Order Number 8914767

The hospital service sector and inflation

Jones, Kristine, Ph.D.

City University of New York, 1988

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

THE HOSPITAL SERVICE SECTOR AND INFLATION

by

KRISTINE JONES

A dissertation submitted to the Graduate Faculty in Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1988

This manuscript has been read and accepted for the Graduate Faculty in Economics in satisfaction of the dissertation requirements for the degree of Doctor of Philosophy.

9/27/88
Date

Michael Grossman
Chairman of Examining Committee

9/27/88
Date

Michael Grossman
Executive Officer

Professor Michael Grossman

Professor Charlotte Muller

Professor Salih Neftci

The City University of New York

Abstract

THE HOSPITAL SERVICE SECTOR AND INFLATION

by

Kristine Jones

Advisor: Professor Michael Grossman

The primary aim of this dissertation is to construct a model of the hospital sector capable of testing various hospital inflation theories. The central theme is that hospital inflation may be a consequence of excess demand for hospital services. The initial argument suggests that the nature of medical services, in this case, hospital services, is such that physicians have sufficient influence on consumers to shift the demand schedule in their own behalf. Next, it is argued, that under some conditions, physicians would benefit from a market characterized by excess demand for hospital services. For instance, the existence of excess demand frees the physician from merely treating those patients who can afford the total costs associated with a stay in a hospital. At a price less than market-clearing, physicians may opt to treat those patients that satisfy their particular taste for risk. Finally, the period following a period with excess demand is likely to be experience a price increase.

A secondary aim of the dissertation is to derive a

maximum likelihood function for the proposed hospital sector model. The market model is atypical since it does not contain a market clearing equation, but an equation which equates the observed number of hospital admissions with the minimum of quantity demanded or quantity supplied. A nuance of the model is that it is never known whether the observed number of admissions is corresponding with quantity demanded, quantity supplied or both. As a result, the model can be treated as an endogenous switching model, with unknown switching. Given the model's design, the likelihood function has the form of a bivariate probit search function.

Acknowledgements

I would like to acknowledge my appreciation to Professors Michael Grossman, Charlotte Muller and Salih Neftci for their suggestions, constructive criticisms and editorial assistance during the various stages of development of my dissertation. Their help did much to improve the organization, content and readability of this paper. I would also like to acknowledge my appreciation to Professors Rachael Boaz and Marjorie Honig for guidance and encouragement during the long process of working on the dissertation. Many fellow students such as Nick Barcia, Mike Massourakis, Homar Bonitsis and Ratna Karmakar influenced me during my stay at the Graduate Center, and I would like to thank them for their years of comraderie. And, I can not finish the program without attending my appreciation to Ann Holzman, who was always helpful in many ways. I would like to acknowledge the great deal of emotional support friends and relatives such as Maureen, Mike, Daniel, Beth, Jean, Lorinda, Nancy, Flo and my parents, provided me during the long years of attending the graduate center. Last, but certainly not least, I would like to acknowledge a fact that I often did not acknowledge during the process, and that is Frank, my husband, was absolutely invaluable to me during the extended time it took me to write the dissertation.

TABLE OF CONTENTS

Chapter

I INTRODUCTION.....1-7

II AN OVERVIEW OF THE HOSPITAL SECTOR.....8-31

III CHOICE THEORETICS OF THE HOSPITAL SECTOR.....32-56

IV THE HOSPITAL SECTOR MODEL.....57-98

V DERIVATION OF THE LIKELIHOOD FUNCTION AND OTHER ISSUES.....99-119

IV CONCLUSION.....120-124

APPENDIX A.....125-128
APPENDIX B.....129-137

BIBLIOGRAPHY.....138-141

Chapter 1

Introduction

Hardly a week goes by without a story appearing in our nation's newspapers lamenting the problems associated with the delivery of medical care to citizens. We cry about how much we pay for health care. Annual per capita expenditures are at a record $1,772. We chastise ourselves that health expenditures as a percent of GNP are closing in on 11 percent. We worry about the fact that a growing portion of Americans have no plans for financing health expenditures. We know that most of us are inadequately prepared for the expense associated with a catastrophic illness. And, none of these woes are new; in the last decade and a half, per capital expenditures rose 447 percent. In 1970, health expenditures represented 7.4 percent of GNP, and by 1986, they represented 10.9 percent.

By far the most troublesome component is the hospital sector. Growth in its expenditures has persistently outstripped growth for any other category of medical service.

As a response to our concerns, legislators have reacted: To provide care to our indigent and our elderly, in 1966, the Medicaid and the Medicare programs were implemented. Currently, Massachusetts' government has

legislated ruled that all employers must offer an health insurance program to their employees. In the U.S. Congress, a movement to extend Medicare benefits to cover long-term care is under debate.

But, many would argue such programs were the initial impetus to our surging demand for health care and new programs will just fuel the upward movement in health expenditures. Instead, we should focus on cost containment and mechanisms for financing our expenditures. Again, such policy has been implemented. The hospital component of the Economic Stabilization Program (ESP) of the mid-1970s did keep growth in the cost of a day spent in a hospital to a rate comparable to that of general inflation. But, once the program was removed, hospital costs galloped ahead, even surpassing the growth rates associated with the period prior to the adoption of ESP.

The most recent chapter in cost containment involves a change in the mechanism of payment for hospital care. With the introduction of the Diagnosis Related Group program (DRG) in 1983, hospitals are reimbursed according to a diagnosis group rather than according to services rendered. But average stay and total discharges from hospital care have recently dropped, in 1980, average stay was 7.3 days and discharges per 1000 persons was 168; in 1986, average stay was 6.5 days and discharges per 1000 persons was 148 (NCHS, 1987). But total expenditures on

health services continue to soar.

A basic contention of this paper is if we do not begin to understand how all the components of the health sector are related, we will continue to design policy which may solve one objective, only to have a negative impact on another aspect of good health care delivery.

This paper is an attempt to design a comprehensive model which will include some of the basic characteristics of the hospital sector. A major hypothesis of the paper is that spiraling hospital cost increases can be explained by a market structure which is persistently in a state of excess demand for hospital services. The paper's outline is the following: First, a scenario is developed which gives a rationale for the claim the sector has excess demand. Second, a model which takes into account assumed characteristics of the hospital sector is designed. Third, a likelihood function which allows the basic hypotheses be tested is derived. And, fourth, a plan for testing the model in the future is provided.

The operative definition of hospital service is that it is a collection of medical and hotel services which are provided in a physical setting known as a 'hospital.' While some services are common to all hospitals, there are many more which are not (Robinson and Luft, 1985). Quantity of hospital services is measured exclusively in terms of total admissions into a community

hospital, defined as a non-government hospital whose patients' average stay is less than thirty days (AHA, 1983).

A basic premise behind any market model is that agents are in rational pursuit of some objective. However, it is concurrently assumed that each agent acts independently. While the agents (consumers, hospital administrators and physicians) may be in rational pursuit of their specific objective functions, given certain characteristics of hospital service, it is reasonable to assume that neither the consumer nor the hospital administrator is capable of acting independently from the physician. This model focuses on the likelihood that the consumer is not independent of one of the major decision-makers on the supply side. Specifically, the consumer requires input from the physician in her decision-making process. As a result, variables reflecting the interest of the physician are found in the explanation of both the supply side and the demand side of the market.

Physician-induced demand is what introduces the hospital service market to the possibility of being in persistent disequilibrium. But physician-induced demand in itself would not guarantee that a market will have this characteristic. The additional requirement is that it must be in the self-interest of the physician for the market to be in persistent disequilibrium. If this is not

the case, the physician would use her influence to set the demand equation such that the observed price would equate quantity demanded with quantity supplied.

Another way of stating this is, if the controlling agent in the hospital sector market is the physician, and if the market is characterized by a state of persistent disequilibrium, in this case, of excess demand, then the physician must be benefiting from its existence. Two types of benefits the physician is experiencing in a market with excess demand are suggested. First, the excess demand may be relaying information concerning quality that the average consumer is not able to discern by evaluating price. An individual may not be fully comfortable relying on the physician's advise as to the consequences of a medical procedure for her health, but observing a queue of other individuals waiting for this medical procedure will give her confidence in the physician's advice. Second, in setting price below a market clearing price, the physician may have more control over the type of patient she services. To the extent that physicians are interested, not just in income, but also in rewards like a challenging case load and providing care to a variety of patients, this non-equilibrating price does allow her to pick and choose her patients.

The above scenario is the theme of Chapter 3, a chapter on the choice theoretics of the hospital sector.

Topics included in the chapter are the non-profit nature of hospitals, the relation between physicians and administrators and its impact on supply decisions, consumers' perception of the effect of hospital care on their health, health insurance, the effect of medical information on the relation between physician and consumers, and the benefits physicians would experience in a market with excess demand for hospital services.

The specific market model that will be presented in Chapter 4 consists of six equations, five which are stochastic in nature. The chapter discusses in great depth the specification of each equation. Special note is taken of predicting values on some of the model's parameters which are included to test whether or not credence should be given to the hypotheses that there is physician-induced demand for hospital service and the market is in disequilibrium.

Since estimation using OLS would result in biased and inconsistent estimators, a method of full-information maximum likelihood is derived in Chapter 5. A second necessary topic for any model which is a simultaneous equation mode is that of identification. The necessary and sufficient conditions for the hospital sector model will be discussed in Chapter 5. While a time-series model would be optimal, because of the size of the model and the limited amount of time-series data, the observations are

pooled over the fifty states and the District of Columbia. The final topic of Chapter 5 concerns the issues surrounding the use of a pooled time-series.

The final chapter, Chapter 6, critiques the model and provides a brief strategy for estimating the hospital sector model. In addition, two appendices are included which discuss the construction and sources of information on variables included in the model. The suggested data set for future estimation can be obtained from the American Hospital Association (AHA) the Health Insurance Institute of America (HIIA), the American Medical Association (AMA), Social Security Bulletins, the National Medical Care Expenditures Survey for 1977, and the National Bureau of Economic Research.

Chapter 2

An Overview of the Hospital Sector

Recent trends of the hospital sector include a 22 percent decline in the average hospital stay (from 7.35 days to 5.71 days) during the 1980 through 1985 period, and a decrease in the occupancy rate from 69.7 percent in 1980 to 56.6 percent in 1985. The average patient in 1985 is reported to be sicker and either a Medicaid or Medicare patient. In 1980, privately insured patients represented 52 percent of hospitals' case load; in 1985, they represent 45 percent. The shifts are attributed to such factors as greater acceptance of outpatient treatment, changes in the reimbursement system and more intense competition between hospitals (New York Times, 5/25/88). But, despite such changes, health expenditures rose 70 percent and they represent a average of 10 percent of GNP for the six year period. Concurrently, health insurance premiums paid by Americans averaged $103 million and grew at an average rate of 13 percent per annum.

The focus of this chapter is to examine the general trends of the hospital sector with special emphases on the 1970 through 1982 period. The statistics for this period are different than the more recent experience. Topics will include trends in hospital revenue and costs, number

of admissions and average stay, hospital regulation and private and public health insurance programs.

Characteristics of Hospitals

Except for a brief period in the 1970s when the Economic Stabilization Program (ESP) was in place, growth in the cost of a day in the hospital outstripped growth in the general price level. Only in the last few years has the cost of a hospital stay been less than growth in the cost of total medical care. (See Table 2.1.) But, as pointed out by M. Feldstein (1977), "the usually rapid

> increase in the cost of a day of hospital care reflects a change in the character of hospital services rather than a higher price for an unchanged product. To understand hospital cost inflation ... requires understanding why hospitals now provide a much more sophisticated and expensive service..."

This section is not concerned with understanding why this phenomenon has occurred, but will merely report on general trends.

As noted in the introduction, total medical expenditures as a percent of GNP have increased from 7.4 percent in 1970 to 10.9 percent in 1986. During this period, hospital expenditures went from $26.0 billion to $173.1 billion in 1986, an annual increase of 17 percent. (See Table 2.2.)

But at least during the 1970s and early 1980s, average stay in a community hospital hovered around 7.6

days, while admissions per 1000 persons actually increased during the early 1970s from 140 to an high of 150 in 1982. (See Table 2.3.) Clearly, not what would be predicted. The increased utilization of hospital services could be explained by better hospital insurance coverage or a change in product. An interesting side note is that the variance of average stay between the states increased drastically during the time frame. With respect to better hospital insurance coverage, the percent of hospital expenditures financed by private insurers did increase from 75.2 percent in 1970 to 82.2 percent in 1982.

After the DRG program was implemented in 1983, a slow decline in admissions and average stay did take place. The coincidence of these events means the decline can not be attributed to a pricing mechanism. Instead, a stricter policy about reimbursing hospitals by insurers probably is the initiator of the decline in utilization of hospital services.

Hospital Regulations

Hospital regulations range from the licensing of professional staff, to mandatory review of cases, to regulation of facilities and service growth, to controlling revenue and costs. While the first two categories of regulation are to insure quality of service

of a good which has merit aspects, and have been 'on the books' in some form for all of this century, the second two categories of regulation are more recent impositions on the behavior of hospitals and can be considered responses to the recent trends of rapidly increasing costs of hospital care.

As summarized by Frank A. Sloan and Bruce Steinwald (1980), "the case for hospital regulation is buttressed by

> a general consensus that the hospital market is far from the competitive ideal. The majority of hospitals are nonprofit and therefore may not be as subject to pressures for efficient production as are profit-maximizing firms in competitive industries. Overinsurance also undermines incentives to implement cost-reducing innovations and has probably led to overuse of hospital services. Many hospitals are local monopolists and perhaps monopsonist as well. Other structural characteristics of the industry, including medical-staffing privileges, severely limit patient and physician choice of hospital. Finally, given the 'merit want' aspect of health and hospital care, society may not be satisfied with a market solution, even if the above impediments were not important."

But this chapter is concerned not with the motivation for hospital regulation but with a description of the major regulations instituted during the 1970 through 1982 period. In particular, the following section will focus on the Certificate of Need program (CON), the Section 1122 program (1122), the Economic Stabilization Program (ESP) and prospective reimbursement programs (PR). Following the description of each program is a brief summary of the estimated impact of the program on quantity

and quality of hospital services and changes in hospital charges and input costs.

State legislated CON programs have been required by Federal law (National Health Planning and Resource Development Act) since 1974. Each state must designate an agency whose function is to review plans for expansion or modernization of a hospital. In some states, in addition, the introduction of a new service by a hospital will warrant review. While CON programs are universally required, they vary substantially between states with respect to comprehensiveness (i.e., some review only hospitals; while in other states, hospitals, nursing homes and physician offices are included in mandatory reviews,) dollar threshold on capital expenditures before review becomes required, and type of appeal process. Furthermore, non-compliance by a medical facility does not result in a penalty (Sloan, 1980).

Studies analyzing the impact of the CON programs have concluded that while they result in no aggregate growth in hospital capital, they will induce a redistribution of types of capital. Specifically, CON programs result in a replacement of bed stock expenditures with capital expenditures which improve quality of service per bed. (This finding concurs with Newhouse's conclusion that hospitals have a quality bias (1970)). In addition, hospitals within states which were in the process of

implementing a CON program had significant anticipatory growth in bed stock. The consequence of CON programs on costs per bed, while inconclusive, appear small, if any (Sloan, 1981).

Similar to the CON programs, the objective of the 1122 programs is to regulate hospital services by review. Unlike the CON programs, the 1122 program criteria are uniform across participating states. The program is voluntary. Any hospital which receives revenue from public programs financed by the Social Security Administration and which is located in a participating state must acquire permission for any hospital expenditures which exceed $100,000, any change in bed size or any change in services offered to the community. Obviously, the larger the proportion of hospital revenues from public programs, the greater the likelihood the hospital will allow its plans to be reviewed. Not to allow review is paramount to not receiving public money. While the 1122 program has a better participation rate than the CON programs, its effects on capital stock expenditures and costs per bed are similar (Sloan, 1981).

The ESP program, with respect to the hospital sector, went through a number of metamorphoses during its existence. Initially, from August 1971 through November 1972, controls were placed on hospital wages and prices charged for specific hospital services. Interestingly,

controls were not placed on charges which would be reimbursed by a third-party payer. Hospitals could circumvent controls by changing the quality of existing services or introducing a substitute service whose costs would be reimbursed by a third-party payer. In December 1972, hospitals were limited to a 6 percent limit on their overall increases in costs per patient-day. Hospitals interpreted this control as a limit on quality of service - which was not part of the legislative mandate of the ESP. The American Hospital Association took the Cost of Living Council to court on this issue. Before a ruling was obtained, the hospital component of ESP expired (in 1974), and no ruling on the issue was made. An interesting fact about the history of ESP is that while the rest of the wage and price controls mandated by the ESP were in effect for approximately five quarters - from August 1971 through December 1972, due to the rapid increase of hospital charges, controls on the hospital sector were in place for a significantly longer period of time - through April 1974 (Sloan, 1980; Feldstein, 1977).

The impact of ESP controls are reported to be ambiguous. Ginsberg (1978) concluded that they had a small negative impact on average wages and no affect on other costs. However, Sloan (1981) found the program to affect hospital costs by 2% to 3.3% in the short-run, and up to 10% in the long-run. Both Ginsberg and Sloan found

a disturbing negative effect on hospital profits. While not conclusive, the ESP program, according to Sloan (1981) may have encouraged hospitals to discriminate toward admission of patients with less severe conditions requiring shorter stays for recuperation and/or use of less hospital resources per day.

A prospective reimbursement program will state the rate of reimbursement for hospital services. The rate structure may be determined by a formula or as a result of a proposed budget submitted by the hospital to some agency. Regardless of which is used, once the PR scheme is in place, overruns are only the responsibility of the hospital and can not be automatically passed on to the consumer in the form of a higher charge (Sloan, 1981)

PR programs vary considerably between states. In particular, as delineated by Sloan (1981), PR programs vary by responsibility of administering the program, being mandatory or voluntary, which third-party payers are included, and how long the program has been in effect. With respect to reducing costs, the most effective PR programs are ones which are mandatory, use a formula, are required by third-party payers, and have been in effect for a substantial period of time. In general, the PR programs have the same consequence as the ESP program with one exception. PR programs are less likely to result in a hospital going bankrupt.

The only substantial PR program during the time period of the study was instituted in New Jersey (established in 1980). As a reaction against the adoption of the Tax Equity and Fiscal Responsibility Act (TEFRA), a nation-wide PR program was adopted as of October 1983. Hospitals which receive reimbursement from either Medicaid or Medicare were constrained to charge for their service in accordance to 467 diagnosis-related groups. The charge schedule is based on average costs in nine regions. Since the paper is concerned with the hospital sector prior to 1983, evaluation of the effectiveness of the DRG program will be ignored.

Health Insurance

Actual increases in hospital charges have allowed hospitals to change their product with respect to volume of personnel, high-tech equipment and other inputs per bed-day. But, many have argued that this is just a market response to consumer demand for such an improved bed-day. And the increase in demand may be considered a consequence to the growth in the scope of health insurance coverage. This section looks at various characteristics of health insurance and its growth.

Expenditures on private health insurance coverage in 1982 represented 5.6 percent of GNP. When public programs are included, the percent of GNP allocated to

health insurance premiums becomes 10.5 percent. Average growth of such expenditures from 1970 through 1982 is 37 percent per annum. Of course during this time frame, the number of people with coverage did increase, but still, the per capita health insurance premium went from $161 in 1970, to $819 in 1982, an increase of 409 percent. Since many Americans have no insurance coverage, the average insurance payment among those people with insurance is surely larger.

Private health insurance coverage used to finance a stay in a hospital can be distributed between a number of different categories of policies. (See Table 2.5.) The categories include hospital expense insurance, surgical expense insurance, major medical expense insurance and physician expense insurance. (See Table 2.6.) People hold various combinations of policies. The most common type of policy, held by over 189.0 million people in 1980, is hospital expense insurance. It usually provides specific benefits for room and board and usual hospital services. The coverage is generally stated as a specific dollar amount per day of stay or as a percent of the daily expense. Coverage has a limit expressed in terms of maximum days of inpatient care per visit and/or a maximum dollar amount. Surgical expense insurance offers coverage for surgical procedures recommended due to illness or accident of the policy's beneficiary. Often, it covers

some portion of an anesthesiologist's fees. Benefits are stated in terms of maximum payment or a percent of payment for a specific surgical procedure. Approximately, 178.2 million people held some sort of surgical expense insurance 1n 1980. To cover physician fee expense, and sometimes diagnostic testing fees, incurred while in a hospital, at home or in an office, over 169.5 million people held physician expense insurance in 1980. Again, coverage is stated either in terms of a maximum reimbursement or as a percent of the fee (HIAA, 1982).

During the 1970 through 1982 period, the fastest growing type of health insurance was Major Medical expense insurance. While the proportion of the U.S. population covered by other types of health insurance remained relatively constant, the proportion of people covered by a Major Medical policy increased from 51.3 percent in 1970 to 68.4 percent in 1982.

Two types of major medical policies are offered by insurers; one which supplements the three previously mentioned types of insurance, and one which provides comprehensive coverage on its own. If major medical policy is the first type, it offers no coverage until the first type of coverage is exhausted, while the second type is in effect starting with the first dollar of hospital related expense. Both types include some stated form of a deductible and copayment. While some portion of major

medical policies provide unlimited benefits, most do state a maximum benefit. However, the maximum benefit varies substantially from policy to policy (HIAA, 1982)

At least for 1977, the year the National Health Expenditure Survey was conducted, if a person had private health insurance, there was a 98.7% probability she was covered for inpatient hospital expenses, a 97.2 percent probability she was covered for surgeon fee expense and a 95.9 percent probability she was covered for physician service expense (Data Preview No. 20, 1985).

A substitute product to health insurance is the health maintenance organization (HMO). For a prepaid fee, an HMO provides nearly universal medical coverage to its members. While some HMOs impose a nominal users' fee for out-of-hospital physician service and/or diagnostic medical services, all in-hospital expenses are non-existent to its members.

While HMOs have been in existence since the 1930s, they have only been popular in the last decade and a half. (See Table 2.7.) From 1972 through 1982, membership in HMO's grew by more than 208 percent (HIAA, 1983). Even more outstanding is the statistic that in one year, from December 1983 to December 1984, membership rose by 34 percent (Excelsior, 1987). Because HMOs emphasize preventative care by caring for the total patient, and consequently, they are considered to be more cost

efficient, it is expected that 25-30 percent of the U.S. population will soon be enrolled in HMOs rather than holding traditional health insurance (D. Waldo, 1986).

HMOs' popularity is not necessarily initiated by the consumer. A common perception of consumers is that HMOs provide lower quality of services with less choice. But because of the drastic increase in medical costs, HMOs are often offered by the insurers to the consumer at a substantially lower price compared to traditional insurance. In other words, insurers have diversified into the HMO market and have made HMOs too good not to be considered as an alternative to health insurance.

Federal and state governments allocate an ever increasing portion of their combined budgets toward health care expenditures. In 1970, total government expenditures were $22.4 billion; in 1980, they were $86.5 billion; and by 1985, expenditures were an amazing $147.5 billion (D. Waldo, 1986).

For the most part, hospital expenditures by the government have been through the auspices of either the Medicaid or Medicare program. Medicaid is a health financing program for our indigent members and Medicare is an insurance program for our elderly.

Medicaid is a federally funded and state administered program. It was instituted under the Social Security Act: Title IV-A to provide funds for the medical

services to groups of 'categorically needy' people. While all states currently participate, participation is voluntary. As a result, states are given a great deal of discretion as to who is eligible, specifying benefit levels and determining the provider payment mechanism. Once a program is in place, states are required to be uniform in their criteria throughout the state.

A minimum eligibility requirement for the class of 'categorically needy' is membership in a cash assistance program. A state can extend Medicaid eligibility to other groups of needy people. All participating states must provide certain basic services, one of which is inpatient hospital service. Medicaid is a 'vendor payment program.' If a hospital accepts Medicaid patients, it must accept Medicaid's reimbursement as full payment for any services rendered. As noted earlier, as of October 1983, reimbursement levels are determined by diagnostic group rather than by services rendered. Finally, expenses incurred by the Medicaid program are shared jointly by the states and the Federal government. The distribution of financial responsibility is determined by the scope of the program in place within a state (NCHS, 1981; Soc.Sec.Adm., 1981).

People, 65 years and older, and disabled people who have received social security payments, are eligible to enroll in Medicare. In fact, approximately 95 percent of

those people eligible do enroll.

Medicare as two components: One, a hospital service insurance program (Part A), is financed through a Federal tax on current earnings. The second, supplemental medical insurance (Part B), is paid for through premiums paid by enrollees. While Part A finances hospital expenditures, and Part B finances physician visits and some other medical expenses, both involve cost-sharing by users. Furthermore, there are many medical services which are not financed by Medicare (prescription drugs, dental care, preventative services and long-term care). As in the Medicaid program, hospital and physicians that accept Medicare patients must accept payment for their service according to a diagnostic group.

Even with the existence of many private and public programs, some Americans have no plan which would finance a hospital stay, should one occur. In fact, a substantial portion of the population is without any health insurance coverage. While there is a wide variation in the estimate of the portion of the population which is uninsured; the Congressional Budget Office claimed the uninsured accounted for five to eight percent of the population (HIAA, 1983) and NCHSR set their numbers to be 12.6 percent of the population; there is agreement as to what characteristics will make a person more or less likely to be uninsured. (See Table 2.8.) A study conducted by

NCHSR found the uninsured tends to be young, poorly educated, black and from rural areas (NCHSR, 1985). Finally, in the 1978 National Health Interview Survey, people who were not covered by any health financing program generally claimed they were in this predicament because health plans were too expensive (54 percent), or because they were unemployed (32 percent) (NCHS, 1981).

Table 2.1
Percent Change in Medical Costs

	General Medical Care[1]	Hospital Room[1]	CPI[2]
1970	6.3%	12.9%	5.9%
1971	6.5	12.2	4.3
1972	3.2	6.6	3.3
1973	3.9	4.7	6.2
1974	9.3	10.7	11.0
1975	12.0	17.2	9.1
1976	9.5	13.8	5.8
1977	9.6	11.5	6.5
1978	8.4	11.0	7.7
1979	9.3	11.4	11.3
1980	10.9	13.1	13.5
1981	10.8	14.8	10.4
1982	11.6	15.7	6.1
1983	8.7	11.3	3.2
1984	6.2	8.3	4.3
1985	6.2	5.9	3.6
1986	7.5	6.0	1.9

1. Source: U.S. Bureau of Labor Statistics, CPI Detailed Report, 1987.

2. Source: Bureau of Economic Analysis, Survey of Current Business, 1987.

Table 2.2
Medical and Hospital
Service Expenditures[1]

	Total[2] Medical Expenditures	Percent of GNP	Hospital[2] Expenditures
1970	$ 65.4B	7.4%	$ 26.2B
1975	120.0	8.3	50.6
1979	198.4	8.6	84.5
1980	230.0	9.1	99.0
1981	266.8	9.4	116.0
1982	301.3	10.2	131.7
1983	333.0	10.5	142.9
1984	365.2	10.4	152.0
1985	395.0	10.6	162.5
1986	430.0	10.9	173.1

1. U.S. Health Care Financing Administration, Health Care Financing Review, 1987.

2. Totals do not include research or construction money.

Table 2.3
Admisions and Average Stays
in Community Hospitals[1]

	Admissions per 1000	Average Stay in days
1970	140	8.2
1975	156	7.7
1979	156	7.6
1980	157	7.6
1981	159	7.6
1982	159	7.6
1983	157	7.6
1984	155	7.5
1985	149	7.3
1986	141	7.1

1. Source: American Hospital Association, Hospital Statistics, 1987.

Table 2.4
Hospital Expenditures and
the Private Sector[1]

	Total Hospital Expenditures	Total Private Expenditures	Percent Paid by Private Insurer
1970	$ 26.2B	$ 12.1B	75.2%
1975	50.6	21.8	81.9
1979	84.5	38.3	83.8
1980	99.0	45.4	83.1
1981	116.0	53.7	81.7
1982	131.7	60.8	82.2
1983	142.9	66.4	80.6
1984	152.0	69.6	78.8
1985	162.5	72.9	80.7
1986	173.1	79.6	79.4

1. Source: U.S. Health Care Financing Administration, Health Care Financing Review, 1987.

Table 2.5
Persons Covered by
Specific Health Insurance Policies[1,2]

	Hospital Expense Insurance	Surgical Expense Insurance	Major Medical	Physician Expense Insurance
1970	158.8	151.4	103.5	138.7
1971	161.8	153.1	108.8	139.4
1972	164.0	154.7	113.8	140.9
1973	168.5	162.6	124.6	151.7
1974	173.1	166.4	131.4	158.2
1975	178.0	168.9	134.1	161.9
1976	176.6	167.4	135.0	163.1
1977	179.0	167.2	139.4	160.4
1978	181.5	172.5	141.5	164.1
1979	183.2	173.8	147.9	163.9
1980	189.0	178.2	153.6	169.5
1981	188.3	176.9	153.7	164.1
1982	191.1	180.3	160.1	171.6
1983	189.9	179.1	159.6	173.1
1984	184.4	NA	167.7	NA
1985	181.3	NA	162.8	NA

1. Source: Health Insurance Association of America, Source Book of Health Insurance Data, 1987.

2. In millions.

Table 2.6
Percent of Population Covered
by Specific Insurance[1]

	Hospital Expense Insurance	Surgical Expense Insurance	Major Medical	Physician Expense Insurance
1970	78.8%	75.0%	51.3%	68.7%
1971	78.9	74.7	53.1	68.1
1972	78.9	74.4	54.1	67.8
1973	79.9	77.1	59.1	71.9
1974	80.9	77.8	61.4	73.9
1975	82.0	77.8	61.8	77.6
1976	80.5	76.3	61.5	74.3
1977	80.7	75.5	62.9	72.4
1978	81.0	76.7	63.1	73.2
1979	80.9	76.8	65.2	72.4
1980	82.5	77.8	67.0	74.0
1981	81.3	76.4	66.4	70.8
1982	81.7	77.1	68.4	73.4
1983	80.4	75.8	67.6	73.3
1984	77.4	NA	70.4	NA
1985	75.4	NA	67.7	NA

1. Source: Health Insurance Association of America, Source Book of Health Insurance Data, 1982, 1987.

Table 2.7
Enrollment in HMOs[1]

	Number(000)	Percentage Change	Percent of Population
1976	6016	-	2.7%
1977	6331	5.2%	2.9
1978	7471	18.0	3.3
1979	8226	10.1	3.6
1980	9100	10.6	4.0
1981	10266	12.8	4.4
1982	10831	5.5	4.6
1983	12491	15.3	5.3
1984	16743	34.0	7.0
1985	18894	12.8	9.8
1986	23664	25.2	9.8

1. Excelsior, National HMO Census, 1987.

Table 2.8
Summary Statistics
on the Unisured[1]

Group	Probability of No Insurance
18-24 Year Olds	20.9%
65 Years and Older	4.3
All Ages	12.6
Non-White	18.1
White	11.7
Adult and No H.S. Diploma	15.5
Adult with a H.S. Diploma	12.1
Rural Residence	18.0
Residence in 16 Largest SMSA	10.4
South or West	16.2
Northeast or Northwest	9.1

1. Source: NHCES, Data Preview 1: Who are the Uninsured?, 1981.

Chapter 3

Choice Theoretic of the Hospital Sector

For the 1970 through 1982 period, two distinct structural models can be considered in the analysis of cost changes in the hospital sector. One model takes as given the notion that the market is either in equilibrium or tending toward equilibrium. As a result the observed cost of a hospital stay is likely to represent one which is equilibrating quantity demanded with quantity supplied. Contrary to the first model is a second structural model which infers no such equilibration characteristic for the observed hospital charge. The hospital sector may or may not be satisfying a market clearing condition. Essential to the second model's design is a notion that desired demand and desired supply may be independent from the actualized or observed number of hospital admissions per time period, though the converse is not true.

Within the first type of market model, any explanation as to the persistent increases in hospital costs requires the introduction of exogenous factors which drive either the demand or supply schedules in an upward direction. Persistent cost increases, then, are interpreted as dynamic adjustments to the outside forces.

One popular scenario relies on the rapid increase in the general price level which characterized the period under study and the United States' progressive tax rate structure. During this period of unusually high inflation, people attempted to avoid the erosion of their total wage basket by negotiating for an ever-increasing portion of their compensation in the form of in-kind fringe benefits including group health insurance coverage. Given the progressive federal income tax structure, their objective, of course, was to avoid some of the otherwise inevitable increase in their marginal tax rate. This was agreeable to employers since income used for health insurance premiums by employers was deductible from their corporate income taxes. Since better health insurance translates into a lower coinsurance rate, the portion of a dollar of health care expenditures paid directly by the consumer, one outcome of the dynamic was an effective decrease in the price of hospital service at time of illness. Consequently, for any potential hospital charge, holding all other things constant, there is an ever larger demand for hospital services (Feldstein and Friedman, 1977). To summarize, the increases in hospital costs, in this case, can only be interpreted as a relative price change.

While the second model's design does not exclude the first model's explanation for hospital cost inflation

- there is always some probability that quantity demanded simultaneously equals quantity supplied - and the inflation is the consequence of an adjustment phenomenon, it does allow for an additional explanation. As a result, the second model can be considered a more general hospital sector model.

The major argument of this paper is that the increases in hospital service costs during the period under study are larger than what can be explained by an inflation-insurance adjustment process, or any other exogenous factor which can be translated into relative price changes, but are, in addition, driven by excess demand for hospital services. To make the argument that the hospital sector is in a permanent state of disequilibrium, it is essential to identify an economic agent who has 'motive and opportunity' to benefit from this market structure. For reasons elaborated upon below, the physician is going to be the 'culprit.'

A basic premise of the paper is that hospital cost inflation is a possible indicator of persistent excess demand for hospital services. The primary objective of this chapter is to provide a rationale for this premise. To make this argument, it must be shown, first, that it is reasonable to believe physicians may have sufficient control over market outcomes to guarantee physician-induced demand, and second, it may be in the interest of

physicians for price to be such that there is excess demand for hospital services. A secondary objective of this chapter is to state some of the assumed characteristics on both the supply and demand sides of the hospital service market.

A sketch of hospital sector relations is the following. Physicians, working jointly with hospital administrators, make hospital supply decisions, locating the supply schedule. But, in addition, because of the nature of health and hospital services, the physician is able to influence the behavior of consumers with respect to their consumption of hospital services; that is, the physician has some influence on the location of the demand schedule. Under some conditions, a market with an agent who has influence over other supply and demand factors may be better off in a market with excess demand than a market which is clearing. The hospital sector is in disequilibrium because the physician, in pursuit of self-interest, is capable of generating a disequilibrium state and benefits from the existence of a disequilibrium market.

The outline of the remainder of this section is the following: First, the dynamic relation between hospital administrators and physicians is discussed. Second, the nature of the 'good' hospital service from a utility perspective and the effects of physicians on the behavior

of consumers with respect to this 'good' are discussed. And, finally, the possible benefits physicians may be experiencing in a market with excess demand are considered.

As defined in the introduction, hospital service is a collection of medical and hotel services which are provided in physical setting known as a hospital. Within each hospital, decisions concerning the exact mix of services along with admissions and treatment (affecting length of stay) must be made. A schedule of hospital charges must be selected. Before hypothesizing how these decisions are determined, some relevant characteristics must be considered along with their influence on the decision-making process. The hospital characteristics which will be discussed below include non-profit status and the interdependence of supply agents. It should be noted that the probable trade-off between quantity and quality of hospital service is left for later research.

While a hospital may have the legal status of being either a for-profit or a non-profit organization, the vast majority of community hospitals are legally non-profit organizations. Hence, a basic assumption of the paper is that hospitals are non-profit organizations providing a merit good. In terms of economic analysis, the ramifications of this assumption is not trivial.

An obvious impact of hospitals' legal status is that they do not finance capital expansion by issuing stock. From one perspective, this means they do not need to be concerned with return to equity. From another perspective, their non-profit status seriously contracts their market for financing. For the most part, hospitals rely on locally floated bond issues, donations from philanthropists and grants from the public sector when financing a capital expansion. In addition, their status allows them exemption from some taxes and labor laws holding down costs and making donating to hospitals attractive (Newhouse, 1970). But, it is a double-edged sword. Because they produce a merit good, and society fears abuse by producers, the hospital sector is among the most regulated of sectors in the U.S. economy. (See Chapter 2 for specifics of hospital regulation.) Regulations such as licensing of professional staff, review of its product, and control of charges, which dictate a range of responses by hospitals, can be interpreted as constraints on their behavior (Sloan, 1981).

As a result of their legal status, hospital decision-makers do not have the objective of profit maximization as in the standard neo-classical model. At a minimum, it will be assumed that they attempt to cover

costs which are partially determined by their regulated status.

While the possible set of hospital decision-makers within a hospital includes a board of trustees, administrators, and physicians, for simplicity of analysis, the interests of the trustees are assumed be identical with those of administrators. Hence, the agent category, administrators, includes trustees. It should be recognized that this simplification may not be accurate for two reasons. One, administrators are employees, and therefore are concerned with their work environment and not just with equating costs with revenue. And, two, hospital trustees lack the important inducement of being able to offer stock options to administrators, which can bring the behavior of managers in a for-profit firm in line with the interest of stockholders.

In addition, administrators in this paper are considered to be partial representatives of the public trust. To the extent that hospital services can be considered merit services, the provision of such hospital services is a social contract. As implied above, the non-profit status and the legal benefits which accompany the status are evidence in favor of this hypothesis.

The majority of physicians are not employees of a hospital. They, instead, have varying levels of privileges to use a hospital's facilities. Their use of

hospital services imposes no cost on them. However, regardless of where a physician service is produced, their income is generated from fees charged directly to consumers. It is this dichotomy which differentiates the objectives of administrators from that of physicians. Administrators consider hospital services as the final product to consumers while physicians consider hospital services as an input into the production of physician services, a final product to consumers (Pauly, 1973).

Expanding upon this notion, the two agents can be considered distinct but mutually dependent with respect to their objectives. Each needs the other in the sense that neither can operate a hospital alone. However, their objectives are not identical, and may be contradictory.

A physician does not produce a homogeneous product. The product will depend on the amount of physician time used, the physician's human capital, office-located inputs and hospital-located inputs. Since technically there is no reason why all services could not be produced within the confines of a physician's office, there must be an economic motivation for observing the use of hospitals by physicians. The benefits to the physicians are twofold: First, by housing non-ambulatory patients under the same roof, physicians are able to economize their time by treating such patients in rapid succession during 'rounds.' And second, the hospital allows a fuller

utilization of specialized equipment and labor by physicians than would occur if each doctor treated her patients in her own office, thus lowering the costs to each physician.

Assuming the physician is a typical rational economic agent, for any given service, she will select an array of inputs for her service which minimizes her cost function subject to a technology constraint. While hospital-located inputs are not free inputs, their cost to the physician is a function of how much of the physician's time is required to use the hospital-located inputs, and their costs are understated to the physician. As a result, a physician will tend to recommend more treatment plans which involve hospitalization than in the case where she was responsible for all costs (Pauly, 1980).

Despite not having property rights to the hospital, physicians are active participants in the process of determining the mix of services offered by a hospital. Physicians will rationally endorse any change in the mix which results in an increase in their income. Hence, any additional hospital service which lowers their per unit cost of producing physician services, or which allows them to offer a profitable physician service, will be endorsed. In summary, physicians have economic motivation to negotiate a larger variety of hospital services than if

they were required to finance by themselves the full cost of such a service.

But, given that physicians have no property rights, a question begging for an answer is why would hospital administrators relinquish such a pivotal, and obviously costly, role to physicians. The answer has more than one dimension.

A hospital without attending physicians is not much more than a hotel. Without attending physicians, administrators would have to rely on their emergency rooms and out-patient clinics for customers. So, administrators rely on physicians for customers. While physicians are drawn to a specific hospital for its array of facilities, physicians allow administrators to bill the consumer directly for use of hospital services. Plus, having specific physicians associated with a hospital enables the hospital to differentiate its product from others, and allows a hospital to set its charges like a monopolist. Having unique services must be combined with having a unique set of physicians. Accommodating physicians in the decision-making process provides hospitals with customers and the ability to set their own price.

The cost to physicians of not having to finance their use of hospital-located inputs is sharing the decision-making process with administrators. Given a different set of motivations from physicians,

administrators will be interested in the total cost of providing additional service, less the physician time costs. It is likely that administrators will have the objective of covering their expenses and will tend to constrain hospital service mix to a least cost level for a given number of patients. This improves their probability of keeping hospitals solvent, but their behavior is constrained by a fear of losing physicians, and subsequently, patients (Robinson and Loft, 1985; Pauly, 1980).

Besides negotiating the mix of hospital services, physicians and administrators will have to also negotiate the price of hospital services. Pragmatically, consumers react to their estimate of total price of hospital services: the hospital charge and the physicians' fees. To the extent that a hospital is a monopolist, total price will be larger than marginal cost, and quantity supplied will be less than what is socially optimal.

A last question is how will the return over cost be distributed between administrators and physicians. The answer depends on the distribution of control between them. If physicians have absolute control, then the return to administrators will be just sufficient to cover costs and to expand in ways which are beneficial to the physicians. The residual will returned as fees to physicians. If, on the other hand, administrators have

absolute control then the return to physicians will be just sufficient to cover their average costs plus an inducement to keep them utilizing the hospital. Any other variation in the distribution of control between physicians and administrators will involve some negotiation process which will weigh their relative strengths.

In the negotiation of both hospital service mix and total price, the outcome will depend on the concentration of hospitals and physicians within a locale. Specifically, the larger the ratio of hospital beds to physicians, the more weight the physicians' objectives will be given and the larger the return to physicians (Robinson and Loft, 1985).

As far as consumers are concerned, hospital service is one of many goods that may be used by the consumer in the production of 'good health' or 'physical well-being.' The demand for any goods or services purchased by the household sector can be considered as derived from a desire to produce some intrinsic good which yields utility to individuals. If the focus of the paper was on demand for hospital care, it would be advantageous to begin with the demand for health, and then derive a demand for hospital service (Grossman, 1972). But, since its focus is on the hospital sector, the determinants of demand for

health will merely be included as determinants of the demand for hospital services.

That demand for hospital services is a derived demand does not explain why its expenditures are financed by insurance or why consumers require outside information concerning its impact on consumer production of 'good health.' It is these characteristics which cause hospital services to be an atypical product.

A person's state of health is more stochastic in nature than other 'goods' found to yield utility to the individual. Even with a specific current level of health and an expected change in the level of health for the next period, an individual faces the possibility of being very wrong in her forecast about her health level. Moreover, a decrease in a person's health state will have a negative effect on a person's ability to earn income. Consequently, variability in a person's level of health introduces variability in a person's ability to earn income. Given this situation, the emergence of insurance to finance expenditures on health producing goods, like hospital service, and to protect income levels, is predictable (Arrow, 1963; Pauly, 1986).

While the uncertain nature of the derived demand for hospital service results in the financing of its expenditures through insurance, health insurance markets tend to be different from other types of insurance

markets. In a world where the insurer and insured have identical information and where the price of insurance is actuarially fair - a price which is based on the odds of having a health status that would require the use of medical products. - people would insure against all losses (Ehrlich and Becker, 1972). If this were the case, people would acquire medical insurance which would cover all costs associated with illness, including a stay in a hospital.

Because of the phenomena of adverse selection and moral hazard, health expenditures are not fully insured and health insurance tends to be purchased by an individual as a member of a group. Specifically, it is likely the individual has better information than the insurer on the probability of incurring a change in his or her health status. Also, an individual's spending on health services in response to such a change varies considerably, and again, the individual has better information on his or her likely behavior than the insurer. The insurer's fear that improved insurance will result in an increase in medical expenditures following any negative change in health status results an offer of less than full coverage. By pooling potential insurance buyers in a group, insurers reduce both types of risk. Insurers are able to offer any level of insurance for less to a member of such a group than to an individual person.

When an consumer purchases group health insurance, there is the clear benefit of a lower price for any given level of insurance (Pauly, 1968). Given our current tax codes, consumers prefer that for the group be an employment group (Feldstein, 1973; Vogel, 1981).

In addition, health has a public and merit good aspect. One person's good health, or lack of it, potentially may affect another's good health. And a popular notion is that, as part of our social contract, our members have a right to good health, or in the very least, there ought not be an unreasonable difference in access to medical care between members of our society. Therefore, to the extent that private insurers do not have the incentive to provide full coverage, and that there is a clear price differential between group and individual acquired insurance, public insurers have been mandated (Arrow, 1963).

Regardless of whether health insurance is acquired privately or through public entitlement, at the time of illness, the consumer is facing a 'price' which is less than the actual market price.

But there is another dimension to 'good health' which makes it distinct from other goods and services. The production of 'good health' is complicated and varies substantially from period to period and from individual to individual. Under some circumstances, the individual will

have a pretty good idea as to what combination of market-purchased goods and services are necessary in its production. And, even if the individual miscalculates, errors in judgment will not make much difference in the actual level of health services the individual consumes. But under different circumstances, this will not be the case. During some time periods, determining what inputs are necessary and in what combination is beyond the comprehension of the consumer. Even if this is known, due to the capital-intensive nature of some medical care and the barriers to entry, the administration of medical inputs by the consumer is impossible. Hospital service is one such input. The consumer requires the advise of an agent - in this case, the physician, both to determine the mix and nature of medical inputs and to administer the production of health (Pauly, 1980).

The inability of the consumer to assess the consequence of hospital services or to access hospital services without the permission of a physician results in the possibility of physician-induced demand and the potential of the sector being in disequilibrium.

The physician can be considered the 'gatekeeper' to hospital services. On the one hand, the physician may act only with the consumer's interest in mind, hence, as a perfect agent, subordinating her own objectives to those of the consumer. If this were the case, the physician's

recommendation to the consumer would be identical to what the consumer would choose if she had the same information and entry rights as the physician. On the other hand, a physician acting purely in this fashion would not be rational from an economic perspective. Logically, it would be expected that the physician will base her recommendation on her perception of the consumer's desires and her own objectives. It may or may not perfectly correspond to the perfect agent case (Pauly, 1980).

By definition, then, physician-induced demand is demand for hospital services which would not have occurred if the consumer had identical information as the physician (Hay and Leahy, 1982). The induced demand may be motivated by a desire to save physician time, to regularize the physician's schedule, to reduce risk of potential lawsuits, misinformation on the part of the physician as to the consumer's attitude toward health and risk, or even physician's own income (Pauly, 1980).

Physician-induced demand might affect the negotiation process between physicians and hospital administrators. As already mentioned, physicians and administrators do not share the same incentives in decisions concerning the mix of hospital services. Specifically, the physician, who treats hospital inputs as free in the production of her services, tends to demand more hospital services than the administrator, who has the

goal of equating expected average cost with expected average revenue. But, if a physician can show steady demand for the hospital service - say, more than current levels can service, then the administrator is likely to permit expansion, in the belief that there is little risk of not satisfying his objective (Feldstein, 1971).

Supporters of the notion of physician-induced demand point to the large interregional variation in per capita utilization of elective services for evidence (Reinhardt, 1978; Fuchs, 1978; Cromwell and Mitchell, 1986). Within regions, high per capita utilization appear to be strongly correlated with above-average physician incomes and densities. By encouraging consumers to use their services more frequently or to use more expensive procedures, physicians may be targeting an income objective. Fee income per physician in a high-density area ought to be significantly higher than fee income of physicians in other regions. Of course, antagonists of the physician-induced demand theory argue that physicians are merely locating in regions with above average demand for their services (Hay and Leahy, 1982). In this case, the larger supply of physicians per capita will result in an increase in the quantity of physicians' services demanded. Concurrently, fee income per physician should have declined. The obvious test for the existence of physician-induced demand for medical services

is whether a change in the number of physicians affects fees.

Assuming that the supporters are correct and that physicians can induce demand for their services, it is not obvious why a physician would prefer a market where the prevailing fees and hospital charges are below market-clearing amounts. In other words, while the theory does provide an explanation for the persistent excess demand scenario, it is not obvious why a market-clearing price will not be observed. The physician can generate the scenario, but will do so only if it is in her interest. Two benefits are considered below:

Excess demand in the form of an actual queue or the necessity of waiting for admission into a hospital may be relaying information to the consumer from the physician that the physician is unable to relay directly to the consumer. When hospitalization is suggested to the consumer, the consumer must decide whether the physician's recommendation falls within one of three cases: First, the recommendation is in her interest (and her physician is a true altruist), second, it is jointly in her interest and the physician's interest, or third, it is merely in the physician's interest. If the consumer believes that either the first or the second is the case, and if the recommendation is her best alternative, then she will accept it. If she believes the third is the case, she

will reject the recommendation and she may terminate her relation with the physician. The physician is best able to exploit the situation and induce demand when the consumer opts to believe the first case. In this case, the physician may recommend hospitalization even if the third case is true without experiencing any penalty. The physician is least able to exploit the situation when the consumer is not able to discern when a recommendation is in her interest or in the physician's interest. Fear that a consumer may choose the third case, may provide an incentive to act solely in the consumer's interest - at least until that period of time the consumer believes the first case to be true.

Under the best of circumstances, the consumer would consider her health state and her information set on medical procedures relevant to her situation, then compare these with what her physician recommended and come to a decision. However, as noted already, hospital service is particularly difficult for the consumer to evaluate and its difficulty renders the consumer exploitable; hence, she may rely on proxy measures of value. While it is not unusual to consider price as a proxy measure for value, price may not be useful in this case.

First, an estimate of the full price for a hospital stay may be difficult to obtain prior to admission. Since hospital service associated with any particular stay has a

significant stochastic component, willingness of either the administrator or the physician to be specific about total cost will depend on their attitudes concerning the risk of miscalculating and the costs associated with such an act. For the most part, given the nature of medical care and its interaction with a person's health, they may be reluctant to project an expected bill due to the large amount of variation in bills, even within a specific diagnostic category.

Second, finding out about the cost experience of previous patients will give the consumer little insight into quality of service. A sample of total hospital bills and physician fees from past patients may reflect the medical and personal idiosyncrasies of the individual patients rather than the quality of medical services being offered jointly by a physician and a hospital. In addition, given the rate at which medical technology changes, it would be expensive for an individual to keep this type of information set current.

An alternative measure of quality of a proposed medical plan which includes admission into a hospital is for the potential patient not only hear about other people who have had a procedure, but also to hear that others have waited for the opportunity to receive a procedure. After receiving a recommendation for a triple bypass operation, and not knowing whether this recommendation is

in a patient's interest or only in a physician's interest, hearing stories of people waiting to be admitted for this type of surgery may influence the consumers' attitude about the intention of the physician. With this information, a consumer has a larger likelihood of concluding that a physician does have her interest in mind in recommending a procedure. Thus, by generating excess demand in the current period, the physician is grooming demand for the future.

Though a majority of Americans have extensive health insurance, significantly reducing out-of-pocket costs for both hospital and physician services, there are many Americans who either do not have the same depth of coverage or have no coverage at all for their medical expenses. They may be willing to incur the time costs associated by hospital stay, but are eliminated from the market when money costs are added in estimating the cost of a stay. When price is less than the market clearing price then the physician does not just include those people who can afford to paid the hospital bills but includes patients she may choose according to some other objective. For instance, to the extent that the physician includes in her objective function such items as the level of challenge in her case load or medical-access equity, setting price below a market clearing price allows her to pursue such objectives.

Along the same lines, the excess demand may allow the physician to choose to service those patients which fit their 'risk-attitude' profile. A risk-averse physician may opt to select patients she estimates will deviate little from her prognosis, while a risk-loving physician may have a taste for patients which are expected to deviate substantially from the prognosis. Again, a price set below the equilibrium price allow the physician to accommodate her taste for risk.

In conclusion, it is possible to provide a rationale for physicians to prefer a market for the hospital services which is characterized by excess demand. However, for this to be the case persistently, the benefits associated with excess demand must be larger than the forgone income the physician is sacrificing when setting price below the market equilibrium price. Otherwise, an equilibrium market would be preferred.

Two disclaimers must be made about the above scenario. First, the hospital market structure may have changed significantly with the introduction of diagnosis-related group reimbursement programs. Their entrance may have altered incentives and the distribution of control between administrators and physicians in the hospital sector. That is the primary reason why all future testing of the scenario will not include data beyond 1982. To do so would introduce the necessity to test whether or not

the structure of the hospital sector had changed after 1982. (However, it should be noted that the design of the model's likelihood function is capable of estimating an excess demand and an excess supply.) Second, all characteristics of insurance have assumed to be exogenous. It is highly unlikely this is the case. For instance, the increase in the size of claims by the insured must have some effect on the premiums charged by insurers. The emergence of health maintenance organizations sponsored by insurance companies during the time period is an indication that the behavior of the insurance industry is influenced by and has influenced, the hospital sector.

As a result of the above theoretic, the market model design of Chapter 4 will include the following characteristics:

First, the observed quantity of hospital service will not be assumed to represent either quantity demanded or quantity supplied.

Second, the demand equation will include variables associated with both the consumer and the physician. In addition, the relevant price at time of illness will reflect the depth of hospital insurance coverage.

Third, the supply equation will include variables representing the interests of both physicians and hospital administrators. Instead of capital market variables, variables associated with philanthropy and government

grants will be included to capture their influence on expansion of hospital services.

Fourth, since the above scenario does exclude a probable dynamic relation between the coinsurance rate and hospital charges, two adjustment equations will be included in the model to pick up their effect on determining both supply and demanded.

One last note, a variable which has not entered the above scenario will be included in the hospital sector model. This variable is the vacancy rate, the number of empty beds relative to the number of available beds, (or, 1-occupancy rate). The sole reason for including the vacancy rate is to measure the amount of excess demand in the hospital sector.

Chapter 4

The Hospital Sector Model

The hospital sector model is designed to explain six endogenous variables simultaneously. Thus it is a structural model consisting of six equations. A highlight of the model is that it is constructed to test the possibility that the sector is in disequilibrium. The objective of this chapter is to specify each equation taking into account the probable variables which explain their respective values.

The six equations are the following: One equation represents the observed quantity of hospital care, a second the demand for hospital care and a third, the supply of hospital care. In all three equations, the annual number of admissions into inpatient care is treated as the quantity variable. In addition, there are two adjustment equations. One represents the adjustment in costs associated with a day of inpatient care and a second represents the adjustment in the coinsurance rate, the proportion of each dollar of hospital charges not financed by a third-party payer. The final equation of the model explains variation in the amount of excess demand or supply through movement in the hospital vacancy rate.

Five of the six equations are stochastic in nature. The exception is the equation for the observed quantity of

hospital service. This equation defines the observed quantity as the minimum of quantity demanded or quantity supplied; thus, it is treated as deterministic. The disturbance terms for each of the stochastic equations are assumed to be normally distributed with a zero mean and a constant variance. For purposes of estimation, all variables are expressed as natural logs. The exception is the time trend variable in Eq.2. The unit of observation is the state over the time period 1971 through 1982, that is, the study uses a pooled time series.

A fuller description of the model's disturbance terms and the actual derivation of a likelihood function will be the central topic of the following chapter. The remainder of this chapter will discuss the six equations in turn. A summary of expected values for all the coefficients in the model follows the text of the chapter (see Table 4.1).

Observed Quantity of Hospital Services

Economists' notion of equilibrium pertains to a market situation where the quantity demanded of a product is just equal to the quantity supplied. Implicit is the notion that the demand and supply schedules were derived under an assumption that the 'resulting price' is that which will clear the market (Hey, 1981). If a discrepancy exists between the quantities, it must be due to omitted

information and transaction costs. Once these omissions are recognized, voila, the market will clear to its neoclassical nirvana solution (Maddala, 1983). For such economists, design and estimation of a disequilibrium market model are superfluous. Disequilibriums are momentary, hence there is no need to understand the condition.

However, since Feldstein's 1971 article on the hospital sector, a number of economists have entertained the hypothesis that some medical markets may be characterized by persistent disequilibrium conditions (Green, 1978; Pauly, 1980; Hu and Yang, 1985; Cromwell and Mitchell, 1986).

This paper has already reasoned that physicians have sufficient control over demand and supply conditions to generate a market which is to their advantage. And they may have a preference for a market with excess demand for their services.

Since the intention is to develop a method to <u>test</u> this hypothesis, the quantity observed equation should neither presuppose a market clearing condition nor be restricted to the case of excess demand. Consequently, the model's equation merely requires the observed quantity to be the minimum of quantity demanded and quantity supplied:

$$(1) \quad Q_{it} = \min (D_{it}, S_{it})$$

Q_{it} represents the observed quantity of hospital services and is constructed as the aggregate number of admissions into a hospital per 1000 persons in the ith state during the year t. (All variables are observed by state over time, thus all variables are subscripted by 'it'.) D_{it} represents the quantity demanded of hospital services and is constructed to measure the number of hospital admissions desired by 1000 persons. S_{it} represents the quantity supplied of hospital services and is constructed to measure the number of hospital admissions administrators and physicians desire to supply per 1000 persons.

The consequence of this specification for quantity observed is that if $D_{it}>S_{it}$, then only S_{it} may be observed and there is excess demand. Conversely, if $S_{it}>D_{it}$, then only D_{it} may be observed and there is excess supply. Unfortunately, this introduces the problem of needing to know when $Q_{it} = D_{it}$ and when $Q_{it} = S_{it}$. If the sample's separation between the two possibilities is known, then the observations associated with $Q_{it} = D_{it}$ can be treated as one regime while $Q_{it} = S_{it}$ can be treated as another. Not knowing this information not only results in at least one endogenous variable not being observed, but it is unknown which endogenous variable is unobserved.

In order to circumvent the separation problem, two strategies are used by researchers. They may introduce a

sorting rule into the model or they may attempt to find the probability of observing quantity demanded or quantity supplied.

Fair and Jaffee (1972) in their seminal paper on the housing market included a rule which sorted the sample depending on the type of change that occurred with the local mortgage rate. Specifically, they hypothesized that a positive change in the mortgage rate was indicative of a market with excess demand and the observed quantity represented quantity supplied.

Closer to the hospital sector model, Cromwell and Mitchell (1986) use a quantity rule to sort their observations of surgeries. They found an average surgeon's workload to be 280 cases per annum and defined an area experiencing a shortage (excess demand) of surgeons if the workload was greater than 280 cases.

This paper rejects the 'rule' approach since, for instance, a positive increase in hospital charges could be explained as an adjustment to a decrease in the average coinsurance rate rather than as a response to persistent excess demand. Similarly, an area with a surgery caseload higher than the average may indicate a sicker population, lower fees, or better quality service rather than excess demand.

Instead this paper has adopted the second strategy. This path will require the derivation of a likelihood

function based on the joint conditional probabilities of the observed variables being associated with either quantity demanded or quantity supplied. The omission of a sorting rule results in the model being classified as an endogenous switching model. (See Chapter 5 for a complete discussion of this issue and the derivation of the likelihood function.) It should be noted that Hu and Yang (1985) also used this approach in their analysis of the physician service market.

An alternative to Eq.1 has been suggested by Jerry Green (1978) who hypothesized that the quantity observed of a product with rising costs (in his study, physician services; in this paper, hospital services) would fall somewhere between S_{it}, the short-side of the market, and D_{it}:

(1)' $Q_{it} = S_{it} + (D_{it} - S_{it})^b$, where $0<b<1$.

Green suggested that the significance of the second term in Eq.1' would depend on such factors as the degree of sensitivity physicians (and administrators) had to the 'merit good' aspect of medical services and their susceptibility to patient pressure to satisfy excess demand for their services.

The alternative expression for Q_{it} is rejected for three reasons: First, Eq.1' does not solve the problem of at least one endogenous variable not being observed. Second, its configuration introduces more non-linearity to

an already complex model. Third and most important, Eq.1' restricts the market to one of either excess demand or equilibrium, disallowing the possibility that the sector could be characterized by excess supply.

Demand for Hospital Services

The derived demand for hospital services is determined by characteristics associated with hospitals, consumers and their physicians. The intention of the following demand for hospital services equation is to incorporate the relevant consumer theories outlined in Chapter 3.

$$(2)\quad D_{it} = a_0 + a_1 NP_{it} + a_2 T_{it} + a_3 Y_{it} + a_4 PCPHY_{it} + a_5 OPHY_{it} + a_6 ELD_{it} + a_7 ED_{it} + a_8 NW_{it} + a_9 G_{it} + a_{10} I_{it} + U_{2it}$$

where D_{it} is the desired aggregate quantity of hospital services demand per 1000 individuals, NP_{it} is the relative net monetary price of hospital services at the time of illness, T_{it} is the time cost associated with an average stay in a hospital, Y_{it} is the average total income per 1000 individuals, $PCPHY_{it}$ is the number of primary care physicians per 1000 individuals, $OPHY_{it}$ is the number of other physicians per 1000 individuals, $OPHY_{it}$ is the number of other physicians per 1000 individuals, ELD_{it} is the number of people 65 years and older per 1000 individuals, ED_{it} is the number of people with at least a

high school education per 1000 individuals, NW_{it} is the number of non-white people per 1000 individuals, G_{it} is the number of women per 1000 individuals and finally, t is a time trend variable. Eq.2 depends solely on current period data. Since the data are annual, this is not an unusual specification for a demand equation.

Demand for hospital services is measured in terms of the total number of admissions desired per 1000 individuals. As far as this study is concerned, quantity demanded is a desired and not necessarily observed variable. At best, during some years, in some states, the actual number of admissions used by consumers corresponds to the demand for admissions. However, in what periods and in what states this event occurs are unknown.

The relevant price to a consumer considering a typical stay in a hospital is comprised on direct hospital charges, physician fees, their respective coinsurance rates and the time costs associated with the stay. The money components and the time components are considered separately in the regression. The objective of this approach is to take into account the probable differing effects of the two components of price on demand for hospital services.

The individual will take into account all monetary expenses whether presented by the hospital's billing department or the physician's secretary. To complicate

matters, the individual may be presented with bills from more than one physician; the physician who recommended the individual for a hospital-based procedure and the physician, often a surgeon or other specialist, who actually is providing the inpatient hospital service. However, when considering money costs of a hospital stay, the consumer will disregard as irrelevant in her decision any portion of hospital expenses which is paid by a third-party payer, generally, some type of insurer (Feldstein, 1971). By definition, the portion of each dollar of hospital expenses the individual must pay out of pocket is the 'coinsurance rate' (Feldstein and Friedman, 1977). Unfortunately, since there are three potential sources of hospital expenses, there are up to three relevant third-party payers, thus, three relevant coinsurance rates. As a result, ideally, the price of an admission into a hospital is the hospital charges for an average stay,less the portion of this covered by hospital insurance; the bill from the physician, less the portion of this covered by physician insurance, and the bill from any involved surgeon, less the portion of this covered by surgery insurance.

While information pertaining to the total number of people with various types of health insurance is available for the sample, no information is available on the average coinsurance rate nor on the utilization rate associated

with specific types of health insurance. In addition, the data on the proportion of physician income associated with inpatient hospital services, or surgeons' aggregate income, is not available for more than an occasional observation point. As a result, the cost of hospital services is defined as the average hospital charge per inpatient day. Similar to a methodology used by Feldstein (1971), the coinsurance rate is constructed from the data on the proportion of a state's population in various insurance groups. their utilization rates for hospital service and their likely coinsurance rate. Taking all the data limitations into account, the relevant price to the consumer is $NP_{it} = COIN_{it}P_{it}$, where NP_{it} is the at-time-of-illness price per stay, $COIN_{it}$ is the average coinsurance rate and P_{it} is the average hospital charge per stay of inpatient service deflated by a state-specific consumer price index.

Any increases in the hospital coinsurance rate or hospital charges will be perceived as an increase in the cost of care facing the consumer at the time of illness and, thus, will have the tendency to decrease demand for hospital care, all other things being equal. As a result, coefficient a_1 in Eq.2 is expected to have a negative value. Since all equations in the model are expressed in log-linear form, the coefficients, including a_1 are elasticities. In the case of a_1, it is the net hospital

charge elasticity of demand. The magnitude of a_1 will indicate whether or not hospitals are considered a necessity by the public. A priori, it is not possible to hypothesize its magnitude. However, if hospital care when recommended was considered a necessity, then the coefficient would be less than one, indicating an insensitivity to changes in net hospital charges. On the other hand, if hospital care is perceived to be elective, then the magnitude ought to reflect a sensitivity to changes in net hospital charges and it should be larger than one in value (Feldstein, 1971). The reality is that hospital care is not a homogeneous product. Some procedures will be perceived as necessities, and others will not. However, since the data is aggregated over an entire state's population, the net classification is impossible to predict.

Even if the estimated value of a_1 is less than one, it may not indicate that hospital service is a necessity. Studies using cross-sectional data have found a large variation in price elasticities, the implication being that most hospital services cannot be categorized definitively as necessary or not. One state's necessary admission seems to be another state's elective admission (Cromwell and Mitchell, 1986). This variation may indicate that populations within different states have different attitudes toward health and health intervention.

Ideally, the time cost component, T_{it} of the relevant price for hospital service would be the sum of time costs associated with the average stay in the hospital and time costs associated with pre- and post-hospital stay physician visits. To be more precise, time costs associated with waiting to see a physician, time spent conferring with the physician, and travel time to and from the physician ought to be included. Unfortunately, due to the limits of available information, time costs will include only direct time costs associated with the average stay in a hospital, S_{it}.

Data limitations also affect the evaluation of the average time spent in a hospital. For this study, time will be evaluated at the state-specific average real wage rate, W_{it}. Since the total population can be divided between employed and non-employed, using the wage rate as a measure of the marginal value of time is incorrect. Ideally, time should be evaluated as a weighted average based on the average opportunity cost associated with both groups. A further consideration is the fact that the non-employed are not a homogenous group. Furthermore, no already-computed statistic representing the value of time for the non-employed group exists. Therefore, measures for the various components of the non-employed would have to be found. Some suggested time values for sub-groups of the non-employed are the following: The household

technician wage rate could be used to evaluate the opportunity cost of time for people who opt to work solely as homemakers. The 'post-career' wage rate could be used to evaluate time for people who have retired from the workforce. Obtaining such estimates for a sample of states over time would be very difficult; hence, the simple wage rate is assumed to be the relevant opportunity cost of time for all members of the population.

In summary, the value of an average hospital stay is computed as the average length of stay, S_{it}, evaluated at the average daily wage, W_{it}: $T_{it} = W_{it}*S_{it}$.

An increase in time costs will impact on the demand for hospital care in a complex fashion. If the increase is a result of an increase in the necessary time spent as an inpatient per admission, then the effect on quantity demanded is an unambiguous decrease. Similarly, if the increase in time costs is a result of an increase in the wage rate then there will be an incentive for quantity demanded to decline. Some of the decrease is a pure shift in allocation of time, while some portion of the decrease is due to substitution toward less time intensive medical procedures. On the other hand, the same increase in the wage rate, since it simultaneously increases the value of the individual's health stock, may induce an increase in utilization of hospital procedures which improve productivity (Grossman, 1972). These contradictory

impulses make it difficult to hypothesize a value for coefficient, a_2, the measure of net responsiveness of demand for hospital services to changes in time costs.

Due to the pervasiveness and dept of third-party payers of hospital expenses, it has been conjectured that the monetary costs at time-of-illness will have less effect than the time costs on the consumer's decision to enter the hospital. If this is the case, then the responsiveness of demand to time costs, a_2, will have a larger magnitude than the responsiveness of demand to monetary costs, a_1.

Obviously, income, Y_{it}, measured as average total income, will impact on the demand for hospital services. Because average total income is a composite of earned and unearned income, the relation between income and demand for hospital services is multidimensional. Unearned income has an unambiguous positive impact on the demand for hospital service. As unearned income increases, demand for 'good health' increases, resulting in an increase in demand for all inputs in the production of 'good health' including the demand for hospital services. However, the consequences associated with an increase in earned income with respect to the demand for hospital services are complex. If earned income is considered to be a proxy measure of labor market efficiency, or the return to time spent in the labor market by an individual,

an increase in earned income should act as an incentive to increase the optimal number of hours in the labor market. While this may be achieved by reducing the number of leisure hours, it can also be achieved by increasing the number of 'healthy hours,' - the number of hours available for either quality leisure activities or quality employment activity - reducing the number of 'sick days'. To acquire these extra 'healthy days,' the individual may employ hospital services in their production (Grossman, 1972).

In addition, higher earned income levels are positively correlated with better insurance coverage. People with higher incomes and their correspondingly better insurance coverage will pay a smaller portion of each dollar of hospital service expenses than their poorer counterpart (Newhouse, 1978). Regardless of whether it is due to an increase in labor efficiency or an improved hospital insurance policy, an increase in earned income will tend to increase the demand for hospital services.

If these were the sole dynamics between earned income and the demand for hospital service, the total effect of income, as measured by coefficient a_3, would have a positive expected value. However, it is not. An increase in earned income will also result in an increase in time costs associated with hospital services. As a consequence of the earned income effect on time costs,

some individuals will substitute toward methods of producing 'healthy days' which do not utilize as many time-intensive hospital services. Whether or not the combined positive effects outweigh the one negative tendency of income on the demand for hospital service is still to be seen.

The variable, number of primary care physicians per 1000 people, $PCPHY_{it}$, is used in Eq.2 as a measure of the influence of a substitute input into the production of 'good health' by the individual. Primary care physicians include general practitioners, family practitioners and internists. The availability of primary care physicians to the public may influence their decision to employ hospital-based techniques in producing 'good health'. Assuming a primary care physician is more likely to recommend non-hospital procedures in prescribing medical care than the specialists who do most of their work in hospitals, an increase in the availability of primary care physicians should result in a decrease in the demand for hospital care. This substitution will be prevalent in situations where there are more than one method of treating a medical problem, hence the coefficient, a_4, has a hypothesized negative value. It should be noted, however, that this effect will be reduced for states where it is common for primary care physicians to have hospital privileges or in states where the majority of hospitals

have an 'open-staff' policy, a policy of allowing staff privileges to any licensed physician. Both practices are observed more frequently in states with large rural populations and in states with low surgeon per population ratios (Cromwell and Mitchell, 1986).

A measure for the degree of control physicians have over demand for hospital services is the number of other physicians per 1000 persons, $OPHY_{it}$. A positive value for coefficient, a_5, is hypothesized to indicate that physicians do, indeed, induce demand, or shift the demand schedule in their favor (Reinhardt, 1978). Specifically, the coefficient is measuring the responsiveness of changes in the demand for hospital services to a percentage change in physician fees.

As already discussed in an earlier section, physicians, like any economic agent, desire to maximize some objective function, in this case, their returns from their professional activities. What distinguishes them from other suppliers in their quest to maximize their own functions is their influence over consumers in demand decisions. Consumers of medical services rely on the recommendations of physicians since they are unable either to assess their own health status, or health care, or even access, without permission of a physician, many types of medical products, especially hospital services. The consequence is that physicians will recommend procedures

which take into account their own income levels. For instance, "where...[they] might have advocated no

> treatment for silent gallstones, [they] now recommend a cholecystectomy. Rather than suggesting drug therapy for menstrual irregularities, [they] propose a hysterectomy. Similarly, the [physicians] now recommend coronary artery bypass graft surgery for patients with sever angina, then previously [they] would have treated such cases medically" (Cromwell and Mitchell, 1980.

While the positive impact on their income is the motivation, the opportunity for physicians to engage in this type of behavior is provided through an ability to induce consumers to follow their recommendations.

Obviously, there are limits to physicians' ability to engage in this type of activity. First, consumers do 'shop' around for physicians. And, if a physician's recommendation appears outside of the consumer's expectations of medical care for a particular ailment, a consumer may choose to bring her business elsewhere. Second, not all hospital procedures are presented to the consumer as being 'emergency,' but 'elective.' And, elective hospitalization is just that, elective. Consumers have the perception that even the most routine hospital procedures have risk, the source of which ranges from receiving anesthesia to being confined with sick people. This risk will discourage some from accepting a physician's recommendation for hospitalization. Moreover, physicians' own conscience will limit the degree to which they consider their own income levels in their

recommendations. Hence, while the coefficient is expected to be positive and significant, it is not without bound.

Sociodemographic variables, (the number of elderly people per 1000 individuals, ELD_{it}, the number of people with at least a high school education per 1000 individuals, ED_{it}, the number of non-whites per 1000 individuals, NW_{it}, the number of women per 1000 individuals, G_{it}), are included to capture the effects of tastes on the demand for hospital services.

Despite an increase in mortality rates associated with a stay in the hospital with an increase in age, an increase in age will result in an increase in demand for hospital services. As people age, the rate at which the health stock depreciates is hypothesized to increase, moving the health stock toward the minimum level required for life at an ever-increasing rate (Grossman, 1972). As this occurs, methods of slowing this movement and keeping health stock above a minimum become more likely to require non-elective hospitalization. Hence, ELD_{it}'s coefficient, a_6, is expected to indicate a positive response to an increase in the proportion of elderly in a population. If hospital stays could be sorted by elective stays versus non-elective stays, an increase in this proportion of elderly in the population would result in a decrease in elective stays and an increase in non-elective stays. But, unfortunately, the data are much too aggregated for

this.

Like income, education can influence demand for hospital services in contrary ways. On one hand, an increase in a person's education level will increase the demand for all medical services since this person is more knowledgeable of the relation between health stock production and medical services. On the other hand, a more educated person will be more efficient in combining all inputs in the production of health stock, hence, for any given health stock level, this person will require less of all inputs, including the hospital service inputs (Grossman, 1972). In the same vein, as education levels of consumers increase, physicians, knowing their patients are now less likely to need their services as an agent, will be less likely to attempt to induce demand to improve their own wealth when advising patients. As a result of all the dynamics associated with education, the most that can be asserted about the net effect of an increase in the proportion of educated people in a population is that it may have some significant effect.

Evidence indicate that a white person is likely to consume more medical services than a non-white person. However, much of the difference in utilization is eliminated once income and education levels are introduced as an explanation in the regression (Cromwell and Mitchell, 1986). In fact, once income and education

variables are introduced in an equation for the demand for medical services, the coefficient measuring its response to changes in the numbers of non-white people is hypothesized to be positive in value. One reason is that it is likely that the quality or productivity of medical services available to non-white people is less than those available to white people. Consequently, non-white people would require a larger quantity of medical services for each unit of 'good health' produced. A second reason may be that non-white people may systematically find themselves living in communities or employed in occupations which expose them to more health hazards than white people, thus increasing their need for medical services. Since both scenarios imply that a larger proportion of non-white people would require more use of all types of health services including hospital services, coefficient, a_8, is expected to be positive.

Empirically, it has been found that, even excluding hospital stays associated with obstetric procedures, women demand more hospital services than men. Therefore, a gender variable, G_{it}, measured as the proportion of women in a population, is included in Eq.2 to measure the effect of variation in this ratio on the demand for hospital services. The expected sign of its coefficient, a_9, is positive.

Cromwell and Mitchell (1983) suggested that the

proportion of women within a state is likely to be similar across states and time. If this is the case, there will be no significant variation in a gender variable. Demand by women would determine a level of demand but would not contribute to an explanation of variation in demand for hospital service across states and time. If this proves to be the case, to save on degrees of freedom, the gender variable will not be included in Eq.2.

Finally, the time trend variable, t, is not included to capture population effects on the demand for hospital services. By expressing variables in terms of units per 1000 persons in a state, population changes will not be a determinant of demand. Instead, the time trend variable is a crude attempt to measure the effect of quality changes in hospital service on demand. Obviously, a day of care in 1970 is a very different product from a day of care in 1982. While time trend's coefficient, a_{10}, is a measure of the average effect of technological change on demand, a priori, it is difficult to guess whether or not its value will be positive or negative.

A medical breakthrough, hypothetically, can generate two possible reactions with respect to admissions into a hospital. Some new procedures will induce people to consumer more hospital service. Such people are those who had lived with a negative medical condition who now believe that the improved service is worth the cost of a

hospital stay. For example, the introduction of joint implants has provided incentive for many people who lived with chronic, debilitating joint conditions to enter a hospital. On the other hand, other medical breakthroughs have freed people from the necessity of using hospital services. New procedures which utilize medication or outpatient facilities, or involve a one-time admission into inpatient service, will reduce the total number of hospital admissions. Such an example is the invention of long-term IV access equipment such as the Hickman catheter or the portacath. Such equipment allow people who require drugs introduced to their systems through IVs to leave the hospital and receive their treatment at home or on an outpatient bases.

It should be noted that possible quality changes between states will be ignored. A weak argument for this omission is an assumption that information about state-of-the-art medical technology moves quickly among the states. But it is acknowledged that this is a weak assumption; while the flow of information about technology may be quick. implementation of new technology is not identical among states. Optimally, changes in quality of care as perceived by the consumer might be measured by quantifying how successful people expect a hospital procedure will be in achieving some health objective. Changes in quality then could be measured in terms of changes in prognosis.

Unfortunately, when a study involves a pooled time-series, it is impossible to quantify this dimension of hospital services.

Other researchers have equated changes in the number of hospital employees per bed (Feldstein, 1977) or the number of CAT scans per state to quality differences and have included such variables in demand for hospital service equations. While the supply equation (Eq.3) does include the number of employees per bed, it is for a different reason.

The Supply Equation

The joint behavior of physicians and administrators with respect to the provision of hospital service will result in the following supply for hospital service equation:

$$(3)\quad S_{it} = b_0 + b_1P_{it} + b_2V_{it} + b_3LX_{it} + b_4IOX_{it} + b_5DON_{it} + b_6GG_{it} + b_7INT_{it} + U_{3it}$$

where S_{it} is the desired total supply quantity per 1000 individuals, P_{it} is the hospital charge per average stay, V_{it} is the vacancy rate, LX_{it} is total labor expense associated with a hospital stay, OIX_{it} is all other input expense associated with an average hospital stay, DON_{it} is total private donations, GG_{it} is total amount of federal grants to hospitals and INT_{it} is an interest rate.

Similar to the demand equation, quantity supplied

is a desired, not necessarily observed variable. Its measure represents the number of admissions physicians and administrators desire to make available per 1000 individuals.

As espoused in Chapter 3, physicians and administrators jointly determine a supply of hospital services. Administrators 'run' the day-to-day operations of a hospital. They appear to be the agent which decides the mix of non-physician inputs in the production of hospital services. However, administrators could not offer any service beyond that of a hotel without the added service of physicians. Physicians may perceive hospital services as a 'free' input into the production of their services and thus, have an incentive to replace medical procedures they provide with medical procedures hospitals provide.

Both supply-side agents are responsive in their decision-making to changes in hospital charges. Administrators, translating an increase in hospital charges into an increase in revenue, and interested in the objective of covering average cost per stay, have contradictory responses. On the one hand, they have an incentive to allow a larger number of admissions per period. But, a concurrent tendency due to a larger charge may encourage administrators to provide more services per admission. Newhouse (1970) hypothesizes that 'hospitals'

(he does not differentiate between types of supply-side agents), have a quality bias. Feldstein (1977) suggests this bias is a result of administrators giving into physicians' pressure for higher quality service when average revenues are rising. Whatever the reason, as long as the cost of providing a more sophisticated product is not sufficient to offset the impulse to increase admissions, an increase in hospital revenue per stay will encourage administrators to endorse an increase in admissions.

Physicians also have contradictory reactions to a potential increase in hospital charges. An increase in hospital charges may be perceived as a signal from hospitals that they will provide more services per admission. Physicians may interpret the increase as an opportunity to expand and diversify their practice. If this is the case, then physicians would advocate more admissions if for no other reason than to allow them to service their existing practice. On the other hand, if they perceive the increase in charges as a reallocation of total money expenditures on medical treatment which utilize hospital services, physicians may attempt to base their recommendations on procedures involving less hospital service. If this is the case, then physicians would not endorse the need for more admissions per period.

Despite some contrary forces, the probable net

effect of a change in hospital charges on desired available supply of hospital service, captured in coefficient, b_1, is hypothesized to be positive.

The vacancy rate, V_{it}, is included in Eq.3 to measure the response in admissions to variation in the vacancy rate, a proxy measure of demand tension, or lack of it, as the case may be, in the hospital service market and its subsequent effect on hospital admissions. Another response to variation in the vacancy rate, a response which is discussed below, is its effect on hospital charges. It will be interesting to see whether a change in the vacancy rate results in a greater variation in admissions or in hospital charges.

Even though the vacancy rate is a measure of empty beds or seemingly excess supply of hospital services, it is, in fact, an operational feature of hospitals not to operate at 100 percent capacity. One function of hospitals is to meet emergency demand for medical services. Emergency demand is demand that if not met would result in death or a considerable reduction in the well-being of a consumer. As a result, some hospital beds remain empty to enable the hospital to meet emergency demand. A second function is to provide a wide variety of specific hospital services. Within a hospital facility, there are many types of hospital 'beds' and these 'beds' are not perfect substitutes. This non-substitution of

beds may either a technical constraint, or, in some circumstances, a legal constraint. Demand for a specific service has some variation, hospitals operate at less than full capacity to accommodate the variation in demand for specific hospital services.

In each period there is some desired vacancy rate, V_{it}^*, which is exogenously determined in this model. Also, in each period there is an actual vacancy rate, V_{it}. This may be greater than, equal to, or less than the desired vacancy rate. In periods where the desired vacancy rate is greater than the actual vacancy rate, physicians and hospital administrators perceive excess demand for hospital services and they may react by adopting a more liberal admissions policy. Similarly, in periods where the desired vacancy rate is less than the actual vacancy rate, physicians and hospital administrators perceive, now, excess supply for hospital services and they may react by making it more difficult to be admitted to a hospital.

In reality, the desired vacancy rate is not known, therefore, the difference between it and the actual vacancy rate as a measure of excess demand cannot be computed. Clearly, the information on the number of empty beds relative to the total supply of beds is merely the actual ratio. So, it is included as some type of proxy measure of excess demand in Eq.3. Its coefficient, b_2, is

hypothesized to have a positive value. This is not unreasonable. Not only do researchers face a data constraint with respect to measuring the excess demand for hospital services, but physicians and hospital administrators have no accurate measure of excess demand. They, like the researcher, have the published vacancy rate and it is worth their while to interpret this statistic much the way it has been suggested above.

Both total labor expense per stay, LX_{it}, and other input expense per stay, OIX_{it}, measure the effects of variation in input costs on admission plans. Therefore, they will be treated together. On one hand, an increase in input costs associated with a day of care should result in a decrease in desired admissions. But, in the construction of either variable, quality of service is not held constant. An increase in labor expense could reflect better medically-trained personnel. Likewise, an increase in other expenses can reflect an improvement in medical technology. While administrators may remain neutral to such improvement, depending on the nature of the technological improvement, physicians may react prescribing more or less admissions. Again, due to contradictory responses to changes in a variable, the coefficients, in this case, b_3 and b_4, are indeterminable. However, if it is assumed that administrators are not concerned with medical improvements and, if either of the

coefficients is positive, then it may be concluded that cost changes were a result of medical advances and physicians were sensitive to such improved hospital services.

The final three variables, the state-specific debt interest rate, INT_{it}, the dollar amount of philanthropic funds, DON_{it}, and government grants to hospitals, GG_{it}, all are variables which help determine the expansion of physical stock associated with the hospital sector. An increase in physical stock can either be in the form of more beds or in a change in the number and/or scope of medical equipment located in hospitals. An increase in the number of beds or quantity of medical equipment will result in an increase in admissions per period. As in the previous case, a change in the scope of medical equipment has an ambiguous effect on admissions.

Since both private donations and government grants are 'free funds,' an increase in their availability will result in an increase in the number of beds, and consequently, admissions. As a result, the expected values for coefficients, b_5 and b_6, are positive. The interest rate associated with debt is, obviously a price. Any increase in it will induce a decrease in the growth of the bed stock and admissions. Hence, the expected value for b_7 is negative.

Adjustment Equations

Included in the model are two adjustment equations. One represents adjustment in hospital charges, the average charge for a hospital stay deflated by a state-specific consumer price index (P_{it}), and a second represents adjustments in the coinsurance rate, the ratio of total direct consumer hospital expenditures to total hospital expenditures ($COIN_{it}$). Not included is an adjustment equation for net hospital charges, the cost to the consumer at time of illness. The rationale for this omission is obvious: Since consumers' cost is a construct of hospital charges and the coinsurance rate, if movements in its components are explained then movement in it will be explained.

To emphasize their dynamic relationship, the two adjustment equations are considered together. An increase in hospital charges, or, maybe news of hospital technology advances, will be perceived by consumers as an increase in expenditure risk associated with any stay in a hospital. In order to reduce new risk, consumers will attempt to improve their hospital insurance coverage. The consequence will be a decrease in the average coinsurance rate. Depending on the extent of the improved insurance coverage, there will be some type of change in the relative price of hospital service. If the improved coinsurance rate results in a relative price decrease or

no change, consumers will either increase their utilization of hospital service or have no change in their use. Hospitals will be able to increase their daily charge. In the short-run, with constant average costs, hospitals will experience an increase in their per-unit returns. The higher return per unit of service will encourage physicians to press for a more sophisticated hospital service product, and for hospital administrators to give in to physician pressure for this change. This adjustment only justifies a further round of increases in hospital charges (Feldstein, 1973).

The story becomes even more complex when tax incentives associated with health insurance are introduced into the dynamics. Before 1983, the United States tax structure allowed both employers and employees to deduct health insurance premium payments form taxable income (Vogel, 1980). Thus, in periods of inflation, the potential of moving into higher tax brackets without a corresponding increase in real income will encourage employees to bargain for increases in their wage basket in the form of improved health insurance coverage, and since employers are able to deduct their premium payments form their income, employers readily agree to such a bargain (Feldstein and Friedman, 1977). A similar incentive exists in the market for individual health insurance policies. Regardless of whether people are considering a

group or a private insurance policy, an increase in the inflation rate will result in a decrease in the coinsurance rate, thus in the effective cost of hospital service to the consumer at time of illness.

As a consequence of the above dynamic relation, the following tow adjustment equation are included in the model:

$$(4)\quad P_{it} = c_0 + c_1(COIN_{it}-COIN_{it-1}) + c_2P_{it-1} + c_3V_{it} + U_{4it}$$

$$(5)\quad COIN_{it} = d_0 + d1(P_{it}-P_{it-1}) + d_2COIN_{it-1} + d_3Y_{it} + d_4PI_{it} + d_5TM_{it} + U_{5it}$$

If the above analysis proves to be correct, $COIN_{it}$ will have a negative effect on hospital charges ($c_1<0$) in Eq.4, while P_{it} will have a positive effect on the coinsurance rate ($d_1>0$) in Eq.5.

Note that each of the adjustment equations includes other variables which may affect the levels of hospital charges or the coinsurance rate. The hospital charge variable is affected not only by the interaction it has with the coinsurance rate but by its own lagged value, P_{it-1}, and a measure of excess demand, the vacancy rater, V_{it}. Similarly, the coinsurance rate is influenced, not only by changes in hospital charges but by its won lagged value, $COIN_{it-1}$, the price for hospital insurance, PI_{it}, and the marginal tax rate, TM_{it}.

Focusing on Eq.4, hospital charges' own lagged

value is included as a variable in the determination of hospital charges to allow for the possibility that adjustment of hospital charges is sluggish. While this is improbable, if it were to be the case, then P_{it-1}'s coefficient, c_2, will be positive in value.

Another variable included in Eq.4, the equation explaining hospital charges, is the vacancy rate,the ratio of empty beds to the total number of beds available for inpatient hospital services. As in Eq.3, it is included to pick up the demand tension, or lack of it in the hospital service market and its subsequent effect on hospital charges.

In each period, the actual vacancy rate may be greater than, equal to, or less than the desired vacancy rate. In periods where the desired vacancy rate is greater than the actual vacancy rate, physicians and hospital administrators perceive excess demand for hospital services and they may react by raising hospital charges. Similarly, in periods where the desired vacancy rate is less than the actual vacancy rate, physicians and hospital administrators perceive, now, excess supply for hospital services and they may react by decreasing hospital charges. The difference between the two rates ought to have a positive impact on hospital charges. Even though only the actual vacancy rate appears in Eq.4, its' coefficient, c_3, is hypothesized to have a negative value.

As mentioned above, the coinsurance rate, Eq.5, besides being a function of a dynamic relationship between itself and hospital charges, is explained by movements in its own lagged values and factors related to the demand for hospital insurance.

The effect of its own lagged value on its current value is a measure of the effect of habit. People with individual insurance policies are not likely to renegotiate their optimal insurance level from period to period. For one reason, it is too costly to undertake this transaction. Moreover, the extent of insurance coverage is a function of convention. Once a person gets used to a certain level of insurance expenditure and its subsequent effect on his or her health maintenance, this person is not likely to make large annual adjustments. Finally, the coinsurance rate is largely determined by membership in groups, especially employment groups and age groups. Insurers can offer policies more cheaply to groups since such policies avoid the costs association with adverse selection and moral hazard (Pauly, 1986). In addition, insurers will pass along some of the savings due to economies associated with managing a large group of people. But there are more significant attachments a person may have to a group than access to health insurance. If a person stays a member of the same group, from one period to the next, she is likely to purchase a

similar level of insurance as a matter of course.

And, since the coinsurance rate is the characteristic representing the depth of insurance coverage, all the reasons for inertia in the movement of insurance will be reflected by sluggish movement in the coinsurance rate. As a result, the coefficient, d_2, is expected to be large and positive.

As noted earlier, some common demand-for-insurance variables are included in the coinsurance rate equation. These variables are, specifically, real income, price of insurance and the marginal tax rate.

Obviously, if insurance is a normal good, an increase in real income ought to increase the demand for insurance. This change translates into the consumer paying a smaller fraction of each dollar of hospital expenditures, or the coinsurance rate will decline. If this is the case, then coefficient, d_3, will be negative. As any good, the magnitude will determine whether or not hospital insurance can, indeed, be categorized as a normal good.

The price of health insurance, the ratio of premiums paid to expected benefits, will influence the demand for insurance. If premiums paid relative to expected benefits increase, the cost of insurance is perceived to have increased, resulting in an increase in the coinsurance rate. Hence, the coefficient, d_4, is

expected to be positive.

Because of the United States tax codes, which exempted from both personal income and corporate income a large part of health insurance premium payments, any increase int he marginal tax rate will be interpreted as a reason to increase expenditures on health insurance, reflected by a decrease in the coinsurance rate. Subsequently, d_5, a measure of the responsiveness of the coinsurance rate to a change in the marginal tax rate, is expected to be negative.

It should be noted that since the coinsurance rate is constructed as a weighted average of coinsurance rates associated with various groups within the population which are likely to have different insurance depth, factors which are likely to change the distribution of people between these groups such as the unemployment rate and changes in age groups or portion of a state's population with Medicaid, are omitted as regressors. This is an unfortunate consequence of the method used in constructing the coinsurance rate variable. If a better measure of the coinsurance rate were available then these variables would be expected to have a behavioral effect on the coinsurance rate and, therefore, would be included in the equation.

In both adjustment equations, the particular lag structure is a result of experimentation rather than theory. Since theory will provide little insight into

speed of adjustment and patterns of lags, rather than relying on it, the final version for both equations will be the result of 'fitting' the data best.

Vacancy and Queue Rates Relationship

The derivation of the final equation, one which estimates excess demand for hospital services as a function of the vacancy rate, borrows a technique used by Rosen and Quandt (1987) in estimating excess labor supply as a function of the official unemployment rate. An AHA-published statistic is the vacancy rate, V_{it}. It does not indicate a vacancy rate which is optimal for the operation of a hospital needing space for emergencies and variation in hospital services, but reports the actual vacancy rate. An unknown statistic associated with the hospital service market is the queue rate. U_{it}. If desired quantity demanded is defined as $D_{it} = Q_{it}(1+U_{it})$, and desired quantity supplied is defined as $S_{it} = Q_{it}(1+V_{it})$, then the ratio of desired demand to desired supply is the following:

$$(1)' \quad D_{it}/S_{it} = (1+U_{it})/(1V_{it})$$

When logarithms are taken of (1)', then the equation becomes:

$$(2)' \quad \ln D_{it} - \ln S_{it} = \ln(1+U_{it}) - \ln(1+V_{it})$$

If U_{it} and V_{it} are fairly small, the right hand of Eq.2' can be expanded around zero, resulting in the following

Taylor approximation equation:

$$(3)' \quad \ln D_{it} - \ln S_{it} = U_{it} - V_{it}$$

Needless to say, there is no data on the queue rate associated with hospital services. However if the queue rate is assumed to be a stable function of the vacancy rate, no bigger an assumption than the labor market's vacancy rate being a stable function of the official unemployment rate, an assumption used by Rosen and Quandt, then it is possible to assume some specific hyperbolic form such as:

$$(4)' \quad U_{it} = K/V_{it}$$

where K is a parameter. Eq.4' implies there is a constant response by the queue rate to changes in the vacancy rate. While the response may or may not be constant in reality, it is reasonable to believe there is some systematic relation between the vacancy rate and the queue rate. As the vacancy rate decreases, physicians and hospital administrators probably become more selective in their admissions policy, resulting in an increase in the queue rate. Likewise, as the vacancy rate increases, physicians and hospital administrators become less selective in their admissions policy, resulting in a decrease in the queue rate. To ignore these admission changes would adversely affect income levels for the physicians and revenue levels for the hospital administrators.

Substituting Eq.4' into Eq.3' gives the model its

final equation:

(6) $D_{it}-S_{it} = K/V_{it} - V_{it} + U_{sit}$

Note, the objective of Eq.6 is to exploit information contained in the published vacancy rate in estimating excess demand for hospital services. Put differently, its objective is to measure excess demand using a reported statistic. Given that this measure depends on a number of heroic assumptions, Eq.6 must be treated as a stochastic equation rather than an identity (Rosen and Quandt, 1986). The estimated value of the parameter should provide a clue to whether the market is characterized by excess demand or not. If K is positive then the market on average is experiencing excess demand for hospital services.

A final word on the vacancy rate: It was mentioned earlier that the actual vacancy rate is not necessarily an optimal one. However, it appears that the vacancy rate does not vary significantly from year to year for a specific state. This is despite the increase in the average number of hospital beds. Therefore, it is reasonable to assume there is not much difference between the actual vacancy rate and the optimal one.

Table 4.1
Summary of Expected
Coefficient Values

Coefficient	Definition	Expected Value
a_1	dlnD/dlnNP	Negative[1]
a_2	dlnD/dlnT	Ambiguous[2]
a_3	dlnD/dlnY	Ambiguous
a_4	dlnD/dlnPCPHY	Negative
a_5	dlnD/dlnOPHY	Positive[3]
a_6	dlnD/dlnELD	Positive
a_7	dlnD/dlnED	Ambiguous
a_8	dlnD/dlnNW	Positive
a_9	dlnD/dlnG	Positive[4]
a_{10}	dlnD/dt	Ambiguous
b_1	dlnS/dlnP	Positive[5]
b_2	dlnS/dlnV	Positive
b_3	dlnS/dlnLX	Ambiguous[6]
b_4	dlnS/dlnIOX	Ambiguous[6]
b_5	dlnS/dlnDON	Positive
b_6	dlns/dlnGG	Positive
b_7	dlnS/dlnINT	Negative
c_1	dlnP/dln COIN	Negative
c_2	dlnP/dlnP_{t-1}	Positive
c_3	dlnP/dlnV	Negative[7]
d_1	dlnCOIN/dln P	Negative
d_2	dlnCOIN/dln$COIN_{t-1}$	Positive
d_3	dlnCOIN/dlnY	Positive

Coefficient	Definition	Expected Value
d_4	dlnCOIN/dlnPI	Positive
d_5	dlnCOIN/dlnTM	Negative
e_1	dln(D-S)/dlnV	Negative[a]

Notes

1. If $|a_1|<1$, then hospital admissions may be necessities.

2. If $|a_2|>|a_1|$, then time costs will appear to have more importance in the decision to be admitted into a hospital than money costs.

3. A positive value for a_5, while not conclusive evidence of physician induced demand, it would open up the possibility.

4. If variation in the proportion of women residing in a state is small, the variable, G, will not be included in estimation of the model.

5. While an increase in hospital charges should induce a more liberal admissions policy, there is some tendency toward quality.

6. If quality were held constant, then b_3 and b_4 would be unambiguously negative.

7. A correct sign would significantly aid in the argument of an excess demand for hospital services.

8. A negative value for e_1 is key in the argument the hospital sector is characterized by excess demand.

Chapter 5

Derivation of the Likelihood Function and Other Issues

Three issues must be resolved prior to the estimation of the parameters in the above system of equations. First, the model is suggesting the endogenous variables switch between two possible states of nature, one where there is excess demand and the other where there is excess supply, and when the switching occurs is unknown. This necessitates an estimation method which incorporates the probabilities of either state occurring. Second, since the model is a system of non-linear equations whose parameters are to be estimated simultaneously, necessary and sufficient conditions for identification of a non-linear system must be considered. Along with this, it is essential to determine the most appropriate method of estimating such models. Two possible estimation methods, full-information maximum likelihood and two-stage maximum likelihood, are considered in turn. And, finally, since the study is a pooled-time series, there is a significant likelihood that the model's disturbance terms are characterized by both serial correlation and heteroscedasticity. Therefore, the final topic of this section is a discussion of possible corrections in the estimation procedure.

Before beginning the discussion on the above issues, the model is restated for the purpose of simplifying the notation:

(1) $Q_{it} = \min(D_{it}, S_{it})$

(2) $D_{it} = a_o + a_1 COIN_{it} + a_2 P_{it} + Z_{2it} + U_{2it}$

(3) $S_{it} = b_o + b_1 P_{it} + b_2 V_{it} + Z_{3it} + U_{3it}$

(4) $P_{it} = c_o + c_1 COIN_{it} + c_2 P_{it-1} + C_3 V_{it} + Z_{4it} + U_{4it}$

(5) $COIN_{it} = d_o + d_1 P_{it} + d_2 COIN_{it-1} + Z_{5it} + U_{5it}$

(6) $D_{it} - S_{it} = K/V_{it} - V_{it} + U_{6it}$

All variables are expressed in log-linear form. Each Z_{jit} represents a vector of exogenous variables. Q_{it}, P_{it} and $COIN_{it}$ are observed endogenous variables, while D_{it}, S_{it} and $(D_{it}-S_{it})$ are unobserved endogenous variables.

Only five of the six equations are stochastic. Q_{it} is an endogenous variable; however, in this model, it is treated as deterministic. For the derivation of the estimation model, the following assumptions pertaining to the nature of the disturbance terms, u_{jit}, are adopted:

8. $U_{jit} \sim N(0,\Phi)$

9. $\Phi = E(U_{jit} U_{kit})$ $\quad j,k = 2\text{-}6$

10. $E(U_{jit} U_{jrt}) = 0$ $\quad r = i$

11. $E(U_{jit} U_{jig}) = 0$ $\quad t = g$

The implication of Eq.8 and Eq.9) is that the disturbance terms are jointly normal with zero means and a constant variance-co-variance matrix. From Eq.10 and Eq.11, it can

be ascertained that the disturbance terms exhibit neither the problem of serial correlation nor heteroscedasticity. However, no assumption is made prohibiting correlation between the equations' disturbance terms.

The preferred method of estimation for the above hospital sector model is maximum likelihood estimation (MLE). Since a likelihood function is derived from the joint probability function of the model's stochastic endogenous variables, MLE will assign values to the model's parameters which take into account the assumed characteristics of the disturbance terms. A central topic of this paper is the derivation of the likelihood function associated with the above model. In later works, the suggested likelihood function coefficients and statistical parameters will be estimated by employing an algorithm such as one associated the gradient method or the Marquardt method. By definition, an algorithm searches for values for unknown parameters by assigning them values which maximize the probability of observing the sample, in this case, a collection of hospital, health insurance and demographic data.

Derivation of the Likelihood Function:

By definition, any likelihood function is derived from the joint probability function of the disturbance

terms, u_{jit}. From Eq.8, each u_{jit} is normally distributed:

(12) $f(U_{jit}) = (2\pi\sigma_j^2)^{-\frac{1}{2}}\exp\{-\frac{1}{2}(U_{jit}/\sigma_j)^2\}$

Therefore, the joint density function will be:

(13) $f(U_{2it},\ldots,U_{6it}) = \prod_{j=2}^{6} f(U_{jit})$

As long as f(.) has an one-to-one correspondence with its respective stochastic endogenous variables, the joint probability function of the stochastic endogenous variables, $g(D_{it},S_{it},P_{it},C_{it},v_{it})$, will be the product of f(.) and a Jacobean transformation factor, A:

(14) $g(.) = A\, f(.)$, where

(14)' $A = |\partial U_{jit}/\partial Y_{it}|$, where Y_{it} is a vector of stochastic endogenous variables. In particular, given

$U_{jit} \sim N(0,\Phi)$;

(15) $g(.) = A(2\pi^3\sigma_2\ldots\sigma_6)\exp\{-\frac{1}{2}(\frac{Dit-ao-\ldots.}{\sigma_2^2} + \frac{Sit-bo-\ldots.}{\sigma_3^2} + \frac{Pit-co-\ldots.}{\sigma_4^2} + \frac{COINit-do-\ldots.}{\sigma_5^2} + \frac{Dit-Sit-K/Vit+Vit}{\sigma_6^2})\}$

with,

(15)' $A = 1 - b_2c_2 + KV_{it}^{-2}$

If each endogenous variable were observed, then the appropriate likelihood function would merely be:

(16) $L = \prod_{T\times N} g(.)$

Unfortunately, in a disequilibrium market model, it is not possible to assume that each stochastic endogenous variable is observed for each time frame. In particular, the quantity variables are truncated. In some periods, when there is excess demand, only S_{it} is observed. Similarly, in other periods, when there is excess supply, only D_{it} is observed. It is for this reason that OLS is an inappropriate method of estimation for disequilibrium market models - the truncation results in the expected value of the disturbance to be non-zero and for the disturbance terms and the exogenous variables to be correlated. The use of OLS estimation would result in biased and inconsistent estimates (Maddala, 1983). But, not only is the sample truncated, it is unknown as to when either situation has occurred, in other words, sample separation is unknown.

Hypothetically, if the change in hospital charges were positive then it could concluded that the market had experienced a period of excess demand. Conversely, if the change in hospital charges were negative then it can be concluded that the market had experienced a period of surplus. This can be expressed as a sorting rule and included in the model's design (Fair and Jaffee, 1972). But since the hospital sector is complex, with a number of other possible factors determining the direction of the

hospital charge change, to use a sorting rule would be premature.

Hence, another approach to the derivation of the likelihood function must be considered, one which does not rely on observed values for both quantity variables or a sorting rule, but is capable of predicting the probability of observing either quantity demanded or quantity supplied using the observed endogenous and exogenous variables.

A nuance of a likelihood function is that not only is it derived from a probability density function of the observed endogenous variables, but the density function is an unconditional one. For this case, it is $g(Q_{it},P_{it},C_{it},v_{it})$. Hence, the likelihood function, L', is

$$(17) \quad L' = \prod h(.)$$

The objective, now, is to explicitly state what is h(.). The derivation of h(.) is dependent on the notion that a unconditional probability function can be expressed as a sum of its conditional probability functions (Maddala and Nelson, 1974):

$$(18) \quad h(.) = Nh_1(Q,P,COIN,V,|Q=D) + (1-N)h_2(Q,P,COIN,V|Q=S)$$

where N is the P(D>S) and $h_1(.)$ is the probability density function of the observed endogenous variables conditioned on Q=D, while (1-N) is the P(D<S) and $h_2(.)$ is

the probability density function of the observed endogenous variables conditioned on Q=S. But, for the hospital sector model with the assumed joint normal distribution for the disturbance terms, the question begging to be addressed is what exactly is h(.).

The model can be considered to be a switching simultaneous equation system with two possible regimes, one where desired demand is less than desired supply, (Regime 1), and one where desired demand is greater than desired supply, (Regime 2) (Goldfeld and Quandt, 1975).

In Regime 1, $D_{it} = Q_{it}$ and S_{it} is unobserved, and the observed variables are Q_{it}, P_{it}, c_{it}, v_{it}, Z_{1it}, ...Z_{5it}. The hospital sector model during this regime is:

$$(2)' \quad Q_{it} = a_o + a_1 COIN_{it} + a_2 P_{it} + Z_{2it} + U_{2it}$$

$$(3)' \quad S_{it} = b_o + b_1 P_{it} + b_2 V_{it} + Z_{3it} + U_{3it}$$

$$(4) \quad P_{it} = c_o + c_1 COIN_{it} + c_2 P_{it-1} + c_3 V_{it} + Z_{4it} + U_{4it}$$

$$(5)' \quad COIN_{it} = d_o + d_1 P_{it} + d_2 COIN_{it} + Z_{5it} + U_{5it}$$

$$(6)' \quad D_{it} - S_{it} = K/V_{it} - V_{it} + U_{6it}$$

Thus, the density function associated with observing the endogenous variables <u>conditioned</u> on the fact that $D_{it} < S_{it}$ is:

$$(19) \quad h_1(Q,P,c,V|D_{it} < S_{it}) = \int g_1(Q,S,P,COIN,V)dS/N$$

where N, a normalizing factor, is the $P(D_{it}<S_{it})$.

Similarly, in Regime 2, $S_{it} = Q_{it}$ and D_{it} is not observed. While the same vector of variables is observed, in this regime, the model is:

$$(3)'' \quad Q_{it} = b_o + b_1P_{it} + b_2V_{it} + Z_{3it} + U_{3it}$$

$$(2)'' \quad D_{it} = a_o + a_1COIN_{it} + a_2P_{it} + Z_{2it} + U_{2it}$$

$$(4)'' \quad P_{it} = c_o + c_1 COIN_{it} + c_2P_{it-1} + c_3V_{it} + Z_{4it} + U_{4it}$$

$$(5)'' \quad COIN_{it} = d_o + d_1P_{it} + d_2COIN_{it-1} + Z_{5it} + U_{5it}$$

$$(6)'' \quad D_{it}-S_{it} = K/V_{it} - V_{it} + U_{6it}$$

The density function associated with observing the endogenous variables conditioned now on the fact that $D_{it}>S_{it}$, is:

$$(20) \quad h_2(Q,P,c,V|D_{it}>s_{it}) = \int g_2(Q,D,P,COIN,V)dD/(1-N)$$

where (1-N) is the $P(D_{it}>S_{it})$.

Hence, the unconditional density function, defined as the sum of two conditional density functions is:

$$(18)' \quad h(Q,P,COIN,V) = \int g_1(.)dS + \int g_2(.)dD.$$

As in the case where all endogenous variables which have

an one-to-one correspondence with their respective disturbance terms, and if $u_{jit} \sim N(0,\Phi)$, then

$$(18)'' \quad h(.) = (2\pi\sigma_2 \ldots \sigma_6)^{-1}\{A_1 \int \exp(-\tfrac{1}{2}\{\frac{Qit-\ldots.}{\sigma_2^2} + \frac{Sit-\ldots.}{\sigma_3^2} + \ldots. + \frac{Dit-Sit\ldots.}{\sigma_6^2}\})dS + A_2 \int \exp(-\tfrac{1}{2}\{\frac{Qit-\ldots.}{\sigma_3^2} + \frac{Dit-\ldots.}{\sigma_2^2} + \ldots. + \frac{Dit-Sit-\ldots.}{\sigma_6^2}\})dD\}$$

where

$$(21) \quad A_1 = \left|\frac{\partial U2it,\ldots,U6it}{\partial Qit,Sit,\ldots..}\right| = 1 - b_2c_2 + KV_{it}^{-2}$$

$$(22) \quad A_2 = \left|\frac{\partial U3it,\ldots,U6it}{\partial Qit,Dit\ldots...}\right| = 1 - b_2c_2 + KV_{it}^{-2}$$

As a result, the likelihood function is:

$$(17)' \quad L' = \prod_{T \times N} h(.) \quad \text{where } h(.) \text{ is from Eq. 18''.}$$

Two substantial issues are ignored in the derivation of the above likelihood function. One issue has already been mentioned, and will be dealt with in the pooling section found later in this chapter, and that is the fact that the

model assumes there is neither serial correlation nor heteroscedasticity.

The second issue is the question whether or not the two regimes are mutually exclusive. The derivation of the likelihood function has assumed the two states are mutually exclusive. If this was not the case, there would have to be a third term which eliminated the effect of both excess demand and excess supply existing simultaneously. Obviously, it makes no sense for there to be both excess demand and excess supply. Hence, for the two regimes not to be mutually exclusive introduces the notion of 'logical inconsistency.' Unfortunately, when an algorithm assigns values to the model's coefficients, unless it is included in the algorithm, it does not subject their assignments to a sense of economic logic. To prevent the occurrence of logical inconsistency, 'coherency conditions' should be assigned to the A_1 and A_2 before actual estimation occurs (Maddala, 1983). Since estimation is a topic beyond the intent of this paper, the necessary restrictions for A_1 and A_2 are not computed.

Nonlinear Identification and Estimation Issues

The above likelihood function (Eq.17') can be estimated either by a full-information maximum likelihood estimation (FIML) method or by a two-stage maximum likelihood (2SML) method. Which method is employed

depends on whether or not the above system can be identified and the statistical properties associated with each method when used in estimating the hospital sector model.

FIML has great intuitive appeal - it involves assigning values to each coefficient in the structural model simultaneously. However, since it does this by solving a system of nonlinear reduced-form equations, FIML can not be used unless the structural equations are exactly identified. If the model were linear in the variables and had only linear restrictions on the parameters then the usual rank and order conditions would apply. But, this is not the case. First of all, by nonlinear in the variables it is meant that at least one of the endogenous variables appear in the model in two or more different forms. In the supply equation, Eq.3, the vacancy rate appears as Vi_t, while in the excess demand equation, Eq. 6, the it appears as v_{it} and v_{it}^{-1}, making the model nonlinear in the variables. Secondly, a model is considered to be nonlinear in the parameters if there exist any nonlinear restrictions on the parameters (Goldfeld and Quandt, 1972). Turning again to Eq.6, the coefficient for v_{it} is required to be '-1', therefore characterizing the model as one which is nonlinear in the parameters.

As noted above, since the model is not linear in either the variable nor in the parameter, the usual rank and order conditions cannot be employed to determined whether or not the individual equations are identified. Specifically, the problem involves the rank condition, a necessary and sufficient condition for linear models. In a linear system, the rank condition guarantees that a member equation in the model cannot be expressed as a linear combination of the remaining equations. If and only if this condition is satisfied, will unique values for the equation's coefficients be obtained. Furthermore, only in the case where all equations satisfy this condition is the model considered to be an identified one. If a model includes a nonlinear equation, the rank condition must be such that when satisfied there exist no linear or nonlinear combination of the remaining equations that can be used to express a member equation (Goldfeld and Quandt, 1972).

While it may be more difficult to state the identification condition in an elegant manner as in the linear case, from a practical perspective, identification may be easier to obtain in the nonlinear case. It appears the rule of thumb is that "nonlinearities never hinder identification and they may help" (Goldfeld and Quandt, 1972) - at least for a local solution.

It is suggested that the first attempt at estimating Eq.17' is by FIML. If it does converge to unique values for the parameters after some reasonable number of iterations, then the model was either exactly identified or over-identified. In other words, the solution may represent either a global solution or a local solution. If the estimated coefficients conform with expected values, fine, the job is done. However, if they do not, hopefully, the estimated values represent a local solution and another solution, one which is both a global solution and one which conforms to economic theory pertaining to the model's coefficients is still out there, waiting to be found. At this point, it may be advantageous to add some rules or restrictions on the nature of the disturbance terms to the model (Goldfeld and Quandt, 1972). This restricted model should be rerun, again using FIML. If the computer is still not converging to appropriate values then some type of two-stage estimation on those equations which are identified may have to be employed.

The FIML estimator has good asymptotic qualities. In other words, as long as the nature of the disturbance terms in the model is as suggested in Eq.8 through Eq.11, then FIML provides consistent estimates of the models parameters. However, given the reasons stated above, a

FIML solution is unlikely. Therefore, estimation may require some notion of a two-stage method.

One suggested two-stage method begins by finding the maximum likelihood estimates of coefficients in the reduced-form versions of the supply and demand equations. If a regime is identified, then the estimated values for quantity demanded and quantity supplied, along with their respected variance and covariance matrices can be generated for each observation. Using the predicted amounts, a weighted value for quantity observed is computed and substituted back into the original model. Now having a value (though estimated value, in some cases,) for each and every endogenous variable, the model the is estimated using ordinary least squares (OLS).

The 2SLS approach has a number of problems. First, weights would need to be assigned to the estimated values for quantity demanded and quantity supplied before quantity observed could be computed. A priori, there is no theory which can determined what the weights ought to be. Second, the disturbance terms in the reduced-form equations for quantity demanded and supplied are correlated. As a result, estimates of their coefficients will be biased and inconsistent. Finally, it is recognized that the standard errors computed in the second-stage are incorrect. If they are used in hypothesis testing, acceptance or rejection decisions will

become erroneous. Hence, the estimated coefficients can not be compared to coefficients estimated in other models, nor can they be used in policy recommendation.

In conclusion, the advantage of FIML estimation - beyond an intuitive appeal - is that it is a consistent estimator. Unfortunately, it is doubtful whether or not it can be used to generate actual estimates. Disequilibrium models with unknown sample separation seldom converge to unique estimated values without the addition of specified relations between the variances of the model's disturbance terms. In addition, the problem of 'logical consistency' will require restrictions placed on the coefficients. The alternative two-stage method results in incorrect standard errors making inference erroneous. A model estimated using this approach has limited value. If one wanted to be pessimistic, one should pay heed to Maddala's words: "Until the problem[s] of estimating [disequilibrium models] with sample separation unknown are resolved, it is not worth undertaking estimating [such models]." (Maddala, 1983) But it is suggested that further research involve attempting both methods (and seeing if new methods of estimation have been suggested) and checking how their estimates of coefficients compare to non-disequilibrium models.

Pooling

Ideally, the testing of a disequilibrium hospital sector model would use data for a smaller unit than a state with observations collected over time. But problems of obtaining data for a smaller unit and not wanting to get involved with the implication of major structural changes that likely occurred in the hospital sector when Medicare and Medicaid were introduced in the mid-1960s and the Diagnostic Related Group program in 1983, required that the state be the unit of observation, and restricted the time frame to the 1971 through 1982 period. If a state were picked as the focus of study, at most, there would have been thirteen observations, and that is only if there were no lagged terms in the model. In this scenario, given the complexity of the model, any estimation method would not have sufficient information for estimation of values for the model's coefficient. Therefore, the reasonable approach is to 'pool' available information on all states of the U.S., which is done in this case.

And, like any second-best approach, it comes with a cost. Specifically, a pooled time-series model with no corrections for it being a time-series model, is making the heroic assumption that the model's coefficients are constant across time and across states. Unfortunately, due to reasons which will be given below, at least during

the first run of the model, the assumptions outlined in Eqs. (8) through (11) are still implied.

Three types of strategies are used to correct the consequence of pooling data: the covariance model strategy, an error component model strategy and a random coefficient model strategy. For this study, any adopted strategy must not result in the loss of many degrees of freedom or require a different structure for the disturbance terms. The objective of this section is to provide a rationale as to why each approach is inappropriate for this model. It should be noted, that an alternative model which requires the omission of one of the model's equation will be suggested at the end of this section.

The first strategy to be discussed is known as the covariance model approach. This strategy introduces dummy variables for both years and states in each equation. The idea is the estimated coefficients on the dummies will measure the average change in the dependent variables due to the observation belonging to a specific state or year. In other words, the introduction of state and year dummies will account for the average effects of states and time on the endogenous variables. The addition of dummy variables adds no insight as to why membership to a state or year will cause a shift in value of a dependent variable. But even more importantly, the covariance model

reduces the number of degrees of freedom by ([N*T-T-N]G where G is the number of equations in the model.), jeopardizing the efficiency of the model, something which can not be afforded in a disequilibrium model with sample separation unknown. Therefore, the dummy variable approach is rejected as a correction for pooling data.

A second strategy (the error component model) requires new assumptions for the structure the disturbance terms. Specifically, it is suggested in a pooled time-series the disturbance terms will correlated across states and time, i.e. $u_{jit} = v_{ji} + w_{jt} + z_{jit}$, where w_{jt} is assumed to follow an autoregressive scheme and the variance of v_{ji} is not a constant 'Pindyck and Rubenfeld, 1981). This would be the preferred strategy since, in fact, it is quite probable the disturbance terms do not adhere to the nice assumptions outlined in Eq.8 through Eq.11. But, problems with serial correlation and heteroscedasticity are purposefully ignored in disequilibrium models. Allowing serial correlation, for instance, except in special circumstances, in equations involving unobserved variables will result an 'intractable' likelihood function (Rosen and Quandt, 1986). Therefore, at best, it could not be included in Eq.3' and Eq.6' in Regime 1 and Eq.2'' and Eq.6'' in Regime 2 (Rosen and Quandt, 1981). The consequence of heteroscedastic disturbance terms on the derivation of a

likelihood function has not even been investigated. Regardless, it would result in a much more complicated likelihood function, one which may have no chance of being estimated. Hence, the error component strategy is rejected. But even if this were not the case, if the model was a simple linear simultaneous equation model with an occasional lagged dependent variable, the error component could not be used. A basic requirement for implementing this model structure is that the model involves but one equation and that equation contains no lagged dependent variables as explanatory variables.

The random coefficient model is the final strategy to be discussed. By allowing the coefficients to vary by state, the model is inferring that different states have different responses to changes in explanatory variables. Specifically, any coefficient in the model, B_i, is comprised of a constant, B, and a random variable, W_i. By definition, W_i is a linear function of a state-specific variable and a random component whose nature is similar to that of the disturbance term, U_{jit}. The estimated coefficient is a weighted average of the state-specific coefficients, where the weights are a function of the estimated variance of the disturbance term U_{jit} and an estimate of the variance associated with the coefficient. As a result, to estimate the model's parameters requires

estimation of each coefficient's variance. The number of estimated parameters actually doubles (Maddala, 1977).

The advantages of the random component model are twofold: First, the loss in degrees of freedom in its estimation process is considerable less than in the covariance model case. Second, it is not necessary to adopt a different set of assumptions concerning the disturbance terms.

Unfortunately, the advantages are accompanied by problems. The theoretical econometricians have only extended the random coefficient model to the case of linear simultaneous models. So, it is unclear whether the above model, with its nonlinearities and latent endogenous variables, stays tractable if coefficients are allowed to have a random component. But, even ignoring this potential problem, a new twist is added to the identification problem with the introduction of random coefficients. Not only must the coefficients be identified, but the coefficients' variances must be identified. Consequently, a condition must be satisfied before the usual rank and order conditions for identification are employed (Kelejian, 1974).

If a correction for pooling the data is going to be made (ignoring the problem may be the best path), the random coefficient method appears to be the only viable approach. But, before implementing this strategy, it is

recommended that the hospital sector model be altered. Specifically, the excess demand equation (Eq.6) should be omitted from the model. If this adaption is made then the model is linear and as long as the condition specified by Kelijian (1974) is satisfied, the model is reducible and the random coefficient model can be adopted.

A final word of cautions, for all the reasons mentioned above, two-stage estimation methods should still be avoided. But, to do so requires each regime in the hospital sector model to contain just linear equations and for all equations to be identified. The reward for this case, is the capability of using a full-information maximum likelihood procedure on the system of random coefficient equations.

Chapter 6

Conclusions

A common thread throughout the paper is that spiraling hospital cost inflation can be explained by a market structure which is persistently in a state of excess demand for hospital services. In Chapter 3, a scenario was developed which gives a rationale for such a claim. The scenario revolves around an argument that physicians, like any economic agent, if given the opportunity to induce demand for their services, and their services sometimes involve hospital services, they will do so. The nature of hospital care provides physicians with such opportunity. With influence on both sides of the market, physicians' desires will have some effect on the final market structure of the hospital sector. Consequently, whether or not the hospital sector has excess demand will depend on the net benefit associated with such a market when compared to one with the typical 'market-clearing' characteristics.

The focus in Chapter 4 is on outlining a general sector hospital market which is not confined to either a market equilibrium or disequilibrium framework. The model itself explains six variables: observed admissions into inpatient hospital service, desired demand, desired

supply, hospital charges, the coinsurance rate and excess demand (or supply).

Given the model developed, Chapter 5's purpose was to present the derivation of the appropriate maximum likelihood function. Since there are two potential states of nature, one characterized by excess demand and one characterized by excess supply, the likelihood function has the form of a bivariate probit search function.

Left for future research is the actual estimation of the likelihood function. But the road to estimation is not free from obstructions. Some substantial problems remain before reliable estimates of the hospital sector model can be made. Furthermore, for the model to be applicable beyond the time frame of the purposed sample, a few adaptations should be made.

For instance, a forgotten agent in the model is the insurer. An implicit assumption operative in the model is that the insurance agent is a passive participant - supplying any requested amount of insurance demanded. This may or may not have been a legitimate assumption for the period under study, 1970 through 1982, a period of not only hypothesized excess demand for hospital services, but also surging increases in the demand for health services.

But, clearly since many of the structural changes in the provision of hospital services have been initiated by health financing institutions, i.e., reimbursement

mechanisms have been altered, and physician sovereignty in recommendation of services has been challenged with the adoption of the DRG program, equations describing the incentives for such institutions ought to be included in the design of the model. This is especially true if testing of the model is to extend beyond 1982, the last year before the imposition of a national DRG program. However, this omission does not prevent testing of the model as developed in this paper for the suggested time frame.

As reported both in Chapter 3 and the Appendices, some of the measures on variables are dubious. An obvious example is that the 'price' of a hospital stay omits many of the additional costs incurred by a consumer. A true price would be not only a function of hospital charges but also all physician fees associated with the possible stay. Another example, as noted by the many critics of the excess demand scenario, is that use of physician density to determine whether a medical service market has this feature, may be circular. In particular, the critics correctly point out that a positive influence of an increase in physician density may not only indicate the ability of physicians to shift demand for their services, but may simply indicate a movement of physicians to locales with more medically needy populations. While notably an excuse, the variables in the model are often

constructed in the manner suggested because there are no better alternatives. This is especially true, given the necessity to pool over states in order to have a large enough sample.

If the model is to be used to test a more recent sample, some additional variables may need to be included. For instance, the portion of GNP allocated to health services has continuously grown during the mid-1980s, while simultaneously, inpatient utilization has declined. This must mean people are substituting toward alternative medical services such as nursing homes, home health care agencies and ambulatory care centers. If this is the case, then price variables associated with these substitutes must be included in the model.

But even accepting the above-mentioned shortcomings, estimation of the model still requires the completion of a few tasks. First of all, the logical consistency restrictions must be specified. Second, a package capable of estimating a bivariate probit search function using a full-information technique must be found. And, even if these two tasks are complete, there is no guarantee that any algorithm will be able to generate stable estimates for the model's coefficient. Since no inference can be drawn for any coefficient estimated via a two-stage estimation technique, some full-information technique is the recommended method of estimation.

A possible solution to an outcome of non-convergence of the suggested model, is to drop the excess demand equation. It is this non-linear equation which introduces many of the estimation problems. But, if this is done, then the likelihood function must be re-derived and each of the remaining equations must meet the identification requirements for a linear system. Maybe Maddala's advise of not attempting to estimate the function, at least in the short run, should be heeded.

Regardless of whether the model described in the paper is correct or whether it can be estimated, health economists need to examine the complex interrelations between the many economic agents in the hospital sector. Not to do so will keep us in the dark about proper policy for a sector which provides an essential service.

Appendix A

Data

To estimate the model requires that hospital insurance and demographic data be collected on all fifty states and the District of Columbia for the period 1970 through 1982. This section briefly defined each variable in the model and indicates the source of the data.

All hospital-related data was obtained from the American Hospital Association's annual publication, Hospital Statistics. Health insurance variables were constructed using data published by the Health Insurance Association of American in their Source Book of Health Insurance Statistics. Since their data set was not adequate in providing information for all insurance variables, various publications on the National Health Care Expenditure Studies was also employed. Data pertaining to physicians was obtained from the American Medical Association's Physician Characteristics and Distribution in the U.S. While most hospital, insurance, and physician variables were constructed using these four sources, various series held by the National Bureau of Economic Research supplemented most of the missing information.

It should be noted that in a substantial number of cases, data on a variable could only be found by state for

a year or two. If this was the only available option, this data will be used in future estimation of the hospital sector.

Table A.1
Sources and Definitions
of Variables

D^1	Desired number of admissions per 1000 persons.	Unobserved
NP^2	Average hospital charge per stay less the average copayment.	AHA and see App. B.
T	Time cost of an average stay.	AHA and NBER
Y	Average total income (earned and unearned) per 1000 persons.	NBER
PCPHY	Number of primary care physicians per 1000 persons.	AMA
OPHY	Number of non-primary care physicians per 1000 persons.	AMA
ED	Number of persons with at least a high school education per 1000.	NBER
NW	Number of non-white persons per 1000.	NBER
G	Number of women per 1000.	NBER
t	Time	-
S^1	Desired admissions policy per 1000 persons by supply agents.	Unobserved
P^1	Averge hospital charge per stay.	AHA
LX	Wage expenses associate with an average stay.	AHA
IOX	All other expenses associated with an average stay.	AHA
V^1	Ratio of empty beds to total available beds.	AHA
DON	Total donations.	
GG	Total Federal grants to	

	hospitals.	
COIN[1]	Average portion of a dollar spent on hospital services.	See App. B.
TM	Average marginal tax rate.	NBER
D-S[1]	Excess demand (or supply) for admission into a hospital.	Unobserved

1. An endogenous variable. All other variables are exogenous.

2. All variables which are measured in dollars have been deflated by a state specific CPI.

Appendix B

Coinsurance Rate Construction

The coinsurance rate, the ratio of the total net hospital charges paid directly by the consumer to the total gross hospital charge is constructed as a weighted average based on nine possible hospital insurance categories. Specifically, it is:

$$c_{it} = \Sigma U_s \lambda_{ts} P_{sit} / \Sigma \lambda_{ts} P_{sit}$$

where U_s is the utilization rate per category, λ_{ts} is the annual average coinsurance rate associated with a category and P_{sit} is the annual portion of a state's population associated with an insurance category.

There are nine insurance categories and they depend on the age and types of insurance coverage: A person younger than age 65 can have group hospital insurance, individual hospital insurance, be eligible for Medicaid or have no hospital insurance coverage. A person older than age 65 can have only Medicare, Medicare and private hospital insurance, only hospital insurance, Medicare and Medicaid, or have no hospital insurance. (Unless it is noted otherwise, all statistics are from NMCES Data Previews.)

The utilization rate is the average of 'at least one admission per year' people within an insurance category have experienced. Because of data limitations, it only varies by insurance category, not by time or by

state. For people younger than age 65, utilization rate estimates were calculated by Aday and Andersen (1978). Using data from the 1976 Health Expenditure Survey, they estimated that 10 percent of the population with either individual or group hospital insurance experienced at least one hospital admission per year. For people eligible for Medicaid, the utilization rate was estimated to be 16 percent, and for people with no insurance, the utilization rate was estimated to be 9 percent. For elderly people, the utilization rate was estimated from NMCES of 1977. It found a utilization rate which is about twice that of younger people, 20.1 percent. While there is a sundry of insurance categories for the elderly, there is but one utilization rate used in the construction of the coinsurance rate in this paper.

The number of people under age 65 in each state with private hospital insurance policies and the number with Medicaid is reported annually by the Health Insurance Association of American (HIAA). The HIAA, however, does not break down the private hospital insurance number into group purchased policies and individually purchased policies. Since depth of coverage differs substantially between these two categories of coverage, it is important to separate the two categories in computing the state-specific coinsurance rate. The NMCES does report that for 1977, 90.4 percent of all people under age 65 with private

health insurance had group health insurance with the remaining 9.6 percent holding individual health insurance policies. The distribution between private group and individual insurance is assumed to remain constant over the period of the study. While these percentages are for 'health' insurance, it is not a concern. NMCES also reported that among people under age 65 with private health insurance, 98.7 percent of them carry hospital insurance. Therefore, the above percentages for people with group and individual health insurance policy are expected to represent the percentages of people with private and group insurance.

People with Medicaid either carry a Medicaid card or participate in a cash welfare program such as AFDC or SSI. One hundred percent of those with Medicaid eligibility are assumed to have hospital insurance coverage. Whether they do have hospital coverage at time of illness, depends on Medicaid approval of their hospital admission. However, given the standards a person must meet to be eligible for Medicaid, if the admission was not approved, the hospital would probably carry such a person's hospital bill as a bad debt.

The number of people under age 65 with no hospital insurance is computed as a residual - it is the number of people under age 65 less those with private hospital insurance, less those eligible for Medicaid. The estimate

of the uninsured probably contains an upward bias. At least some of the people with no hospital insurance, if they did required hospitalization would then apply for Medicare and be accepted. Technically, these people ought to be classified as part of those eligible for Medicaid, but it is not possible to guess what this number may be. Therefore, these people are included among the uninsured.

The HIAA also reports annually the number of elderly people with Medicare in each state. But this is it. The HIAA provides no data on the number of elderly with just private health plans, private plans and Medicare, or Medicare and Medicaid. In fact, it provides no information o the private health policies held by the elderly. The NMCES, however, does provide enough information to compute the portion of the elderly population (and the total population) which fall into these various categories. Unfortunately, their data is based on a national survey and is for just one year (1977). As a result, the proportions computed using the NMCES data are assumed to not vary over time or by state. The same logic is used in estimating the number of elderly who are uninsured.

Again, much of the data is in reference to 'health' rather than 'hospital' insurance. But this is even less significant for the elderly. If an elderly person has a private health insurance policy, there is a 98.9 percent

probability it is a hospital insurance policy. Plus, 92.2 percent of all Medicare enrollees have Part A - the hospital insurance component. It should be noted that many of the categories pertaining to the elderly represent such a small percentage of the total population, that they will not influence the computation of the annual state-specific coinsurance rate (holding only private health insurance - .17 percent, holding both Medicare and Medicaid - .35 percent, holding no health insurance - .08 percent.)

In all the above cases, a 'portion,' P_{it}, of the state population for each year is computed. It is the ratio of the number of people placed in a category to the state's population and it is the only component of the coinsurance rate which consistently varies by both state and over time.

For each of the categories that represent at least one percent of the total population, an average coinsurance rate is computed. (Thus, the coinsurance rate is an average of five insurance groups.) Since the only annual data that is available on the hospital coinsurance rate is a national ratio of direct consumer expenditures on hospital services to total national hospital expenditures, and no information exists on a state level, either annually or even for one year, each coinsurance rate is computed by piecing together information form the

NMCES. As a result, the coinsurance rate is assumed to vary only among the categories. (A slight exception is in the computation of the relevant coinsurance rate for people with only Medicare.) While it is true that for the period under study, a period with high inflation and significant tax incentives to substitute improved health coverage for money wage increases, it still may not be such a bad assumption. There is some evidence that when people acquire 'better' health coverage it has meant that they acquired coverage for health services that they had not previously covered rather than improving on the depth of their coverage on a health service they already had some coverage. For instance, during the 1970s, the number of people with dental insurance, an unusual type of insurance prior to 1970, increased significantly, while the number of people which moved from an 80 percent copayment to something better did not change significantly (Feldstein, 1973).

Hospital insurance, in general, is differentiated by varying amounts od deductibles, copayments, ceilings and exemptions. The coinsurance rate, however, is used as an indicator of the depth of coverage associated with insurance policies. In ths study, for people under age 65, the private hospital insurance coinsurance rate takes into account the copayment factor associated with basic coverage plans, Major Medical plans and Health Maintenance

Organizations (HMOs), and the distribution of policyholders between these plans.

It is assumed that the copayment factor for HMOs is 100 percent; for any plan which includes a Major Medical component, it is 80 percent; and the copayment factor associated with basic coverage plans was found by setting the weighted average of copayment factors for all plans in 1977 and equating this with the 1977 national ratio of total direct consumer expenditures to total hospital expenditures. According to the NMCES, the distribution for group health insurance policies among the three categories was the following: 4.9 percent of the policies were for HMO plans, 81.7 percent were Major Medical plans and 13.4 percent were basic coverage plans. For individual health insurance policies, the distribution was the reverse: 2.6 percent of the policies were in the form of HMOs, 34.5 percent were Major Medical plans while 62.9 percent were basic coverage plans.

Obviously, the coinsurance rates for group health insurance will differ substantially from private health insurance. Again, because of data limitations, the coinsurance rates for these groups is computed once for the nation and, consequently, they are not allowed to vary by state or over time.

As noted above, the hospital coverage for Medicaid-eligible people is total. Hence, the estimated coinsurance

rate is zero. At the other extreme, people with no coverage experience a 100 percent coinsurance rate. This is an exaggeration. As mentioned earlier, some people with no insurance, if ill, would find themselves eligible for Medicaid. In addition, some will simply receive hospital services without paying for them. In both cases, the coinsurance rate for this portion of the uninsured would be zero, not 100 percent. By ignoring these possible events, the coinsurance rate for the uninsured is overstated.

The remaining two insurance groups both include elderly with Medicare coverage: One group represents elderly with just Medicare and a second group representing elderly with Medicare and private health insurance. People with Medicare receive total coverage for hospital expense after paying a deductible equal to the cost of one day of hospital service for the first 60 days of a hospital stay. Beyond 60 days, the elderly are subject to a schedule of copayments with increase with length of stay. For the purposes of constructing a coinsurance rate associated with an average stay in a hospital for recipients of Medicare, it is not necessary to include in its' estimate the increasing coinsurance rate schedule. Even for the elderly, whose average hospital stay is longer than the average stay of people younger than age 65, the average stay does not approach 60 days.

Therefore, it is unnecessary to consider the copayment schedule for stays longer than 60 days.

For elderly people with only Medicare coverage, the coinsurance rate is the ratio of an average cost of one bed-day of hospital service to the gross cost of an average stay. Since the AHA does provide annual information on state-specific hospital expenditures and the average stay, the coinsurance rate for this group will vary by state and over time.

Private health insurance for the elderly has the characteristic of filling in the no-coverage gaps of Medicare. As it is, the only hospital expense they face is the cost associated with the first day of a hospital stay. The NMCES of 1977 found that 92.9 percent of the elderly with private insurance had full coverage for the cost of the first day of hospital stay Medicare deductible. The remaining group had some copayment factor for this expense. Assuming this copayment is 80 percent, it is possible to estimate an average copayment of 98.58 percent, or a 1.42 percent coinsurance rate for the Medicare deductible. Together with the Medicare coverage, elderly in this category experience approximately a zero coinsurance rate.

Bibliography

Aday, LuAnn and Ronald Andersen. "Insurance Coverage and Access: Implications for Health Policy." Health Services Research. Vol. 13, Winter 1978. p. 369-377.

American Hospital Association. Hospital Statistics. AHA. Chicago. Issues 1972-1983.

_____. Hospitals: Guide Issue. AHA. Chicago. 1971.

Arrow, Kenneth J. "Uncertainty and the Welfare Economics of Medical Care." American Economic Review. Vol. 53, December 1963. p. 941-973.

Barro, Robert J. and Chaipat Sahasakul. "Average Marginal Tax Rates from Social Security and Individual Income Tax." NBER Working Paper No. 1214. October 1983.

Cromwell, Jerry and Janet B. Mitchell. "Physician-Induced Demand for Surgical Operations." Journal of Health Economics. Vol. 5, 1986. p. 293-333.

Eeckhoudt, L., T. LeBrun and J.C. Sailly. "Risk Aversion and Physicians' Medical Decision-Making." Journal of Health Economics. Vol. 4, 1985. p. 273-281.

Ehrlich, Isaac and Gary S. Becker. "Market Insurance, Self-Insurance and Self Protection." Journal of Political Economy. Vol. 80, July 1982. p. 623-648.

Fair, R.C. and D.M. Jaffee. "Methods of Estimation for Markets in Disequilibrium." Econometrica. Vol. 40, 1972. p. 497-514.

Feldstein, Martin S. "Quality Changes and the Demand for Hospital Care." Econometrica. Vol. 45, 1977. p. 1681-1702.

_____. "The High Cost of Hospital Care, and What to Do About It." The Public Interest. Vol. 48, Summer 1977. p. 40-54.

_____. "The Welfare Loss of Excess Health Insurance." Journal of Political Economy. Vol. 81, March 1973. p. 251-280.

_____. "Hospital Cost Inflation: A Study of Non-Profit Price Dynamics." American Economic Review. Vol. 61, December 1971. p. 853-872.

_____. "The Rising Price of Physicians' Services." Review of Economics and Statistics. Vol. 52, May 1970. p. 121-133.

Feldstein, Martin S. and Bernard Friedman. "Tax Subsidies, the Rational Demand for Insurance and the Health Care Crisis." Journal of Political Economy. Vol. 85, April 1977. p. 155-178.

Fisher, Franklin M. The Identification Problem in Econometrics. McGraw-Hill. New York. 1966.

Fuch, Victor R., Robert T. Michael and Sharon R. Scott. "A State Price Index." NBER Working Paper No. 320. February 1979.

Goldfeld, S.M. and R.E. Quandt. Nonlinear Methods in Econometrics. North-Holland Press. London. 1972.

Green, Jerry. "Physician-Induced Demand for Medical Care." Journal of Human Resources. Vol. 13, Supplement, 1978. p. 21-32.

Grossman, Michael. The Demand for Health: A Theoretical and Empirical Investigation: Occasional Paper No. 119. NBER. New York. 1972.

Hay, Joel, and Michael J. Leahy. "Physician-Induced Demand: An Empirical Analysis of the Consumer Information Gap." Journal of Health Economics. Vol. 1, 1982. p. 231-244.

Health Insurance Association of American. Source Book of Health Insurance Data. HIAA. Washington. 1970-1983.

Hey, John D. Economics in Disequilibrium. New York University Press. New York. 1981.

Kelejian, H. H. "Random Parameters in a Simultaneous Equation Framework: Identification and Estimation." Econometrica. Vol. 42, May 1974. p. 517-527.

Kmetta, Jan. Elements of Econometrics. Macmillan. New York. 1971.

Maddala, G. S. Econometrics. McGraw-Hill, Inc. New York. 1977.

_____. Limited Dependent and Qualitative Variables in Econometrics. Cambridge University Press. New York. 1983.

National Center for Health Services Research. Private Insurance and Public Programs: Coverage of Health Services, Data Preview 20. U.S. Department of Health and Human Resources. Washington. March 1985.

_____. Private Health Insuance: Premium Expenditures and Sources of Payment: Data Preview 17. U.S. Department of Health and Human Resources. Washington. November 1983.

_____. Employer and Employee Expenditures for Private Health Insurance: Data Preview 7. U.S. Department of Health and Human Resources. Washington. June 1981.

Newhouse, Joseph P. "Toward a Theory of Non-Profit Institutions: An Economic Model of a Hospital." American Economic Review. Vol. 60, March 1970. p. 64-73.

Newhouse, Joseph P. and Charles E, Phelps. "New Estimates of Price and Income Elasticities." in The Role of Health Insurance in the Health Services Sector. ed. Richard Rosett. NBER. New York. 1976.

Pauly, Mark V. "Taxation, Health Insurance and Market Failure in the Medical Economy." Journal of Economic Literature. Vol. 24, June 1986. p. 629-675.

_____. "Policy Lessons from Studying Hospital Costs." Canadian Journal of Economics. September 1985. p. 22-26.

_____. Doctors and their Workshop. University of Chicago Press. Chicago. 1980.

_____. "The Economics of Moral Hazard." American Economic Review. Vol. 58, June 1968, p. 281-291.

Pauly, Mark V. and Michael Redisch. "The Not-For-Profit Hospital as a Physician Cooperative." American Economic Review. Vol. 63, March 1973. p.87-99.

Phelps, Charles E. and Joseph P. Newhouse. "Coinsurance, the Price of Time and the Demand for Medical Services." Review of Economics and Statistics. p. 334-342.

Quandt, Richard E. "Autocorrelated Errors in Simple Disequilibrium Models. Economic Letters. Vol. 7, 1981. p. 55-61.

_____. "Tests of the Equilibrium vs. Disequilibrium Hypothesis." International Economic Review. Vol. 19, June 1978. p. 435-452.

Quandt, Richard E. and Harvey S. Rosen. "Unemployment, Disequilibrium and the Short-Run Phillips' Curve: An Econometric Approach." Financial Research Center Memorandum, No. 61. Princeton University. Princeton. June 1985.

Robinson, James C. and Harold S. Luft. "The Impact of Hospital Market Structure on Patient Volume, Average Length of Stay and the Cost of Care." Journal of Health Economics. Vol. 4, 1985. p. 333-356.

Rosen, Harvey S. and Richard E. Quandt. "Estimation of a Disequilibrium Aggregate Labor Market. Review of Economics and Statistics. p. 371-379.

Sloan, Frank A. "Hospital Rate Review: A Theory and an Empirical View." Journal of Health Economics. Vol. 3, 1984. p. 83-86.

_____. "Regulation and the Rising Cost of Hospital Care." Review of Economics and Statistics. Vol. 63, November 1982. p. 479-487.

Sloan, Frank A. and Bruce Steinwald. "Effects of Regulation on Hospital Costs and Input Use." Journal of Law and Economics. Vol. 23, April 1980. p. 81-109.

Vogel, Ronald J. "The Tax Treatment of Health Insurance Premiums as a Cause of Overinsurance." National Health Insurance: What Now, What Later, What Never. ed. Mark V. Pauly. American Enterprise Institute. Washington. 1980.

Waldo, Daniel R., Katharine R. Levit and Helen Lazenby. "National Health Expenditures, 1985." Health Care Financing Review. Vol. 8, Fall 1986. p. 1-21.