Learning-by-Doing and Infant Industries

by

C. Keith Head

B.A. (Economics), Swarthmore College (1986)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

October 1991

© C. Keith Head. All rights reserved.

The author hereby grants to MIT permission to reproduce and
to distribute copies of this thesis document in whole or in part.

Author .................................................................

Department of Economics
October 18, 1991

Certified by .............................................................

Paul Krugman
Professor of Economics
Thesis Supervisor

Accepted by .............................................................

Dick Eckaus
Chairman, Departmental Committee on Graduate Students
To Jeremy W. Head 1940–1990
Learning-by-Doing and Infant Industries
by
C. Keith Head

Submitted to the Department of Economics
on October 18, 1991, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Chapter 1 presents historical and statistical evidence establishing the existence and magnitude of learning-by-doing in steel rail production. The econometric evidence comes from price equations that attempt to control for static scale effects, exogenous technological progress, changing input costs, and physical capital accumulation. In an appendix to the chapter, I use simulations to show how optimal pricing by a monopolist would radically alter the paths of steel rail prices and their relationship to cumulative output.

Chapter 2 examines the joint roles of learning-by-doing, changing resource endowments, and tariff protection in the growth of the domestic steel industry. It specifies and estimates a demand model. Then, making use of the cost parameters from Chapter 1, it simulates what would have happened under free trade. Finally, the optimal path of duties are calculated and compared to the duties actually enacted. I conclude with a discussion of the relevance of this case study for current trade policy formulation.

While the first chapter examines the evidence of learning-by-doing, and the second chapter explores its implications for trade policy, Chapter 3 inquires into the nature of the learning process itself. After reviewing evidence that learning curves are linear when expressed in logs, it compares that form with alternatives using data from steel rail production. Next it provides a model of the learning process consistent with the stylized facts. Finally, it presents modifications to the model that address concerns that the classic formulation of learning-by-doing makes technological progress too mechanistic. The revised models show how costs of generating and implementing ideas for technical improvements lead to endogenous bounds on learning-by-doing.

Thesis Supervisor: Paul Krugman
Title: Professor of Economics
Acknowledgments

I would like to thank Paul Krugman, Peter Temin, Jean Tirole and Alwyn Young for advising me on this thesis. I would also like to express my gratitude to my family and friends for their support and encouragement throughout this process.
Contents

Introduction 10

1 Learning-by-Doing in Steel Rail Production 16
   1.1 The Literature on Learning-by-Doing 17
   1.2 The Bessemer Process 18
   1.3 Input Costs 20
   1.4 Static Returns to Scale 21
   1.5 Technological Progress 22
   1.6 Spillovers 24
   1.7 Market Structure and Pricing 25
   1.8 Estimation of Price Equations 28
   1.9 Physical Capital Accumulation 35

Appendix: Implications of Forward Pricing 40

2 The Tariff on Steel Rails 47
   2.1 Natural Resources or Government Intervention? 51
   2.2 Demand 54
   2.3 The Demand Model 55
   2.4 The Baseline Simulation 58
   2.5 Simulation of Free Trade 59
   2.6 Exogenous or Endogenous Growth? 59
   2.7 Welfare Consequences of the Steel Rail Tariff 62
      2.7.1 General Equilibrium Welfare 62
2.7.2 Consumer Surplus ........................................ 65
2.7.3 Profits .................................................... 66
2.7.4 Simulation Results ....................................... 67
2.8 Optimal Path of Protection .................................. 69
2.9 Choice of Instruments ...................................... 73
2.10 Computation of the Optimal Duty Path ................... 74
2.11 Conclusions ................................................ 77

3 Learning Curves: A Closer Look ............................. 80

3.1 Introduction ................................................. 80
3.2 The Power Law ............................................. 81
3.3 The Limits of Learning-by-Doing .......................... 82
   3.3.1 Quadratic Returns to Experience ..................... 84
   3.3.2 The Kinked Learning Curve .......................... 87
3.4 Models of the Learning Process ............................ 93
3.5 Learning-by-Doing as Costless Search .................... 97
3.6 Occam’s Razor .............................................. 103
3.7 Learning-by-Doing without the Free Lunch ................. 105
   3.7.1 Effort and Learning .................................. 105
   3.7.2 Embodied Progress .................................. 112
   3.7.3 Trial and Error ..................................... 113
   3.7.4 Screening of New Techniques ....................... 115
3.8 Conclusions and Extensions ............................... 117

Appendix: The Expected Minimum of n Draws from a Weibull Distribution 119

References ................................................... 120
List of Figures

1-1 Prices of American steel rails in current (—) and 1860 (---) dollars . 17
1-2 Entry Year and Size of Nine Large Firms in 1880 .................. 26
1-3 The Productivity of Physical Capital .............................. 38
1-4 Optimal Dynamic Pricing ........................................ 43
1-5 The Lerner Index under various discount rates. ................. 44
1-6 Pricing and the Learning Curve: A scatterplot of efficiency on cumulative output under myopia and foresight. ................. 45

2-1 The Steel Rail Tariff 1866-1913 .................................. 50
2-2 Import Share (—) and Price Difference (---) .................... 51
2-3 Real Pig Iron Prices in the U.S.(—) and the U.K. (— —) ...... 52
2-4 Steel rail(—) and pig iron(---) price differences in constant 1860 dollars 53
2-5 The Baseline Simulation ......................................... 60
2-6 The Simulation of Free Trade .................................... 61
2-7 Simulation of Constant Local Pig Iron Prices .................. 63
2-8 American steel rail prices with (—) and without (---) experience accumulation ........................................... 64
2-9 Welfare Consequences of Protection .............................. 68
2-10 Optimal and Historic Protection ................................. 76
2-11 Optimal Protection with a Revenue Motive ..................... 77
2-12 Optimal Protection and Comparative Advantage ............... 78

3-1 Learning-by-Doing in Steel Rail Production (U.S. = o, U.K. = *) . 91
3-2 Percentage Cost Reduction Attributable to Learning-by-Doing (Solid line indicates domestic industry. Dashed line indicates foreign.) . . . 93
3-3 Expectation of the minimum of $n$ draws from a Weibull distribution (plotted in normal and log-log scales) .............................. 102
3-4 The Weibull probability density function .............................. 103
3-5 The Distribution of Learning Elasticities. (Estimated from simulated data.) ................................................................. 110
3-6 Simulated Learning Relationships ...................................... 111
3-7 Simulated Aggregate Learning Curve ($\lambda = 0.19$) .............. 113
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Price Equation for American Steel Rails</td>
<td>30</td>
</tr>
<tr>
<td>1.2</td>
<td>Price Equations for British Steel Rails</td>
<td>32</td>
</tr>
<tr>
<td>1.3</td>
<td>Price Equations with Capital Accumulation</td>
<td>39</td>
</tr>
<tr>
<td>3.1</td>
<td>The Functional Form of U.S. Learning</td>
<td>86</td>
</tr>
<tr>
<td>3.2</td>
<td>The Functional Form of British Learning</td>
<td>88</td>
</tr>
<tr>
<td>3.3</td>
<td>Is Learning Bounded?</td>
<td>90</td>
</tr>
</tbody>
</table>
Introduction

The infant-industry argument for protection states that temporary government assistance—often in the form of import tariffs—may allow an industry to make the technological improvements necessary for it to become internationally competitive. A stronger form of the argument contends that, in the absence of intervention, countries will retain primitive trade patterns.

Skeptics counter with three major criticisms of the infant-industry argument. First, they assert that so-called infant industries do not grow up. Instead, they stagnate and retain distortionary protection as a result of political pressure, not economic potential. Second, the models that formalize the infant-industry argument depend on industry-wide learning-by-doing. Little evidence exists that this phenomena deserves the crucial role in trade determination that the new trade theorists have assigned to it. Finally, even if learning-by-doing does matter, there is no guarantee that long-run benefits from protection will outweigh the short-run allocative costs under reasonable discount rates.

This thesis provides both an illustration and assessment of the infant-industry argument and its criticisms in the context of import protection offered to American producers of steel rails in the 19th century. While I find that each of the criticisms have some validity, the results reported here provide some support for both the assumptions and implications of the infant-industry argument for protection.
The Theory

While the argument dates back to the 18th century, the notion of infant-industry protection did not gain intellectual respectability until John Stuart Mill's eloquent statement of the case.

The only case in which, on mere principles of political economy, protecting duties can be defensible, is when they are imposed temporarily (especially in a young and rising nation) in hopes of naturalizing a foreign industry, in itself perfectly suitable to the circumstances of the country. The superiority of one country over another in a branch of production often arises only from having begun it sooner. There may be no inherent advantage on one part, or disadvantage on the other, but only a present superiority of acquired skill and experience.

One defect of the older analyses, including Mill's, is the failure to specify what market failure prevents private firms from expanding production to accumulate experience and become internationally competitive under free trade. Recently, economists writing in the strategic trade policy and endogenous growth theory literatures have updated and formalized the old infant-industry argument for protection. The essence of the new argument for infant-industry protection is that the dependence of costs on cumulative output induces path dependence in the determination of trade patterns. If externalities (as emphasized by Krugman (1987), Lucas (1988), and Young (1989)) or myopic firm behavior (as emphasized in Dasgupta and Stiglitz (1988)) lead to initially low production levels, the domestic industry's subsequent performance may be permanently retarded even if that country has basic cost conditions appropriate for the industry. These considerations create some scope for temporary intervention that has lasting benefits to domestic producers and consumers.

In "The Narrow Moving Band..." Paul Krugman presents a model of trade in which learning-by-doing, not the underlying country attributes invoked by standard trade theory, determines the pattern of comparative advantage. Krugman demonstrates that, in such a model, "arbitrary patterns of specialization, once established,
tend to become more entrenched over time.” This feature allows temporary protection of selected sectors to permanently transform a country’s pattern of trade. Commercial policy can thus become a means of acquiring comparative advantage rather than simply a cause of deviations from a given “natural” comparative advantage (Grossman and Helpman 1988). More recently, Alwyn Young (1989) demonstrated in a model featuring “bounded learning-by-doing with spillover effects” that autarky—the most extreme form of import protection—may raise the rates of technical progress and growth experienced by an LDC. However, he found that autarky may reduce domestic welfare because it deprives the LDC’s consumers of the benefits of technical progress in developed countries. In a two-good model Lucas (1988) argues that even if the government manipulates producer incentives in such a way as to raise growth rates the effects on welfare remain uncertain.

Mathematical formalizations of the infant industry argument do not constitute empirical validation. The primary beneficiaries of protection have ample incentive to frame their pleas in terms of appeals to long-run public interest. It remains an empirical issue whether learning-by-doing has a major effect on costs and, if so, whether tariffs can exploit this to generate present value gains.

**Empirical Evidence**

Casual observation indicates that infant industry protection has been used successfully in several important industries. In steel and auto production, the Japanese, Korean, and Brazilian governments have all participated actively in the promotion of their domestic industries. Each country practiced a large amount of import protection and each of them now exports substantial amounts. European governments have subsidized Airbus to promote European competition against the dominant American aircraft producers. However, it is unclear to what extent subsidies and protection contributed to the success of these industries, and especially in the case of Airbus, to what extent these industries are internationally competitive in the absence of assistance.
In general, several factors make it difficult to assess the effects of recent episodes of infant industry protection. First, governments have employed multiple policy instruments. For instance, Korea’s infant industry promotion policies included rationed low-interest-rate credit, tax and intermediate-import tariff exemptions, import quotas, restrictive import licensing, export subsidies, and the fostering of the large conglomerate groups known as chaebols.\(^1\) Disentangling the separate contributions of each of these policies seems like an almost impossible task. Second, intervention frequently remains pervasive. Hence, one often cannot determine when—if ever—the domestic industry matured. In order to conduct a normative assessment of the use of protection, one should look at complete episodes in order to be as certain as possible that all of the costs and benefits have been included in welfare calculations.

Very little statistical evidence exists. Two fairly recent papers adopt a crosssection approach. Krueger and Tuncer (1982) compare the productivity growth rates of various Turkish industries. They found no systematic relationship between the effective rate of protection and the rate of growth of total factor productivity. Although the authors interpret this absence of a statistical relationship as the failure of a necessary condition for the desirability of infant industry protection, this conclusion does not seem warranted. The infant-industry argument does not assert that protected infant industries necessarily grow faster than mature industries in which a country already possesses a comparative advantage. A variety of forces contribute to productivity growth and these forces have varying degrees of importance at different stages in an industry’s development. The infant-industry argument simply asserts that protection can induce learning-based productivity growth that would not exist otherwise.

Bell, Westphal, and Ross-Larson (1984) compare productivity growth rates within industries across countries and find large variation. While Brazil and Korea’s steel industries attained productivity growth rates in excess of 10% during their infancies, several Indian and Turkish industries experienced low, and in some cases negative, rates. Although these results could be interpreted as indication that India and Turkey

\(^1\)Westphal 1990 p. 47-8
lacked dynamic comparative advantage in the industries they protected, the authors conclude that the firms in those countries failed to exert the "technological effort" required for productivity advances.

Evaluating the performance of an infant promotion policy involves a counterfactual analysis of what would have happened in the absence of intervention. To conduct that kind of analysis generally requires detailed knowledge of the industry. Hence, it seems desirable to accumulate a number of quantitative case studies. One such study, by Baldwin and Krugman (1988), concluded that although Japanese import protection aided the growth of their semiconductor industry, it did not increase welfare. However, the continuing evolution of technology and competition in the semiconductor industry make it difficult to draw final conclusions at this time. To avoid this problem, it would help to look at a completed episode of international competition. The growth of the American steel industry between the Civil War and World War I provides such an episode. This thesis focuses on the role of the steel rail tariff in fostering the domestic industry.
Overview

Chapter 1 presents historical and statistical evidence establishing the existence and magnitude of learning-by-doing in steel rail production. The econometric evidence comes from price equations that attempt to control for static scale effects, exogenous technological progress, changing input costs, and physical capital accumulation. In an appendix to the chapter, I use simulations to show how optimal pricing by a monopolist would radically alter the paths of steel rail prices and their relationship to cumulative output.

Chapter 2 examines the joint roles of learning-by-doing, changing resource endowments, and tariff protection in the growth of the domestic steel industry. It specifies and estimates a demand model. Then, making use of the cost parameters from Chapter 1, it simulates what would have happened under free trade. Finally, the optimal path of duties are calculated and compared to the duties actually enacted. I conclude with a discussion of the relevance of this case study for current trade policy formulation.

While the first chapter examines the evidence of learning-by-doing, and the second chapter explores its implications for trade policy, Chapter 3 inquires into the nature of the learning process itself. After reviewing evidence that learning curves are linear when expressed in logs, it compares that form with alternatives using data from steel rail production. Next it provides a model of the learning process consistent with the stylized facts. Finally, it presents modifications to the model that address concerns that the classic formulation of learning-by-doing makes technological progress too mechanistic. The revised models show how costs of generating and implementing ideas for technical improvements lead to endogenous bounds on learning-by-doing.
Chapter 1

Learning-by-Doing in Steel Rail Production

Temporary protection, coupled with learning-by-doing, may cause domestic costs to decline permanently relative to the costs of foreign rivals. This chapter sets the stage for assessing the effects of the steel rail tariff by first establishing the existence and magnitude of experience effects in steel rail production.

Figure 1-1 plots the data that inspired this thesis. Initially set at $166.00 per ton in 1867, steel rail prices fell at an annual rate of 5.6 % until they settled at $28.00 per ton after the formation of U.S. Steel Corp. in 1901.

Price declines of this magnitude are almost unheard of today except in certain segments of the electronics industry. However, the period following the Civil War was one of substantial deflation of the general price level. One might wonder if the constant dollar price of steel rails fell at all. The dashed line in Figure 1-1 shows a substantial decline in the real price of steel rails as well. Real steel rail prices fell at an annual rate of 4.1 % for a total decline of 75%.\(^1\)

\(^1\) The Consumer Price Index comes from David and Solar (1977).
1.1 The Literature on Learning-by-Doing

A broad definition of learning-by-doing would include all technological improvements—and corresponding cost reductions—associated with experience in an activity. Thus learning-by-doing differs from progress arising from research and development or the introduction of radically new processes. In practice, learning-by-doing tends to be defined as the mathematical relationship between unit costs or total factor productivity and cumulative output.

Economists have found statistical evidence indicating the presence of learning-by-doing in numerous industries including air-frames, Liberty ships, machine tools, and semi-conductors, and power plant construction. Marvin Lieberman provides a good survey as well as one of the more complete studies in his 1984 paper on the
learning curve in the chemical processing industries. Joskow and Rose (1985) find evidence that architect-engineer and utility experience reduce the costs of building new coal-burning generating units.

The empirical research estimating learning curves frequently suffers from a number of serious problems: omitted input prices (instead use general deflators), omitted output, the use of price instead of marginal cost regardless of market structure considerations. Also the articles generally ignore capital accumulation and changes in the quality of products or radical shifts in production processes. Furthermore, the issue of spillovers is often left unaddressed.

In light of these problems it is surprising that similar estimates of learning elasticities turn up for such a diverse group of products and methodological approaches. Pankaj Ghemawat's 1985 survey of estimates of learning curves for 97 different products found that in 79 of the 97 products examined, a doubling of cumulative would result in cost reductions ranging between 16 and 41 percent. About one third of the products had learning elasticities between −0.25 and −0.32. Ghemawat's survey found an average elasticity of −0.23.

1.2 The Bessemer Process

The Bessemer process, invented in the mid-1850s, made mass production of steel feasible for the first time. Rails were initially the primary product made with the new technology and rail production led to developments in steel rolling technologies.

The Bessemer process consists in placing molten pig iron inside a large pear-shaped vessel known as the "converter". Workers pump cold air through the converter which reacts with the pig iron, removing carbon and silicon. At the end of the process, some spiegeleisen is added to restore the carbon lost during combustion. Next, workers pour the molten metal into molds where it solidifies into steel ingots. Finally, heavy rollers flatten the ingots, first into blooms and then into rails.

An advantage of this study of learning-by-doing is that I do not have to distinguish between technological progress caused by improvements in a process versus progress
arising from the use of new processes. Although the first American open-hearth plant was built in 1875, the process was not used for rail production until 1897. As late as 1903, the Bessemer process accounted for over 98% of U.S. rail production.

The chief inputs are pig iron, labor, plant and equipment. Interestingly, as long as the pig iron comes directly from the blast furnace, the Bessemer process requires no additional fuel.

Total variable costs for rails produced in country $i$ are

$$C_{it} = \Phi_{it} \exp(-\gamma_{it})(1 + E_{it})^{-\lambda_{i}}q_{it}^{\theta_i}$$

where $\theta$, $\lambda$, and $\gamma$ represent the effects of contemporaneous production, cumulative production, and exogenous technological progress on costs. $\Phi_{it}$ is a function of the input prices in country $i$.

$$E_{us}(t) = \sum_{\tau=1867}^{t-1} q_{us}(\tau)$$

$$E_{uk}(t) = \sum_{\tau=1859}^{t-1} q_{uk}(\tau)$$

Differences in data availability lead to an asymmetry in the definitions of output ($q_{us}$ and $q_{uk}$) in the American and British cost equations.

British data on rail production is spotty and only begins in 1877. Hence, I assume that British costs depend on cumulative total steel output. Unfortunately, even steel output data are not available for Great Britain prior to 1868. As a result, I had to generate values for British steel output between 1859, the first year of English production, and 1868 by means of a constant growth rate interpolation. Cumulating and taking the log (the regressions are linear in logs) should smoothen out the inevitable errors to a large degree.

American data exists for both steel rail production and Bessemer ingot production. By using steel rail output, I omit the possibility that production experience in other steel goods led to improvements in the Bessemer process that raised the productivity
of rail production. As a remedy, one could assume that learning is based on cumulative production of Bessemer steel ingots. However, this would omit the contributions to rolling technologies associated with the use of open-hearth ingots. Fortunately, during this sample period most steel was used for rails and most rails were made from Bessemer steel. In practice, the estimates of the learning curve change by negligible amounts depending on which measure of cumulative output I use. Since I use steel rail prices in the estimation section, cumulative steel rail output seems like the natural measure of experience.

To avoid taking the log of zero in the first year of U.S. production, I add one to cumulative output. I add one in the British case as well in order to preserve symmetry. Sections 1.3, 1.4, and 1.5 examine the role of input costs, static returns, and technological change in the determination of costs.

1.3 Input Costs

The principal input cost for steel rail production are pig iron prices. Metal costs—which include relatively small expenses for manganese—accounted for 78% of the cost of steel rail production in 1879 and 82% in 1910.²

The U.S. imported pig iron. Hence, it is possible that the U.S. steel producers used both domestic and imported pig iron. Let z be the portion from American blast furnaces and 1 − z be the share from England. The domestic pig iron has price pipₜₚ, and the imported pig iron costs pipₜₚ + D₂ + taf₂, where pipₜₚ is the price of pig iron in Great Britain, D₂ is the U.S. duty on imported pig iron, and taf₂ is an AISA estimate of the per ton transatlantic shipping cost. The per unit domestic pig iron cost is pipₜₚ ≡ z(pipₜₚ) + (1 − z)(pipₜₚ + D₂ + taf₂). I have found no direct evidence on how much imported pig iron steel producers used. In later years vertical integration became prevalent and U.S. imports became very small relative to domestic production of pig iron. To obtain z, I assume that rail industry uses imported and domestic pig iron in the same proportions as the aggregate use of pig iron. Hence, z is total U.S.

²Computed from Table 1 in Allen (1981)
pig iron production divided by the sum of U.S. production and imports of pig iron. Note that \( z \) exceeds 90\% every year except 1880; pig iron imports were not overly important, so \( \pi_{c_{us}} \) is generally quite close to \( \pi_{p_{us}} \).

Wage data for steel workers is spotty. In census years, labor expenses divided by labor input (measured in man-years) provide a rough measure of the price of labor. Allen (1981) reports wages of $432, $577, $609, and $697 per man-year in 1879, 1889, 1902, and 1910. The total wage increase of 61\% does not differ much from the 58\% growth in the David and Solar (1977) low-skilled-labor wage index during the same period. Moreover, the steel wage remains at a roughly constant proportion to the wage index in all four census years. These considerations suggest that the wage index may provide a decent proxy for steel wages.\(^3\)

1.4 Static Returns to Scale

A large part of the empirical work on learning-by-doing assumes constant static returns to scale. However, under non-constant returns, this omission will lead to a biased estimate of the learning elasticity. Since current and cumulative output are usually positively correlated, the learning elasticity will be biased upward (in absolute value) under increasing returns and downward under decreasing returns.

Fixed plant and equipment costs and the well known volume-surface area relationship seem to indicate increasing returns. However, the fact that over a dozen firms coexisted for long periods of time casts doubt on the idea of large economies of scale at the industry level. From 1867 to 1881, the average size of converters remained around 6 tons per heat. By 1901, plants were using larger converters but around half had capacities below ten tons and only 9 out of 81 had capacities of 15 tons or more. This suggests that the gains from larger converters were not very large. The speed of each heat seemed to matter much more than the size of the heat.

There are reasons to expect rising marginal costs at both the plant and industry

---

\(^3\)In comparison with average annual earnings for non-farm employees (data from Historical Statistics of the United States), steel workers earned more but their wage premia remained fairly constant at 1.16 in 1879, 1.23 in 1889, 1.17 in 1902, and 1.10 in 1910.
level. For given equipment and organization of production, attempts to raise annual output at a plant will tend to cause more work-delaying accidents. This will tend to raise marginal costs. Moreover, the physical capacity of available Bessemer converters places limits on short-term production. If firms differ in their efficiency and have short-run capacity constraints, then the industry as a whole may have increasing marginal costs even if individual firms have constant returns to scale. Efficiency differentials existed because of differing levels of managerial and engineering talent across firms and because improved production methods were often embodied in new plant designs.

The price regressions include the log of output on the right-hand side. The implied constant elasticity of costs with respect to output facilitates estimation. A positive coefficient indicates a rising marginal cost of production. Two possible interpretations exist for negative coefficients. One is that there are external static increasing returns that no individual firm can affect. The other possibility is that increasing returns are internal to the firms but the firms in the industry set price equal to average cost as in the contestable markets model.

1.5 Technological Progress

The central feature of this model is the inclusion of dynamic returns to scale as a source of cost reductions. The regression model allows for exogenous technological progress but the nature of the technological changes observed in the steel industry make endogenously generated learning-by-doing a more likely source of improvements.

The learning process can be seen as consisting in a series of small discoveries about how to organize production in a more efficient manner (See Chapter 3 for an explicit model along these lines). These lead to the increases in the speed of production that the trade publication Iron Age frequently lauded. Learning-by-doing is likely to be a major source of productivity improvements in cases where production knowledge cannot be advanced through laboratory research. This is the case where underlying mechanisms are poorly understood and rules-of-thumb are selected through trial and
error experimentatation. "[T]he infancy of metallurgical, chemical, and thermodynamic science did not allow the experimental, and therefore inexpensive, testing of all relevant factors under laboratory conditions." "One could make a model train of rolls of new design, but it took the passing of real metal, under actual heat and pressure, to determine whether flanges would tear or rails would split."\(^4\)

Accidents plagued the Bessemer process. If the lining along the bottom of the converters wore out, hot metal would leak out and come into contact with the water kept around the converters for cooling purposes. The pressure caused by the evaporation of the water under the metal caused serious explosions. A second source of accidents was spillage during the pouring of the hot steel out of the converter. Production experience led to improvements in equipment and organization that reduced accidents and the accompanying production delays.

Innovations such as removable converter bottoms and improved refractory linings lowered the time spent repairing converters. Simultaneously, improved methods for using the hydraulical cranes reduced accidents and the time per heat. Thus speed seems much more important than pure scale. Improved organization and machinery probably made it possible to increase the size of converters. Another source of increased speed seems to be the eight-hour day established at the Edgar Thomson works around 1881.\(^5\)

Strassman provides a few concrete examples of technological improvements that look like learning-by-doing. "Fuel was also saved by avoiding reheating between smelting and rolling."\(^6\) The Jones mixer was used to combine metal from different furnaces prior to pouring the molten pig iron into the Bessemer converter. "Other innovations included the use of high-pressure steam to keep ingots under pressure while cooling, and later the use of soaking pits for bringing ingots to a uniform temperature."\(^7\) Livesay provides another example:

\(^4\)Quotes come from Strassman p. 51.
\(^5\)Hogan, pp. 218-221
\(^6\)Strassmann p. 42
\(^7\)Strassmann p. 44
Shinn and Jones discovered, while poring over the cost sheets 'to find some detail capable of judicious pruning,' that broken ingot molds (into which the Bessemer converters poured molten steel) added 60 cents to the cost of each ton of steel. Experiments soon developed molds that cut the cost to 15 cents, a saving that quickly amounted to $40,000 a year.

Production experience lowers costs at a diminishing rate, as the opportunities for incremental improvements in a given process are exhausted. Following the empirical literature on learning-by-doing, I assume the log-linear functional form which exhibits a constant elasticity of learning. Chapter 3 explores the justifications and consequences for this assumption.

1.6 Spillovers

Following the theoretical literature on infant industry protection, I assume (1) perfect spillovers of learning-based knowledge across domestic firms and (2) zero spillovers between foreign and domestic firms. These assumptions seem appropriate if we think of accumulated knowledge as residing within entrepreneurs, engineers, and workers who either communicate with each other or move from firm to firm but never move overseas.

One cannot test (1) due to a lack of firm-level data. However, several historical factors make it seem like a reasonable approximation. Managers at newer firms were often trained at one of the older ones. When Carnegie began operations he obtained a large number of the workers, including his chief engineer, Captain Jones, during a strike at Cambria works. Iron Age frequently described the latest technical developments in steel production. Presumably, this induced firms to attempt to copy the new ideas. Strassmann argues that firms rarely kept their innovations private.

Many important blowing-engine and rolling-mill innovations were not patented at all. Such ironmasters as Abram Hewitt and John Fritz often allowed others to inspect their plants and even lent original castings to visitors who wanted to copy machinery.8

8Strassmann p. 50
Another source of spillovers was the fact that one engineer, Alexander Holley, played a major role in the design of the first eleven domestic Bessemer plants. Furthermore Holley, John and George Fritz, the superintendents at Bethlehem and Cambria, Captain Hunt, chemist at Cambria, and Captain Jones, chief engineer at Carnegie Steel, met regularly to discuss technological difficulties. John Fritz wrote in his autobiography that “What each of us knew was common to all.”

The assumption of no international learning spillovers is a necessary simplification. However, in order to explain cross-country differences in productivity, some limitations on the transfer of technological knowledge must exist. Furthermore, the factors that helped information to flow between American firms concentrated in the Chicago and Pittsburgh areas seem unlikely to function over the large distances between British and American plants.

1.7 Market Structure and Pricing

To complete the model, I must characterize the the nature of competition in the American and British steel industries. One option is to assume that perfect competition prevailed at home and abroad and that it drove prices to costs. In Britain, this seems plausible given that, as early as 1878, twenty four firms operated in the industry.

Characterizing U.S. market structure poses a more difficult challenge. In the later part of my sample (1901-1913), the perfect-competition assumption would seem at complete odds with the facts. U.S. Steel had a two-thirds market share in steel rails and used the famous “Gary Dinners” to persuade fringe firms not to price aggressively. However, the formation of U.S. Steel caused prices to stabilize at $28.00 in 1901, and they did not rise again until World War I (see Figure 1-1). Hence, the degree to which U.S. Steel exploited its price leadership power remains uncertain.

9Strassman p. 50
10The delayed adoption of the reverse blooming mill provides anecdotal evidence of poor communication between American and British steel makers. The English developed this device—which allows blooms to move through rolls in both directions—in 1868 but it was not installed in the
During the nineteenth century, there is little evidence of market power. In the 1860s and 1870s, imports provided ample competition for the domestic firms. By 1880, 11 domestic firms made steel. According to Hogan nine of them had capacities over 100 thousand tons. Figure 1-2 depicts the sizes of these firms and the year each firm began operations. The absence of a close relationship between early entry and subsequent success provides additional evidence that technological advances were non-appropriable.

Starting in 1887, the domestic steelmakers formed a series of “pools” which set
quotas for each member. Most commentators expressed skepticism about the effectiveness of these early attempts at collusion.

Suppose we interpret the cost functions as the industry cost of producing a given aggregate output. Under perfect competition, the price should be set equal to the industry's marginal cost of production.

\[ p_{it} = mc_{it} = \theta_i \Phi_i \exp(-\gamma_i t_i) (1 + E_{it})^{-\lambda_i} q_{it}^{\theta_i - 1} \]  

(1.2)

Market power changes the pricing equation in two ways. First, it means the firm will charge a markup over marginal cost. Under constant elasticity of demand, \( \epsilon \), the profit maximizing markup is \( \frac{\epsilon}{\epsilon - 1} \). Second, the true marginal cost of production now must include the shadow value of experience. As shown in the appendix, this shadow value, denoted \( \mu_t \), consists of the present value of the future cost savings associated with an extra unit of current production. Taking these considerations into account, the general pricing equation becomes

\[ p_t = \begin{cases} 
mc_t & \text{under perfect competition} \\
\frac{\epsilon}{\epsilon - 1} (mc_t - \mu_t) & \text{under domestic monopoly}
\end{cases} \]  

(1.3)

If the shadow value of experience were known to be very close to zero in the years where monopoly pricing may have occurred, the pricing equation will remain linear in logs. The constant will measure \( \ln(\theta \phi) \) where \( \phi \) is the constant in the cost function. The equation should also contain a dummy variable set equal to one during the years where monopoly pricing may have occurred and zero otherwise. Monopoly pricing predicts that the coefficient on the constant will equal \( \epsilon/(\epsilon - 1) \) where \( \epsilon \) comes from the estimation of a constant-elasticity demand curve for domestic rails.

In the case of domestic steel rails, the logical period in which to expect monopoly is the period after the formation of U.S. Steel which had 66% of the steel rail making capacity and allegedly performed the role of price leader. By 1901, the year of the merger, it seems quite likely that most learning opportunities had been exhausted. Hence, the assumption that \( \mu_t \approx 0 \) for \( t > 1901 \) seems reasonable.
1.8 Estimation of Price Equations

Contemporaneous output is an endogenous variable. Hence in order to obtain consistent coefficients one must instrument for it using variables that shift the demand curve for domestic rails. Fortunately, a number of potential instruments exist.

One important shifter is the import price. Since it may be endogenous as well, it may prove desirable to use the variables that affect import prices: foreign pig iron prices and cumulative output, the duty and the exchange rate.

The aggregate demand for rails depends on overall demand for railroad services as well as the cost and availability of funds to finance new construction. Railroad traffic affects demand for rails in two ways. First, the higher the level of traffic, the quicker rails wear out and need to be replaced. Second, higher demand tends to raise the profitability of new capital expenditures on rails.

Railroad passenger traffic is chiefly influenced by consumer incomes. Rail freight traffic depends on the general level of business activity. It seems reasonable to regard railroad traffic as predeterminined with respect to rail prices.

Demand depends not only on the monetary cost of rails, but also on the financing cost. U.S. Steel Corporation reported in the 1930's that 70% of its investment funding came from railroad income, 25% came from the issue of securities, and the rest from working capital. Hence, both the cost of external and internal funds should influence the demand for rails.

Instruments for Domestic Output: Demand Shifters

- **Sum of railroad freight and passenger revenues (rev):** proxy for state of demand. (deflated into 1860 dollars)
- **Railroad dividends declared (div):** proxy for availability of internal funds and for profitability. (deflated into 1860 dollars)
- **Interest rate on railroad bonds (irate):** proxy for cost of external funds.
- **Time trend (t):** proxy for other demand factors such as the extent of comple-
tion of the railroad system and the geographical distribution of rail demand.\textsuperscript{11}

- **Mileage of rails operated (mro):** Provides an idea of the total demand for replacing iron rails and worn out steel rails. Also, in partial adjustment models of investment firms adjust their capital stocks gradually from current levels to the desired level: $I_t = \rho(K^*_{t+1} - K_t)$. Thus, depending on the relative magnitudes of the speed of adjustment and the rate of depreciation, the coefficient on railroad mileage could take either sign.

All the price equations employ the demand shifters as instruments for output.\textsuperscript{12} Results for two-stage least squares estimates of the price equations are presented in Table (1.1). The results point strongly towards the existence of large and statistically significant effects of the accumulation of production experience on costs.

The first column assumes competitive (or myopic monopoly) pricing for the entire sample. Although it includes both pig iron costs and wages, only the former has a significant impact on prices. The absence of a wage effect may occur because labor is a relatively unimportant part of costs or because the wage index is poorly correlated with true wages. Since pig iron costs include wage payments to blast furnace workers, it is possible that they capture most of the movements in steel worker wages.

The time trend is small in magnitude and insignificant statistically. This corroborates our prior belief that exogenous technological progress had no influence on costs. The positive coefficient may proxy for a omitted cost determinants that trend steadily upwards.\textsuperscript{13}

Columns (1), (2), and (4) of Table 1.1 suggest that there are static diminishing returns to scale.

\textsuperscript{11}Of the 62,000 miles of railroad added between 1870 and 1880 32\% was west of the Mississippi river while only 2\% was in New England, indicating a westward shift in rail demand as the older markets became saturated.

\textsuperscript{12}I also experimented with miles of road added in the previous year as an instrument. Theoretically, it should be positively correlated with demand if railroad lines take more than a year to complete or if there is gradual adjustment to the optimal railroad size. However, it entered the first stage with a negative and insignificant coefficient.

\textsuperscript{13}A negative coefficient could have been interpreted as either exogenous progress or an alternate measure of experience, cumulative years of production.
Table 1.1: Price Equation for American Steel Rails

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IV</td>
<td>IV</td>
<td>OLS</td>
<td>IV</td>
<td>IV</td>
<td>IV</td>
</tr>
<tr>
<td>constant</td>
<td>2.830</td>
<td>2.906</td>
<td>2.348</td>
<td>2.792</td>
<td>2.152</td>
<td>2.152</td>
</tr>
<tr>
<td></td>
<td>(0.534)</td>
<td>(0.487)</td>
<td>(0.368)</td>
<td>(0.482)</td>
<td>(0.465)</td>
<td>(0.466)</td>
</tr>
<tr>
<td>trend</td>
<td>0.006</td>
<td>0.007</td>
<td>-0.001</td>
<td>0.007</td>
<td>-0.005</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.009)</td>
<td>(0.008)</td>
<td></td>
</tr>
<tr>
<td>ln(picubic)</td>
<td>0.594</td>
<td>0.565</td>
<td>0.772</td>
<td>0.603</td>
<td>0.801</td>
<td>0.794</td>
</tr>
<tr>
<td></td>
<td>(0.198)</td>
<td>(0.142)</td>
<td>(0.092)</td>
<td>(0.138)</td>
<td>(0.128)</td>
<td>(0.128)</td>
</tr>
<tr>
<td>ln(1 + Eus)</td>
<td>-0.287</td>
<td>-0.314</td>
<td>-0.114</td>
<td>-0.291</td>
<td>-0.127</td>
<td>-0.174</td>
</tr>
<tr>
<td></td>
<td>(0.106)</td>
<td>(0.099)</td>
<td>(0.017)</td>
<td>(0.112)</td>
<td>(0.089)</td>
<td>(0.054)</td>
</tr>
<tr>
<td>ln(qus)</td>
<td>0.206</td>
<td>0.239</td>
<td>0.210</td>
<td>0.046</td>
<td>0.093</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.116)</td>
<td>(0.127)</td>
<td>(0.099)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(wage)</td>
<td>-0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.540)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USSteel</td>
<td></td>
<td></td>
<td></td>
<td>-0.022</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.110)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>0.951</td>
<td>0.950</td>
<td>0.959</td>
<td>0.951</td>
<td>0.965</td>
<td>0.965</td>
</tr>
<tr>
<td>S.E.R.</td>
<td>0.129</td>
<td>0.131</td>
<td>0.119</td>
<td>0.130</td>
<td>0.119</td>
<td>0.119</td>
</tr>
<tr>
<td>D.W.</td>
<td>1.294</td>
<td>1.327</td>
<td>1.097</td>
<td>1.301</td>
<td>1.149</td>
<td>1.157</td>
</tr>
<tr>
<td>Begin</td>
<td>1867</td>
<td>1867</td>
<td>1867</td>
<td>1867</td>
<td>1867</td>
<td>1867</td>
</tr>
<tr>
<td>End</td>
<td>1913</td>
<td>1913</td>
<td>1913</td>
<td>1913</td>
<td>1900</td>
<td>1900</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. S.E.R. stands for the standard error of the regression. $q_{us}$ is the sum of U.S. production for the domestic market and U.S. exports to foreign markets. The instrumental variable list consists of the included exogenous variables, as well as the demand shifters, British experience and pig iron prices, the exchange rate, and the duty. (All in natural logs.)
The large standard error on output implies that we cannot draw precise statistical inferences on the existence of rising marginal costs. However, the sharp decline in the learning coefficient that occurs current output is excluded—column (3)—adds support for the diminishing static returns hypothesis.

Column (4) allows for a change in pricing behavior following the formation of U.S. Steel Corporation in 1901. The dummy variable USSteel takes a value of zero from 1867 to 1900 and a value of one thereafter. Hence, it allows the constant to shift up as a result of U.S. Steel's alleged price leadership policies. Contrary to expectation, U.S. Steel appears to have had no effect at all on the level of steel rail prices. Nevertheless, the total lack of price movement following the formation of U.S. Steel probably implies that the competitive model is not appropriate for the 1900s.

Columns (5) and (6) restrict the sample to the pre-U.S.S. years. Not surprisingly, the coefficient on pig-iron costs increases in the reduced sample. This is because steel rail prices were held constant in the face of quite variable pig iron prices between 1901 and 1913. The reduced sample causes a marked decline in the t-statistics on current and cumulative output. The time trend switches sign without becoming significant. Apparently, in the reduced sample it becomes difficult to disentangle the effects of these three variables, each of which has a strong upward trend. Since the time trend never appears significantly, and since we have a priori grounds to doubt the importance of exogenous technological influences, I omit the time trend in the regression reported in the sixth column. This restores the statistical significance of the learning effect without any noticeable reduction in the “fit” as measured by either the standard error of the regression or the $\bar{R}^2$.

Table 1.2 presents results for two samples of the British data. Here the reason for restricting the sample comes from suspect data rather than changes in market structure. The data prior to 1867 comes from a different source and since there were large fluctuations in the general price level during the civil war, there was a possibility of introducing noise. As it turns out, exclusion of the years 1862-1866 slightly reduces the standard error of the regression without having a major impact on the other coefficients.
Table 1.2: Price Equations for British Steel Rails

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>OLS</th>
<th>IV</th>
<th>IV</th>
<th>OLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>\text{constant}</td>
<td>2.977</td>
<td>3.019</td>
<td>2.929</td>
<td>3.061</td>
<td>3.156</td>
</tr>
<tr>
<td></td>
<td>(0.199)</td>
<td>(0.208)</td>
<td>(0.225)</td>
<td>(0.277)</td>
<td>(0.250)</td>
</tr>
<tr>
<td>\text{trend}</td>
<td>0.014</td>
<td>0.016</td>
<td>0.018</td>
<td>0.018</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>\ln(p_{\text{ip}uk})</td>
<td>0.840</td>
<td>0.820</td>
<td>0.821</td>
<td>0.848</td>
<td>0.864</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.065)</td>
<td>(0.067)</td>
<td>(0.060)</td>
<td>(0.057)</td>
</tr>
<tr>
<td>\ln(1 + E_{uk})</td>
<td>-0.210</td>
<td>-0.258</td>
<td>-0.319</td>
<td>-0.301</td>
<td>-0.233</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.068)</td>
<td>(0.088)</td>
<td>(0.087)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>\ln(q_{uk})</td>
<td>0.054</td>
<td>0.136</td>
<td>0.099</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.075)</td>
<td>(0.102)</td>
<td>(0.123)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>\text{R}^2</td>
<td>0.976</td>
<td>0.976</td>
<td>0.974</td>
<td>0.961</td>
<td>0.961</td>
</tr>
<tr>
<td>\text{S.E.R.}</td>
<td>0.100</td>
<td>0.101</td>
<td>0.100</td>
<td>0.090</td>
<td>0.090</td>
</tr>
<tr>
<td>\text{D.W.}</td>
<td>1.550</td>
<td>1.526</td>
<td>1.577</td>
<td>1.828</td>
<td>1.841</td>
</tr>
<tr>
<td>Begin</td>
<td>1862</td>
<td>1862</td>
<td>1863</td>
<td>1867</td>
<td>1867</td>
</tr>
<tr>
<td>End</td>
<td>1913</td>
<td>1913</td>
<td>1913</td>
<td>1913</td>
<td>1913</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. \text{S.E.R.} stands for the standard error of the regression. \( q_{uk} \) is the sum of English exports to the United States and their production for all other markets. In equation (2) lagged output instruments for \( \ln(q_{uk}) \). The instrumental variable list for equations (3) and (4) consists of the include exogenous variables, British steel output not destined for the U.S. market, American demand shifters, experience, and pig iron prices.
The British price equations have coefficients that are remarkably similar to the U.S. coefficients. The learning elasticities differ by small amounts and the elasticity with respect to pig iron prices lies only a small amount over the pre-U.S.S. sample result. This divergence may arise from the use of lower quality pig iron in Britain.

The coefficients from the two countries differ most in terms of the time trend and current output. British prices have a residual one and a half percent annual rate of increase. This trend is statistically significant. On the other hand, current output does not appear to have a significant effect. Column (3) uses lagged British steel output as an instrument since the U.S. demand shifters do not extend back before 1867. Column (4) uses the U.S. demand shifters as well as U.S. cumulative output and pig iron prices. In neither case do we find large or significant effects of current output. Column (5) shows that a constant returns specification seems to fit the British data well while providing an eminently sensible estimate of the learning elasticity: 0.23.

Since the hypothesis of constant returns is not rejected by the British data, the import price (rather than its determinants: U.K. pig iron prices, the exchange rate, and the duty) can probably be used as an instrument for domestic production. I reestimate column (6) of Table 1.1 and obtain similar qualitative results, but more appealing coefficient estimates.\textsuperscript{14}

\[
\ln(s_r p_{us}) = 2.36 + 0.71 \ln(p_{ic_{us}}) - 0.23 \ln(1 + E_{us}) + 0.17 \ln(q_{us})
\]

(4.7) \hspace{1cm} (5.1) \hspace{1cm} (3.7) \hspace{1cm} (2.1)

The t-statistics in parentheses show that each right-hand-side variable is statistically significant. This regression has an $R^2$ of 0.96 and a standard error of 0.12. A Hausman regression-based test \textit{does not} reject the 5 overidentifying restrictions at the 10% confidence level. This regression provides an estimate of $\lambda$, 0.23 that is very close to the one generated in column (5) for the British data. It also corresponds to the mean of the 96 products reviewed in Ghemawat (1985). Furthermore, we obtain a fairly precise and reasonable estimate on the degree of rising marginal costs. Under

\textsuperscript{14}T-statistics in parentheses.
marginal cost pricing the mark-up over average variable costs depends on $\theta$. 

\[
\frac{p - auc}{p} = \frac{\theta - 1}{\theta} \tag{1.4}
\]

Substituting in $\theta = 0.1738$ the implied mark-up over average variable costs is 0.148. This number lies quite close to the mark-ups calculated from the census of manufacturing data supplied in Allen (1981). The actual steel rail markups over all variable costs were 0.15 in 1879 and 0.12 in 1889. The proximity of these numbers provides some support for the estimate of $\theta$ and the assumption of competitive pricing.

For comparison purposes, it seems worthwhile to consider the crude specification frequently used on modern data. This deflates the product price by the consumer price index and computes a linear-in-logs regression on the constant and cumulative output only.\(^{15}\)

\[
\ln(\frac{SPR_{pa}}{cpi}) = 4.81 - 0.15 \ln(1 + E_{us}) \tag{50.2}
\]

\[
-0.15 \ln(1 + E_{us}) \tag{13.0}
\]

The $R^2$ of this regression, 0.84, is considerably lower and the standard error of the regression, 0.19, considerably higher than the specifications used above. However, it turns out that despite the use of a poor deflator (the cpi contains agricultural prices which have little to do with the cost of steel production) and the omission of static scale effects, the learning effect remains in the correct neighborhood.

The extremely robust result of the price regressions is that production experience, measured as cumulative output, has a large and statistically significant impact on steel rail prices. Moreover, the range of estimated elasticity estimates corresponds closely to the values reported in Ghemawat (1985) and Argote and Epplle’s (1990) survey articles on learning-by-doing.

The existence of strong learning effects makes it plausible, but by no means necessary, that protection could be welfare-improving. Before proceeding to the simulations that provide evidence on the positive and normative effects of the duty, we need to address the issue of whether cumulative output is proxying for an omitted factor. The

\(^{15}\)T-statistics in parentheses.
most obvious candidate is the physical capital stock.

1.9 Physical Capital Accumulation

An alternate explanation for the decline of domestic steel rail prices involves capital accumulation. Gradual expansion of the physical capital stock could explain two of the empirical findings of this paper: (1) long-run declining prices not fully explainable by declines in input costs, (2) short-run rising marginal costs. If capital accumulation explains the price reductions, then the argument that the duty promoted the industry's acquisition of comparative advantage becomes less plausible. We would expect the removal of the duty to lead to deaccumulation of capital and consequent decline of the domestic industry. Since the industry continued to prosper and expand without import protection until the 1960's, the capital-accumulation explanation does not seem that likely. Nonetheless, it merits further consideration.

It is difficult in practice to distinguish between the roles of capital accumulation and learning-by-doing in causing cost reductions. If we use full capacity output as a measure of the capital stock we include changes in capacity that arose from learning rather than capital investments. However, even if we had a measure of the monetary value of the capital stock it would not represent a distinct source of cost reductions. This is because many of the innovations derived from learning-by-doing were embodied in new plant and equipment.

If capital accumulation in a constant returns world had caused the observed reductions in steel rail prices then one would expect to see the productivity of capital decline over time. Furthermore, one would expect the size of the capital stock to play an important explanatory role in the price determination equation. I illustrate these points within a simple Cobb-Douglas production function. Let \( q_t = A_tK_t^\alpha M_t^{1-\alpha} \)
where \( K \) is the capital stock, \( M \) represents pig iron, and \( A \) indicates the state of technology. Let the costs of capital and pig iron be \( r \) and \( v \). Assume that the capital stock can only be increased gradually and that it initially lies below the optimal level given factor prices, i.e. \( K_0 < K^*(r,v) \). Note first that the average product of capital,
\(q_t/K_t\) is a decreasing function of \(K_t\). Second, note that the short run cost function is

\[C_t(q_t) = rK_t + \nu M(K_t, q_t)\]

where \(M(K_t, q_t) = (q_t/A_t)^{1-\alpha} K_t^{\alpha-\delta}\). Assume that technological change, if any, arises from learning-by-doing: \(A_t = E_t^\lambda\) where \(E_t\) is cumulative output. Differentiating the cost function, taking logs and assuming competition yields

\[
\log(p_t) = \log(C_t'(q_t)) = \log(\nu) + \frac{\alpha}{1-\alpha} \log(q_t) - \frac{\alpha}{1-\alpha} \log(K_t) - \frac{\lambda}{1-\alpha} \log(E_t). \tag{1.5}
\]

If cumulative output in the previous estimates were simply a proxy for physical capital one would expect that capital stock measures would enter the equation with a negative and significant coefficient and eliminate the statistical significance of the cumulative output term.

There is no consistent data series on the capital stock. However, for the first 18 years of the industry's existence, I have constructed several measures of the capital stock using firm level data. During that period all steel rails were made using the Bessemer process. Furthermore, the new steel works produced few other products. Hence, Bessemer capacity should correspond fairly closely to steel rail capacity. One might measure the aggregate Bessemer capital stock as the number of Bessemer works. However, virtually all steel works had multiple converters. Furthermore the converters existed in various sizes, having capacities that ranged between 4 and 10 tons of metal per heat. This suggests that we might want to measure the capital stock as the sum of the total tonnages per heat that the existing converters are capable of producing. Alternatively, it may be that aggregate capital stock does not matter as much as capital per work so we could divide total tonnage by the number of works. It turns out that all these methods of measuring the capital stock lead to the conclusions that the average productivity of capital increased by large amounts during the capital accumulation period and none of the measures detract from the explanatory role of cumulative output in the price equations.

Output of steel rails per Bessemer converter rose from 575 tons in 1867 to 39,580
tons in 1882. Meanwhile output per converter-ton rose from 90 tons in 1867 to 6,000 tons in 1882. Figure 1-3 shows the growth of Bessemer ingot output per converter and per converter-ton. After 1884, data on the number of converters existed for 1892 and 1901 only (A dotted line interpolates between these points).

Capital Stock Measures

- **works**: Number of Bessemer steel works.
- **convs**: Number of Bessemer steel converters.
- **conv-tons**: Total converter tonnage, i.e. the aggregate capacity of steel output per "heat" of the existing stock of converters.
- **work size**: Converter tonnage per steel work (conv-tons/works).

The price regressions reported in Table (1.3) fail to support the idea that physical capital accumulation played an important role in cost reductions. None of the measures of the capital stock enter significantly and only two, the number of converters and converter-tons per steel work, even have the correct sign. Most importantly, the coefficient on cumulative output retains its statistical significance and approximate magnitude. Thus learning-by-doing explains price reductions relative to input costs much better than measures of the physical size of the capital stock.
Figure 1-3: The Productivity of Physical Capital

A. Output per converter (—) and the number of converters (---) where 1867 = 1.0

B. Output per converter-ton (—) and converter capacity (---) where 1867 = 1.0
Table 1.3: Price Equations with Capital Accumulation

<table>
<thead>
<tr>
<th></th>
<th>Dependent Variable: ln(stp_u)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>IV</td>
</tr>
<tr>
<td>constant</td>
<td>1.749</td>
</tr>
<tr>
<td></td>
<td>(0.633)</td>
</tr>
<tr>
<td>ln(pic_u)</td>
<td>0.846</td>
</tr>
<tr>
<td></td>
<td>(0.157)</td>
</tr>
<tr>
<td>ln(1 + E_u)</td>
<td>-0.278</td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
</tr>
<tr>
<td>ln(q_u)</td>
<td>0.220</td>
</tr>
<tr>
<td></td>
<td>(0.133)</td>
</tr>
<tr>
<td>ln(works)</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>(0.133)</td>
</tr>
<tr>
<td>ln(cons)</td>
<td>-0.044</td>
</tr>
<tr>
<td></td>
<td>(0.203)</td>
</tr>
<tr>
<td>ln(cons)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(worksize)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>R²</td>
<td>0.943</td>
</tr>
<tr>
<td>S.E.R.</td>
<td>0.123</td>
</tr>
<tr>
<td>D.W.</td>
<td>1.668</td>
</tr>
<tr>
<td>Begin</td>
<td>1867</td>
</tr>
<tr>
<td>End</td>
<td>1884</td>
</tr>
</tbody>
</table>

Note: Standard errors in parentheses. S.E.R. stands for the standard error of the regression. q_u is the sum of U.S. production for the domestic market and U.S. exports to foreign markets. The instrumental variable list consists of the included exogenous variables, as well as the demand shifters, British experience and pig iron prices, the exchange rate, and the duty. (All in natural logs.)
Implications of Forward Pricing

Consider a monopoly with revenue function $R(\cdot)$ and cost function $C(\cdot)$. Suppose that costs each period depend on current output, i.e. static returns to scale, and cumulative output, i.e. dynamic returns to scale. The value function at time $t$ is:

$$V(E_t, t) = \max_{q(s)} \left\{ \sum_{s=t}^{T} \delta^s [R(q(s), s) - C(q(s), E(s), s)] \right\}$$

(1.6)

where $\delta = \frac{1}{1+r}$ and $E_t$ is cumulative output.\(^{16}\) This can be rewritten as

$$V(E_t, t) = R(q_t, t) - C(q_t, E_t, t) + \delta V(E_{t+1}, t + 1)$$

(1.7)

The constraint is that

$$E_{t+1} = E_t + q_t$$

(1.8)

Maximizing 1.7 with respect to $q_t$ subject to 1.8 leads to the following condition:

$$R_q(q_t, t) - C_q(q_t, E_t, t) + \delta V_q(E_t + q_t, t) = 0$$

(1.9)

The envelope theorem implies that only the direct effect of current output on future costs matters. Hence, $V_q$ is just $- \sum C_E$. Thus, the condition becomes

$$R_q(q_t, t) - C_q(q_t, E_t, t) + \mu_t = 0$$

(1.10)

where $\mu_t$ is the shadow value of extra production experience.

$$\mu_t \equiv - \sum_{s=t+1}^{T} \delta^{s-t} C_E(q_s, E_s, s)$$

where $E_s = E_t + q_t + \sum_{r=t+1}^{s-1} q_r$.

The profit-maximizing choice of $q_t$ depends on static marginal costs, denoted $mc(q_t; E_t, t)$, and the shadow value of experience, denoted $\mu(q_t; E_t, q_t)$ where $q_t$

\(^{16}\)Let $q_0 = 0$ and $E_0 = 1$ in order to avoid division by zero in the first period of production.
is a vector of the optimal choices for output from period $t+1$ to $T$.\textsuperscript{17} Thus the optimal choice in any period depends on past and future choices. Iteration starting with an initial guess for the path of output and continuing until a fixed point is reached will solve the dynamic optimization problem.

A firm that chooses the optimal dynamic price will hold down markups initially when experience is low and costs are high and raise markups over time. Hence it is not clear that prices will even fall as output accumulates. Spence (1981) showed that in continuous time, under constant returns to scale and constant input prices, the optimal price will be a constant set equal to a markup over final-period marginal costs. For the general case, prices will decline over time but not nearly as much as they would if the firm priced myopically.

To make further progress, we need to assume a functional form for demand. Isoelastic demand curves have two advantages. First, the optimal pricing equation does not depend directly on the variables that shift the demand curve.

\[ p_t = \frac{\epsilon}{\epsilon - 1}(mc(q_t; E_t, t) - \mu(q_t; E_t, q_t)) \]  

(1.11)

Second, the price equation can be estimated using linear regression if the monopolist is myopic with respect to the learning curve. Hence, we can easily obtain benchmark parameters under the myopia assumption and use them to illustrate the effects of forward-looking monopoly behavior.

In the myopic case given isoelastic demand and cost functions we can solve for prices and output each period. However, for $\mu_t > 0$ we can no longer take logs and obtain a linear equation. Hence, each period we must use an iterative procedure to solve for domestic output. This procedure will solve for quantities in periods 1 through $T$. Then we substitute those new values of future quantities into $\mu_t$ and repeat the process until the new and old vectors of quantities differ by negligible amounts.

In order to determine the optimal path of prices, I must estimate a demand curve

\textsuperscript{17}$q_t \equiv [q_{t+1}, q_{t+2}, \ldots, q_{T-1}]$
for domestic production. I use domestic pig iron costs and cumulative output to instrument for the domestic price and take the import price as exogenous. The estimation uses a sample beginning in 1867 and ending in 1913.

\[
\ln(q_1) = -1.71 -0.265 \text{ trend} +1.05 \ln(\text{div}) +6.46 \ln(\text{rev})
\]

(5.847) (0.033) (0.500) (1.45)

\[-0.607 \text{irate} -4.70 \ln(\text{mro}) +2.02 \ln(\text{srp2}) -4.09 \ln(\text{srp1})\]

(0.203) (1.803) (0.994) (1.000)

All of the t-statistics in the above regression exceed 2.0. Moreover, price effects seem quite strong. The own price elasticity of -4.09 implies an optimal (under myopia) Lerner index (i.e. markup over marginal costs) of 24%. Dividends, revenues and the interest rate on railroad bonds each enter with the expected sign. The strong significance of the interest rate is an unusual result for an investment equation. The negative time trend may be interpreted as a sign that the American rail system was gradually becoming complete and hence leading to a gradual reduction in demand for new rails. In terms of the overall fit the demand equation achieves a corrected $R^2$ of 0.934. However, the 0.45 standard error of the regression seems quite high.

Using this demand curve and cost parameters derived from the pricing equations of Table 1.1 under the assumption of myopic monopoly pricing, I stimulate the path of domestic price and output under various discount factors. Figure 1-4 shows that the myopic pricing model fits the historic data quite well. This is not too surprising since the parameters were derived under the assumption of myopic pricing. Fully forward-looking prices differ substantially. They start much lower and decline much less. This is true for learning elasticities of both 31% (Figure 1-4A) and 11% (Figure 1-4B).

Figure 1-5 shows that the Lerner index starts out very negative for all reasonable interest rates. Markups then move fairly rapidly towards the myopic level of 24%. By 1900 forward pricing appears indistinguishable from myopic pricing.

Figure 1-6 deflates steel rail prices with the pig iron price and shows how this real price relates to cumulative output. Under myopic monopoly pricing there is easily observable negative linear relationship between the log of price and the log of
Figure 1-4: Optimal Dynamic Pricing

A. Historical Price of steel rails(—) and simulated price for $\delta = 1$ (---) and $\delta = 0$ (⋅⋅⋅). Coefficients from column (2) of Table 1.1

B. Historical Price of steel rails(—) and simulated price for $\delta = 1$ (---) and $\delta = 0$ (⋅⋅⋅). Coefficients from column (4) of Table 1.1
cumulative output. Under forward pricing, this relationship seems to disappear from the third observation onwards. This occurs despite the fairly high discount rate of 10%.

Thus forward pricing seems to contradict the frequently observed pattern of a strong linear-in-logs relationship with elasticity of approximately 25%. There are several possible explanations for the failure of most data to conform with the predictions of forward pricing. One is simply that learning is a public good and industries are fairly competitive. This seems reasonable for the steel rail case. However, in cases where each firm produces a highly differentiated product, the incentives to forward price must be quite high. Hence explanations for pricing based on current marginal costs need to explain why the future seems to have no effect. One possibility is some form of managerial myopia. Majd and Pindyck (1989) have advanced an alternate
Figure 1-6: Pricing and the Learning Curve: A scatterplot of efficiency on cumulative output under myopia and foresight.

One way to understand this is to recognize that when a firm faces a learning curve, part of its cost of producing is in fact an investment expenditure: the firm is investing in reduced future costs. This is an irreversible investment, i.e. the expenditure is sunk. When future prices (and hence, the payoff from investing) are uncertain, the firm incurs an option of waiting to see how the price will evolve. The greater the uncertainty about future prices, the greater this opportunity cost, and the less investing the firm should do.

Although all three explanations of the absence of forward pricing behavior generate the same empirical prediction, the policy implications may differ substantially. Only
in the case of external learning and competitive pricing is it clear that the stimulation of output has the potential to increase welfare.
Chapter 2

The Tariff on Steel Rails

It will be readily acknowledged from a perusal of the duties levied by the United States government on metallic products, that it is little short of marvelous that England should send any iron at all into that country. It must be very discouraging to those who strive for the adoption of true politico-economic principles, to find that the youngest and in some respects most advanced country in Christendom should place the greatest restrictions upon commerce. And yet so it is.


The steel manufacture of this country is eminently the child of the protective policy, and its healthy growth and beneficent influence illustrate most signally the wisdom of that policy.


This paper examines the experience of a protected industry that developed rapidly during the years it was protected and maintained its competitive position after the removal of government assistance. I will focus on two questions:

1. Would the industry have developed in the absence of protection?

2. Did protection raise domestic welfare?

This chapter studies the joint roles of learning-by-doing and protection from imports in the early development of the American steel industry. I concentrate on a specific
product, steel rails. Rails were the principal product of the new steel industry which was one of the major manufacturing sectors in post bellum U.S. The greater strength and durability of steel rails led to rapid displacement of iron rails and assisted the late 19th century construction of a national transportation network. The steel rails "industry" possesses several characteristics that make it a plausible case of infant industry protection.

- **U.S. was initially disadvantaged:** English had 8 year head start (1859 vs. 1867)\(^1\) and a 100% import share between 1862 and 1866. The initial U.S. price was $166 per ton. The same year, 1867, British rails cost $135.60 at U.S. ports and $90.67 in England.

- **U.S. ultimately obtained comparative advantage:** In 1913, the year the duty was eliminated, steel rails were priced at $28.00 in the U.S. and $32.23 in England. Imports remained at almost negligible levels long after the elimination of the duty. Furthermore, American firms eventually dominated many export markets. The U.S. exported 450,000 tons of rails in 1913 compared to 500,000 from the United Kingdom and 493,000 from Germany (The fourth largest exporter, Belgium exported only 162,000 tons.)\(^2\)

- **Substantial Evidence of Industry-Wide Learning-by-doing:** Chapter 1 presented both technological anecdotes and statistical corroboration for the premise that cumulative domestic production led to major reductions in the costs of producing steel rails.

- **Government intervened in a visible and temporary manner:** The U.S.

---

\(^1\)The delayed start of the U.S. industry may be attributed to several factors. First, the Bessemer process was invented in England. An American, William Kelley, conceived a similar technique for steel production in Kentucky and conducted numerous experiments in the 1850s. Although his techniques never became commercially useful he did obtain a patent which became part of a second impediment to the early establishment of the U.S. steel industry: patent conflicts. In order to produce steel using the Bessemer process, one needed control of patents owned by Kelley, Bessemer, and another English inventor, Robert Mushet. Consolidation of these patents did not occur until 1866. The interruptions of commerce associated with the Civil War may have contributed as well to the late U.S. adoption of the process. (Strassmann p.27-28)

\(^2\)Caplin and Tarr pp.248,252
government confined its intervention to a specific duty on imported rails.\(^3\)

The specific goals of this paper are to determine the importance of the duty in the development of the American steel rail industry and to reach some conclusions on the welfare consequences of the duty. The method followed will be to:

1. Present a simple model of the steel rail industry in which input costs, learning-by-doing, and the duty jointly determine prices and the import share.
2. Estimate the model using time-series data from the U.S. and Great Britain.
3. Simulate what would have happened to prices, domestic production and imports if the U.S. had pursued a policy of free trade.
4. Calculate the impact of the duty on domestic welfare.
5. Examine the interaction between changing relative resource endowments and the consequences of protection.
6. Using hindsight, calculate the optimal form and path of intervention.

Advocates of protection often take the behavior of foreign governments and firms as a given, unaffected by the tariff. Two different types of foreign response will substantially weaken the case for protection. The first and obvious response is retaliatory duties or export subsidies. In a trade war, both countries almost inevitably lose. The second response to domestic protection could be a reduction in foreign learning. The first response was not an important aspect of the case studied here because of England’s avowed policy of free trade and because the English market for rails was much smaller than the American one. The second effect will be explicitly modeled.

Unlike many modern cases of infant industry promotion, steel rails were affected by a single, explicit policy instrument that varied over time. The U.S. government confined its intervention to a specific duty on imported rails.

\(^3\)This simplifies modeling as well as counterfactual simulations. It also avoids the interpretation problem present in Baldwin and Krugman’s work of whether low U.S. share of the Japanese market reflected American firm behavior or implicit Japanese protection.
Figure (2-1) displays the duty expressed in both dollars per ton and as percentage of the pre-duty import price. 4

Figure 2-1: The Steel Rail Tariff 1866-1913

At the same time as the price decline occurred, the American steel rail manufacturers rapidly captured the domestic industry. Figure 2-2 depicts the dwindling importance of imports. The absence of data for the 1867-1870 period occurs because initially the U.S. customs department did not differentiate between iron and steel rails. However we do know that all steel rails installed in the U.S. between 1862 and 1866 came from Britain.

4Unfortunately, history does not provide us with a "before" period in order to clarify the effects of the duty. Steel rails were protected prior to the resolution of the patent dispute in 1866 which provided the first opportunity for the establishment of the industry. For an industry where an error by the Treasury department left a steel product unprotected until 1890, see my forthcoming chapter on the tinplate duty.
2.1 Natural Resources or Government Intervention?

The export success and absence of significant import competition from roughly 1890 onwards suggests that something must have happened to change the competitive status of the U.S. steel rail producers. One possibility is an exogenous shift in primary factor abundances. The other is that learning-by-doing led to endogenous advances of the domestic industry.

Writing around 1915, F.W. Taussig argued for the former position. I reformulate his views as follows:
Discoveries of new natural resources (iron ore and coal) and the construction of infrastructure (railroads and canals) shifted U.S. comparative advantage towards metal production. This shift is reflected in the falling price of U.S. pig iron which accounted for over 3/4 of the production costs of steel. The decline in U.S. pig iron prices relative to foreign pig iron prices (see Figure 2-3) can account for the rise of the American industry. The steel rail duty should therefore be seen as a wasteful and redistributive attempt to hasten a change in competitiveness that would have occurred in any case.

![Figure 2-3: Real Pig Iron Prices in the U.S. (—) and the U.K. (− − −)](image)

The alternative hypothesis asserts the shifts in relative resource abundance reflected in the decline in domestic pig iron prices relative to foreign pig iron prices may well have caused a shift in latent comparative advantage. However, in order to realize this potential, the domestic firms would need to catch up to foreign pro-
ducers in terms of the efficiency of their techniques of production. Unfortunately, a large component of technological capability resides in “tacit knowledge” which can only be learned through a groping trial and error process. Hence, a protective duty was necessary to provide an initial boost to domestic production levels which led to learning-based cost reductions, which enabled the domestic industry to compete with its older and more experienced international rival.

Figure 2-4: Steel rail(—) and pig iron(-- -) price differences in constant 1860 dollars

Figure (2-4) presents some preliminary evidence on the issue. It illustrates the movement in the difference between the American and British prices of pig iron (dashed line) and steel rails (solid line). The differences are measured in constant U.S. currency. While the price data suggest that movements in the relative input costs can explain some of the movement in steel rail price differences, there is still an
unexplained long-term downward trend in U.S. rail prices relative to foreign prices. The question is how much of that trend can be attributed to tariff-induced learning-by-doing.

2.2 Demand

The American Iron and Steel Association (AISA) collected two steel rail price series, one for domestic rails at Pennsylvania mills and the other for imported British rails at U.S. ports. The often substantial gaps between these prices and the correlation between the price gap and the import share provide indirect evidence for imperfect substitutability between American and British steel rails (see Figure 2-2). This raises the question of what factor could differentiate such a seemingly homogeneous good as a steel rail.

The cost of transporting rails over land provides one explanation for such differentiation. James Swank of AISA wrote, “Heavy products of iron and steel, for instance, can be carried much more cheaply from Liverpool to the Gulf ports of the U.S. than from our own rolling mills and blast furnaces which are not situated on the sea coast or on the Mississippi river, and very few of them are so situated.”

Writing in 1890, when the pre-duty price of imported steel rails was slightly less than the domestic price, Taussig argued that the “cost of transportation from the sea-board to the interior is such that, even in the absence of the duty, steel rails would be imported only to supply railways near tide-water.” (Taussig 1931 p.272)

Railroad builders had to pay the cost of transporting rails from the ports or mills to the construction site. As a result, one would expect them to consider the relative distance to purchase point as well as relative prices in determining whether to buy from foreign or domestic suppliers.

Initially the bulk of railroads were concentrated in the Mid-Atlantic and New England. The expansion of the railroads into the midwest did tend to benefit the domestic mills located in the interior. However, this cannot in itself explain the growth

---

5 Transatlantic freight costs were small partly because rails were used as ballast.
of the domestic industry. Shifts in demand that favor the U.S. should cause rises in the U.S. price relative to import prices. However, the reverse occurred. Moreover, despite the distance of the U.S. mills to the ports, the U.S. became a major exporter. Furthermore, demand also grew along the Gulf and Pacific Coasts which the British could reach at lower cost than the American producers. The expanding rail networks themselves may have added to the facility of serving interior markets.

Geographical dispersion of rail demand need not have been the only source of differentiation between domestic and foreign rails. There were probably differences in rail durability as well as supply reliability.\footnote{The first data on railroad freight revenue per ton-mile states that railroads received $0.01236 per ton-mile in 1882. This implies a $3.76 per ton charge for the 305 mile trip from Philadelphia to Pittsburgh. By 1902, the implied cost of such a trip had fallen to $2.31. These charges by themselves do not seem sufficient to explain the large gaps between Pittsburgh domestic prices and imported prices at the ports.}

\section*{2.3 The Demand Model}

In order to quantify the effects of the duty on domestic output and imports, I require a specific model of steel rail demand. The constant elasticity model used in the appendix to Chapter 1 fit the data reasonably well and had fairly sensible own and cross price elasticities. However, it suffers from at least two serious defects. As shown in Head (1990) the corresponding constant-elasticity import demand function does not fit the data well at all. Moreover, the constant elasticity demand curves lack utility-function foundations. This makes them vulnerable to the Lucas Critique, and, of more practical importance, it forces the use of numerical methods (rather than evaluation of a a well-defined closed-form function) to calculate the effects of the duty on consumer surplus.

Furthermore, the linear demand curve, unlike the constant elasticity demand curve, has a choke-off point. A sufficiently large difference between domestic and import prices there will eliminate all demand for domestic or imported rails. This seems like a desirable feature given the large number of years where low U.S. prices combined with high import duties kept out virtually all imports.
I have come to the conclusion that alternate demand model is required. The one developed here still fits the data reasonably well. Following estimation, one can back out the key parameters of preferences which are assumed to be unaffected by trade policy.\footnote{However, these improvements come at the cost of making the aggregate demand for steel rails insensitive to prices. The duty only affects market share, not the domestic market size. This seems fairly reasonable given that there is no good substitute for rails in building and operating a railroad system. Iron rails were so inferior to steel rails that virtually production of them ceased in the early 1880s. While it is possible to defer replacement of rails that have worn out, this short term effect does not seem very important in analyzing the long run consequences of protection. This still leaves the effect of steel rail prices on the derived demand for railroad services. In subsequent sections I will comment on how the omission of derived demand effects might affect the results.}

Suppose each consumer obtains a base level of utility \( v \) from consuming the good plus a bonus (or deduction) of \( \theta \) if she consumes the domestic good which will be denoted with a subscript of \( i = 1 \). From this she must subtract the price of the good. Hence

\[
\text{cs}(\theta, i) = v + \theta I(\theta) - p_i, \tag{2.1}
\]

where \( I(\theta) \) maps consumers of various types, \( \theta \), to consuming either domestic good 1, \( I = 1 \), or imported good 2, \( I = 0 \). In principle, there exists a type \( \tilde{\theta} \) such that a consumer of that type is indifferent between goods 1 and 2, i.e. \( \text{cs}(\tilde{\theta}, 1) = \text{cs}(\tilde{\theta}, 2) \). Not surprisingly, the utility bonus from consuming good 1 must exactly offset the extra monetary cost. Hence, \( \tilde{\theta} = p_1 - p_2 \). The taste parameter \( \theta \) is uniformly distributed between \( \underline{\theta} \) and \( \bar{\theta} \). Suppose there are \( Q \) consumers who each purchase 1 unit. Then the density for each \( \theta \) is \( f(\theta) = \frac{Q}{\bar{\theta} - \underline{\theta}} \). For interior values of \( \bar{\theta} \), i.e. \( \underline{\theta} < p_1 - p_2 < \bar{\theta} \), demand can be determined by integrating over the density function.

\[
q_1 = \int_{\underline{\theta}}^{\bar{\theta}} f(\theta) d\theta = \frac{Q(\bar{\theta} - p_1 + p_2)}{\bar{\theta} - \underline{\theta}} \tag{2.2}
\]

\[
q_2 = \int_{\underline{\theta}}^{\tilde{\theta}} f(\theta) d\theta = \frac{Q(p_1 - p_2 - \theta)}{\bar{\theta} - \underline{\theta}} \tag{2.3}
\]

Noting that \( q_1 + q_2 = Q \), it is clear that market share is a linear function of the
price difference.

\[ x_1 = \frac{a}{Q} = (1 - a) - b(p_1 - p_2) \]  
(2.4)

\[ x_2 = \frac{a}{Q} = a + b(p_1 - p_2) \]  
(2.5)

where \( a \equiv \frac{\theta}{\bar{\theta}} \) and \( b \equiv \frac{1}{\bar{\theta}} \).

Note that this model exhibits pure vertical differentiation if \( \theta \) and \( \bar{\theta} \) have the same sign. If they are both positive, then good 1 is superior to good 2. However, some consumers—those with high marginal utilities of income—will prefer good 2 if it is sufficiently less expensive than good 1. Conversely good 2 is the intrinsically superior good if \( \theta < \bar{\theta} < 0 \). The model becomes one of pure horizontal differentiation if \( \theta = -\bar{\theta} \). In that case, increases in \( \bar{\theta} - \theta \) may be seen as increases in “transportation costs.”

The model’s implication that market share is a linear function of the price difference allows us to deal with the years with negligible imports by coding them as zero import shares and performing a Tobit estimation. As a first pass, I defined import shares of less than 0.05% as zero. Since the data on steel rail imports begins in 1871, I use the sample 1871-1913 to estimate the import share equation.\(^9\)

\[
impshare(in\%) = 24.64 + 2.25(p_{us} - (p_{uk} + D + tafi))
\]

(4.50) (0.44)

These results indicate that most consumers prefer domestic rails. If the price at the point of purchase were equal only one quarter of the customers would purchase imported rails. However, price has an important effect on demand. Each ten dollar increase in the price of U.S. rails relative to imported rails raises the import share by 22.5 percentage points. These regression coefficients imply \( \bar{\theta} = 33.45 \) and \( \theta = -10.94 \). This means the consumer who most prefers domestic rails is willing to pay no more than $33.45 extra for them and the consumer who most prefers imported rails is willing to pay $10.94 extra.

\(^9\)Standard errors in parentheses
Quantity demanded depends on the import share and total consumption. In order to perform the simulation we need estimates of total steel rail consumption for the period 1867-1870. These were obtained using the predicted values from the import share equation and known quantities of domestic production.

2.4 The Baseline Simulation

To predict what prices and production would have been in the absence of the duty we cannot just look at the predicted values from the regressions because, under alternate trade regimes, the whole path of cumulative output changes. Hence, I examine the effects of the duty using a dynamic—but deterministic—simulation. Thus, the simulations should trace the expected consequences of trade policies assuming that the residuals from the estimated equations were not predictable.

Since Chapter 1 provides a number of possible specifications for the price equations, some selectivity is required. For British prices, I use column (5) of Table 1.2. This constant-returns specification fit the data well and simplifies calculation of the import price. For American prices, I use the reestimated version of column (6) of Table 1.1. These specification choices ensure that the slopes of the two learning curves do not differ much. Without strong prior reasons to expect different slopes in the two countries, one would not want the results to be driven by the relative size of learning elasticities.

Prior to simulating counterfactual regimes, I used the simulation to attempt to reproduce what actually happened under historical duties. Note that the linearity of the share equation coupled with the constant-elasticity cost functions imply that finding the equilibrium each period requires the solution of an equation that is nonlinear in the import share. I use a Newton-based iterative procedure. However, the solution is constrained by the bounds of zero and 100% import shares.

The baseline simulation suggests that the model actually fits the data reasonably well. Figure 2-5B shows that the price equations manage to capture most of the sources of short- and long-run price variation. The simulated import share does not
fit the historical data as tightly. However, the simulated path does exhibit the main rises and declines of the historic import share.

2.5 Simulation of Free Trade

Next, I explore what would have happened if the U.S. had pursued a policy of free trade (see Figure 2-6). The simulations suggest that free trade would have sharply reduced the domestic industry's growth path. Import share remains above 80% until the late 1880s. Although it gradually declines after that, import share under free trade remains above that under protection even in 1913 once all duties had been removed. The expansionary effect that protection had on domestic output did not come costlessly. Increases in domestic learning-by-doing caused the U.S. price under protection to fall below the price that would have occurred under free trade in 1874. However, the average cost of rails to U.S. railroads remained above free trade levels until 1890. The lower average prices under free trade in the 1890s occurred despite import duties of $13.44 for much of that decade. Expenditures (at the point of purchase) on steel rails totalled $1.981 billion (in 1860 dollars) under protection and $1.907 billion under free trade between 1867 and 1913.

2.6 Exogenous or Endogenous Growth?

The simulation method also helps address the issue of the role of changes in input costs in explaining the rise of American steel rail production. Suppose the United States had not had the ore discoveries and improvements in infrastructure that caused its pig iron prices to decline relative to British pig iron prices. Was protection-induced learning enough to dominate the standard determinants of comparative advantage? The answer seems to be no.

I simulated the paths of steel rail prices and the import share under the hypothetical no reductions in the real price of American pig iron relative to English pig iron, i.e. both $p_{ic_{us}}$ and $p_{ip_{uk}}$ are fixed in constant dollars at their 1867 levels. The
Figure 2-5: The Baseline Simulation

A. Historic import share (---) baseline simulation (----).

60
Figure 2-6: The Simulation of Free Trade

A. Import share: Baseline (—) and Free Trade (---) simulation.

B. Weighted Average Steel Rail Price: Baseline (—) and Free Trade (---) simulation.
simulations offer some support for Taussig's argument that reductions in domestic input costs relative to foreign costs accounted for a large part of rise of the American steel industry (see Figure 2-7). Although the import share seems stuck around 40% under constant pig iron prices, the average price of steel rails is not as strongly affected. Railroads simply buy imported rails in much larger amounts.\footnote{Figure 2-3 shows that British real pig iron prices actually rose in absolute terms over the period. This is why importing large amounts of steel rails does not turn out too costly under constant real pig iron prices.}

The role of domestic learning-by-doing in explaining price reductions and output growth, however, far outweighs the role of changing input costs. To demonstrate this, I run a simulation that uses historical values for pig iron prices while holding domestic cumulative output equal to zero throughout. The results show that prices would have risen substantially over the sample, ending up over $200 per ton rather than under $30 per ton (see Figure 2-8). At such prices, imports displace all domestic production.

\section*{2.7 Welfare Consequences of the Steel Rail Tariff}

The simulations of free trade suggest that the duty raised the short- and long-run market share of the domestic steel rail producers. Moreover, it lowered the long run price of domestic steel rails. However, even if we infer from this that the tariff “worked”, we need not conclude that it raised welfare. An import restraint that only encouraged inefficient production and distorted consumer choices would certainly not increase the sum of producer and consumer surplus. Moreover, the mere existence of long-run benefits from protection does not imply that they inevitably outweigh the short-run costs.

\subsection*{2.7.1 General Equilibrium Welfare}

The estimated model contains enough information to calculate the partial equilibrium effects of the tariff. However, there may be important—but unmeasurable—effects on
Figure 2-7: Simulation of Constant Real Pig Iron Prices

A. Import share: Baseline (—) and Fixed Pig Iron Prices (---).

B. Weighted Average Steel Rail Price: Baseline (—) and Fixed Pig Iron Prices (---).
other sectors. In a paper on targeted export promotion Dixit and Grossman (1976) make the point that a subsidy on one sector may end up becoming a tax on another sector. Their model addresses high technology industries and it assumes that each sector uses scientists as a factor of production. The total domestic endowment of scientists is assumed to be fixed. Hence promoting the expansion of one sector will bid up the wages of scientists and cause a contraction of a different high technology sector. In general, partial equilibrium analysis will overestimate the gains from export subsidies wherever sectors use common, yet scarce inputs. In a dynamic model, a tariff that encourages labor to migrate to one infant industry might reduce output, and, as a consequence, learning in another infant industry.

While these effects are important in principle, they do not seem particularly rel-
event in the case of steel rails. One primary input, iron ore, has no use outside of the iron and steel industry. Increased demand for iron ore results in increased efforts to discover new resources and exploit more fully the currently known supplies. On the other hand, the steel industry does not seem large enough to seriously affect total demand for unskilled labor or investment funding. Production experience allowed a given set of factor inputs to produce larger and larger amounts of steel rails. Absent another infant industry whose growth was retarded, these productivity gains are essentially free. It seems far more likely that the promotion of steel rails had a net positive effect on other industries. To the extent that cheap domestic steel rails spurred the expansion of the national rail system, all producers and consumers of transportable goods are better off. Advances in converter and rolling technology benefit all steel users. The late 19th century saw dramatic growth in the construction, shipbuilding, canning, and machine tool industries. Later on, automobiles would become leading consumers of steel. The partial-equilibrium measure of welfare fails to capture the benefits these future industries would receive as a result of lower domestic steel prices.

2.7.2 Consumer Surplus

In 1870 a group of railroad owners wrote to Congress supporting the steel rail duty. In fact, they asked for a higher specific duty than the one Congress actually legislated. Were they crazy, patriotic, or very far-sighted? Although the simulation procedure cannot answer this question, it can at least suggest whether rail "consumers" could have supported the duty out of pure long-run self-interest.

Measuring the impact of protection on consumer welfare poses a problem because rails are not a final good. There are several "consumers": railroad companies, passengers, companies and farmers that transport freight on the railroads, and consumers of those companies' and farmers' products. The model abstracts from these issues and instead imagines that rails are a small enough share of total income that the marginal utility of income is approximately 1. This implies that net utility from rail consumption can be obtained by subtracting out the amount paid. I make the further

65
assumption that consumers have a large enough valuation for steel rails that they will consume the same aggregate amount for any prices in the relevant range.

To obtain aggregate consumer surplus one integrates equation (2.1).

\[
CS = \int_{\theta}^{\hat{\theta}} u(\theta, 2)f(\theta)d\theta + \int_{\theta}^{\hat{\theta}} u(\theta, 1)f(\theta)d\theta \\
= (v - p_1)q_1 + (v - p_2)q_2 + \int_{\theta}^{\hat{\theta}} \theta f(\theta)d\theta \\
= (v - p_1)q_1 + (v - p_2)q_2 + \frac{(q_1 + q_2)}{2(\hat{\theta} - \theta)} \left( \hat{\theta}^2 - (p_1 - p_2)^2 \right) .
\]  

(2.6)

### 2.7.3 Profits

Domestic steel rail profits are revenues minus variable costs minus total fixed costs:

\[
\pi_t = p_t q_t - C_1(q_{us}, E_1) - n_t fc
\]  

(2.7)

where \( n_t \) is the number of firms and \( fc \) represents the average fixed costs per firm. Since I do not have a measure of total fixed costs, I assume that neither the number of domestic firms nor average fixed costs per firm are affected by trade policy.

On the other hand, if entry occurs freely, profits will be driven to zero and welfare calculations will depend on changes in consumer surplus and duty revenues only. Both approximations seem extreme; hence the constant-total-fixed-costs assumption probably provides an upper bound on the increase in profits caused by protection.\(^{11}\)

Producer surplus under competitive pricing corresponds to rents that accrued to the most efficient suppliers such as Carnegie Steel. As mentioned in Chapter 1, with constant static elasticity of cost with respect to output of 1.174, profits are a constant 15% of domestic revenues. This corresponds well to profit margins during the early years of the industry; however, it may be an underestimate for the twentieth century.\(^{12}\)

\(^{11}\)A potential exception is the case where free trade delays or thwarts all domestic entry because firms cannot borrow to cover their initial losses. In this case endogeneity of \( n \) could result in an underestimate of the additional profits caused by protection.

\(^{12}\)Livesay provides some fragmentary evidence that prices far exceeded Carnegie's costs in some years. Allen (1981) provides some production cost data culled from the census of manufactures. I calculated an approximation for the Lerner index using average variable costs instead of marginal
2.7.4 Simulation Results

Each period the total change in welfare caused by import protection consists of the sum of the protection's effect on consumer welfare, domestic profits, $\Delta \pi_t$, and government welfare. The change in government welfare caused by protection is usually just the revenue raised by the duty.

$$\Delta W_t = \Delta CS_t + \Delta \pi_t + D_t Q_t$$  \hspace{1cm} (2.8)

Figure 2.9A plots $\Delta W_t$ as a solid line and $\Delta CS_t$ as a dashed line. Welfare losses from protection remained fairly small until the rail boom of the early 1880s. During this period U.S. pig iron prices soared relative to the prices in Britain. Under free trade this would have led to large purchases of the cheaper imported rails. However, protection prevented such efficient substitution. As a result, large losses in consumer welfare occurred. By the 1990s increases in domestic experience resulted in such low domestic steel rail prices that consumers were better off as a result of protection despite the fact that the duty was not fully removed until 1913.

The net effect of the duty on welfare depends on whether the final gains, discounted back to 1867 outweigh initial consumer losses. Define $V_t$ as the present value of the discounted gains (losses, if negative) from free trade accumulated from 1867 ($t = 0$) to period $t$.

$$V_t = \sum_{r=0}^{t} \frac{\Delta W_r}{(1 + r)^r}$$  \hspace{1cm} (2.9)

A benevolent government with a time horizon of $T$ years will prefer the protectionist policies actually implemented to free trade if and only if $V_T < 0$. In computing $V_T$, I deflate all changes in consumer surplus, profits, and duty revenue to 1860 dollars. I use the rational expectations assumption to estimate the real interest rate, $r$. Assuming that there are no systematic biases in inflation prediction errors, the real interest rate costs. This index took the values 0.15, 0.12, 0.27 and 0.28 in 1879, 1889, 1902 and 1910.
Figure 2-9: Welfare Consequences of Protection

A. Annual Differences in welfare (—) and consumer surplus (---) between historical protection and free trade.

B. Running Discounted Sum of welfare (—) and consumer surplus (---) differences.
should equal the average ex post real interest rate. i.e.

\[ r = \frac{\sum_{t=1}^{T} i_t - \frac{cpi_t - cpi_{t-1}}{cpi_{t-1}}}{T}. \]

Using railroad bond rates (irate) as the nominal interest rate, I obtain \( r = 0.060 \). This real interest rate seems high by twentieth century standards so I made the present value calculation using other real interest rates as well.\(^{13}\)

Figure 2-9B plots \( V_t \) for a real interest rate of 6\%. The duty leads to welfare gains of close to $30 million 1860 dollars discounted back to 1867. Welfare gains still accrue for interest rates on 12\% although they are smaller.

The evidence on consumer surplus suggests that gains in the form of lower domestic prices after 1890 were probably too small and too late to yield a present value gain for rail consumers. Hence, we should probably seek an alternative explanation for railroaders' early support for protection. As Temin (1964) notes, pervasive financial ties existed between the early steelmakers and the railroads. Hence, railroad owners may have added domestic steel profits to their own in the private welfare calculations that led to their calls for a high steel rail duty.

### 2.8 Optimal Path of Protection

Welfare (\( W \)) is measured as the sum of consumer surplus (\( CS \)), profits (\( \pi_1 \)), and the government's fiscal surplus (\( GS \)).

\[
CS = U(q_1, q_2) - (p_1 q_1 + p_2 q_2) \tag{2.10}
\]

\[
\pi_1 = p_1 (q_1 + q_3) - C_1(q_1 + q_3, E_1) \tag{2.11}
\]

\[
GS = (1 + \kappa) D q_2 \tag{2.12}
\]

\(^{13}\)For example, Stanley Fischer (1977) notes that the average real rate of return on treasury bills between 1954 and 1973 was 1\% and the average real yield on long-term corporate debt for a sample of 16 large firms was less than 2\% during the same period. The high real interest rate that I obtain occurs mainly as result of the sharp deflation following the Civil War.
$U()$ is a subutility function. $q_1$ and $q_3$ are domestically produced steel rails sold at home and abroad. $q_2$ are steel rail imports.

The specification of the government surplus requires some explanation. As in the Laffont-Tirile models of optimal regulation, I assume that government taxation causes distortionary losses to welfare of $\kappa$ per dollar of revenue raised. This seems reasonable given that virtually all government revenue in the nineteenth century came from customs duties and from a few internal excise taxes.\textsuperscript{14} Hence, revenue from the tariff can be used to reduce revenue requirements for taxes on other goods.\textsuperscript{15}

I assume competitive import pricing with constant returns. Hence,

$$p_2 = mc_2 + D + taf_i \quad \text{where} \quad mc_2 \equiv \frac{\partial C_2(q_{uk}, E_2)}{\partial q_{uk}}.$$

Note that $taf_i$ represents transatlantic freight and insurance costs. I make the additional assumption of constant returns in foreign production, i.e. $\frac{\partial^2 C_2(q_{uk}, E_2)}{\partial q_{uk}^2} = 0$. This assumption implies that the duty has only a direct effect on pricing. Adding imperfect competition or rising marginal costs would create secondary effects of the duty. Increases in the duty would cause the foreign price to decline and the duty inclusive price would not rise by the full valued of the duty. Such effects complicate interpretation of the results while distracting from the primary purposes of the duty: protection and revenue.

I also assume marginal cost pricing by domestic firms.

$$p_1 = mc_1 \quad \text{where} \quad mc_1 \equiv \frac{\partial C_1(q_{us}, E_1)}{\partial q_{us}}.$$

This assumption allows me to use the parameters from the domestic pricing equations (estimated in Chapter 1) in the cost functions. It seems reasonable in that the early period of protection featured the entry of over a dozen firms. Furthermore, the price of domestic steel rails declined sharply as cumulative output increased a fact that is hard

\textsuperscript{14}Congress introduced the corporate and personal income taxes in 1909 and 1913
\textsuperscript{15}In the optimal commodity taxation problem $\kappa$ is the multiplier on the government’s balanced budget constraint.
to square with monopoly pricing (see appendix to Chapter 1). After incorporating the pricing assumptions, welfare in a single period can be rewritten as

\[ W = U(q_1, q_2) + mc_1 q_3 + (\kappa D - mc_2 - taf_i)q_2 - C_1(q_1 + q_3, E_1). \]  (2.13)

Note that we can express \( q_1 \) and \( q_2 \) as functions of the duty and domestic and foreign cumulative output. Making the appropriate substitutions provides an expression of \( W \) as a function of the same three arguments.

\( E_1 \) and \( E_2 \) evolve according to the following equations of motion:

\[ \dot{E}_1 = q_1 + q_3 \]
\[ \dot{E}_2 = q_2 + q_4 \]

\( q_4 \) represents all the production of steel by British firms that is not destined for sale as rails in the United States. I treat both \( q_3 \) and \( q_4 \) as exogenously determined. While this may seem less realistic in the case of exports \( (q_3) \), their relatively small volume throughout most of the sample means that we probably do lose much by not endogenizing exports.

Forming the Hamiltonian, we have \( \mathcal{H} = W + \mu_1 q_1 + \mu_2 q_2 \). \( E_1 \) and \( E_2 \) are the state variables and the control variable is the specific duty, \( D \). The first-order conditions yield

\[ \frac{\partial \mathcal{H}}{\partial D} = \frac{\partial U}{\partial D} + (\kappa D - mc_2 - taf_i + \mu_2) \frac{\partial q_2}{\partial D} + \left( \frac{\partial mc_1}{\partial q_1} q_3 - mc_1 + \mu_1 \right) \frac{\partial q_1}{\partial D} + \kappa q_2 = 0 \]  (2.14)

\[ \frac{\partial \mathcal{H}}{\partial E_1} = \frac{\partial mc_1}{\partial E_1} q_3 - \frac{\partial C_1}{\partial E_1} = -\dot{\mu}_1 \]  (2.15)

\[ \frac{\partial \mathcal{H}}{\partial E_2} = -\frac{\partial mc_2}{\partial E_2} q_2 = -\dot{\mu}_2 \]  (2.16)

\[ ^{16} \text{Although Chapter 1 found no evidence of it, monopoly pricing seems more likely after the formation of U.S. Steel in 1901. This would effect the optimal duty if it resulted in a restriction of output during that period which would tend to lower the shadow value of experience in earlier periods.} \]
Integrating (2.15) and (2.16) and imposing $\mu_{1T} = \mu_{2T} = 0$ results in

$$
\mu_{1t} = -\int_t^T \left( \frac{\partial C_1(q_{us\tau}, E_{1\tau})}{\partial E_1} - \frac{\partial^2 C_1(q_{us\tau}, E_{1\tau})}{\partial E_1 \partial q_{us}} q_{3\tau} \right) d\tau
$$

(2.17)

$$
\mu_{2t} = -\int_t^T \frac{\partial^2 C_2(q_{uk\tau}, E_{2\tau})}{\partial E_2 \partial q_{uk}} q_{2\tau} d\tau.
$$

(2.18)

Now, given competitive pricing of both goods, $\frac{\partial q_1}{\partial D} = \frac{\partial q_2}{\partial p_2}$. Taken together with the fact that $\frac{\partial U}{\partial q_i} = p_i$ this implies

$$
\frac{\partial U}{\partial D} = p_1 \frac{\partial q_1}{\partial p_2} + p_2 \frac{\partial q_2}{\partial p_2}
$$

After some substitution and manipulation of (2.14), I obtain

$$
\frac{\partial H}{\partial D} = \left( \mu_1 + \frac{\partial m c_1}{\partial q_1} q_3 \right) \frac{\partial q_1}{\partial p_2} + (D + \mu_2) \frac{\partial q_2}{\partial p_2} + \kappa (q_2 + D \frac{\partial q_2}{\partial p_2}) = 0
$$

(2.19)

Examining (2.19) shows that a rise in the duty has several effects on welfare. On the positive side, it raises domestic output and hence lowers future domestic costs. Such cost reductions have a shadow value of $\mu_1$. On the other hand, the higher duty hurts consumers by raising the current and future prices of imports. Finally, the increased duty affects tariff revenue which the government values for $\kappa > 0$.

To obtain the optimal duty, I now make use the information from the demand specification. In section 2.3 I obtained the following results: $\frac{\partial q_1}{\partial p_2} = -\frac{\partial q_2}{\partial p_2} = \frac{Q}{\delta - \beta}$ and $q_2 = \frac{Q(\beta - \delta - \beta)}{\delta - \beta}$. Substitution of these expressions and the pricing equations into (2.19) leads to the following solution for the optimal path for import duties.

$$
D_t^* = \frac{1}{1 + 2\kappa} \left( \mu_{1t} - \mu_{2t} + \frac{\partial m c_{1t}}{\partial q_1} q_{3t} \right) + \frac{\kappa}{1 + 2\kappa} (-\theta + m c_{1t} - m c_{2t} - t a f i_t)
$$

(2.20)

In the absence of exports and a revenue motive for protection, i.e. if $q_{3t}$ and $\kappa$ equal zero, the duty should be set equal to the differences in the shadow values of domestic and foreign output.
Note that the shadow values $\mu_1$ and $\mu_2$ only count experience-based cost reductions that benefit domestic consumers and firms. The duty that maximizes world welfare could be larger or smaller. By increasing future costs to British consumers and non-American purchasers of British exports, the duty poses a negative externality. On the other hand, by lowering future costs to foreign purchasers of American steel rails, the duty has a positive externality. Since American mineral resources made it the long-run low cost steel producer, it is not inconceivable that the optimal duty from the world's point of view might have exceeded the optimal one from the U.S. point of view.

2.9 Choice of Instruments

A frequent criticism of infant-industry protection argues that tariffs are an inefficient policy instrument. Since the problem is suboptimal domestic production, the solution should address the market failure directly. The government should use production subsidies to raise the private return to domestic production until it equals the social return. The basic drawback of duties is that they raise prices in addition to stimulating domestic production. Hence, they distort consumer choices in an undesirable manner. In apparent disregard for these admonitions, I will solve for the optimal import duty, not production subsidy, for steel rails. Several considerations lie beyond this decision.

First, in the absence of lump-sum taxes, the revenue needs of subsidies induce a variety of distortions in other sectors of the economy. As a practical matter, the 19th century America and many less developed countries today rely on import tariffs as major sources of revenue. Although protective tariffs are often set at prohibitive levels and consequently raised little or no revenue, they do not require any government outlays and corresponding tariff increases on other goods.

Second, given the demand model employed in this chapter, subsidies do not lead to more efficient decisions than duties. The reason is that under competition, output depends only on the price difference, not the price levels. For given domestic and for-
eign marginal costs, the optimal level of domestic production may be reached through a specific subsidy or tariff. Abstracting from government revenue considerations, all that matters is that the sum of the domestic production subsidy and the import duty equal the difference between the domestic and foreign shadow values of experience.

Finally, policy instruments differ in their political economy consequences. Tariffs benefit producers and impose costs on consumers. Subsidies benefit producers and consumers at the expense of the general taxpayer. If the protected industry produces an intermediate input or capital good, its consumers may have sufficient political clout to influence policy formulation. In that case as the shadow value of experience declines consumers will pressure for lower tariffs. They would have no similar incentive to ask for reductions in production subsidies. Although tax-payers have the incentive to reduce subsidies, their incentive is not proportional to the shadow value of experience. Since subsidies for individual industries are probably small relative to the total tax burden, tax payers will care only about the aggregate level of subsidies to all industries. This incentive structure implies that selective intervention may be much more feasible when the government employs import duties.

2.10 Computation of the Optimal Duty Path

Section 2.8 solved for the optimal duty in a continuous time framework without discounting. While these omissions simplify the presentation, they are not appropriate for the actual calculation of the optimal steel rail tariff. Fortunately, adding discounting and switching to discrete time require only a minor modifications in the definition of the shadow values of experience, $\mu_1$ and $\mu_2$.\(^{17}\)

\[
\mu_{1t} = - \sum_{\tau = t+1}^{T} \delta^{\tau-t} \left( \frac{\partial C_1(q_{us,\tau}, E_{1\tau})}{\partial E_1} - \frac{\partial^2 C_1(q_{us,\tau}, E_{1\tau})}{\partial E_1 \partial q_{us,\tau}} q_{3\tau} \right) \tag{2.21}
\]

\[
\mu_{2t} = - \sum_{\tau = t+1}^{T} \delta^{\tau-t} \frac{\partial^2 C_2(q_{uk,\tau}, E_{2\tau})}{\partial E_2 \partial q_{uk,\tau}} q_{2\tau}. \tag{2.22}
\]

\(^{17}\)For the derivation of $\mu$ in a very similar problem, see the Appendix to Chapter 1.
The discount factor, $\delta$, relates to the real interest rate $r$ according to the formula $\delta = \frac{1}{1+r}$.

Figure 2-10A shows the optimal duty expressed in ad valorem terms for a variety of different interest rates. In each case the optimal tariff starts substantially higher than historic tariffs (the solid line) but rapidly declines. The excess protection dating from the 1870s may be interpreted as evidence of political power of the steel industry. By the time duties finally declined, they had ceased to have any important impact on the domestic industry.

Figure 2-10B compares the import shares associated with the various optimal duties with the import shares under historical protection and free trade. Under optimal duties import share would have been driven down more quickly than under historic protection. However, the near prohibitive duties would give way to levels that allowed for modest import shares, especially during the 1880s rail construction boom. It is interesting to note that the present value of consumer surplus is higher under optimal duties than it would have been under free trade. Hence if railroads had exerted more political power they would have had the incentive to shorten the life of the duty, not eliminate it altogether.

Perhaps the excess protection occurred not because of political pressure from rail producers, but rather because of the governments desire for revenues. To investigate this hypothesis I calculated the optimal path of duties for $\kappa = 0.5$. Under this assumption, the government values a dollar's worth of funds at $\$1.50$ because it reduces the need for distortionary taxes elsewhere. Figure 2-11 shows that the revenue motive can explain part of the high ad valorem duties during the 1880s but still the duties do not differ too much from the optimal ones in the absence of revenue motivations.

The simulations reported in Figure 2-12 turn to the issue of how the shifts in relative ore availabilities affects optimal policy. Somewhat counter-intuitively it turns out that under fixed real pig iron prices the optimal duty is generally larger than under historic trends. However, as shown in Figure 2-12B, the implied optimal import share is still much higher under constant pig iron prices despite the higher duties. Since
A. *Ad Valorem* Duty: Historic (—) and optimal for interest rates of 0% (---), 6% (···), and 12% (---). Between historical protection and free trade.

B. Import Share: Historic (···), free trade (also ···) and optimal for interest rates of 0% (···), 6% (—), and 12% (---).
Figure 2-11: Optimal Protection with a Revenue Motive

*Ad Valorem* Duty: Historic (—) and optimal for \( \kappa = 0.0 \) (---), and \( \kappa = 0.5 \) (···).

railroads desire to consume domestic rails, it is welfare improving to lower their costs. Absent protection and favorable trends in relative input costs much less U.S. learning would occur. The optimal duty helps to counteract that but it does recognize that the optimal size of the domestic industry depends on relative pig iron prices.

### 2.11 Conclusions

The traditional case for free trade facilitated trade policy decisions by providing a blanket denouncement of all import restraints. Theoretical work by Krugman and others suggested a set of conditions where tariffs *might* improve welfare. This paper presents an example from American history showing that those conditions may ex-
Figure 2-12: Optimal Protection and Comparative Advantage

A. Optimal *Ad Valorem* Duty: Historic (—) and fixed real (---) pig iron prices.

B. Optimal Import Share: Historic (—) and fixed real (---) pig iron prices.
ist in some industries. This means that trade policy formulators will require more information in order to reach a verdict on proposed import restraints.

Was the steel rail duty a successful case of infant industry protection? On the one hand, the domestic industry did "grow up" and the duty was eventually removed. Hence, protection certainly did not cause stagnation and gross inefficiencies. Furthermore, the duty led to permanent increases in domestic production and total steel rail consumption. While historic duties hurt railroad builders, the overall effect on welfare appears to have been positive. Under optimal duties, the gain would have been larger and it would have been shared by consumers.

As a final note, I would like to emphasize that the experience of the U.S. steel rail industry provides no support for the indiscriminant use of protection practiced by countries pursuing import-substitution development strategies. The steel rail tariff promoted domestic production at little long-run expense to rail consumers as a result of several important factors that distinguish this case from many others where protection has failed miserably. First of all, the basic resource base in the United States gave the American industry a latent comparative advantage. Second, the American market was large and growing. Between 1867 and 1912 total rail mileage in the U.S. rose from 39,250 miles to 246,777 miles compared to an increase from 12,319 miles to 20,038 miles in Great Britain during the same period. The size of the U.S. market allowed for plant level scale economies to be realized while many domestic firms could coexist providing both competition and a continued flow of technological innovations for each other. The market growth facilitated the embodiment of new ideas in new plants and equipment. Finally, on a more speculative note, the political power of the railroads may have constituted a credible threat that high levels of protection would not remain in place even if the industry failed to become competitive.
Chapter 3

Learning Curves: A Closer Look

3.1 Introduction

Virtually all empirical work on learning-by-doing—including Chapter 1—stipulate that some measure of productivity depends on some measure of production experience in a linear-in-logs manner:

\[ \ln(c_t) = \gamma - \lambda \ln(E_t) \]

In this case unit costs, \( c_t \), serve as the productivity measure while experience, \( E_t \) generally corresponds to cumulative output. Ghemawat (1985) Argote and Eppe (1990) present histograms summarizing the learning-curve estimates for hundreds of products. This comparability across results for so many different products arises because all of the studies estimate essentially the same linear-in-logs regression.

The concept of learning-by-doing by itself only asserts the sign of the experience effect—not the specific functional form that it should take. Nevertheless, no paper would omit the linear-in-logs formulation and few even experiment with alternatives. The ubiquitous use of this form suggests that it provides a reasonably good fit for data from a variety of settings. However, the constant-elasticity learning curve possesses two characteristics that should make us leery of using it to formulate strategy for businesses and governments. First, it suggests a mechanical and automatic relationship between experience and productivity. Second, it appears to allow for perpetual
productivity improvements. Acceptance of the first leaves firms and governments surprised and dismayed when industrial plants accumulate large amounts of experience but few productivity improvements. Acceptance of the second characteristic may lead to long-run planning policies that rely too much on scale as an instrument for progress and too little on the introduction of new products and processes. An improved view of learning-by-doing would include a description of factors that determine the magnitude of learning-based cost reductions as well as a better idea of how quickly firms will exhaust such opportunities.

The first part of this paper compares the performance of the power function formulation with specifications that have the same general form but allow the learning elasticity to decline over time. The second part presents several variations of a model of the underlying process that generates learning curves. In these formulations, learning is not the automatic consequence of production experience. Moreover, the models allow for endogenously determined bounds to learning. Nevertheless, they still predict a relationship between costs and cumulative output that closely resembles the classic learning curve.

3.2 The Power Law

The notion of learning-by-doing has become virtually inseparable from the functional form by which production experience is thought to affect efficiency. It appears axiomatic that the log of some efficiency measure—labors hours per unit of output, unit costs, or total factor productivity—has a negative linear relationship with the log of cumulative output.

Economists, engineers, and management consultants have presented evidence of a linear-in-logs relationship between productivity or unit costs and measures of experience such as cumulative output. Learning curves have been reported for products as varied as aircraft, the Model T Ford, memory chips, chemicals, and power plant construction. Some analysts went on to assert that learning curves shared a common slope as well as functional form. In particular it was claimed that each time cumula-
tive output doubles, unit costs fall to 80% of their previous levels. While Argote and Epple (1990) did find that the modal slope (or progress ratio) was 81 to 82%, their survey found large numbers of products with slopes in the range between 70% and 90%. This corresponds to an elasticity range of 0.15 to 0.51.¹

Psychologists have found the same phenomenon—with the same functional form—in experimental settings. In 1936, the same year Wright documented the existence of learning-by-doing in aircraft production, the psychologist J.M. Blackburn reported his experimental findings on the learning curves for card-sorting, canceling 'E's in nonsense French words, adding digits, and solving pencil mazes. Researchers later discovered that the log of the time taken to complete many tasks was a linear function of the log of the number of trials. In a survey of the research on practice, Newell and Rosenbloom (1986) wrote, “There is a ubiquitous regularity underlying human practice, referred to as the power law of practice.” “The law states that when human performance is measured in terms of the time needed to perform a task, it improves as a power-law function of the number of times the task has been performed... This law was originally recognized in the domain of motor skills, but it has recently become clear that it holds over a much wider range of human tasks [including] tasks involving perceptual-motor skills, perception, motor behavior, elementary decisions, memory, routine cognitive skills, and problem solving.” Examples include the time it takes to read inverted text, detect a target, or supply a geometric proof. Crossman (1959) found that the log of the time the operator of a cigar-making machine took to make a cigar declined in a linear manner with the log of the number of cigars produced for over 2 years and 3 million cigars.

3.3 The Limits of Learning-by-Doing

The phenomenon of learning-by-doing has become a staple in the literature on endogenous growth and trade patterns. Unlike the empirical micro literature, the growth models have made use of several different functional forms. Lucas (1987) and Quah

¹The slope or progress ratio equals $2^\lambda$ where $\lambda$ is the elasticity estimate.
and Rauch (1990) assume constant returns from learning. Krugman (1987) assumes a constant elasticity. Young's models (1990, 1991) allow the initial effects of learning to be proportional to cumulative output but eventually the learning curve reaches a kink and becomes flat. These functional form choices have important implications for the growth models. As with the case of accumulation of physical capital, diminishing returns from the accumulation of production experience may eliminate endogenous growth.

Lucas's specification intentionally violates the diminishing returns to learning that he acknowledges is found at the product level. He makes this choice because his goal is a model in which human capital accumulation, his specification of experience, can serve as an engine of growth.

What I want [this specification] to 'stand for,' then, is an environment in which new goods are continually being introduced, with diminishing returns to learning on each of them separately, and with human capital specialized to old goods being 'inherited' in some way by new goods. In other words, one would like to consider the inheritance of human capital within 'families' of goods as well as within families of people.

One might maintain that since Lucas is not making an empirical claim that learning is linear, the evidence that learning opportunities for are effectively finite has no relevance. However, if the only way out of the diminishing returns trap is the introduction of new products then one cannot accurately call learning-by-doing the engine of growth. If invention and adoption of products invented elsewhere provide the crucial role in growth, then that role should be modeled explicitly rather than left in the background.

Alwyn Young (1991) provides such a model which illustrates the important interactions between invention and learning-by-doing. In his model small economies may have insufficient incentives for invention and therefore find their growth grind to a halt as they reach the physical limitations on efficiency in the established set of products. On the other hand, in large economies, products may be invented that,
due to insufficient learning, are so expensive that consumers do not demand them.

The Young model assumes free entry in the invention area so potential inventors care only about the rents generated by new products or processes. In models with few firms, invention may have a replacement effect whereby it undermines existing profits. In this case firms may push down the learning curve and become uninterested in further invention. At the firm level this may have occurred with Ford's concentration on cost reductions in the production of Model T cars. General Motors innovated and overtook Ford. Abernathy and Wayne (1974) use Ford as an example of a company whose excessive reliance on the learning curve eventually caused it to be surpassed by a more innovative rival.²

These considerations suggest that it would be useful to know whether real learning curves have definite bounds and how rapidly industries and firms exhaust the opportunities for cost-reductions based solely on production experience. The power function formulation contains no boundaries and, by definition, it precludes even a gradual decline in the learning elasticity. The following sections propose two formulations that allow the learning elasticity to decline as experience accumulates. I apply the methods to the case of learning-by-doing in steel rail production in the 19th century.

3.3.1 Quadratic Returns to Experience

The exponential function offers a simple alternative to the power function which, when generalized to include a second-order term, generates a region where the elasticity of learning declines. This quadratic formulation for the rate of return on experience states that $c = \exp(-\alpha E + \beta E^2)$. Calculation of the elasticity with respect to experience yields $\alpha E - 2\beta E^2$. This implies that in the range $\frac{\alpha}{4\beta} < E$, the elasticity of unit costs or output with respect to experience is decreasing in the amount of experience.

²Another interaction between new product introduction and learning-by-doing occurs when there are no learning spillovers. Then the reward for product introduction is the rents earned as a result of moving down the learning curve first. In this model patents are not required to create private incentives for invention. However, market size will still turn out to be a primary determinant of the innovation rate as will market structure.
However, for $E > \frac{a}{2b}$ accumulated experience raise costs rather than lowering them. Hence, the quadratic form would make sense for such large values of $E$.

Tables 3.1 and 3.2 provide results for comparing the linear-inlogs and the quadratic returns specifications of the learning curve.

In the case of American rails allowing experience to enter quadratically flips the signs of the time trend and current output. Cumulative output enters with the wrong sign and both it and the squared term have statistically insignificant effects. The regressions presented in columns (3) and (4) omit the time trend which was insignificant in both regressions (1) and (2). Now the cumulative output terms enter with the expected signs and statistical significance. Moreover, the negative sign on current output reappears with a larger magnitude and strong statistical significance. This appears to support the existence of large static returns to scale, the opposite of what prior estimates (see Chapter 1) had indicated. Moreover, this specification seems to offer a superior fit as measured by the adjusted $R^2$ and the standard error of the regression.

Quadratic experience enters the British equation in Column (2) of Table 3.2 significantly but with the incorrect signs. In order for there to be diminishing returns to experience, the first term should receive a negative sign and the squared term should enter negatively. However, the reverse was found. The explanation for the unexpected signs may again arise from the inclusion of the time trend. As with the U.S. data, it reverses signs in the presence of the quadratic experience formulation. When I leave the time trend out in columns (3) and (4), the cumulative output terms take the predicted signs. However, the quadratic specification reduces the adjusted $R^2$ from 0.94 to 0.84.

An analysis of the implications of the parameter estimates helps to explain the different U.S. and British results. The quadratic returns model predicts that the accumulation of production experience reduces costs up until the point where cumulative output equals $\frac{a}{2b}$. One hopes that this point will occur outside the sample and hence be irrelevant. Unfortunately, experience begins to raise costs within both the American and the British steel rail data sample. In the American case, the
Table 3.1: The Functional Form of U.S. Learning

<table>
<thead>
<tr>
<th></th>
<th>Dependent Variable: ln(srp_{us})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>constant</td>
<td>2.737</td>
</tr>
<tr>
<td></td>
<td>(6.04)</td>
</tr>
<tr>
<td>trend</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(1.09)</td>
</tr>
<tr>
<td>ln(pic_{us})</td>
<td>0.631</td>
</tr>
<tr>
<td></td>
<td>(4.88)</td>
</tr>
<tr>
<td>ln(q_{us})</td>
<td>0.157</td>
</tr>
<tr>
<td></td>
<td>(1.63)</td>
</tr>
<tr>
<td>ln(1 + E_{us})</td>
<td>-0.248</td>
</tr>
<tr>
<td></td>
<td>(2.94)</td>
</tr>
<tr>
<td>E_{us}</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>E_{us}^2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>R^2</td>
<td>0.954</td>
</tr>
<tr>
<td>S.E.R.</td>
<td>0.125</td>
</tr>
<tr>
<td>D.W.</td>
<td>1.270</td>
</tr>
<tr>
<td>Begin</td>
<td>1867</td>
</tr>
<tr>
<td>End</td>
<td>1913</td>
</tr>
</tbody>
</table>

Note: Absolute values of t-statistics in parentheses.

S.E.R. stands for the standard error of the regression. q_{us} is the sum of domestic production for the home market and U.S. exports. All equations estimated with two-stage least squares due to endogeneity of q_{us}. The instrumental variable list consists of the included exogenous variables as well as the demand shifters, foreign experience and pig iron prices, the exchange rate, and the duty.
quadratic returns model predicts that all experience accumulated after 1903 raises costs. Similarly, British experience begins to raise costs in 1905. Although the timing is similar, the corresponding levels of cumulative output differ by a large amount. In the U.S. cumulative output becomes detrimental after 40 million tons whereas the British data suggests that experience continues to reduce costs until 88 million tons have been produced. In both cases, by 1913 a % increase in cumulative output leads to almost a % increase in unit costs. The point at which the sign of the experience effect changes seems to depend on sample size as well. When I restricted the U.S. sample to 1867-1900, the new parameters indicate that experience becomes harmful after 1896. To prevent the model from grossly overpredicting costs after experience becomes harmful, the U.S. regressions flip the sign of current output. Since current output is high in the later years of the sample, the quadratic returns model can still generate a good fit. The British data, where Chapter 1 found constant returns to scale, could not use current output to correct the implicit errors coming from cumulative output. This may explain the decline in the adjusted $R^2$ that occurs when the exponential form is imposed on the British data.

These findings suggest that quadratic returns provides a poor alternative to the constant-elasticity learning curve. It yields results that may fit the data reasonably well but nonetheless have nonsensical predictions regarding the effects of production experience on costs. Alternate methods for allowing for declining learning elasticities seem desirable.

### 3.3.2 The Kinked Learning Curve

A kink in the the learning curve offers a potentially superior method for introducing the empirical possibility of bounded learning. Suppose that the elasticity of learning begins at $\lambda$ for cumulative output levels $E_t < \bar{E}$ and then drops to $\lambda'$ for $E_t > \bar{E}$. The bounded learning hypothesis asserts that some $\bar{E}$ is the maximum amount of useful experience. Hence, we should expect $\lambda' = 0$. I am not aware of any studies of learning-by-doing that employ this specification. Conway and Schultz (1959) and Baloff (1966) present a number of scatterplots relating the efficiency measures to
Table 3.2: The Functional Form of British Learning

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dependent Variable: ln(srpu)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>constant</strong></td>
<td>3.156</td>
<td>2.569</td>
<td>2.113</td>
<td>0.777</td>
</tr>
<tr>
<td></td>
<td>(12.63)</td>
<td>(12.16)</td>
<td>(9.83)</td>
<td>(2.80)</td>
</tr>
<tr>
<td><strong>trend</strong></td>
<td>0.017</td>
<td>-0.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.58)</td>
<td>(11.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(pipuk)</td>
<td>0.864</td>
<td>0.723</td>
<td>1.006</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>(15.17)</td>
<td>(10.83)</td>
<td>(15.22)</td>
<td>(11.18)</td>
</tr>
<tr>
<td>ln(1 + Eu)</td>
<td>-0.233</td>
<td>-0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(10.83)</td>
<td>(13.00)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euk</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.912e-05</td>
<td>-1.122e-05</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.74)</td>
<td>(4.78)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eu^2</td>
<td>-7.188e-11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.71)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R^2</td>
<td>0.961</td>
<td>0.959</td>
<td>0.935</td>
<td>0.838</td>
</tr>
<tr>
<td>S.E.R.</td>
<td>0.090</td>
<td>0.092</td>
<td>0.116</td>
<td>0.183</td>
</tr>
<tr>
<td>D.W.</td>
<td>1.841</td>
<td>1.579</td>
<td>1.312</td>
<td>0.760</td>
</tr>
<tr>
<td>Begin</td>
<td>1867</td>
<td>1867</td>
<td>1867</td>
<td>1867</td>
</tr>
<tr>
<td>End</td>
<td>1913</td>
<td>1913</td>
<td>1913</td>
<td>1913</td>
</tr>
</tbody>
</table>

Note: Absolute values of t-statistics in parentheses. 
S.E.R. stands for the standard error of the regression. Estimated using ordinary least squares.
cumulative output. A large portion of the cases they examined seem to exhibit kinks. However, neither paper provided statistical evidence.

Steel rail data provide an interesting opportunity to look for kinks in the learning curve since the product is well-defined, and production without major changes in the process occurred over a very long period. The complete lack of price movement following the formation of U.S. Steel makes this hypothesis difficult to test using the American data. However, the British data offer an opportunity. By 1913, the British had produced 141 million tons of steel. How much of that was redundant production experience? Table 3.3 reports estimates of $\lambda$ and $\lambda'$ for various values of $\bar{E}$. The general regression takes the following form:

$$\ln(srp_{uk}) = \beta_0 + \beta_1 \text{trend} + \beta_2 \ln(pip_{uk}) - \lambda X_1 - \lambda' X_2$$

where

$$X_{1t} = \begin{cases} 
\ln(1 + E_t) & \text{for } t \leq t' \\
\ln(1 + \bar{E}) & \text{for } t > t' 
\end{cases}$$

and

$$X_{2t} = \begin{cases} 
0.0 & \text{for } t \leq t' \\
\ln(1 + E_t) - \ln(1 + \bar{E}) & \text{for } t > t' 
\end{cases}$$

Table 3.3 presents regressions for the entire sample where U.K. price data is available and for a restricted sample starting in 1867. The data in the first 5 years may be less trustworthy for two reasons. First, the price information was pieced together from several sources. Second, British output data begins in 1868. The output data prior to that was generated via a constant-growth-rate interpolation using the data after 1868 and the fact that British commercial production of steel was negligible prior to 1859. Fortunately both samples produce similar results. Furthermore, the first 5 years do not appear to be outliers in the scatterplot presented in Figure 3-1.

The results in Table 3.3 suggest that around 1888, after the British had produced

---

3 The price of steel rails remained fixed at $28.00 for 12 years, despite fairly large swings in the price of pig iron. This created the large variation in the ratio of steel rail to pig iron prices seen in Figure 3-1.
25 million tons of steel that the learning curve elasticity becomes statistically indistinguishable from zero. However, this should not persuade us to accept the kinked learning curve specification. The last column for each sample contains F statistics for the hypothesis that \( \lambda = \lambda' \). The critical values for the 5% F statistic are 4.05 for the 1862–1913 regressions and 4.07 for the 1867–1913 ones. The 1% critical values are 7.20 and 7.27 respectively. In the 1867–1913 sample one cannot reject \( \lambda = \lambda' \) for any value of \( \bar{E} \). When the entire sample is included, we may reject the hypothesis for only one value of \( \bar{E} \) and then the rejection barely occurs at the 5% level. From the mid-1880s onward the data provide no compelling statistical evidence for or against the bounded-learning hypothesis.

Table 3.3: Is Learning Bounded?

<table>
<thead>
<tr>
<th>( \bar{E} )</th>
<th>( t' )</th>
<th>( t'/T )</th>
<th>Sample: 1862–1913</th>
<th>Sample: 1867–1913</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{E} )</td>
<td>( t' )</td>
<td>( t'/T )</td>
<td>( \lambda )</td>
<td>( \lambda' )</td>
</tr>
<tr>
<td>1</td>
<td>1872</td>
<td>%21</td>
<td>0.2128 (16.11)</td>
<td>0.3129 (5.95)</td>
</tr>
<tr>
<td>10</td>
<td>1881</td>
<td>%38</td>
<td>0.2152 (9.60)</td>
<td>0.2437 (2.09)</td>
</tr>
<tr>
<td>25</td>
<td>1888</td>
<td>%52</td>
<td>0.1729 (7.89)</td>
<td>-0.0929 (0.64)</td>
</tr>
<tr>
<td>50</td>
<td>1896</td>
<td>%67</td>
<td>0.1932 (12.06)</td>
<td>-0.0028 (0.02)</td>
</tr>
<tr>
<td>100</td>
<td>1906</td>
<td>%87</td>
<td>0.2095 (14.49)</td>
<td>0.1822 (0.73)</td>
</tr>
</tbody>
</table>

Note: T-statistics in parentheses. \("t'/T"\) indicates portion of sample where \( E_t < \bar{E} \). The entire sample extends from 1862 where \( E = 1500 \) tons or 1867 where \( E = 54,800 \) to 1913 where \( E = 141 \) million tons.

The scatterplot evidence in Figure 3-1 helps to explain the result that for the last half of the sample we cannot discern statistically whether learning continues at the
initial elasticity or whether learning opportunities have been completely exhausted. Figure 3-1 deflates steel rail prices by the major input cost, pig iron, and shows how they relate to cumulative output in the U.S. and Britain. The linearity when both variables are expressed in natural logs indicates a constant elasticity of learning. One may draw a line through the data starting from any of the first years and it will fit then entire data set fairly well. However, starting at one of the later years, flat lines will also fit. The bunching of data for large values of experience suggests that it will be difficult to discern at what point, if any, learning opportunities completely disappear. However, there does seem to be strong evidence of continued learning-by-doing through 1881.

Figure 3-1: Learning-by-Doing in Steel Rail Production (U.S. = o, U.K. = *)

While the case of steel rails may not persuade us that learning curves level off completely after a certain amount the accumulation of a certain amount of experience,
it does support the idea that cumulative output gradually loses its importance as a source of cost reductions. Consider the simple learning curve $c_t = a_t E_t^{-\lambda}$. Taking natural logs provides the approximate percentage change in costs attributable to changes in $a$ and $E$.

\[
\ln(c_{t+1}/c_t) = \ln(a_{t+1}/a_t) - \lambda \ln(E_{t+1}/E_t).
\]

The second term can be rewritten as $\lambda \ln(1 + q_t/E_t)$.

The constant elasticity framework leads to effectively bounded learning if $q_t/E_t$ goes to zero as $t$ gets large. Let $s_t$ be the expenditure share on a product and $I_t$ equal total income. Hence, $p_t = s_t I_t$. Add the assumptions of a single factor of production with input price $w_t$ and marginal cost pricing: $p_t = w_t E_t^{-\lambda}$. Combining the assumptions, $q_t = \frac{s_t L_t}{w_t E_t^{1-\lambda}}$. Given a single factor of production with initial level $L_0$ and growth rate $\eta$, $I_t = w_t L_0 e^{\eta t}$. Taken together, these assumptions yield

\[
q_t/E_t = s_t L_0 e^{\eta t} E_t^{-\lambda (1-\lambda)}
\]

Since $s_t$ may be constant (as in the case of Cobb-Douglas utility) and certainly has an upper bound of 1, this expression implies that $\frac{q_t}{E_t}$ will converge to a limiting value of $\frac{\eta}{1-\lambda}$. Only with zero resource growth will constant-elasticity learning eventually cease to affect productivity. However, with normal rates of growth, the contribution of learning-by-doing will ultimately become quite small. Plugging this result back into the expression for cost reductions attributable to learning with population growth rate ($\eta$) of 2% and learning elasticity ($\lambda$) of 0.25, we find that learning-by-doing ultimately leads to annual cost reductions of 0.66% per year.

Figure 3-2 plots $\lambda \ln((1+E_{t+1})/(1+E_t))$ for the U.S. and Britain for learning curves with elasticities of $\lambda = 0.15$ and $\lambda = 0.30$ using the historical values of cumulative output. In each case, accumulation of experience generates annual cost reductions over 10% initially, but gradually diminishes despite the fact that British and American
output grew throughout the sample.\textsuperscript{4}

Figure 3-2: Percentage Cost Reduction Attributable to Learning-by-Doing (Solid line indicates domestic industry. Dashed line indicates foreign.)

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure32a.png}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure32b.png}
\end{figure}

\subsection{3.4 Models of the Learning Process}

Although learning-by-doing has played a major role in models of growth, infant-industry protection, and business strategy, economists have devoted almost no attention to understanding the process itself. By specifying the underlying process, testable predictions regarding the shape of learning curves may emerge. Moreover,

\textsuperscript{4}The spikes in the domestic line are caused by production booms that temporarily raise the contribution of learning to cost reduction. These booms in U.S. rail consumption had little effect on the British learning pace since they sold only a small amount of their steel output in the United States.
the nature of the process may have implications for business or government policies.

Psychologists have proposed a variety of models to explain why skills depend on practice according to the power law. Some models stress that practice helps to codify knowledge. Others assert that practice saves time because it causes an increasing percentage of the sub-tasks involved in a skill to become automatic and hence not require conscious thought. These concepts do not have obvious analogies in industrial settings. However, the psychologist Crossman (1959) has proposed a model of learning-by-doing that could, in principle, apply to economic situations. It models learning as an adaptive process in which the learner experiments with multiple techniques and gradually selects the superior ones. Crossman assumes the existence of a finite number of possible methods, denoted $M_i$. Each method "occurs" with probability $p_i$ and has a completion time associated with it of $t_i$. While the characteristics of the methods do not change, the probabilities of choosing particular methods depend on the number of trials. After each trial the probability of using the last-tried technique will rise or decline depending on whether it took less or more time than the expectation generated using initial probabilities. Thus the model has an evolutionary flavor; techniques found to be inferior are gradually phased out. At a psychological level, Crossman's model seems to assume imperfect recall. Bad methods may be used by accident, but the likelihood of that declines as an individual's experiments with superior techniques "reinforce" them, making subsequent use more likely.

Crossman shows that cost reductions in his model are proportional to the variance of the methods in each period. As more weight is placed on the successful methods, the variance will decline and productivity will plateau. Although this model predicts learning curves of the standard shape, it seems quite bizarre from an economic point of view. If the learner was aware of all the possible techniques then $p_i = 1$ for the method with the lowest time $t_i$ and $p_j = 0$ for all other techniques. Inferior techniques should be dropped immediately, not slowly phased out in accordance with some parameter measuring the strength of the selective process.

Although skill acquisition exhibits the same power function relationship as indus-
trial learning curves one should distinguish learning-by-doing that involves increases in speed at replicating a set of well-understood steps and learning that involves conscious discovery. Skills may improve without the agent’s awareness of any changes in method. For instance the time I take to type the previous paragraph without error will be a decreasing function of the number of times I type it. However, I will not consciously learn anything new about the process of typing. On the other hand, if I throw sand into the steel furnace and discover that this calms the molten steel and reduces the likelihood of explosions, “practice” has generated new knowledge.

The distinction seems important in two respects. First, one would expect pure increases in motor skills to reach physical limitations fairly quickly, whereas increases in knowledge may continue for a much longer duration. Second, discoveries may be communicated. Recipe books are essentially the communication of knowledge generated from experience. However, there is little I could tell someone about typing the previous paragraph that would decrease the time he or she required to type it.

Hence, economic theories of learning-by-doing should probably focus on learning as conscious discovery. To qualify as learning-by-doing, such discoveries should occur while the firm is engaged in actual production rather than activities such as research and development which may be thought of as “learning-by-studying”.

Grossman, Kihlstrom and Mirman (1977) proposed a model of experience-based information generation that they assert relates to learning-by-doing. In this model the learner acquires information about a parameter of interest through experimentation. They consider two examples. In the first, a consumer varies his consumption of a drug in order to assess its efficacy. In the second, a monopolist varies output levels to determine the slope of the demand curve. Although they do not provide any examples of firms becoming more efficient at producing a good, one possibility would be a firm that does not know the productivity of some input. By using that input in large amounts, the firm would update it’s beliefs and ultimately choose an input mix closer to the full-information optimum.

---

5Conway and Schultz (1959) studied improvements in manufacturing in a number of plants and concluded that “operator learning in the true sense of performance of a fixed task is of negligible importance in most manufacturing processes.”
The advantage of the Grossman et al model is that it builds a complete model of learning. However, since they do not explicitly address learning-by-doing in a production setting, their model makes no functional form predictions regarding the relationship between experience and efficiency. A more fundamental criticism is that one cannot accurately characterize learning-by-doing in firms as the generation of information regarding some parameter. In the cases where the improvements associated with learning-by-doing have been studied, one cannot reduce them to learning about the optimal setting for some continuous variable.

For instance, Chapter 1 provides a number of cases where experimentation led to minor, but concrete and self-contained, modifications in the steel production process which had a large cumulative effect on costs.

Henry Ford’s description of a series of modifications made in the assembly of the fly-wheel magneto provides another example of an actual learning process that one should have in mind while constructing a stylized model.

We had previously assembled the fly-wheel magneto in the usual method. With one workman doing a complete job he could turn out from thirty-five to forty pieces in a nine-hour day, or about twenty minutes to an assembly. What he did alone was then spread into twenty-nine operations; that cut down the assembly time to thirteen minutes, ten seconds. Then we raised the height of the line eight inches—this was in 1914—and cut the time down to seven minutes. Further experimenting with the speed that the work should move at cut the time down to five minutes. In short the result is this: by the aid of scientific study one man is now able to do somewhat more than four did only a comparatively few years ago. *Henry Ford.*

Finally, Holmander’s (1965) study of Rayon cited examples of minor technical improvements including changes in the cellulose content of the wood pulp and increasing the number of nozzles per spinning machine.

In each of these cases learning-by-doing involved the discovery that some change in the way things were done would lower costs. However, once that change had been
made, future cost reductions generally came from altogether different modifications
of existing methods rather than continued calibration of a single choice variable.

In summary, economic models of learning-by-doing should possess the following
characteristics:

1. Improvements arise from new knowledge.

2. Improvements comprise a large set of diverse, unrelated modifications to existing
   practices.

3. A necessary, but not necessarily sufficient, condition for the generation of ideas
   for improvements is the act of production itself.

The rest of the paper is devoted to the development of a model that meets these con-
ditions while also being consistent with the empirical finding that unit costs depend
on cumulative output in a linear-in-logs manner.

3.5 Learning-by-Doing as Costless Search

The fundamental assumption in the model advanced here is that a unit of produc-
tion experience generates an idea. After producing a batch of output using current
methods, the chief engineer conceives an alternate method. The idea may or may not
constitute an improvement over existing production techniques. I make the following
assumptions regarding ideas for new methods:

An infinite variety of potential techniques exist. The unit costs associated with
these techniques are distributed between $c$ and $\bar{c}$ according to unchanging cumulative
and marginal distributions $F(c)$ and $f(c)$.

The simplest case occurs if the engineer is able to assess the costs associated with
a new idea at the moment of conception. Actually this assumption is stronger than
necessary. All the engineer needs to know is whether the production costs associated
with the new technique are lower than the costs of the old technique. Under this
assumption, in the absence of costs of generating or implementing ideas, the firm will
always adopt a new technique if its costs are less than current costs. Hence, the costs of
producing the \( n \)th batch will be the costs associated with the most efficient technique from the set comprising all previously used techniques and the idea currently under consideration.

The definition of a "batch" will depend on the specific production context. For example, in Bessemer steel production a batch is the cycle of filling the converter with molten pig iron, blasting it with cold air, allowing the reaction to take place, and the pouring the molten steel into molds. In the case of autos or aircraft, a batch would be the complete construction of a single car or plane. Other cases will be more difficult to define. For instance, what is a batch of cloth? In those cases, a batch might be measured by some unit of time actually spent on productive activities, e.g. an 8-hour shift. The central concept is that entire "batch" must be completed to generate an idea. This suggests that cumulative output may need to be measured in different ways for different products. In some cases, cumulative labor-hours of input will be more appropriate than cumulative units of output.

Another concept in the model that remains vague by intention is originator of the idea. The simplest way to imagine it is that the chief engineer or production foreman generates ideas while the rest of the workers simply perform their duties. However the model is also consistent with the notion that the group of workers jointly propose the new technique. The key is that one batch generates one and only one idea. Otherwise, the firm would be able to devote more people to each batch and to obtain more ideas and this would look much more like a research lab than it would like learning-by-doing. However, it is possible that multiple teams working simulatanecously on different batches each produce ideas. The model assumes sequential production of batches in an effort to maintain simplicity.

Expected costs of the \( n \)th batch will be the expected minimum of \( n \) draws from the distribution:

\[
E[c_n] = E[\min\{c_0, c_1, ..., c_{n-1}\}]
\]

Define \( F_n() \) and \( f_n() \) as the cumulative and marginal distributions of the minimum
of a sample of size $n$. They are derived as follows.

$$F_n(c_n) = Pr(c < c_n) = 1 - Pr(c_0 \geq c_n, \ldots, c_{n-1} \geq c_n) = 1 - (1 - F(c_n))^n$$

$$f_n(c_n) = F'_n(c_n) = n(1 - F(c_n))^{n-1} f(c_n)$$

Substitution of the formula $f_n(c_n)$ yields the expectation of minimum of a sample of $n$ draws.

$$E[c_n] = \int_\xi^\overline{c} cf_n(c)dc = \int_\xi^\overline{c} cn(1 - F(c))^{n-1} f(c)dc$$

(3.1)

Closed form expressions for the expected minimum exist for at least two distributions.

The first case is where costs are uniformly distributed between $c$ and $\overline{c}$. Simple integration yields $E[\overline{c} - c_n] = \frac{n(\overline{c} - c)}{n+1}$. Hence,

$$E[c_n] = \overline{c} - \frac{n(\overline{c} - c)}{n+1} = \frac{\overline{c} + nc}{n+1}.$$  

This function shares several features with the conventional wisdom regarding learning curves. First, accumulation of experience lowers costs. Second, there are diminishing returns, i.e. the second derivative with respect to $n$ is positive. Third, costs tend towards a lower bound as production experience becomes very large. Finally, if $c$ is zero and $n$ is large, $E[c_n]$ is approximately a constant elasticity function of experience. Unfortunately, the elasticity of approximately unity lies far above the range normally found in empirical estimations of learning curves.

Since negative costs do not make sense, we may restrict $c$ to be greater than or equal to zero. In that case we can apply the standard statistical result that

---

For the uniform distribution, costs are a decreasing convex function of experience.

$$\frac{\partial E[c_n]}{\partial n} = - \frac{(\overline{c} - c)}{(n+1)^2} < 0,$$

$$\frac{\partial^2 E[c_n]}{\partial n^2} = 2 \frac{(\overline{c} - c)}{(n+1)^3} > 0.$$  

(3.2)
E[x] = \int_0^\infty (1 - F(x))dx for non-negative x. Noting that \(1 - M(c_n) = (1 - F(c))^n\) we obtain the following expression for the expectation of the sample minimum.

\[
E[c_n] = \int_0^\infty (1 - F(c))^n dc
\]  

(3.3)

Using Equation (3.3), it becomes easy to find expected minimum costs under the exponential distribution \(F(x) = 1 - \exp(-\eta x)\). The expectation of minimum costs is just \(\frac{1}{\eta n}\). Note that once again expected minimum costs are a declining convex function of the number of draws. Moreover, as with the uniform, there is a unit elasticity relationship between expected costs and cumulative output.

Examination of the derivatives of (3.3) shows why such diverse distributions as the uniform and exponential yield similar formulations for the expected sample minimum.

\[
\frac{\partial^i E[c_n]}{\partial n^i} = \int_0^\infty (1 - F(c))^n (\ln(1 - F(c)))^i dc
\]  

(3.4)

Compare this to the derivatives of the “classic” learning curve:

\[
\frac{\partial^i c}{\partial n^i} = (-1)^i n^{-\eta - i} \prod_{j=1}^i (\eta + i - 1)
\]  

(3.5)

Both specifications of the learning curve imply negative odd derivatives and even positive derivatives. Although we do not have strong prior beliefs regarding higher derivatives, it is reassuring to find that the technique-sampling model generates diminishing returns to experience for all continuous non-negative distributions.

Although the uniform and exponential distributions generate linear-in-logs learning curves, they fail to deliver learning curves capable of taking on the range of elasticities found in virtually all empirical studies. The intuitive explanation for this failure lies in the large amounts of density that both distributions contain near their lower bounds on costs. Observed learning curves are characterized by continual steady progress towards lower costs. However, this seems unlikely in distributions where it is relatively easy to happen on costs very close to the distribution’s lower bound. To obtain learning curves with sensible elasticities we require distributions with thinner
lower tails.

One possibility is the Weibull distribution, of which the exponential is a special form. The cumulative density for the Weibull is $1 - \exp(-x/\beta)$. Hence, application of equation (3.3) yields

$$E[c_n] = \int_0^\infty \exp(-n(c/\beta) ) dc. \quad (3.6)$$

Figure 3-3 plots this expression on normal and log-log scales for the first 10,000 draws. The Weibull appears to generate learning curves with the right shapes and slopes.

Integration of Equation (3.6) by parts and then by substitution yields the following result.\(^7\)

$$E[c_n] = \frac{\beta}{\alpha} \Gamma(1/\alpha)n^{-1/\alpha}. \quad (3.7)$$

Taking logs, we discover see that the Weibull distribution delivers the classic learning-curve formulation!

$$\ln(E[c_n]) = \gamma - \lambda \ln(n) \quad (3.8)$$

where $\gamma \equiv \ln(\frac{\beta}{\alpha} \Gamma(1/\alpha))$ and $\lambda \equiv -1/\alpha$.

On the first draw ($n = 1$) expected costs are just the first moment of the Weibull distribution, i.e. $\frac{\beta}{\alpha} \Gamma(1/\alpha)$. As the sample size—which measures production experience—increases, expected costs decline with a constant elasticity. Values of $\alpha$ in the range from 3 to 5 correspond to the range of learning curve elasticities most often found in empirical estimations. As shown in Figure 3-4, increasing $\alpha$ while holding $\beta$ constant has the effect of narrowing the distribution and therefore reducing the likelihood that a new idea will yield a cost reduction. Hence, the return to experience—expressed as the learning elasticity—declines (see Figure 3-3). An industry with a steep learning curve is one where the unit costs associated with various methods do not cluster tightly.

---

\(^7\)The steps are laid out in the appendix to this chapter. Euler’s Gamma function is defined as $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$. 

101
Figure 3-3: Expectation of the minimum of $n$ draws from a Weibull distribution (plotted in normal and log-log scales)
3.6 Occam's Razor

The learning-as-search model can generate learning curves with the same shape and slopes as those found in empirical work. A major difference seems to be that the model makes the learning process stochastic. However, researchers routinely make the classic learning curve model stochastic when they add an error term. If $c_t = e^{\gamma} n_t^{-\lambda} e^{\epsilon_t}$ then the estimated model will be

$$\ln(c_t) = \gamma - \lambda \ln(n_t) + \epsilon_t.$$  \hspace{1cm} (3.9)

Occam's razor suggests that theories should not not "multiply entities unnecessarily." A complex model that merely replicates results from a simpler one serves little purpose. However, the learning-as-search model differs from (3.9) in at least two important ways. First, it introduces technological uncertainty as a primitive rather than an afterthought and this has important implications for planning. Second, it provides a framework to introduce a number of factors that may have an effect on the magnitude or duration of learning-by-doing. The first point is considered below while the final sections of the chapter illustrate the second point.

In the standard learning curve illustrated in equation (3.9) all production experi-
ence is equal in the sense of providing the same contribution to cost reduction. In the learning-as-search model much experience is fruitless. Consider the expected costs of the next batch of production.

\[ E[c_{n+1}|c_n] = c_n(1 - F(c_n)) + E[c|c < c_n]F(c_n) \]  \hspace{1cm} (3.10)

This equation shows that, given costs of the preceding batch, the amount of experience accumulated conveys no additional information useful for predicting future costs. The exact opposite is true under the classic learning curve formulation. The practical importance of this is that while both conceptualizations of learning-by-doing predict the same relationship between the accumulation of experience and cost reductions, the model presented here suggests that, \textit{ex post}, much experience will have proved utterly unproductive. Such experience should be disregarded in the firm's predictions of future costs.

One of the empirical puzzles posed by learning curves has been the large variation in learning rates across organizations or organizational units producing the same product. Argote and Epple (1990) cite the evidence from Liberty ships during World War II where "productivity gains varied more within shipbuilding production programs than between production programs." They also present learning-curve plots from three truck plants that were part of the same company. The plots reveal large differences in the pace of learning and the level at which productivity gains seemed to flatten out. Argote and Epple point out that these cases seem to contradict the conventional specification of learning curves. However, the learning-as-search model is quite consistent with varying progress rates \textit{provided} that barriers to communication between plants exist. Otherwise, at any point in time costs at all three plants would be identical and equal to the cost associated with the most efficient technique yet discovered by any of the plants.
3.7 Learning-by-Doing without the Free Lunch

Although the "luck of the draw" may explain some of the variation in learning-by-doing in seemingly similar situations, we should also examine factors that relate to actions the firm may take which affect the learning process. If such actions are costly, then variations in their costs and benefits will provide another avenue for explaining variations in learning curves.

One way that learning may be accelerated is by expanding production above the short-run profit maximizing level. This will tend to cause cost reductions that ultimately raise profits. Such an "investment" in over-production will be more or less profitable depending on the firm’s discount rate, its ability to appropriate the cost reductions, and the strategic consequences of observed cost-reduction. Using the conventional model of the learning curve, Spence (1981) and Fudenberg and Tirole (1983) have studied the optimal path of output in the presence of learning-by-doing. For that purpose, the conventional learning curve formulation seems entirely appropriate. The models in this paper abstract from the decision of how much to produce each period and focus instead on actions the firm must take to generate and implement new ideas associated with each individual batch.

3.7.1 Effort and Learning

The pace of learning probably depends on efforts to discover new techniques suggested by production experience. Bell et. al. (1984) comment in their review of the technological progress made by various protected infant industries that "productivity growth in infant industries appears to be highly variable, even in apparently similar economic conditions." They argue that a large part of these differences can be explained by variation in the level of efforts made to generate a "flow of information and understanding [from production in one period] that allows improvements in a subsequent period." The authors conclude that "Costless learning-by-doing thus has limited relevance for the accumulation of technological capability."

Suppose that ideas do not flow automatically from the production process. In-
Instead, each batch must be studied to see how things work and what might improve the process. Assume that "technological effort" and production experience are required in fixed proportions to generate an idea. This provides a middle case between typical models of learning-by-doing where "practice" alone perfects the production process and R&D where efforts to improve technology occur independently from production.

Both anecdotal and statistical evidence exist to support the notion that technological effort may complement experience in generating progress. Hollander (1965) found that minor technical improvements collectively accounted for the majority of cost reductions at the Rayon plants he studied. He commented that many of the changes depended on the efforts of Technical Assistance Groups at the plant sites. He went on to note that "the work of the Engineering Department depended less upon its function as a generator of technology based on new formal research and more on its function as a source of consultation and advice for the plants." Lieberman (1984) provided some corroborative statistical evidence. In his estimations of learning curves for chemical products, Lieberman found that the cumulative output caused larger cost reductions for products manufactured by firms with large research and development expenditures.

I will assume that after each production batch, the engineer in charge may conceive a new technique by expending effort in amount \( e \). Firms will implement new techniques that have costs less than current costs and ignore the rest. Hence, as before, the unit cost of each batch will be the minimum of the costs of all previous batches. However, the firm may eventually decide that the effort costs required for learning-by-doing outweigh the expected benefits.

Assume that the firm has an infinite horizon. Denote the value function for a firm that experiments as \( V \) and for a firm that sticks with the same technique forever as \( W(c) \). Let \( \pi(c) \) be profits as a function of unit costs and \( \delta \) the discount factor. \( W(c) \), the value of the firm if never experiments again, is just an infinite sum: 
\[
\frac{\pi(c)}{1-\delta}
\]
To calculate the value of the firm when it experiments, one considers three different cases. For most values of \( e \) and \( f(c) \) there will exist a positive cost level, \( c^* \), below which the firm will not find it profitable to continue to expend \( e \) in order to search for
new techniques. There is also a cost range between current costs, $c_0$, and $c^*$ where the firm adopts the new technique and continues to search. Finally, for cost realizations above current costs, the firm remains at the same position as it started with value $V(c_0)$. These three possibilities and the probabilities of each are displayed in the value function below.

\[
V(c_0) = \pi(c_0) - e + \frac{\delta}{1 - \delta} F(c^*) E[\pi(c)|c < c^*] + \delta(F(c_0) - F(c^*))E[V(c')|c^* < c' < c_0] \\
+ \delta(1 - F(c_0))V(c_0)
\] (3.11)

To obtain the expression defining $c^*$, we need to find the costs that make the firm indifferent between experimenting again and keeping the same costs forever, i.e. we need to solve $V(c^*) = W(c^*)$.

\[
V(c^*) = \pi(c^*) - e + \frac{\delta}{1 - \delta} (F(c^*)E[\pi(c)|c < c^*] + (1 - F(c^*))V(c^*)). 
\] (3.12)

Noting that $V(c^*) = W(c^*) = \frac{\pi(c^*)}{1 - \delta}$, I obtain

\[
V(c^*) - W(c^*) = \frac{\delta}{1 - \delta} F(c^*)E[\pi(c) - \pi(c^*)|c < c^*] - e = 0. 
\] (3.13)

To make further progress, I assume unit demand for prices less than or equal to $v$ and zero demand otherwise. Now Equation (3.13) reduces to

\[
\frac{\delta}{1 - \delta} F(c^*)((v - E[c|c < c^*]) - (v - c)) = e.
\]

Assume that costs are normally distributed with mean $\mu$ and variance $\sigma^2$. A normal distribution, with the mean a sufficient number of standard deviations from zero will closely approximate the Weibull distribution which generates an exact linear-in-logs relationship between expected costs and experience. The normal distribution has the advantage that it generates simple, intuitive expressions for the conditional
expectations of unit costs.

\[ E[c|c < c^*] = \mu - \frac{\sigma \phi(x)}{\Phi(x)} \]

where \( \phi() \) and \( \Phi() \) are the marginal and cumulative distributions of the standard normal and \( x \equiv \frac{c^* - \mu}{\sigma} \). Substitution and simple manipulation yields

\[ c^* = \frac{e}{\Phi} \left( \frac{1}{\delta} - 1 \right) + \mu - \sigma \frac{\phi}{\Phi} \] (3.14)

Ignoring for a moment that \( \Phi \) and \( \phi \) are functions of \( c^* \), the above equation reveals several determinants of the optimal cost at which to terminate search. The higher the effort/search cost, \( e \), the higher will be \( c^* \). Increases in the variance not only raise the learning elasticity as argued in the discussion of the Weibull distribution, they also cause learning efforts to cease at a later point. Finally, the more patient the firm is, the lower will be \( c^* \).8

I conducted numerical simulations to investigate the effect of including effort costs. In addition, the simulations explore the degree that data generated by the learning process match the analytical expectations calculated for the Weibull distribution.

Using a pseudo-random number generator I produced draws from a normally distributed technology with mean cost of 100 and varying standard deviations. I generated 200 different “histories”, in each of which the firm produces a total of 100 production batches. The firm samples new techniques each period for which initial

---

8Market structure and information flow may affect the shape of the learning curve as well. Continue the assumption of unit demand. Now suppose that the price lies just below the cost of the second most efficient firm. Since the price cancels in the expected gain from effort, the optimal condition for \( c^* \) does not change. However, now it becomes possible that an unlucky firm will exit the industry due to negative expected profits. The firm’s effort decision does change if the technique it discovers may be “leaked” to other firms with probability \( \nu \) each period. The firm’s probability of still being in sole possession of the new technique after \( t \) periods is \( (1 - \nu)^t \). Thus the expected discounted gain if a superior technique is found is \( \frac{\delta}{1 - \delta(1 - \nu)} E[c^* - c|c < c^*] \). This implies a simple change in the optimal stopping rule.

\[ c^* = \frac{e}{\Phi} \left( \frac{1}{\delta} + \nu - 1 \right) + \mu - \sigma \frac{\phi}{\Phi} \]

Since \( c^* \) is an increasing function of the \( \nu \), anything that makes “leaks” less likely—for instance patents or worker loyalty—will cause the learning curve to flatten out at a lower cost level.
costs exceed $c^*$ where $c^*$ is defined by Equation 3.14. Since Equation 3.14 is nonlinear in $c^*$, I solve for it using an iterative method. For an effort cost $e = 10$ and a mean of $\mu = 100$ standard error for the technology of $\sigma = 25$, the optimal stopping cost is 42.

Final observed costs each period equal the sum of the technical cost rating and an additional zero mean normally distributed disturbance. This error term, to which I assigned a variance of 4 is designed to capture the various non-technological determinants of the costs economists actually observe. It gives the simulated data a more realistic appearance than the one that would occur if costs moved down in a strict, stair-like manner.

After generating the cost paths for each history, I estimate a simple linear-in-logs learning-curve relationship:

$$\ln(c_t) = \gamma - \lambda \log(1 + E_t) + \epsilon_t$$

where $c_t$ is unit cost and $E_t$ is cumulative batches prior to the current one. The histogram in Figure 3-5 depicts the distribution of the estimated elasticities for 200 different histories. In each case draws came from the same underlying technology ($N[\mu = 100, \sigma = 25]$). However the estimates vary widely. The mean elasticity was 0.23 which corresponds to the mean for 96 different products summarized in Ghemawat (1985). The average learning elasticity seems to be roughly proportional to the standard deviation of the technology. When the standard deviation of the technology declines to 20, the mean elasticity declines to 0.16. A further reduction to $\sigma = 15$ shrinks the mean learning elasticity to 0.10. On the other hand, raising the standard deviation to 30, causes the learning elasticity to rise to 0.375. However, it generates a large number of outliers since the high standard deviation makes large downward cost jumps more likely. The median value of $\lambda$ was 0.28. The standard error was a very high 0.29.

The linear-in-logs relationship fits the data reasonably well. The average $R^2$ of the regression is 61%. However, the quality of the fit varies drastically across histories. Figure 3-6 presents two cases at opposite ends of the spectrum. Part A presents
Figure 3-5: The Distribution of Learning Elasticities. (Estimated from simulated data.)

![Histogram of Learning Curve Elasticity](image)

A history that matches our notions of a standard learning curve, except that costs decline in a somewhat choppy manner. Costs never reach the critical level $c^*$ at which it becomes optimal to stop trying to find technical improvements. Given the data in Part B, few observers would suspect learning-by-doing. Rather, they would concentrate their analysis on the "drastic innovation" that occurs after the 14th production batch. At that point costs decline by almost 50% to a level below $c^*$ (the horizontal dotted line). Thus, if a single task dominates final costs, the same process may generate very different histories.

The model simulated above corresponds to cases where all attention revolves around improving the performance of a single process. Naturally, there will be cases where progress moves steadily and others where it moves haltingly. Most real p...
Figure 3-6: Simulated Learning Relationships

A. A "classic" learning curve.

B. Different Data Generated by the Same Process.
ucts require a number of components or steps in the production process. The costs of producing the product itself will be the sum of the costs of such sub-processes. It seems sensible to presume that learning occurs in each sub-process. If each subprocess is independent, one would expect aggregation to smoothen out the path of costs over time. I conducted simulations to confirm that intuition. I divided the production process into 10 tasks with identical technology distributions. Aggregate costs for 20 different histories follow very similar paths. The average and median learning elasticity was 0.24. The standard deviation of the estimated λs dropped to 0.04 while the $R^2$ jumped to 95%. I plot a representative case (with $\lambda = 0.19$ and $R^2 = 0.9566$) in Figure 3-7. As expected, aggregation smoothenes the observed learning pattern considerably. If most observed unit costs really are aggregates of independent tasks, then the differences in learning elasticities across industries probably arise more from differences in the variance of the underlying technology than from differences in the particular realizations of the learning process.

### 3.7.2 Embodied Progress

The implementation of new ideas may involve additional outlays beyond those required to generate the ideas. This section and the following two develop three different “costs” associated with the introduction of new techniques.

The simplest cost to analyze occurs when new ideas can not be put in place without purchases of new equipment or modifications of existing machinery. Hollander's (1965) study of technological progress at DuPont's rayon factories found that most technical changes “required for their introduction investment in plant and equipment.” Hollander discovered that most outlays were made to replace existing machinery. As in the effort-assisted learning model, embodiment costs will lead the eventual cessation of technical progress. However, embodiment costs also add technological inertia throughout the learning process. This is because new techniques

---

9 However, Hollander comments that technical improvements did not add large costs "because even in the absence of technical change the existing mechanisms required relatively frequent replacements."
now must offer large enough improvements over existing methods to provide present value gains in excess of the investment cost. In addition, embodiment costs generate a somewhat spurious correlation between unit production costs and cumulative investment. Instead of investment itself reducing costs, both cost reductions and investments are being driven by the same force: ideas generated through production experience.

3.7.3 Trial and Error

Up to this point, I have assumed that at the moment the engineer conceives a new technique, she also knows the unit production cost associated with the idea. If the production process is as poorly understood as is implicit in the learning-by-doing
model, new ideas may be no more than groping, tentative attempts to solve a problem. As an example, consider the following quote from a German steel maker, “Today, there is a little bit of nickel in the steel. Sometimes it works, sometimes not. You want to make progress, so you try different things.”

Each batch of production the firm may stick with the technique it has or experiment with a new technique. However the properties of new techniques only become known once the firm has produced a batch using the new method. Sometimes a technique will turn out to be inferior and result in higher costs. This is why the designation “trial-and-error” seems appropriate.

As before $V$ is the value of a firm that experiments and $W(c)$ is its value if it sticks with the same technique forever. $W(c)$ is the present value of current profits discounted out to infinity. Each period the firm experiments it will receive the expected value of profits and then it will be able to decide whether to experiment or stick with the technique whose cost the firm just discovered.

$$V = E[\pi(c)] + \delta E[\max(W(c), V)]$$

where $\pi(c)$ represent per period profits and and $\delta$ is the discount factor.

Since the problem the faces remains the same each period, once it decides not to experiment, the firm will never want to experiment again. Conversely prior to that point it will experiment every period.  

Suppose a firm discovers a technique with cost $c'$. There exists a $c^*$ such that for $c' \leq c^*$, $V < W$, and therefore the firm will terminate experimentation and stick with cost $c'$ for eternity. The probability of such a event is $F(c^*)$ where $F()$ is the cumulative density for the distribution from which the firm draws ideas. Making use of this information we may rewrite the value function as follows.

---

10 *New Yorker* February 25, 1991 p. 70 Preston, Richard “Hot Metal”

11 These stark predictions arise because because the firm learns the costs of a new technique after producing one batch. If there were uncertainty about the costs, then the firm might wish to stick with a technique while it gradually updates it’s beliefs about the unit costs of the new method.
\[ V = E[\pi(c)] + \delta \left( F(c^*) E \left[ \frac{\pi(c)}{1 - \delta} \mid c < c^* \right] + (1 - F(c^*)) V \right) \]

Rearranging, \( V \) becomes

\[
V = \frac{E[\pi(c)] + \frac{\delta}{1 - \delta} F(c^*) E[\pi(c) \mid c < c^*]}{1 - \delta(1 - F(c^*))} \]

\[
= \frac{\int_{\xi} \pi(c)f(c)dc + \frac{\delta}{1 - \delta} F(c^*) \int_{\xi} \pi(c)f(c)dc}{1 - \delta(1 - F(c^*))} \tag{3.15}
\]

For normally distributed techniques and unit demands, I obtain the following expression for \( c^* \).

\[
c^* = \mu - \frac{\delta \sigma \phi}{1 - \delta(1 - \Phi)} \tag{3.16}
\]

where all the parameters are defined as in Equation (3.14). Note that just as in the effort model, high variance of techniques and large discount factors lead to the termination of learning at lower cost levels.

The trial-and-error model predicts that costs will bounce around the mean of the distribution of \( c \) until they happen to reach a level below \( c^* \) and then they will stay at that level forever. This does not correspond to the traditional notion of gradual but continuous cost reductions that eventually become very small. Even after aggregating over several tasks, plants, or firms, it seems unlikely that this sort of behavior would yield cost-experience relationships that look like conventional learning curves. I conclude that this trial and error model, while superficially appealing, does not accurately capture the process underlying learning-by-doing.

### 3.7.4 Screening of New Techniques

The extreme results of the trial and error model arose because the only way the firm could assess the costs of new techniques was to implement them at full scale. This may not be necessary. For instance one might be able to develop some form of prototype to predict costs of new techniques. Similarly, a firm might establish a small-scale pilot project to test a new technique. To take an example from wine
production, the firm might experiment with one barrel of wine each harvest while using tried and true techniques on the remainder.

The case of the Model T provides some anecdotal evidence on the use of small-scale implementation to screen out bad ideas.

We try everything in a little way first—we will rip out anything once we discover a better way, but we have to know absolutely that the new way is going to be better than the old before we do anything drastic. *Henry Ford*\(^{12}\)

One easy way to introduce small-scale experimentation is to stipulate that the firm can build a model to accurately determine the costs of an idea generated by production experience if it is willing to incur a cost of \(m\). This corresponds quite closely to the set-up in which the firm had to expend effort in amount \(e\) to generate an idea. In both cases the firm will wish to spend \(m\) or \(e\) until they exceed the expected gain from experimentation. The difference is that the firm retains the option of generating and testing ideas without making the upfront expenditure if it is willing to implement the new idea at full scale.

Now there are three stop points to consider. As in previous models there are the two points where one form of experimentation is no longer profitable. Denote as \(\bar{c}\) and \(\tilde{c}\), the stopping points for full-scale experimentation and technique-screening. Under normally distributed technology they will be defined according to Equations (3.16) and (3.14) respectively. Denote as \(c^*\) the cost level at which the firm decides to switch from experimenting at full scale to screening new ideas. This switching point will depend on the magnitude of screening costs and where current costs are relative to their mean value: \(c^* = \mu - m\). For \(c > c^*\), the firm will implement at full scale. The intuition is that costs are so high that the firm knows there is a high probability that *any* new idea will constitute an improvement over existing practice. For costs in between \(c^*\) and \(\bar{c}\) the firm will use the minimum cost technique discovered up to that point and test out new ideas by building prototypes. In this region, under a Weibull

\(^{12}\)Quote from p. 39 of *Giant Enterprise* by Alfred D. Chandler, Jr.
distribution, expected costs will evolve according to the power function. Eventually the firm will find a technique with costs below \( \bar{c} \) and it will retain that technique of production thereafter. Thus, this model predicts a relationship between costs and experience which starts out flat, then becomes roughly linear-in-logs, and then finally becomes flat again. This suggests that empirical studies of learning curves should look for cost "plateaus" for products at both their early-infant and mature stages.\textsuperscript{13}

3.8 Conclusions and Extensions

Belief in the existence and importance of learning-by-doing does not imply that one take the classic constant-elasticity formulation literally. It is possible to generate the standard relationship between unit costs and cumulative output with a model of repeated technique sampling. This model provides a framework to analyze the ways that chance, technological effort, and uncertainty affect the contributions of production experience to cost reductions. Linear-in-logs learning curves may fit the data generated by models in which productivity improvements are neither free nor unbounded. As a result, policy makers at the firm and government level should consider policies that shift out the limits to learning when possible and induce new innovations once all opportunites have been exhausted.

The learning-as-search model suggests that the shape of learning curves may vary with market structure because the possibility of information "leaks" will reduce learning incentives. However, for a given number of firms, aggregate costs will tend to fall more quickly if firms get together to share experience-based discoveries. The model also suggests that internal wage-incentive policies may be able to expand the bounds of learning curves. For instance, in the effort model, manager who receive a share of profit increases will be more willing to exert the unobservable effort necessary for the generation of ideas. On the other hand, salaries tied closely to cost outcomes will not work in the trial and error model since risk averse managers would probably be

\textsuperscript{13}Note that the dual kinks of \( \bar{c} \) and \( \bar{\bar{c}} \) will only exist for intermediate values of of the screening cost \( m \). Very high-\( m \) products will exhibit only the two plateaus, while very low-\( m \) products will exhibit the normal learning relationship until the lower plateau is attained.
unwilling to accept the high variability of earnings during the experimentation phase. These statements are quite speculative; however, they do suggest the usefulness of having an underlying model of learning-by-doing rather than a mechanistic formulation where the elasticity of costs with respect to experience enters as some kind of unanalyzed, reduced-form parameter.
Appendix: The Expected Minimum of $n$ Draws from a Weibull Distribution

Let $x$ be distributed according to a Weibull distribution with parameters $\alpha$ and $\beta$. Denote the expected minimum after $n$ draws as $m$. As shown in the text,

$$m = \int_0^\infty e^{-n(x/\beta)^\alpha} \, dx$$

Integration by parts yields

$$e^{-n(x/\beta)^\alpha} x + \int n\alpha(x/\beta)^\alpha e^{-n(x/\beta)^\alpha} \, dx.$$

Evaluating this expression at $(0, \infty)$ causes the first term to disappear and we are left with

$$m = \int_0^\infty n\alpha(x/\beta)^\alpha e^{-n(x/\beta)^\alpha} \, dx.$$

Let $w = n(x/\beta)^\alpha$. Substituting out $x$ and $dx$, one obtains

$$m = \beta n^{-1/\alpha} \int_0^\infty \frac{\beta}{1/\alpha} w^{1/\alpha} e^{-w} \, dw.$$

Noting that the Gamma function is defined as $\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} \, dx$ one can rewrite the above expression as

$$m = \beta n^{-1/\alpha} \Gamma(1 + 1/\alpha).$$

Finally, since it is a property of the Gamma function that $\Gamma(1 + r) = r\Gamma(r)$, one achieves the desired result.

$$m = \frac{\beta}{\alpha} \Gamma(1/\alpha) n^{-1/\alpha}.$$
References


American Iron and Steel Association Annual Statistical Reports

Argote, Linda and Dennis Epple 1990 “Learning Curves in Manufacturing” Science, Vol. 247 920–924


British Iron Trade Association Annual Reports


HOLLANDER, SAMUEL 1965 The Sources of Increased Efficiency: A Study of Du Pont Rayon Plants The M.I.T. Press


KRUGMAN, P.R., 1986, “Market Access and competition in high technology industries: a simulation exercise.” MIT Mimeo

KRUGMAN, P.R., 1987, “The Narrow Moving Band, the Dutch Disease, and the Competitive Consequences of Mrs. Thatcher.” Journal of Development Economics 27 pp. 41-55


LIEBERMAN, MARVIN “Learning by doing and pricing in the Chemical Processing industry.” Rand Journal of Economics


QUAH, DANNY AND JAMES E. RAUCH 1990 “Openness and The Rate of Economic Growth” M.I.T. mimeo


TAUSSIG, F.W., 1915, Some Aspects of the Tariff Question

TEMEN, P., 1964, Iron and Steel in Nineteenth Century America


YOUNG, ALWYN 1989 “Learning by Doing and the Dynamic Effects of International Trade.” mimeo

YOUNG, ALWYN 1991 “Invention and Bounded Learning by Doing” *M.I.T.* mimeo