DEPARTMENT OF ECONOMICS WORKING PAPER SERIES

Medical Expenditure Growth and the Diffusion of Medical Technology

Justin Polchlopek

Working Paper No: 2011-10

University of Utah
Department of Economics
260 S. Central Campus Dr., Rm. 343
Tel: (801) 581-7481
Fax: (801) 585-5649
http://www.econ.utah.edu

Medical Expenditure Growth and The Diffusion of Medical Technology

Justin Polchlopek*
Department of Economics
University of Utah

March 29, 2011

Abstract

The general consensus among health economists is that the increasing capability of medical providers—often called medical "technology"—is responsible for the majority of growth in medical expenditure. And yet, the principle means of understanding medical technology is through the use of total factor productivity, which, despite giving reasonable estimates of the magnitude of the effects, is not a theory of technology, leaving policymakers without effective tools for prediction.

This paper develops a descriptive model of technology that may have interesting implications for health economics. The model suggests that the manner of diffusion of technology is critical, and when technology diffuses haphazardly, the effects on expenditure can be unexpectedly large.

Keywords: Health Economics, Health Care Production, National Health Expenditures, Sraffian Economics, Total Factor Productivity, Input-Output Economics, Technological Diffusion Processes

JEL Classification: B51, C67, D24, D57, I11, I12, O33

Acknowledgements: Thanks to Gabriel Lozada and Brian England for their helpful comments in the preparation of this paper. Additional thanks go to Scott Carter for arranging the presentation of this work at the 2011 Eastern Economic Association meeting.

*justin@economics.utah.edu

1 Introduction

It is now well known among health economists that health care expenditures are on the rise. As of 2008, health expenditures comprised over 16% of GDP, and historically, the growth in health expenditures almost always outpaces the growth in GDP [3]. This means that, generally speaking, all economic growth is being soaked up by the medical sector and then some. The present state of affairs, where the US pays more than twice the OECD average (in USD PPP), could be excused if it were paying dividends in terms of the health of the public, but by many standards, medical care in the US lags behind many countries [11].[1] Thus, Americans are facing a progressive erosion of their disposable income to pay for health care that is not delivering particularly good results in many basic dimensions. Clearly, determining the sources of the increase in health expenditure is of key importance.

Health economists have, by and large, settled on medical technology—more aptly called medical *capability*—as the primary motivator of increasing expenditure. But for as important as it is to determine the causes of the growth, the methods used are surprisingly unsatisfying. Some of the most significant works that attempt to assess the magnitude of technology's effect center on total factor productivity (TFP), notably Newhouse [10] and Smith, Newhouse, and Freeland [14]. However, TFP-based approaches should be considered with extreme caution as they may be doing nothing more than testing an underlying accounting identity [5]. Even when considered separately from such esoteric theoretical concerns, TFP is fundamentally a method of residuals, giving a name to that which we don't yet fully understand as a means to claim mastery over it.

It should be noted that it is not the results of the TFP methods that are in question. Certainly other authors have likely come quite close to the true effect of technological advances in health. The larger concern is that these aggregate methods are failing to illuminate the vectors by which technology is influencing the economics of health care. One can imagine a world in which new medical breakthroughs could reduce the overall cost of care; so to say that technology, in a blanket sense, is responsible for the bulk of expenditure growth is somewhat reductionist.

So, rather than identifying a negative space and progressively chipping away at it, this paper proposes a positive method for discussing medical technology. At the heart of this discussion is a constructive model of the economy that makes it possible to describe the technology employed in the production of health care in a real and measurable sense. The model leads to measures for technology that connect to broader economic phenomena, including the expansion of medical expenditures.

[1]Note that even though the excess growth of health expenditures of many OECD nations is exceeding the US [17], the absolute magnitude of the difference implies that many years at the current levels would be required to bring US expenditures back in line.

The narrative arc of this paper will begin with a presentation of the model that forms the center of this work. Regrettably, this model requires some discussion that can be quite abstract at times, but in the interest of readability, most of the esoteric mathematical details will be relegated to an appendix. In lieu of a precise treatment of the model, the discussion will attempt to focus on the interpretation and implications of the model. And although the model will be introduced in a generic sense, the most interesting application domain is that of health care markets, as these markets tend to have characteristics that can lead to the undesirable behavior that the model captures. Thus, the main thrust of this discussion will focus on how the model can be used to analyze that sector of the economy.

2 A Model of Production

If one is to talk about technology, then one needs an operational definition for technology. The strategy used in this paper is to consider the *production* of health care goods and services as the central function of the health care sector. In this frame, technology can be simply defined as the manner in which goods and services are produced—specifically, the amount of labor and materials needed to produce a commodity. The model of production that follows from this conception of the economy is a simple general equilibrium framework that avoids the aggregation of disparate commodities while still allowing for meaningful characterizations to be made about the economy. Such a definition is not all-encompassing, since certain capital investments and the prevailing institutional framework, including the laws governing medical practitioners, are not accounted for. But, from the perspective of the day-to-day operation of the economy, such a description is sufficient.[2]

The model at the center of this paper is an input-output system wherein commodities are produced via the use of other commodities. This description leads to a linear system:

$$\vec{x} = A\vec{x} + \vec{d}, \tag{1}$$

where $\vec{x}$ is the total production of the economy, A is an $n \times n$ matrix describing the technical requirements of production, n is the number of commodities produced by the economy, and $\vec{d}$ is the vector describing the demand experienced within the economy.

This model is an accounting method. The economy produces in order to satisfy demand, both the demand that is internal to the producers and the final demand of consumers. Internal demand is counted within the matrix A and final demand is captured in the vector $\vec{d}$. The vector $\vec{d}$ is not individual demand, but an aggregate-

[2]Note that resources spent to counteract capital depreciation and the patterns of behavior that emerge due to the influence of the institutional structure are accounted for in this simple representation.

level demand that is supported by purchasing power; in short, $\vec{d}$ represents *effective demand*. The components of effective demand respond to prices, but the response is the result of aggregating (usually binary) choices by a vast number of consumers. Thus, the demand side of this model is not radically individualistic, but more about the social decision-making process.

The matrix A has a dual purpose in this system. In its primal form, the rows of this matrix describe how the outputs of each industry are divided among the other industries of the economy. To wit, each element, a_{ij}, of the matrix A describes how many units of good i are needed to produce one unit of good j. These flows, of course, only describe the requirements of production and not final (consumer) demand.

On the monetary side, the dual form of (1) describes the prices of production:

$$\vec{p} = (1+r)\left[A^T\vec{p} + m\vec{\ell}\right] \tag{2}$$

with $\vec{p}$ being the vector of prices, r the normal rate of profit for all firms in the economy, m the normal nominal wage rate, and $\vec{\ell}$ being the vector of labor time needed in all sectors. In this case, the term $A^T\vec{p}$ describes the cost of the inputs needed to produce a unit of output in each sector.

The price equations in (2) utilize a uniform rate of profit, r, and a uniform money wage m. These are equalized through the interaction of competitive market forces with the free mobility of capital and labor.[3]

This model also offers an intuitive interpretation of the economy, in that it highlights the nature of the economy as the engine of *reproduction* of society. From period to period, the economy must be producing in excess of the requirements of reproduction; that is, the economy must produce a surplus. The surplus is the source of all economic growth as well as the source of all profits. Also, the interconnection of producers is quite plain in this formulation. If each industry were conceived of as a node in a graph describing commodity flows, then A is an adjacency matrix that represents the edges of the graph. Similarly, A^T describes the reciprocal monetary flows.

One peculiar feature of the price and quantity equations is the absence of the labor vector from equation (1) and the absence of $\vec{d}$ from equation (2). The omission of labor from the quantity equation is deliberate as the quantity equation deals exclusively with commodities, and it is not proper to consider human labor time in the same sense as toasters and light bulbs. In contrast, $\vec{d}$ is implicitly contained in the price equation; as $\vec{d}$ captures

[3]It is true that wage differentials exist—most often between different skill categories of labor—and rates of profit may differ between industries—often due to barriers to entry; but such concerns can be allayed to a large extent. Skill-weighted labor time can be used in conjunction with the normal wage. Further, the uniform rate of profit will give rise to behaviors that are rather pathological, and the non-uniform case is likely to reproduce those same pathologies.

the (material) surplus of the economy, the value of that surplus is divided among r and m.

One might wonder why the present work is concerned with introducing a non-standard model in place of the conventional supply and demand model that pervades modern economics. But it is the shortcomings of the total factor productivity method mentioned in the introduction that provides the impetus to look for alternative formulations. Simply speaking, it is the aggregation of capital required to form the macro-level production function that is central to any study of TFP that is at issue. Capital aggregation leads to many problems that were thoroughly debated in the Cambridge Capital Controversies [7], and so the input-output system described above is a way to avoid aggregation.[4] Ultimately, this approach allows for some novel results while simultaneously skirting the pitfalls that plague TFP.

2.1 Analysis of the Model

The purpose of this paper is to eventually undertake an analysis of a general economy, with a particular emphasis on the interaction between the medical and non-medical sectors. But the general model operates in a space with an arbitrary number of dimensions which can be hard to visualize, making the results of this paper more difficult to believe. Moreover, the model being employed here is likely unfamiliar to many readers. Therefore, it is sensible to start with a very simple instance of the model: one describing an economy that is only engaged in the production of two commodities.

A two-commodity model is appropriate as a narrative device since all of the important characteristics needed for the general model are already present in this very simplified model. Once the necessary concepts have been described, some minimal discussion of the general model will be undertaken in section 2.1.2. For those readers wishing to see more detail, the full derivation can be found in appendix A.

2.1.1 Two-Commodity Case

To begin, consider an economy with two industries: one that produces corn, and one that produces beef. The technology matrix is written as

$$A = \begin{bmatrix} a_{CC} & a_{CB} \\ a_{BC} & a_{BB} \end{bmatrix}.$$

Recall that each coefficient, a_{ij}, describes the flow of output from sector i to sector j. In the corn industry, part of the output is set aside to be used as seed for the following year, while in the beef industry, part of the herd must be retained in order to allow for reproduction of the stock. Thus, both a_{CC} and a_{BB} are nonzero.

[4] Also note that the input output system used here lies at the heart of computable general equilibrium models that have been on the rise in modern economics.

These quantities must also be less than 1 for each sector to be able to produce a usable surplus. Moreover, if corn is defined as the wage good for the economy while beef is a luxury good, then it is useful to include the subsistence requirements of the labor force in the coefficients of the matrix. Thus, the coefficient a_{CC} can be considered to be the sum of seed requirements and the food requirements of labor. Furthermore, the beef sector requires corn to feed the cattle and the ranchers. As a result, a_{CB} is positive. In this scheme, however, a_{BC} is zero since cattle are not required to produce corn, or to feed workers in either sector since beef is a luxury good. Thus, the technology matrix simplifies to an upper triangular form.[5] As a matter of definition, *basic goods* are those that serve as inputs to other productive activities while *luxury goods* do not.

The quantity equations do not provide much insight into the economic problem, but they are important to consider as they ground the economic process in physical reality. The price equations, on the other hand, give rise to more interesting results.

Two fundamental expressions fall from the price equation, (2). First is a relationship between the normal wage rate and the profit rate:

$$w = \frac{1-(1+r)a_{CC}}{\ell_C}. \tag{3}$$

Note that the wage rate is found by taking the price of corn as the numeraire, $w \equiv \frac{m}{p_C}$.[6] Similarly, the relative price of the luxury good can be found as

$$\hat{p} = \frac{(1+r)a_{CB}\ell_C + (1-(1+r)a_{CC})\ell_B}{(1-(1+r)a_{BB})\ell_C}. \tag{4}$$

Equations (3) and (4) are of central concern as they summarize the dynamics existing in the economy. First of all, (3) describes the *wage-profit frontier*. This curve captures the fact that the product of any economy must be distributed among capital owners and labor, and the support of the wage-profit frontier describes the valid range of profit rates and the corresponding wages that may exist with the extant technology. Notably there are a range of possible values, indicating that the shares of labor and capital are not fixed by the productive process but through other means, usually political influence or general bargaining power. This is in contrast to the standard theory that dictates a single wage/profit split based on the marginal productivity of factors.[7]. In the case of the simple corn/beef economy, the wage-profit frontier is linear, but in general the wage-profit

[5] This restricted form of the model that arises when A is upper triangular is of particular interest to the narrative of this paper. Although results of general interest can be derived from the complete 2×2 matrix, those results will be left to other, more worthy references [9].

[6] Also consider that the wage rate specified by (3) in this simple model refers to wages in excess of (biological) subsistence as subsistence wages were provided directly in the technology matrix.

[7] Such determinism may be unjustified, however, due to an implicit assumption of linearity in the wage-profit frontier [2].

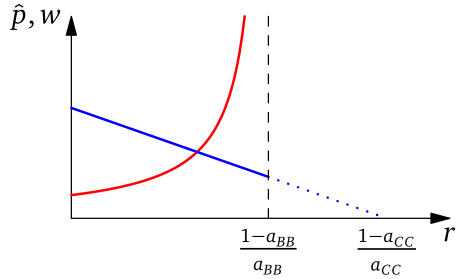

Figure 1: Increasing $\hat{p}$, shown with the w-r frontier.

frontier will be polynomial.

One other interesting feature is that the wage-profit frontier is defined solely in terms of the basic good. This will reappear in the general model and may have noteworthy implications for the behavior of firms.

Turning now to relative prices, an examination of (4) shows that the relative price of beef to corn is a function of the rate of profit and the technology in use, as captured by A and $\vec{\ell}$. Thus, one must consider the response of $\hat{p}$ to the technical conditions of production, as well as market conditions proxied by r. This amounts to an investigation of the shape of $\hat{p}(r)$ within the support of the wage-profit frontier, $r \in \left[0, \frac{1-a_{CC}}{a_{CC}}\right]$.

The form of the price curve gives rise to one particularly interesting case which occurs when $a_{BB} > a_{CC}$. This corresponds to the condition where the own-use of beef exceeds that of corn. This technological configuration leads to a rather peculiar price curve. Figure 1 shows both the wage-profit frontier and a sample $\hat{p}$ curve. The unexpected result is that the relative price of beef experiences unbounded growth within the plausible range of profit rates.

That this phenomenon has such humble origins as a simple and reasonable relationship between a_{CC} and a_{BB} should be an indication that although the shape of $\hat{p}$ seems pathological, it is actually to be expected.

Now, the astute reader might be inclined to complain that, since the profit rate is fixed and generally slow to change, that the changes in $\hat{p}$ that follow due to changes in r are of little interest. This is certainly true, but what does tend to change is technology itself. On the one hand, assume that corn production becomes more efficient in terms of its own-use, then this will have an effect on the wage-profit frontier in the form of an outward shift—simply speaking, the economy has more surplus to divvy up between the participants in the economy. The outward shift in the w-r curve will mean that a new distribution will need to be decided upon.

If one assumes that the economy is operating at the distribution $(r^{\star}, w^{\star})$, then an outward shift of the w-r curve means that the segment of the shifted frontier lying toward the northeast of $(r^{\star}, w^{\star})$ offers a range of new distributions that are, in some sense, Pareto efficient. Through a bargaining process, a new distribution,

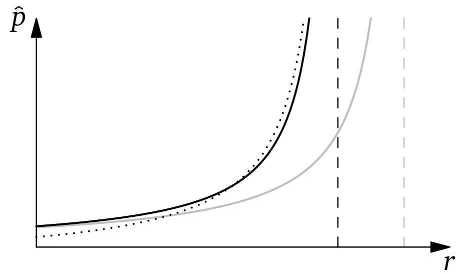

Figure 2: The effect on $\hat{p}$ of increasing own-use in the beef sector.

(r', w'), will be settled upon. Due to the superior bargaining position of capital owners, one expects that $r' > r^{\star}$. Thus, improvements in the production process of basic goods can drive r towards the asymptote of $\hat{p}$. A similar situation emerges if corn becomes more efficient in terms of labor use.

On the opposing side, the own-use of the beef sector determines the location of the vertical asymptote in the price curve. If, for some reason, a_{BB} were to rise, the vertical asymptote would move to the left, causing a rise in $\hat{p}$ for any given r in the support. This situation is depicted in figure 2. In that figure, the gray curve is the original w-r frontier and the solid black curve is the result of raising the value of a_{BB}. For all values of r to the left of the new asymptote, the value of $\hat{p}$ has risen. One can see that this is the case by examining $\frac{\partial \hat{p}}{\partial a_{BB}}$. Thus, one can never arrive at the dotted curve shown in figure 2.

For this simplified model, it is difficult to see a realistic situation that would cause a_{BB} to rise, but for the more complex multi-commodity model, "own-use" has a richer analog which more readily accepts the possibility of increases. That discussion will be deferred until section 3.1.1

2.1.2 General Case

The two-commodity model serves a pedagogical purpose. It is not useful in and of itself, but it does illustrate the kinds of features that are present in the general multi-commodity setting. The addition of an arbitrary number of industries merely changes the form of the system and requires the use of more complex linear algebraic techniques to arrive at all of the same conclusions in a more general setting. Because the tools are more complex, most of the underlying theory has been pushed off into the appendix, but some details must be included here in order to facilitate the discussion regarding medical markets.

To begin, consider that the economy is composed of n industries partitioned into two categories, basic and luxury (non-basic). The technology matrix is formed as a block upper-triangular matrix as shown:

$$A = \left[\begin{array}{c|c} A_0 & A_{0\to 1} \\ \hline 0 & A_1 \end{array}\right]. \tag{5}$$

Here, A_0 is an $m \times m$ technology matrix describing flows internal to the basic sector, A_1 is the $k \times k$ technology matrix internal to the luxury sector, and $A_{0\to 1}$ describes the flows from the basic sector to the luxury sector. Note that $m + k = n$.

The quantity system for the economy is still described by equation (1), and the price system is nearly identical to (2). In the two-commodity setting, the choice of numeraire was easy, since there was only one basic good, and using luxury goods as the numeraire makes little sense. But in the context of many commodities, the numeraire must be a bundle of basic goods. To some extent, the choice of the numeraire bundle, $\vec{b}$, is irrelevant, but as long as the bundle is composed only of basic goods and the money value of the bundle is normalized so that $\vec{b} \cdot \vec{p} = 1$, then that bundle will suffice. With this definition, the price equation in (2) can be rewritten in terms of real wage instead of the money wage:

$$\vec{p} = (1 + r)\left[A^T \vec{p} + w\vec{\ell}\right]. \tag{6}$$

From (6), the wage-profit frontier can again be derived, and just as before, this frontier will only be defined in terms of the technology of the basic sector. However, instead of being linear, it will be a polynomial of order m. The w-r curve will still be downward sloping, but given the indeterminate sign of the second derivative, the frontier may be bowed out or bowed in.

At this point, it is desirable to characterize the price vector. A quick transformation of (6) leads to the equation

$$\vec{p} = w\left(\tfrac{1}{1+r} I - A^T\right)^{-1} \vec{\ell}. \tag{7}$$

In this setup, one wishes to guarantee that a non-negative price vector results from this multiplication. This is guaranteed if $r \in [0, 1/\lambda^\star - 1]$ where $\lambda^\star$ is the maximal eigenvalue of A.[8]

Now, given the structure of A shown in (5), the support of the w-r frontier is $\left[0, 1/\lambda_0^\star - 1\right]$ while the maximal eigenvalue is $\lambda^\star = \max\left\{\mathrm{eig}(A_0), \mathrm{eig}(A_1)\right\}$. Just as in the corn/beef model, if the maximal eigenvalue arises from

[8]Technically, the left endpoint of the valid range of r extends to -1, but negative values of r indicates that the industry is running at a loss, and can never be a long-period equilibrium.

the non-basic sector, then vertically-asymptotic price curves for luxury goods arise. One can see that the values a_{CC} and a_{BB} from the two-commodity case correspond directly to the eigenvalues of the diagonal blocks, and $a_{BB} < a_{CC}$ did in fact lead to a vertical asymptote.

It is the relationship between eig(A_0) and eig(A_1) that determines the price behaviors in this two-sector model. If there were only one industry in a sector, then the eigenvalue for that sector would simply be interpreted as "own-use" of that industry's output. However, in a sector with many industries, there is a richer interpretation for the sectoral eigenvalues: the inverse efficiency of each sector. Naturally, it is the efficiency that is of particular importance, and so the quantities of interest are the $(1/\lambda_i^\star - 1)$'s.

As a final technical note, the Perron-Frobenius theorem[9] states that the $\lambda_i^\star$'s are non-decreasing functions of the elements of the A_i's. Or more intuitively, as resource use rises, efficiency falls. If $\lambda_0^\star$ rises, the support of the w-r frontier shrinks; and if $\lambda_1^\star$ rises, the asymptotes of the relative price curves for all non-basic commodities (with respect to the numeraire bundle, $\vec{b}$) shift to the left.

3 Discussion

The model above captures certain potentially interesting phenomena, but the question remains whether the model enables any useful analyses. The narrative of this piece, unsurprisingly, claims that it does. The model presentation has made connections between technological efficiency and price effects, and a minor extension to the model will also allow for a fuller discussion of the total expenditure picture as well as quality of care. Even though these areas are all well-trodden ground, having a single model to unify them may be a novel contribution to the field.

In order to carry on the discussion, a definition must be made: *medical goods* must be considered as those products which find use only within the medical sector while non-medical goods can find use in any sector. This may lead to some peculiar categorizations. For example, some antibiotics will be considered non-medical goods as they find use in livestock, for example.[10] Medical goods, as a general matter, will find no use in the non-medical markets.[11] In this sense, A_0 from the generic model now represents the flows internal to the non-medical market while A_1 corresponds to the medical market's internal flows.

This model is useful to explain the kinds of results that have been found by earlier TFP-based studies. How-

[9] See proposition 1 in the appendix.

[10] Sub-therapeutic doses of antibiotics are used to encourage more rapid growth of cattle.

[11] Although the argument can be made that workers are more productive when healthy, because A captures only those goods and services that are used in the production of other goods, then the effect of health production will be seen in the labor usage vector, $\vec{\ell}$. To be sure, a decline in the labor requirements will lower prices, but the location of any vertical asymptote in the price curves depends entirely on the contents of A. Therefore, even with a healthy, productive workforce, rapid expansion of prices is possible.

ever, without the use of capital aggregation, this model is not subject to the same kind of criticisms leveled at previous works. The implications of this model are also of some interest, as they offer a unified framework to describe issues involving changes to medical technology as well as the efficiency of medical production.

3.1 The Effect of Increasing Capabilities

The general notion that technology is responsible for the majority of expenditure increases is best explained in terms of an expansion of the capabilities of medical providers. Specifically, with an increasing understanding of various disorders, new treatments come on line. These treatments require certain resources, both medical and non-medical, to provide the treatment.

Suppose that commodity n is a new medical service. Before the treatment is developed and commodified, the n^{th} column of the technology matrix is all zeros; the procedure has not yet been put into production. But, as the procedure becomes commodified, the n^{th} column gains some non-zero entries. In terms of the technology matrix, A, the following structure is present:

$$\left[\begin{array}{ccc|cccc} \times & \cdots & \times & \times & \cdots & \times & \square \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \times & \cdots & \times & \times & \cdots & \times & \square \\ \hline & & & \times & \cdots & \times & \square \\ & & & \vdots & \ddots & \vdots & \vdots \\ & 0 & & \times & \cdots & \times & \square \\ & & & \square & \cdots & \square & \square \end{array}\right]. \tag{8}$$

The boxed entries are the flows involved in the production and use of the new commodity.

The boxed column describes the resources needed to produce the commodity. As production of the new commodity begins, some of these entries will become positive. In section 2.1.2, it was noted that the eigenvalue of A_1, the medical technology matrix, is non-decreasing as a function of the elements of A_1. But, as long as the n^{th} row of A is has zeros in all but the ${a_{nn}}^{\text{th}}$ position, this addition is unlikely to affect the eigenvalue of A.[12]

It is not until the new product has achieved widespread use that the effect of technology can be seen. At that time, the boxed row of A will begin to attain non-zero values. This will drive down the efficiency of the medical sector, potentially leading to unexpectedly severe price effects for medical goods.

[12] This is due to the fact that the matrix A can now be considered a block upper-triangular matrix with three diagonal blocks, A_0, A_1, and A_2, where A_2 is the 1×1 block whose element is the coefficient a_{nn}. So long as the new commodity's own-use does not exceed the maximal eigenvalue of A_1, then the addition of the new commodity will not affect the inverse efficiency of the whole economy.

Clearly, it is not so much the creation of new technologies *per se* that drives cost increases. It is in the diffusion of those technologies that the economic problems begin to arise. For example, Shih and Berliner [13] give a case study of the coronary stent describing how this device has been rapidly adopted in medical practice, with new types of stents and stenting procedures being developed and employed faster than medical researchers can assess the efficacy of those new procedures. This type of progress illustrates the process of technological diffusion, and highlights how quickly such diffusion can occur. In the case of stents, Shih and Berliner show that these devices were frequently utilized even before appropriate structures for remuneration were put in place.

This is the general progress of medical technology. New capabilities are added to the menu of medical treatments and diffuse out, sometimes quite rapidly, into the medical markets. Without a concomitant reduction of utilization in other medical commodities, the efficiency of the production of medical care will fall.

3.1.1 Price Dynamics and Market Behavior

In section 2.1.1, it was shown that the prices of non-basic goods relative to the price of a numeraire bundle may experience a vertical asymptote. The location of that vertical asymptote was defined by the efficiency of production of the non-basic good. In the context of the simple two-commodity model, it might appear difficult to imagine how such a dynamic could possibly arise. Surely, producers must expend significant effort on streamlining production and would not willingly suffer a decrease in efficiency! However, the multicommodity setting—especially the particular character of medical commodity markets—easily allows for a decline in the efficiency of a sector. As a result of perverse incentives toward rapid innovation and technical diffusion, the asymptote of the relative price curve for medical commodities is likely to shift left, drawing nearer to the existing rate of profit.

In the case of non-basic, non-medical commodities, the dynamic of the market imposes some discipline on producers to become more efficient. Although non-basic commodity producers cannot affect the shape of the wage-profit frontier, they can attempt to earn larger absolute levels of profit. This means that there is an interplay between maximizing revenues and minimizing costs. Since cost reduction is a time- and resource-intensive process, those producers who can increase revenues without process changes may do so. This highlights the role of the price elasticity of demand for this class of producer.

For those non-basic commodities that can be considered luxury goods, demand is price elastic. Thus, producers in that sector of the economy must pay heed to the efficiency of production. In the case of medical commodity producers, however, elasticity of demand is generally low. This means that rising prices are a net positive for those producers, and there may be little motivation toward process improvements. Time will be spent on new products and towards pushing existing products into new areas. This priority towards innovation and broader diffusion has already been shown to have adverse effects on prices within the model employed by this paper.

Of course, the type of technological change that has been described earlier is the kind in which the technological advance is *non-substituting*. Such technological advances only describe the beginning of the life cycle of a new technology; in many cases, substitution away from other procedures will likely eventually occur. One must suspect that angioplasty with coronary stents has led to fewer coronary artery bypass grafts.[13] However, the degree to which substitution (a slow process) can counterbalance innovation and diffusion (a fast process) is an open question.

In the end, the model of this paper provides support for the notion that the process of innovation in medical markets leads to undesirable price effects. As prices are half of the expenditure picture, this is a partial explanation for why 16% of US GDP can be spent on health care. However, there are clearly quantity effects that must be considered as well.

3.2 The Role of Income and Insurance

Understanding the quantity side of the medical expenditure is, fundamentally, a study of effective demand—the vector $\vec{d}$ from equation (1). Effective demand arises both as a product of needs and wants as well as the ability of people to pay for those commodities. In medical markets, the fundamental driver of effective demand is need arising from the incidence of illness. However, some medical conditions are of lesser severity; in which case, the ability to pay is a greater determinant of how often treatment will be sought. But, the nature of medical care is that it is often provided through a physician. These agents possess the knowledge to translate an illness into a required bundle of care. These two facets of medical care provision lead to a decomposition of effective demand:

$$\vec{d}_1 = \Lambda \vec{\phi} \tag{9}$$

where $\vec{\phi}$ is a q-dimensional vector of propensities toward particular illnesses, and Λ is a $k \times q$ matrix of procedure utilizations for each condition.

The vector $\vec{\phi}$ is measured as the number of cases of each condition in which treatment was sought. And while rates of incidence are largely stable, certain environmental and social effects can lead to increasing demand for services—specifically, asthma from air pollution, or diabetes, hypertension, and coronary disease from obesity.

[13] In fact, there may be empirical support for this notion in papers such as Cutler *et al.* [4].

But beyond the usual rates of incidence, the magnitude of any ϕ_i is affected by ability to pay, and so the effects of income and insurance coverage will be seen in this vector.

The matrix Λ, on the other hand, captures the behaviors of the physicians, those imperfect agents acting on the behalf of patients. Λ captures many of the phenomena that have been of great interest to economists and policy researchers: demand induction, defensive medicine, procedure duplication, and over-utilization. Insurance and income will also play a role in the formation of Λ as well, since systems of remuneration and the ability of patients to request lower-intensity treatment plans will affect the course of action taken by physicians. Similarly, industrial organization may affect how physicians practice. For example, some staff model MCOs may emphasize the importance of scripts or standard operating procedures in the provision of care.

3.2.1 The Health Care Quadrilemma

It is important to recognize that changes in $\vec{\phi}$ and Λ are not arbitrary or strictly stochastic, but follow from particular patterns of behavior that are both economic and social in origin. In Burton Weisbrod's excellent 1991 paper [16], he describes the interplay between social values, technological development, and insurance. All of these factors are at play in this model as well.

For a start, many societies take the normative stance that health care is a basic right of the citizens of any nation. Thus, safety nets, legal structures, and professional codes of conduct are developed to guarantee that those suffering from illnesses will receive care, with different standards being employed in different countries. Beyond this, insurance insulates individuals from the risk of illness and allows consumption decisions to be made in an environment that is, in many ways, free of the typical cost/benefit comparisons that normally accompany expenditures of the magnitudes often encountered in health care. These characteristics, in addition to the basic desire to prolong and improve the quality of life leads to an aggregate demand for health care that is generally price-inelastic [12]. But more than this, the commitment to tend to the health of its citizens means that certain categories of illness are much more likely to be covered. And as the wealth of the society grows, the number of conditions that will be allowed under this umbrella of coverage will grow as well. Thus, there is a relationship between the magnitude of the elements of $\vec{\phi}$ and the aggregate income of a nation.

Weisbrod also makes plain the connections between the incentives presented by the insurance system and the types of technical developments that occur in the medical sector. Namely, it is only those developments that are likely to be paid for that will be brought to market. This assures that those products that become commodified will be supported by effective demand.

On top of this, it is important to recognize that the focus among many medical R&D firms is towards treatments and not cures. Cures would tend to be substituting technologies that have the potential to lower the utilizations in A, while treatments are more likely to be non-substituting, leading to higher utilizations. This phenomenon interacts with patent law to lead to an even more pronounced tendency to push incremental development where the R&D costs that enable high profits are substantially lowered.

There are other phenomena beyond this that result in a system that fails to consider system-wide efficiency. As one example, consider the possibility that hospitals adopt technology in order to bolster their reputation as a technical leader, and not necessarily to provide lower cost care. This point was illustrated by Teplensky, *et al.* for the case of MRI machines [15].

3.3 Quality and Efficiency

So far, this paper has focused very narrowly on the particular details of expenditure and production, and has pointed a stern finger in the direction of haphazard technical diffusion. But one question has been avoided: isn't it worth it? Don't the benefits of all these new technologies and techniques merit paying the price that we do? After all, if money can't be spent to ensure a better quality of life, then what good is all that wealth?

In principle, there can be no counter argument to this claim. The resulting debate would be a clash of normative stances; and in wars of ideology, there are no winners. But a responsible society must consider the questions of distribution that arise within this debate—especially in the United States, where medical care is increasingly two tiered. But even if one is to set aside the fact that the individuals who give rise to effective demand are not uniformly sampled from the population, the problem of superinflation of medical prices with respect to the rest of the economy will mean fewer medical services, or a less vibrant core economy.

Thus, if one wishes to design good health policy, a link between the technology of production and quality of care must be established. But such an understanding will not come easily. Again referring back to the expansion of use of coronary stents, Shih and Berliner show that adoption of new techniques is motivated by the desire to reduce complications. And yet, the manner of technical diffusion is fast and somewhat haphazard, potentially characterized by a broad adoption of techniques before clinical trials justify their use. Unfortunately, studies of quality are slow and laborious. They require willing patients, resources, and time. And more troubling, the speed of technical diffusion can be so rapid that it is difficult to get a meaningful measure of the quality/efficiency tradeoff. Compound this with the problem that one is unlikely to measure the A matrix, or even the A_1 submatrix, and it would appear that any hope of quality/efficiency comparisons of a particular technology of production are remote, indeed.

Despite the difficulties of measurement and of assessing the effectiveness of a treatment, the input-output scheme may be able to assist the determination of the economic ramifications of a new technological advance. In many cases, new technology is not simply an add-on to an existing procedure. Stenting, for example, is a relatively non-invasive surgical procedure that in many cases can obviate the need for the more intensive coronary artery bypass graft (CABG). If stents were used largely as a substitute for CABG, then this is likely to be an economic boon, as captured by an increase in efficiency. However, stents are often installed as a prophylactic against future complications. Such use may not be warranted in a large number of cases, and projecting the effect on efficiency of this type of change may aid in the process of determining best practices for physicians.

There is also further opportunity to consider quality within the model: via the Λ matrix. The bundle of procedures used on a per-illness basis may in fact be measurable from publicly (albeit, not freely) available data. If one were to consider a large sample of episode-of-care information down to the level of procedure codes, estimates of Λ could be obtained. However, to make this data useful, one needs a baseline, and thankfully such baselines for a wide variety of conditions were developed by Kerr *et al.* [8] for use in the RAND QA tools system. The authors develop a system of quality metrics based on a retrospective analysis of the treatments performed for patients presenting with a particular diagnosis. The quality metrics are based on the procedures recommended by a panel of experts. The closer the treatment to the panel recommendations, the better the care is considered to be. As such, if one were to encode the recommended bundle into a Λ-like matrix, call it $\bar{\Lambda}$, then the difference between the two matrices could serve as the basis for a measurement of quality.

However, these quantity effects are linear, while price changes are superlinear; and so, such a measure, while useful, does not go far enough towards providing a tool for assessing technological change. Clearly, further study is needed to obtain useful, practical tools from this model.

4 Conclusion

The knowledge among health economists that the costs of providing health care are rising at a dangerous rate is pervasive, so much so that an understanding of this basic fact is even seeping out into the non-economist population. The culprit for the increase has long been identified as technology, and yet, after years of study, the best estimates are based on less-than-satisfying estimates of residuals. Approaches such as these can trace out the broad outlines of the present moment, but can say much less about the future. Moreover, the effect of particular innovations on the trajectory of cost are difficult to assess.

This paper presented an inductive model of technology that offers some notion of the specific vectors by which technology affects expenditure. It is clear that the interaction between the determinants of expenditures is complex, and the position that the United States finds itself in is the result of an aggregation of small choices made over many decades. Illness rates, practice behavior, and the specific shape of medical technology are all to blame. To reverse course, a deliberate pattern of reforms must occur over a similarly long time span. One result of this paper is to suggest that specific avenues of reform will bear more fruit than others: those reforms that are most likely to raise efficiency.

Many proposals for cost containment have been put forth, with many only affecting the utilization of medical goods and services in the treatment of a given condition (the elements of Λ in the terminology given above). But, improvements in rates of utilization are linear contributors to the expenditure picture while the effect of changes to technology are potentially unbounded. In some ways, the idea of price inflation as a major motivator for the growth of expenditure has been explored [1], but that work does not make an explicit connection to the role of technology.

Clearly, any improvement in the expenditure picture is worth the effort, but the changes that will net the greatest returns are those that will most forcefully impose discipline on the manner of production. This highlights the role of specialty clinics that focus on providing a very small number of procedures with the most streamlined production process available. But, the medical system as a whole must respect the fact that the superlinear growth in expenditure is a system-wide phenomenon that demands a holistic approach to containment, one in which individual providers need to consider surrendering some autonomy in the interest of making coordinated and well-reasoned changes to a system in dire need of repair.

References

[1] Anderson, G., Reinhardt, U., Hussey, P., and Petrosyan, V. It's the prices, stupid: why the United States is so different from other countries. *Health Affairs 22*, 3 (2003), 89.

[2] Bhaduri, A. On the significance of recent controversies on capital theory: A Marxian view. *The Economic Journal 79*, 315 (1969), 532–539.

[3] Centers for Medicare and Medicaid Services. National health expenditure accounts. `https://www.cms.gov/nationalhealthexpenddata/02_nationalhealthaccountshistorical.asp`, 2008. [Online; accessed October, 2010].

[4] Cutler, D. M., McClellan, M., Newhouse, J. P., and Remler, D. Are medical prices declining? evidence from heart attack treatments. *The Quarterly Journal of Economics 113*, 4 (November 1998), 991–1024.

[5] Felipe, J., and McCombie, J. Some methodological problems with the neoclassical analysis of the East Asian miracle. *Cambridge Journal of Economics 27*, 5 (2003), 695.

[6] Graham, A. *Nonnegative matrices and applicable topics in linear algebra*. Ellis Horwood, 1987.

[7] Harcourt, G. *Some Cambridge controversies in the theory of capital*. Cambridge University Press, 1972.

[8] Kerr, E., Asch, S., Hamilton, E., and McGlynn, E. *Quality of Care for General Medical Conditions: A Review of the Literature and Quality Indicators*. RAND Health, 2000.

[9] Kurz, H., and Salvadori, N. *Theory of Production: A Long-Period Analysis*. Cambridge University Press, 1995.

[10] Newhouse, J. An iconoclastic view of health cost containment. *Health Affairs 12*, Supplement 1 (1993), 154–71.

[11] OECD Indicators. Health at a Glance 2009.

[12] Ringel, J. S., D.Hosek, S., Vollaard, B. A., and Mahnovski, S. *The elasticity of demand for health care: a review of the literature and its application to the military health system*. Rand Corp, 2002.

[13] Shih, C., and Berliner, E. Diffusion Of New Technology And Payment Policies: Coronary Stents. *Health Affairs 27*, 6 (2008), 1566.

[14] Smith, S., Newhouse, J., and Freeland, M. Income, Insurance, And Technology: Why Does Health Spending Outpace Economic Growth? *Health Affairs 28*, 5 (2009), 1276.

[15] Teplensky, J., Pauly, M., Kimberly, J., Hillman, A., and Schwartz, J. Hospital adoption of medical technology: an empirical test of alternative models. *Health Services Research 30*, 3 (1995), 437.

[16] Weisbrod, B. The health care quadrilemma: an essay on technological change, insurance, quality of care, and cost containment. *Journal of economic literature 29*, 2 (1991), 523–552.

[17] White, C. Health care spending growth: how different is the United States from the rest of the OECD? *Health affairs 26*, 1 (2007), 154.

A Mathematical Appendix

The following sections show the full derivations of the mathematical theory used in section 2. This will proceed from the 2-commodity model to the general multi-sectoral model.

A.1 Two-Commodity Model

To begin, consider the simple economy producing two commodities: corn and beef. Corn is the basic good and the numeraire. Beef is a luxury good and a final good, i.e., beef does not enter into the production of corn. This is distilled into the technology matrix

$$A = \begin{bmatrix} a_{CC} & a_{CB} \\ 0 & a_{BB} \end{bmatrix}. \tag{10}$$

The quantity equations are of significantly less interest than the price equation

$$\vec{p} = (1+r)\left[A^T \vec{p} + m\vec{\ell}\right]. \tag{11}$$

Here, $\vec{p}$ is the 2-vector comprised of p_C and p_B, the individual commodity prices, while $\vec{\ell}$ is the vector of labor time needed per unit of output in each sector. The rate of profit, r, is taken to be uniform across both industries as is the money wage, m. The equalization is a result of competition and the free mobility of capital and labor. Equation (11) can be manipulated to read

$$\vec{p} = m\left(\tfrac{1}{1+r} I - A^T\right)^{-1} \vec{\ell}. \tag{12}$$

This expands into the system of two equations

$$p_C = m\frac{\ell_C}{1-(1+r)a_{CC}} \tag{13}$$

$$p_B = m\frac{(1+r)a_{CB}\ell_C + (1-(1+r)a_{CC})\ell_B}{(1-(1+r)a_{CC})(1-(1+r)a_{BB})}. \tag{14}$$

Dividing both equations by the numeraire, p_C, gives

$$1 = w\frac{\ell_C}{(1-(1+r)a_{CC}} \tag{15}$$

$$\hat{p} = \frac{(1+r)a_{CB}\ell_C + (1-(1+r)a_{CC})\ell_B}{(1-(1+r)a_{BB})\ell_C}. \tag{16}$$

where $w \equiv \frac{m}{p_C}$ is the real wage, and $\hat{p} \equiv \frac{p_B}{p_C}$ is the relative price of beef.

Equation (15) can be manipulated to yield

$$w = \frac{1-(1+r)a_{CC}}{\ell_C}, \tag{17}$$

the so-called *wage-profit frontier*, showing the profit rate and the wage rate to have a range of possible pairings determined by the technical coefficients of the economy. In other words, wages and profits are a distribution of the surplus product of the society, and the size of the surplus is technically determined. The support of the wage-profit frontier is $[0, \frac{1-a_{CC}}{a_{CC}}]$.

In the two-commodity model, the wage-profit frontier is defined solely in terms of the corn industry, but the relative price of beef is defined in terms of both industries. The value of $\hat{p}$ must be positive, and so it is necessary to consider the range over which $\hat{p} \geqslant 0$. Note that the denominator of (16) is positive so long as r is less than $r_{crit} \equiv \frac{1-a_{BB}}{a_{BB}}$. At that critical value, there is a vertical asymptote in the price curve.[14] The specific shape will depend on the derivative of $\hat{p}$:

$$\frac{\partial \hat{p}}{\partial r} = \frac{a_{CB}\ell_C + (a_{BB}-a_{CC})\ell_B}{(1-(1+r)a_{BB})^2\,\ell_C}. \tag{18}$$

Note that $\frac{\partial \hat{p}}{\partial r}$ will have a constant sign (barring $r = r_{crit}$) whose value depends entirely on the sign of the numerator, which does not vary with r. Therefore, equation (18) gives rise to the following statement:

$$\frac{\partial \hat{p}}{\partial r} > 0 \iff \frac{\ell_C}{\ell_B} > \frac{a_{CC}-a_{BB}}{a_{CB}}, \tag{19}$$

and the technology of production dictates the behavior of the relative price of beef.

To begin, consider the initial value of the relative price:

$$\hat{p}(r=0) = \frac{a_{CB}\ell_C + (1-a_{CC})\ell_B}{(1-a_{BB})\ell_C}.$$

This positive starting point means that there will be a non-empty intersection of the support of (17) and (16). That interval depends entirely on sign changes of $\hat{p}$.

If $\hat{p}$ has a negative slope, then the upper end of the support of $\hat{p}$ is given by the r-intercept of that function.

[14]It should be noted that the asymptote is not of the removable kind so long as $a_{CB} > 0$. This should be assumed to be the general form of the economy, as economies with two entirely disjoint sectors are unlikely.

This is found by locating the roots of the numerator of (16):

$$0 = (1+r)a_{CB}\ell_C + (1-(1+r)a_{CC})\ell_B.$$

This distills down to the inequality

$$r \leqslant \frac{1-a_{CC}}{a_{CC}} \leqslant \frac{(1-a_{CC})\ell_B + a_{CB}\ell_C}{a_{CC}\ell_B - a_{CB}\ell_C}. \tag{20}$$

Thus the support of $\hat{p}$ is a superset of the support of the wage-profit frontier.

In the event that $\hat{p}$ is an increasing function, then $\hat{p}$ becomes negative at r_{crit}. However, when $a_{BB} > a_{CC}$, then it is the case that the valid range of r is not the full support of the wage-profit frontier, but rather the half-open interval $[\,0, r_{crit}\,)$. Also note that the condition $a_{BB} > a_{CC}$ guarantees a positively-sloped price curve as per the condition in (19).

A.2 General Case

Consider the case where the economy is comprised of n different industries, and those industries are partitioned into two groups. The partitioning is set up so that the outputs of one group of industries does not feed back into any industry of the other group. In this case, the group of industries whose outputs are consumed by the other group constitute the *basic sector* of the economy. The group whose outputs are only consumed by end users is called the *non-basic sector*.

Such a partitioning of the economy leads to an $n \times n$ matrix with a 2×2 block upper-triangular structure:

$$A = \begin{bmatrix} A_0 & A_{0\to 1} \\ 0 & A_1 \end{bmatrix} \tag{21}$$

where A_0 is an $m \times m$ matrix and A_1 is a $k \times k$ matrix with $n = m + k$. For this matrix structure to be of interest, $A_{0\to 1}$ must be nonzero.

The price equations for the economy are still described in general form by equations (11) and (12). But rather than approach this portion of the analysis in the system of equations form, a more generic linear algebraic solution will be employed.

Recall that equation (12) can be rewritten as

$$\vec{p} = (1+r)\left[A^T\vec{p} + m\vec{\ell}\,\right].$$

In the more general case, one may wish to consider differentiated labor types, which are not included in the previous expression. Different labor types could be incorporated by using $L\vec{\omega}$ in place of $w\vec{\ell}$, where $\vec{\omega}$ is a vector of differing wages and the columns of L show the amount of one type of labor. However, this adds complications that are unnecessary since one can simply define $\vec{\ell}$ as skill-weighted labor time, in which case the normal wage rate, w, can be defined so that $w\vec{\ell} \equiv L\vec{\omega}$.

Again, (12) is found by simple algebraic manipulation:

$$\left(\frac{1}{1+r}I - A^T\right)\vec{p} = m\vec{\ell}.$$

Matrices of the form $sI - A$ where A is non-negative are in the class of M-matrices, which is rigorously defined as

$$\mathscr{M} = \left\{M \in R^{n\times n} \middle| M = sI - A, A \geqq 0, s > \rho(A)\right\},$$

where $\rho(A)$ is the spectral radius of A.[15]

M-matrices have many interesting properties. For the purposes of this discussion, however, the properties of interest involve the positivity of the inverse—meaning that because $\vec{\ell}$ is positive and $\vec{p}$ must also be positive, then a non-negative inverse of $\frac{1}{1+r}I - A^T$ is required. Some knowledge about the eigenstructure of the M-matrix is also needed due to the connections between the eigenvalues and the wage-profit frontier. These results will then permit a characterization of prices akin to the two-commodity model.

To begin, the inverse of an M-matrix is positive. The following propositions demonstrate:

Lemma 1 *Given a square matrix $B \geqq 0$ with $\rho(B) < 1$, then*

$$(I - B)^{-1} = I + B + B^2 + B^3 + \dots$$

Proof: Starting with the expression

$$(I - B)(I + B + B^2 + B^3 + \dots + B^k) = I - B^{k+1}$$

for some k, left-multiply both sides by $(I - B)^{-1}$:

$$I + B + B^2 + B^3 + \dots + B^k = (I - B)^{-1}(I - B^{k+1}).$$

[15]For clarity, $A > B$ is shorthand for $a_{ij} > b_{ij}$ $\forall i, j = 1, \dots, n$, $A \geq B$ indicates $[a_{ij}] \geqslant [b_{ij}]$ $\forall i, j = 1, \dots, n$ with at least one strict inequality, and $A \geqq B$ indicates that $A \geq B$ or $A = B$.

Taking the limit of both sides as $k \to \infty$ reveals

$$I + B + B^2 + B^3 + \ldots = (I - B)^{-1},$$

noting that since $\rho(B) < 1$, B^k approaches 0 when k is large. •

This last proposition leads to a general statement about the non-negativity of the inverse of an M-matrix:

Lemma 2 *Given a matrix $M \in \mathscr{M}$, M^{-1} exists and $M^{-1} \geq 0$.*

Proof: Since $M \in \mathscr{M}$, then $M = sI - A$ for some $A \geqq 0$. This implies

$$M = s\left[I - \frac{A}{s}\right]$$

where $\rho\left(\frac{A}{s}\right) < 1$, by construction. Due to the spectral radius of $\frac{A}{s}$, $I - \frac{A}{s}$ is obviously non-singular. Further, Lemma 1 indicates that $(I - A/s)^{-1}$ is positive, so M^{-1} is positive as a result. •

The existence and positivity of the inverse of an M-matrix is very useful if one can demonstrate that $\frac{1}{1+r}I - A^T$ is in the class of M-matrices. To show this, consider the case where the economy pays no wages.[16] Since the present model is concerned with distributing the surplus value of the economy to capital and labor, the zero-wage case corresponds to the maximum profit case—a fact that will be verified in due course.

When there are no wages, the price equations become

$$\left(\frac{1}{1+r}I - A^T\right)\vec{p} = \vec{0}. \tag{22}$$

This homogeneous equation shows that the eigenvalues of A^T relate to the rate of profit and the eigenvectors correspond to the equilibrium price vectors. However, this implies that, for an economy with n industries, there are n viable rates of profit. The key to reducing this problem is to consider the Perron-Frobenius Theorem, previously shown as Proposition 1.

Several aspects of the Perron-Frobenius Theorem are useful, but for the moment, notice that the largest eigenvalue of A^T is positive and real, with a corresponding price vector that is non-negative. No other eigenvalue is guaranteed to have a corresponding non-negative eigenvalue, and as such, the value $\lambda^\star$ should be treated as the unique value for which the economic interpretation is valid. This eigenvalue gives the maximal rate of profit for the given technology: $r^\star = \frac{1}{\lambda^\star} - 1$.

[16] It is possible to incorporate a subsistence wage basket into the technology matrix. The inclusion of subsistence wages into the technology matrix would mean that w represents wages above subsistence.

Combined with the definition of an M-matrix, this says that any value of $\frac{1}{1+r} > \rho(A^T) = \lambda^\star$ will admit a positive wage and the corresponding equilibrium price vector will be positive to boot. Thus, the matrix $\frac{1}{1+r}I - A^T$ is in the class of M-matrices for the meaningful range of values $-1 < r \leqslant 1/\lambda^\star - 1$.

With this background of mathematical knowledge, the analysis of the economic model can continue. Two features are of particular interest: the definition of the wage-profit frontier and the shape of the price curves.

In order to develop the wage-profit frontier in the two commodity model, the price equations were normalized through the use of a numeraire. That will be the approach taken here as well, the only difference being that the numeraire will no longer be a single commodity but a bundle of commodities. Thus, the vector $\vec{b}$ will be a collection of commodities for which

$$\vec{b} \cdot \vec{p} = 1.$$

In order to simplify the later analysis, the numeraire basket will be composed entirely of basic goods. That is,

$$\vec{b} = \left[\vec{b}_0^T \,|\, \vec{0}^T\right]^T,$$

where $\vec{b}_0$ is $m \times 1$.

Taking the numeraire good into account, one can find, starting from (12), that

$$\vec{b}^T \vec{p} = 1 = w \vec{b}^T \left(\tfrac{1}{1+r} I - A^T\right)^{-1} \vec{\ell}$$

which leads to the wage-profit frontier:

$$w = \frac{1}{\vec{b}^T \left(\frac{1}{1+r} I - A^T\right)^{-1} \vec{\ell}}.$$

One benefit of taking the numeraire to be defined only in basic commodities is that the wage profit frontier reduces to

$$w = \frac{1}{\vec{b}_0^T \left(\frac{1}{1+r} I - A_0^T\right)^{-1} \vec{\ell}_0}, \tag{23}$$

or in other words, the wage profit frontier is only defined in terms of the basic sector of the economy.

Once again, the properties of M-matrices and the Perron-Frobenius theorem can be used here to show that the maximal eigenvalue of A_0^T, $\lambda_0^\star$ sets the upper bound for values of r that have positive wage rates associated to them. Again notice that only the technological specification of the basic sector of the economy participates in this definition.

With the wage-profit frontier defined, some understanding of prices must also be derived. Starting with equation (12), dividing by the price of the numeraire bundle gives

$$\vec{p} = w\left(\tfrac{1}{1+r}I - A^T\right)^{-1}\vec{\ell}$$

which combines with (23) to give

$$\vec{p} = \frac{\left(\tfrac{1}{1+r}I - A^T\right)^{-1}\vec{\ell}}{\vec{b}_0^T\left(\tfrac{1}{1+r}I - A_0^T\right)^{-1}\vec{\ell}_0}. \tag{24}$$

As the current focus is to generalize the results of the two-good case, the object of study now becomes the prices of the non-basic goods. If $\vec{p}$ is partitioned into an m-vector, $\vec{p}_0$, and a k-vector, $\vec{p}_1$, the former corresponding to the basic commodity prices, and the latter corresponding to non-basic goods, then it is the response of the elements of $\vec{p}_1$ to changes in r that is of particular interest.

Note that in equation (24), the denominator only involves $\frac{1}{1+r}I - A_0$. This M-matrix is invertible and the inverse is positive as long as $\frac{1}{1+r} > \rho(A_0)$. The numerator of (24), on the other hand, has participation from all of A, and the resulting M-matrix is therefore invertible with a positive inverse for the range $\frac{1}{1+r} > \max\{\rho(A_0), \rho(A_1)\}$ since the eigenvalues of a block triangular matrix is the union of the eigenvalues of the diagonal blocks. Thus, if $\rho(A_1)$ is larger, then the valid range for which a non-negative price vector is guaranteed is a subset of the support of the whole wage profit frontier. This same phenomenon was also observed in the two-commodity case.

The next result that must be derived will be akin to the result shown in (19); that is, under what conditions do the prices of non-basic goods experience unbounded growth? The condition can be stated simply as

$$\rho(A_1) > \rho(A_0) \implies \frac{\partial p_i}{\partial r} > 0, \forall i = m+1, \ldots, n. \tag{25}$$

That is, if the non-basic sector has a larger eigenvalue than the basic sector, then the prices in the non-basic sector will tend to have the same asymptotic behavior as in the two-commodity case. But, in order to prove this, one preliminary result is needed.

Lemma 3 *If $A, B \in \mathscr{M}$ and $B \geqslant A$, then*

(a) $A^{-1} \geqslant B^{-1}$, and

(b) $|B| \geqslant |A| > 0$.

This proposition is Lemma 5.6 in [6], and proof can be found in that reference.

Corollary 1 *If* $M = sI - A \in \mathscr{M}$*, then* $\operatorname{adj}(M) \geqslant 0$.

Proof: Since $\operatorname{adj}(M) = |M| \cdot M^{-1}$ and both $|M| > 0$ and $M^{-1} \geqslant 0$, the stated claim follows immediately. •

In preparation for the main result of this paper, one further mathematical theorem is required:

Proposition 1 (Perron-Frobenius Theorem) *If* $A \geqslant 0$ *and* A *is* $n \times n$*, then*

(a) *A has a non-negative real eigenvalue, $\lambda^\star$;*

(b) $\lambda^\star = \rho(A)$*;*

(c) *there is at least one eigenvector, $\vec{v}$, associated with $\lambda^\star$ so that $\vec{v} \geqslant 0$;*

(d) *$\lambda^\star$ is a non-decreasing function of all the elements of A.*

This proposition is a fundamental result in the theory of non-negative matrices, and as such many different proofs are available; see, for example, [9] or [6].

For the purposes of this paper, it is part (d) that is most relevant. Notice that this statement says that the aggregate measure of inverse efficiency in any sector cannot decline if that sector increases the utilization of any of its own products. This observation becomes more interesting when one considers that the location of the vertical asymptote of the price curves for non-basic goods is a function of $\operatorname{eig}(A_1)$; and as such, this relationship between utilization (the entries of A_1) and the maximal eigenvalue is very important. It becomes even more significant when couched in terms of the production of medical care.

Proposition 2 *For any technology matrix structured as in* (21)*, if* $\rho(A_1) > \rho(A_0)$*, then the price curves for goods* $m+1$ *to* k *have a vertical asymptote at* $r^\star = \frac{1}{\rho(A_1)} - 1$ *with positive prices to the left of* $r^\star$*.*

Proof: Starting with equation (24), one can utilize the block lower-triangular form of A^T to write $\vec{p}_1$, the vector of non-basic commodity prices, as

$$\begin{aligned}
\vec{p}_1 &= \frac{\left(sI - A_1^T\right)^{-1}\left(A_{0\to 1}\left(sI - A_0^T\right)^{-1}\vec{\ell}_0 + \vec{\ell}_1\right)}{\vec{b}_0^T\left(sI - A_0^T\right)^{-1}\vec{\ell}_0} \\
&= \frac{\widehat{\left(sI - A_1^T\right)}\left(A_{0\to 1}\widehat{\left(sI - A_0^T\right)}\vec{\ell}_0 + \left|sI - A_0^T\right|\vec{\ell}_1\right)}{\left|sI - A_1^T\right|\vec{b}_0^T\widehat{\left(sI - A_0^T\right)}\vec{\ell}_0}
\end{aligned} \tag{26}$$

where $s \equiv \frac{1}{1+r}$ and $\widehat{M}$ is shorthand for the adjoint of M. Both the numerator and denominator are guaranteed to be positive so long as $s > \max\left(\rho(A_0), \rho(A_1)\right)$.

Returning to (24), it is evident that each commodity's price is a multilinear rational function. The numerator of (24) is a sum with each term being a multiple of the entries in $\vec{\ell}$; the denominator is also a summation, but there is only participation from the elements of $\vec{\ell}_0$. Thus, the numerator will generally include terms that are multiples of the elements of $\vec{\ell}_1$. Furthermore, the terms in the denominator will also have participation from the elements of $\vec{b}_0$. Given these observations, it is clear that only under the most pathological conditions will any discontinuity in the prices of a commodity be removable.

In the specific case of $\rho(A_1) > \rho(A_0)$, equation (26) reveals that when $s = \rho(A_1)$, the denominator will equal zero, but not all elements of $\widehat{sI - A_1^T}$ are likely to be zero. Because the discontinuities of a rational function are only removable if both the numerator and denominator are simultaneously zero, this is further evidence to show that the discontinuity at $s = \rho(A_1)$ will give rise to a vertical asymptote.

Since all the elements of $\vec{p}_1$ are guaranteed to be positive for $s > \rho(A_1)$, this means that as s approaches $\rho(A_1)$ from the right, the elements of $\vec{p}_1$ exhibit unbounded growth. Or, with respect to r, the unbounded growth appears as r approaches $r^\star$ from the left. •

As an aside, note that equation (26) will not guarantee increasing prices at $s = \rho(A_0)$ if $\rho(A_0) > \rho(A_1)$ since the denominator is not guaranteed to have a root at $\rho(A_0)$. At values of s less than $\rho(A_0)$, it is merely possible to have negative contributions from some of the terms in the numerator and denominator. This would indicate that decreasing values of the elements of $\vec{p}_1$ are possible for $s < \rho(A_0)$ when $\rho(A_0) > \rho(A_1)$. This jibes with the results from the two-commodity model.

Finally, the graph shown in figure 2 indicates that if $\text{eig}(A_1) > \text{eig}(A_0)$, then the prices of non-basic goods rise, regardless of the extant level of r. In the two-commodity case, this is a simple derivative. In the case of an arbitrary number of commodities, matters become more complex, requiring the following proposition:

Proposition 3 *If $\lambda_1^\star > \lambda_0^\star$, then the price of any luxury commodity is non-decreasing given an increase in any element of A_1 at any level of $0 \leqslant r < 1/\rho(A_1') - 1$ where A_1' is the perturbed value of A_1.*

Proof: Let k be the index of any non-basic commodity, and let a_{ij} be any element in the matrix A_1. To find $\frac{\partial p_k}{\partial a_{ij}}$, the following derivation applies:

$$
\begin{aligned}
\vec{p} &= (1+r)\left[A^T\vec{p} + w\vec{\ell}\right] \\
&= w\left(\frac{1}{1+r}I - A^T\right)^{-1}\vec{\ell} \\
&= w\begin{bmatrix} M_0^{-1} & 0 \\ M_1^{-1}A_{0\to 1}^T M_0^{-1} & M_1^{-1}\end{bmatrix}\vec{\ell}
\end{aligned}
$$

where the notation $M_i \equiv \frac{1}{1+r}I - A_i^T$ is adopted for brevity. This leads to the expression

$$\vec{p}_1 = wM_1^{-1}\left(A_{0\to 1}^T M_0^{-1}\vec{\ell}_0 + \vec{\ell}_1\right). \tag{27}$$

Next define A_1' to be A_1 with the element at a_{ij} increased by some arbitrarily small positive value ϵ. This implies that

$$\frac{1}{1+r}I - \left(A_1'\right)^T < \frac{1}{1+r}I - A_1^T,$$

and according to lemma 3, this in turn implies that

$$\left(\frac{1}{1+r}I - \left(A_1'\right)^T\right)^{-1} \geqslant \left(\frac{1}{1+r}I - A_1^T\right)^{-1}.$$

As a consequence, due to equation (27), the value of $\vec{p}_1$ rises as A_1 becomes perturbed to A_1'. Note that this result holds for any value of r for which $\frac{1}{1+r}I - (A_1')^T \in \mathscr{M}$. •

A.2.1 A Note on Many-Sector Economies

In the broadest case, the economy can be partitioned into multiple cascading sectors where the products of the less basic sectors do not feed into the production of more basic sectors, but do feed into the production of even less basic commodities. Such an organization leads to a general $n \times n$ block upper-triangular matrix form of A:

$$A = \begin{bmatrix} A_0 & A_{0\to 1} & \cdots & A_{0\to n} \\ 0 & A_1 & \cdots & A_{1\to n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_n \end{bmatrix}. \tag{28}$$

In this matrix structure, the wage-profit frontier is still defined in the most basic sector of the economy, the sector whose internal flows are represented by A_1. It is also true that the maximal eigenvalue, $\lambda^\star$, will still determine the valid range of r, and $\lambda^\star$ will be the largest eigenvalue of the set

$$\{\lambda | \lambda \in \mathrm{eig}(A_i), \forall i = 1, \ldots, n\}.$$

If this eigenvalue is not in $\mathrm{eig}(A_1)$, but in $\mathrm{eig}(A_j)$, then the price functions of the j^{th} sector have a vertical asymptote. This asymptote may also cascade to the k^{th} sector, $\forall k > j$ depending on whether the k^{th} sector employs commodities produced by the j^{th} sector.