ABSTRACT

Baumol's Cost Disease and Physician Shortages:
An Analysis of Rising Healthcare Expenditures from the Supply Side

Qian Wu, M.S.Eco.

Mentor: Mark C. Kelly, Ph.D.

Over the last two decades, the U.S. experienced a stagnant supply of physicians, as well as rapidly rising healthcare expenditures. Based on a novel version of Baumol's unbalanced growth model, this paper addresses the relationship between physician supply and healthcare expenditures. Applying a fixed-effect estimation on a panel data set consisting of 50 U.S. states over 2008-2016, we confirm the existence of Baumol's cost disease in the healthcare sector. Besides, a negative correlation between the growth of relative physician supply and unit healthcare cost is found, which according to the theory model, implies a less-than-one substitution elasticity between physicians and non-physicians. Followingly, a translog production function is estimated using seemingly unrelated regression, obtaining a magnitude of the elasticity of substitution which is roughly 0.23. Our study shows that due to the weak substitution between two groups of healthcare workers, the physician shortage can harm healthcare productivity substantially, which further cause the healthcare price to rise dramatically.

Baumol's Cost Disease and Physician Shortages:
An Analysis of Rising Healthcare Expenditures from the Supply Side

by

Qian Wu, B.Eco.

A Thesis

Approved by the Department of Economics

____

Charles North, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Master of Science in Economics

Approved by the Thesis Committee

____

Mark Kelly, Ph.D., Chairperson

____

James West, Ph.D.

____

Steve Bradley, Ph.D.

Accepted by the Graduate School

May 2019

____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank Dr. Mark Kelly for his great mentorship. I would also like to thank the other members of my committee, Dr. James West and Dr. Steve Bradley, for their support and suggestions. To Baylor admissions office and graduate school, thank you for allowing me to become a part of this excellent program. And finally, I would like to thank my family members for all the love and care you have been giving to me during the abroad study.

CHAPTER ONE

Introduction

In 1980, a report by the Graduate Medical Education National Advisory Committee concluded that the number of medical school positions, as well as that of international medical graduates entering the United States, should be restricted due to a predicted surplus of physicians. This conclusion was supported by a series of reports between 1992 and 1998 published by the national Council on Graduate Medical Education. Despite little indicative evidence, it was widely accepted that the surplus of

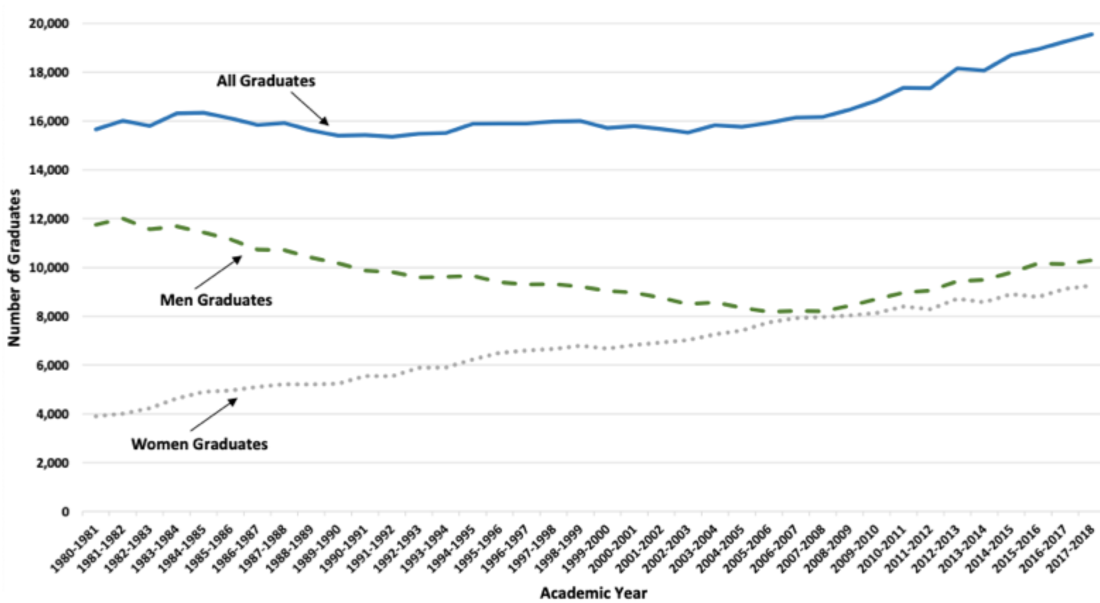

Figure 1.1 Graduates of U.S. Medical Schools. Source: AAMC.

physicians would arise by the end of the last century[1]. Influenced by such expectations, Medicare began capping graduate medical education funding. As a result of these caps, the number of medical school graduates experienced almost no change during the period of 1980-2000. This fact is depicted in Figure 1.1.

Throughout this time period healthcare expenditures in the United States rose rapidly. Illustrated by Figure 1.2, the share of healthcare expenditures to GDP in the United States has been increasing dramatically from 8.23% in 1980 to 17.15% in 2017. Due to the rapid increase in the demand for healthcare, the expected physician surplus never materialized. Instead, fears of physician surpluses have been replaced by concerns of a massive physician shortage. According to a recent study by the Association of American Medical Colleges (AAMC) concludes that by the year 2030, the nation will face a shortage of physicians sizing between 42600 and 121300[2].

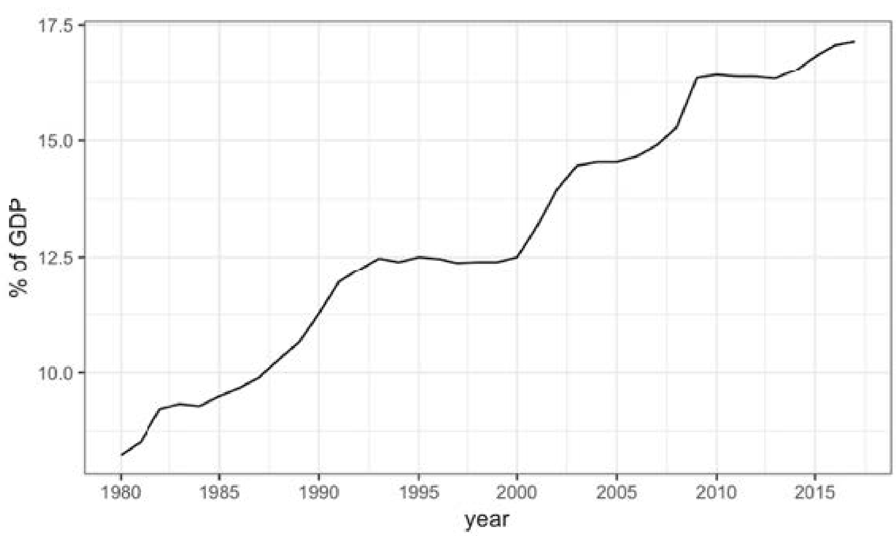

Figure 1.2. Share of Healthcare Expenditure in GDP.

[1] See Salsberg and Forte (2002).

[2] See AAMC Workforce Studies DATA SNAPSHOT (May 2018).

Our study is theoretically based on Baumol (1967), who explains the rapid growth of healthcare expenditures as a consequence of unbalanced productivity growth between the healthcare sector and the rest of the economy. We modify the original model by using a Constant Elasticity of Substitution (CES) production function for the healthcare sector, dividing healthcare laborers into physicians and non-physician, or "auxiliary workers", and allowing the elasticity of substitution between the two types of workers to vary. The first purpose of this paper is to empirically test for the existence of Baumol's cost disease in the healthcare sector. In order to do so, we build on the empirical strategy developed by Hartwig (2008) and expanded by Colombier (2012) and Bates and Santerre (2012) to include the ratio of physicians to auxiliary workers. The estimation results suggest that the healthcare industry in the U.S. does experience Baumol's cost disease. What's more, a larger relative physician supply tends to slow down the growth of healthcare expenditures. These two findings collectively imply that a larger ratio of physicians to auxiliary workers significantly improves the productivity of the healthcare sector, which should be reflected in the CES production function by the substitution elasticity between the two groups of laborers being smaller than one. Therefore, the second half of our work is designed to test whether this implication is true. A translog production function is used as a second-order Taylor approximation for the CES production function. Applying a Seemingly Unrelated Regression (SUR) strategy, we find that the substitution elasticity between physicians and auxiliary workers is roughly 0.23. In other words, the two types of workers are strongly complementary, implying that nurses, nurse practitioners, and physician assistants are generally not able to undertake the tasks requiring highly specialized skills that physicians process. Therefore, if the projected physician shortage materializes, healthcare providers will hire more non-physician labor to support existing physicians, reducing average labor productivity in the healthcare sector as well as the physician to auxiliary worker ratio, resulting in faster growth of nominal healthcare expenditures.

The next chapter provides a review of the literature with regard to the determinates of healthcare expenditures and Baumol's cost disease. In the third chapter, the modified Baumol's model is introduced in detail. Chapter four describes our empirical methodology, which presents the fixed effect model and translog production function estimated in our analysis, as well as the data used in regressions. The fifth chapter presents empirical results and analysis. The final part makes conclusions, pointing out policy implications, restriction of our work, and future improvement.

CHAPTER TWO

Literature Review

Determinants of Healthcare Expenditures

There have been numerous studies to investigate the rising healthcare expenditures from the demand-side. Most studies identify the income level as the main driver, though they hold different views on whether medical care is a luxury good. Relevant research starts with Newhouse (1977). Newhouse used data from 13 developed countries in 1970 and found that the income elasticity of medical care demand was between 1.15 and 1.31, implying that medical care is luxury. Likewise, Leu (1986), Parkin et at. (1987), and Brown (1987) used data of different OECD countries to conduct cross-national studies and obtained similar magnitudes of income elasticity of healthcare service demand[1].

On the other hand, Hitiris and Posnett (1992) used a panel data set and obtained a close-to-one income elasticity, suggesting that healthcare is more necessary[2]. Hall and Jones (2017) contribute to the literature by explaining why a higher income level causes a rising share of healthcare spending in GDP. They conclude that as people become wealthier, the marginal utility of consumption falls since the consumption rises. However, the utility of spending on medical care service to extend life does not decrease. As a result, a bigger share of income shifts toward healthcare.

---

[1] The estimated elasticity by Leu (1986), Parkin et al. (1987), and Brown (1987) are 1.18-1.36, 1.12-1.18, and 1.39 respectively.

[2] The estimated elasticity by Hitiris and Posnett (1992) is 1.026, with a confident interval (p=0.95) of 1-1.05.

Medical progress is suggested to be another significant cause of soaring healthcare spending. For instance, Newhouse (1992) analyzed healthcare expenditures in the U.S. from 1960 to 1987 and found a strong positive correlation between expenditures and medical technologies. A similar result has been given by Okunade and Murthy (2002). Developing a stochastic life-cycle model, Fonseca et al. (2013) concluded that the technological progress on its own explains 30% of the rise in healthcare spending. Frankovic et al. (2017) presented an overlapping generations model to address the impacts of medical innovation. They showed that a medical innovation that increases the remaining life expectancy at age 20 by 1.1 years, results in a 12.2% increase in the healthcare expenditure per capita.

Several studies attribute the rising healthcare expenditures to determinants other than income level and medical progress, such as age structure, territorial decentralization, and election activity[3]. Nevertheless, as stated in the beginning, past literature focuses on the analysis of the demand side, leaving the supply shocks unstudied. Thus, our study is a pioneer in addressing the healthcare spending problem from the supply side, as we suggest that the relative supply of physicians is one of the significant explanatory variables in affecting healthcare expenditures.

Baumol's Cost Disease

Baumol (1967) presents an unbalanced growth model that is adopted as the theoretical foundation of this paper. Baumol assumes that there are two sectors that grow at disparate rates. Labor serves as the sole input in each sector. Firms in each sector compete for workers in a single labor market. In order to compete with the "progressive"

[3] See Blomqvist and Carter (1997), Koening et al. (2003), and Potrafke (2010).

sector, “nonprogressive” firms must pay the market clearing wage rate which, in equilibrium, is set equal to the marginal product of labor in the progressive sector. As a result, the wage rate will grow at the same rate as labor productivity in the progressive sector, causing the unit cost in the nonprogressive sector to rise. If the demand for nonprogressive goods and services are sufficiently price inelastic, then nonprogressive firms will respond by raising their prices to keep up with the increase in their costs. This is the so-called “Baumol’s cost disease”. Healthcare is considered a typical nonprogressive industry by Baumol, implying that it can suffer from the Baumol’s cost disease.

While most of the literature has focused on demand-side causes of rising medical expenditures, supply-side effects such as Baumol’s cost disease may also play a significant role in explaining the persistent rise in the cost of healthcare. Testing for the existence of cost disease in the healthcare sector has proven challenging, given the difficulty of obtaining reliable medical prices[4]. However, in a recent study, Hartwig (2008) introduced an innovative methodology to empirically test for the presence cost disease in the healthcare sector. Hartwig concludes that if Baumol’s model is correct, we should observe that the increase in healthcare prices is driven by the wage growth in excess of the overall productivity growth. This methodology works particularly well on cross-country studies since it requires no use of the healthcare price index, which has been acknowledged to suffer from measurement issues. Hartwig (2008) applies a fixed effect analysis on an unbalanced panel data set consisting of 19 OECD countries. He defines the gap between the growth rate of nominal wages and the labor productivity as

[4] Boskin et al. (1996), Triplett (1999), Berndt et al. (2000), and Newhouse (2001) suggest that the medical care price index in the U.S. is substantially upward-biased.

the “Baumol variable” and obtains a coefficient for that variable very close to one, which is supportive to the previous assumption that the wage growth in excess of the productivity growth affects the medical care price proportionally.

Subsequent work by Colombier (2012) and Bates and Santerre (2012) correct a flaw in Hartwig’s empirical strategy. In short, the one-to-one causal relationship between the wage growth in excess of productivity growth and the rise in healthcare spending describes only a special case where the share of healthcare employment in total labor force approaches to one. In response, Colombier generates an “adjusted Baumol variable” that is equal to the original Baumol variable multiplied by the inverse of the share of healthcare sector in the total labor force. Guided by the modified version of Hartwig’s methodology, Colombier conducts the empirical test using an unbalanced panel data set including 20 OECD countries for the period from 1965 to 2007. Controlling for other possible drivers of healthcare spending such as income level, age structure, and medical progress, the results given by fixed effect models confirm the existence of the Baumol’s cost disease. The estimated coefficient for the adjusted Baumol variable is between 0.16 and 0.2.

In addition to cross-country investigations, Bates and Santerre (2012) employ the Colombier (2012) strategy to test for the existence of Baumol’s cost disease for the healthcare industry in the U.S. with a balanced panel data set of 50 states over the 1980-2009 period. Like previous scholars, Bates and Santerre use a fixed effect model and do find a positive correlation. However, Bates and Santerre expand on Hartwig’s and Colombier’s work by utilizing a two-stage least squares (2SLS) technique to confirm the correlation between the adjusted Baumol variable and the U.S. health expenditures is a causal relationship.

In summary, there are several previous studies that attribute the rapidly rising healthcare expenditures to the existence of Baumol's cost disease. However, most studies end up with obtaining a significant and positive coefficient, without taking the next step to examine what could be the reason for relatively slower productivity growth in the healthcare industry. In response to this concern, this paper first retests the Baumol's cost disease for the healthcare industry in the U.S., then takes a further step to examine the correlation between the change rate of relative physicians supply and the growth rate of healthcare spending, as well as the substitution elasticity between the physicians and auxiliary workers. The relationship between the physician supply and healthcare costs has not been widely studied. The only relevant work is done by Koening et al. (2003), who find a positive correlation between the supply of physicians, specialists in particular, and the unit healthcare cost. However, this study has two limitations. First, in their regression, the number of physicians is highly correlated with another explanatory variable, technological progress[5]. Thus, suffering from a large variance, the estimated coefficient may have implausible magnitude or even a wrong sign. Additionally, their study didn't exclude the possibility of reverse causality, a rapidly rising healthcare can prompt more M.D. students and therefore increase the overall supply of physicians.

Intuitively, one might argue that the strict residency caps stated above could help reduce the cost of healthcare, as training medical school residents is acknowledged to be costly. Limiting the physician supply then forces the healthcare industry to hire more lower-cost laborers such as nurse practitioners, physician assistants, and technicians.

[5] See Koening et al. (2003, p41) for more details.

However, many physicians are highly specialized, and in most states are the only ones who are able to diagnose conditions, write prescriptions, and perform surgeries, which cannot be done by non-physicians. Thus, it is more appropriate to assume that physicians and non-physician workers are weak substitutes or even complements. Therefore, when medical residency caps are applied, and a physician shortage is created, the healthcare sector will be less productive, driving up the price of medical service.

CHAPTER THREE

Theory Model

Baumol (1967) suggests that the economy can be divided into a progressive and a non-progressive sector, with labor being the sole input in each sector. Output in the progressive sector, or what we refer to the final goods sector, can be written as follow:

$$Y_t = Ze^{rt}L_t \tag{1}$$

Where L_t is the labor force engaged in the progressive sector at time *t*. Note that the marginal labor productivity in the progressive sector is Ze^{rt}, implying that labor productivity in the final goods sector grows at the rate *r* over time.

Following Baumol (1967), we assume that the healthcare sector is the non-progressive sector, which is characterized by very occasional productivity growth. Different from Baumol (1967), our paper employs a CES production function for the non-progressive sector.

$$H_t = \left[\alpha\left(B_{P,t}P_t\right)^{\frac{\theta-1}{\theta}} + (1-\alpha)\left(B_{A,t}A_t\right)^{\frac{\theta}{\theta-1}}\right]^{\frac{\theta}{\theta-1}} \; ; \; \theta \geq 0 \tag{2}$$

It can be shown by equation (2) that the labor inputs in the healthcare sector are divided into two groups, physicians, P_t, and auxiliary workers, A_t. The parameter α represents the income share of physicians, sizing between 0 and 1. $B_{P,t}$ and $B_{A,t}$ correspond to the productivity of physicians and non-physicians respectively. Baumol (1967) suggests that healthcare, as a typical non-progressive industry, features a negligible productivity growth. To make the analysis simple, we further assume that $B_{P,t}$

and $B_{A,t}$ are constants. The elasticity of substitution between physicians and auxiliary workers, θ, measures the level of easiness for the physicians and auxiliary workers to be replaced by each other. A smaller-than-one θ implies the physicians and auxiliary workers act as complements, while a larger-than-one θ implies that physicians and auxiliary workers are substitutes. The magnitude of θ is of particular interest to us because it can affect the overall labor productivity in the healthcare sector. To take a closer look, we rewrite the healthcare production function as:

$$H_t = B_t M_t \tag{3}$$

The rewritten healthcare production function formulates the real healthcare output as the product of average labor productivity, B_t and the total labor force in the healthcare sector, M_t. Note that by assumption, $\frac{\dot{B}_t}{B_t} < r$. What's more, we generate μ_t as the ratio of physicians to auxiliary workers to provide the convenience of analyzing the effect of the relative supply of physicians. By substituting P_t and A_t by $\mu_t M_t$ and $(1 - \mu_t)M_t$ respectively, the average healthcare labor productivity B_t then can be rewritten as follows:

$$B_t = \left[\alpha(B_P \mu_t)^{\frac{\theta-1}{\theta}} + (1-\alpha)(B_A(1-\mu_t)^{\frac{\theta}{\theta-1}}\right]^{\frac{\theta}{\theta-1}} \tag{4}$$

Equation (4) implies that if physicians are more productive than auxiliary workers ($B_P > B_A$) and the elasticity of substitution between physicians and auxiliary workers is sufficiently low ($\theta < 1$), then a bigger share of physician workers (μ_t) will induce a higher level of average healthcare labor productivity.

Note that equation (1) and equation (2) collectively imply that the ratio of real healthcare output to real final goods output is equal to $\frac{B_t M_t}{Ze^{rt}L_t}$. If the price elasticity of demand is relatively large for both sectors, then this ratio decreases towards zero, suggesting the healthcare sector tends to vanish over time. However, this ratio can be kept constant if we assume the demand for healthcare goods is highly price-inelastic. This case can be written as follows:

$$\Gamma = \frac{ZH_t}{B_0 Y_t} = \frac{B_t M_t}{B_0 e^{rt} L_t} \tag{5}$$

Note that the parameters Z and B_0 denote the initial levels of labor productivity for the final goods sector and healthcare sector respectively. Rewriting $\frac{ZH_t}{B_0 Y_t}$ as $\frac{H_t}{B_0} / \frac{Y_t}{Z}$, the parameter Γ measures the initial ratio of healthcare labor force to final goods labor force. Rearranging equation (5) to solve for M_t as a function of L_t yields equation (6) below:

$$M_t = \left(\frac{\Gamma B_0 e^{rt}}{B_t}\right) L_t \tag{6}$$

Let N_t be the total labor supply on the market such that $N_t = L_t + M_t$, we can derive both L_t and M_t as functions of N_t by substituting for M_t using equation (6).

$$L_t = \left(\frac{B_t}{B_t + \Gamma B_0 e^{rt}}\right) N_t;\ M_t = \left(\frac{\Gamma B_0 e^{rt}}{B_t + \Gamma B_0 e^{rt}}\right) N_t \tag{7}$$

Furthermore, Baumol (1967) assumes that the average wage rate in healthcare sector tends to rise at the same speed as in the final goods sector. According to equation (1), the uniformed wage rate in the economy then can be written as $w_t = we^{rt}$, with the wage rate goes up in line with final goods productivity. Keeping in mind that labor is the only input in production, the unit costs for both healthcare and final goods sectors are derived as below:

$$C_{Y,t} = \frac{w_t L_t}{Y_t} = \frac{w}{Z}; C_{H,t} = \frac{w_t M_t}{H_t} = \frac{we^{rt}}{B_t} \tag{8}$$

Equation (8) describes the so-called Baumol's cost disease, where the unit cost remains unchanged in final goods sector but soars over time in healthcare sector since r is assumed to be much bigger than $\frac{\dot{B}_t}{B_t}$. If we further assume that $\frac{\dot{B}_t}{B_t} = 0$, then the growth rate of unit cost in the healthcare sector is equal to the growth rate of final goods productivity. Hartwig (2008) figures out an empirical methodology to test Baumol's theory under these assumptions. He suggests that the growth of unit cost in the healthcare sector is directed proportional to the growth of average wages in excess of the aggregate labor productivity, which Hartwig refers to as the "Baumol variable".

However, the assumption that the labor productivity of healthcare sector experiences no growth can be problematic. Later studies by Colombier (2012) and Bates and Santerre (2012) allow the healthcare sector to have a positive growth rate that is much smaller than that of the final goods sector. They demonstrate that Hartwig's test deals with a special case where the share of healthcare employment converges to unity. In the general case, unit cost growth should instead be proportional to the adjusted Baumol variable which is equal to Hartwig's Baumol variable weighted by the inverse of aggregate healthcare employment share. If cost disease exists in the healthcare industry, then the coefficient for the adjusted Baumol variable is positive.

We update Colombier (2012) and Bates and Santerre (2012) theory by replacing the single-type healthcare workers with two different groups: physicians and non-physicians. As shown earlier in this chapter, if physicians and non-physicians are weak substitutes or complements, then a physician shortage will cause a decrease in healthcare productivity, resulting in an acceleration of healthcare costs.

To test this theory, we will first derive the adjusted Baumol variable. Let Q_t be real GDP, then the aggregate labor productivity is calculated as follows:

$$q_t \equiv \frac{Q_t}{N_t} = \left(\frac{Z + \Gamma B_0}{B_t + \Gamma B_0 e^{rt}}\right) B_t e^{rt} \tag{9}$$

Differentiating q_t with respect to time and dividing by q_t, we obtain the growth rate of aggregate labor productivity:

$$\frac{\dot{q}_t}{q_t} = \frac{\dot{B}_t}{B_t} + r - \frac{\dot{B}_t + r\Gamma B_0 e^{rt}}{B_t + \Gamma B_0 e^{rt}} \tag{10}$$

As the growth rate of average wage rate equals *r*, the unadjusted Baulmol variable is given by:

$$\frac{\dot{w}_t}{w_t} - \frac{\dot{q}_t}{q_t} = r - \left(\frac{\dot{B}_t}{B_t} + r - \frac{\dot{B}_t + r\Gamma B_0 e^{rt}}{B_t + \Gamma B_0 e^{rt}}\right) = \frac{M_t}{N_t}\left(r - \frac{\dot{B}_t}{B_t}\right) \tag{11}$$

We will next derive the growth rate of unit cost in healthcare sector. Using the second half of equation (8), we obtain:

$$\frac{\dot{C}_{H,t}}{C_{H,t}} = r - \frac{\dot{B}_t}{B_t} = \frac{N_t}{M_t}\left(\frac{\dot{w}_t}{w_t} - \frac{\dot{q}_t}{q_t}\right) \tag{12}$$

Equation (12) shows that the healthcare unit cost growth is linked to the overall wage growth over the aggregate productivity growth weighted by the inverse of healthcare labor share (i.e. the adjusted Baumol variable). Like Colombier (2012) and Bates and Santerre (2012), we will use it as the foundation of the first part of the empirical analysis in the next chapter. Note that in our version, the healthcare cost will also be affected by μ_t, the physician share of healthcare labor, since it influences the level of productivity of the healthcare sector.

CHAPTER FOUR

Empirical Methodology

Testing for Baumol's Cost Disease

Empirical Model

In this section, we will demonstrate our empirical strategy of testing for the existence of Baumol's cost disease in the U.S. healthcare sector. Our analysis uses a balanced panel data set of 50 U.S. states over 2008-2016. From equation (12), our empirical model is expressed as follows:

$$\Delta \log(HCE_{s,t}) = \beta_0 + \beta_1 \left(\frac{N_{s,t}}{M_{s,t}}\right) [\Delta \log(w_{s,t}) - \Delta \log(q_{s,t})] + \beta_2 \Delta \log\left(\frac{P_{s,t}}{A_{s,t}}\right) + \gamma \Delta \log X_{s,t} + \sigma_s + \tau_t + \varepsilon_{s,t} \quad (13)$$

Note that in equation (13), the s and t subscripts specify the state and time for variables respectively. The left-hand side variable represents the growth rate of unit nominal healthcare expenditures. $\frac{N_{s,t}}{M_{s,t}}$ denotes the inverse of healthcare labor share, and $[\Delta \log(w_{s,t}) - \Delta \log(y_{s,t})]$ denotes the wage growth over the average productivity growth. Note that the product $\left(\frac{N_{s,t}}{M_{s,t}}\right) [\Delta \log(w_{s,t}) - \Delta \log(y_{s,t})]$ corresponds to the adjusted Baumol variable. Different from previous literature, we include the change in the ratio of physicians to non-physicians as one of the explanatory variables as well, which is denoted by $\Delta \log\left(\frac{P_{s,t}}{A_{s,t}}\right)$. Our analysis focuses on the effect of two specific explanatory variables; the adjusted Baumol variable and the ratio of physicians to non-physicians. Specifically, a positive β_1 justifies the existence of Baumol's cost disease in the healthcare sector, and a negative β_2 implies that nominal healthcare expenditure growth will be aggravated if the projected physician shortage is realized. Note that in our model, the biggest gains from increasing the physician labor supply ratio will occur when the elasticity of substitution between physicians and non-physicians is less than one. Thus, a negative β_2 also implies a weak substitution relationship between physicians and non-physicians.

To isolate the effect of the adjusted Baumol variable and the ratio of physicians to non-physicians, controlling for other possible drivers of healthcare cost is necessary. Thus, we create a vector $X_{s,t}$ to include these potential drivers. Besides, state fixed effects, σ_s, and time fixed effects, τ_t, are also included in the estimation equation. The state fixed effects capture all unobservable state-specific elements that affect the growth of healthcare spending. On the other hand, the time fixed effects are intended to control for time-specific variations that are common to all states. Finally, $\varepsilon_{s,t}$ represents the error term.

The control variables contained in vector $X_{s,t}$ are chosen according to past literature. As stated earlier in chapter 2, income level and age structure have been widely identified as drivers of healthcare expenditures. Thus, we include real GDP per capita[1] and the percentage of people over 65 years old in vector $X_{s,t}$. In addition, we also include the labor union coverage rate, the unemployment rate, and the poverty rate[2] to work

[1] We differ from Bates and Santerre (2012) by replacing nominal GDP per capita with real term, as the price element has already been captured by the wage term. Regression results using nominal GDP as a control are included in Appendix.

[2] The share of people whose income is below federal poverty line.

together as a proxy for the insurance coverage rate, which is thought to be a significant driver by Newhouse (1992).

Data

We utilize data from several sources. The dependent variable, the unit cost of healthcare goods, is calculated by dividing total nominal healthcare expenditures by total population for each state year. The Census Bureau provides data for population estimates. Two different measures of the total healthcare expenditures are adopted, the Gross Domestic Product (GDP) and the Personal Consumption Expenditures (PCE). The healthcare expenditures are calculated as the sum of three components: Ambulatory Healthcare Service, Hospitals, and Nursing and Residential Care Facilities. All data is collected from the Bureau of Economic Analysis (BEA), available at state level from 1997 to 2017.

The computation of the adjusted Baumol variable requires data of total employment, healthcare employment, total nominal wage rate, and total real GDP, and are all obtained from BEA. Consistent with healthcare expenditures, healthcare employment, $M_{s,t}$, is composed of workers in three industries, Ambulatory Healthcare Service, Hospitals, and Nursing and Residential Care Facilities. The average nominal wage rate, $w_{s,t}$, is measured by dividing total nominal wages by the total employment. We use total compensation instead of total wages and salaries as our estimate of the average wage, since wages and salaries alone do not include bonuses or benefits, which will be reflected in the cost to employers. Likewise, the economy-wide real output per worker, $q_{s,t}$, is computed as total real GDP divided by total employment, including the self-employed.

To our best knowledge, the credible state-level data of physician numbers is not publicly available from any source. Therefore, we utilize the American Medical Association (AMA) Physician Masterfile which provides current and historical data for all of the more than 1.4 million physicians in the United States. We purchased AMA data for the period 2007-2016 through Medical Marketing Services, Inc. The number of non-physicians, $A_{s,t}$, is then calculated as the difference between healthcare employment and physician workers.

Table 4.1. Descriptive Statistics for Testing Baumol's Cost Disease

VARIABLES	(1) mean	(2) sd	(3) min	(4) max
Growth of healthcare PCE per capita	0.0357	0.0170	-0.00720	0.0804
Growth of Healthcare GDP per capita	0.0374	0.0214	-0.00804	0.113
Adjusted Baumol variable	0.184	0.257	-0.973	1.487
Growth of real GDP per capita	0.00255	0.0263	-0.0990	0.179
Growth of unemployment rate	0.00786	0.206	-0.315	0.725
Growth of poverty rate	0.00778	0.0552	-0.143	0.190
Growth of union coverage rate	-0.0129	0.108	-0.384	0.398
Growth of old percentage	0.0226	0.0107	-0.0109	0.0597
Growth of ratio of physicians to non-physicians	-0.00195	0.0130	-0.0697	0.0361
Number of states	50	50	50	50

Number of Observations: 450.

The control variables include GDP per capita, the percentage of people over 65 years old, the labor union coverage rate, the unemployment rate, the poverty rate, and the labor union coverage rate, all in growth rate form. The calculation of GDP per capita involves dividing total GDP by total population. State-level data for the old-aged population rate and the poverty rate are obtained from the Census Bureau. The labor union coverage rate, compiled from the Current Population Survey, is provided by the

Union Membership and Coverage Database.[3] The only remaining variable, unemployment rate, is obtained from the Bureau of Labor Statistics. Descriptive statistics of all variables used in testing Baumol's Cost Disease can be found in Table 4.1.

Estimating Substitution Elasticity Between Physicians and Non-physicians

Empirical Model

In section 4.1, we briefly discussed the significance of the elasticity of substitution between physicians and auxiliary workers in determining the magnitude of the effect that the physician supply has on the growth rate of nominal health expenditures. If the elasticity of substitution is sufficiently low, then increasing the relative supply of physicians will have a large, positive effect on the average labor productivity in the healthcare sector. Therefore, to better understand the potential effect of a physician shortage on the growth rate of nominal healthcare expenditures, we estimate the substitution elasticity between physicians and non-physicians.

In the empirical analysis, instead of using the original CES production function, we take the logarithms of equation (2) and adopt the second-order Taylor Approximation about the point $lnP = lnA = 0$ to obtain a translog production function. In fact, the translog function has a more general form, and it can be viewed as an approximation for any arbitrary twice-differentiable production function[4]. We proceed to write our translog function as follows:

$$lnH = \alpha_0 + \alpha_P lnP + \alpha_A lnA + \frac{1}{2}\delta_{PP}(lnP)^2 + \frac{1}{2}\delta_{AA}(lnA)^2 + \delta_{pA} lnPlnA \qquad (14)$$

[3] http://www.unionstats.com/

[4] See Griffin and Gregory (1976) for more details.

The α's and δ's in equation (14) are parameters that need to be estimated. Note that the derivation of equation (14) involves assuming the symmetry of second derivatives, $\frac{\partial^2 lnH}{\partial lnP \partial lnA} = \delta_{PA} = \frac{\partial^2 lnH}{\partial lnA \partial lnP} = \delta_{AP}$. Differentiating equation (14) logarithmically with respect to inputs and applying the assumption of perfectly competitive input and output markets under Constant Returns to Scale (CRS), we obtain the cost share functions for *P* and *A*:

$$S_P = \frac{\partial lnH}{\partial lnP} = \alpha_P + \delta_{PP} lnP + \delta_{PA} lnA \tag{15}$$

$$S_A = \frac{\partial lnH}{\partial lnA} = \alpha_A + \delta_{AA} lnA + \delta_{PA} lnP \tag{16}$$

The parameters S_P and S_A stand for the cost share of physicians and non-physicians respectively. Note that the cost shares should sum up to one, so that $S_P + S_A$=1. Of primary policy interest is the elasticity of substation between physicians and non-physicians. Berndt and Wood (1975) suggest the Allen partial elasticity of substitution can be computed as follows:

$$\sigma_{PA} = \frac{\delta_{PA} + S_P S_A}{S_P S_A} \tag{17}$$

Note that in equation (17), S_P and S_A are cost shares, the dependent variables in equations (15) and (16), and δ_{PA} is a parameter that needs to be estimated. Equations (14), (15), and (16) form the basis of our empirical analysis. Consistent with past literature, we estimate the equation system using Seemingly Unrelated Regression (SUR) strategy since the errors may be correlated with each other. To make the estimation operational, some theoretical restrictions have to be applied. As stated above, the symmetry of the second-order derivative has to be imposed, besides, the cost shares should sum up to one. In addition, implied by the unit-sum cost shares, the changes in cost shares in response to an input change must sum up to zero. The restriction system can be written as follows:

$$\delta_{PA} = \delta_{AP} \tag{18}$$

$$\alpha_P + \alpha_A = 1 \tag{19}$$

$$\delta_{PP} + \delta_{PA} = 0;\ \delta_{AA} + \delta_{AP} = 0 \tag{20}$$

Equations (14)-(20) form the basis of our empirical work. Note that equations (16) will be dropped in the estimation, because having all restrictions applied, either one equation alone includes all information, in other words, the parameters in one equation can be derived by those in the other.

Data

A panel data set that includes 50 U.S. states during of period of 2008-2016 is used in this section. The estimation of equation (14) requires data of real healthcare production and the numbers of physicians and non-physicians. Real healthcare output, *H*, is obtained from BEA. Consistent with the first part of our empirical work, we restrict healthcare production to three industries: Ambulatory Healthcare Service, Hospitals, and Nursing and Residential Care Facilities. The source and calculation for numbers of physicians and non-physicians have been stated in detail in the previous section.

Note that in equation (15), the physicians' cost share is calculated as the share of physicians' wages in total healthcare wages, $S_P = \frac{w_p * P}{Total\ Health\ Wages}$. Hence, estimating equation (15) requires data of the average wage rate for physicians, total nominal healthcare wages, as well as the numbers of physicians and non-physicians. As far as we know, the credible state-level data of average physicians' income is not available from any source. As a result, we approximate using the data from the Occupational

Employment Statistics (OES) program, which is produced by the BLS. Restricted by considerable missing data, we proxy for average physician wages utilizing two specific occupation categories, "Family and General Practitioners" (FGP) and "Physicians and Surgeons, all other" (Other)[5], for which the employment and wage rate data is available in most years and states. Total healthcare wages are given by BEA. The data of the remaining terms, the numbers of physicians and non-physicians, have been introduced in detail by previous sections. Descriptive statistics of all variables used in estimating substitution elasticity can be found in table 4.2.

Table 4.2. Descriptive Statistics for Estimating Substitution Elasticity.

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
lnH	489	23.30	1.028	20.82	25.66
Sp	489	0.199	0.0292	0.127	0.293
lnP	489	9.315	1.075	6.958	11.70
lnA	489	12.13	1.004	9.726	14.27
1/2*(lnP)^2	489	43.96	10.03	24.21	68.46
1/2*(lnA)^2	489	74.03	12.17	47.30	101.8
lnP*lnA	489	114.0	22.33	67.68	167.0

Inputs are defined as P: number of physicians; A: number of non-physicians.

[5] OES provides wage estimates for over 800 occupations. There are seven physician relative occupations: Family and General Practitioners, Internists, Obstetricians and Gynecologists, Pediatricians, Psychiatrists, Surgeons, and Physicians and Surgeons, All other (Occupation Code: 29-1062 through 29-1069).

CHAPTER FIVE

Empirical Results

Baumol's Cost Disease

The existence of Baumol's Cost Disease is tested by applying a fixed-effect estimation on a balanced panel data set consisting of 50 U.S. states over 2008-2016. Two different definitions have been used for the unit healthcare cost, healthcare PCE per capita and healthcare GDP per capita, in order to measure costs from both the production and spending sides.

Table 5.1 presents the estimation results when using healthcare PCE per capita as the dependent variable. A first glance at the table suggests that both the adjusted Baumol variable and relative physician supply are significant at 10% level, which is consistent with our previous assumptions. In addition, it should be noted that the magnitudes of these two coefficients estimated by the seven regressions are reasonably close.

To begin with, we embark on our analysis with the baseline models obviating all control variables. Though the magnitude of the coefficients of interest may not be accurate, their signs are pretty informative for the following investigation. Column (1) indicates that the healthcare industry in the U.S. does suffer from Baumol's Cost Disease, which justifies our theoretical basis. Column (3), on the other hand, shows a negative correlation between the change in relative physician supply and the growth of healthcare expenditures. Note that including information about physician numbers results in a loss

Table 5.1. Results of Baumol's Cost Disease, Spending.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Adjusted Baumol variable	0.00607**	0.0140***			0.00731**	0.0150***	0.0147***
	(0.00298)	(0.00313)			(0.00316)	(0.00264)	(0.00269)
Growth of ratio of physicians to non-physicians			-0.192***	-0.187***	-0.189**	-0.195***	-0.173**
			(0.0710)	(0.0667)	(0.0714)	(0.0607)	(0.0669)
Baumol*Ratio of physicians to non-physicians							-1.545
							(2.231)
Growth of real GDP per capita		0.170***		0.0171		0.134***	0.133***
		(0.0370)		(0.0272)		(0.0333)	(0.0343)
Growth of unemployment rate		-0.00143		-0.0244**		-0.0110	-0.0114
		(0.00678)		(0.00991)		(0.00917)	(0.00929)
Growth of poverty rate		0.0213		-0.00457		-0.00168	0.000447
		(0.0129)		(0.0159)		(0.0152)	(0.0163)
Growth of union coverage rate		-0.00543		-0.00148		-0.00136	-0.00150
		(0.00518)		(0.00572)		(0.00565)	(0.00568)
Growth of old percentage		-0.0169		-0.216		-0.0841	-0.0773
		(0.114)		(0.145)		(0.155)	(0.158)
Constant	0.0315***	0.0729***	0.0433***	0.0522***	0.0409***	0.0430***	0.0428***
	(0.00274)	(0.00302)	(0.00218)	(0.00395)	(0.00239)	(0.00472)	(0.00482)
Observations	950	850	450	450	450	450	450
R-squared	0.594	0.651	0.481	0.502	0.490	0.521	0.522
Number of states	50	50	50	50	50	50	50
Year FE	YES	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES	YES

Standard errors clustered at the state level.
*p<.1; **p<.05; ***p<.01.
The dependent variable is healthcare PCE per capita.

of observations since the data about physician numbers starts from 2008, restricting our sample to 450 observations altogether. Regression (2) and (4) improve the results of (1) and (3) by including control variables that could affect the growth of healthcare expenditures. What's reflected by column (5) and (6) is that the coefficients on the adjusted Baumol variable and the relative physician supply are both significant. In other words, the rapidly rising healthcare spending cannot be explained solely by either Baumol's Cost Disease or a relative shortage of physician supply. Furthermore, with the magnitudes of the two coefficients of interest keeping reasonably similar across the seven regressions, the estimated relationships tend to be robust. Column (7) incorporates the interaction term between the adjusted Baumol variable and the growth of relative physician supply. As the interaction term is insignificant, we can conclude that the degree to which increasing the relative physician supply could abate the healthcare inflation problem does not depend on the severity of the Baumol's Cost Disease. Specifically, a 1% increase in the ratio of physicians to non- physicians is related with a 0.173% decrease in the healthcare PCE per capita, regardless of the degree of Baumol's Cost Disease.

Table 5.2 reveals the estimation results when using nominal healthcare GDP per capita as the dependent variable. Note that the coefficient on the growth rate of the relative physician supply has a greater magnitude when nominal healthcare output is the dependent variable. Specifically, a 1% increase in the ratio of physicians to non-physicians is related with a 0.35% decrease in the growth of nominal healthcare GDP per capita. This occurs because the relative supply of physicians affects the healthcare price as a supply shock. While in Table 5.1, the PCE used as the dependent variable measures healthcare expenditures from the spending side that also reflects demand shocks.

Table 5.2. Results of Baumol's Cost Disease, Value Added.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Adjusted Baumol variable	0.00494	0.0204***			0.00547	0.0210***	0.0205***
	(0.00308)	(0.00356)			(0.00418)	(0.00376)	(0.00386)
Growth of ratio of physicians to non-physicians			-0.350***	-0.374***	-0.348***	-0.385***	-0.350***
			(0.0620)	(0.0595)	(0.0629)	(0.0592)	(0.0661)
Baumol*Ratio of physicians to non-physicians							-2.430
							(1.733)
Growth of real GDP per capita		0.250***		0.0989***		0.262***	0.260***
		(0.0425)		(0.0331)		(0.0488)	(0.0518)
Growth of unemployment rate		0.0143**		-0.00742		0.0112	0.0106
		(0.00592)		(0.0108)		(0.00698)	(0.00688)
Growth of poverty rate		0.0122		0.0404**		0.0444***	0.0478***
		(0.00986)		(0.0164)		(0.0143)	(0.0146)
Growth of union coverage rate		0.00349		0.00275		0.00291	0.00270
		(0.00439)		(0.00471)		(0.00419)	(0.00428)
Growth of old percentage		-0.0222		-0.168		0.0170	0.0278
		(0.125)		(0.172)		(0.160)	(0.157)
Constant	0.0397***	0.0620***	0.0731***	0.0772***	0.0713***	0.0644***	0.0641***
	(0.00290)	(0.00261)	(0.00199)	(0.00460)	(0.00197)	(0.00375)	(0.00380)
Observations	900	800	450	450	450	450	450
R-squared	0.622	0.706	0.726	0.739	0.729	0.762	0.763
Number of states	50	50	50	50	50	50	50
Year FE	YES	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES	YES

Standard errors clustered at the state level.
*p<.1; **p<.05; ***p<.01.
The dependent variable is nominal healthcare GDP per capita.

Therefore, a smaller portion of variation in the growth of healthcare expenditures is explained by the change in relative physician supply.

Real GDP per capita is significant in both regressions, confirming the past literature that income level is a significant driver of healthcare expenditures. However, in Table 5.2, the adjusted Baumol variable tends to be statistically insignificant when estimated without controls, but significantly positive when estimated with controls. This implies that after controlling for other explainers such as the real income level and the unemployment and poverty rates, the effect of the adjust Baumol variable is substantially lessened. In other words, there might be multicollinearity between the four independent variables above. To get a sense of the interactions between independent variables, we first calculate the pairwise correlations, which are presented by table 5.3. The two variables of interest, the adjusted Baumol variable and the ration of physicians to non-physicians, do not present a significant correlation between each other. Revisiting equation (12), faster growth of relative physician supply increases the growth of overall

Table 5.3. Pairwise Correlations Between Selected Independent Variables

Variable	Adjusted Baumol variable	Physician ratio	Real GDP per capita	Unemployment rate	Poverty rate	Union coverage rate	Old percentage
Adjusted Baumol variable	1.000						
Physician ratio	-0.0794	1.000					
Real GDP per capita	-0.3046***	-0.0128	1.0000				
Unemployment rate	-0.2928***	0.0516	-0.5757***	1.0000			
Poverty rate	-0.1964***	0.1543**	-0.3962***	0.5805***	1.0000		
Union coverage rate	0.0003	-0.0160	-0.0184	0.0647	-0.0053	1.0000	
Old percentage	0.2592***	-0.0292	-0.0004	-0.4074***	-0.3297***	-0.0375	1.0000

Variables are in the growth rate form.
* p<0.05; ** p<0.01; *** p<0.001.

productivity ($\frac{\dot{q}_t}{q_t}$), causing the difference between the wage growth and productivity growth to shrink. Nonetheless, since the physicians are much more skilled than non-physicians, increasing physician supply leads to a more than one-for-one decline in the employment of non-physicians, which will cause the total healthcare employment (M_t) to fall in the end. These two contradictory effects may explain partly why our results do not reflect any correlation between the two variables of interest. The adjusted Baumol variable tends to be correlated with a bunch of control variables, which explains why its magnitude and significance level swing across regressions. If the multicollinearity is severe, the estimates will be unreliable because of widely inflated standard errors. Also, the t-statistics will be much smaller, making the null hypothesis harder to decline. As a result, the coefficient of the adjusted Baumol variable could be falsely insignificant.

A widely-used diagnostic for multicollinearity is the Variance Inflation Factor (VIF). VIF measures how much the variance of an estimated coefficient is inflated due to the multicollinearity. If no correlation exists, VIF will equal one. A variable having a greater-than-10 VIF merits further investigation. The VIF for coefficient *k* can be calculated as follows:

$$VIF_k = \frac{1}{1 - R_k^2} \tag{21}$$

In equation (21), R_k^2 represents the R-squared of the auxiliary regression where predictor *k* is regressed on all other predictors. To obtain the VIF for each independent variable in our analysis, we need to run seven auxiliary regressions in total. Table A.3 in Appendix includes results of all auxiliary regressions. The calculated VIF's are given by Table 5.4. Note that the VIF's for unemployment rate and old percentage are bigger than

10, suggesting the multicollinearity brings quite inflated variance to coefficients of these two variables. However, since these two variables serve as controls, it is less problematic to have large VIF's. The two variables of primary interest, adjusted Baumol variable and physician ratio, have VIF's both smaller than 10, implying some variance inflation though not enough to be overly concerned.

Comparing the results in Table 5.1 with those in Table 5.2, we find that using healthcare PCE per capita as the dependent variable obtains more stable coefficients for the adjusted Baumol variable (i.e. smaller standard errors). Thus, for the future studies that address the growth of healthcare expenditures from the view of unbalanced growth, the use of PCE as the dependent variable is recommended. Note that Bates and Senterre (2012) use healthcare GDP per capita as the only dependent variable, so the magnitudes and p-values of their estimates could be less accurate.

Table 5.4. VIF's for Selected Independent Variables.

Variable	VIF
Adjusted Baumol variable	3.01
Physician ratio	2.02
Real GDP per capita	3.98
Unemployment rate	10.35
Poverty rate	3.19
Union coverage rate	1.09
Old percentage	12.16

VIF's for year and state dummies are omitted.
Variables are in the growth rate form.

Substitution Elasticity Between Physicians and Non-physicians

The results in the last section collectively suggest a negative correlation between the change in relative physician supply and the growth of healthcare expenditures. Revisiting equation (14) of our theory model, this implies a weak substitution relationship between physicians and non-physicians (i.e. θ is below one), whereby a higher ratio of physicians in total healthcare labor employment (μ_t) increases the average labor productivity in the healthcare sector (B_t). In order to test this inference, we write our CES production function in translog form, apply SUR on the unbalanced panel data set, and estimate parameters jointly to calculate the substitution elasticity.

Following the strategy described in chapter 4.2, the SUR results are presented in Table 5.5. Including the data of physicians' wage rate restricts our sample to 489 observations in total. Column (1) and (2) represent two linear regressions, equation (14) and (15), where the dependent variable is lnH and S_p respectively. Note that the same variable in two regressions is connected with different coefficients by the nature of our design. For instance, the coefficient for lnP in regression (1) is α_P, however, in regression (2) it is δ_{PP}. A first glance at Table 5.5 suggests that all variables are significant at 1% level. The Breusch-Pagan test of independence returns a chi-squared statistics 69.05 with a close-to-zero p-value, thus the null hypothesis is rejected, implying that the error terms in the two equations are significantly correlated. Therefore, estimating the two equations jointly is more efficient than running OLS separately, which justifies the use of SUR. To show the magnitudes of parameters more clearly, we reorganize Table 5.5 to be Table 5.6.

Table 5.5. Parameter Estimates, Seemingly Unrelated Regression.

VARIABLES	(1) lnH	(2) Sp
lnP	0.550***	0.123***
	(0.0143)	(0.00514)
lnA	0.455***	-0.123***
	(0.0145)	(0.00514)
1/2*(lnP)^2	0.123***	
	(0.00514)	
1/2*(lnA)^2	0.123***	
	(0.00514)	
lnP*lnA	-0.123***	
	(0.00514)	
Constant	12.17***	0.545***
	(0.0393)	(0.0145)
Observations	489	489
R-squared	0.994	0.474

Standard errors in parentheses.
Inputs are defined as P: number of physicians; A: number of non-physicians.
*p<.1; **p<.05; ***p<.01.
Breusch-Pagan test: chi2(1)=69.05, Pr=0.0000.

Table 5.6. Parameter Estimates, Rearranged.

Parameter	Estimate	Standard Error
α_0	12.17	0.0393
α_P	0.55	0.0143
α_A	0.455	0.0145
δ_{PP}	0.123	0.00514
δ_{AA}	0.123	0.00514
δ_{PA}	−0.123	0.00514

Inputs are defined as P: number of physicians; A: number of non-physicians.
All variables are significant at 1% level.

Using the estimates in Table 5.6, we are able to calculate the predicted value of the cost shares, which will be used in the computation of elasticity of substitution. Note that each observation in our sample will obtain a specific group of predicted cost shares, and we take the arithmetic mean to use in the final computation. The fitted cost shares can be obtained by the following equations:

$$\widehat{S_P} = \widehat{\alpha_P} + \widehat{\delta_{PP}} lnP + \widehat{\delta_{PA}} lnA \tag{21}$$

$$\widehat{S_A} = \widehat{\alpha_A} + \widehat{\delta_{AA}} lnA + \widehat{\delta_{PA}} lnP \tag{22}$$

The cost shares for physicians and non-physicians will sum up to one since they are assumed to be the only inputs in the production of the healthcare sector. Using equation (21) and (22), we obtain 440 groups of fitted cost shares, of which the mean for physicians' and non-physicians' cost share is roughly 0.2 and 0.8 respectively. Since both numbers are positive, this result supports our model. Inserting values in equation (17), we solve for the Allen partial elasticity of substitution between physicians and non-physicians:

$$\widehat{\sigma_{PA}} = \frac{\widehat{\delta_{PA}} + \widehat{S_P}\widehat{S_A}}{\widehat{S_P}\widehat{S_A}} \approx \frac{-0.123 + 0.2 \times 0.8}{0.2 \times 0.8} \approx 0.23 \tag{23}$$

Our results suggest that the elasticity of substitution is 0.23, considerably smaller than one, which indicates a complementary relationship between physicians and non-physician healthcare workers. This result goes in line with our findings in section 5.1, where a negative correlation between the growth of relative physician supply and the growth of medical care price is identified.

CHAPTER SIX

Conclusions and Discussions

Conclusions

For more than two decades, the annual number of medical school graduates in the United States has been held relatively constant due to severe caps placed on the total number of medical school residencies. As a result, there is growing concern of a physician shortage in the future. In the meantime, healthcare expenditures in the United States began rising rapidly. This leads us to ask: Is there a correlation between healthcare spending and the physician supply? To address this question, we adopt a modified version of Baumol's unbalanced growth model, which attributes the rising healthcare expenditures to the gap in productivity growth between the healthcare sector and the final goods sector. Innovatively, we group healthcare laborers into physicians and non-physicians to investigate the interaction between physician supply and unit healthcare cost. Our theoretical model suggests that when physicians and non-physicians work as weak substitutes, a larger share of physicians in total healthcare employment helps decrease the growth rate in the cost of providing healthcare.

Our empirical analysis consists of two parts. The first part is designed to test for the existence of Baumol's Cost disease for the healthcare sector in the U.S., which serves validate the theory that underpins our model. We build on the work of Colombier (2012) and Bates and Santerre (2012) by including the change in the ratio of physicians to non-physicians as an independent variable. Applying a fixed-effect model on a panel data set including 50 states over 2008-2016, we find support for the existence of Baumol's cost disease in the healthcare sector. Moreover, our results indicate that a higher growth rate of relative physician supply relates to a lower growth rate of medical care price, which implies a smaller-than-one elasticity of substitution between physicians and auxiliary workers. Thus, the second part of the empirical analysis is designed to test the implication above. Specifically, we use a SUR procedure to estimate a system of equations that are derived from the translog production function. We estimate that the elasticity of substitution between physicians and auxiliary workers is roughly 0.23, which is consistent with the result from the first part of empirical work. Therefore, if the projected physician shortage comes true, the overall labor productivity in the healthcare sector will be harmed, even if more auxiliary workers are hired. As a result, the gap of productivity growth between the healthcare sector and the final goods sector will be enlarged, which will drive the healthcare price to soar more dramatically.

Past literature attributes the rising healthcare expenditures to the rising demand for medical care service, so it's relatively difficult, if not impossible at all, to solve the problem without making people worse off. Our findings provide a new explanation for the rapid rise of healthcare expenditures. The real-world implication is that governments and organizations can to some extent control the soaring healthcare price by increasing physician supply, by funding more medical school residencies, subsidizing graduate medical education, and training non-physicians to be more skilled, which are achievable with relatively low costs.

Limitations and Future Improvements

Essentially, there is one major limitation in our work, the approximation of physicians' wage rate could be problematic. Recall that we use two categories, FGP and Other, to approximate the physicians' wage rate as the weighted mean, abandoning other groups of specialized physicians. Since specialized physicians' income tends to be much higher than general practitioners, our approximated number might be lower biased. This further affects the calculation of cost shares, and finally could result in a wrong magnitude of substitution elasticity.

The limitation above shapes the direction for future studies. More trustworthy data about physicians' wages is needed to obtain a reliable estimate of substitution elasticity between physicians and auxiliary workers. Besides, future work can further divide physicians into two groups: general practitioners and specialists. These two groups of workers are different in terms of productivity: Specialists are highly skilled while general practitioners serve more like non-physicians. It would be helpful to see how the structure of physicians affects healthcare productivity, which further impacts medical care spending. Furthermore, if the structure of the projected physician shortage can be figured out, we will be able to predict how much extra healthcare spending will be caused, if governments and organizations do not intervene purposefully.

APPENDIX

Additional Tables

Table A.1. Revisiting Table 5.1, Nominal GDP as Control

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Adjusted Baumol variable	0.00607**	0.00562**			0.00731**	0.00767***	0.00743***
	(0.00298)	(0.00218)			(0.00316)	(0.00244)	(0.00258)
Growth of ratio of physicians to non-physicians			-0.192***	-0.193***	-0.189**	-0.191***	-0.172**
			(0.0710)	(0.0648)	(0.0714)	(0.0651)	(0.0681)
Baumol*Ratio of physicians to non-physicians							-1.271
							(2.251)
Growth of nominal GDP per capita		0.0632***		0.0403**		0.0513***	0.0494***
		(0.0233)		(0.0197)		(0.0179)	(0.0180)
Growth of unemployment rate		-0.00977		-0.0211**		-0.0176**	-0.0180**
		(0.00703)		(0.00902)		(0.00804)	(0.00823)
Growth of poverty rate		0.0146		-0.00377		-0.00846	-0.00674
		(0.0126)		(0.0160)		(0.0151)	(0.0161)
Growth of union coverage rate		-0.00486		-0.00156		-0.00141	-0.00152
		(0.00502)		(0.00569)		(0.00562)	(0.00566)
Growth of old percentage		-0.104		-0.184		-0.162	-0.157
		(0.110)		(0.152)		(0.148)	(0.149)
Constant	0.0315***	0.0760***	0.0433***	0.0501***	0.0409***	0.0464***	0.0463***
	(0.00274)	(0.00325)	(0.00218)	(0.00411)	(0.00239)	(0.00410)	(0.00412)
Observations	950	850	450	450	450	450	450
R-squared	0.594	0.644	0.481	0.505	0.490	0.515	0.515
Number of code	50	50	50	50	50	50	50
Year FE	YES	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES	YES

Standard errors clustered at the state level.
*p<.05; **p<.01; ***p<.001.
The dependent variable is healthcare PCE per capita.

Table A.2. Revisiting Table 5.3, Nominal GDP as Control

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Adjusted Baumol variable	0.00494	0.00752***			0.00547	0.00640*	0.00602*
	(0.00308)	(0.00246)			(0.00418)	(0.00341)	(0.00359)
Growth of ratio of physicians to non-physicians			-0.350***	-0.377***	-0.348***	-0.375***	-0.345***
			(0.0620)	(0.0591)	(0.0629)	(0.0596)	(0.0615)
Baumol*Ratio of physicians to non-physicians							-1.996
							(1.741)
Growth of nominal GDP per capita		0.0795**		0.0811***		0.0904***	0.0874**
		(0.0299)		(0.0292)		(0.0335)	(0.0344)
Growth of unemployment rate		0.000386		-0.00574		-0.00277	-0.00344
		(0.00719)		(0.00739)		(0.00650)	(0.00624)
Growth of poverty rate		0.00177		0.0345*		0.0306*	0.0333**
		(0.0109)		(0.0179)		(0.0164)	(0.0161)
Growth of union coverage rate		0.00439		0.00271		0.00283	0.00267
		(0.00451)		(0.00464)		(0.00451)	(0.00462)
Growth of old percentage		-0.142		-0.165		-0.147	-0.140
		(0.124)		(0.161)		(0.159)	(0.159)
Constant	0.0397***	0.0674***	0.0731***	0.0748***	0.0713***	0.0717***	0.0715***
	(0.00290)	(0.00372)	(0.00199)	(0.00398)	(0.00197)	(0.00387)	(0.00397)
Observations	900	800	450	450	450	450	450
R-squared	0.622	0.691	0.726	0.740	0.729	0.744	0.744
Number of code	50	50	50	50	50	50	50
Year FE	YES	YES	YES	YES	YES	YES	YES
State FE	YES	YES	YES	YES	YES	YES	YES

Standard errors clustered at the state level.
*p<.05; **p<.01; ***p<.001.
The dependent variable is nominal healthcare GDP per capita.

Table A.3. Results of Auxiliary Regressions for Multicollinearity Test.

VARIABLES	(1) Adjusted Baumol variable	(2) Physician ratio	(3) Real GDP per capita	(4) Unemployment rate	(5) Poverty rate	(6) Union rate	(7) Old percentage
Growth of ratio of physicians to non-physicians	0.542		0.108	0.339	0.306*	0.629	-0.0166
	(0.994)		(0.139)	(0.403)	(0.163)	(0.575)	(0.0231)
Growth of real GDP per capita	-7.785***	0.0519		-2.200***	-0.395***	0.0344	-0.0571***
	(1.128)	(0.0563)		(0.303)	(0.124)	(0.325)	(0.0116)
Growth of unemployment rate	-0.886***	0.00695	-0.0936***		0.0934***	0.0752	-0.00301
	(0.325)	(0.00778)	(0.0202)		(0.0328)	(0.0823)	(0.00327)
Growth of poverty rate	-0.192	0.0269*	-0.0720**	0.400***		0.0146	0.0147***
	(0.294)	(0.0139)	(0.0281)	(0.140)		(0.176)	(0.00530)
Growth of union coverage rate	-0.00772	0.00494	0.000562	0.0288	0.00131		0.000701
	(0.0662)	(0.00442)	(0.00530)	(0.0314)	(0.0158)		(0.00142)
Growth of old percentage	-8.796***	-0.147	-1.053***	-1.306	1.492**	0.792	
	(2.407)	(0.190)	(0.262)	(1.261)	(0.618)	(1.519)	
Adjusted Baumol variable		0.00208	-0.0618***	-0.165***	-0.00835	-0.00376	-0.00379***
		(0.00342)	(0.00945)	(0.0206)	(0.0138)	(0.0329)	(0.00131)
Constant	0.331***	0.00146	0.0278***	0.0410	-0.0831***	-0.0490	0.0260***
	(0.0542)	(0.00463)	(0.00684)	(0.0343)	(0.0145)	(0.0440)	(0.000624)
Observations	450	450	450	450	450	450	450
R-squared	0.668	0.505	0.749	0.903	0.687	0.082	0.918
Year Dummy	YES	YES	YES	YES	YES	YES	YES
State Dummy	YES	YES	YES	YES	YES	YES	YES

Standard errors clustered at the state level.
*p<.1; **p<.05; ***p<.01.
Column names refer to dependent variables.

BIBLIOGRAPHY

Baumol, W. J. "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis." *The American Economic Review* 57, No. 3 (1967): 415-426.

Berndt, E. R. and Christensen, L. R. "The Translog Function and the Substitution of Equipment, Structures, and Labor in U.S. Manufacturing 1929-68." *Journal of Econometrics* 1, No. 1 (1973): 81-113.

Blomqvist, A. and Carter, R. "Is Health-Care Really a Luxury?" *Journal of Health Economics* 16, No. 2 (1997): 207–229.

Brown, M. C. *Caring for Profit: Economic Dimension of Canada's Health Industry*, (1987).

Colombier, C. "Drivers of Health Care Expenditure: Does Baumol's Cost Disease Loom Large?" *FiFo Discussion Papers*, No. 12-5 (2012).

Cooper, R., T. Getzen, H. McKee, and P. Laud. "Economic and Demographic Trends Signal an Impending Physician Shortage." *Health Affairs* 21, No. 1 (2002): 140-154.

Fonseca, B., Raquel, Michaud, Pierre-Carl, Kapteyn, Arie and T. J. Galama. "Accounting for the Rise of Health Spending and Longevity." *IZA Discussion Paper* No. 7622 (2013).

Frankovic, I., M. Kuhn and S. Wrzaczek. "Medical Progress, Demand for Health Care, and Economic Performance." *ECON WPS-Vienna University of Technology Working Papers in Economic Theory and Policy* No. 8 (2017): 1-54.

Griffin, J. and R Gregory, P. "An Inter Country Translog Model of Energy Substitution Responses." *American Economic Review* 66. No. 5 (1976): 845-857.

Hall, R. E., Jones, C. I. "The Value of Life and the Rise in Health Spending" *The Quarterly Journal of Economics* 122, No. 1 (2007): 39–72.

Hartwig, J. "What Drives Health Care Expenditure? Baumol's Model of Unbalanced Growth Revisited." *Journal of Health Economics* 27, No. 3 (2008): 603–623.

Hitiris, T. and Posnett, J. "The Determinants and Effects of Health Expenditures in Developed Countries". *Journal of Health Economics* 11, No. 2 (1992): 173–181.

J Bates, L. and Santerre, R. “Does the U.S. Health Care Sector Suffer from Baumol's Cost Disease? Evidence from the 50 States.” *Journal of Health Economics* 32, No. 2 (2012): 386-391.

Kelly, M. C. “The Aggregate Consequences of Physician Shortages.” (2018).

Kim, H Y. "The Translog Production Function and Variable Returns to Scale." *The Review of Economics and Statistics* 74, No. 3 (1992): 546-552.

Koening, L., Siegel, J. M., Dobson, A., Hearle, K., Ho, S. and Roduwitz, R. “Drivers of Healthcare Expenditures Associated with Physician Services.” *The American Journal of Managed Care* 9, No.1 (2003): 34–42.

Leu, R. E. “The Public-Private Mix and International Health Care Cost.” *Public and Private Health Services*, (1986): 41–63.

Martín Martín, José J., M. Puerto López del Amo González and M. Dolores Cano García. “Review of the Literature on the Determinants of Healthcare Expenditure” *Applied Economics* 43, No. 1 (2011): 19-46.

Mosca, I. “Decentralization as a Determinant of Health Care Expenditure: Empirical Analysis for OECD Countries.” *Applied Economics Letters* 14, No. 7 (2007): 511–515.

Newhouse, J. P. “Medical care cost: how much welfare loss?” *Journal of Economics Perspectives* 6, No. 3 (1992): 3–21.

Newhouse, J. P. “Medical-Care Expenditure: A Cross-National Survey” *The Journal of Human Resource* 12, No. 1 (1977): 115–125.

Okunade, A. A. and Murthy, V. N. “Technology as a ‘Major Driver’ of Health Care Costs: A Cointegration Analysis of the Newhouse Conjecture.” *Journal of Health Economics* 21, No. 1 (2002): 147–59.

Parkin, D., Mcguire, A. and Yule, B. “Aggregate Health Expenditure and National Income: Is Health Care a Luxury Good.” *Journal of Health Economics* 6, No. 2 (1987): 109–127.

Potrafke, N. “The Growth of Public Health Expenditures in OECD Countries: Do Government Ideology and Electoral Motives Matter?” *Journal of Health Economics* 29, No. 6 (2010): 797-810.

Salsberg, E. and G. Forte. “Trends in the Physician Workforce, 1980-2000.” *Health Affairs* 21, No. 5 (2002): 165-173.

Thurston, N. K. and Libby, A. M. “A Production Function for Physician Services Revisited.” *Review of Economics and Statistics* 84, No. 1 (2002).