J Prod Anal (2011) 35:119–128
DOI 10.1007/s11123-009-0167-1

A multi-product cost function for physician private practices

Timothy S. Gunning · Robin C. Sickles

Published online: 24 December 2009

Abstract This paper provides an empirical analysis of the production of physician services using a multi-product cost function. Prior studies examine the physician production process in a theoretical setting and do not provide empirical insight. We expand upon the theoretical work in the literature by specifying a four-product generalized Leontief cost function for physician services that recovers measures of marginal cost, scale, scope, and elasticity. Our study is based on physician survey data from the 1998 American Medical Association Physician Socioeconomic Monitoring Survey and motivates a scientific framework for advancing the existing reimbursement fee schedule. Our analysis indicates that physician office visits are generally priced above marginal cost, implying there may be evidence of market power in physician private practices. Furthermore, our analysis lends to the policy debate over whether the use of a Resource-Based Relative Value Scale system is the most appropriate mechanism for facilitating Medicare reimbursements.

Keywords Physician · Multi-product cost function · Medicare reimbursement

JEL Classification C30 · D24 · I12

Note: The views expressed in this paper are those of the author and not those of Ernst & Young LLP.

T. S. Gunning (✉)
Ernst & Young LLP, Transfer Pricing and Economics, 5 Times Square, New York, NY 10036, USA
e-mail: timothy.gunning@ey.com

R. C. Sickles
Department of Economics, Rice University, 6100 Main Street South, Houston, TX 77005, USA

1 Introduction

The production of physician services relies heavily upon revenue reimbursements from third-party payers. Figures reported by the American Medical Association in 2006 revealed that 560,000 physicians billed Medicare for $54 billion in federal reimbursements, accounting for 17% of all Medicare payouts. It has been reported that ~45% of physicians in the American Medical Association plan to decrease or stop their acceptance of new Medicare beneficiaries, due to the general belief that existing reimbursement schedules poorly approximate the marginal cost for physician services.[1] The Bush administration reduced government payments to physicians as a means for slowing the spending growth of Medicare. As of July 2008, the U.S. Congress approved a 1.1% increase in the payment rates associated with Medicare payouts; well below an inflationary adjustment. Unsurprisingly, this has not been a policy without detractors. Many in the opposition have suggested that continued cuts in spending will destabilize government programs and risk the availability of patient health care; an issue at the forefront of U.S. politics. Prior to these developments, the Justice Department asserted that large physician practices exhibit within- and between-network market power, which may result in fee schedules set at prices well above marginal cost.[2]

There are widely acknowledged differences among the political parties and health care providers with respect to the effectiveness of the existing reimbursement schedule,

[1] http://seniorjournal.com/NEWS/Medicare/6-07-19-ReductionInMedicare.htm. See the 2006 American Medical Association report on Medicare physician payments.

[2] See Haas-Wilson and Gaynor (1998) for an examination and summary of competition in physician markets.

with much of the debate focused around the Resource-Based Relative Value Scale (RBRVS) system. The RBRVS is an index that is used to approximate the relative value of the services performed by physicians when facilitating Medicare reimbursements. The relative values are updated by the Centers for Medicare and Medicaid Services (CMS) every 5 years to reflect changes in physician practice characteristics. The inputs used to construct the RBRVS index, as well the differences in reimbursement prices between generalists and specialists have been subject to criticism. With respect to the former, the RBRVS index is constructed using three inputs: the type of physician work performed, practice expense, and malpractice expense.[3] This alone is problematic because there is no compensation for the physician's quality of service and therefore no incentive for the physician to provide more than a minimal amount of attention to the patient. A second issue is the degree of specialization. In particular, there has been debate over differences in reimbursement for generalists versus specialists and whether these differences appropriately account for economic cost.[4]

The goal of this paper is to address these issues by proposing an econometric framework that could be used in a service-based fee setting. It has been noted that most estimates of physician practice costs are not developed by formal econometric modeling.[5] Instead they rely upon an inspection of accounting data that examines various economic measures of interest. Our analysis employs and expands upon many of the theoretical contributions in the literature to examine the issue empirically.

Escarce and Pauly (1998) provided a seminal contribution to the physician cost function literature by specifying a theoretical framework for the production process of physician services. This extended previous work by Gaynor and Pauly (1990). Prior physician cost studies, such as Pope and Burge (1995), were limited due to a lack of available data and the inability of the functional form to adhere to well-established properties consistent with economic theory. The theoretical work of Escarce and Pauly (1998) provided an illustrative empirical application; unfortunately, it too was limited due to many of these same issues.[6] The production study by Reinhardt (1972) provided the motivation for economic studies of cost and scale; however, it preceded many of the theoretical advancements in functional form flexibility, most notably, those of Diewert (1971). The recent work by Thurston and Libby (2002) revisited the earlier contributions of Reinhardt (1972) by specifying a production function framework that examines substitutability between various inputs in the physician production process. Gaynor and Vogt (2000) summarized many of the production specifications for health services. Escarce and Pauly (1998) suggested physician private practices exhibit regional market power, while Gunning and Sickles (2008) showed evidence of regional monopolies. This study applies the theoretical contributions of Escarce and Pauly (1998) to recover significant economic measures that could be used in a resource-based fee setting. In particular, we use a more recent and robust data set from the American Medical Association to examine the revenues received by physicians for their services and assess to what degree the existing fee schedules for physician reimbursements accurately reflect physician costs at the margin. Our study advances prior findings in the literature due to the increased flexibility of our cost specification and the precision of our estimates. We provide a comparison of our results against those of Escarce and Pauly (1998) to examine the change in marginal costs over time and examine how differences in methodology affect our findings. Our paper contributes to the highly publicized health care debate over an appropriate mechanism for compensating physicians with a market return for their services.

We propose a multi-product cost specification that is locally flexible and places no a priori restrictions on factor substitution elasticities. Escarce and Pauly (1998) detailed a theoretical framework examining these properties in their study of physician private practices. Li and Rosenman (2001) highlighted these advantages when they examined production function flexibility as opposed to other more restricted approaches used in policy discussions (Cowing et al. 1983). Elbasha and Messonnier (2004) highlighted the notion of constant returns to scale in health technologies due to the potential inability to replicate essential resources that are fixed in the short run. Fischer et al (2006) extended this framework by providing evidence of non-linearities in measures of cost and scale of health services, suggesting the need for a functional form that exhibits local flexibility. The primary advantage of functional flexibility in both theoretical and empirical settings is the measurements of scale, scope, and elasticity. Our specification identifies the aforementioned key measures and seeks to implement them into a relevant policy setting. In the next section, we provide the motivation for our empirical study. This is followed by Sect. 3, where we present our functional form and the econometric specification of the model. The data is described in Sect. 4 and the results are presented in Sect. 5 with an example and discussion around the implementation of our results. We conclude in Sect. 6.

[3] "Medicare physician fees geographic adjustment indices are valid in design, but data and methods need refinement." Washington, DC: Government Accountability Office, March 2005. Publication GAO-05-119.

[4] See Hsiao et al. (1988).

[5] Escarce and Pauly (1998).

[6] Escarce and Pauly (1998) note that their analysis is purely illustrative and should not be relied upon for empirical purposes.

2 Background

The theoretical specification of our model relies upon the observation that physician labor supply is the result of a utility maximization problem by the physician. In particular, because the physician production model is primarily that of a self-employed entrepreneur, physician time is most appropriately treated as an endogenous input into the production process. Therefore, the level of physician labor supply is presumed to be quasi-fixed; an assumption contained within the model. The result is a two-stage utility-maximization problem, yielding a system of two equations: the first equation optimizing physician time; and the second equation optimizing utility. An overview of the framework follows.

Physicians maximize a utility function comprised of net income and leisure, $U(I, L)$. They treat $Y = (Y_1, Y_2,\ldots, Y_N)$ as the outputs of production, representing each of the N outputs produced by the physician; $X = (X_1, X_2,\ldots, X_M)$ as the M inputs used in production, with each facing a cost of $W = (W_1, W_2,\ldots, W_M)$, respectively; and T as the supply of physician labor input. This results in the physician's net income, $I = R(Y) - (W \cdot X)$, and the physician's leisure, $L = t - T$, where t is the total time available to the physician.

In the first stage, output (Y) and the supply of physician labor (T) are treated as parametric, resulting in the physician selecting the quantity of physician practice inputs that maximize utility, yielding the conditional input demands, $X^*(Y, W, T)$. The conditional input demands are the quantities of non-physician labor inputs the physician purchases as a function of total output, input prices, and labor supply.[7] The first stage is represented by a traditional cost function, which has a dependent cost variable that is the sum of all of the costs incurred by the practice. Given the labor supply of the physician (T), the minimum cost of the inputs is $C(Y, W, T) = W \cdot X^*(Y, W, T)$. In the second stage, outputs are taken as parametric and the physician selects the optimal supply of labor (T) that maximizes the utility function of the physician, yielding the conditional demand for physician labor, which is equivalent to the level of physician labor supply produced, subject to the price of the inputs and the quantity of the outputs. Thus, the physician maximization problem is $U(R(Y) - C(Y, W, T), (t - T))$ over T. The solution to the second stage problem yields the conditional demand for physician labor, $T^*(Y, W)$. The marginal condition for utility maximization is $\frac{\partial C(Y,W,T*)}{\mathrm{d}t} = -\left(\frac{U_L}{U_I}\right)$, implying the physician can substitute his or her labor for non-physician labor inputs (i.e. all inputs except the physician's labor supply). Hence, the marginal opportunity cost of physician labor is the rate of substitution between leisure and income. The second stage is represented by a physician labor supply equation.[8]

There are differing views as to how a multi-product cost function should be specified, including what outputs to consider.[9] The primary output of a general practice physician is a single office visit. Office visits often constitute detailed consultations with new patients or regular examinations with established patients. Depending on the specialty of interest, physicians may also spend a considerable amount of time in a hospital setting, including regular emergency room trips and scheduled hospital rounds. Our model considers four primary outputs of interest: established patient office visits, new patient office visits, emergency room visits, and hospital visits. We considered a fifth output for miscellaneous services such as x-rays, test interpretations, and phone calls. However, these miscellaneous services are sparse, resulting in $\sim$10% of total cost to the practice. Data limitations prevented their inclusion.

Our model applies the generalized Leontief cost function first proposed by Diewert (1971). The generalized Leontief provides a direct theoretical representation of the physician data and the underlying production technology due to its treatment of the outputs and the ability to determine input substitutability with little difficulty. This is in contrast to the more common transcendental logarithmic (translog) functional form.[10] The translog in its original form is a second-order Taylor-series approximation to an arbitrary cost function (Caves et al. 1980) and is used most frequently in an empirical setting due to the relative ease required to impose the restrictions required by economic theory and the additional benefit of easily interpretable coefficients. Escarce and Pauly (1998) specified a translog with Box–Cox transformations on the outputs, with the Box–Cox transformation addressing the issue of many physicians reporting zero outputs.[11] However, in their illustration, many of the higher-order terms and cross-product terms were dropped to preserve degrees of freedom in their limited sample, leading to bias in their estimates and destroying the flexibility of the cost function. In the past, many applications of the translog have deferred to a "small-value translog", where the researcher arbitrarily assigns small values to outputs corresponding to zeros. However, that approach has a history of producing erratic

[7] Examples of non-physician inputs include nurses, clerical support, malpractice rents, and office supplies.

[8] We refer the reader to Escarce and Pauly (1998) for a formal derivation and a detailed discussion.

[9] See McFadden (1978).

[10] Or a variety of its extensions. See Berndt and Christensen (1979) or Christensen et al. (1971).

[11] We note these are true zero measures and not missing values.

measures (Berger et al 1999; Weninger 2003). An alternative to both the Leontief and the translog is the proper quadratic cost function or composite cost function first proposed by Roller (1990) and extended by Pulley and Braunstein (1992); however, the implementation of this specification is scarce and does not necessarily provide advantages over the Leontief and translog.[12]

Guilkey et al. (1983) noted that it is often the data and the underlying technology that determines the most appropriate specification, not the functional form itself.[13] Our specification is reminiscent of the multi-product generalized Leontief cost function used by Li and Rosenman (2001) in their two-product study of the hospital industry and thus reconciles issues related to true zero outputs.[14]

3 Model

We estimate the structure of production in the physician services market by jointly estimating the cost function and the physician labor function. Our motivation to estimate the entire system rather than the cost function in isolation is due to the empirical evidence of potential endogeneity bias in physician practice cost functions (Escarce 1996). The resulting specification is a restricted (variable) cost function, reminiscent of Sickles and Streitwieser (1998). The quasi-fixed input of physician labor is based on the underlying assumption that the physician is always at (or close to) their optimal level of labor supply, conditional upon peripheral market conditions. We note that the empirical interpretation of the theoretical model essentially allows for the existence of temporary disequilibria due to the presence of the quasi-fixed input. Temporary disequilibria may simply be due to unexpected demand shocks in the market, a presumption implicitly preserved in the model.

The objective of the physician is to minimize a cost function of physician and non-physician inputs, subject to a technological constraint,

$$\min \sum W_i X_i \quad \text{subject to} \quad G(Y, X, T) \le 0 \tag{1}$$

where G represents the transformation function of the production technology. In Eq. (1), Y represents the measures of output: annual office visits with established patients, new patients, emergency room trips, and hospital rounds.[15] Input prices, W, are for office rent, non-physician employee payments, and malpractice premiums and correspond to their respective quantities of input (X). The annual hours of labor provided by the physician are represented by T. The solution to Eq. (1) yields the cost function,

$$C = C(Y, W, T) \tag{2}$$

where C is the summation of practice costs incurred by the physician for his or her production of services. The cost function (C) is homogeneous of degree one and symmetric in factor prices (W), non-increasing and convex in the levels of the quasi-fixed factor (T), and non-negative and non-decreasing in output (Y). We approximate C by a multi-product generalized Leontief cost function, including practice-level characteristics that are believed to influence practice cost and dummy variables to control for differences in areas of physician specialization.[16]

The physician labor supply equation is represented by:

$$T = T(Y, W). \tag{3}$$

Output (Y) and input prices (W) are defined in Eq. (1). In addition to practice-level characteristics and dummy variables to control for heterogeneity in areas of physician specialization, we include physician-level characteristics that are expected to influence the supply of physician services, independent of their effects on non-physician costs. As with Escarce and Pauly (1998), we specify the physician labor supply function in the same manner as the cost function, in our case, a multi-product generalized Leontief. The decision to specify the labor supply equation is inspired by the observation that the physician is

[12] See McFadden (1979) for a summary.

[13] The globally flexible Symmetric Generalized McFadden proposed by Diewert and Wales (1987) was also considered. However the authors found its "flexibility" was distorted by the need to impose global concavity. Moreover, with the presence of several categories of outputs, the Symmetric Generalized McFadden produced estimates of technology that were unstable as well as economically unmeaningful without the restrictions. We opted instead for a generalized Leontief flexible functional form which circumvented the issues we encountered with the Symmetric Generalized McFadden.

[14] Li and Rosenman (2001) noted that the generalized Leontief has traditionally been used in a single output setting; however its extension to the multi-product setting has been seen in studies dating back to Hall (1973).

[15] We use office visits as physician output. There are other measures of physician output that are more comprehensive but less easily observed and measured. For a discussion of these alternatives, health outcomes that can be constructed for these alternatives, and their use in static and intertemporal models of health production see, for example, Sickles and Taubman (1997), Behrman et al. (1998), and Sickles and Yazbeck (1998).

[16] The authors note that hospital visits may be endogenous if physicians can accurately plan the time they commit in a hospital setting. We performed a Durban-Wu-Hausman test to examine the endogeneity of the hospital visits. Our results indicate that hospital visits are exogenous. This is most likely attributable to the unpredictable nature of the time required to consult with patients in a specialized setting, due to the potential complications that often arise in hospital visits.

maximizing their labor supply, subject to many of the non-physician input prices that appear in the cost function.[17]

The econometric specification for the cost equation faced by the physician is represented by a non-homothetic generalized Leontief:

$$
\begin{aligned}
C(Y,W,T) = \alpha_T T &+ \sum_{i=1}^{4}\left(\sum_{j=1}^{3}\sum_{k=1}^{3}\alpha_{i,j,k}W_j^{\frac{1}{2}}W_k^{\frac{1}{2}}\right)Y_i \\
&+ \sum_{i=1}^{4}\sum_{j=1}^{3}\beta_{i,j}W_jY_i + \sum_{i=1}^{4}\left(\sum_{j=1}^{3}\phi_{i,j}W_j\right)Y_i^2 \\
&+ \sum_{1\le i<l\le 4} Y_{il}\left(\sum_{j=1}^{3}\phi_{j,il}W_j\right) + \sum_{j=1}^{3}\gamma W_j \\
&+ \Gamma + \Psi \qquad (4)
\end{aligned}
$$

where Γ are practice controls that include binary variables for whether the physician practices in a metropolitan location and if the physician is board-certified. In addition, we include the percentage of patients that pay with Medicaid and the number of physicians in the practice. The four physician-reported dummy variables are represented by Ψ and describe the area of physician specialization: general practice, medical specialty, surgical specialty, and a fourth category for all remaining specialties.[18]

Differentiating the cost function with respect to each input price, yields the unconditional factor demands for the inputs. Hence, the factor demand function for any price (W_k) is:

$$
\begin{aligned}
X_k &= \frac{\partial C}{\partial W_k} \\
&= \frac{1}{2}\sum_{i=1}^{4}\sum_{j\neq k}\alpha_{i,j,k}\frac{W_j^{\frac{1}{2}}}{W_k^{\frac{1}{2}}}Y_i + \sum_{i=1}^{4}\sum_{k=1}^{3}\beta_{i,k}Y_i + \sum_{i=1}^{4}\phi_iY_i{}^2 \\
&\quad + \sum_{1\le i<l\le 4}\phi_{il}Y_{il} + \gamma W_k \qquad (5)
\end{aligned}
$$

The physician labor equation is specified as a non-homothetic generalized Leontief, yielding:

$$
\begin{aligned}
T(Y,W) = \sum_{i=1}^{4}\left(\sum_{j=1}^{3}\sum_{k=1}^{3}\alpha_{i,j,k}W_j^{\frac{1}{2}}W_k^{\frac{1}{2}}\right)Y_i &+ \sum_{i=1}^{4}\sum_{j=1}^{3}\beta_{i,j}W_j \\
&+ \sum_{i=1}^{4}\left(\sum_{j=1}^{3}\phi_{i,j}W_j\right)Y_i^2 \\
&+ \sum_{1\le i<l\le 4} Y_{il}\left(\sum_{j=1}^{3}\phi_{j,il}W_j\right) + \sum_{j=1}^{3}\gamma W_j + \Gamma \\
&+ \Psi + \zeta \qquad (6)
\end{aligned}
$$

where ζ are physician-specific heterogeneity controls. We include linear and quadratic terms for age and years of experience.

Our specification of (4) is advantageous because it maintains desirable theoretical characteristics. The generalized Leontief is homogeneous in input prices by construction and symmetry is imposed prior to estimation. It is a second-order approximation that is locally flexible and imposes no a priori restrictions on the elasticities of factor inputs. The flexible nature of the cost function allows our model to capture the degree of scale and scope economies, in addition to the potential non-linearity of marginal costs.

4 Data

The data for this study comes from the 1998 American Medical Association (AMA) Physician Socioeconomic Monitoring Survey (SMS) and the 1998 Geographic Practice Cost Index (GPCI) originally proposed by Zuckerman

Table 1 Summary statistics

	Mean	Standard deviation
Annual total cost	\$270,972	\$255,537
Annual hours of labor	2,336	724
Annual office visits (established patients)	4,174	2,492
Annual office visits (new patients)	637	620
Annual ER visits	243	586
Annual hospital visits	693	1,028
Relative non-physician price	1.02	0.03
Relative office price	1.07	0.13
Relative malpractice price	1.14	0.53
Percent practicing in metropolitan area	0.83	0.37
Number of physicians in practice	3.08	2.98
Percent of patients paying with medicaid	0.12	0.14
Percent graduated from board certified school	0.79	0.41
Experience* (scaled)	6.81	1.04
Age* (scaled)	2.52	1.03
Percent in general practice	0.25	0.43
Percent in nonsurgical specialty	0.61	0.48
Percent in surgical specialty	0.07	0.25
Percent in other specialty	0.07	0.26

* The SMS asks physicians whether they fall within a range of years for the experience and age variables. For instance, a "1" represents a physician with 1–5 years of experience, a "2" represents 6–10 years of experience, and so on. A similar approach is employed for the age variable with a "1" representing 20–29 years of age, a "2" representing 30–39" years of age and so on

[17] However, it is important to note that T is a representation of physician labor and does not satisfy the traditional properties of a cost function, hence there is no need to impose restrictions.

[18] This category represents only 7% of the sample.

et al (1990). Table 1 summarizes the data. The SMS is a telephone survey of 3,700 private practice physicians that are members of the AMA and practice in the United States. The SMS is a geographically comprehensive and occupationally detailed survey that provides physician, practice, and demographic characteristics for the physician private practice setting. Total cost is the sum of non-physician employee payments, insurance expenses related to malpractice, and office expenses. Non-physician employee payments include secretarial support, nurses, and assistants. Insurance expenses consist of malpractice premiums and any additional malpractice costs associated with the practice.[19] Office expense is the cost of leasing, renting, or owning the infrastructure that the practice is located in. The office expense variable also includes any rents related to the lease or ownership of technological equipment. The physician labor variable is constructed by multiplying the average hours worked per week, as reported by the physician, by the number of hours practiced in the 1998 calendar year. We use the same methodology to construct the four outputs, by way of multiplying reported weekly outputs by the number of weeks worked. All demographic and sociological characteristics are reported by the physician and are detailed in the previous section.

The SMS does not contain information on input prices. The GPCI is a Laspeyres index that is designed to control for price fluctuations in health markets by assigning weighted values to three input categories: office rent, non-physician wages, and malpractice rents. Pope and Burge (1990) suggested deflating the dependent cost variable by the GPCI to preserve linear homogeneity in prices. However, that method has a number of limitations.[20] Escarce and Pauly (1998) used two sub-components of the index; office rent and non-physician wages, allowing the malpractice subcomponent to be the numeraire. We use all three subcomponents of the GPCI for our input prices. Our study examines the demographic variables of the SMS to determine the geography of the practice and assigns geographic prices to the practice accordingly. Since there is reasonable variation in malpractice premiums by location, we include the malpractice subcomponents as a unique measure of price, rather than presume it as the implicit numeraire.

Our sample considered only those physicians who practiced at least 20 hours or more a week. Those respondents who spent the majority of their time in a hospital or a school setting were not considered for this study. Our final sample consists of 939 private practice physicians that practice strictly in the United States.

5 Estimation

The four-equation system consists of the cost equation (4), two of the three demand equations (5), and the physician labor equation (6).[21] We append additive error terms to the cost, labor, and demand equations. The system is estimated via three-stage least squares (3SLS).[22] Table 2 provides parameter estimates with standard errors in parentheses. Escarce and Pauly (1998) have showed that the physician characteristics unique to the physician labor equation but excluded from the cost equation, identify the cost function. In our case, those are linear and quadratic terms for age and experience (ζ).[23] Linear homogeneity in prices is achieved by construction, while symmetry is imposed prior to estimation. The cost function is concave in prices and in the quasi-fixed factor. We test the null hypothesis of non-homotheticity of the cost equation and the labor supply equation by testing the joint significance of the price/output terms and reject it at the 99% confidence level, implying there is evidence of price/output interaction in both the cost function and labor supply function.

Table 3 reports our measures for marginal cost for each of the four outputs. We also report the marginal opportunity cost of an hour of physician labor. As noted by Li and Rosenman (2001), there are two common approaches for determining marginal cost estimates at the mean. The traditional approach is to use the mean estimates of the data and the estimated coefficients from the cost equation to derive a point estimate for the four outputs of interest. Since the marginal costs are non-linear in our model, we prefer the method by Li and Rosenman (2001), which consists of using the estimated coefficients from the cost equation to recover the marginal costs by observation,

$$\mathrm{MC}_s(Y_i) = \sum_{j=1}^{3} \beta_{i,j} W_j + 2Y_i\left(\sum_{j=1}^{3} \phi_{i,j} W_j\right) + \sum_{i \neq l} Y_l \left(\sum_{j=1}^{3} \phi_{j,l} W_j\right) \tag{7}$$

where $\mathrm{MC}_s(Y_i)$ represent the S data points for output Y_i. The S data points are then averaged over the sample with

[19] Malpractice insurance may include legal fees associated with malpractice cases.

[20] The method by Pope and Burge (1990) is rather problematic, since it precludes measurement of share equations, economies of scale and scope, and the interaction of second-order prices. The result is a functional form that does not adhere to economic theory.

[21] One of the demand equations is excluded to avoid singularity.

[22] 3SLS was selected due its notable history in simultaneous equation models. It is also worth noting that in our case, 3SLS is asymptotically equivalent to the Generalized Method of Moments estimator.

[23] A regression test of the overidentifying restrictions found that these restrictions could not be rejected (χ^2=2.60; p=0.63).

Table 2 Estimation results

	Cost	Physician labor
Labor	−53.511 (35.494)	–
P1	848,933 (777,956)	−4,014.2 (3,675.6)
P2	228,677 (238,845)	151.05 (1,149.2)
P3	63,100 (33,380)	270.48* (154.5)
P1Y1	−327.18 (967)	2.325 (4.61)
P2Y1	−19.57 (899.4)	2.546 (4.29)
P3Y1	−6.56 (20.3)	−0.055 (0.097)
P1P2Y1	363.24 (1,862)	−4.77 (8.89)
P1P3Y1	339.85* (185.8)	0.052 (0.09)
P2P3Y1	−336.56** (170.6)	−0.006 (0.818)
P1Y2	−84,383* (4,930.9)	−22.44 (23.20)
P2Y2	−5,400.6 (4,342.3)	−16.27 (20.59)
P3Y2	37.78 (125.98)	−0.162 (0.604)
P1P2Y2	14,129 (9,231.6)	39.3 (43.59)
P1P3Y2	988.15 (1,066.1)	3.01 (5.06)
P2P3Y2	−1,201.9 (944.2)	−3.3 (4.48)
P1Y3	11,533.13** (5,111)	32.69 (23.7)
P2Y3	9,219.9** (4,843.7)	30.32 (22.6)
P3Y3	149.87 (173.68)	−0.589 (0.827)
P1P2Y3	−20,554.6** (9,876.6)	−63.3 (45.92)
P1P3Y3	−276.09 (1,287.4)	−0.423 (6.18)
P2P3Y3	−63.79 (1,160.2)	1.487 (5.574)
P1Y4	−1,812.7 (2,211.7)	−7.34 (10.49)
P2Y4	−843.22 (2,132.15)	−6.916 (10.159)
P3Y4	−7.367 (72.63)	−0.102 (0.348)
P1P2Y4	2,691.3 (4,333.5)	14.672 (20.61)
P1P3Y4	636.71 (538.3)	−0.563 (2.572)
P2P3Y4	−623.63 (447.3)	0.651 (2.136)
P1Y1Y1	−0.0.22 (0.016)	−0.00007 (0.00007)
P2Y1Y1	0.008 (0.009)	0.00011 (0.00008)
P3Y1Y1	−0.001 (0.126)	−0.00005 (0.00004)
P1Y2Y2	−0.165 (0.185)	−0.00093 (0.00087)
P2Y2Y2	0.127 (0.259)	0.0011 (0.0012)
P3Y2Y2	−0.001 (0.126)	−0.00024 (0.0006)
P1Y3Y3	−0.107 (0.33)	−0.0017 (0.0016)
P2Y3Y3	0.197 (0.46)	0.0022 (0.0022)
P3Y3Y3	−0.095 (0.169)	−0.00051 (0.00081)
P1Y4Y4	0.078 (0.062)	0.00036 (0.00029)
P2Y4Y4	−0.099 (0.088)	−0.0005 (0.00041)
P3Y4Y4	0.016 (0.038)	0.0001 (0.0002)
P1Y1Y2	0.134*** (0.057)	0.00025 (0.00027)
P1Y1Y3	−0.054 (0.567)	−0.0004 (0.0003)
P1Y1Y4	−0.006 (0.034)	0.00031** (0.00015)
P1Y2Y3	−0.161 (0.34)	0.0037 (0.0015)
P1Y2Y4	0.335* (0.196)	−0.0012 (0.0009)
P1Y3Y4	0.087 (0.119)	0.0004 (0.0006)
P2Y1Y2	−0.122*** (0.052)	−0.0003 (0.0002
P2Y1Y3	0.052 (0.51)	0.0003 (0.00024)
P2Y1Y4	0.005 (0.031)	−0.0003 (0.0001)
P2Y2Y3	0.225 (0.305)	−0.0034*** (0.0013)
P2Y2Y4	−0.395** (0.175)	0.001 (0.0008)
P2Y3Y4	−0.059 (0.123)	−0.0004 (0.0006)
P3Y1Y2	0.002 (0.006)	0.0002 (0.00003)
P3Y1Y3	−0.005 (0.008)	0.000003 (0.00039)
P3Y1Y4	0.003 (0.004)	−0.000003 (0.0002)
P3Y2Y3	0.003 (0.036)	0.0002 (0.0002)
P3Y2Y4	0.023 (0.025)	0.0002 (0.0001)
P3Y3Y4	−0.007 (0.024)	−0.00003 (0.0001)
SPEC1	−37,178** (19,265)	5,749.1** (2,882.7)
SPEC2	78,361*** (17,338)	5,856.27** (2,885.6)
SPEC3	–	5,876.9** (2,884.4)
SPEC4	820.65 (22,955)	5,779** (2,890.2)
METRO	−25,434* (15,889)	−1.244 (76.39)
DOCNUM	1,938.4 (1,514.1)	−8.532 (7.132)
MEDCAD	−127.9 (309.6)	−0.861 (1.478)
CERT	1,548.6 (11,082)	−99.42* (53.69)
AGE	–	129.36 (120.19)
AGESQ	–	−30.13 (22.19)
EXP	–	−200.3 (306)
EXPSQ	–	17.39 (22.33)
CONSTANT	853,515 (591,736)	–

Standard errors appear in parentheses

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

standard errors appearing in parentheses below the estimates[24]:

$$\mathrm{MC}(Y_i) = \frac{\sum_{s=1}^{S} \mathrm{MC}_s(Y_i)}{S}. \tag{8}$$

All four of our reported measures of output are significant at the 1% level. The marginal cost to the physician for an additional visit with an established patient is \$27.83, while the additional cost of an office visit with a new patient is \$75.05. The cost of an emergency room visit at the margin is \$27.73 and the marginal cost of a hospital visit is only \$16.02. Table 3 also compares our estimates to Escarce and Pauly (1998) after inflating their 1987 estimates to 1998 dollars with the Consumer Price Index (CPI). The reported measure for hospital visits and emergency room visits is almost the same. Our measure for new patient

[24] Standard errors are calculated by taking the square root of the variance for all four measures of output and then dividing by the square root of the sample size. An alternative methodology is to interact the coefficients derived from the cost function against the statistical means of the data to derive a marginal cost "at the mean", with standard errors derived by way of the Delta Method. We note that this approach produced roughly the same results.

Table 3 Marginal costs calculated at the mean

Outputs	1998 G&S	1998 E&P (adjusted for inflation)
Established patient	\$27.83*** (1.43)	\$32.12*** (0.69)
New patient	\$75.05*** (3.30)	\$87.03* (35.71)
Emergency room	\$27.73*** (3.00)	\$29.55 (59.91)
Hospital	\$16.02*** (2.05)	\$17.61** (9.62)

Standard errors appear in parentheses

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

visits is noticeably lower, while our marginal cost measure for established patient office visits is slightly higher.[25]

Since the labor variable and cost variable are entered into the cost specification linearly, the labor coefficient is the marginal cost for an additional hour of physician labor, holding all other factors constant. The labor coefficient is negative, consistent with theory, implying that the physician would have to substitute his or her labor to reduce total practice costs while maintaining the same level of productivity. Our estimate for the marginal cost of an additional hour of physician labor is \$53.51.

We test the null hypothesis of constant returns to scale. Escarce and Pauly (1998) found evidence of increasing returns with respect to physician outputs.[26] The literature in health services is reviewed in Gaynor and Vogt (2000), with Kass (1987) finding little evidence of economies of scale in the market for home health services and scale economies in the market for hospital production being found by Vita (1990) and Cowing and Holtmann (1983).[27] Economies of scale are measured by inverting the sum of the elasticities of output. Panzar and Willig (1977) showed that economies of scale for a production process with a quasi-fixed factor reduces to

$$\mathrm{SCALE} = \frac{C(W, Y, T^{*}) - t^{*}\left(\frac{\partial C}{\partial T}\right)}{\sum_{i=1}^{4} \frac{\partial C}{\partial Y_i} \cdot Y_i}. \tag{9}$$

A value greater (less) than unity implies a greater (lesser) than proportional expansion along the vector of outputs. Pauly and Escarce (1998) found a point estimate of 1.67 for the degree of ray economies of scale, thus implying increasing returns. Our results also yielded a positive point estimate of 2.95 ($z = 18.66$, $p < 0.001$), indicating increasing returns.[28] Therefore a 10% increase in output implies a 3.4% increase in practice cost. This can also be interpreted as 34% of practice costs vary directly with output, while the remaining 66% are fixed.

Economies of scope can be computed a number of ways. Li and Rosenman (2001) suggest the method by Vita (1990), which indicates that weak cost complementarities are a sufficient condition for economies of scope (i.e. if $\frac{\partial^2 C}{\partial Y_i \partial Y_j} < 0$ for all $i \neq j$). However, summing the coefficients of the cross-product terms provides little economic interpretation for the sensitivity that private practices exhibit with respect to specialization. We favor a more traditional approach:

$$\mathrm{SCOPE} = \frac{\sum_{i=1}^{4} C(y_i, 0) - C(Y)}{C(Y)}. \tag{10}$$

The above percentage describes the additional cost to the practice for producing the outputs separately, rather than together.[29] Expression (10) yielded a modest measure of 0.007 and was not statistically significant from zero, implying that costs could increase 0.7% if the practice produced the outputs separately. We note that our cost function controlled for fields of specialty. Our result implies there is little advantage *within* a specialty to produce outputs separately, but does not necessarily imply that physician practices can not exhibit scope economies *across* specialties.

Allen partial own-price and cross-price elasticities are reported in Table 4 (Allen 1938). Table 5 reports Allen-Uzawa partial elasticities of substitution (Uzawa 1962). Due to symmetry in prices, we report only the upper triangular matrix. The own-price elasticities for non-physician wages and office rent have the correct sign (−) and are statistically significant at the 1% level. Non-physician wages and office rents are relatively inelastic, implying physician private practices are rather unresponsive to a change in price.[30] The own-price elasticity for malpractice insurance is incorrectly signed (+), but statistically insignificant. Cross-price elasticities associated with non-physician wages are positive and significant at the 1% level, implying that practices substitute non-physician employees for office rent and malpractice premiums. Malpractice premiums are typically determined by the level of skill and duties of the physician. Hence a decrease in non-physician labor requires an increase in physician labor, resulting in higher demand for malpractice associated inputs. The

[25] The limited sample size and inflexibility of the Escarce and Pauly (1998) specification is most likely a contributing factor in the reported differences.

[26] In theory, physician private practices exhibit returns based on outputs and practice size. Our study focuses on the former.

[27] See Elbasha and Messonnier (2004) for a summary.

[28] It is well-known that these measures, as ratios of terms whose denominator is not bounded away from zero, do not have finite

Footnote 28 continued
moments. We use a standard trimming proportion of 2.5% in each tail of the empirical distribution in calculating the mean and standard deviation of the sample statistics for scale.

[29] We note that values of zero are within the range of outputs.

[30] This result is most likely attributable to long-term contractual obligations on office rent and office equipment, in addition to a high demand for specialized assistants.

Table 4 Own-price and cross-price elasticities

Inputs	Non-physician wages	Office rent	Malpractice
Non-physician wages	−0.222***	0.175***	0.047***
Office rent	0.450**	−0.237***	−0.214*
Malpractice	1.250	−1.287	0.037

Standard errors appear in parentheses

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

Table 5 Allen-Uzawa own-price and cross-price elasticities of substitution

Inputs	Non-physician wage	Office rent	Malpractice
Non-physician wages	−0.419***	0.463***	0.506***
Office rent	–	−0.626***	−2.310*
Malpractice	–	–	0.405

Standard errors appear in parentheses

* $p < 0.10$; *** $p < 0.01$

cross-price elasticity on office rent and malpractice is positive, but barely achieved significance at the 10% level.

6 Conclusion

This paper has examined the physician production process in a multi-product cost setting. We found the generalized Leontief functional form responds well to our physician-level data. Gaynor and Pauly (1990) and Escarce and Pauly (1998) made substantial steps in developing a theoretical model for physician private practices with the primary motivation being the potential endogeneity bias of physician labor. This paper demonstrates how to use much of the theoretical motivation of the literature to derive plausible estimates of the marginal cost for a variety of outputs that a private practice physician performs on a regular basis. Our results indicate that marginal costs are the highest for office visits with new patients and lowest for visits in a hospital setting. We found evidence of rising returns, with respect to scale economies. However, it does not appear there are economies of scope within specialties. In addition, our elasticity findings revealed that physicians rely heavily upon the inputs of production and are relatively insensitive to changes in price. Our results could be viewed as a motivation for the development of a resource-based fee schedule that relies on formal econometric modeling, rather than a study of accounting data.

The existing RBRVS system reimburses physicians based on the geographic practice costs they face and the relative-value units (RVU) of procedures. The RVUs are determined by the AMA Relative Value Update Committee and are updated by CMS every 5 years. The interaction of the GPCIs and RVUs are then multiplied by a conversion factor to yield the final reimbursement price. For example, an office visit reimbursement would be determined by the physician's GPCIs and RVUs corresponding to the three components of the RBRVS index (work, practice expense, and liability insurance), in addition to a conversion factor:

$$\left((\mathrm{GPCI}_{\mathrm{work}} \cdot \mathrm{RVU}_{\mathrm{work}}) + (\mathrm{GPCI}_{\mathrm{prac}} \cdot \mathrm{RVU}_{\mathrm{prac}}) + (\mathrm{GPCI}_{\mathrm{ins}} \cdot \mathrm{RVU}_{\mathrm{ins}})\right) \cdot (\mathrm{CF}) = \text{payment.}$$

The RBRVS methodology is intended to *competitively* reimburse physicians for their cost of service based on the aforementioned three key components. As discussed earlier, this approach is problematic because there is no compensation for the physician's quality of service and therefore no incentive for the physician to provide more than a minimal amount of attention to the patient. Our econometric specification suggests that physician-level data could be used to construct an empirical model that controls for such differences and is consistent with economic theory. This model could be used to corroborate or possibly replace the existing accounting-based system, thus resulting in more competitive reimbursements that encourage the efficient allocation of resources across private practices. Specifically, it could be estimated by specialty to control for differences in competition. Moreover, with a more detailed survey that examines malpractice lawsuits and settlements, the issue of quality could be addressed as well.[31]

As noted in our introduction, it has been a contentious debate as to whether physicians are compensated close to marginal cost. Much of this debate is around the effectiveness of the existing reimbursement schedule, in particular, the RBRVS system. Table 6 suggests that physicians are compensated well above marginal cost for office visits with established patients, office visits with new patients, and hospital-based visits. This is most likely due to two factors: the distribution of fixed costs versus variable costs and possible market power exhibited by large physician groups and specialists. Our findings of economies of scale suggest that compensating physicians at marginal cost may not appropriately reimburse physicians for the high fixed costs often faced. However, with mark-ups on marginal cost on the order of 135% and 390% for new patient office visits and hospital visits, respectively, one can not dismiss the implication of market power as proposed by the Justice Department in 1998.[32] Our empirical findings suggest a more in-depth study on the impact of collusion in physician

[31] The existing survey lacks the necessary data to examine this.

[32] For example, see The United States of America vs. The Marshfield Clinic.

Table 6 Price markups (price vs. marginal cost at the mean)

Outputs	1998 Price	1998 MC	% Markup
Established patient	$65.30	$27.83***	135
New patient	$110.00	$75.05***	46
Hospital visit	$78.52	$16.02***	390

Standard errors appear in parentheses

*** $p < 0.01$

private practices is warranted and could prove highly beneficial in the ongoing debate over health care and its costly implications.

References

Allen R (1938) Mathematical analysis for economists. Macmillan, London

Behrman J, Sickles RC, Taubman P (1998) The causes, some methodological approaches and substantive analysis: correlates and consequences of death among older adults. Kluwer, Massachusetts

Berger A, Demsetz R, Strahan P (1999) The consolidation of the financial services industry: causes, consequences and implications for the future. J Bank Finance 23:135–195

Berndt E, Christensen L (1979) The translog function and substitution of equipment, structures, and labor in U.S. manufacturing. J Econom 1:81–114

Caves D, Christensen L, Trethaway M (1980) Flexible cost functions for multi-product firms. Rev Econ Stat 62:477–481

Christensen L, Jorgensen D, Lau L (1971) Conjugate duality and the transcendental logarithmic production function. Econometrica 39:255–256

Cowing T, Holtmann A (1983) Multi-product short-run hospital cost functions: empirical evidence and policy implications from cross-section data. South Econ J 49:637–653

Cowing T, Holtmann A, Powers S (1983) Hospital cost analysis: a survey and evaluation of recent studies. In: Scheffler R (ed) Advances in health economics and health services research, vol 4. JAI Press, Greenwich, pp 257–303

Diewert W (1971) An application of the Shephard duality theorem: a generalized Leontief production function. J Pol Econ 79:481–507

Diewert W, Wales T (1987) Flexible functional forms and global curvature conditions. Econometrica 55:43–68

Elbasha E, Messonnier M (2004) Cost effectiveness analysis and health care resource allocation: decision rules under variable returns to scale. Health Econ 13:21–35

Escarce J (1996) Using physician practice cost functions in payment policy: the problem of endogeneity bias. Inquiry 33:66–78

Escarce J, Pauly M (1998) Physician opportunity costs in physician practice cost functions. J Health Econ 17:129–151

Fischer A, Laking G, Lord J (2006) Non-linearity in the cost-effectiveness frontier. Health Econ 15:565–577

Gaynor M, Pauly M (1990) Compensation and productive efficiency in partnerships: evidence from medical group practice. J Pol Econ 98:544–573

Gaynor M, Vogt W (2000) Antitrust and competition in health care markets. In: Culyer A, New House J (eds) Handbook of health economics, vol 1B. Amsterdam, North-Holland

Guilkey D, Lovell C, Sickles RC (1983) A comparison of the performance of three flexible functional forms. Int Econ Rev 24:591–616

Gunning T, Sickles R (2008) Competition and market power in physician private practices. Mimeo, Rice University

Haas-Wilson D, Gaynor M (1998) Physician networks and their implications for competition in health care markets. Health Econ 7:179–182

Hall R (1973) The specification of technology with several kinds of output. J Pol Econ 81:878–892

Hsiao W, Braun P, Dunn D, Becker E, DeNicola M, Ketcham T (1988) Results and policy implications of the resource-based relative-value study. N Engl J Med 13:881–888

Kass D (1987) Economies of scale and scope in the provision of home health services. J Health Econ 6:129–146

Li T, Rosenman R (2001) Estimating hospital costs with a generalized Leontief cost function. Health Econ 10:523–538

McFadden D (1978) Cost, revenue, and profit functions. In: Fuss M, McFadden DL (eds) Production economics: a dual approach to theory and applications, vol 1. North Holland, Amsterdam, pp 2–109

Panzar J, Willig R (1977) Economies of scale in multi-output production. Quart J Econ 91:481–493

Pope G, Burge R (1995) The marginal practice cost of physicians' services. J Socioecon Plann Sci 29:1–16

Pulley L, Braunstein Y (1992) A composite cost function for multi-product firms with an application to economies of scope in banking. Rev Econ Stat 74:221–230

Reinhardt U (1972) A production function for physician services. Rev Econ Stat 54:55–56

Roller L-H (1990) Proper quadratic cost functions with an application to the bell system. Rev Econ Stat 72:202–210

Sickles RC, Streitwieser M (1998) The structure of technology, substitution and productivity in the interstate natural gas transmission industry under the natural gas policy act of l978. J Appl Econom 13:377–395

Sickles RC, Taubman P (1997) Mortality and morbidity among adults and the elderly. In: Rosenzweig M, Stark O (eds) The handbook of population and family economics. North-Holland, Amsterdam

Sickles RC, Yazbeck A (1998) On the dynamics of demand for leisure and production of health: evidence from the retirement history survey. J Bus Econ Stat 16:187–197

Thurston N, Libby A (2002) A production function for physician services revisited. Rev Econ Stat 84:184–191

Uzawa H (1962) Production functions with constant elasticities of substitution. Rev Econ Stud 29:291–299

Vita M (1990) Exploring hospital production relationship with flexible functional forms. J Health Econ 9:1–21

Washington, DC: Government Accountability Office (2005) Medicare physician fees geographic adjustment indices are valid in design, but data and methods need refinement. Publication GAO-05-119

Weninger Q (2003) Estimating multi-product costs when some outputs are not produced. Empir Econ 28:753–765

Zuckerman S, Welch W, Pope G (1990) A geographic index of physician practice costs. J Health Econ 9:39–69