free rate of return, and is tangent to the boundary of the risky portfolios at the point $M = (\sigma_M, \mu_M)$. This point of tangency is referred to as Portfolio M, which is discussed below. This line is called the Capital Market Line (CML). The significance of the CML equation

$$\mu = r + \left(\frac{\mu_M - r}{\sigma_M}\right)\sigma \tag{7}$$

is that the expected return μ on a portfolio is equal to the risk-free rate r plus a risk premium, which is the market price of risk $\frac{\mu_M - r}{\sigma_M}$ times the quantity of risk σ.

Problem 4.1 In Section 3.1, we considered a portfolio consisting of two risky assets. The weighted mean (1) and the standard deviation (2) can be regarded as parametric equations of the weight x. The parametrized curve $\langle\sigma(x), \mu(x)\rangle$ is generally not a straight line.

(a) To see the effect of including a risk-free asset with Portfolio M, use $\mu_a = r$ and $\mu_b = \mu_M$ in (1), and $\sigma_a = 0$ and $\sigma_b = \sigma_M$ in (2). What type of functions are $\sigma(x)$ and $\mu(x)$?

(b) Show that by eliminating the parameter, you obtain the CML equation.

Economists have demonstrated that the point of tangency $M = (\sigma_M, \mu_M)$ on the CML must consist of all risky securities available to investors. Therefore, M is the market portfolio, or simply the market. In practice, the market portfolio is approximated by a suitable stock exchange index.

Sharpe next considered an individual stock i in the market, and found the Security Market Line (SML) for pricing it in terms of relative risk to the market.

Problem 4.2 Consider a hypothetical portfolio consisting of a stock i with expected return μ_i and standard deviation σ_i and the market portfolio with expected return μ_M and standard deviation σ_M. Let ρ_{iM} be the correlation coefficient between stock i and the market portfolio M. Then, from (1) and (2), this portfolio has mean and standard deviation

$$\mu = x\mu_i + (1-x)\mu_M, \quad \sigma = \left[x^2\sigma_i^2 + (1-x)^2\sigma_M^2 + 2x(1-x)\rho_{iM}\sigma_i\sigma_M\right]^{1/2},$$

where x is the portion of wealth invested in Stock i. We want to find out how μ_i and σ_i are related.

(a) Use the above parametric equations in x for μ and σ to show that the derivative $\frac{d\mu}{d\sigma}$ evaluated at $x = 0$ is

$$\left.\frac{d\mu}{d\sigma}\right|_{x=0} = \frac{\mu_i - \mu_M}{\rho_{iM}\sigma_i - \sigma_M}.$$

(b) At $x = 0$, the portfolio is just the market portfolio M. Because the CML is tangent to the boundary of risky portfolios at M, we conclude that $\left.\frac{d\mu}{d\sigma}\right|_{x=0}$ is the same as the slope of the CML in the (σ, μ)-plane. By equating $\left.\frac{d\mu}{d\sigma}\right|_{x=0}$ to the slope of the CML, show that

$$\mu_i = r + (\mu_M - r)\frac{\rho_{iM}\sigma_i}{\sigma_M}. \tag{8}$$

This is the SML equation.

The ratio $\frac{\rho_{iM}\sigma_i}{\sigma_M}$ in (8), referred to as β among financial professionals, can be estimated empirically. It is the slope of the regression line for the returns of Stock i and the market portfolio M. Therefore, the value β is a measure of how the stock reacts to the market's rise and fall. Betas are calculated and published by Merrill Lynch, Standard & Poor's, and many other organizations. The rewritten SML equation

$$\mu_i = r + (\mu_M - r)\beta \tag{9}$$

is the heart of Sharpe's model. It shows that the expected return on an individual asset is a positive linear function of its β, construed as risk relative to the market. Investors are compensated for exposure to this relative risk; the higher the β value, the higher the expected return. For example [6], suppose the stock of Hot TechCo has a β of 1.5. Using U.S. Treasury bill's risk-free return rate $r = 2\%$, and S & P 500's 11% return rate for the market's expected return μ_M, the SML equation expects Hot TechCo stock's return rate to be 15.5%.

5 The Black-Scholes Formula

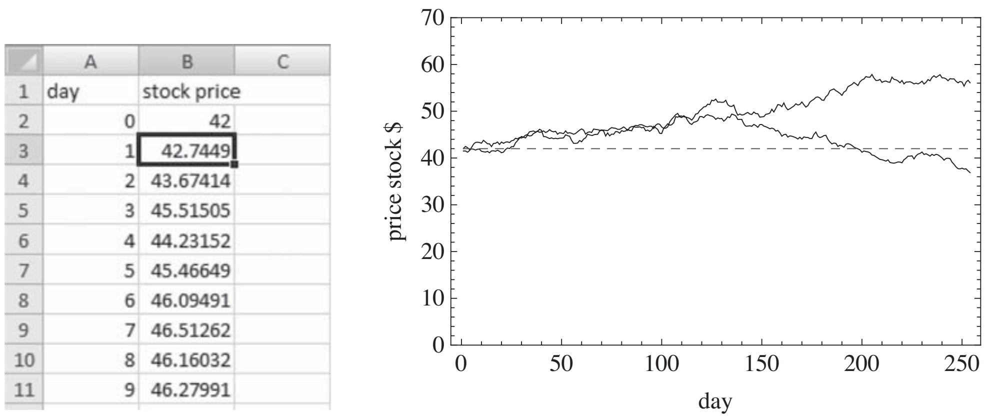
The CAPM in the preceding section shows that the expected return on an asset is a linear function of its β value, a measure of the stock's sensitivity to the fluctuations of the market. See equation (9). Following the introduction of the CAPM, Fisher Black teamed up with Myron Scholes and discovered an equation that gave the *exact* value of a risky derivative production called an "option." In this section, we use a computer to simulate the trajectory of a stock's prices, and use it to estimate the stock's option price. The results from simulation will be compared with the celebrated Black-Scholes formula.

Stock options are contracts to buy or sell a company's stock at a fixed price within a given duration. We focus on the European call option that has a specific due date, which works as follows. Suppose you owned 100 shares of XYZ stock, which is priced at \$138 per share. You could sell someone the right to buy your shares 2 months from now at a price of \$140 per share. The \$140 is called the *exercise price*. (Such options are traded on a number of exchanges, with the Chicago Board Options Exchange being the oldest and the largest.) If the stock price on this date is less than \$140, one will clearly choose not to exercise and lose only the purchase price paid for the option. However, if the price per share is \$155, the investor can buy your 100 shares for \$14,000 and immediately sell them for \$15,500 for a gain of \$1,500. Options are often a form of compensation for many executives or employees. They are also used for insurance. For some speculators they are another way to bet on the stocks themselves. The option's huge capital gains potential, combined with its loss limitation, is clearly worth paying something for. The Black-Scholes formula gives the exact value of the European option. For an excellent introduction to options and the Black-Scholes formula, see the widely used textbook [4].

One of the assumptions in the Black-Scholes model is that stocks advance as with geometric Brownian motion. The discrete-time model is

$$S_{n+1} = S_n + \mu S_n \Delta t + \sigma S_n \epsilon \sqrt{\Delta t}, \tag{10}$$

where S_n is the stock price at time $n\Delta t$ ($n = 0, 1, 2, \ldots$), μ is the stock's expected rate of return, σ is the stock's risk or volatility, and ϵ is a random number drawn from a standardized normal distribution (using a computer).



	A	B	C
1	day	stock price	
2	0	42	
3	1	42.7449	
4	2	43.67414	
5	3	45.51505	
6	4	44.23152	
7	5	45.46649	
8	6	46.09491	
9	7	46.51262	
10	8	46.16032	
11	9	46.27991	

Figure 4. Simulation of two stock trajectories based on data in Problem 5.1. To use Excel for this purpose, after entering the initial price at `B2`, the formula in `B3` reads `=B2+0.12*B2*1/254+0.2*B2*norminv(rand())*sqrt(1/254)`. This formula is copied down column `B`. To recalculate another trajectory, Press the function key F9.

Problem 5.1 Consider a European call option on a stock when the current stock price is $S = \$42$, the exercise price is $X = \$40$, the risk-free interest rate is $r = 10\%$ per annum, the stock's expected return is $\mu = 12\%$ per annum and volatility is $\sigma = 20\%$ per annum, and the time to maturity T is 1 year.

(a) To see how probable the stock price will fall after one year despite a rather high expected return, use (10) to simulate 20 trajectories each beginning with $S = \$42$ on day 0. Use $\Delta t = 1/254$, where 254 is the number of trading days in a year. What is the relative frequency that a year from now the stock's price S_T actually drops below its initial price S?
Hint: With *Maple*, normally distributed random numbers are generated using the `Generate(distribution (Normal(0,1)))` command under `RandomTools` library. With *Mathematica*, it is done using `RandomReal [NormalDistribution[0,1]]`. See Figure 4 for Excel usage.

(b) It can be shown [4] that the change of the logarithm of a stock's prices during the time interval T is normally distributed,

$$Y \equiv \ln S_T - \ln S \sim \phi[(\mu - \sigma^2/2)T, \sigma\sqrt{T}],$$

where S and S_T are defined in the previous part, and $\phi(m, s) = \frac{1}{\sqrt{2\pi s^2}} \exp[-\frac{(Y-m)^2}{2s^2}]$ is the probability density function of a normal distribution. Verify that after one year, the probability that the stock's price drops below its initial price ($S = \$42$) is 0.31.

(c) To find the value of the European call option on this stock, generate 10,000 trajectories using the risk-free interest rate r in place of μ in (10). (This is in accordance with the important discovery by Black and Scholes.) For each trajectory, the payoff is $\max(S_T, 0)$ at a future time T. The expected payoff at T is the mean of these 10,000 possible outcomes. To obtain the current value of the option, multiply the mean by $\exp(-rT)$.

(d) The Black-Scholes formula for a European call option is

$$c = S\Phi(d_1) - Xe^{-rT}\Phi(d_2), \tag{11}$$

where

$$d_1 = \frac{1}{\sigma\sqrt{T}}\left[\ln(S/X) + (r + \sigma^2/2)T\right], \quad d_2 = d_1 - \sigma\sqrt{T}$$

and $\Phi(x)$ is the cumulative probability distribution function for the standard normal distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt. \tag{12}$$

Use the formula to verify that the option's price is $c = \$6.83$.

(e) One way to estimate a stock's volatility is to observe an option price of the stock in the market. Suppose the value of a European call option is \$1.88 when $S = \$21$, $X = \$20$, $r = 0.1$ per annum, and $T = 0.25$ year. Use (11) to produce a table of values (σ, c) for $\sigma = 0.1 + 0.01n$ $(n = 0, 1, 2, \ldots, 30)$. Based on your table, what is your estimate of σ?

The Black-Scholes formula for pricing a stock option is exact, and dependent on observable variables only (the risk-free rate r and the stock's volatility σ). But we should not be blinded by the formula and forget the rather unrealistic assumptions of perfect market conditions. The Black-Scholes formula has permitted an entirely new type of trading, not only in stocks, indices, or currencies themselves, but in their volatility. Traders can construct elaborate combinations of options in order to cash in when prices swing wildly. Some analysts blame such products, called derivatives, for causing major financial crises.

6 Afterwords

When Robert Merton and Myron Scholes received the Nobel Prize in 1997, they were partners of Long-Term Capital Management (LTCM), a hedge fund company. Many hedge fund companies used complex mathematical models and have engaged in excessively risky practices for potentially great return. In 1998, some bets went terribly wrong and LTCM crashed. This threatened to unleash a chain reaction of hedge fund and bank failures around the world, but was prevented by a massive bailout by the Federal Reserve. The LTCM case prompted economists to reexamine the premises of standard financial models. One of the criticisms concerns the normal distribution assumption as in the stock price model in (10) for the Black-Scholes equation. If changes in the market's values were indeed characterized by normal distributions, it would be almost impossible that moves beyond 5 or 6 standard deviations could ever occur. Using the method in the following problem, Benoit Mandelbrot [6] exposed the serious problem of the normal distribution assumption. He illustrated that the occurrence that the Dow Jones Industrial Average swinging more than 7% should come once every 300,000 years, but the twentieth century saw more than forty such days.

Problem 6.1 Download the Dow Jones Industrial Average (symbol: ˆDJI) from 1928 to 2003 from `http://finance.yahoo.com`. Use a spreadsheet to take the logarithm of each day's average. Subtract each day's logarithm from that of the day following, to get y that approximates the daily change. According to (10), y's should follow the normal distribution.

(a) Verify that the mean is $\bar{y} = 0.02\%$ and the standard deviation is $s = 1.15\%$ for these y's.

(b) Calculate the z scores $z = (y - \bar{y})/s$ of the daily changes. How many times do you see $|z| \geq 6$?

In 2007, Goldman Sachs chief financial officer David Viniar was quoted saying, "We were seeing things that were 25-standard deviation moves, several days in a row." For the normal distribution, the probability is 3.06×10^{-138} that an event having 25-standard deviation or great occurs. If we trade 254 days a year, this event should occur about once every 1.3×10^{135} years. The age of the Universe is only 1.4×10^{10} years.

In addition to the criticism of the normal distribution assumption, experts also pointed out that to construct efficient portfolios using Markowitz's theory, one needs good forecasts of earnings, share prices, and volatility for thousands of stocks. Otherwise, garbage in, garbage out [6]. But good economic forecasts are nonexistent. A deeper problem is that many investors do not act rationally, as the models assume. The 2007–2010 housing bubble is a perfect example of the apparent irrationality of investors. After the 2008 crisis, Paul Krugman wrote an essay [5] beginning with the section title "Mistaking Beauty for Truth." As he sees it, the central cause of the economics profession's failure was the desire for an all-encompassing, intellectually elegant approach based on their idealized version of an economy in which rational individuals interact in perfect markets—the assumptions behind the models we studied in this paper. Krugman contends that we will have to acknowledge the importance of irrational and often unpredictable behavior, face up to the often idiosyncratic imperfections of markets, and accept that an elegant economic "theory of everything" is a long way off.

7 Solutions

Problem 2.1

(a) The variance is 0.0218 and the standard deviation is $\sqrt{0.0218} = 0.148$.

(b) The sequence of annual returns on the HP stock is $\{0.342, -0.0731, 0.381, 0.450, 0.233\}$. The mean is $(0.342 - 0.0731 + 0.381 + 0.450 + 0.233)/5 = 0.267$. The variance is $[(0.342 - 0.267)^2 + (-.0731 - 0.267)^2 + (0.281 - 0.267)^2 + (0.450 - 0.267)^2 + (0.233 - 0.267)^2]/(5-1) = 0.0422$, and the standard deviation is $\sqrt{0.0422} = 0.205$.

(c) The covariance s_{ab} is

$$\begin{aligned}&\frac{1}{5-1}[(0.204 - 0.0881)(0.342 - 0.267) + (0.0712 - 0.0881)(-0.0731 - 0.267)\\ &+ (-0.158 - 0.0881)(0.381 - 0.267) + (0.195 - 0.0881)(0.450 - 0.267)\\ &+ (0.128 - 0.088)(0.233 - 0.267)] = 0.00116.\end{aligned}$$

The correlation coefficient is $\rho_{ab} = 0.00116/[(0.148)(0.205)] = 0.0382$.

Problem 3.1 For the given values,

$$\begin{aligned}\sigma^2 &= x^2 0.365^2 + (1-x)^2 0.436^2 + 2x(1-x)(0.694)(0.365)(0.436)\\ &= 0.103x^2 - 0.159x + 0.190.\end{aligned}$$

Since $\frac{d\sigma^2}{dx} = 0.206x - 0.159 = 0$ at $x = 0.77$, and $\frac{d^2\sigma^2}{dx^2} > 0$, the minimal risk is $[0.103(0.77)^2 - 0.159(0.77) + 0.190]^{1/2} = 0.359$, which is less than both σ_a and σ_b.

Problem 3.2

(a) We write (2) as

$$\sigma^2 = \big(\sigma_a^2 + \sigma_b^2 - 2\rho_{ab}\sigma_a\sigma_b\big)x^2 + 2\sigma_b(\rho_{ab}\sigma_a - \sigma_b)x + \sigma_b^2$$

and solve

$$\frac{d\sigma^2}{dx} = 2\big(\sigma_a^2 + \sigma_b^2 - 2\rho_{ab}\sigma_a\sigma_b\big)x + 2\sigma_b(\rho_{ab}\sigma_a - \sigma_b) = 0$$

for the critical value $x_c = \frac{\sigma_b(\sigma_b - \rho_{ab}\sigma_a)}{\sigma_a^2 + \sigma_b^2 - 2\rho_{ab}\sigma_a\sigma_b}$.

(b) The coefficient of x^2 (which is also $\frac{1}{2}\frac{d^2\sigma^2}{dx^2}$) is greater than zero because

$$\sigma_a^2 + \sigma_b^2 - 2\rho_{ab}\sigma_a\sigma_b = (\sigma_a - \sigma_b)^2 + 2(1 - \rho_{ab})\sigma_a\sigma_b > 0.$$

(Recall that $\rho_{ab} \in [-1, 1]$.)

(c) With the assumption $\sigma_b > \sigma_a$, the numerator of x_c is greater than zero. (The denominator of x_c is greater than zero as seen above.) Solving $x_c < 1$ for ρ_{ab}, we obtain $\rho_{ab} < \frac{\sigma_a}{\sigma_b}$. Therefore, if $-1 \le \rho_{ab} < \frac{\sigma_a}{\sigma_b}$, then $0 < x_c < 1$. Furthermore, the minimum at x_c is the global minimum, so σ evaluated at this point is less than σ_a and σ_b.

(d) Using the expression for x_c in part (a),

$$x_c = \frac{17.2(17.2 - 0.44 \times 14.9)}{14.9^2 + 17.2^2 - 2(0.44)(14.9)(17.2)} = 0.63$$

for investing in US large cap equity. (EAFE portion will be $1 - 0.63 = 0.37$.) We might also substitute the numerical values into (2) to obtain $\sigma^2 = 0.0292x^2 - 0.0366x + 0.0296$, and use calculus to find the critical value $x_c = 0.63$.

Problem 3.4

(a) For the given values, (4) becomes

$$\begin{aligned}\sigma^2 &= x^2(0.28)^2 + y^2(0.24)^2 + (1-x-y)^2 0.25^2 + 2xy(-0.10)(0.28)(0.24)\\ &\quad + 2y(1-x-y)(0.20)(0.24)(0.25) + 2x(1-x-y)(0.25)(0.28)(0.25)\\ &= 0.106x^2 + 0.0526xy + 0.0961y^2 - 0.090x - 0.101y + 0.0625.\end{aligned}$$

Solving the equations

$$\frac{\partial \sigma^2}{\partial x} = 0.212x \; + \; 0.0526y \; - \; 0.090 = 0$$

and

$$\frac{\partial \sigma^2}{\partial y} = 0.526x \; + \; 0.192y \; - \; 0.101 = 0,$$

we obtain the critical point $(x_c, y_c) = (0.32, 0.44)$. The second derivatives test yields $D = \sigma^2_{xx}\sigma^2_{yy} - (\sigma^2_{xy})^2 = 0.0379 > 0$ and $\sigma^2_{xx} = 0.212 > 0$, so the critical point gives a minimum. Because x_c and y_c are greater than zero and less than 1, this ordered pair is in the attainable set. The minimum variance is

$$\begin{aligned}\sigma^2_{\text{min}} &= 0.106(0.32)^2 + 0.0526(0.32)(0.44) + 0.0961(0.44)^2\\ &\quad - 0.09(0.32) - 0.101(0.44) + 0.0625 = 0.0261,\end{aligned}$$

and $\sigma_{\text{min}} = \sqrt{0.0261} = 0.162$ is less than σ_1, σ_2, and σ_3.

(b) Using σ^2 from (a), the Lagrangian is

$$\begin{aligned}L &= 0.106x^2 + 0.0526xy + 0.0961y^2 - 0.090x - 0.101y + 0.0625\\ &\quad - \lambda[\mu - x(0.10) - y(015) - (1-x-y)(0.20)].\end{aligned}$$

Solving the three equations

$$\begin{aligned}\frac{\partial L}{\partial x} &= 0.212x + 0.0526y - 0.1\lambda - 0.090 = 0,\\ \frac{\partial L}{\partial y} &= 0.526x + 0.192y - 0.05\lambda - 0.101 = 0,\\ \frac{\partial L}{\partial \lambda} &= -\mu - 0.1x - 0.05y + 0.2 = 0,\end{aligned}$$

we obtain

$$x_c = 1.58 - 8.62\mu, \quad y_c = 0.845 - 2.77\mu, \quad \lambda = 2.89 - 19.7\mu,$$

which are linear in μ.

(c) With the expressions for x_c and y_c in (4), we obtain $\sigma^2_{\text{min}} = 9.85\mu^2 - 2.89\mu + 0.237$. The Markowitz bullet is the parametric curve $\langle \sigma_{\text{min}}, \mu \rangle = \langle \sqrt{9.85\mu^2 - 2.89\mu + 0.237}, \mu \rangle$ shown in Figure 5 (a).

(d) Figure 5 (b) shows points (σ, μ) corresponding to some arbitrary x and y within the attainable set ($x \geq 0$, $y \geq 0$, and $1 - x - y \geq 0$). These arbitrarily constructed portfolios are on or to the right of the Markowitz bullet, meaning that they have greater risk for a given expected return.

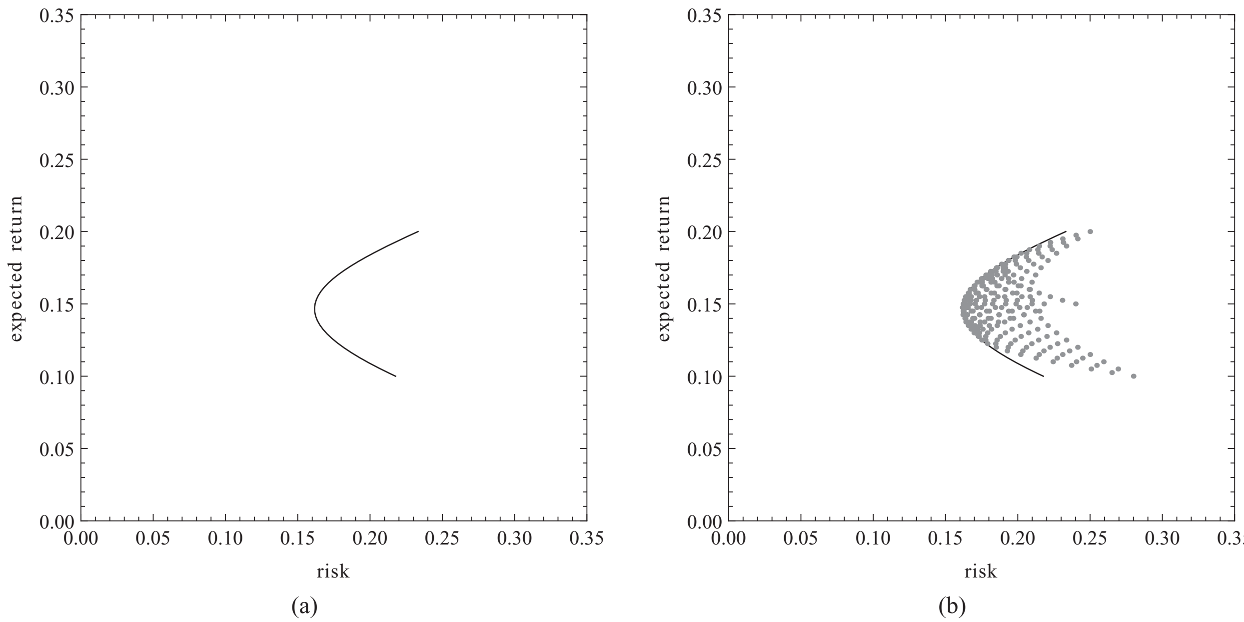


Figure 5. Solution to Problem 3.4 (c), (d)

Problem 3.5

(a) For $\mathbf{m} = (\mu_1 \quad \mu_2 \quad \mu_3)$ and $\mathbf{w} = (x \quad y \quad 1 - x - y)$,

$$\mathbf{m}\mathbf{w}^T = x\mu_1 + y\mu_2 + (1 - x - y)\mu_3 = \mu.$$

Using

$$C = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{31}\sigma_3\sigma_1 & \rho_{32}\sigma_3\sigma_2 & \sigma_3^2 \end{pmatrix},$$

we have

$$\begin{aligned} \mathbf{w}C\mathbf{w}^T &= x^2\sigma_1^2 + y^2\sigma_2^2 + (1 - x - y)^2\sigma_3^2 + 2xy\rho_{12}\sigma_1\sigma_2 \\ &\quad + 2y(1 - x - y)\rho_{23}\sigma_2\sigma_3 + 2x(1 - x - y)\rho_{31}\sigma_1\sigma_3 \\ &= \sigma^2. \end{aligned}$$

(b) For the given values, the covariance matrix is

$$\begin{aligned} C &= \begin{pmatrix} 0.28^2 & (-0.10)(0.28)(0.24) & (0.25)(0.28)(0.25) \\ (-0.10)(0.24)(0.28) & 0.24^2 & (0.20)(0.24)(0.25) \\ (0.25)(0.25)(0.28) & (0.20)(0.25)(0.24) & 0.25^2 \end{pmatrix} \\ &= \begin{pmatrix} 0.0784 & -0.00672 & 0.0175 \\ -0.00672 & 0.0576 & 0.012 \\ 0.0175 & 0.012 & 0.0625 \end{pmatrix}, \end{aligned}$$

and its inverse matrix is

$$C^{-1} = \begin{pmatrix} 13.95 & 2.544 & -4.396 \\ 2.544 & 18.55 & -4.274 \\ -4.396 & -4.274 & 18.05 \end{pmatrix}.$$

Using (6), we obtain the weights

$$\mathbf{w} = (1.58 - 8.62\mu, \quad 0.845 - 2.77\mu, \quad -1.42 + 11.4\mu)$$

and minimum variance $\sigma^2_{\min} = \mathbf{w}C\mathbf{w}^T = 9.85\mu^2 - 2.89\mu + 0.237$.

(c) For the given values, the covariance matrix is

$$C = \begin{pmatrix} 0.047^2 & (0.32)(0.047)(0.149) & (0.06)(0.047)(0.196) & (0.17)(0.047)(0.172) \\ (0.32)(.149)(.047) & 0.149^2 & (0.76)(0.149)(0.196) & (0.44)(0.149)(0.172) \\ (0.06)(0.196)(0.047) & (0.76)(0.196)(0.149) & 0.196^2 & (0.38)(0.196)(0.172) \\ (0.17)(0.172)(0.047) & (0.44)(0.172)(0.149) & (0.38)(0.172)(0.196) & 0.172^2 \end{pmatrix}$$
$$= \begin{pmatrix} 0.00221 & 0.00224 & 0.000553 & 0.00137 \\ 0.00224 & 0.0222 & 0.0222 & 0.0113 \\ 0.000553 & 0.0222 & 0.0384 & 0.0128 \\ 0.00137 & 0.0113 & 0.0128 & 0.0296 \end{pmatrix},$$

and its inverse matrix is

$$C^{-1} = \begin{pmatrix} 555.3 & -110.0 & 58.64 & -9.273 \\ -110.0 & 135.6 & -71.57 & -15.59 \\ 58.64 & -71.57 & 68.20 & -4.975 \\ -9.273 & -15.59 & -4.975 & 42.33 \end{pmatrix}.$$

Using (6), we obtain

$$\mathbf{w} = (2.36 - 20.8\mu, \quad -0.771 + 9.89\mu, \quad -0.0343 + 2.08\mu, \quad -0.553 + 8.84\mu).$$

Therefore, $\sigma^2_{\min} = \mathbf{w}C\mathbf{w}^T = 7.48\mu^2 - 0.978\mu + 0.0340$.

Problem 4.1

(a) Combining Portfolio M and a risk-free asset, we have $\mu = xr + (1-x)\mu_M$ and $\sigma = \sqrt{\sigma^2} = \sqrt{(1-x)^2\sigma_M^2} = (1-x)\sigma_M$. Notice that σ becomes a linear function of x if one asset is risk-free.

(b) From $\mu = xr + (1-x)\mu_M$, we get $x = (\mu_M - \mu)/(\mu_M - r)$. And from the standard deviation $\sigma = (1-x)\sigma_M$, we get $x = 1 - \sigma/\sigma_M = (\sigma_M - \sigma)/\sigma_M$. Therefore, $(\mu_M - \mu)/(\mu_M - r) = (\sigma_M - \sigma)/\sigma_M$, which can be simplified to the CML equation $\mu = r + (\frac{\mu_M - r}{\sigma_M})\sigma$.

Problem 4.2

(a) For the parametric equations μ and σ with parameter x, the derivative of μ with respect to σ is $\frac{d\mu}{d\sigma} = \frac{d\mu/dx}{d\sigma/dx}$. Since

$$\frac{d\mu}{dx} = \mu_i - \mu_M \quad \text{and} \quad \frac{d\sigma}{dx} = \frac{1}{2}\frac{2x\sigma_i^2 - 2(1-x)\sigma_M^2 + 2(1-x)\rho_{iM}\sigma_i\sigma_M - 2x\sigma_i\sigma_M}{\sqrt{x^2\sigma_i^2 + (1-x)^2\sigma_M^2 + 2x(1-x)\rho_{iM}\sigma_i\sigma_M}},$$

$$\left.\frac{d\sigma}{dx}\right|_{x=0} = \frac{1}{2}\frac{2\rho_{iM}\sigma_i\sigma_M - 2\sigma_M^2}{\sigma_M} = \rho_{iM}\sigma_i - \sigma_M.$$

Thus,

$$\left.\frac{d\mu}{d\sigma}\right|_{x=0} = \frac{\mu_i - \mu_M}{\rho_{iM}\sigma_i - \sigma_M}.$$

(b) Equating $\left.\frac{d\mu}{d\sigma}\right|_{x=0}$ and the slope of the CML, we have $(\mu_i - \mu_M)/(\rho_{iM}\sigma_i - \sigma_M) = (\mu_M - r)/\sigma_M$. This can be expressed as the SML equation $\mu_i = r + (\mu_M - r)\frac{\rho_{iM}\sigma_i}{\sigma_M}$.

Problem 5.1

(a) See Figure 4 for Excel implementation.

(b) The mean of the normal distribution is $(\mu - \sigma^2/2)T = 0.12 - \frac{0.20^2}{2} = 0.1$, and the standard deviation is $\sigma = 0.20 \times \sqrt{1} = 0.2$. The probability density function for Y is $Y \sim \phi = \frac{1}{\sqrt{2\pi(0.2)^2}} \exp[-\frac{(Y-0.1)^2}{2(0.2)^2}]$. Therefore, the probability for the price to drop below the initial value is

$$P\{S_T \leq 42\} = P\{Y \leq 0\} = \int_{-\infty}^{0} \phi \, dY = 0.3085.$$

(c) The *Mathematica* code is shown below.

```
s0=42;
sigma=.2;
r=.1;
T=1.0;
X=40;
NSteps=254;
dt=T/NSteps;
l=10000;
Table[p[0,i]=s0, {i,1,l}];
Table[For[n=0, n<NSteps, n++, p[n+1,i]=p[n,i]+r p[n,i] dt
+sigma p[n,i] Sqrt[dt] RandomReal[NormalDistribution[0,1]]], {i,1,l}];
Exp[-r T] Mean[Table[Max[{p[NSteps,i]-X,0}], {i,1,l}]]
```

(d) For the given values,

$$d_1 = \frac{\ln(42/40) + 0.10 + 0.20^2/2}{0.20\sqrt{1}} = 0.8439, \quad d_2 = 0.8439 - 0.20\sqrt{1} = 0.6439.$$

The cumulative probability distribution function evaluated at d_1 and d_2 are $\Phi(0.8439) = 0.8006$ and $\Phi(0.6439) = 0.7402$. The Black-Scholes formula gives $c = 42(0.8006) - 40e^{-0.10\times 1}(0.7402) = 6.83$.

(e) For the table and plot of the data, the estimated volatility is $\sigma = 0.24$ when $c = 1.88$.

σ (volatility)	0.01	0.05	0.10	0.15	0.20	0.25	0.30
c (call option)	1.49	1.49	1.53	1.62	1.76	1.93	2.10

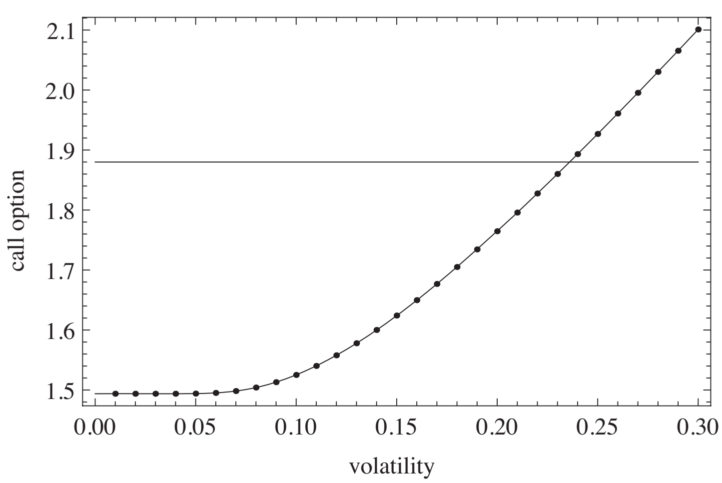
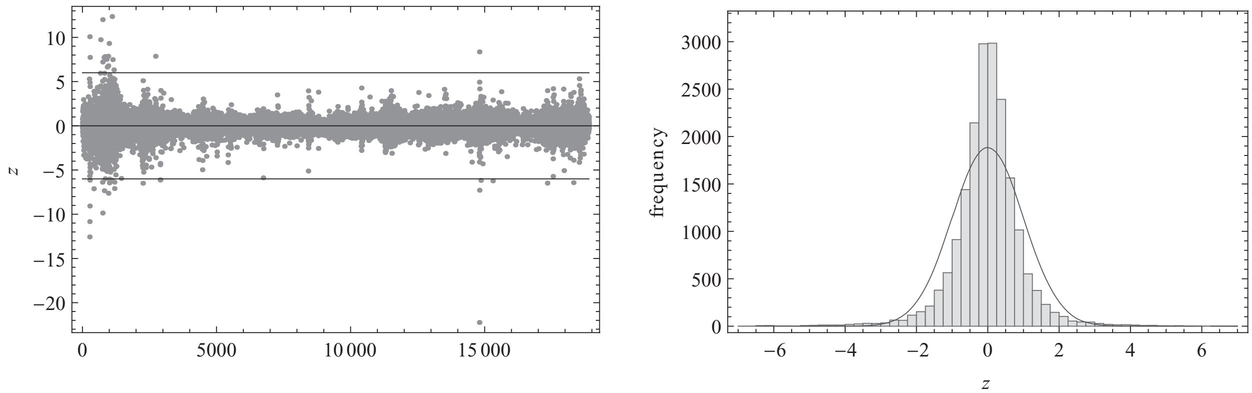


Figure 6. Call option price c versus volatility σ.

Problem 6.1 The daily changes (approximated by y's) of the Dow Jones Industrial Average from 1928 to 2003 in standard deviations is shown in Figure 7 (a). We can see quite a number of "unusual events," such as $|z| \geq 6$. (There are 40 of them if you count.) The bell curve in Figure 7 (b) does not fit the histogram well, as there are too many changes that are very small, and too many that are very large.



(a) (b)

Figure 7. Daily changes of the Dow Jones Industrial Average, 1928–2003.

References

[1] Fisher Black and Myron Scholes, "The Pricing of Option and Corporate Liabilities," *Journal of Political Economy* **81** (1973) 637–654.

[2] Marek Capiński and Tomasz Zastawniak, *Mathematics for Finance: An Introduction to Financial Engineering*, Springer, New York, 2005.

[3] Sergio M. Focardi and Frank J. Fabozzi, *Mathematics of Financial Modeling and Investment Management*, Wiley, Hoboken NJ, 2004.

[4] John C. Hull, *Options, Futures, and other Derivative Securities*, 7th edition, Pearson Prentice Hall, Upper Saddle River NJ, 2008.

[5] Paul Krugman, "How Did Economists Get It So Wrong?" *The New York Times Magazine* (September 6, 2009) 39; available at `http://www.nytimes.com/2009/09/06/magazine/06Economic-t.html`.

[6] Benoit B. Mandelbrot and Richard L. Hudson, *The (mis)Behavior of Market*, Basic Books, New York, 2004.

[7] Harry Markowitz, "Portfolio Selection," *Journal of Finance* **12** (1952) 77–91.

[8] William F. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance* **19** (1964) 425–442.

[9] *Standard & Poor's 500 Guide 2008 Edition*, McGraw-Hill, New York, 2008.

9

Maximizing Profit with Production Constraints

Jennifer Wilson[1]
Eugene Lang College the New School for Liberal Arts

1 Introduction

Mathematical applications to economics are rarely introduced in Calculus II or III. This is a missed opportunity since so many important concepts in second and third semester calculus courses can be discussed in terms of production, profit, utility, and social welfare functions, which are central to microeconomics. In this paper, we focus on mathematical techniques for optimizing profit functions with and without constraints. We illustrate these techniques with examples, and provide additional problems at the end of each section for student use.

Section 2 (Production Functions) introduces production functions and discusses several of their key properties. Section 3 (Unconstrained Optimization) looks at profit maximization problems and shows how a function's matrix of second-derivatives can be used to determine the function's concavity. In Section 4 (Constrained Optimization), we use the method of Lagrangian multipliers to solve optimization problems with one or more constraints, and explain how to modify the second-derivative test for a constrained maximization problem. We also discuss the significance of the Lagrange multiplier as a shadow price, and how it measures the amount of increase in the objective function as the constraint increases (for instance, the increase in maximal production resulting from an increase in the total budget). Section 5 (Optimization with Inequality Constraints) provides a brief introduction to optimizing production when the constraints are expressed as inequalities as, for example, when a firm can use some, but not all, of its resources. We state and illustrate the Karush-Kuhn-Tucker theorem that defines the necessary conditions for solving these problems, and state one version of the sufficient conditions. Section 6 contains answers to the sections' problems.

To provide greater insight, most of the analysis is carried out in generality leading to messy algebraic calculations. Specific values for the parameters can easily be substituted in all the examples, making the problems more tractable to students. Software (such as Mathematica or Maple) can also be used to explore these functions numerically and graphically, and to handle some of the more involved algebraic calculations.

[1] The author wishes to thank Duncan Foley, Department of Economics, New School for Social Research, for his insightful conversations and help in preparing this article.

2 Production Functions

Production functions typically take the form $Q = f(\mathbf{x}) = f(x_1, x_2, \ldots x_n)$, where Q describes the quantity of some product or output that a firm can produce using x_1 units of input 1, x_2 units of input 2, etc. One of the best-known production functions is

$$f(x_1, x_2, \ldots, x_n) = \left[\alpha_1 x_1^\rho + \alpha_2 x_2^\rho + \cdots + \alpha_n x_n^\rho\right]^{\frac{1}{\rho}},$$

where $0 \neq \rho \leq 1$ and $\{\alpha_i \in (0,1) : i = 1, 2, \ldots, n\}$ are parameters such that $\sum_{i=1}^n \alpha_i = 1$. This is known as the *constant elasticity of substitution* (CES) production function.[2] (See [2] for additional information about the elasticity of substitution.)

Example 1 Consider the CES production function $f(x_1, x_2) = [\alpha x_1^\rho + (1-\alpha) x_2^\rho]^{\frac{1}{\rho}}$, where x_1 is the amount of labor and x_2 is the amount of land available to a firm. The partial derivatives

$$\frac{\partial f}{\partial x_1} = \alpha x_1^{\rho-1}\left[\alpha x_1^\rho + (1-\alpha)x_2^\rho\right]^{\frac{1-\rho}{\rho}} \quad \text{and} \quad \frac{\partial f}{\partial x_2} = (1-\alpha) x_2^{\rho-1}\left[\alpha x_1^\rho + (1-\alpha)x_2^\rho\right]^{\frac{1-\rho}{\rho}}$$

are referred to respectively as the *marginal product with respect to labor* (MP_1) and the *marginal product with respect to land* (MP_2). The marginal products measure the rate of change of production with respect to the input quantities; evaluated at a specific (x_1, x_2), they can be interpreted as the approximate amount of additional production obtained when labor or land is increased by one unit, with the other held fixed.

The partial derivatives of the CES function are positive for $x_1, x_2 > 0$, as is typical of production functions (since increased input levels lead to increased production). And if $\rho \neq 1$, the non-mixed second-derivatives $\partial^2 f/\partial x_i^2 = -(1-\rho)\alpha(1-\alpha)\frac{(x_1 x_2)^\rho}{x_i^2}[\alpha x_1^\rho + (1-\alpha)x_2^\rho]^{\frac{1}{\rho}-2}$ are negative. This latter fact is known as the property of *diminishing returns* – a consequence of the fact that if one input increases while the other input is held constant, the firm's marginal production decreases. If $\rho = 1$, the production function reduces to a linear combination of the inputs, and its marginal products are constant.

The CES production function is a generalization of two other well-known production functions. The *Cobb-Douglas* function, $f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$ can be derived as the limit of the CES function as $\rho \to 0$, and the Leontief production function $f(x_1, x_2) = \min\{x_1, x_2\}$ corresponds to the limit as $\rho \to -\infty$. (See Problem 2.2.)[3]

An important characteristic of a production function is how it behaves when the input variables are simultaneously scaled up or down. In the simplest case, the output changes in proportion to the input so that, for example, when all input variables are doubled, production is doubled. However there are many instances when an increase in the available input yields more than a proportional increase in production (as when a firm's average costs decrease as its output increases), or no increase (as when a manufacturing plant is already acting at maximum capacity). These ideas are captured in the following definitions.

Definition 1 A function $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$ is said to exhibit *constant* returns to scale if $f(t\mathbf{x}) = tf(\mathbf{x})$ for all $t \geq 0$. It is said to exhibit *increasing* returns to scale if $f(t\mathbf{x}) > tf(\mathbf{x})$ for all $t > 1$, and *decreasing* returns to scale if $f(t\mathbf{x}) < tf(\mathbf{x})$ for all $t > 1$.

[2] The CES function was first introduced by Solow ([4]) who specified the general form of the function without explaining the economic significance of the $\sum_{i=1}^n \alpha_i = 1$ condition. Since that time, a number of variants of the CES function have been introduced, allowing for different degrees of "efficiency" in each input (land, labor, etc.). See [3] for a fuller discussion.

[3] While these functions are often used to model production, it is important to remember that real production may differ from these models in significant ways. Production levels may not depend continuously on the value of the inputs (input quantities may also be discrete). Even when a continuous model is appropriate, these functions may not apply over all portions of their domain. The partial derivatives of the Cobb-Douglas function, for instance, become unbounded as the input values approach zero, where in reality the dependence of production on the inputs may be finite. Production functions also frequently exhibit regions of both concavity and convexity. Nonetheless, such examples are useful because they shed light on the relationship between production, profit, and input variables.

Example 2

(a) The CES production function has constant returns to scale since for $t \geq 0$,

$$f(tx_1, tx_2, \ldots, tx_n) = \left[t^\rho\left(\alpha_1 x_1^\rho + \alpha_2 x_2^\rho + \cdots + \alpha_n x_n^\rho\right)\right]^{\frac{1}{\rho}} = tf(x_1, x_2, \ldots, x_n).$$

(b) If $f(x_1, x_2) = x_1^\alpha x_2^\beta$, $(0 < \alpha, \beta < 1)$, then $f(tx_1, tx_2) = t^{\alpha+\beta} x_1^\alpha x_2^\beta$. Therefore, f exhibits increasing returns to scale if $\alpha + \beta > 1$, decreasing returns to scale if $\alpha + \beta < 1$, and constant returns to scale if $\alpha + \beta = 1$.

Of particular interest for production functions is the amount by which one input can be substituted for another input. Suppose there are only two inputs and production is fixed at $f(x_1, x_2) = Q_0$. Differentiating implicitly with respect to x_1, we obtain

$$\frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial x_1} = 0 \quad \text{or} \quad \frac{\partial x_2}{\partial x_1} = -\frac{\frac{\partial f}{\partial x_1}}{\frac{\partial f}{\partial x_2}} = -\frac{MP_1}{MP_2}.$$

The quantity $\frac{MP_1}{MP_2}$ is referred to as the *marginal rate of technical substitution*. It measures the negative of the change in x_2 with respect to a change in x_1 when the production level is fixed. Thus, the marginal rate of technical substitution represents the approximate amount that x_2 can be reduced for each additional unit of x_1, while maintaining a constant production level. As the derivation indicates, the marginal rate of technical substitution is also the negative of the slope of the level curve $f(x_1, x_2) = Q_0$.

Example 3 In Example 1, $\partial x_2/\partial x_1 = (\alpha/(1-\alpha))(x_1/x_2)^{\rho-1}$ (See problem 2.3.) Suppose $\alpha = 2/3$, $\rho = 1/2$, and $(x, y) = (200, 800)$. Then $f(200, 800) = 3200/9$ and $\partial x_2/\partial x_1 = -4$. So for two hundred units labor and eight hundred units land, the production is $f(200, 800) = 3200/9$; if labor is increased by ten units, the production level will remain at 3200/9 if the land used is decreased by approximately 40 units. In particular, setting $3200/9 = f(210, x_2)$ yields $x_2 = 769.98$, a decrease of 39.02 units in land used.

Problems

Problem 2.1 A function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ is *homogeneous* of degree k if $f(t\mathbf{x}) = t^k f(\mathbf{x})$ for all $t \geq 0$.

(a) For each of the following functions, determine if it is homogeneous. If so, find its degree of homogeneity, and indicate whether the function exhibits increasing, decreasing, or constant returns to scale.
   (i) $f(x_1, x_2) = \ln g(x_1, x_2)$ where $g(x_1, x_2)$ is homogeneous of degree k
   (ii) $f(x_1, x_2) = \min\{x_1, x_2\}$.

(b) By differentiating both sides of the equation $f(t\mathbf{x}) = t^k f(\mathbf{x})$, show that the level curves of a homogenous function have constant slopes along rays coming out of the origin. What does this say about the marginal rate of technical substitution along these rays?

Problem 2.2 Let $f(x_1, x_2) = [\alpha x_1^\rho + (1-\alpha)x_2^\rho]^{\frac{1}{\rho}}$ for $0 \neq \rho \leq 1$.

(a) Show that the Cobb-Douglas function $g(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$ corresponds to $\lim_{\rho \to 0} f(x_1, x_2)$ by taking a natural logarithm and applying L'Hopital's rule.

(b) Show that the Leontief function $g(x_1, x_2) = \min\{x_1, x_2\}$ corresponds to $\lim_{\rho \to -\infty} f(x_1, x_2)$.

Problem 2.3

(a) Show that the marginal rate of technical substitution for the CES function is equal to $\frac{\alpha}{1-\alpha}(\frac{x_2}{x_1})^{1-\rho}$. If $\rho < 1$, what happens to the marginal rate of technical substitution as $\frac{x_2}{x_1}$ increases?

(b) What is the marginal rate of technical substitution if $\rho = 1$? (In this case, the inputs are said to be *perfect substitutes*.) How does the marginal rate of technical substitution relate to the level curves of $f(x_1, x_2)$?

(c) Find the marginal rate of technical substitution of the Leontief function. (In this case, the two inputs are said to have *no substitutability* between them.) Explain why this is consistent with the behavior of the Leontief function.

3 Unconstrained Optimization

Given a production function f, a natural question to ask is how a firm can maximize the profits associated with selling its product. If the per unit price P of the product and the per unit costs w_i of the inputs x_i are constant, then assuming that the firm sells all its production, the profit π is given by

$$\begin{aligned}\pi(x_1,\dots,x_n) &= \text{Revenue} - \text{Cost}\\ &= Pf(x_1,\dots,x_n) - [w_1x_1+\cdots+w_nx_n].\end{aligned}$$

(Note that economists use the symbol π to denote profit; π is not the number approximated by $\frac{22}{7}$.)

Example 4 Let $f(x_1,x_2) = [\alpha x_1^\rho + (1-\alpha)x_2^\rho]^{\frac{\mu}{\rho}}$, where $0 \neq \rho \leq 1$ and $\alpha \in (0,1)$, and μ is constant. Then

$$\pi(x_1,x_2) = P\big[\alpha x_1^\rho + (1-\alpha)x_2^\rho\big]^{\frac{\mu}{\rho}} - w_1x_1 - w_2x_2.$$

To determine the amount of labor and land that will maximize the firm's profit, we set the first-derivatives

$$\frac{\partial\pi}{\partial x_1} = \alpha\mu P x_1^{\rho-1}\big[\alpha x_1^\rho + (1-\alpha)x_2^\rho\big]^{\frac{\mu-\rho}{\rho}} - w_1 \quad\text{and}\quad \frac{\partial\pi}{\partial x_1} = (1-\alpha)\mu P x_2^{\rho-1}\big[\alpha x_1^\rho + (1-\alpha)x_2^\rho\big]^{\frac{\mu-\rho}{\rho}} - w_2 \tag{3.1}$$

equal to 0, and divide the equations to obtain $\frac{x_2}{x_1} = [\frac{w_1(1-\alpha)}{w_2\alpha}]^{\frac{1}{1-\rho}}$. Thus, in order for the input quantities, x_1^*, x_2^*, to correspond to a maximum, x_1^* and x_2^* must be in a fixed proportion. Substituting the expression for x_2 in terms of x_1 into either of the equations $\frac{\partial\pi}{\partial x_1} = 0$ or $\frac{\partial\pi}{\partial x_2} = 0$, we obtain

$$x_1^{*\mu-1} = \frac{w_1(w_2\alpha)^{\frac{\mu-\rho}{1-\rho}}}{\mu\alpha PC} \quad\text{and}\quad x_2^{*\mu-1} = \frac{w_2(w_1(1-\alpha))^{\frac{\mu-\rho}{1-\rho}}}{\mu(1-\alpha)PC},$$

where $C = [\alpha(w_2\alpha)^{\frac{\rho}{1-\rho}} + (1-\alpha)(w_1(1-\alpha))^{\frac{\rho}{1-\rho}}]^{\frac{\mu-\rho}{\rho}}$. This yields the maximum profit $\pi(x_1^*,x_2^*) = (1-\mu)PC^{\frac{\mu}{\mu-\rho}}(\frac{w_1w_2}{\mu PC})^{\frac{\mu}{\mu-1}}$.

Note that if $\mu = 1$, there is no solution to the equations $\frac{\partial\pi}{\partial x_1} = 0$ and $\frac{\partial\pi}{\partial x_2} = 0$ (no stationary point for π). This makes sense because if $\mu = 1$, then $f(x_1,x_2)$ in Example 4 exhibits constant returns to scale, and hence so must $\pi(x_1,x_2) = Pf(x_1,x_2) - [w_1x_1 + w_2x_2]$. Thus, $\pi(x_1,x_2) = x_1\pi(1,\frac{x_2}{x_1})$ is fixed along rays with a constant $\frac{x_2}{x_1}$ ratio.

If $\mu \neq 1$, then the second-derivative test can be used to determine if $\pi(x_1^*,x_2^*)$ is a maximum. To illustrate this, we introduce the following terminology.

Definition 2 The Hessian of a function $f(x_1,x_2,\dots,x_n)$ is the function's matrix of second-derivatives, denoted

$$D^2f = \begin{bmatrix} \frac{\partial f^2}{\partial x_1^2} & \frac{\partial f^2}{\partial x_1\partial x_2} \\ \frac{\partial f^2}{\partial x_2\partial x_1} & \frac{\partial f^2}{\partial x_2^2} \end{bmatrix}.$$

Example 5 In Example 4, let $P = 10$, $w_1 = 3$, $w_2 = 4$, $\alpha = \frac{3}{5}$, $\rho = \frac{1}{2}$ and $\mu = \frac{1}{4}$. Then

$$x_1^* = \left(\frac{25}{256}\right)^{1/3}, \quad x_2^* = \frac{1}{4}\left(\frac{25}{256}\right)^{1/3}, \quad \text{and } \pi(x_1^*,x_2^*) = 3\left(\frac{25}{4}\right)^{1/3}.$$

Moreover, the Hessian of π is given by

$$D^2\pi = \begin{bmatrix} \frac{\partial\pi^2}{\partial x_1^2} & \frac{\partial\pi^2}{\partial x_1\partial x_2} \\ \frac{\partial\pi^2}{\partial x_2\partial x_1} & \frac{\partial\pi^2}{\partial x_2^2} \end{bmatrix} = -\,Q^{-3}\begin{bmatrix} x_1^{-3/2}\big(\frac{27}{40}x_1^{1/2} + \frac{3}{10}x_2^{1/2}\big) & \frac{3}{20}x_1^{-3/2}x_2^{-1/2} \\ \frac{3}{20}x_1^{-1/2}x_2^{-1/2} & x_2^{-3/2}\big(\frac{3}{10}x_1^{1/2} + \frac{3}{10}x_2^{1/2}\big) \end{bmatrix},$$

where $Q = [\frac{3}{5}x_1^{1/2} + \frac{2}{5}x_2^{1/2}]^{1/2}$ is the production level at (x_1, x_2). A little algebra yields

$$\frac{\partial^2\pi}{\partial x_1^2}\frac{\partial^2\pi}{\partial x_2^2} - \left[\frac{\partial^2\pi}{\partial x_1 \partial x_2}\right]^2 = \frac{9}{16}x_1^{-3/2}x_2^{-3/2}Q^{-2} > 0 \quad \text{and} \quad \frac{\partial^2\pi}{\partial x_1^2} = -Q^{-3}x_1^{-3/2}\left[\frac{27}{40}x_1^{1/2} + \frac{3}{10}x_2^{1/2}\right] < 0$$

for all $x_1, x_2 > 0$. Thus, $\pi(x_1^*, x_2^*)$ is a relative maximum by the second-order test discussed in Section 3.1. Since there is only one stationary point, the relative maximum is also the absolute maximum – a crucial requirement for a firm seeking to maximize its profit.

3.1 Optimization, Second-Order Conditions and Convexity

As indicated in Example 5, it is often vital to determine a function's *absolute* maxima or minima. If there are a small number of relative extrema, the absolute extrema can be determined by evaluating the function at each point. In practice, however, this problem rarely arises because production and profit functions are usually assumed to have fixed *concavity*, and hence unique maxima or minima. (This is true of most of the standard production models, such as those discussed here. However, the assumption does not hold universally, and thus must be tested empirically or by some other means for more complex models.) To understand the relationship between a function's concavity, the existence of unique extrema, and the second-derivative test, we use the following.

Definition 3 An $n \times n$ real-valued matrix H is *positive (negative) definite* if $\mathbf{y}^T H\mathbf{y} > 0$ (if $\mathbf{y}^T H\mathbf{y} < 0$) for all non-zero vectors $\mathbf{y} \in \mathbb{R}^n$ (where $\mathbf{y}^T$ is the transpose of the vector $\mathbf{y}$). The matrix H is *positive (negative) semi-definite* if $\mathbf{y}^T H\mathbf{y} \geq 0$ (if $\mathbf{y}^T H\mathbf{y} \leq 0$) for all $\mathbf{y} \in \mathbb{R}^n$.

There is a well-known test of positive (negative) definiteness [2], although no simple corresponding test for positive (negative) semi-definiteness. It relies on the *principal minors* of matrix H. The k^{th} principal minor of H is the top left $k \times k$ submatrix of H.

Proposition 1 *An $n \times n$ real-valued matrix H is positive definite if and only if the determinants of all principal minors are positive. An $n \times n$ real-value matrix H is negative definite if and only if the determinant of the k^{th} principal minor has sign $(-1)^k$ for $k = 1, 2, \ldots, n$.*

This proposition, whose proof we omit, can be used to determine the convexity of a function as follows. Let $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_n)$ be a function of n variables with continuous second derivatives and consider $\mathbf{x}$ to be a $n \times 1$ column matrix. Then the second order Taylor expansion of f is given by

$$f(\mathbf{x}) = f(\mathbf{y}) + Df(\mathbf{y}) \cdot (\mathbf{x} - \mathbf{y}) + \tfrac{1}{2}(\mathbf{x} - \mathbf{y})^T D^2 f(\mathbf{y})(\mathbf{x} - \mathbf{y}) + O((\mathbf{x} - \mathbf{y})^3). \tag{3.2}$$

Suppose $\mathbf{y} = \mathbf{x}^*$, is a stationary point of $f(\mathbf{x})$ so that $Df(\mathbf{x}^*) = 0$. If $f(\mathbf{x}^*)$ is a local maximum, then $(\mathbf{x} - \mathbf{x}^*)^T D^2 f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \leq 0$; conversely, if $(\mathbf{x} - \mathbf{x}^*)^T D^2 f(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) < 0$, then $f(\mathbf{x}^*)$ must be a (local) maximum. The analogous statements are true for $f(\mathbf{x}^*)$ a minimum. Thus, a *necessary* condition for $f(\mathbf{x}^*)$ to be a maximum (minimum) is for $D^2 f(\mathbf{x}^*)$ to be negative (positive) semi-definite. A *sufficient* condition for $f(\mathbf{x}^*)$ to be a maximum (minimum) is for $D^2 f(\mathbf{x}^*)$ to be negative (positive) definite.

As Example 5 illustrates for $n = 2$, the sufficiency requirement for $f(\mathbf{x}^*)$ to be a maximum is $\frac{\partial^2\pi}{\partial x_1^2}(\mathbf{x}^*) < 0$ and $\frac{\partial^2\pi}{\partial x_1^2}\frac{\partial^2\pi}{\partial x_2^2} - [\frac{\partial^2\pi}{\partial x_1 x_2}]^2(\mathbf{x}^*) > 0$.

The Hessian matrix $D^2 f(\mathbf{x})$ of $f(\mathbf{x})$ can be used to describe the concavity of $f(\mathbf{x})$.

Definition 4 A function $f(\mathbf{x})$ is *concave* if $f(t\mathbf{x} + (1-t)\mathbf{y}) \geq tf(\mathbf{x}) + (1-t)f(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $0 \leq t \leq 1$. A function is *strictly concave* if $f(t\mathbf{x} + (1-t)\mathbf{y}) > tf(\mathbf{x}) + (1-t)f(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $0 \leq t \leq 1$. A function f is *convex* or *strictly convex* if the inequalities are reversed.

Geometrically, a function $f(\mathbf{x})$ is concave if its value at the weighted average of two points is greater than or equal to the weighted average of the function at the two points. If the function $f(\mathbf{x})$ is differentiable, then concavity is equivalent to the inequality $f(\mathbf{x}) \leq f(\mathbf{y}) + Df(\mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})$. Thus (Figure 1), a strictly concave function lies everywhere below its tangent plane, and a strictly convex function lies everywhere above its tangent plane. By comparing this inequality to Equation (3.2), we see the following.

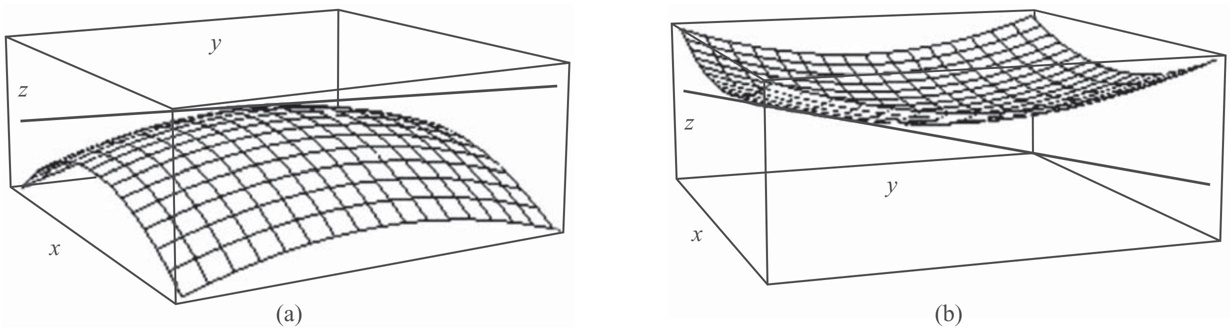


Figure 1. Graphs of a concave function (a) and a convex function (b).

Proposition 2 *Suppose $f(\mathbf{x})$ has continuous second-derivatives. Then $f(\mathbf{x})$ is concave (convex) if and only if its Hessian is everywhere negative (positive) semi-definite. And $f(\mathbf{x})$ is strictly concave (strictly convex) if its Hessian is everywhere negative (positive) definite.*

Thus, Proposition 1 can be used to determine whether a twice continuously differentiable function is strictly concave or convex. A strictly concave (convex) function has at most one stationary point and achieves an absolute maximum (minimum) at that point. In Example 5, the Hessian of $f(x_1, x_2) = (\frac{3}{5}x_1^{1/2} + \frac{2}{5}x_2^{1/2})^{1/2}$ is negative definite for all $x_1, x_2 > 0$. Therefore, $f(x_1, x_2)$ is strictly concave and has a unique maximum.

Example 6 Let $f(x_1, x_2) = [\alpha x_1^\rho + (1-\alpha)x_2^\rho]^{\frac{1}{\rho}}$ be the CES production function, and let

$$\pi(x_1, x_2) = P\big[\alpha x_1^\rho + (1-\alpha)x_2^\rho\big]^{\frac{1}{\rho}} - w_1 x_1 - w_2 x_2.$$

In Example 4, we showed that the profit function

$$\pi(x_1, x_2) = P\big[\alpha x_1^\rho + (1-\alpha)x_2^\rho\big]^{\frac{\mu}{\rho}} - w_1 x_1 - w_2 x_2$$

does not have a stationary point and thus cannot be strictly concave. Using $\pi's$ first-derivative in (3.1), we can express π's Hessian as

$$D^2\pi = C\begin{bmatrix} -x_1^2 & x_1x_2 \\ x_1x_2 & -x_2^2 \end{bmatrix},$$

where $C = \alpha(1-\alpha)(1-\rho)P[\alpha x_1^\rho + (1-\alpha)x_2^\rho]^{\frac{1-2\rho}{\rho}} x_1^{\rho-2}x_2^{\rho-2}$. Since the determinant of $D^2\pi$ is zero, $D^2\pi$ is not negative definite. However, it is negative semi-definite. To see this, let $\mathbf{y} = (y_1, y_2)^T$. Then $\mathbf{y}^T D^2\pi\mathbf{y} = -C(x_2y_1 - x_1y_2)^2 \leq 0$ for all y_1 and y_2. Thus, π is concave for $x_1, x_2 > 0$.

Problems

Problem 3.1 Let $f(x_1, x_2) = x_1^{1/3}x_2^{1/3}$ be a production function. Suppose that the unit price of the product is \$8 and that the input costs are \$2 per unit x_1, and \$1 per unit x_2. Determine how much of each input is required to maximize the profit.

Problem 3.2 Let $f(x_1, x_2) = x_1^\alpha x_2^\beta$ be the Cobb-Douglas function, where $\alpha, \beta > 0$ and $\alpha + \beta \leq 1$. Assume that the unit price of the product is P, and that the input costs are w_1 per unit x_1, and w_2 per unit x_2. For what conditions on α and β will the Hessian matrix be negative definite? Show that the profit function is concave in the remaining cases by checking for negative semi-definiteness directly. What can you conclude about the stationary points of the profit function?

Problem 3.3 Let $f(x_1, x_2) = [\alpha x_1^\rho + (1-\alpha)x_2^\rho]^{1/\rho}$ be the CES production function, where $0 < \alpha < 1$ and $0 \neq \rho \leq 1$. In Example 6, it was shown that f was concave on the region $x_1, x_2 > 0$. Use the definition of concavity to show that f is concave on the closed region $x_1, x_2 \geq 0$.

Problem 3.4 Suppose that the production function is given by $f(x_1, x_2) = Ax_1^2 + Bx_1x_2 + Cx_2^2$. Let P be the unit price of the product, and let w_i be the cost of input x_i. Determine what conditions on A, B, C guarantee that the profit function is concave, and hence that it achieves a maximum at any stationary point.

4 Constrained Optimization

The production of a firm depends on many factors: materials, budget, time, etc. Often these resources are limited, and some inputs may not be available on the market. To maximize the output of a production function $f(\mathbf{x})$ subject to a set of constraints $\{g_k(\mathbf{x}) \leq C_k : k = 1, \ldots, m\}$, we form the *Lagrangian* $\mathfrak{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \Sigma_k \lambda_k [g_k(\mathbf{x}) - C_k]$, where the unknown constants λ_k are referred to as *Lagrange multipliers.* The expression $g_k(\mathbf{x}) - C_k$ can be interpreted as the amount by which the firm over (or under) utilizes an internal resource, with λ_k acting as an internally derived price of that resource. Because of this, the Lagrange multipliers are often referred to as *shadow prices.*

In this section, we focus on the case when there is a single constraint expressed as an equality, so the problem is to maximize $f(\mathbf{x})$ on $\mathbb{R}^n_+$ such that $g(\mathbf{x}) = C$. The set $\Omega = \{\mathbf{x} \in \mathbb{R}^n_+$ such that $g(\mathbf{x}) = C\}$ is known as the *feasibility region.* If Ω is a closed bounded region and f is continuous, then f achieves an absolute maximum and absolute minimum on Ω. Recall that for an extrema of f to lie in the interior of Ω, it must occur at a point $\mathbf{x}$ where the level curves of f are parallel to the graph of $g(\mathbf{x}) = C$, so that $Df(\mathbf{x}^*) = \lambda Dg(\mathbf{x}^*)$ for some constant λ. And if $\mathfrak{L}(\mathbf{x}) = f(\mathbf{x}) - \lambda[g(\mathbf{x}) - C]$, then $\frac{\partial \mathfrak{L}}{\partial x_i}(\mathbf{x}^*) = 0$ is equivalent to $\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = \lambda \frac{\partial g}{\partial x_i}(\mathbf{x}^*)$ for each $i = 1, \ldots, n$. Moreover, if we regard $\mathfrak{L}(\mathbf{x}) = \mathfrak{L}(\mathbf{x}, \lambda)$ as a function in $n+1$ variables, then $\frac{\partial \mathfrak{L}}{\partial \lambda}(\mathbf{x}^*) = 0$ is equivalent to $g(\mathbf{x}^*) = C$. Thus, if $\mathbf{x}^*$ is a stationary point of $\mathfrak{L}$, it satisfies the usual first-order requirements for $\mathbf{x}^*$ to be an extrema of $f(\mathbf{x})$ given $g(\mathbf{x}) = C$.

Example 7 Consider the production function $f(x_1, x_2) = x_1^\alpha x_2^\beta$ $(\alpha, \beta > 0)$ with budget constraint $w_1x_1 + w_2x_2 = C$, where w_i is the per unit cost of x_i. The feasibility region Ω is the portion of the line $w_1x_1 + w_2x_2 = C$ lying in the first quadrant. Since $f(0, x_2) = f(x_1, 0) = 0$, and $f(x_1, x_2) \geq 0$ in Ω, the absolute minimum of $f(x_1, x_2)$ is 0. To find f's absolute maximum, let $\mathfrak{L} = x_1^\alpha x_2^\beta - \lambda(w_1x_1 + w_2x_2 - C)$. Setting the first-derivatives of $\mathfrak{L}$ equal to zero, we have

$$\alpha x_1^{\alpha-1} x_2^\beta - \lambda w_1 = 0, \quad \beta x_1^\alpha x_2^{\beta-1} - \lambda w_2 = 0, \quad w_1x_1 + w_2x_2 = C.$$

Solving, we obtain

$$x_1^* = \frac{\alpha C}{(\alpha+\beta)w_1}, \quad x_2^* = \frac{\beta C}{(\alpha+\beta)w_2}, \quad \lambda = \left(\frac{\alpha}{w_1}\right)^\alpha \left(\frac{\beta}{w_2}\right)^\beta \left(\frac{C}{(\alpha+\beta)}\right)^{\alpha+\beta-1}.$$

As this is the only extrema in the interior of Ω, it must correspond to the absolute maximum of f.

4.1 Second-Order Conditions when there are Equality Constraints

In Example 7, it was not necessary to appeal to the second-order conditions since there was only one candidate for absolute maximum. In general, however, the second-order conditions are necessary to determine if the solutions to the first-order conditions correspond to an absolute maxima or minima. In analogy with the unconstrained case, we have the following definition.

Definition 5 Let H and B be real-valued matrices with dimensions $n \times n$ and $1 \times n$ respectively. Then H is *positive (negative) definite subject to* $B\mathbf{y} = 0$ if $\mathbf{y}^T H\mathbf{y} > 0$ (if $\mathbf{y}^T H\mathbf{y} < 0$) for all non-zero vectors $\mathbf{y} \in \mathbb{R}^n$ such that $B\mathbf{y} = 0$. The matrix H is *positive (negative) semi-definite subject to* $B\mathbf{y} = 0$ if $\mathbf{y}^T H\mathbf{y} \geq 0$ (if $\mathbf{y}^T H\mathbf{y} \leq 0$) for all $\mathbf{y} \in \mathbb{R}^n$ such that $B\mathbf{y} = 0$.

The test for positive (negative) definiteness of H subject to the constraint $B\mathbf{y} = 0$ is given in terms of the principal minors of the *border preserving* matrix

$$BH = \begin{bmatrix} 0 & b_1 & b_2 & \cdots & b_n \\ b_1 & h_{11} & h_{12} & \cdots & h_{1n} \\ b_2 & h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_n & h_{n1} & h_{n2} & \cdots & h_{nn} \end{bmatrix},$$

where $H = [h_{ij}]$ and $B = [b_1, \ldots b_n]$. The k^{th} principal minor of BH, denoted BH_k, is the top left $(k+1) \times (k+1)$ submatrix of BH.

Proposition 3 *An $n \times n$ real-valued matrix H is positive definite subject to $B\mathbf{y} = 0$ if and only if the determinants of all principal minors of BH are negative. An $n \times n$ real-valued matrix H is negative definite if and only if the determinant of the $(k+1) \times (k+1)$ principal minor BH_k of BH has sign $(-1)^k$ for all $k = 1, \ldots, n$.*

The second-derivative test for a constrained optimization problem is based on the Taylor expansion of $\mathfrak{L}$. We consider only the case when g is linear. (For the non-linear case, see [2].) If $\mathbf{x}^*$ is a stationary point of $\mathfrak{L}$, then $\mathfrak{L}(\mathbf{x}^*) = 0$, and so

$$\begin{aligned} \mathfrak{L}(\mathbf{x}) &= \mathfrak{L}(\mathbf{x}^*) + \tfrac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T D^2\mathfrak{L}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + O((\mathbf{x} - \mathbf{x}^*)^3) \\ &= \tfrac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T D^2\mathfrak{L}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + O((\mathbf{x} - \mathbf{x}^*)^3). \end{aligned}$$

Thus, if $f(\mathbf{x}^*) = \mathfrak{L}(\mathbf{x}^*)$ is a constrained maximum, $(\mathbf{x} - \mathbf{x}^*)^T D^2\mathfrak{L}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \leq 0$ for all $\mathbf{x}$ satisfying $g(\mathbf{x}) = C$. Conversely, if $(\mathbf{x} - \mathbf{x}^*)^T D^2\mathfrak{L}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) < 0$ for all $\mathbf{x}$ satisfying $g(\mathbf{x}) = C$, then $f(\mathbf{x}^*)$ is a (local) constrained maximum. But if g is linear, then $g(\mathbf{x}) = g(\mathbf{x}^*) + Dg(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*)$. So $g(\mathbf{x}) = C$ is equivalent to $Dg(\mathbf{x}^*) \cdot (\mathbf{x} - \mathbf{x}^*) = 0$. Thus, using $B = Dg(\mathbf{x}^*)$ and $y = \mathbf{x} - \mathbf{x}^*$ in Definition 4, we see that a necessary condition for $f(\mathbf{x}^*)$ to be a constrained maximum (minimum) is for $D^2\mathfrak{L}(\mathbf{x}^*)$ to be negative (positive) semi-definite subject to $Dg(\mathbf{x}^*) = 0$. A sufficient condition for $\mathfrak{L}(\mathbf{x}^*)$ to be a constrained maximum (minimum) is for $D^2\mathfrak{L}(\mathbf{x}^*)$ to be negative (positive) definite subject to $Dg(\mathbf{x}^*) = 0$.

Example 8 Applying Proposition 3 to Example 7, we have

$$BH = \begin{bmatrix} 0 & g_1 & g_2 \\ g_1 & f_{11} & f_{12} \\ g_2 & f_{21} & f_{22} \end{bmatrix} = \begin{bmatrix} 0 & w_1 & w_2 \\ w_1 & \alpha(\alpha-1)x_1^{\alpha-2}x_2^{\beta} & \alpha\beta x_1^{\alpha-1}x_2^{\beta-1} \\ w_2 & \alpha\beta x_1^{\alpha-1}x_2^{\beta-1} & \beta(\beta-1)x_1^{\alpha}x_2^{\beta-2} \end{bmatrix},$$

where $g_i = \frac{\partial g}{\partial x_i}$ and $f_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$. Thus, $|BH_1| = -w_1^2 < 0$ (note that $|BH_1|$ is always nonpositive) and $|BH_2| = x_1^{\alpha-2}x_2^{\beta-2}[\beta(\beta-1)(w_1x_1)^2 - 2\alpha\beta(w_1x_1)(w_2x_2) + \alpha(\alpha-1)(w_2x_2)^2]$. At the stationary point, $|BH_2| = C\alpha\beta(x_1^*)^{\alpha-2}(x_2^*)^{\beta-2}/(\alpha+\beta) > 0$. Therefore, the bordered Hessian is negative definite, and $f(x_1^*, x_2^*)$ is a maximum, as expected.

We now consider a problem with two constraints.

Example 9 Suppose that a firm produces three products Q_1, Q_2, Q_3 and that the profit it makes on Q_1 and Q_2 depends on their level of production, so that the respective unit profits of Q_1, Q_2, Q_3 are $46 - \frac{1}{2}Q_1$, $43.25 - \frac{1}{4}Q_2$, and 20. Suppose also that each unit of Q_1 requires 2 units of labor and 3 units of land; each unit of Q_2 requires 1 unit of labor and 2 units of land; each unit of Q_3 requires 3 units of labor and 2 units of land; and the total amounts of labor and land available are 230 units and 280 units, respectively. Then the firm's profit is given by

$$\pi(Q_1, Q_2, Q_3) = \left(46 - \tfrac{1}{2}Q_1\right)Q_1 + \left(43.25 - \tfrac{1}{4}Q_2\right)Q_2 + 20Q_3$$

and the constraints are

$$2Q_1 + Q_2 + 3Q_3 = 230 \quad \text{and} \quad 3Q_1 + 2Q_2 + 2Q_3 = 280.$$

Note that the feasibility region

$$\Omega = \{(Q_1, Q_2, Q_3) : Q_i \geq 0 \text{ such that } 2Q_1 + Q_2 + 3Q_3 = 230 \text{ and } 3Q_1 + 2Q_2 + 2Q_3 = 280\}$$

is a line segment in the first octant with endpoints $(0, 95, 45)$ and $(76, 0, 26)$ corresponding to respective profits $\pi(0, 95, 45) = 2752.5$ and $\pi(76, 0, 26) = -1760$. To investigate the interior of Ω, we form the Lagrangian

$$\mathfrak{L} = \left(46 - \tfrac{1}{2}Q_1\right)Q_1 + \left(43.25 - \tfrac{1}{4}Q_2\right)Q_2 + 20Q_3 - \lambda_1[2Q_1 + Q_2 + 3Q_3 - 230] - \lambda_2[3Q_1 + 2Q_2 + 2Q_3 - 280].$$

Setting $\frac{\partial \mathfrak{L}}{\partial Q_i} = 0$ for each $i = 1, 2, 3$, yields the three equations

$$46 - Q_1 - 2\lambda_1 - 3\lambda_2 = 0, \quad 43.25 - \tfrac{1}{2}Q_2 - \lambda_1 - 2\lambda_2 = 0, \quad 20 - 3\lambda_1 - 2\lambda_2 = 0.$$

Together with the two constraints, we have five equations that must be solved for $Q_1, Q_2, Q_3, \lambda_1, \lambda_2$. After some routine linear algebra, we obtain

$$Q_1 = 26, \quad Q_2 = 62.5, \quad Q_3 = 38.5, \quad \lambda_1 = \lambda_2 = 4.$$

Since $\pi(26, 62.5, 38.5) = 3016.56$, this must be an absolute maximum. For details on how to apply the second-derivative test when there is more than one constraint, see Problem 4.2.

4.2 Interpretation of the Lagrange Multiplier

One benefit of studying optimization problems in the context of production functions is that the Lagrange multiplier has a useful interpretation.

If we regard the expressions for $\mathbf{x}^* = (x_1^*, x_2^*)$ and λ in Example 7 as functions of C, then $f = f(\mathbf{x}^*(C))$, $g = g(\mathbf{x}^*(C))$, and $\mathfrak{L} = \mathfrak{L}(\mathbf{x}^*(C), \lambda(C))$ also depend on C. Differentiating $\mathfrak{L}$ with respect to C, we obtain

$$\begin{aligned}
\frac{\partial \mathfrak{L}}{\partial C} &= \frac{\partial f}{\partial C} - \frac{\partial \lambda}{\partial C}(g - C) - \lambda\left(\frac{\partial g}{\partial C} - 1\right) \\
&= \frac{\partial f}{\partial x_1^*}\frac{\partial x_1^*}{\partial C} + \frac{\partial f}{\partial x_2^*}\frac{\partial x_2^*}{\partial C} + \frac{\partial \lambda}{\partial C}(g - C) - \lambda\left[\frac{\partial g}{\partial x_1^*}\frac{\partial x_1^*}{\partial C} + \frac{\partial g}{\partial x_2^*}\frac{\partial x_2^*}{\partial C} - 1\right] \\
&= \left[\frac{\partial f}{\partial x_1^*} - \lambda\frac{\partial g}{\partial x_1^*}\right]\frac{\partial x_1^*}{\partial C} + \left[\frac{\partial f}{\partial x_2^*} - \lambda\frac{\partial g}{\partial x_2^*}\right]\frac{\partial x_2^*}{\partial C} + \frac{\partial \lambda}{\partial C}(g - C) + \lambda.
\end{aligned}$$

At the stationary point, the first three terms drop out, leaving $\frac{\partial \mathfrak{L}}{\partial C}(\mathbf{x}^*) = \lambda$. In addition since $\mathbf{x}^*$ satisfies the constraint, $\mathfrak{L}(\mathbf{x}^*(C), \lambda(C)) = f(\mathbf{x}^*(C)) - \lambda(g(\mathbf{x}^*(C)) - C) = f(\mathbf{x}^*(C))$, and so $\frac{\partial f}{\partial C}(\mathbf{x}^*) = \frac{\partial \mathfrak{L}}{\partial C}(\mathbf{x}^*) = \lambda$. Thus, when the firm is producing optimally, λ represents the rate of change in production with respect to the total budget. In other words, at optimal production, λ represents the approximate amount by which production is increased for every additional unit in the total budget.

The relationship between λ and the function to be maximized also holds in more general cases. For instance, in Example 9, the λ_i represent the increased profit available to the firm with each additional unit of labor or land. In general, if the stationary point $\mathbf{x}^*$ corresponds to a constrained maxima of f, then the function $V(C) = f(\mathbf{x}^*(C))$ is called the *maximum value function of* $f(\mathbf{x})$ given the constraint $g(\mathbf{x}) = C$. A similar calculation to that above, yields $\frac{\partial V}{\partial C} = \frac{\partial \mathfrak{L}}{\partial C}$. (These results illustrate how a maximum value function increases as a parameter increases.)

4.3 Expansion Paths and Dual Problems

As described above, the maximum value function of a production function measures how the optimal level of production changes in response to a change in the total budget. The *expansion path* measures how the optimal inputs $\mathbf{x}^*$ change in response to a change in the total budget.

Recall that a stationary point $\mathbf{x}^*$ of $\mathfrak{L}$ can be interpreted geometrically as the point where the level curves of f are tangent to the level curves of g. This can be observed in Figure 2, where the level curves of the $f(x_1, x_2) = cx_1^{\alpha}x_2^{\beta}$ are superimposed with budget lines $g(x_1, x_2) = w_1x_1 + w_2x_2 = C$ for different values of C. From Example 7, we have $x_1^* = \alpha C/(\alpha + \beta)w_1$ and $x_2^* = \beta C/(\alpha + \beta)w_2$. As C increases, the point of tangency $\mathbf{x}^* = (x_1^*, x_2^*)$ traces out the diagonal line $x_2^* = \frac{\beta}{\alpha}\frac{w_1}{w_2}x_1^*$, representing how the optimal levels of labor and land change as the budget increases.

Expansion paths can be defined in terms of the costs associated with an increase in production.

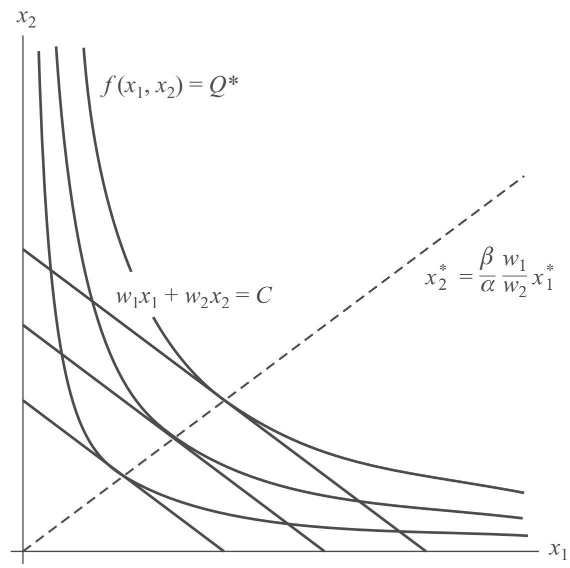


Figure 2. (From Example 7). Level curves of the production function $f(x_1, x_2) = x_1^{\alpha}x_2^{\beta}$ superimposed with the budget constraint $w_1x_1 + w_2x_2 = C$ for several values of C. For each C, the point (x_1^*, x_2^*) lies at the intersection of the budget constraint and a level curve of f. The dotted line represents the expansion path: the curve traced out by (x_1^*, x_2^*) as the value of C changes.

Example 10 Consider the production function $f(x_1, x_2) = \ln c + \alpha \ln x_1 + \beta \ln x_2$ (the natural logarithm of a Cobb-Douglas production function) with associated costs $C = w_1x_1 + w_2x_2$. To determine the input levels that minimize the cost for a fixed level of production Q^*, we let $\mathfrak{L} = w_1x_1 + w_2x_2 - \lambda(\ln c + \alpha \ln x_1 + \beta \ln x_2 - Q^*)$. Setting the first-order derivatives equal to 0, we have

$$w_1 - \lambda\alpha/x_1 = 0, \quad w_2 - \lambda\beta/x_2, \quad \ln c + \alpha \ln x_1 + \beta \ln x_2 = Q^*.$$

Solving, we obtain $x_1^* = \lambda\frac{\alpha}{w_1}$ and $x_2^* = \lambda\frac{\beta}{w_2}$, where $\lambda = [\frac{e^{Q^*}}{c}(\frac{w_1}{\alpha})^{\alpha}(\frac{w_2}{\beta})^{\beta}]^{1/(\alpha+\beta)}$.

To check that this corresponds to a maximum, we consider the bordered Hessian

$$BH = \begin{bmatrix} 0 & \frac{\alpha}{x_1} & \frac{\beta}{x_2} \\ \frac{\alpha}{x_1} & \frac{\lambda\alpha}{x_1^2} & 0 \\ \frac{\beta}{x_2} & 0 & \frac{\lambda\beta}{x_2^2} \end{bmatrix},$$

where $B = (\frac{\partial g}{\partial x_1}, \frac{\partial g}{\partial x_2}) = (\frac{\alpha}{x_1}, \frac{\beta}{x_2})$ for the constraint $g(x_1, x_2) = \ln c + \alpha \ln x_1 + \beta \ln x_2$. The principal minors have determinants $|BH_1| = -(\frac{\alpha}{x_1})^2 < 0$ and $|BH_2| = -\frac{\lambda\alpha\beta}{(x_1x_2)^2}(\alpha + \beta) < 0$ since $\lambda > 0$. Thus, the bordered Hessian is positive definite and $\mathfrak{L}(x_1^*, x_2^*)$ is a minimum. The point (x_1^*, x_2^*) represents the inputs required to minimize the cost for a fixed production level of Q^*. Viewed as functions of Q^*, the point (x_1^*, x_2^*) also traces out an expansion path indicating how the minimum cost input levels change as production increases.

The optimization problems in Examples 7 and 10 are known as *dual problems*. In the first problem, the goal is to maximize production while holding costs fixed; in the second problem, the goal is to minimize costs while holding production fixed. When applied to the same production and cost functions, the solutions to the two problems are related. Namely, if the production level in the second problem is fixed at the maximum production level obtained in the first problem, then the solutions (x_1^*, x_2^*) will be identical for both. (Similarly, if the cost is fixed in the first problem at the minimum level obtained in the solution of the second problem, the solutions will be identical.) The values of the Lagrange multipliers, however, will be different. Problem 4.3 explores this relationship.

Problems

Problem 4.1 Given a profit function $f(\mathbf{x})$ and a linear budget constraint $g(\mathbf{x}) = w_1x_1 + \cdots + w_nx_n = C$, show that at the output maximizing combination of inputs, the ratio of the marginal products of f with respect to inputs x_i and x_j is equal to the ratio of the corresponding input prices.

Problem 4.2 Check that $\pi(26, 62.5, 38.5) = 3016.56$ is a maximum in Example 8 by applying the second-derivative test. *Use the fact that the general form of the Bordered Hessian for a function $f(\mathbf{x}) = f(x_1, x_2, x_3)$ with constraints $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ is given by*

$$BH = \begin{bmatrix} 0 & 0 & (g_1)_1 & (g_1)_2 & (g_1)_3 \\ 0 & 0 & (g_2)_1 & (g_2)_2 & (g_2)_3 \\ (g_1)_1 & (g_2)_1 & \mathfrak{L}_{11} & \mathfrak{L}_{12} & \mathfrak{L}_{13} \\ (g_1)_2 & (g_2)_2 & \mathfrak{L}_{21} & \mathfrak{L}_{22} & \mathfrak{L}_{23} \\ (g_1)_3 & (g_2)_3 & \mathfrak{L}_{31} & \mathfrak{L}_{32} & \mathfrak{L}_{33} \end{bmatrix}.$$

Let BH_k denote the principal minor of dimension $(k+2) \times (k+2)$ for $k = 2, 3$. Then, a sufficient condition for $f(\mathbf{x}^*)$ to be a maximum given the constraints $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ is for the determinants $|BH_k|$ to have sign $(-1)^k$. A sufficient condition for $f(\mathbf{x}^*)$ to be a minimum given the constraints $g_1(\mathbf{x})$ and $g_2(\mathbf{x})$ is for the determinants $|BH_k|$ to all be positive.

Problem 4.3 Let $f(x_1, x_2) = 4(x_1^{1/2} + x_2^{1/2})^2$ be a production function with \$5 per unit cost of x_1 and \$4 per unit cost of x_2.

(a) Determine the maximum production if the fixed cost is \$170.

(b) Determine the minimum cost for the fixed maximum production level found in (a). Check that the stationary points (x_1^*, x_2^*) of each problem are equal.

(c) Compare the value and meaning of λ in (a) and (b).

Problem 4.4 Assume $f(x_1, x_2, x_3) = 2x_1^2 + 5x_1x_2 + x_2^2 + x_1x_3 + 2x_3^2$ is a production function, where the per unit costs of its inputs x_1, x_2, and x_3 are \$2, \$1, and \$2, respectively. Suppose that the total budget is \$220, and that the production process requires that the amount of input x_3 be equal to three times the amount of input x_1 plus the amount of input x_2. Determine the maximum production, assuming that the total budget is spent.

5 Optimization with Inequality Constraints

In Example 9, we determined production levels that maximize profit while using all available resources of labor and land. A more natural formulation of this problem is to allow the firm to use some, but not necessarily all, of its resources. In general, suppose the problem is to maximize the output of a production function $f(x_1, x_2, \ldots, x_n)$ subject to some constraints:

$$\max f(\mathbf{x}) \quad \text{such that} \quad \mathbf{x} \geq 0 \quad \text{and} \quad g_k(\mathbf{x}) \leq C_k, \quad (k = 1, \ldots, m). \tag{5.1}$$

As discussed in the last section, we form the Lagrangian $\mathfrak{L} = f - \sum_{k=1}^{m} \lambda_k(g_k - C_k)$ and interpret the λ_k as shadow prices. The values of $\mathbf{x}^*$ and $\{\lambda_k : k = 1, \ldots, m\}$ are *dual* variables in the sense that we seek values so that the λ_k *minimize* $\mathfrak{L}$ (given $\mathbf{x}^*$), and the values of $\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)$ *maximize* $\mathfrak{L}$ (given λ_k).

If $\mathbf{x}^*$ is an interior point so that each $x_i^* > 0$, then $\frac{\partial \mathfrak{L}}{\partial x_i}(\mathbf{x}^*) = 0$. If $x_i^* = 0$, it still must be the case that increasing x_i must not increase $\mathfrak{L}$, and so $\frac{\partial \mathfrak{L}}{\partial x_i}(\mathbf{x}^*) \leq 0$. Similarly, if $\lambda_k > 0$, then $\frac{\partial \mathfrak{L}}{\partial \lambda_k}(\mathbf{x}^*) = 0$. And if $\lambda_k = 0$, then $\frac{\partial \mathfrak{L}}{\partial \lambda_k}(\mathbf{x}^*) \geq 0$.

Taken together, this yields

$$\frac{\partial \mathfrak{L}}{\partial x_i}(\mathbf{x}^*) \leq 0, \quad x_i^* \geq 0, \quad \text{and } \frac{\partial \mathfrak{L}}{\partial x_i}(\mathbf{x}^*)x_i = 0 \tag{5.2}$$

$$\frac{\partial \mathfrak{L}}{\partial \lambda_k}(\mathbf{x}^*) \geq 0, \quad \lambda_k \geq 0, \quad \text{and } \frac{\partial \mathfrak{L}}{\partial \lambda_k}(\mathbf{x}^*)\lambda_k = 0. \tag{5.3}$$

These observations are encapsulated in the well-known Karush-Kuhn-Tucker Theorem

Theorem 4 *(W. Karush, H. W. Kuhn, and A. W. Tucker) If* $\mathbf{x}^*$ *satisfies (5.1), and if the gradients* $\{Dg_k(\mathbf{x}^*) : g_k(\mathbf{x}^*) = 0\}$ *are linearly independent, then there exist a set of multipliers* $\{\lambda_k : k = 1, \ldots, m\}$ *such that if* $\mathfrak{L} = f - \sum_{k=1}^{m} \lambda_k(g_k - C_k)$*, then* $\mathbf{x}^*$ *and* λ_k *satisfy* (5.2) *and* (5.3).

Note that unlike optimization with equality constraints, the λ_k must be nonnegative.

Example 11 (This is a variant of Example 9.) Suppose that a firm produces three products Q_1, Q_2 and Q_3, and that its profit per unit is $60 - \frac{1}{2}Q_1$ for Q_1, $60 - \frac{1}{4}Q_2$ for Q_2, and 20 for Q_3. Assume that each unit of Q_1 requires 2 units of labor and 3 units of land; each unit of Q_2 requires 1 unit of labor and 2 units of land; and each unit of Q_3 requires 3 units of labor and 2 units of land. Assume also, that the total available labor is 230 units and the available land is 280 units, and that the firm is not required to use all its resources. Then the Lagrangian is

$$\mathfrak{L} = \left(60 - \tfrac{1}{2}Q_1\right)Q_1 + \left(60 - \tfrac{1}{4}Q_2\right)Q_2 + 20Q_3 - \lambda_1(2Q_1 + Q_2 + 3Q_3 - 230) - \lambda_2(3Q_1 + 2Q_2 + 2Q_3 - 280).$$

Since $Dg_1 = (2, 1, 3)$ and $Dg_2 = (3, 2, 2)$ are linearly independent, we can apply the Karush-Kuhn-Tucker Theorem to get

$$\begin{aligned}
60 - Q_1 - 2\lambda_1 - 3\lambda_2 \leq 0; &\quad Q_1 \geq 0; \quad [60 - Q_1 - 2\lambda_1 - 3\lambda_2]Q_1 = 0\\
60 - \tfrac{1}{2}Q_2 - \lambda_1 - 2\lambda_2 \leq 0; &\quad Q_2 \geq 0; \quad \left[60 - \tfrac{1}{2}Q_2 - \lambda_1 - 2\lambda_2\right]Q_2 = 0\\
20 - 3\lambda_1 - 2\lambda_2 \leq 0; &\quad Q_3 \geq 0; \quad [20 - 3\lambda_1 - 2\lambda_2]Q_3 = 0\\
230 - 2Q_1 - Q_2 - 3Q_3 \geq 0; &\quad \lambda_1 \geq 0; \quad [230 - 2Q_1 - Q_2 - 3Q_3]\lambda_1 = 0\\
280 - 3Q_1 - 2Q_2 - 2Q_3 \geq 0; &\quad \lambda_2 \geq 0; \quad [280 - 3Q_1 - 2Q_2 - 2Q_3]\lambda_2 = 0.
\end{aligned}$$

To solve these equations, we first assume that Q_1, Q_2, and Q_3 are positive. Then, the first three lines above imply

$$60 - Q_1 - 2\lambda_1 - 3\lambda_2 = 0, \quad 60 - \tfrac{1}{2}Q_2 - \lambda_1 - 2\lambda_2 = 0, \quad \text{and} \quad 20 - 3\lambda_1 - 2\lambda_2 = 0. \tag{5.4}$$

Rewriting the last equation in (5.4) as $\lambda_2 = 10 - \frac{3}{2}\lambda_1$, and substituting into the first two equations in (5.4), we obtain $Q_1 = 30 + \frac{5}{2}\lambda_1$ and $Q_2 = 80 + 4\lambda_1$.

Moving to the constraints, we begin by assuming that λ_1 and λ_2 are positive. Then

$$2Q_1 + Q_2 + 3Q_3 = 230 \quad \text{and} \quad 3Q_1 + 2Q_2 + 2Q_3 = 280. \tag{5.5}$$

Eliminating Q_3 from the equations in (5.5), we obtain $5Q_1 + 4Q_2 = 380$. Substituting $Q_1 = 30 + \frac{5}{2}\lambda_1$ and $Q_2 = 80 + 4\lambda_1$ into this expression, we find that $\lambda_1 < 0$, which is a contraction. Therefore, we cannot have both λ_1 and λ_2 positive.

Next, we assume that $\lambda_2 = 0$, so that $\lambda_1 = 20/3$, $Q_1 = 140/3$ and $Q_2 = 320/3$. Note that in this case, only the first constraint in (5.5) must be satisfied, which can be solved to get $Q_3 = 10$. To determine whether this is a solution, we must check the sign of $\frac{\partial \mathfrak{L}}{\partial \lambda_2}$. Since $3Q_1 + 2Q_2 + 2Q_3 = 1120/3 > 280$, this is not a solution.

Finally, we assume that $\lambda_1 = 0$, so that $\lambda_2 = 10$, $Q_1 = 30$, and $Q_2 = 80$. Substituting these values into the second constraint in (5.5), we obtain $Q_3 = 15$. Again, we check the sign of $\frac{\partial \mathfrak{L}}{\partial \lambda_1}$. Since $2Q_1 + Q_2 + 3Q_3 = 185 < 230$, we have a solution $Q_1 = 30$, $Q_2 = 80$, $Q_3 = 15$, $\lambda_1 = 0$, $\lambda_2 = 10$.

In a similar manner, we can check for solutions for which one of the $Q_i = 0$. It is easily determined that there are none. Thus, the only solution of the Karush-Kuhn-Tucker equations is $Q_1^* = 30$, $Q_2^* = 80$, $Q_3^* = 15$, $\lambda_1 = 0$, $\lambda_2 = 10$. The shadow prices $\lambda_1 = 0$ and $\lambda_2 = 10$ indicate that the firm's profit would increase by \$10 for every additional unit of land, but would not increase with additional labor. This is because $(Q_1^*, Q_2^*, Q_3^*) = (30, 80, 15)$ requires all the current resources in land, but not in labor. Figure 3 shows a level curve of the profit function

$$\pi(Q_1, Q_2, Q_3) = \left(60 - \tfrac{1}{2}Q_1\right)Q_1 + \left(60 - \tfrac{1}{4}Q_2\right)Q_2 + 20Q_3$$

in the Q_2Q_3-plane when Q_1 is equal to 30. The shaded area corresponds to the feasible region which is bounded above by the two constraint lines. As expected, the level curve of f is tangent only to the second constraint since the first constraint is not binding.

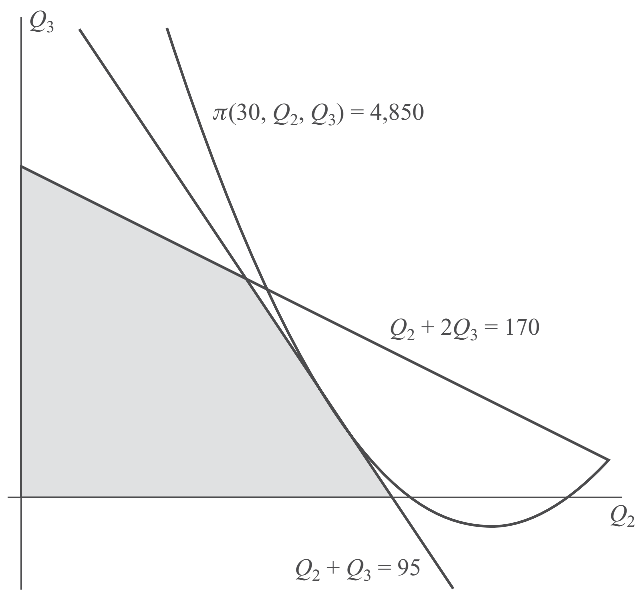


Figure 3. (From Example 11). Level curve of Profit function when $Q_1 = 30$.

The Karush-Kuhn-Tucker Theorem describes conditions *necessary* for $\mathbf{x}^*$ to solve the maximization problem, but these conditions are not sufficient. There are several formulations of sufficiency conditions. The most common is the following.

Theorem 5 *Suppose that $f(\mathbf{x})$ is concave and that $\{g_k(\mathbf{x}) \le C_k : k = 1, 2, \ldots, m\}$ are convex constraints such that $\{Dg_k(\mathbf{x}^*) : g_k(\mathbf{x}^*) = 0\}$ are linearly independent. Suppose also that there exist $\mathbf{x}^*$ and $\{\lambda_k : k = 1, 2, \ldots, m\}$ satisfying* (5.2) *and* (5.3), *where $\mathfrak{L} = f - \sum_{k=1}^{m} \lambda_k(g_k - C_k)$. Then $f(\mathbf{x}^*)$ is an absolute maximum.*

Example 12 In Example 11, the profit function

$$\pi(Q_1, Q_2, Q_3) = \left(60 - \tfrac{1}{2}Q_1\right)Q_1 + \left(60 - \tfrac{1}{4}Q_2\right)Q_2 + 20Q_3$$

is concave (see Problem 5.1), and the constraints are linear. Hence, $\pi(Q_1, Q_2, Q_3)$ is convex. So we can apply Theorem 5 to determine that $f(30, 80, 15)$ is an absolute maximum.

Problems

Problem 5.1 Show that the function $\pi(Q_1, Q_2, Q_3) = (60 - \frac{1}{2}Q_1)Q_1 + (60 - \frac{1}{4}Q_2)Q_2 + 20Q_3$ in Example 11 is concave.

Problem 5.2 For the production function $f(x_1, x_2, x_3) = (\frac{2}{5}x_1^{1/2} + \frac{3}{5}x_2^{1/2})^2$, the respective per unit costs for x_1 and x_2 are \$2 and \$1. Suppose that the total budget is \$220 and the firm must use at least four times as much input x_2 as input x_1. Determine the maximum production, given that the firm must stay within its budget.

6 Solutions

Problem 2.1

(a) (i) The function is not homogeneous since

$$f(t\mathbf{x}) = \ln g(tx_1, tx_2) = \ln[t^k g(x_1, x_2)] = \ln t^k + \ln g(x_1, x_2) = \ln t^k + f(x_1, x_2) \neq t_k f(x_1, x_2).$$

(ii) The function is homogeneous of degree 1; it shows constant returns to scale.

(b) Differentiating both sides of the equation $f(t\mathbf{x}) = t^k f(\mathbf{x})$ gives $t\frac{\partial f(t\mathbf{x})}{\partial x_i} = t^k \frac{\partial f(\mathbf{x})}{\partial x_i}$ for each x_i. Thus,

$$\frac{\frac{\partial f(t\mathbf{x})}{\partial x_i}}{\frac{\partial f(t\mathbf{x})}{\partial x_j}} = \frac{\frac{\partial f(\mathbf{x})}{\partial x_i}}{\frac{\partial f(\mathbf{x})}{\partial x_j}} = \frac{MP_i}{MP_j},$$

and so the slope of the level curve at the point $t\mathbf{x}$ is equal to the slope of the level curve at $\mathbf{x}$ for all $t > 0$. The marginal rate of technical substitution is also constant along these rays.

Problem 2.2

(a) Taking a natural logarithm of $f(x_1, x_2) = [\alpha x_1^\rho + (1-\alpha)x_2^\rho]^{1/\rho}$ gives $\ln f(x_1, x_2) = \frac{1}{\rho}[\ln(\alpha x_1^\rho + (1-\alpha)x_2^\rho)]$. Applying L'Hopital's rule, yields

$$\lim_{\rho\to 0} \ln f(x_1, x_2) = \lim_{\rho\to 0} \frac{\alpha x_1^\rho \ln x_1 + (1-\alpha)x_2^\rho \ln x_2}{\left(\alpha x_1^\rho + (1-\alpha)x_2^\rho\right) \times 1} = \alpha \ln x_1 + (1-\alpha)\ln x_2 = \ln x_1^\alpha x_2^{1-\alpha}.$$

Hence $\lim_{\rho\to 0} f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$.

(b) If $x_1 < x_2$, then $(\frac{x_2}{x_1})^\rho \to 0$ as $\rho \to -\infty$ and $f(x_1, x_2) = x_1[\alpha + (1-\alpha)(\frac{x_2}{x_1})^\rho]^{1/\rho} \to x_1$. If $x_1 > x_2$, then $f(x_1, x_2) = x_2[\alpha(\frac{x_1}{x_2})^\rho + (1-\alpha)]^{1/\rho} \to x_2$.

Problem 2.3

(a) If $f(x_1, x_2) = [\alpha x_1^\rho + (1-\alpha)x_2^\rho]^{1/\rho}$, then

$$\frac{MP_1}{MP_2} = \frac{\alpha x_1^{\rho-1}\left[\alpha x_1^\rho + (1-\alpha)x_2^\rho\right]^{1/\rho-1}}{(1-\alpha)x_2^{\rho-1}\left[\alpha x_1^\rho + (1-\alpha)x_2^\rho\right]^{1/\rho-1}} = \frac{\alpha}{1-\alpha}\left(\frac{x_2}{x_1}\right)^{1-\rho}.$$

For $\rho < 1$, the marginal rate of technical substitution increases with the ratio $\frac{x_2}{x_1}$.

(b) If $\rho = 1$ then the marginal rate of technical substitution reduces to the constant $\frac{\alpha}{1-\alpha}$. This is the negative of the slope of the level curves of f, which consist of lines of the form $\alpha x_1 + (1-\alpha)x_2 = Q_0$, for some constant Q_0.

(c) The marginal rate of technical substitution of the Leontief function $f(x_1, x_2) = \min\{x_1, x_2\}$ is

$$\lim_{\rho\to-\infty} \frac{MP_1}{MP_2} = \lim_{\rho\to-\infty} \frac{\alpha}{1-\alpha}\left(\frac{x_2}{x_1}\right)^{1-\rho} = \begin{cases} 0, & \text{if } x_1 > x_2 \\ \infty, & \text{if } x_1 < x_2 \end{cases}.$$

This is consistent with the behavior of f since if $x_1 > x_2$, then $f(x_1, x_2) = x_2$ and an increase in x_1 requires no change in x_2 to maintain constant production. Likewise, if $x_1 < x_2$, then $f(x_1, x_2) = x_1$ and an increase in x_1 can be substituted by no finite change in x_2 to maintain constant production.

Problem 3.1 The profit function is $\pi(x_1, x_2) = 8[x_1^{1/3}x_2^{1/3}] - 2x_1 - x_2$. The stationary point is $(x_1, x_2) = (1/54, 1/27)$. Since the Hessian is negative definite, $f(1/54, 1/27) = 2/[3(18^{1/9})]$ is an absolute maximum.

Problem 3.2 The profit function $\pi(x_1, x_2) = Px_1^\alpha x_2^\beta - w_1x_1 - w_2x_2$ has Hessian matrix

$$D^2\pi = -Px_1^{\alpha-2}x_2^{\beta-2}\begin{bmatrix} \alpha(1-\alpha)x_2^2 & -\alpha\beta x_1x_2 \\ -\alpha\beta x_1x_2 & \beta(1-\beta)x_1^2 \end{bmatrix}.$$

Hence, $\frac{\partial^2\pi}{\partial x_1^2}\frac{\partial^2\pi}{\partial x_2^2} - [\frac{\partial^2\pi}{\partial x_1\partial x_2}]^2 = P^2x_1^{2\alpha-2}x_2^{2\beta-2}\alpha\beta[-(\alpha+\beta)+1]$. If $\alpha + \beta < 1$, then $\frac{\partial^2\pi}{\partial x_1^2} \leq 0$ and $D^2\pi > 0$, so the Hessian is negative definite.

If $\alpha + \beta = 1$,

$$\begin{aligned}
\mathbf{y}^T D^2\pi\mathbf{y} &= -\ Px_1^{\alpha-2}x_2^{\beta-2}\big[\alpha(1-\alpha)x_2^2y_1^2 - 2\alpha\beta x_1x_2y_1y_2 + \beta(1-\beta)x_1^2y_2^2\big] \\
&= -\ \alpha\beta Px_1^{\alpha-2}x_2^{\beta-2}\big[x_2^2y_1^2 - 2x_1x_2y_1y_2 + x_1^2y_2^2\big] \\
&= -\ \alpha\beta Px_1^{\alpha-2}x_2^{\beta-2}[x_2y_1 - x_1y_2]^2 \le 0.
\end{aligned}$$

Hence, $D^2\pi$ is negative semi-definite. So for $\alpha + \beta \le 1$, the function π is concave and will have an absolute maximum at any stationary point.

Problem 3.3 If suffices to show that $f(t\mathbf{x} + (1-t)\mathbf{y}) \ge tf(\mathbf{x}) + (1-t)f(\mathbf{y})$ for $\mathbf{x} \ge 0$ and $\mathbf{y} = \mathbf{0}$. Since $f(\mathbf{0}) = 0$ and f is homogeneous of degree 1,

$$f(t\mathbf{x} + (1-t)\mathbf{y}) = f(t\mathbf{x}) = tf(\mathbf{x}) = tf(\mathbf{x}) + (1-t)f(\mathbf{0}).$$

Thus, f is concave on $x_1, x_2 \ge 0$.

Problem 3.4 Let $\pi(x_1, x_2) = P(Ax_1^2 + Bx_1x_2 + Cx_2^2) - w_1x_1 - w_2x_2$. The first derivatives are

$$\frac{\partial\pi}{\partial x_1} = 2PAx_1 + PBx_2 - w_1 \quad \text{and} \quad \frac{\partial\pi}{\partial x_2} = PBx_1 + 2PCx_2 - w_2.$$

The Hessian matrix is $P\left[\begin{smallmatrix} 2A & B \\ B & 2C \end{smallmatrix}\right]$. So the profit function is negative definite if $A < 0$ and $4AC - B^2 > 0$. A straightforward calculation shows that there is a stationary point at

$$x_1^* = (2w_1C - w_2B)/P(4AC - B^2), \quad x_2^* = (2w_2A - w_1B)/P(4AC - B^2).$$

Therefore, π achieves an absolute maximum at this stationary point. (Note, however, that $A < 0$ and $4AC - B^2 > 0$ imply that $C < 0$, so both x_1^* and x_2^* are negative.)

Problem 4.1 Given a profit function $f(\mathbf{x})$ and a linear budget constraint $g(\mathbf{x}) = w_1x_1 + \cdots + w_nx_n = C$, we have $\frac{\partial f}{\partial x_i}(\mathbf{x}^*) = \lambda\frac{\partial g}{\partial x_i} = \lambda w_i$ for each i. Thus,

$$\frac{\frac{\partial f}{\partial x_i}(\mathbf{x}^*)}{\frac{\partial f}{\partial x_j}(\mathbf{x}^*)} = \frac{w_i}{w_j}.$$

So, at the output maximizing combination of inputs, the ratio of the marginal products of f with respect to inputs x_i and x_j is equal to the ratio of the corresponding input prices.

Problem 4.2 The Bordered Hessian is

$$BH = \begin{bmatrix} 0 & 0 & 2 & 1 & 3 \\ 0 & 0 & 3 & 2 & 2 \\ 2 & 3 & -1 & 0 & 0 \\ 1 & 2 & 0 & -1/2 & 0 \\ 3 & 2 & 0 & 0 & 0 \end{bmatrix}.$$

Since $|BH_2| = 1$ and $|BH_3| = -28.5$, $\pi(26, 62.5, 38.5)$ is a maximum.

Problem 4.3

(a) The stationary point is $(272/18, 425/18)$, yielding a maximum production of $f(272/18, 425/18) = 306$.
(b) The stationary point is again $(272/18, 425/18)$, yielding a minimum cost of $5(272/18) + 4(425/18) = \$170$.
(c) The respective Lagrange multipliers for problems (i) and (ii) are $9/5$ and $5/9$ (reciprocals). The former represents the increase in the production as the budget increases; the latter represents the decrease in cost as production increases.

Problem 4.4 The Lagrangian is

$$\mathfrak{L} = 2x_1^2 + 5x_1x_2 + x_2^2 + x_1x_3 + 2x_3^2 - \lambda_1(2x_1 + x_2 + 2x_3 - 220) - \lambda_2(3x_1 + x_2 - x_3).$$

Setting $\mathfrak{L}$'s first-derivatives equal to 0, and solving the system of equations, we get $x_1 = x_2 = 20$, $x_3 = 80$, $\lambda_1 = 160$ and $\lambda_2 = -20$. The Bordered Hessian is

$$BH = \begin{bmatrix} 0 & 0 & 2 & 1 & 2 \\ 0 & 0 & 3 & 1 & -1 \\ 2 & 3 & 4 & 5 & 1 \\ 1 & 1 & 5 & 2 & 0 \\ 2 & -1 & 1 & 0 & 4 \end{bmatrix}.$$

Since $|BH_2| = 1$ and $|BH_3| = -66$, the maximum production is $f(20, 20, 80) = 11,200$.

Problem 5.1 The function $\pi(Q_1, Q_2, Q_3) = (60 - \frac{1}{2}Q_1)Q_1 + (60 - \frac{1}{4}Q_2)Q_2 + 20Q_3$ has Hessian

$$H = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1/2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since the determinant is 0, the Hessian is not negative definite. However, it is negative semi-definite since $\mathbf{y}^{\mathbf{t}} H\mathbf{y} = -y_1^2 - \frac{1}{2}y_2^2 \le 0$ for all $\mathbf{y}$. Thus, π is concave.

Problem 5.2 The function f is concave on $x_1, x_2 \ge 0$ (see Example 6 and Problem 3.3). Moreover, the constraints, $g_1(x_1, x_2) = 2x_1 + x_2$ and $g_2(x_1, x_2) = 4x_1 - x_2$ are linear, and hence convex, and the set $\{Dg_1, Dg_2\}$ is linearly independent. Thus, by the Karush-Kuhn-Tucker Sufficiency Theorem, it suffices to find $\mathbf{x}^*$, λ_1 and λ_2 satisfying the following inequalities, based on the Lagrangian $\mathfrak{L} = (\frac{2}{5}x_1^{1/2} + \frac{3}{5}x_2^{1/2})^2 - \lambda_1(2x_1 + x_2 - 220) - \lambda_2(4x_1 - x_2)$:

$$\begin{aligned}
\frac{2}{5}x_1^{-1/2}\left(\frac{2}{5}x_1^{1/2} + \frac{3}{5}x_2^{1/2}\right) - 2\lambda_1 - 4\lambda_2 \le 0, &\quad x_1 \ge 0, \quad \text{and } \left[\frac{2}{5}x_1^{-1/2}\left(\frac{2}{5}x_1^{1/2} + \frac{3}{5}x_2^{1/2}\right) - 2\lambda_1 - 4\lambda_2\right]x_1 = 0 \\
\frac{3}{5}x_2^{-1/2}\left(\frac{2}{5}x_1^{1/2} + \frac{3}{5}x_2^{1/2}\right) - \lambda_1 + \lambda_2 \le 0, &\quad x_2 \ge 0, \quad \text{and } \left[\frac{3}{5}x_2^{-1/2}\left(\frac{2}{5}x_1^{1/2} + \frac{3}{5}x_2^{1/2}\right) - \lambda_1 + \lambda_2\right]x_2 = 0 \\
-2x_1 - x_2 + 220 \ge 0, &\quad \lambda_1 \ge 0, \quad \text{and } [-2x_1 - x_2 + 220]\lambda_1 = 0 \\
-4x_1 + x_2 \ge 0, &\quad \lambda_2 \ge 0, \quad \text{and } [-4x_1 + x_2]\lambda_2 = 0.
\end{aligned}$$

The only solution is $x_1 = 20$, $x_2 = 180$, $\lambda_1 = 11/25$, and $\lambda_2 = 0$. Therefore, $f(20, 180) = 484/5$ is the maximum.

References

[1] W. Baumol and A. Blinden, *Microeconomics, Principles and Policy, 12th edition*, South-Western Cengage Learning, Mason, 2012.

[2] Alpha C. Chiang and Kevin Wainwright, *Fundamental Methods of Mathematical Economics, 4th edition*, McGraw-Hill Irwin, New York, 2005.

[3] Rainer Klump and Harold Preissler, "CES Production Functions and Economic Growth," Scandanavian Journal of Economics 102(1) (2000), 41–56.

[4] R. Solow, "A Contribution to the Theory of Economic Growth," Quarterly Journal of Economics 70 (1956) 65–94.

[5] H. Varian, *Microeconomic Analysis*, W.W. Norton & Company, New York, 1992.

About the Editor

Warren Page (M.S. and Ph.D. from Polytechnic University of New York) is a Professor Emeritus of New York City College of Technology, City University of New York (CUNY). He also taught at Yeshiva University and was a Visiting Professor at Ohio State University and Baruch College of CUNY.

Professor Page was a Second Vice-President of the Mathematical Association of America and, for eighteen years, Secretary of the American Association for the Advancement of Science's mathematics section. He has also been a member of, or advisor to, the National Research Council's United States Commission on Mathematical Instruction, CUNY's Advisory Commission on Teacher Education, EDUCOM's National Center for Research to Improve Postsecondary Teaching and Learning, the National Steering Committee of Discrete Mathematics and Computer Science (DIMACS), and the Consortium for Mathematics and its Applications (COMAP) and its High School and its Mathematics Applications Project (HIGHMAP). He was a consultant and materials developer for the New York State Initiative to its Science and Technology Foundation, and an MAA consultant to the Curriculum Foundation's Workshop on Mathematics and Economics. A former Advisory Board member and judge for the Mathematical Competition in Modeling, he has recently been serving as a judge for SIAM in Moody's Mega Mathematical Competitions.

Dr. Page is also widely known through his talks and invited addresses at local, national, and international conferences and from his more than fifty publications in pure and applied mathematics and mathematics education. He is the author of Topological Uniform Structures (Wiley-Interscience) and the editor of MAA's volumes Two-Year College Mathematics Readings and American Perspectives on the Fifth International Congress on Mathematics Education. He has served on the editorial board of DIMACS Mathematics and Computer Science Monographs, been chair of MAA's NOTES editorial board, editor of The College Mathematics Journal, and the founding editor of its columns *Classroom Capsule*s (which he edited for twelve years) and *Media Highlights* (which he has been editing since1984).

His honors and recognition include: NYSMATYC's Distinguished Achievement Award (1981), MAA's George Pólya Award (1982), CUNY's Salute to Scholars (1983), AMATYC's Mathematical Excellence Award (1990), and Metropolitan NY MAA's Distinguished Service Award (2008). He led an American delegation of mathematics professors to the People's Republic of China in 1987, and was a National Science Foundation ICME Fellow in 1984 and 1988. In 2007, he was elected a Fellow of the American Association for the Advancement of Science.

The editor can be contacted at wxpny@aol.com.

About the Authors

Julie Suzanna Glass is a Professor of Mathematics at California State University, East Bay. She received her undergraduate degree from Smith College and doctorate from the University of California, Santa Cruz. She is a mathematician and educator with substantial experience working on cross sector projects and has done extensive work in the area of professional development for university faculty as Director of the Office of Faculty Development and the Faculty Center for Excellence in Teaching. Dr. Glass has also served as the Project Manager for the Gateways Education Workforce Partnership, which seeks to improve educational and career outcomes for students "cradle to career" in Alameda and Contra Costa Counties.

Rae Jean Goodman is a Professor of Economics at the U.S. Naval Academy. She has taught Economic Statistics, Microeconomic Theory, Macroeconomic Theory, Monetary Policy and Theory, Macroeconomics Forecasting, and Research Seminar. She is currently the Director of Teaching and Learning and teaches half time. Goodman was an undergraduate mathematics major and has consistently integrated and emphasized mathematics in her economics courses. She has presented at the Lilly Conference on College Teaching and at the American Economic Association (AEA) conference, and served on the AEA Committee on Economic

Kevin J. Hastings discovered his love for applications of probability as an undergraduate student at Knox College in the 1970's. He earned his M.S. (1977) and Ph.D. (1981) degrees in Mathematics at Northwestern University specializing in Stochastic Processes. After moving to the University of Delaware in the early 1980's, the growing area of Mathematical Finance began to interest him, resulting in three papers on the optimization of investments. Since then he has written respected texts on Probability, Statistics, and Operations Research, focusing on interesting applications of the theory and the appropriate use of technology in pedagogy and problem solving in these areas.

Mary Lesser is professor of economics and Chair of the Charles M. Snipes School of Business at Lenoir-Rhyne University in Hickory, NC and professor emeritus, Iona College (New Rochelle, New York). From 1999 to 2009 she was the Executive Director of the Eastern Economic Association, one of the major regional associations of economists in the United States. Dr. Lesser received her Bachelors Degree, Masters Degree and Ph.D. from Fordham University. Her subject areas of interest are money and banking and financial markets, and her teaching interests are in the uses of various forms of technology in and outside of the classroom to facilitate learning.

Jane Schaerges Lopus is Professor of Economics and Director of the Center for Economic Education at California State University, East Bay. She holds an undergraduate degree from the University of Michigan, Ann Arbor, and a PhD from the University of California, Davis. Through the Council for Economic Education, she has conducted seminars and workshops in 15 former communist and developing countries to train teachers to teach market-based economics. She was named the Cal State East Bay Outstanding Professor in 2006 and has received national awards for both research and leadership in economic education from the National Association of Economic Educators.

Michael P. Murray, the Charles Franklin Phillips Professor of Economics at Bates College, is a microeconomist who has published three books and scores of articles. In addition to his writings on economics, he has published in operations research, demography, criminology, and statistics. Within economics, his work spans urban economics, public economics, labor economics, development economics, behavioral economics, and econometrics.

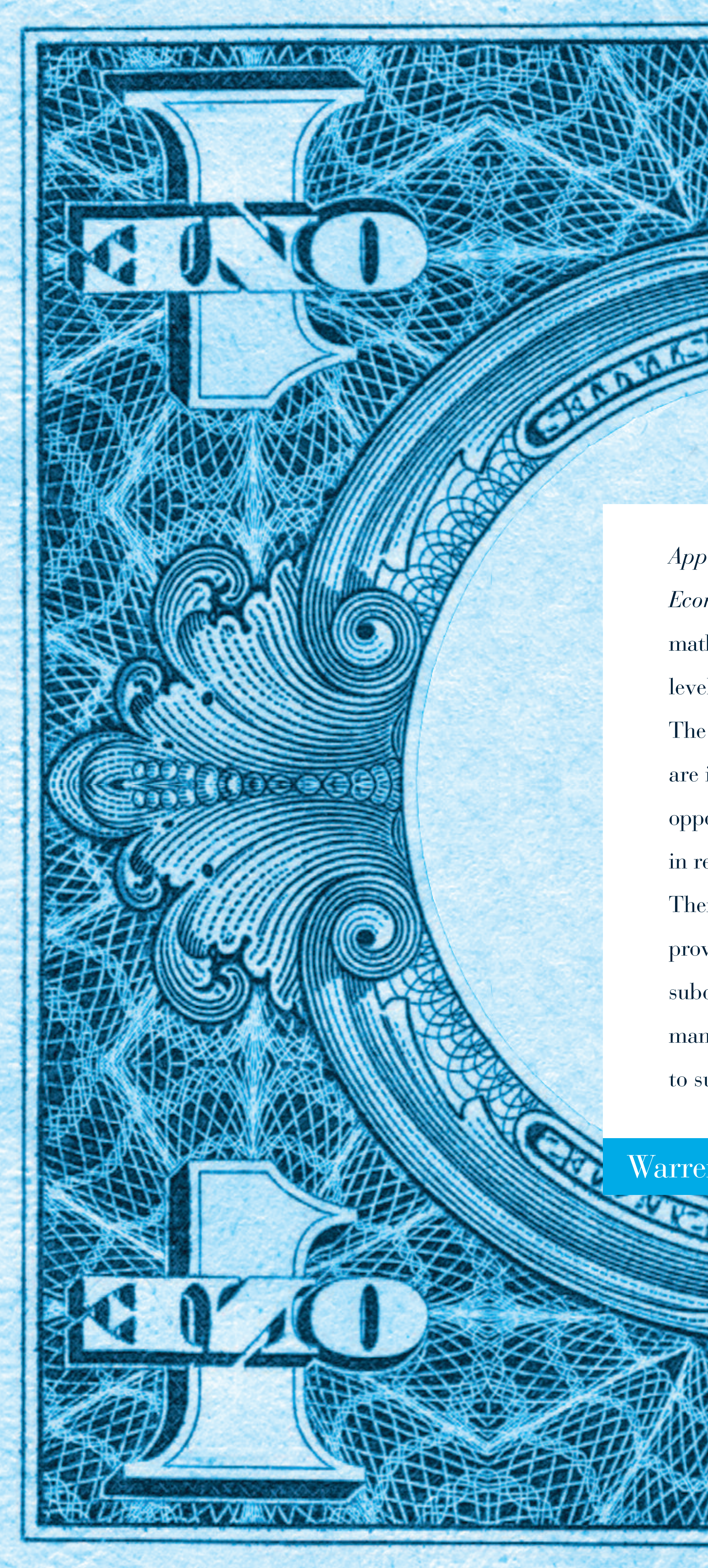
Lynn Paringer is a Professor of Economics at California State University, East Bay and is a Certified Financial Planner™. She holds a Ph.D. in economics from the University of Wisconsin, Madison. Lynn has served as principal investigator on numerous grants in the field of health services research and public health policy and has published articles on the impact of health insurance on access to care and the quality of care received. Lynn was a Fulbright Fellow at the Prague School of Economics and teaches in domestic and international MBA programs

Alan Parks is professor of mathematics and computer science at Lawrence University, where he has taught for 26 years. He is interested in many areas of applied mathematics, for instance: the theory of error correcting codes, linear optimization, and differential equations. Parks obtained his degrees at the University of Wisconsin—Madison, going on to a post-doctoral position at Michigan State University and a year at Ohio University before returning to Wisconsin. He greatly appreciates the goals and methods used at small liberal arts colleges. Parks is also a serious amateur musician interested in everything from baroque to modern worship music.

Michael K. Salemi is Professor and Chair of Economics at the UNC–Chapel Hill. Salemi is the author of more than sixty published articles in macroeconomics, domestic and international monetary theory, and economic education. Salemi was chair of the AEA Committee on Economic Education and Co-PI for the NSF-funded Teaching Innovations Program. He was awarded the Bowman and Gordon Gray Professorship for Excellence in Undergraduate Instruction by UNC-CH in 1987 and again in 2004, the Bower Medal by the National Council on Economic Education in 1998, and the Villard Research Award by the Association of Economic Educators in 2001.

Frank Wang received his Ph.D. from Columbia University and is currently a professor of mathematics at LaGuardia Community College of the City University of New York. His research interests include relativistic mechanics and dynamical systems. As an ardent advocate of computer algebra in education, Dr. Wang is the author of *Physics with Maple: The Computer Algebra Resource for Mathematical Methods in Physics*. He contributes frequently to Media Highlight column of the *College Mathematics Journal*. At LaGuardia he co-leads several initiatives that promote student success, include *Project Quantum Leap* funded by U.S. Department of Education and *Achieving the Dream* sponsored by Lumina Foundation. In 2009, he won the CUNY Chancellor's Award for Undergraduate Mathematics Instruction.

Jennifer Wilson is an Associate Professor of Mathematics in the Department of Natural Sciences and Mathematics at Eugene Lang College the New School for Liberal Arts. Her research interests lie in game theory and voting theory, and more generally, mathematics applied to the social sciences. At Eugene Lang College, she teaches mathematics courses for the Interdisciplinary Science and Economics majors, as well as elective courses in Game Theory, Ethnomathematics and Discrete Math. She became interested in integrating economics into the mathematics curriculum when she was asked to teach a course for Economics majors that combined parts of Calculus I and II, Linear Equations and Differential Equations. This led to many stimulating conversations with colleagues in the Economics department as well as an increased appreciation for the way optimization can be used as a unifying thread to teach many topics in undergraduate mathematics.



MAA Notes #82

Applications of Mathematics in Economics shows instructors what mathematics is used at the undergraduate level in various parts of economics. The volume's self-contained sections are intended to provide students with opportunities to apply their mathematics in relevant economics contexts. There is currently no known work that provides such an overview of economic subdisciplines or that brings together so many different mathematics applications to such varied economics topics.

Warren Page, EDITOR

ISBN: 978-0-88385-192-0