2019-03

Measuring Physicians' Response to Incentives: Labour Supply, Multitasking, and Earnings

Bruce Shearer
Nibene Habib Somé
Bernard Fortin

Avril / April 2019

Centre de recherche sur les risques
les enjeux économiques et les politiques publiques

www.crrep.ca

Abstract

We measure the response of physicians to monetary incentives using matched administrative and time-use data on specialists from Québec (Canada). These physicians were paid fee-for-service contracts and supplied a number of different services. We model physician behaviour and derive a conditional earnings function that returns the maximum earnings a physician can generate in the labour market, conditional on total hours worked. The earnings function is estimated using both limitedinformation methods and full-information methods. Limited-information methods impose fewer restrictions on the data, but are less informative over incentive effects. Le Chatelier effects imply that they identify lower bounds to the own-price substitution effects. Full-information methods explain earnings and hours simultaneously. They identify the full response to incentives, including income effects. Our results confirm that physicians respond to financial incentives. The own-price substitution effects of a relative price change are both economically and statistically significant. Income effects are present, but are overridden when prices are increased for individual services. They are more prominent in the presence of broad-based fee increases. In such cases, the income effect empirically dominates the substitution effect, which leads physicians to reduce their supply of services.

JEL Classification: I10, J22, J33, J44.
Keywords: Physician labour supply, multitasking, earnings function, Le Chatelier effects, incentive pay.

Bruce Shearer : Département d'économique, Université Laval, CRREP, CIRANO and IZA. bruce.shearer@ecn.ulaval.ca

Nibene Habib Somé : Department of epidemiology and biostatistics, Western University and Institute for clinical evaluative sciences (ICES). nsome@uwo.ca

Bernard Fortin : Département d'économique, Université Laval, CRREP, Industrial Alliance Research Chair on the Economics of Demographic Change, CIRANO. bernard.fortin@ecn.ulaval.ca.

Financial support from CIHR and FRQSC is gratefully acknowledged. We thank participants to workshops and seminars in Paris School of Economics and Canadian Health Economics Study Group for useful discussion.

1 Introduction

Physician labour supply is distinct from many other settings. Physicians are typically paid according to their output, which implies hours worked are an input into the production of services. Moreover, physicians must allocate their time across different services, introducing a multitasking element to their labour-supply decisions.[1] Within this setting, physicians can alter the supply of services by changing total hours of work or by reallocating a fixed number of hours to different services. Knowledge of the sensitivity of these responses to changes in prices has important policy implications, particularly when health care is provided in the public sector. At least two issues emerge. First, beginning with Feldstein (1970), Rizzo and Blumenthal (1994) and, more recently, Baltagi, Bratberg, and Holmas (2005) and Kalb, Kuehnle, Scott, Cheng, and Jeon (2018), much attention has been paid to characterizing the shape of physician labour-supply curves and the resulting implications for policies aimed at increasing the total supply of services (*e.g.*, Sloan, 1975). A second issue is whether a change in the relative fee paid for a particular service leads physicians to increase the supply of that service. An ageing population is likely to increase the demand for services such as cardiovascular treatments, cataract surgeries, and hip replacements. Since training more physicians takes time, monetary incentives can provide the government with a policy tool to meet short term demand changes.

In this paper we develop and estimate a physician labour-supply model which incorporates the production of services as a function of hours worked. We analyse physician choices over the total hours they spend seeing patients and the manner in which those hours are allocated to different services, which we refer to as multitasking. Our model gives rise to a conditional earnings function, which returns the maximum earnings a physician can generate in the labour market for a given number of hours worked. The conditional earnings function takes total hours as given and explains their allocation across different services. It depends on a wage index that measures the marginal return to an hour worked when that hour is optimally allocated across different services.

We characterize the properties of the conditional earnings function, concentrating on the information it contains regarding physicians' reaction to monetary incentives. We show that, when evaluated at optimal hours worked, it identifies a lower bound to the own-price substitution effects of the physician supply of services. The lower bound results from the fact that changes in optimal hours, subsequent to a change in relative prices, reinforce the own-price substitution effect – these are Le Chatelier effects (Samuelson,

[1]Multitasking was introduced to economics by Holmstrom and Milgrom (1991), who analysed workers' allocation of effort across different tasks within an agency framework.

1947; Milgrom and Roberts, 1996).

We estimate the earnings function for a specific economic model, using a sample of physicians working in the province of Québec, Canada. Our data contain information on the number of services completed by individual physicians, along with the fees paid for those services and hours worked. We pay particular attention to how physicians react to changes in the relative prices paid for services. Relatively little is known about these effects, particularly with respect to the importance of income and substitution effects. Some studies have looked at the supply of isolated services in response to variation in remuneration rates. For example, Allin, Baker, Isabelle, and Stabile (2015) found that the propensity to deliver babies by Cesarean section across Canadian provinces was sensitive to the relative price paid to physicians for completing that service.[2] The natural-experiment empirical approach exploited by the authors provides robust evidence of the total reaction to incentives, but does not distinguish between income and substitution effects (see Blundell and Macurdy, 1999). Other studies have relied on geographically-aggregated service data to analyse the effect of fee changes. Hurley and Labelle (1995) considered how changes in the relative fee paid for given services affected the completion rates of those services in Canadian provinces. They found little consistency in results across services, either in terms of the statistical significance of the relative fee as a determinant of the utilization rate, or in the direction of the effect.

Our model specifies a CES utility function for physicians, defined over income and leisure. Observable heterogeneity in preferences is accounted for as the utility weight for leisure depends on physician characteristics. CES preferences have a rich history of use in empirical labour-supply models, beginning with Stern (1976) and Zabalza (1983). This function is general enough to permit unrestricted responses to incentives, and to identify both income and substitution effects, yet it is parsimonious in parameters, allowing for simple and direct interpretations of the results.[3] We model the production of each service as a power function of the hours devoted to that service and physician characteristics.

The multitask setting generalizes the traditional notion of substitution effects to capture the manner in which hours worked are allocated across services. The substitution effect operates through two separate channels in this setting. First, as prices change, physicians alter their supply of services to maximize income. For a given number of hours worked, physicians allocate more time towards those services for which the relative price

[2]Gruber and Owings (1996) found that the propensity to complete Cesarean sections across American states was inversely related to fertility rates, consistent with physician-induced demand (Evans, 1974).

[3]Stern (1986) provides an excellent discussion of the properties of CES and other utility functions and their uses in labour-supply models.

has risen. Second, the reallocation of hours across services increases the marginal return to an hour of work, causing total hours to increase. These additional hours are then allocated across different services optimally, including the service whose price has changed.

We estimate two versions of the conditional earnings function, one using limited-information methods and the other using full-information ones. First, we condition on observed total hours worked, to estimate earnings and the conditional supplies of services. Conditional demands and supplies have been introduced in other settings by Pollak (1969, 1971) and estimated notably by Lundberg (1988) and Browning and Meghir (1991). Here we use limited-information estimation which places relatively few restrictions on the data and can be accomplished using (non-linear) least squares or instrumental variables. Yet, since the variation in hours is not explained within the model, income effects are not identified. The lower bound to the substitution effect is identified, but the full effect operates through changes in hours worked. Explaining their variation requires modelling the choice of hours by individual physicians. We therefore turn to full-information methods which explain earnings and hours (through an hours function) simultaneously. This requires additional assumptions, but has the advantage of identifying the full response of physicians to changes in relative prices, including both income and substitution effects. Our model is non linear, without a closed form solution for optimal hours. We therefore use numerical methods and incorporate them into our estimation procedure through the simulated method of moments.

We apply our model to administrative data collected on specialist physicians working in the province of Québec between 1996 and 2002. These data include detailed information on the number of services provided per quarter by individual physicians, the prices paid for these services and physician earnings. These data were matched to time-use survey data provided by the physicians which include information on the number of hours worked per week. The prices for services are set by the government and apply to all physicians in our sample. One important advantage of this feature for the econometrician is that it is reasonable to assume that these prices are exogenous at the physician level. In addition, the Québec government altered the fee schedule in 2001, changing the relative prices paid for different services. We exploit this variation in prices and incentives to identify our model.

Our results suggest that physicians do react to incentives. The own-price elasticities vary across services, but are positive and statistically significant for all services. We find some evidence that observed price variation results from technological changes which alter the productivity of physicians. Yet, taking account of technological change has little effect on our results. We exploit our model to highlight the difference in reaction to changes in the fees for individual services and to broad-based fee increases. Changes in the fees for individual services have positive effects on the supply of those services as substitution effects outweigh income effects. However, broad-based fee increases have negative effects on the supply of services as the income effect dominates. This result is consistent with a developing consensus on the importance of income effects in determining physician behaviour.[4]. We discuss the policy implications of our results for using the compensation system to meet short-term demand shocks in health care. We also simulate the effects of the recent decision of the Québec government to increase all fees by 30%. Our simulations point to this increase leading to a reduction in services in the order of 2.6%-2.8%.

The rest of the paper is organized as follows. The next Section describes the institutional setting related to physician compensation in Québec and the sources of our data. Section 3 develops the general properties of the earnings function in the presence of multitasking. Section 4 develops our empirical model. Section 5 derives comparative statics, elasticities and the lower-bound for our empirical model. Section 6 presents the details of our sample used for estimation and the descriptive statistics. Section 7 presents our estimation results, while Section 8 presents the incentive effects and discusses policy simulations. The last section concludes.

2 Data and Institutional Details

The data used for this study contain information on specialist physicians practicing in Québec between 1996 and 2002. These data are derived from two sources: the Québec College of Physicians (CMQ) and the Health Insurance Organization of Québec (RAMQ). During this time, the Québec College of Physicians conducted an annual time-use survey of its members. This survey contained information on labour supply behaviour, captured by time spent at work, measured as the average (over the whole year) number of hours per week and time devoted to seeing patients. Our second source of data comes from the RAMQ administrative files used to pay physicians. These files give information on the medical fees paid to physicians for services completed, and the number of services performed by each physician. These data are available on a quarterly basis for each physician. The data from the Québec College of Physicians and from RAMQ were matched on the basis of an anonymous payroll number attributed to each physician.

[4]McGuire and Pauly (1991) provide an overview of early evidence.

In 2001, the government of Québec changed the prices paid to physicians for completed services, increasing the prices paid for services by up to 25%. Documented evidence suggests the main motivating factor behind the fee change was to narrow the income gap between Québec physicians and other Canadian physicians. In 1998-1999 Québec's physicians had the lowest average incomes in comparisons across Canadian provinces (Lemieux, Bergeron, Bégin, and Bélanger, 2003). Reducing this income gap was a principal bargaining point of the Québec Federation of Specialist Physicians (FMSQ) in the 2000-2001 bargaining agreement with the government (Hebert (2016)).

We restrict our sample to specialists who were present both before and after 2001, the year in which prices were changed and to those who were paid fee-for-service contracts.[5] We removed services for which prices increased between the years 1996 and 2000 as we suspect these price changes reflect technological changes and are hence endogenous. There were 85 such services. Each physician conducts a large number of medical services. To render our empirical model tractable, we aggregate services based on the composite commodity theorem. With this in mind, we dropped medical services which are not present over the whole sample period – 98 services are concerned. Finally, for empirical tractability, we restrict our estimation sample to physicians who supply two aggregate services.[6]

3 Multitasking and Earnings

To fix ideas and investigate the general properties of the earnings function, we consider the labour supply problem in the presence of multitasking. Individuals select hours h_s, and the manner in which those hours are allocated across tasks $j = 1, 2, \ldots, J$ to produce a quantity of services $A_1, A_2, \ldots, A_J$ via the production functions $f_j(h_j)$, where h_j is the hours devoted to service j.

A physician's preferences are represented by a well-behaved strictly quasi-concave utility function defined over income, M, and leisure $\ell = T - h_s$:

$$U(M, T - h_s), \text{ with } \quad \frac{\partial U}{\partial M} > 0, \quad \frac{\partial U}{\partial \ell} > 0,$$

[5]In 2000, the government of Québec introduced a mixed remuneration system, under which physicians were paid a reduced (relative to fee-for-service contracts) fee for services completed and a per-diem rewarding hours worked. Extending our model to include these physicians is an important extension that we leave for future work.

[6]A complete description of these data are available in Somé (2016).

where

$$
\begin{aligned}
M &= \sum_{j=1}^{J} \alpha_j A_j + y = \mathcal{E} + y, && (1) \\
A_j &= f_j(h_j) \quad j \in \{1, \ldots, J\}, && (2) \\
h_s &= \sum_{j=1}^{J} h_j, \quad h_j > 0 \quad \forall j, && (3) \\
f_j'(h_j) &> 0, \quad f_j''(h_j) < 0 \quad \forall j. && (4)
\end{aligned}
$$

The variable T is time endowment and h_s is total hours worked. The constraint (1) describes income as generated from providing services $j \in \{1, \ldots, J\}$, (we take J as fixed) at prices, α_j, and non-labour income y. The constraint (2) describes the production of service A_j from the physicians input of time, h_j. The constraints (3) specify that total hours of work is allocated across services and we only consider interior solutions. The production functions $f_j(h_j)$ are increasing and concave in h_j, $\forall j$ (see (4)) .

The physician's optimization problem can be analysed in two steps. First, conditional on h_s, the physician chooses $h_1, h_2, \ldots, h_J$ to maximise income M. Second, the physician chooses total hours worked, h_s, to maximize utility. In this section, we focus on the first problem to derive the conditional earnings function and its properties. We show that this function contains economic information over the reaction to incentives.

Maximizing income conditional on h_s yields $J - 1$ first-order conditions

$$
\alpha_j f_j'(h_j) - \alpha_J f_J'(h_s - \sum_{j=1}^{J-1} h_j) = 0, \quad j = 1, \ldots J - 1. \tag{5}
$$

The optimal solution, denoted $h_j^* \quad j \in \{1, 2, J - 1\}$, solves (5). The J^{th} term comes from the constraint (3). Replacing h_j^* into (5) gives the identity

$$
\alpha_j f_j'(h_j^*) - \alpha_J f_J'(h_J^*) \equiv 0, \quad j = 1, \ldots J - 1. \tag{6}
$$

Lemma 1 follows from differentiation of (6).[7]

Lemma 1: An increase in h_s increases hours allocated to all J services, $h_j^* \quad j \in \{1, 2, \ldots, J\}$, with

$$\frac{\partial h_j^*}{\partial h_s} = \frac{\prod_{k \neq j} \alpha_k f_k''(h_k^*)}{\sum_{j=1}^{J} \left[\prod_{k \neq j} \alpha_k f_k''(h_k^*)\right]} > 0.$$

The conditional (labour-market) earnings function is defined as

$$\mathcal{E}(\boldsymbol{\alpha}; h_s) = \sum_{j=1}^{J} \alpha_j f_j(h_j^*(\boldsymbol{\alpha}; h_s)),$$

where $\boldsymbol{\alpha}$ denotes the vector of prices, $(\alpha_1, \alpha_2, \ldots, \alpha_J)'$. It represents the maximum value of earnings from providing services that a physician can generate at prices $\boldsymbol{\alpha}$ for a given number of total hours worked, h_s.

The conditional earnings function can be evaluated at any h_s. Evaluating it at h_s^*, optimal hours worked, generates information over substitution effects as developed in property (4), below. Solving for the utility maximising $h_s^*(\boldsymbol{\alpha}, y)$ as a function of its underlying arguments (prices and non-labour income), and evaluating the earnings function at $h_s^*(\boldsymbol{\alpha}, y)$ gives the unconditional earnings function, which is a function of prices and non-labour income.

3.1 Properties of the Conditional Earnings Function

The conditional earnings function has the following properties:

1. The partial derivative of the conditional earnings function with respect to α_j is equal to the conditional supply of service j:

$$\frac{\partial \mathcal{E}}{\partial \alpha_j} = A_j^*(\boldsymbol{\alpha}, h_s).$$

2. The second partial derivative of the conditional earnings function with respect to α_j

[7] All proofs are in Appendix A1.

is equal to the slope of the conditional supply of service j:

$$\frac{\partial^2 \mathcal{E}}{\partial \alpha_j^2} = \frac{\partial A_j^*}{\partial \alpha_j}.$$

3. The conditional earnings function is convex in prices. Since physicians select h_j^* to maximize earnings, when prices adjust, they can increase their earnings by more than the simple price change by reoptimizing.

4. The second partial derivative of the earnings function with respect to α_j, evaluated at h_s^*, provides a lower bound to the own-price substitution effect of α_j on A_j.

   The complete reaction to a change in price α_j involves a substitution effect and an income effect. The income effect operates solely through h_s – changes in non-labour income do not affect the relative return to activity j. The partial derivative of the earnings function, holding h_s constant, is therefore independent of the income effect.

   The substitution effect also operates, in part, through h_s but this reinforces the direct effect on A_j that is measured in the earnings function. An increase in α_j increases the return to hours worked h_s. The full utility maximization problem implies the substitution effect on h_s is positive. Moreover, any increase in h_s is distributed across all services, $\frac{\partial h_j^c}{\partial h_s} > 0$ through Lemma 1. This is the Le Chatelier effect (Samuelson, 1947; Milgrom and Roberts, 1996), well-known within the analysis of input demands in the presence of fixed factors of production. Here the fixed factor is total hours worked, which are set at their optimal level.[8] Allowing hours to vary reinforces and magnifies the conditional price elasticity.

4 Empirical Model

To pursue our analysis of physician labour supply we derive the earnings function for a specific empirical model. The production of service j per time unit by physician i is given by the production function

$$A_{i,j} = b_j(x_{b,i}) h_{i,j}^{\delta} \epsilon_{i,j}, \quad \epsilon_{i,j} > 0,\ b_j > 0, \tag{7}$$

[8]This assumption can be generalized to the case where total hours worked are set at any given level by making use of the concept of virtual prices, that is, the hypothetical prices at which total hours worked correspond to their nonconstrained optimal level (Neary and Roberts, 1980).

where $\delta \in (0,1)$ captures the marginal return to time spent by the physician to provide service j. The value of δ is common across services. The production shock captures random elements that are specific to the physician (such as state of health), and affect his/her productivity; they are also common across services i.e. $\epsilon_{i,j} = \epsilon_i \quad \forall j$. The function $b_j(\mathbf{x}_{b,i})$ captures average production when one hour is supplied to the service. Allowing b_j to depend on observable individual characteristics, $\mathbf{x}_{b,i}$, captures elements, such as age (or experience) and gender, that can affect the productivity or speed at which a physician completes the service. For example, experienced physicians may perform services more quickly, through learning-by-doing effects. Alternatively, male physicians may work at different speeds than female physicians.[9]

Physician utility is defined over consumption (which is assumed to be equal to income, M) and leisure, denoted by ℓ. Physician preferences are CES:

$$U(M, \ell_o, \ell_p) = \left(\gamma(\mathbf{x}_\gamma)M^\rho + 0.5(1-\gamma(\mathbf{x}_\gamma))\ell_o^\rho + 0.5(1-\gamma(\mathbf{x}_\gamma))\ell_p^\rho\right)^{\frac{1}{\rho}}, \quad \rho < 1. \tag{8}$$

Here, on the job leisure is $\ell_o = h_T - h_s$, where h_T is total time worked and h_s, is time spent at work providing services to patients. Traditional leisure is $\ell_p = T - h_T$, where T is total time available. The relative weight a physician places on income and leisure is determined by γ. We restrict the relative weight of both types of leisure to be the same. Allowing γ to depend on characteristics, $\mathbf{x}_\gamma$, captures observable heterogeneity in preferences for leisure.

Income is given by

$$M = \sum_{j=1}^{J} \alpha_j A_j + y(\mathbf{x}_y), \tag{9}$$

where α_j represents the fee paid for service A_j and $y(\mathbf{x}_y)$ is non-labour income. This can depend on observable factors $\mathbf{x}_y$ which are related to asset returns.

We list the key assumptions that we impose to simplify the model's resolution and the empirical analysis.

A1. **Exogenous Service Mix:** A key assumption of the model is that the group of services that a particular physician provides is exogenously fixed. This is equivalent to assuming that each physician is trained to provide a fixed number of services. It allows us to search for interior solutions that examine how the supply of those

[9]For notational simplicity, save for a few exceptions, the index indicating physician i is suppressed until the section (8) on estimation.

services varies as prices change, ignoring services outside of this set. We ignore the decision to provide certain services and not others.[10]

A.2 **Common Shocks:** We assume common shocks across services for a given physician i: $\epsilon_{i,j} = \epsilon_i$ for $j = \{1, 2, \ldots, J\}$. We therefore interpret the production shock purely in terms of elements that affect the physician's productivity across all services. This can be due to elements affecting a physician's personal health. It can also reflect physician ability (or inherent productivity) which is constant across periods. Common shocks simplify the estimation as they drop out of the optimization problem for allocating time across services. Since the shock affects physician productivity independent of the service completed, only hours worked decisions are affected by the shock. An alternative would be to allow for demand shocks that vary across services.[11]

A.3 **Perfectly Elastic Demand for Services:** We rule out any demand shocks as determinants of the observed number of services provided. This is a strong assumption, but one that allows us to focus completely on physician behaviour.

A.4 **Full Information for physicians:** The physician observes ϵ, the price of each service, and the technology parameters, b_j, before choosing hours. Given the restriction to common shocks that represent physician health, it seems reasonable to assume that the physician observes the value of the shock before selecting hours of work.

A.5 **Stationary distribution of shocks:** The mean and variance of the distribution of shocks are constant over time. Our data takes the general form of a before-after natural experiment. Given prices are revised annually, any change in unobservable shocks is not separately identified from the effect of the change in prices. Given our interpretation of the shocks as a health shock, drawn across a relatively broad population of physicians, we feel comfortable in assuming that its general characteristics do not change over time.

A.6 **Exogenous Participation:** We assume that participation decisions are independent of potential physician productivity, ϵ. This allows us to ignore modelling the par-

[10] A more general model would examine the choice of which services to provide and allow for corner solutions – possibly due to demand shocks – to explain the fact that certain services are not provided.

[11] An agency interpretation would allow for service-specific shocks, perhaps due to the complexity of individual cases and asymmetric information. Physicians could then hide low effort levels behind low values of production shocks, generating agency costs. We discuss this, and other possible extensions, further in the conclusion of the paper.

ticipation decision in estimating the model, focussing solely on the choice of hours worked and services provided.

A.7 **Independence:** Productivity shocks are independent of personal characteristics, $\mathbf{x} = \left\{\mathbf{x}_b \cup \mathbf{x}_\gamma \cup \mathbf{x}_y\right\}$.

Physicians choose their total time at work, h_t, the amount of time devoted to providing services to patients, h_s, and the manner that those hours are allocated across different services, $h_j, j = \{1, 2, \ldots, J\}$. Substituting (9) and (7) into (8) and taking account of the definition of leisure and that $h_s = \sum_{j=1}^{J} h_j$, utility is

$$\begin{aligned} U &= \left\{\gamma(\mathbf{x}_\gamma)\left[\sum_{j=1}^{J-1} \alpha_j b_j(\mathbf{x}_b) h_j^{\delta}\epsilon + \alpha_J b_J(\mathbf{x}_b)\left(h_s - \sum_{j=1}^{J-1} h_j\right)^{\delta}\epsilon + y(\mathbf{x}_y)\right]^{\rho}\right. \\ &+ \left. 0.5(1-\gamma(\mathbf{x}_\gamma))\left(h_t - h_s\right)^{\rho} + 0.5(1-\gamma(\mathbf{x}_\gamma))\left(T - h_t\right)^{\rho}\right\}^{\frac{1}{\rho}}. \end{aligned} \tag{10}$$

For notational simplicity, we temporarily suppress dependence of γ, b and y on $\mathbf{x}$. These will be reintroduced in the empirical section.

4.1 Conditional Earnings

Conditional on clinical hours h_s, the optimal time spent providing service j is

$$h_j^*(\boldsymbol{\alpha}, h_s) = \frac{P_j}{\sum_{k=1}^{J} P_k} h_s, \tag{11}$$

where

$$P_j = (\alpha_j b_j)^{\frac{1}{1-\delta}}.$$

The optimal number of services of type j is

$$A_j(\boldsymbol{\alpha}, h_s) = b_j \left[\frac{P_j}{\sum_{k=1}^{J} P_k}\right]^{\delta} h_s^{\delta}\epsilon$$

Substituting into (7), multiplying by α_j and summing over all j gives the conditional (labour-market) earnings function

$$\mathcal{E}(\boldsymbol{\alpha}, \mathbf{x}_b, h_s, \epsilon) = \omega(\boldsymbol{\alpha}, \mathbf{x}_b) h_s^{\delta}\epsilon.$$

The conditional earnings function can be evaluated at any hours, h_s. Taking logarithms, and evaluating at optimal hours, h_s^*,

$$\ln \mathcal{E}(\boldsymbol{\alpha}, \mathbf{x}_b, h_s^*, \epsilon) = \ln \omega(\boldsymbol{\alpha}, \mathbf{x}_b) + \delta \ln h_s^* + \epsilon. \tag{12}$$

The term

$$\omega(\boldsymbol{\alpha}, \mathbf{x}_b) = \left(\sum_{j=1}^{J} P_j \right)^{1-\delta}$$

determines the marginal return to an hour worked when that hour is optimally allocated across services, given relative prices. The term w is not a wage in the traditional sense, but a wage index. Earnings are not linear in hours worked. Rather hours are an input to the production of services and exhibit decreasing marginal productivity. Notice as well, each hour worked is replicated and distributed across different services. This reflects the decreasing returns to the production of any given service and common shocks giving rise to interior solutions within the set of services that the physician provides – in the absence of increasing returns there are no gains to specialization among services.

The incentive (or substitution) effects, δ, are identified from two sources. First, exogenous variation in h_s and second (12), through w, by measuring the second-order effects of an increase in the price of service j on earnings when total hours are fixed. In the model, δ captures the sensitivity of the supply of hours to a particular service to changes in the price for that service. If $\delta = 0$, hours and services provided are fixed and outside of the control of physicians. If this were the case, an increase in the price of service j would induce a linear (accounting) increase in earnings with no change in physician behaviour. If physicians react to incentives, $\delta > 0$, then an increase in the price of service j will lead to a change in earnings that is convex in price since both the price of service j and hours devoted to service j increase.[12]

[12] This is analogous to the detection of substitution effects from cost functions in the theory of the firm, discussed in microeconomic textbooks; see, for example, Varian (1992).

4.2 Optimal Hours and the Hours Function

The complete utility maximisation problem solves for optimal hours, h_s^* as a function of prices, $\boldsymbol{\alpha}$ and alternative income y. The optimal time spent working, conditional on h_s, is

$$h_t(h_s) = \frac{T + h_s}{2}. \tag{13}$$

Substituting from (11) and (13) back into (10) gives indirect utility as a function of h_s:

$$V(h_s) = \left[\gamma(\omega h_s^{\delta}\epsilon + y)^{\rho} + 0.5(1-\gamma)2^{1-\rho}(T-h_s)^{\rho}\right]^{\frac{1}{\rho}},$$

The physician's optimal hours spent seeing patients, h_s^*, solves

$$\gamma\omega\delta h_s^{*\delta-1}\epsilon(\omega h_s^{*\delta}\epsilon + y)^{\rho-1} - 0.5(1-\gamma)2^{1-\rho}(T-h_s^*)^{\rho-1} = 0. \tag{14}$$

We note, h_s^* depends on: prices $\boldsymbol{\alpha}$ through ω, y, $\mathbf{x}_\gamma$ through $\gamma(\mathbf{x}_\gamma)$, $\mathbf{x}_b$, which enters ω, $\mathbf{x}_y$ through y and ϵ.[13] We write

$$h_s^* = h_s^*(\boldsymbol{\alpha}, y, \mathbf{x}, \epsilon). \tag{15}$$

The second-order condition is

$$\begin{aligned} V_{h_s h_s} &= \gamma\widetilde{\omega}\delta(\delta-1)h_s^{*\delta-2}(\widetilde{\omega}h_s^{*\delta} + y)^{\rho-1} + \gamma(\rho-1)(\widetilde{\omega}\delta h_s^{*\delta-1})^2(\widetilde{\omega}h_s^{*\delta} + y)^{\rho-2} \\ &\quad + 0.5(1-\gamma)2^{1-\rho}(\rho-1)(T-h_s^*)^{\rho-2} < 0 \end{aligned}$$

for $\delta \in (0,1)$ and $\rho < 1$.

While (14) does not give rise to an explicit functional form for h_s^*, it can be solved numerically. Evaluating (12) using (15) gives the conditional earnings function at h_s^* as solved by our model:[14]

$$\ln \mathcal{E}^*(\boldsymbol{\alpha}, \mathbf{x}, \epsilon) = \ln \omega(\boldsymbol{\alpha}, \mathbf{x}_b) + \delta \ln h_s^*(\boldsymbol{\alpha}, y, \mathbf{x}, \epsilon) + \epsilon. \tag{16}$$

Estimation of (12) and (16) is complicated by three elements. First, they are non-linear functions of δ since $\omega = \left(\sum_j P_j\right)^{1-\delta}$ and $P_j = (\alpha_j b_j)^{\frac{1}{1-\delta}}$. Second, h_s is potentially cor-

[13] It will also depend on tax rates (see Section 7.3.1, below).

[14] This is the unconditional earnings function since optimal hours are expressed as a function of prices $\boldsymbol{\alpha}$ and non-labour income, y. The h^* in (16) solves (14), rather than being pre-allocated at observed hours.

related with ϵ, since physician's choose hours worked, implying $h_s^*(\boldsymbol{\alpha}, y, \mathbf{x}, \epsilon)$ maximises (8). Finally, individual characteristics can affect both preferences for hours worked and productivity, rendering identification of these different paths problematic. Identification of these separate channels may be through the non-linearities inherent in earnings, exclusion restrictions, and cross-equation restrictions. We return to these points below in discussing the empirical model.

5 Comparative statics and Lower Bound

A physician's reaction to incentives can be analyzed using comparative-static techniques. Price changes imply income and substitution effects for the supply of services. Within the context of our model, these effects operate through multiple channels since physicians choose the number of hours to work and the manner in which those hours are allocated across services. We present the relevant equations in the text, suppressing dependence on $\mathbf{x}$. Complete derivations are given in Appendix A.2.

We make the following definitions:

1. Let
$$\tilde{V}(h_s, \boldsymbol{\alpha}, y, \epsilon) = \gamma\omega\delta h_s^{\delta-1}\epsilon\left(\omega h_s^{\delta}\epsilon + y\right)^{\rho-1} - 0.5(1-\gamma)2^{1-\rho}(T-h_s)^{\rho-1}.$$

2. h_s^* solves
$$\tilde{V}(h_s^*, \boldsymbol{\alpha}, y, \epsilon) = 0;$$

3. The second-order condition
$$\tilde{V}_{h_s} \equiv \frac{\partial \tilde{V}(h_s^*, \boldsymbol{\alpha}, y, \epsilon)}{\partial h_s} < 0.$$

5.1 Own-price elasticities

The own-price elasticity of hours devoted to service j is given by

$$\eta_{h_j,\alpha_j} = \underbrace{\left[\sum_{k\neq j}\frac{P_k}{\sum_k P_k}\frac{1}{(1-\delta)} - \gamma\frac{\alpha_j A_j \delta M^{\rho-1}}{h_s^2\tilde{V}_{h_s}}\right]}_{\text{Substitution effect}} + \underbrace{\frac{\alpha_j A_j}{y}\eta_{h_s,y}}_{\text{Income effect}} > 0. \tag{17}$$

From (11),

$$\left.\frac{\alpha_j}{h_j}\frac{\partial h_j}{d\alpha_j}\right|_{h_s} = \frac{1}{(1-\delta)}\sum_{k\neq j}\frac{P_k}{\sum_k P_k} > 0,$$

the first term of the substitution effect. This captures physicians reallocating a fixed number of hours towards those services that have higher relative prices in order to maximize earnings. The effect of this reallocation is to increase the wage index ω. The second term of (17) is positive, since $\tilde{V}_{h_s}$ is negative from the second-order condition. The increase in ω leads physicians to work more hours, which are then allocated across all services. The fact that the total substitution effect is positive allows us to state its lower bound as

$$\mathcal{L}_j = \frac{1}{(1-\delta)}\frac{\sum_{k\neq j} P_k}{\sum_k P_k}, \tag{18}$$

which is independent of ρ.

Income effects are also present. The optimal allocation of time implies that the price-weighted marginal utility of each service is equated across services. Income effects therefore operate only through total hours worked – they do not affect the relative supply of different services.

5.2 Cross-price elasticities

The cross-price elasticity is

$$\eta_{h_k,\alpha_j} = \underbrace{-\left[\frac{1}{1-\delta}\frac{P_j}{\sum_k P_k} + \gamma\frac{\delta\alpha_k A_k M^{\rho-1}}{h_s^2\tilde{V}_{h_s}}\right]}_{\text{Substitution effect}} + \underbrace{\frac{\alpha_j A_j}{y}\eta_{h_k,y}}_{\text{Income effect}}.$$

Again the substitution effect has two components, but unlike the own-price effect, these components operate in different directions. If the price of service j increases, *ceteris paribus*, the change in relative prices causes physicians to substitute away from services whose relative price has decreased. But the resulting increase in ω leads to an increase in hours worked which is distributed across all services, including those with lower prices. The overall cross-price substitution effect is ambiguous. Again, the second term of the substitution effect operates through hours worked and hence depends on ρ which is not identified from the conditional earnings equation.

Notice, as well, the substitution effects are not symmetric, even conditional on h_s. This is due to the nonlinearities in the production of services that enter the budget constraint (*e.g.*, Kalman and Intriligator, 1973; Blomquist, 1989). Changes in prices cause first and second-order effects that determine the elasticity of substitution.

5.3 Wage index elasticities

We can also consider the impact of a proportional increase in all prices on physician behaviour. From (14), this can be approximated by the effect of the wage index on clinical hours worked, as given by:

$$\eta_{h_s,\omega} = \frac{\omega}{h_s}\Bigg[- \underbrace{\frac{\gamma\delta h_s^{\delta-1}\epsilon M^{\rho-1}}{\tilde{V}_{h_s}}}_{\text{Substitution Effect}} + \underbrace{\frac{(1-\rho)\gamma\delta\omega h_s^{2\delta-1}\epsilon^2 M^{\rho-2}}{\tilde{V}_{h_s}}}_{\text{Income Effect}}\Bigg], \tag{19}$$

where

$$\begin{aligned}\tilde{V}_{h_s} =& w\epsilon\gamma\delta(\delta-1)h_s^{\delta-2}M^{\rho-1} + \gamma(\rho-1)(w\delta h_s^{\delta-1}\epsilon)^2 M^{\rho-2} \\ &+0.5(1-\gamma)2^{1-\rho}(\rho-1)(T-h_s)^{\rho-2} < 0,\end{aligned}$$

and

$$M = \omega h_s^{\delta}\epsilon + y.$$

The substitution effect is positive and reflects the compensated effect of a change in the wage index on total clinical hours, while the income effect is negative as (pure and on-the-job) leisure is a normal good.

Using (11), the effect of an increase in ω on hours devoted to a given service can are the same as the effect on total hours

$$\begin{aligned}\frac{\partial h_j}{\partial\omega}\frac{\omega}{h_j} =& \frac{P_j}{\sum_{k=1}^{K} P_k}\frac{\partial h_s}{\partial\omega}\frac{\omega}{h_j} = \frac{P_j}{\sum_{k=k}^{K} P_k}\frac{\partial h_s}{\partial\omega}\frac{\omega}{\frac{P_j}{\sum_{k=1}^{K} P_k}h_s} \\ =& \frac{\partial h_s}{\partial\omega}\frac{\omega}{h_s}.\end{aligned} \tag{20}$$

Similarly,

$$\frac{\partial A_j}{\partial \omega}\frac{\omega}{A_j} = \delta \frac{\partial h_s}{\partial \omega}\frac{\omega}{h_s}.$$

6 Data description and sample construction

The data used for this study contain information on specialist physicians practicing in Québec between 1996 and 2002. Physicians provide a large number of different services, at a variety of prices. To render our empirical problem tractable, we aggregated services based on the percentage change of their price using the composite commodity theorem.[15] This gave 6 aggregate services, depending on whether the price increased by 0%, 5%, 10%, 15%, 20% or 25%. These aggregate services are denoted 1, 2, 3, 4, 5 and 6, respectively. Our data include physicians from different specialties. The specialties include: cardiac and vascular surgery, nephrology, radio-oncology, anesthesiology, endocrinology, gastroenterology, cardiology, pediatrics, internal medicine, neurology, general surgery, dermatology, gynecology and obstetrics, orthopedics surgery and otorhinolaryngology. We grouped physicians according to the number of aggregate services that they supplied. This gave three separate groups of physicians: those who provided 2 services, those who provided 3 services and those who provided 4 services.[16] For empirical tractability, we restrict our sample here to physicians who provided 2 services. Note, these physicians do not all provide the same services. In fact, there are two groups of physicians who provide 2 services: those who provide services 1 and 2, and those who provide services 1 and 3. There are 242 physicians in our data set. These physicians were present both before and after 2001, the year in which prices were changed. A summary of these data are given in Table 1.

6.1 Descriptive Statistics

Our data contain physicians who provided two aggregate services. This includes physicians who provided services 1 and 2 and physicians who provided services 1 and 3. The raw data are shown in Table 1. It shows summary statistics on the main variables of in-

[15]See Appendix A3 for the application of the composite commodity theorem for our particular case. The details of the aggregation process and variable construction are given in are given in Appendices A4 and A5. Further details are available in Somé (2016).

[16]A description of the complete data set, which includes physicians who provided 2, 3 and 4 services, a description of its construction is available in Shearer, Somé, and Fortin (2018) and Somé (2016).

terest for our model (hours worked, prices and earnings). We provide statistics for each period of the sample data, separated by the number of services provided. Hours worked are reported on a weekly basis; earnings are annual and in thousands of dollars.

The prices of all goods are the same before the price increases in the year 2000. This reflects the fact that these are the prices of the aggregate services (measured by the revenue generated from those services). Under the aggregation theorem, their prices are equal to the rate of increase of the prices within the relevant group of services. As all prices were stable before the year 2000, their nominal prices are equal to one for those years. The variation across years reflects changes in the rate of inflation. The price increases for services two and three are evident in years 2001-2002, raising average earnings in the process. The real price of service two increases by 3.7% in 2001 relative to 2000 prices, and by 8.5% in 2002. The real price of service three increases by 8.8% in year 2001 relative to 2000, and by 19% in 2001. raising average earnings in the process. Subsequent to the fee changes, physician incomes increased by 21% . There is a slight decrease in clinical hours worked between the years 2000 and 2002, in the order of 3.5% .

The lower part of Table 1 presents average clinical hours and earnings for different characteristics of physicians. There is little difference between male and female physicians in terms of annual earnings, yet males spend less time seeing patients. This suggests that males earn more per hour worked. Physicians whose native language is French work more and earn more than do those whose native language is English. Middle-aged physicians, those whose age is between 40 and 60, display no tendency to work more than do those aged less than 40, but they earn considerably more. This suggests they are more efficient in their diagnoses and practices. Physicians who are older than 60 spend less time seeing patients and earn less than middle-aged physicians, but more than their younger counterparts. We will use these facts in specifying our econometric model.

Table 1: Descriptive statistics : Prices, Earnings and Hours

Physicians providing 2 services						
Year	Obs	Prices			Annual Earnings	Clinical Hours
		Service 1	Service 2	Service 3	(000's)	(weekly)
1996	197	1.104	1.104	1.104	96.27	42.02
1997	201	1.086	1.086	1.086	93.68	46.47
1998	188	1.057	1.057	1.057	96.55	43.12
1999	192	1.035	1.035	1.035	95.75	44.74
2000	189	1.017	1.017	1.017	93.32	44.98
2001	168	1.005	1.055	1.106	106.94	44.15
2002	165	1.000	1.103	1.210	112.94	43.39
Male	1188	–	–	–	99.00	43.84
Female	201	–	–	–	97.90	47.37
French	1082	–	–	–	99.50	44.90
English	218	–	–	–	95.95	40.44
Age $<$ 40	482	–	–	–	87.33	44.45
40 $<$ Age $<$ 60	602	–	–	–	108.67	44.74
60 $<$ Age	216	–	–	–	97.49	41.81

7 Estimation

Our empirical model consists of three basic equations, explaining for physician i in period t: total earnings, (21a), earnings on services j, (21b), and hours worked, (21c):

$$\ln \mathcal{E}_{i,t} = \ln \omega_t + \delta \ln h_{i,t} + \epsilon_{i,t}, \tag{21a}$$

$$\ln \mathcal{E}_{i,j,t} = \ln(P_{j,t}) - \delta \ln(\sum_j P_{j,t}) + \delta \ln h_{i,t} + \epsilon_{i,t} \quad j = 2, \dots J, \tag{21b}$$

$$0 = (1-\gamma(x))\omega_t \delta h_{i,t}^{\delta-1} \epsilon_{i,t} (\omega_t h_{i,t}^{\delta} \epsilon + y(\mathbf{x}_y))^{\rho-1} - 0.5\gamma(x) 2^{1-\rho} (T - h_{i,t})^{\rho-1} \tag{21c}$$

where $\omega_t = \left(\sum_{j=1}^{J} P_{j,t}\right)^{(1-\delta)}$ and $P_{j,t} = \left(b_j \alpha_{j,t}\right)^{\frac{1}{1-\delta}}$.

7.1 Specification

Recall, we allow physician choices to be affected by two sources of exogenous variation: variables on the personal characteristics of physicians and variables on non-labour income. We use $\mathbf{x}_\gamma$, to denote personal characteristics which affect preferences for leisure, $\mathbf{x}_b$ to denote characteristics which affect productivity and $\mathbf{x}_Y$ to denote variables which affect non-labour income.

The function $b_j(\mathbf{x}_b)$ includes a service-specific constant term $b_{0,j}$ as well as terms capturing the effect of personal characteristics (age, and gender) which plausibly affect physician productivity. Age is included to capture the effect of experience on the ability to perform diagnoses and perform services. We also experiment with including a trend to capture changes in productivity through time due, for example, to technological change.

We ensure $b(\mathbf{x})$ is positive using the exponential function:

$$b(\mathbf{x}) = exp(b_{0,j} + \mathbf{x}_b' \boldsymbol{b}),$$

where $\mathbf{x}_b' \boldsymbol{b}$ is independent of j.

The function $\gamma(\mathbf{x}_\gamma)$ is specified as logistic:

$$\gamma(\mathbf{x}_\gamma) = \frac{exp(\mathbf{x}_\gamma' \boldsymbol{\gamma})}{1 + exp(\mathbf{x}_\gamma' \boldsymbol{\gamma})},$$

where $\mathbf{x}_\gamma$ contains age, gender and native language.[17] We exclude the middle-age dummy

[17] In practice we restricted $\gamma(\mathbf{x}_\gamma) \in [0, 2/3]$, forcing its value to equal $1/3$ when $\mathbf{x}_\gamma' \gamma = 0$. The latter captures

from $\mathbf{x}_\gamma$, forcing their preferences for leisure to be identical to young physicians, since these two groups display no differences in hours worked in Table 1.

The function $y(\mathbf{x}_y)$ is specified to capture non-labour income in year t that can affect hours choices through income effects. We specify $y(\mathbf{x}_y)$ to be a linear function of the stock market return during the year. Returns are plausibly correlated with asset income, which has often been used to capture non-labour income (*e.g.*, Heckman, 1974).[18]

7.2 Limited Information Estimation

Limited-information methods estimate the conditional earnings equations (21a) and (21b). Hours, h_s, are taken as given and the model is used to explain how those hours are allocated across different services, given changes in prices. This has certain advantages. Principally, estimation does not require solving for optimal hours through (21c) and is therefore easier. It also relies on fewer restrictions from the model and therefore may provide more robust estimates. Yet there are also costs. Ignoring variation in hours worked precludes the identification of ρ (which only enters (21c)). Since income effects and part of the substitution effect depend on changes in h_s, these full effects are not identified from limited-information estimation. The limited-information approach does allow us to estimate δ and to construct a lower bound to the own-price substitution effect, based on (18). It therefore provides an answer the question of whether or not physicians respond to monetary incentives.

To estimate the parameters, we treat the earnings of each service as a separate equation and estimate a multivariate non-linear regression model:

$$\begin{aligned}
\ln \mathcal{E}_{1,12_{i,t}} &= \ln(P_{1,12_t}) - \delta \ln(P_{1,12_t} + P_{2_t}) + \delta \ln h_{i,t} + \epsilon_{i,t}, \\
\ln \mathcal{E}_{1,13_{i,t}} &= \ln(P_{1,13_t}) - \delta \ln(P_{1,13_t} + P_{3_t}) + \delta \ln h_{i,t} + \epsilon_{i,t}, \\
\ln \mathcal{E}_{2_{i,t}} &= \ln(P_{2_t}) - \delta \ln(P_{1,12_t} + P_{2_t}) + \delta \ln h_{i,t} + \epsilon_{i,t}, \\
\ln \mathcal{E}_{3_{i,t}} &= \ln(P_{3_t}) - \delta \ln(P_{1,13} + P_{3_t}) + \delta \ln h_{i,t} + \epsilon_{i,t},
\end{aligned} \tag{22}$$

the equal shares case.

[18]We have no information on spousal income. Nor was there a major tax reform during the time period under study, another possible instrument for hours worked; see, for example, Blundell, Duncan, and Meghir (1998), and Showalter and Thurston (1997). Also, to be fully consistent with an inter-temporally separable life-cycle model involving a two stage budgeting process, the non-labour income should be net of savings (Blundell and Walker, 1986). Unfortunately, we have no information on physicians' savings in our data set. Admittedly, this may be a source of measurement errors in our estimates.

where $P_{j_t} = \left(b_j \alpha_{j_t}\right)^{\frac{1}{1-\delta}}$. We allow the parameter b_1 to differ depending on whether the physician provides services 1 and 2 or services 1 and 3. $\mathbf{x}_b$ contains a dummy variable for male physicians, a dummy variable for middle aged physicians, a dummy variable for old physicians and a trend term. We define a set of instruments for equation ℓ as $\mathbf{Z}_\ell$.[19] We include in $\mathbf{Z}_\ell$ the relevant service prices for equation ℓ, $Dmale, Dmid, Dold, DFrench$, the annual market return and its interaction with *Dmid* and *Dold*. We note that the earnings equations do not include a constant term, hence no constant is included in the instrument set.[20]

Equations (22) can then be estimated by minimizing

$$\left(\boldsymbol{\mathcal{E}}^{o} - \boldsymbol{\mathcal{E}}(\boldsymbol{\beta})\right)' \mathbf{W} \left(\left(\boldsymbol{\mathcal{E}}^{o} - \boldsymbol{\mathcal{E}}(\boldsymbol{\beta})\right)\right),$$

where $\boldsymbol{\mathcal{E}}^{o}$ represents the stacked vector of observed earnings, $\boldsymbol{\mathcal{E}}(\boldsymbol{\beta})$ represents the stacked vector of predicted earnings, from the model, and $\mathbf{W}$ is a weighting matrix. In the case of non-linear least squares, $\mathbf{W}$ is set to the identity matrix. For non-linear instrumental variables estimation, (Amemiya, 1985), $\mathbf{W}$ is a block diagonal matrix with block ℓ given by $\mathbf{P}_{\mathbf{Z}_\ell} = \mathbf{Z}_\ell \left(\mathbf{Z}_\ell' \mathbf{Z}_\ell\right)^{-\mathbf{1}} \mathbf{Z}_\ell$.

7.2.1 First-Stage Results

We present first-stage regressions in Table 2. These are regressions of the model's endogenous variable ($\ln h$) on the exogenous prices in equation ℓ and the instruments. We present the results from two separate specifications of the model: with and without a trend term among the instruments. Results are presented separately for physicians providing services one and two and physicians providing services one and three. The price variables are generally statistically significant, though not always positive. The market return variable is negative and statistically significant in the versions without a trend. It loses its significance when the trend is included. The reported F statistics are for the restriction that all coefficients apart from the prices are equal to zero. These are significant in all cases except for the specification without trend on physicians providing services one and three. This shows that the instruments are correlated with the endogenous variable.

Often tests for weak instruments concentrate on the F statistics for the subset of instru-

[19]Valid instruments affect h_s, but are independent of ϵ. From the model, non-labour market income is correlated with hours through the income effect. Yet, conditional on hours, the earnings function is independent of these effects.

[20]Notice that even if b_j contains a constant, its derivative multiplies the price α_j.

ments that are excluded from the equation of interest – in our case the earnings equation. These tests will depend on the specification of the model. Below we estimate two versions of the earnings equation. The first version allows $b(\mathbf{x})$ to depend on observable characteristics: *male*, *Dmid*, and *Dold*. The excluded instruments for this version are: *DFrench*, *Market Return*, *Market* $\times$ *Dmid* and *Market* $\times$ *Dold*. The F-statistics is 6.88 with 4 and 184 degrees of freedom for the case of physicians providing services one and two; the p-value is essentially zero. It is 2.17 with 4 and 56 degrees of freedom for the case of physicians providing services one and three; the p-value is 0.084. The second version adds a trend term to $b(\mathbf{x})$. Here, the excluded instruments are the same as for the first version. The F-statistics for the case of physicians providing services one and two is 0.92; the p-value is 0.456. For physicians providing services one and three the F-statistic is 1.31 with a p-value of 0.276.

Overall, the evidence for weak instruments is mixed and depends on the version of the model estimated. The version of the model without a trend in $b(\mathbf{x})$ shows strong correlation between the excluded instruments and hours worked, particularly for physicians providing services one and two. The version with a trend, shows weak correlation.[21]

7.2.2 Conditional Earnings Equation Estimates

Table 3 provides results for two different specifications of the earnings function. Specification (1) allows $b(\mathbf{x})$ to depend on gender and age, but no trend. Specification (2) adds a trend term to service 3.[22] For each specification, we provide least-squares estimates (LS) and generalised method of moments estimates (GMM).

The least-squares estimate of δ is close to 0.5 in each specification and are statistically significant. The GMM estimate of δ is close to 0.6 in specification (1), but falls to 0.544 when the trend term is included in specification (2). They are also statistically significant. Of the individual characteristics, only b_{Dmid} is consistently significant. It is also positive, reflecting that physicians in the middle of their careers are more productive and have higher earnings for a given level of hours. This is consistent with the summary statistics presented in Table 1. The trend term is insignificant in both least-squares and GMM estimation.

We test the overidentifying restrictions in the model using a Sargan test, based the value of the objective function. These are given in the last row of the Table. These restric-

[21] We remind the reader that it is actually the correlation between instruments and the derivative of the conditional mean function that is important for non-linear instrumental variables.

[22] Allowing the trend to affect services 1 and 2 did not change the results.

Table 2: First-Stage Estimates
(Dependent variable: $\ln h$)

	Services 1 and 2		Services 1 and 3	
Parameter	I	II	III	IV
α_1	2.266*** (0.325)	4.920*** (0.325)	3.043*** (0.252)	4.263*** (0.350)
α_2	1.208*** (0.311)	-1.690 (0.311)		
α_3			0.484* (0.265)	-1.003*** (0.365)
Male	0.026 (0.109)	-0.005 (0.107)	-0.209 (0.159)	-0.283 (0.149)
DFrench	0.129* (0.077)	0.102 (0.078)	0.153 (0.121)	0.136 (0.121)
Dmid	0.005 (0.047)	-0.014 (0.048)	0.096 (0.084)	0.072 (0.084)
Dold	-0.088 (0.088)	-0.117 (0.088)	0.082 (0.124)	0.060 (0.126)
MarketReturn	-0.295*** (0.086)	0.046 (0.086)	-0.539*** (0.211)	-0.217 (0.174)
$Market \times Dmid$	-0.049 (0.108)	-0.022 (0.107)	0.516* (0.285)	0.504 (0.275)
$Market \times Dold$	-0.158 (0.170)	-0.217 (0.168)	0.409 (0.374)	0.466 (0.376)
Trend		0.090*** (0.007)		0.093*** (0.015)
F-Statistic[1]	4.70***	30.63***	1.81	7.79***
Degrees of Freedom	(7,184)	(8,184)	(7,56)	(8,56)
P-Value	0.000	0.000	0.104	0.000
Observations	983	983	317	317

1. The reported F-Statistics correspond to the restriction that all coefficients, apart from the αs, are equal to zero.
2. Estimated standard errors in parentheses.
3. Standard errors are robust and clustered over individuals.
4. ***, ** and * indicate statistical significance at levels: 0.01, 0.05 and 0.1, respectively.

tions are not rejected – the P-values are well above 0.05 in all cases. We also perform a Durbin-Wu-Hausman test on the difference between the least-squares estimates and the GMM estimates.[23] Rejections of the null suggest that the least-squares estimates are not consistent and instrumental variables should be used. The evidence is consistent with endogeneity causing inconsistency in our case. The p-values are 0.042, and 0.050 for the different specifications, suggesting inconsistency of least-squares estimates.

The limited-information estimates allow for the estimation of the lower bound to the own-price substitution effect of a price change on service j, from (18). These are presented in Table 4. The estimated elasticities are all positive, although they are considerably larger for service 1 than for services 2 and 3.

The full substitution effects and the income effects of a price change, given in (17), as well as the elasticities of hours worked, depend on the parameter ρ which does not enter the conditional earnings function. These effects are therefore not identified from limited information estimation. We now turn to full-information methods to estimate all parameters and identify these effects.

7.3 Full-Information Estimation

The full-information model adds the hours equation (21c) and evaluates the conditional earnings equations (21a) and (21b) at optimal hours given by (15). Explaining observed variation in hours worked in each period identifies the parameter ρ. The fact that $b(\mathbf{x})$ affects hours choices through P_j generates cross-equation restrictions that can help identify its parameters as well. We denote the parameter vector as

$$\mathbf{\Gamma} = (\rho, \delta, b_{1,12}, b_{1,13}, b_2, b_3, \boldsymbol{b}_{\mathbf{x}_b}, \boldsymbol{b}_{\mathbf{x}_\gamma}, \sigma^2_\epsilon). \tag{23}$$

We estimate the model using simulated method of moments (SMM), generating ϵ from a lognormal distribution. Let $h^*_{s,i,t,r}(\Gamma, \boldsymbol{\alpha}_t, \mathbf{X}_i, \epsilon_r)$, denote the hours worked that solves (21c), given prices, $\boldsymbol{\alpha}_t$, observed characteristics $\mathbf{X}_i$ and a particular draw of ϵ_r. Similarly, let $\mathcal{E}^*_{i,j,t,r}(\Gamma, \boldsymbol{\alpha}_t, \mathbf{X}_i, \epsilon_r)$ denote the resulting earnings on service j and $\mathcal{E}^*_{i,t,r}(\Gamma, \boldsymbol{\alpha}_t, X_i, \epsilon_r)$, total

[23] The form of the test is given in Davidson and MacKinnon (1993). It is calculated as the F-test for $c = 0$ in the artificial regression

$$\mathbf{y} - \mathbf{x}(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{X}}\mathbf{b} + \mathbf{M_z}\hat{\mathbf{X}}^*\mathbf{c} + \textit{residuals},$$

where $\hat{\mathbf{X}} = \partial\mathbf{x}(\hat{\boldsymbol{\beta}})/\partial\boldsymbol{\beta}$, evaluated at the NLS estimates $\hat{\boldsymbol{\beta}}$, and $\hat{\mathbf{X}}^*$ are the columns of $\hat{\mathbf{X}}$ that are correlated with the error term: $\partial\mathbf{x}(\hat{\boldsymbol{\beta}})/\partial\delta$ in our case. $\mathbf{M_z}\hat{\mathbf{X}}^*$ are the residuals from regressing the columns of $\hat{\mathbf{X}}^*$ on the set of instruments.

Table 3: Limited-Information Estimates
(Dependent variable: $\ln \mathcal{E}$)

	Estimates			
Specification	(1)		(2)	
Parameter	LS	GMM	LS	GMM
δ	0.505*** (0.075)	0.626*** (0.122)	0.508*** (0.074)	0.544** (0.159)
$b_{1,12}$	0.784*** (0.127)	0.799*** (0.184)	0.781*** (0.126)	0.866*** (0.205)
$b_{1,13}$	-0.063 (0.138)	-0.004 (0.206)	-0.065 (0.137)	0.089 (0.236)
b_2	2.222*** (0.304)	1.856** (0.487)	2.211*** (0.301)	2.157*** (0.615)
b_3	1.268*** (0.305)	0.980* (0.485)	1.142*** (0.299)	1.277** (0.612)
b_{Male}	0.150 (0.110)	-0.010 (0.143)	0.150 (0.110)	-0.015 (0.143)
b_{Dmid}	0.177*** (0.068)	0.221*** (0.087)	0.177*** (0.068)	0.202** (0.091)
b_{Dold}	0.200*** (0.091)	0.060 (0.130)	-0.200*** (0.092)	0.002 (0.143)
b_{Trend_3}			0.030 (0.033)	0.006 (0.008)
P-Value Overidentification		0.697		0.622
P-Value DWH Test	0.042		0.050	
Observations	1,300	1,300	1,300	1,300

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Own-Price Elasticity Lower Bounds

	Estimates		
Specification	(1)	(2)	(3)
Parameter	GMM	GMM	GMM
$\eta_{1,1_{12}}$	1.673 [1.491, 1.811]	1.581 [1.360, 1.698]	1.235 [1.020, 1.357]
$\eta_{2,2}$	0.130 [0.032, 0.387]	0.090 [0.028, 0.421]	0.071 [0.021, 0.363]
$\eta_{1,1_{13}}$	1.661 [1.484, 1.809]	1.566 [1.351, 1.694]	1.221 [1.013, 1.349]
$\eta_{3,3}$	0.142 [0.035, 0.395]	0.105 [0.034, 0.415]	0.085 [0.025, 0.362]

earnings. Given N_r repeated draws of $\epsilon_{i,r}$ for observation i, we calculate the simulated expectations, conditional on prices $\boldsymbol{\alpha}_t$, and observable $\mathbf{X}_i$, as

$$m^*_{\ln h_{i,t}}(\boldsymbol{\Gamma}, \boldsymbol{\alpha}_t, \mathbf{X}_i) = \frac{1}{N_r}\sum_{r=1}^{N_r} \ln h^*_{i,t,r}(\Gamma, \boldsymbol{\alpha}_t, \mathbf{X}_i, \epsilon_r),$$

$$m^*_{\ln \mathcal{E}_{i,t}}(\boldsymbol{\Gamma}, \boldsymbol{\alpha}_t, \mathbf{X}_i) = \frac{1}{N_r}\sum_{r=1}^{N_r} \ln \mathcal{E}^*_{i,t,r}(\boldsymbol{\Gamma}, \boldsymbol{\alpha}_t, \mathbf{X}_i, \epsilon_r)$$

$$m^*_{ln \mathcal{E}_{i,j,t}}(\boldsymbol{\Gamma}, \boldsymbol{\alpha}_t, \mathbf{X}_i) = \frac{1}{N_r}\sum_{r=1}^{N_r} \ln \mathcal{E}^*_{i,j,t,r}(\Gamma, \boldsymbol{\alpha}_t, \mathbf{X}_i, \epsilon_r),$$

for services $j = 2, 3$.

Recall that physicians provide either services one and two or services one and three. We let $\mathscr{S} \in \{2, 3\}$ denote these two sets of physicians. We match each simulated expectation to its observed counterpart, giving simulated residuals for observation i, t

$$m^*_{\mathscr{S}}(\Gamma)_{i,t} = \begin{bmatrix} \ln h_{t,i} & - & m^*_{\ln h_{t,i}}(\mathbf{\Gamma}, \boldsymbol{\alpha}_i, \mathbf{X}_i) \\ \ln \mathcal{E}_{t,i} & - & m^*_{\ln \mathcal{E}_{t,i}}(\mathbf{\Gamma}, \boldsymbol{\alpha}_i, \mathbf{X}_i) \\ \ln \mathcal{E}_{t,i,j} & - & m_{\ln \mathcal{E}^*_{t,i,j}(\mathbf{\Gamma}, \boldsymbol{\alpha}_i, \mathbf{X}_i)} \\ & & \end{bmatrix} \quad j \in \{2, 3\}.$$

Let $N^{\mathscr{S}}$ denote the number of observations on physicians of type $\mathscr{S}$ and let $m^*_{\mathscr{S}}(\mathbf{\Gamma})$ denote the $3N^{\mathscr{S}} \times 1$ vector which stacks the $N^{\mathscr{S}} \times 1$ vectors: $\ln h_t - m^*_{\ln h_t}(\mathbf{\Gamma}, \boldsymbol{\alpha}, \mathbf{X})$, $\ln \mathcal{E}_t - m^*_{\ln \mathcal{E}_t}(\mathbf{\Gamma}, \boldsymbol{\alpha}, \mathbf{X})$ and $\ln \mathcal{E}_{t,j} - m^*_{ln\mathcal{E}_{t,j}}(ln\mathcal{E}_{t,j}; \mathbf{\Gamma}, \boldsymbol{\alpha}, \mathbf{X})$. The $N^{\mathscr{S}} \times k$ instrument matrix, $\mathbf{Z}_{\mathscr{S}}$ is the same for all equations.

We form the objective function for each set of physicians $\mathscr{S} \in \{1, 2\}$

$$m^*_{\mathscr{S}}(\mathbf{\Gamma})' \mathbf{W}_{\mathscr{S}} m^*_{\mathscr{S}}(\mathbf{\Gamma}). \tag{24}$$

$\mathbf{W}_{\mathscr{S}}$ is a block diagonal matrix with block k given by $\mathbf{P}_{\mathbf{Z}_{\mathscr{S}}} = \mathbf{Z}_{\mathscr{S}} \left(\mathbf{Z}'_{\mathscr{S}} \mathbf{Z}_{\mathscr{S}}\right)^{-1} \mathbf{Z}'_{\mathscr{S}}$, the projection matrix associated with the instruments of the set $\mathscr{S}$ physicians. Since the error term is common to all equations of the model, the instrument set is the same across equations, within each set of physicians. It includes the relevant prices as well as the elements of $\mathbf{X} = \{\mathbf{x}_b \cup \mathbf{x}_\gamma \cup \mathbf{x}_y\}$.[24] To identify the variance of ϵ, which is assumed constant across time and sets of physicians, we add the unconditional second moment of total earnings. Here we match the simulated second moment of log of total earnings

$$m^*_{(\ln \mathcal{E})^2} = \frac{1}{N_r} \frac{1}{N^{\mathscr{S}}} \sum_{i=1}^{N^{\mathscr{S}}} \sum_{r=1}^{N_r} (\ln \mathcal{E}^*_{i,r}(\mathbf{\Gamma}, \epsilon_r))^2$$

to the observed second moment of the log of earnings.

$$\frac{1}{N^{\mathscr{S}}} \sum_{i=1}^{N^{\mathscr{S}}} (\ln \mathcal{E}_{i,t})^2$$

We weight this moment in the objective function by the inverse of its standard deviation.

Our estimation approach is similar in spirit to a pooled statistical model, in which ignored correlation across periods, due to random effects for example, causes an efficiency loss. In our case, random effects enter ϵ. The correlation between those random effects and

[24] Estimation of (24) is equivalent to the non-linear instrumental variables estimator of Amemiya (1985).

hours worked is captured through (21c), providing consistent estimates of the parameters.

7.3.1 Income Taxes and Billing Ceilings

In order to estimate the model we take account of the institutional incentives imposed on physicians by the government through income ceilings and taxes. These affect the budget constraint and hence hours worked. We describe briefly here the institutions and method. Details are also presented in the Appendix; see also Somé (2016). Note, given we observe gross earnings, we solve for optimal hours given the tax rates and income ceilings. We then match the implied gross earnings that optimal hours implies to observed earnings.
Billing Ceilings:

Prior to 1999, the government of Québec imposed half-yearly billing ceilings[25] on physicians. Payment for billed services, beyond the ceiling, was reduced by 75%.

Let $\overline{E}_{w,c}$ denote the weekly income ceiling.[26] The weekly earnings derived from seeing patients,

$$\mathcal{E} = wh_s^{\delta}\epsilon$$

allows us to calculate the number of weekly hours needed to obtain $\overline{E}_{w,c}$,

$$\overline{h}_{s,c} = \left(\frac{\overline{E}_{w,c}}{w\epsilon}\right)^{1/\delta}.$$

Let $\tau_c = 0.75$ be the penalty for exceeding the billing ceiling. The potential earnings (or budget constraint) of the physician is then given by

$$\mathcal{E} = \begin{cases} wh_s^{\delta}\epsilon & \text{if} \quad h_s \leq \overline{h}_{s,c} \\ (1-\tau_c)wh_s^{\delta}\epsilon & \text{if} \quad h_s > \overline{h}_{s,c}. \end{cases}$$

The penalty implies a kink in potential earnings at $\overline{h}_{s,c}$ which depends on both δ and ϵ.

Income Taxes:

The budget constraint becomes more complex when taking account of income taxes. We calculated the marginal tax rates, including both provincial and federal income taxes.

[25] The income ceilings for specialists was set at 150 thousand CAN dollars per semester between 1996 and 1999, except for neurologists, the ceiling was 142.5 thousand CAN dollars per semester.

[26] We convert to a weekly ceiling by dividing the annual income ceiling by the average weeks worked per year in the sample. The average weeks worked per year is 45.83 for physicians providing 2 services, 45.70 for physicians providing 3 services and 44.2 for physicians providing 4 services.

For example, in 2001 the tax structure is:

$$\text{Tax rate} = \begin{cases} \tau_1 = 33\% & \text{if} \quad 0 \leq E < 26,000 \\ \tau_2 = 37.25\% & \text{if} \quad 26,000 \leq E < 30,754 \\ \tau_3 = 43.25\% & \text{if} \quad 30,754 \leq E < 52,000 \\ \tau_4 = 46.5\% & \text{if} \quad 52,000 \leq E < 61,509 \\ \tau_5 = 50.5\% & \text{if} \quad 61,5090 \leq E < 100,000 \\ \tau_6 = 53.5\% & \text{if} \quad E \geq 100,000. \end{cases}$$

Since the marginal tax rate depends on income, it will depend on hours worked (and ϵ). We proceed by calculating the virtual budget constraints associated with each marginal tax rate, ignoring at first any billing ceilings. For example, let $\overline{h_{s,1}}$ be the maximum number of hours a physician can work and still be in the lowest income-tax bracket, taxed at τ_1. Then, for $h_s > \overline{h_{s,1}}$, we solve for virtual income, $\overline{B_2}$, that equates

$$\begin{aligned} \overline{B}_2 + (1 - \tau_2) w \overline{h}_{s,1}^{\delta} \epsilon =& (1 - \tau_1) w \overline{h}_{s,1}^{\delta} \epsilon \\ \Leftrightarrow \overline{B}_2 =& (\tau_2 - \tau_1) w \overline{h}_{s,1}^{\delta} \epsilon \\ =& (\tau_2 - \tau_1) \overline{E}_{w,1}. \end{aligned}$$

This generalizes easily to find the virtual income that equates earnings at between the j^{th} and $(j-1)^{th}$ income-tax bracket:

$$\overline{B}_j = \sum_{k=1}^{j-1} (\tau_{i+1} - \tau_i) \overline{E}_{w,i}.$$

Billing ceilings are easily added by noting that physicians are taxed on income received. Once the billing ceiling is attained, after tax earnings become

$$(1 - \tau_j)(1 - \tau_c) w h^{\delta} \epsilon$$

, where τ_j is the marginal tax rate at the j^{th} income-tax bracket. To calculate the optimal hours in this context we proceed piecewise throughout the composite budget constraint following Hausman (1979) and Moffitt (1990). Given the kink points, $\bar{h}_{s,c}$ depend on ϵ, the program must be solved for each draw of ϵ, for each individual.

7.3.2 Full-Information Results

The results are presented in Table 5. We present two versions of the model. In the first version, $b(\mathbf{x})$ and $\gamma(\mathbf{x})$ depend on gender and age. The second version adds a trend term to $b(\mathbf{x})$ for service 3. One interpretation of this specification is to control for possible endogeneity of the changes to service prices due to technological change that affects the production of type 3 services relative to type 1 and type 2 services. Given the prices of services 3 increase by more that service 1 or 2, one would expect the coefficient on the trend term to be negative if technological change is affecting prices.[27] We exclude $dmid$ from $\gamma(x)$ since the statistics from Table 1 suggest that middle-aged physicians do not differ in hours worked from young physicians. We also exclude a french-speaking dummy variable from $\gamma(\mathbf{x})$ as it caused collinearity (and convergence) problems. We specify non-labour income in period t as a (non stochastic) function of the market return in period t. For the version without a trend we use the same instruments as in the limited-information model: the relevant prices and $dmale, dfrench, dmid, dold, market, market * dmid, market * dold$. When the trend is included, the estimation algorithm did not converge with these instruments. We therefore dropped the market interaction terms and dfrench.

The parameter estimates are similar across specifications. The utility function parameters δ and ρ are precisely estimated, as are the constant terms in $b(\mathbf{x})$. The estimated value of δ is 0.647 without the trend, and 0.623 with the trend. The value of ρ is somewhat more sensitive to the inclusion of the trend. Its value is -0.195 without a trend and -0.242 with the trend. The coefficients which determine the dependence of production on characteristics are generally less precisely estimated. There is no evidence that male physicians in this sample differ in productivity from their female counterparts.[28] Middle aged physicians are more productive than young (inexperienced) physicians. This is consistent with learning by doing as experienced physicians are able to perform diagnoses and services more quickly. Older physicians, display no productivity differences from their young counterparts, suggesting that the productivity profile is concave in age (or experience). The preference for leisure displays little variation across gender, although older physicians have a lower value of γ which leads to working fewer hours. The p-value on this coefficient is 0.107 for the version without a trend and 0.104 for the version with a trend.

[27] The coefficient estimates on the trend terms for services one and two were very close to zero, statistically insignificant and their inclusion led to imprecise estimates of other coefficients in the model. We also exclude the french-language dummy and the interactions between age and the market return from the instrument set as their inclusion caused convergence problems due to multicollinearity.

[28] The reader should bear in mind that our sample is restricted to physicians providing 2 aggregate services. It will be interesting to see if similar results are found on other samples of physicians.

The inclusion of the trend does not alter the parameter estimates in sign or the order of magnitude. The market return coefficient is positive and statistically significant, yet it is small. Income is measured thousands of dollars annually, suggesting that a one percent increase in the stock market return leads to a 20$ increase in non-labour income. Taken literally, this suggests that physicians had very little money invested in the stock market. An alternative interpretation is that this coefficient reflects perceived income generated in the stock market. Under this interpretation physicians adjust their hours worked in response to stock-market changes as if a one percent increase in the stock market generated 20$ of income. Inclusion of the trend term nearly doubles the stock market coefficient. Interestingly, the trend term is negative and statistically significant. While its inclusion does not significantly affect our results (or the calculated elasticities), this is consistent with prices being set, at least in part, in response to factors, such as technological change, that affect physician productivity, something that warrants further investigation in future work.

8 Incentive Effects

Estimation of the full-information model allows us to provide a complete characterization of the reaction of physicians to monetary incentives. We use our parameter estimates from the version without a trend to calculate the income and substitution effects of price changes on total hours providing services, h_s and services supplied. We calculate the effects for each observation in the year 2002 and then average over these observations. In each case, we report the overall effect of the price change, along with its income and substitution effect.[29] Simulated 95% confidence intervals are reported below the calculated estimate.[30]

The own-price elasticities for services are presented in Table 6. The estimated substitution effects are all positive and larger than their corresponding estimated lower bounds from Table 4. The estimated income effects are all negative, indicating that leisure is a normal good. For all cases the overall elasticity is significantly positive, suggesting that the substitution effect dominates the income effect when a single price is changed.

[29] The income effects are calculated using the derivative of utility with respect to non-labour income, y. If our estimates of y include a preference parameter, capturing perceived income for example, then the elasticity should be interpreted in those terms.

[30] The bootstrap is parametric. Following Krinsky and Robb (1986) and Krinsky and Robb (1990), we repeatedly draw, parameters from a normal distribution, setting the mean to the estimated parameter vector and using the estimated variance-covariance matrix of the parameter estimates. The reported confidence intervals are based on 999 replications.

Table 5: Full-Information Estimates

Parameter	Model and Estimates	
	(1) No Trend	(2) Trend
δ	0.647*** (0.083)	0.623*** (0.075)
ρ	-0.195*** (0.052)	-0.242*** (0.049)
$b_{1,12}$	1.316*** (0.211)	1.387*** (0.192)
$b_{1,13}$	-0.957*** (0.274)	-0.932** (0.270)
b_2	1.948*** (0.291)	2.029*** (0.244)
b_3	0.952*** (0.301)	1.228*** (0.247)
b_{Male}	-0.094 (0.155)	-0.087 (0.153)
b_{Dmid}	0.199* (0.107)	0.206* (0.109)
b_{Dold}	0.093 (0.151)	0.103 (0.153)
Υ_{Market}	0.023*** (0.010)	0.046*** (0.010)
b_{Trend_3}	– –	-0.052*** (0.018)
Γ_{Male}	0.003 (0.170)	0.003 (0.173)
Γ_{Dold}	-0.200 (0.124)	-0.209* (0.127)
σ_ϵ	0.876*** (0.244)	0.775*** (0.263)
Observations	1300	1300

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7 presents the cross-price elasticities for services. The overall elasticities are all negative, indicating services are gross substitutes. The lack of symmetry reflects the non-linear effects in the changing prices, noted in table 7. Table 8 presents the changes in hours worked, devoted to providing services, due to changes in prices and the wage index. The substitution effects are all positive, but are dominated by negative income effects; hours elasticities are negative.

Table 6: Own-Price Service Elasticities 2002

	Estimates		
Specification	(1)	(2)	(3)
Parameter	Elasticitiy	Income Effect	Substitution Effect
$\eta_{A1,P1_{12}}$	1.521 [1.292, 1.694]	-0.636 [-1.059, -0.481]	2.156 [1.922, 2.542]
$\eta_{A2,P2}$	0.193 [0.106, 0.377]	-0.081 [-0.181, -0.403]	0.273 [0.156, 0.534]
$\eta_{A1,P1_{13}}$	1.537 [1.238, 1.708]	-0.689 [-1.205, -0.487]	2.227 [1.836, 2.691]
$\eta_{A3,P3}$	0.166 [0.040, 0.552]	-0.074 [-0.263, -0.019]	0.240 [0.061, 0.799]

8.1 Model Fit

The model fit is presented for the version without trend in Figure 1. We concentrate on the predicted and observed aggregate first moments of log earnings and log hours. Predicted moments are given by the hollow symbols and observed moments, the solid symbols. While a statistical test, such as one based on the value of the overidentification statistic, is technically rejected by the data, it is clear that the model replicates the observed moments quite well. In particular it matches very well the increase in earnings following the rise in prices. It is notable, however, that there is a tendancy to overestimate both hours and earnings in year 4.

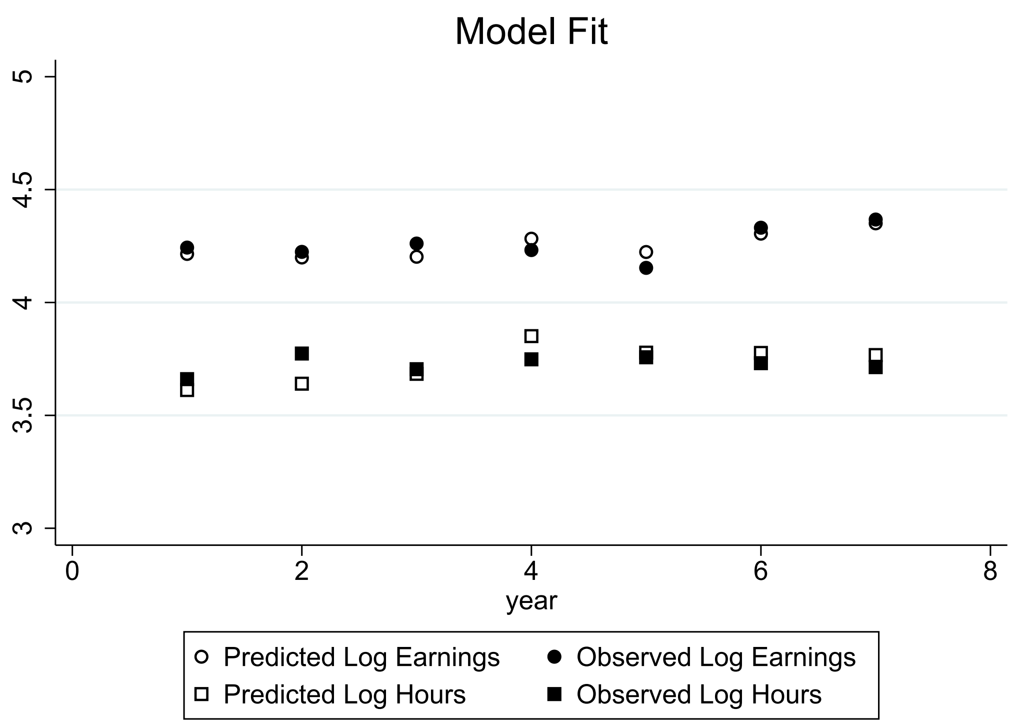
Table 7: Cross-Price Service Elasticities 2002

	Estimates		
Specification	(1)	(2)	(3)
Parameter	Elasticitiy	Income Effect	Substitution Effect
$\eta_{A1,P2_{12}}$	-1.639 [-1.788, -1.458]	-0.081 [-0.181, -0.043]	-1.558 [-1.710, -1.362]
$\eta_{A2,P1}$	-0.311 [-0.515, -0.208]	-0.636 [-1.059, -0.481]	0.325 [-0.076, 0.722]
$\eta_{A1,P3_{13}}$	-1.665 [-1.842, -1.389]	-0.074 [-0.263, -0.019]	-1.592 [-1.774, -1.170]
$\eta_{A3,P1}$	-0.294 [-0.671, -0.164]	-0.689 [-1.205, -0.487]	0.395 [-0.115, 0.830]

Table 8: Hours Elasticities 2002

	Estimates		
Specification	(1)	(2)	(3)
Parameter	Elasticitiy	Income Effect	Substitution Effect
$\eta_{h_s,P1_{12}}$	-0.162 [-0.302, -0.091]	-0.983 [-1.653, -0.739]	0.821 [0.619, 1.391]
$\eta_{h_s,P1_{13}}$	-0.179 [-0.352, -0.088]	-1.066 [-1.882, -0.739]	0.887 [0.621, 1.548]
$\eta_{h_s,P2}$	-0.021 [-0.046, -0.009]	-0.125 [-0.273, -0.068]	0.104 [0.057, 0.230]
$\eta_{h_s,P3}$	-0.019 [-0.058, -0.006]	-0.115 [-0.392, -0.030]	0.096 [0.024, 0.339]
$\eta_{h_s,w_{12}}$	-0.117 [-0.158, -0.074]	-0.717 [-0.750, -0.677]	0.599 [0.547, 0.654]
$\eta_{h_s,w_{13}}$	-0.116 [-0.152, -0.074]	-0.689 [-0.729, -0.645]	0.573 [0.512, 0.641]

Figure 1



8.2 Policy Simulation

Estimation of the structural model allows us to predict how physicians would respond to policy changes by the government. As the data are historic, we can take advantage of past price increases enacted by the government and compare the model's predictions to reported actual responses. Between 2007 and 2011, the Québec government increased the prices paid for physician services by 30%. Contandriopoulos and Perroux (2013) presented aggregate evidence that this increase led physicians to reduce their supply of services.

To evaluate the effects of this policy, we calculated (19) at the estimated parameter values. The results are presented in bottom two rows of Table 8 for physicians providing services 1 and 2, $\eta_{h_s,w_{12}}$, and for physicans providing services 1 and 3, $\eta_{h_s,w_{13}}$. In both cases, hours worked and the volume of all services are predicted to decrease. The hours elasticities are -0.143 and -0.132. Multiplying by 30 and by the relevant estimate of δ gives estimates of the percent service response to the 30% increase in all prices. This is -2.78% for physicians providing services 1 and 2 and -2.56% for physicians providing services 1 and 3.

These results contrast with those in which the price of single service is increased (see Table 6), which give positive own-price effects. The difference here is due to the lack of a substitution effect on any specific service. A broad-based price increase does not change relative prices, but only affects the return to an hour's work (the wage index). It therefore introduces an income and substitution effect on hours devoted to services, which are then distributed over all services.[31]

9 Discussion and Conclusion

We have developed and estimated a structural labour supply model that incorporates the production of medical services and multitasking into the standard consumption/leisure trade-off. Our model gives rise to a conditional earnings function – the maximum earnings a physician can generate for a given number of hours worked. When evaluated at optimal hours, the conditional earnings function shares many properties of other maximum-value functions. Most importantly for our purposes are Le Chatelier effects: with total hours fixed, the second-order effects of price changes on earnings identify a lower bound to the own-price substitution effect.

The conditional earnings function can be estimated using limited information and full-information methods. Limited information methods are sufficient for identifying the lower bound to the own-price substition effect. Full-information methods explain earnings and hours simultaneously. They impose more restrictions on the data, but have the advantage of identifying the complete reaction to incentives, including both substitution and income effects.

We have applied our model to a sample of physicians who were paid fee-for-service contracts in Québec, Canada. Our results suggest that physicians react to monetary incentives. The lower-bound to the own-price substitution effect is positive and statistically significant. While income effects are present, and tend to reduce hours worked and services provided, substitution effects outweigh them when the price of a single service is changed. Changing many prices in the same proportion however, introduces a large in-

[31]This result is also consistent with a " target income hypothesis" (see Kantarevic, Kralj, and Weinkauf, 2008; Rizzo and Blumenthal, 1994; McGuire and Pauly, 1991).

come effect which reduces the supply of services. These results have policy implications for the provision of heath services. Governments (or other health care providers) who are faced with increased demand for particular medical services (and accompanying waiting times) can use price controls to increase the supply of those services. Meanwhile, broad-based price increases induce income effects (with implications for the access to health care) that should be taken into account in government negotiations with physicians. We note that, while our approach to modelling behaviour differs, our results pointing to the importance of the income effect are qualitatively consistent with those of Fortin, Jacquemet, and Shearer (2017) who used flexible functional forms to approximate the utility function and discretized the choice set over practice variables.

The simplicity of our model is one of its attractive features. It is parsimonious, leading to a relatively small number of estimated parameters and easily interpretable comparative statics. Yet it is powerful enough to predict physician behaviour, capturing both income and substitution effects. Nevertheless, the model is limited and it can be extended in various ways to allow for a richer analysis of physician behaviour. While our sample only includes physicians who are present before and after the price changes, we have ignored participation in physician's labour-supply decision. Incorporating participation decisions into our model shifts attention to moments that are conditional on working. To the extent that participation decisions depend on potential productivity, this can affect the parameter estimates.

Part of the model's parsimony is due to the aggregation of services and the assumption of common shocks. Eliminating aggregation and introducing service-specific shocks would be an interesting extension, but would increase the numerical intensity of solving and estimating the model. Doing so would allow the incorporation of demand and technology shocks as determinants of the variation in observed services. It would also allow consideration of agency questions as service-specific shocks might be observed uniquely by the physician. This would allow the measurement of the extent of asymmetric information in the medical profession.

While our results demonstrate the usefulness of the earnings function in the econometric analysis of incentives, particularly in multitasking settings, our particular application has been limited by the available instruments in our data. More extensive data sets may have a richer set of instruments and provide more precise results. A popular instrument in labour-supply models is based on changes in the tax rate (*e.g.,* Blundell, Duncan, and Meghir, 1998; Showalter and Thurston, 1997). While no tax reform was present in the years covered by our data, recent reforms have taken place. It would be interesting to investigate the power of these reforms as instruments in the earnings function. Other approaches may also be available to estimate the earnings function. We have based our empirical work on a parameterized economic model. Flexible functional forms may offer a less restrictive approach to identifying incentive effects.

Finally, we have concentrated on evaluating the quantity (volume-increase) response of physicians to fee increases. It would be interesting to extend this model to account for the quality of services provided. Estimating a model that takes account of the quality of care will require data on the health outcomes of patients and following patients through time.

References

ALLIN, S., M. BAKER, M. ISABELLE, AND M. STABILE (2015): "Physician Incentives and the Rise in C-Sections: Evidence from Canada," *NBER Working Papers*, Working Paper 21022.

AMEMIYA, T. (1985): *Advanced Econometrics*. Harvard University Press.

BALTAGI, B. H., E. BRATBERG, AND T. H. HOLMAS (2005): "A panel data study of physicians' labor supply: the case of Norway," *Health Economics*, 14(10), 1035–1045.

BLOMQUIST, N. S. (1989): "Comparative Statics for Utility Maximization Models with Nonlinear Budget Constraints," *International Economic Review*, 30(2), 275–296.

BLUNDELL, R., A. DUNCAN, AND C. MEGHIR (1998): "Estimating Labor Supply Responses Using Tax Reforms," *Econometrica*, 66(4), 827–861.

BLUNDELL, R., AND T. MACURDY (1999): "Labor Supply: a Review of Alternative Approaches," in *Handbook of Labor Economics, Volume 3A*, ed. by O. C. Ashenfelter, and D. Card. Elsevier.

BLUNDELL, R., AND I. WALKER (1986): "A Life-Cycle Consistent Empirical Model of Family Labour Supply Using Cross-Section Data," *The Review of Economic Studies*, 53(4), 539–558.

BROWNING, M., AND C. MEGHIR (1991): "The Effects of Male and Female Labor Supply on Commodity Demands," *Econometrica*, 59(4), 925–951.

CONTANDRIOPOULOS, D., AND M. PERROUX (2013): "Fee Increases and Target Income Hypothesis: Data from Quebec on Physicians' Compensation and Service Volumes," *Healthcare Policy*, 9, 30–35.

DAVIDSON, R., AND J. G. MACKINNON (1993): *Estimation and Inference in Econometrics*. Oxford University Press.

DEATON, A., AND J. MUELLBAUER (1980): *Economics and Consumer Behavior*. Cambridge University Press.

EVANS, R. G. (1974): "Supplier-Induced Demand: Some Empirical Evidence and Implication," in *The Economics of Health and Medical Care*, ed. by M. Perlman. Macmillan.

FELDSTEIN, M. S. (1970): "The Rising Price of Physicians' Services," *The Review of Economics and Statistics*, 52(2), 121–33.

FORTIN, B., N. JACQUEMET, AND B. SHEARER (2017): "Labour Supply and Service Intensity: Theory and Evidence on Physicians," *mimeo*.

GRUBER, J., AND M. OWINGS (1996): "Physician Financial Incentives and Cesarean Section Delivery," *The RAND Journal of Economics*, 27(1), 99–123.

HAUSMAN, J. A. (1979): "The econometrics of labor supply on convex budget sets," *Economics Letters*, 3(2), 171–174.

HECKMAN, J. J. (1974): "Life Cycle Consumption and Labor Supply: An Explanation of the Relationship Between Income and Consumption Over the Life Cycle," *American Economic Review*, 64(1), 188–94.

HOLMSTROM, B., AND P. MILGROM (1991): "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics & Organization*, 7, 24–52.

HURLEY, J., AND R. LABELLE (1995): "Relative fees and the utilization of physicians' services in Canada," *Health Economics*, 4(6), 419–438.

HEBERT, G. (2016): "La rémunération des médecins québécois," *Institut de recherche et d'informations socioéconomiques*.

KALB, G., D. KUEHNLE, A. SCOTT, T. C. CHENG, AND S.-H. JEON (2018): "What factors affect physicians' labour supply: Comparing structural discrete choice and reduced-form approaches," *Health Economics*, 27(2), e101–e119.

KALMAN, P. J., AND M. D. INTRILIGATOR (1973): "Generalized Comparative Statics with Applications to Consumer Theory and ProducerTheory," *International Economic Review*, 14(2), 473–486.

KANTAREVIC, J., B. KRALJ, AND D. WEINKAUF (2008): "Income effects and physician labour supply: evidence from the threshold system in Ontario," *Canadian Journal of Economics*, 41(4), 1262–1284.

KRINSKY, I., AND L. ROBB (1986): "On Approximating the Statistical Properties of Elasticities," *Review of Economics and Statistics*, 68(4), 715–719.

——— (1990): "On Approximating the Statistical Properties of Elasticities: Correction," *Review of Economics and Statistics*, 72(1), 189–190.

LEMIEUX, V., P. BERGERON, C. BÉGIN, AND G. BÉLANGER (2003): *Le système de santé au Québec: Organisations, acteurs et enjeux*. Les presses de l'Université Laval, Québec, Canada.

LUNDBERG, S. (1988): "Labor Supply of Husbands and Wives: A Simultaneous Equations Approach," *The Review of Economics and Statistics*, 70(2), 224–235.

MCGUIRE, T. G., AND M. V. PAULY (1991): "Physician response to fee changes with multiple payers," *Journal of Health Economics*, 10(4), 385–410.

MILGROM, P., AND J. ROBERTS (1996): "The Le Chatelier Principle," *The American Economic Review*, 86(1), 173–179.

MILLIGAN, K. (2016): *Canadian Tax and Credit Simulator. Database, software and documentation, Version 2016-2.*

MOFFITT, R. (1990): "The Econometrics of Kinked Budget Constraints," *The Journal of Economic Perspectives*, 4(2), 119–139.

NEARY, J. P., AND K. W. S. ROBERTS (1980): "The Theory of Household Behaviour Under Rationing," *European Economic Review*, 13, 25–42.

POLLAK, R. A. (1969): "Conditional Demand Functions and Consumption Theory," *The Quarterly Journal of Economics*, 83(1), 60–78.

——— (1971): "Conditional Demand Functions and the Implications of Separable Utility," *Southern Economic Journal*, 37(4), 423–433.

RIZZO, J. A., AND D. BLUMENTHAL (1994): "Physician labor supply: Do income effects matter?," *Journal of Health Economics*, 13(4), 433–453.

SAMUELSON, P. A. (1947): *Foundations of Economic Analysis*. Harvard University Press.

SHEARER, B., N. SOMÉ, AND B. FORTIN (2018): "Measuring Physicians' Response to Incentives: Evidence on Hours Worked and Multitasking," *CRREP Working Paper 2018-09, Université Laval*.

SHOWALTER, M. H., AND N. K. THURSTON (1997): "Taxes and labor supply of high-income physicians," *Journal of Public Economics*, 66(1), 73–97.

SLOAN, F. A. (1975): "Physician supply behavior in the short run," *Industrial and Labor Relations Review*, 28(4), 549–569.

SOMÉ, H. (2016): "Modelling and Estimating Models of Physician Labour Supply and Productivity," *Phd. Thesis, Laval University*.

STERN, N. (1986): "On the Specification of Labour Supply Functions," in *Unemployment, Searn and Labour Supply*, ed. by R. Blundell, and I. Walker. Cambridge University Press.

STERN, N. H. (1976): "On the Specification of Models of Optimum Income Taxation," *Journal of Public Economics*, 6, 123–62.

VARIAN, H. R. (1992): *Microeconomic Analysis, third edition*. Norton & Company.

ZABALZA, A. (1983): "The CES utility function, non-linear budget constraints and labour supply. Results on female participation and hours," *Economic Journal*, 93, 312–330.

Appendix A1: The Conditional Earnings Function

Proof of Lemma 1.

The proof is by induction. We derive the result for the cases of physicicans providing two and three services and then generalize to J services.

With two services, the first-order conditions (5) can be written

$$\alpha_1 f_1'(h_1^*) - \alpha_2 f_2'(h_s - h_1^*) = 0.$$

Differentiating with respect to h_s gives

$$\alpha_1 f_1''(h_1^*) \frac{\partial h_1^*}{\partial h_s} - \alpha_2 f_2''(h_s - h_1^*) + \alpha_2 f_2''(h_s - h_1^*) \frac{\partial h_1^*}{\partial h_s} = 0$$

or

$$\frac{\partial h_1^*}{\partial h_s} = \frac{\alpha_2 f_2''(h_s - h_1^*)}{\alpha_1 f_1''(h_1^*) + \alpha_2 f_2''(h_s - h_1^*)} = \frac{\alpha_2 f_2''(h_2^*)}{\alpha_1 f_1''(h_1^*) + \alpha_2 f_2''(h_2^*)} > 0,$$

since $f_j'' < 0$, $\alpha_j > 0$, $j = 1, 2$; and $h_2^* = h_s - h_1^*$ from the constraint. Solving from the constraint $h_2^* = h_s - h_1^*$,

$$\begin{aligned} \frac{\partial h_2^*}{\partial h_s} =& 1 - \frac{\partial h_1^*}{\partial h_s} \\ =& 1 - \frac{\alpha_2 f_2''(h_2^*)}{\alpha_1 f_1''(h_1^*) + \alpha_2 f_2''(h_2^*)} \\ =& \frac{\alpha_1 f_1''(h_1^*)}{\alpha_1 f_1''(h_1^*) + \alpha_2 f_2''(h_2^*)} > 0. \end{aligned} \tag{25}$$

For three services, the first-order conditions (5) can be written

$$\begin{aligned} \alpha_1 f_1'(h_1^*) - \alpha_3 f_3'(h_s - h_1^* - h_2^*) =& 0 \\ \alpha_2 f_2'(h_2^*) - \alpha_3 f_3'(h_s - h_1^* - h_2^*) =& 0. \end{aligned} \tag{26}$$

Differentiating with respect to h_s we have

$$\left[\alpha_1 f_1''(h_1^*) + \alpha_3 f_3''(h_3^*)\right] \frac{\partial h_1^*}{\partial h_s} + \alpha_3 f_3''(h_3^*) \frac{\partial h_2^*}{\partial h_s} = \alpha_3 f_3''(h_3^*), \tag{27}$$

$$\alpha_3 f_3''(h_3^*) \frac{\partial h_1^*}{\partial h_s} + \left[\alpha_2 f_2''(h_2^*) + \alpha_3 f_3''(h_3^*)\right] \frac{\partial h_2^*}{\partial h_s} = \alpha_3 f_3''(h_3^*). \tag{28}$$

Solving (28) gives

$$\frac{\partial h_2^*}{\partial h_s} \quad = \quad \alpha_3 f_3''(h_3^*) \frac{\left(1 - \frac{\partial h_1^*}{\partial h_s}\right)}{\left[\alpha_2 f_2''(h_2^*) + \alpha_3 f_3''(h_3^*)\right]}. \tag{29}$$

Substituting into (27) and rearranging, we have

$$\frac{\partial h_1^*}{\partial h_s} = \frac{\alpha_2\alpha_3 f_2''(h_2^*) f_3''(h_3^*)}{\left[\alpha_1\alpha_2 f_1''(h_1^*) f_2''(h_2^*) + \alpha_1\alpha_3 f_1''(h_1^*) f_3''(h_3^*) + \alpha_2\alpha_3 f_2''(h_2^*) f_3''(h_3^*)\right]} > 0.$$

Substituting back into (29) gives

$$\frac{\partial h_2^*}{\partial h_s} = \frac{\alpha_1\alpha_3 f_1''(h_1^*) f_3''(h_3^*)}{\left[\alpha_1\alpha_2 f_1''(h_1^*) f_2''(h_2^*) + \alpha_1\alpha_3 f_1''(h_1^*) f_3''(h_3^*) + \alpha_2\alpha_3 f_2''(h_2^*) f_3''(h_3^*)\right]} > 0.$$

Finally, using the constraint, $h_3^* = h_s - h_1^* - h_2^*$, gives

$$\frac{\partial h_3^*}{\partial h_s} = \frac{\alpha_1\alpha_2 f_1''(h_1^*) f_2''(h_2^*)}{\left[\alpha_1\alpha_2 f_1''(h_1^*) f_2''(h_2^*) + \alpha_1\alpha_3 f_1''(h_1^*) f_3''(h_3^*) + \alpha_2\alpha_3 f_2''(h_2^*) f_3''(h_3^*)\right]} > 0.$$

With J services, these formulas generalize to:

$$\frac{\partial h_j^*}{\partial h_s} = \frac{\prod_{i \neq j} \alpha_k f_k''(h_k^*)}{\sum_{j=1}^{J} \left[\prod_{i \neq j} \alpha_k f_k''(h_k^*)\right]} > 0,$$

which is positive since each term $\prod_{k \neq j} \alpha_k f_k''(h_k^*)$ has the same sign.

Properties of the Conditional Earnings Function
The conditional earnings function has the following properties.

1. The partial derivative with respect to α_j is equal to $A_j^*(\boldsymbol{\alpha}, h_s)$, the conditional supply of service j.

$$\begin{aligned}
\mathcal{E}(\boldsymbol{\alpha}; h_s) &= \sum_{j=1}^{J} \alpha_j A_j^*(\boldsymbol{\alpha}; h_s) \\
&= \sum_{j=1}^{J} \alpha_j f_j(h_j^*(\boldsymbol{\alpha}; h_s)) \\
&= \alpha_j f_j(h_j^*(\boldsymbol{\alpha}; h_s)) + \sum_{k \neq j}^{J-1} \alpha_k f_k(h_k^*(\boldsymbol{\alpha}; h_s)) + \alpha_J f_J(h_J^*(\boldsymbol{\alpha}; h_s)) \\
&= \alpha_j f_j(h_j^*(\boldsymbol{\alpha}; h_s)) + \sum_{k \neq j}^{J-1} \alpha_k f_k(h_k^*(\boldsymbol{\alpha}; h_s)) + \alpha_J f_J(h_s - \sum_{k=1}^{J-1} h_k^*(\boldsymbol{\alpha}; h_s)).
\end{aligned}$$

Taking the partial derivative with respect to α_j gives

$$\begin{aligned}
\frac{\partial \mathcal{E}}{\partial \alpha_j} &= f_j(h_j^*) + \alpha_j f_j'(h_j^*) \frac{\partial h_j^*}{\partial \alpha_j} + \sum_{k \neq j}^{J-1} \alpha_k f_k'(h_k^*) \frac{\partial h_k^*}{\partial \alpha_j} - \alpha_J f_J'(h_s - \sum_{k=1}^{J-1} h_k^*) \sum_{k=1}^{J-1} \frac{\partial h_k^*}{\partial \alpha_j} \\
&= f_j(h_j^*) + \sum_{k=1}^{J-1} \left[\alpha_k f_k'(h_k^*) - \alpha_J f_J'(h_J^*) \right] \frac{\partial h_k^*}{\partial \alpha_j} \\
&= f_j(h_j^*) = A_j^* \quad \text{by (6).}
\end{aligned}$$

2. The second partial derivative of the earnings function with respect to α_j is equal to the slope of the conditional supply of service j:

$$\frac{\partial^2 \mathcal{E}}{\partial \alpha_j^2} = \frac{\partial A_j^*}{\partial \alpha_j} = f'(h_j^*) \frac{\partial h_j^*}{\partial \alpha_j}.$$

This follows directly from property 1.

3. Convex in prices. Since physicians select h_j^* to maximize earnings, when prices adjust, they can increase their earnings by more than the simple price change by reoptimizing.
Let $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}$ be two price vectors and $\boldsymbol{\alpha_3} = \theta \boldsymbol{\alpha_1} + (1-\theta)\boldsymbol{\alpha_2}$ for $\theta \in (0,1)$. Let $h_{1,j}^*, h_{2,j}^*, h_{3,j}^*$ denote the optimal hours allocated to service j under price vectors $\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2},$ and $\boldsymbol{\alpha_3}$, respectively. The

earnings function evaluated at $\boldsymbol{\alpha_3}$ is

$$\begin{aligned}
\mathcal{E}(\boldsymbol{\alpha_3}; h_s) &= \sum_{j=1}^{J} \alpha_{3,j} f_j(h^*_{3,j}) \\
&= \sum_{j=1}^{J} (\theta\alpha_{1,j} + (1-\theta)\alpha_{2,j}) f_j(h^*_{3,j}). \qquad (30)
\end{aligned}$$

Since $h^*_{j,1}$ is optimal at $\boldsymbol{\alpha_1}$, and $h^*_{j,2}$ is optimal at $\boldsymbol{\alpha_2}$ it must be the case that

$$\sum_{j=1}^{J} \alpha_{1,j} f_j(h^*_{1,j}) \geq \sum_{j=1}^{J} \alpha_{1,j} f_j(h^*_{3,j}) \quad \forall j \quad \text{and} \quad \sum_{j=1}^{J} \alpha_{2,j} f_j(h^*_{2,j}) \geq \sum_{j=1}^{J} \alpha_{2,j} f_j(h^*_{3,j}) \quad \forall j.$$

Substituting into (30) gives

$$\begin{aligned}
\mathcal{E}(\boldsymbol{\alpha_3}; h_s) &= \sum_{j=1}^{J} (\theta\alpha_{1,j} f_j(h^*_{3,j}) + (1-\theta)\alpha_{2,j} f_j(h^*_{3,j})) \\
&\leq \sum_{j=1}^{J} \theta\alpha_{1,j} f_j(h^*_{1,j}) + (1-\theta)\alpha_{2,j} f_j(h^*_{2,j}) = \theta\mathcal{E}(\boldsymbol{\alpha_1}; h_s) + (1-\theta)\mathcal{E}(\boldsymbol{\alpha_2}; h_s)
\end{aligned}$$

so:

$$\mathcal{E}\left(\theta\boldsymbol{\alpha_j} + (1-\theta)\boldsymbol{\alpha_2}; h_s\right) \leq \theta\mathcal{E}\left(\boldsymbol{\alpha_1}; h_s\right) + (1-\theta)\mathcal{E}\left(\boldsymbol{\alpha_2}; h_s\right).$$

4. The second derivative of the earnings function with respect to α_j provides a lower bound to the own-price substitution effect of α_j on A_j.

   Let $\bar{u}$ denote the level of utility attained when supplying optimal hours h_s^*. We use $h_s^{*,\bar{u}}$ to denote the hicksian supply of hours. Following Pollak (1969), evaluate the conditional supply for service j, $h_j^*(\boldsymbol{\alpha}; h_s)$, at $h_s^{*,\bar{u}}$. This gives the identity

$$h_j^*(\boldsymbol{\alpha}; h_s^{*,\bar{u}}) \quad \equiv \quad h_j^{*\bar{u}}(\boldsymbol{\alpha}; \bar{u}),$$

   where $h_j^{*\bar{u}}(\boldsymbol{\alpha}; \bar{u})$ is the hicksian supply of hours to service j. Note, given f_j is a monotonic increasing function, the following are implied:

$$\begin{aligned} A_j^* = f_j(h_j^*(\boldsymbol{\alpha}, h_s^{*\bar{u}})) \quad &\equiv \quad f_j(h_j^{*\bar{u}}(\boldsymbol{\alpha}; \bar{u})) = A_j^{\bar{u}}, \\ f_j'(h_j^*) \quad &= \quad f_j'(h_j^{*\bar{u}}). \end{aligned} \tag{31}$$

   Differentiating the identity (31) with respect to α_j gives

$$\begin{aligned} & f_j'(h_j^*)\frac{\partial h_j^*}{\partial \alpha_j} + f_j'(h_j^*)\frac{\partial h_j^*}{\partial h_s}\frac{\partial h_s^{*\bar{u}}}{\partial \alpha_j} = f_j'(h_j^{*\bar{u}})\frac{\partial h_j^{*\bar{u}}}{\partial \alpha_j} \\ \text{or} \quad & f_j'(h_j^*)\frac{\partial h_j^*}{\partial \alpha_j} + f_j'(h_j^*)\frac{\partial h_j^*}{\partial h_s}\frac{\partial h_s^{*\bar{u}}}{\partial \alpha_j} = f_j'(h_j^*)\frac{\partial h_j^{*\bar{u}}}{\partial \alpha_j}, \end{aligned} \tag{32}$$

   since $f_j'(h_j^*) = f_j'(h_s^{*\bar{u}})$. The term on the right-hand side of (32) is the own-price substitution effect of α_j on A_j. Rearranging gives

$$f_j'(h_j^*)\frac{\partial h_j^*}{\partial \alpha_j} = f_j'(h_j^*)\frac{\partial h_j^{*\bar{u}}}{\partial \alpha_j} - f_j'(h_j^*)\frac{\partial h_j^*}{\partial h_s}\frac{\partial h_s^{*\bar{u}}}{\partial \alpha_j}.$$

   The term on the left-hand side is the second partial derivative of the earnings function with respect to α_j, from 2. The term

$$\frac{\partial h_j^*}{\partial h_s}\frac{\partial h_s^{*\bar{u}}}{\partial \alpha_j} > 0.$$

   Any increase in α_j increases the return to hours worked h_s. Hence $\frac{\partial h_s^{*\bar{u}}}{\partial \alpha_j} > 0$ is positive since the substitution effect on hours worked is positive when indifference curves are strictly convex. Moreover, any increase in h_s is distributed across all services, $\frac{\partial h_j^*}{\partial h_s} > 0$ as shown in Lemma 1.

Appendix A2: Elasticities

Let

$$F(h_s, \alpha_1, \alpha_2, ..., \alpha_J, y, \epsilon) = \boldsymbol{\Gamma}\omega\delta h_s^{\delta-1}\epsilon\left(\omega h_s^{\delta}\epsilon + y\right)^{\rho-1} - 0.5(1-\gamma)2^{1-\rho}\left(T - h_s\right)^{\rho-1},$$

where

$$\omega = \left(\sum_{j=1}^{J}(\alpha_j b_j)^{\frac{1}{1-\delta}}\right)^{1-\delta}$$

and note that optimal h_s^* solves

$$F(h_s, \alpha_1, \alpha_2, ..., \alpha_J, y, \epsilon) = 0,$$

with the second-order condition

$$\frac{\partial F(h_s^*, \alpha_1, \alpha_2, ..., \alpha_J, y, \epsilon)}{\partial h_s} = V_{h_s h_s}(h_s^*) < 0.$$

By the implicit function theorem, we can write

$$h_s^* = \psi(\alpha_1, \alpha_2, ..., \alpha_J, y, \epsilon).$$

Furthermore,

$$\frac{\partial h_s^*}{\partial \alpha_j} = -\frac{\frac{\partial F}{\partial \alpha_j}}{\frac{\partial F}{\partial h_s}},$$

and

$$\frac{\partial h_s^*}{\partial y} = -\frac{\frac{\partial F}{\partial y}}{\frac{\partial F}{\partial h_s}}.$$

We use the following notation:

(i) $M = \omega h_s^{\delta}\epsilon + y$.

(ii) $\omega = \left(\sum_{j=1}^{J}(\alpha_j b_j)^{\frac{1}{1-\delta}}\right)^{1-\delta}$.

(iii) $P_j = (\alpha_j b_j)^{\frac{1}{1-\delta}}$.

We note

$$\frac{\partial \omega}{\partial \alpha_j} = b_j\left(\frac{P_j}{\sum_k P_k}\right)^{\delta} > 0.$$

We will use the following resutls which rely on the parameters satisfying the second-order condition, i.e., $\rho < 1$ and $\delta \in (0,1)$.

$$\begin{aligned}
F_{h_s} &= \gamma(\delta-1)\omega\delta h_s^{\delta-2}\epsilon M^{\rho-1} + \gamma(\rho-1)(\omega\delta h_s^{\delta-1}\epsilon)^2 M^{\rho-2} + 0.5(\rho-1)(1-\gamma)2^{(1-\rho)}\left(T-h_s\right)^{\rho-2}. \\
F_y &= \gamma(\rho-1)\omega\delta h_s^{\delta-1}\epsilon M^{\rho-2}. \\
F_{\alpha_j} &= \gamma\frac{\partial\omega}{\partial\alpha_j}\delta h_s^{\delta-1}\epsilon M^{\rho-1} + \gamma(\rho-1)\omega\delta h_s^{\delta-1}\epsilon\frac{\partial M}{\partial\alpha_j}M^{\rho-2} \\
&= \frac{\partial\omega}{\partial\alpha_j}h_s^{\delta-1}\epsilon\left(\gamma\delta M^{\rho-1} + h_s F_y\right) \gtrless 0.
\end{aligned}$$

Income elasticity, h_s

$$\begin{aligned}
\eta_{h_s,y} &= \frac{y}{h_s}\frac{dh_s}{dy} = \frac{y}{h_s}\left(-\frac{F_y}{F_{h_s}}\right) \\
&= \frac{y}{h_s}\left(\frac{\gamma(1-\rho)\omega\delta h_s^{\delta-1}\epsilon M^{\rho-2}}{F_{h_s}}\right) < 0.
\end{aligned} \tag{33}$$

Income elasticity, h_j
Recall,

$$h_j(h_s) = \frac{P_j}{\sum_{k=1}^{J} P_k}h_s. \tag{34}$$

It follows that

$$\frac{dh_j}{dy}(h_s) = \frac{P_j}{\sum_{k=1}^{J} P_k}\frac{dh_s}{dy} < 0, \quad \text{since} \quad \frac{dh_s}{dy} < 0 \quad \text{from} \quad (33). \tag{35}$$

Using (34) we have

$$\eta_{h_j,y} = \eta_{h_s,y}.$$

Income elasticity, A_j
Recall,

$$A_j = b_j(x_b)h_j^{\delta}\epsilon.$$

It follows that

$$\frac{dA_j}{dy} = \delta b_j(x_b)h_j^{\delta-1}\frac{dh_j}{dy}\epsilon < 0, \quad \text{since} \quad \frac{dh_j}{dy} < 0 \quad \text{from} \quad (35). \tag{36}$$

Using (36), we have the elasticity form

$$\eta_{A_j,y} = \delta\eta_{h_j,y} = \delta\eta_{h_s,y}.$$

Price elasticity, h_s

$$\begin{aligned}
\frac{dh_s}{d\alpha_j} &= -\frac{F_{\alpha_j}}{F_{h_s}} \\
&= \frac{\partial\omega}{\partial\alpha_j}h_s^{\delta-1}\epsilon\left(-\frac{\gamma\delta M^{\rho-1}}{F_{h_s}} - h_s\frac{F_y}{F_{h_s}}\right) \\
&= \frac{\partial\omega}{\partial\alpha_j}h_s^{\delta-1}\epsilon\left(-\frac{\gamma\delta M^{\rho-1}}{F_{h_s}} + h_s\frac{dh_s}{dy}\right), \quad \text{from} \quad (33) \\
&= -\frac{\partial\omega}{\partial\alpha_j}\frac{h_s^{\delta-1}\epsilon\gamma\delta M^{\rho-1}}{F_{h_s}} + h_s^{\delta}\epsilon\frac{\partial\omega}{\partial\alpha_j}\frac{dh_s}{dy} \\
&= -b_j\left(\frac{P_j}{\sum_k P_k}\right)^{\delta}\frac{h_s^{\delta-1}\epsilon\gamma\delta M^{\rho-1}}{F_{h_s}} + h_s^{\delta}\epsilon b_j\left(\frac{P_j}{\sum_k P_k}\right)^{\delta}\frac{dh_s}{dy} \quad \text{from} \quad (9) \\
&= -b_j h_j^{\delta}\epsilon\frac{\gamma\delta M^{\rho-1}}{h_s F_{h_s}} + \epsilon b_j h_j^{\delta}\frac{dh_s}{dy} \\
&= -\gamma\frac{\delta A_j M^{\rho-1}}{h_s F_{h_s}} + A_j\frac{dh_s}{dy}.
\end{aligned}$$

To convert to elasticity terms, multiply by α_j/h_s. After adjusting the income effect, we get:

$$\eta_{h_s,\alpha_j} = -\gamma\frac{\delta\alpha_j A_j M^{\rho-1}}{h_s^2 F_{h_s}} + \frac{\alpha_j A_j}{y}\eta_{h_s,y}.$$

Own-price elasticity, h_j
Recall,

$$h_j(h_s) = \frac{P_j}{\sum_{k=1}^{J} P_k}h_s.$$

We then have:

$$\frac{\partial h_j}{\partial\alpha_j} = \frac{P_j}{\sum_k P_k}\frac{\partial h_s}{\partial\alpha_j} + h_s\frac{\partial}{\partial\alpha_j}\left(\frac{P_j}{\sum_k P_k}\right), \tag{37}$$

but

$$\begin{aligned}\frac{\partial h_s}{\partial \alpha_j} &= -\frac{F_{\alpha_j}}{F_{h_s}} \\ &= -\gamma \frac{\delta A_j M^{\rho-1}}{h_s F_{h_s}} + A_j \frac{dh_s}{dy}.\end{aligned}$$

Substituting back into (37) gives

$$\begin{aligned}\frac{\partial h_j}{\partial \alpha_j} &= \frac{P_j}{\sum_k P_k}\left(-\gamma \frac{\delta A_j M^{\rho-1}}{h_s F_{h_s}} + A_j \frac{dh_s}{dy}\right) + h_s \frac{\partial}{\partial \alpha_j}\left(\frac{P_j}{\sum_k P_k}\right) \\ &= \frac{h_j}{h_s}\left(-\gamma \frac{\delta A_j M^{\rho-1}}{h_s F_{h_s}} + A_j \frac{dh_s}{dy}\right) + \frac{h_s}{\alpha_j(1-\delta)} \frac{P_j \sum_{k\neq j} P_k}{\left(\sum_k P_k\right)^2}; \\ \frac{\alpha_j}{h_j}\frac{\partial h_j}{\partial \alpha_j} &= \left(\frac{1}{1-\delta}\frac{\sum_{k\neq j} P_k}{\sum_k P_k} - \gamma \frac{\delta \alpha_j A_j M^{\rho-1}}{h_s^2 F_{h_s}}\right) + \frac{\alpha_j A_j}{y}\frac{dh_s}{dy}\frac{y}{h_s}.\end{aligned}$$

Finally

$$\eta_{h_j,\alpha_j} = \left(\frac{1}{1-\delta}\frac{\sum_{k\neq j} P_k}{\sum_k P_k} - \gamma \frac{\delta \alpha_j A_j M^{\rho-1}}{h_s^2 F_{h_s}}\right) + \frac{\alpha_j A_j}{y}\eta_{h_s,y}.$$

Own-price elasticity, A_j
Recall,

$$A_j = b_j(x_b) h_j^{\delta} \epsilon.$$

It follows that

$$\begin{aligned}\frac{\partial A_j}{\partial \alpha_j} &= \delta b_j(x_b) h_j^{\delta-1} \epsilon \frac{\partial h_j}{\partial \alpha_j} \\ &= \delta \frac{A_j}{h_j}\frac{\partial h_j}{\partial \alpha_j}; \\ \frac{\alpha_j}{A_j}\frac{\partial A_j}{\partial \alpha_j} &= \delta \frac{\alpha_j}{h_j}\frac{\partial h_j}{\partial \alpha_j}. \\ \eta_{A_j,\alpha_j} &= \delta \eta_{h_j,\alpha_j}.\end{aligned}$$

Cross-price elasticity, h_j
We have:

$$
\begin{aligned}
\frac{\partial h_j}{\partial \alpha_k} &= \frac{P_j}{\sum_k P_k}\frac{\partial h_s}{\partial \alpha_k} + h_s \frac{\partial}{\partial \alpha_k}\left(\frac{P_j}{\sum_k P_k}\right) \\
&= \frac{P_j}{\sum_k P_k}\frac{\partial h_s}{\partial \alpha_k} + h_s \frac{\partial}{\partial \alpha_k}\left(\frac{P_j}{\sum_k P_k}\right) \\
&= \frac{P_j}{\sum_k P_k}\left(-\gamma \frac{\delta A_k M^{\rho-1}}{h_s F_{h_s}} + A_k \frac{dh_s}{dy}\right) - \frac{h_s}{\alpha_k(1-\delta)}\frac{P_k P_j}{(\sum_k P_k)^2}.
\end{aligned}
$$

Using $h_j = \frac{P_j}{\sum_{k=1}^{J} P_k} h_s$ we have:

$$
\begin{aligned}
\frac{\partial h_j}{\partial \alpha_k} &= \left(-\gamma \frac{\delta h_j A_k M^{\rho-1}}{h_s^2 F_{h_s}} + A_k \frac{P_j}{\sum_k P_k}\frac{dh_s}{dy}\right) - \frac{1}{\alpha_k(1-\delta)}\frac{h_k h_j}{h_s} \\
&= -\gamma \frac{\delta h_j A_k M^{\rho-1}}{h_s^2 F_{h_s}} - \frac{1}{\alpha_k(1-\delta)}\frac{h_k h_j}{h_s} + A_k \frac{P_j}{\sum_k P_k}\frac{dh_s}{dy}.
\end{aligned}
$$

Finally, using $\frac{\partial h_j}{\partial y} = \frac{P_j}{\sum_k P_k}\frac{dh_s}{dy}$, we have

$$
\begin{aligned}
\frac{\partial h_j}{\partial \alpha_k} &= -\frac{1}{\alpha_k(1-\delta)}\frac{h_k h_j}{h_s} - \gamma \frac{\delta h_j A_k M^{\rho-1}}{h_s^2 F_{h_s}} + A_k \frac{\partial h_j}{\partial y} \\
&= -\frac{h_j}{\alpha_k}\left(\frac{1}{(1-\delta)}\frac{h_k}{h_s} + \gamma \frac{\delta \alpha_k A_k M^{\rho-1}}{h_s^2 F_{h_s}}\right) + A_k \frac{\partial h_j}{\partial y}.
\end{aligned}
$$

or in elasticity form,

$$
\begin{aligned}
\frac{\alpha_k}{h_j}\frac{\partial h_j}{\partial \alpha_k} &= -\left(\frac{1}{1-\delta}\frac{h_k}{h_s} + \gamma \frac{\delta \alpha_k A_k M^{\rho-1}}{h_s^2 F_{h_s}}\right) + \frac{\alpha_k A_k}{y}\frac{\partial h_j}{\partial y}\frac{y}{h_j}; \\
\eta_{h_j,\alpha_k} &= -\left(\frac{1}{1-\delta}\frac{P_j}{\sum_k P_k} + \gamma \frac{\delta \alpha_k A_k M^{\rho-1}}{h_s^2 F_{h_s}}\right) + \frac{\alpha_k A_k}{y}\eta_{h_j,y},
\end{aligned}
$$

where the last line uses the fact that $h_j = \frac{P_j}{\sum_{k=1}^{J} P_k} h_s$.

Cross-price elasticity, A_j

$$
A_j = b_j(x_b) h_j^{\delta} \epsilon.
$$

It follows that

$$
\begin{aligned}
\frac{\partial A_j}{\partial \alpha_k} &= \delta b_j(x_b) h_j^{\delta-1} \epsilon \frac{\partial h_j}{\partial \alpha_k} \\
&= \delta \frac{A_j}{h_j} \frac{\partial h_j}{\partial \alpha_k}; \\
\frac{\alpha_j}{A_j} \frac{\partial A_j}{\partial \alpha_k} &= \delta \frac{\alpha_k}{h_j} \frac{\partial h_j}{\partial \alpha_k}; \\
\eta_{A_j,\alpha_k} &= \delta \eta_{h_j,\alpha_k}.
\end{aligned}
$$

Appendix A3: Composite Services

To aggregate services we use the hicks composite commodity theorem.[32]

Given n services that can be provided by a physician, the vector of service quantities is $(A_1, A_2, ..., A_n)$ and the associated price vector is $(\alpha_1, \alpha_2, ..., \alpha_n)$. Note, for example, if prices i and j move in the same proportion θ with respect to their base-period prices, denoted α_k^0, α_j^0, then we can write

$$\alpha_{k,t} = \theta_t \alpha_k^0 \quad \text{and} \quad \alpha_{j,t} = \theta_t \alpha_j^0.$$

The relative prices of services i and j are constant in each period:

$$\frac{\alpha_{it}}{\alpha_{jt}} = \frac{\alpha_k^0}{\alpha_j^0}.$$

Now let $q < n$ be the number groups of services with distinct changes in service prices. Let $\theta_1, \theta_2, ..., \theta_q$ denote those price changes and let Θ_j denote the group of services associated with each $\theta_j, \quad j \in \{1, 2, ..., q\}$.

Proposition : *If $(A_1, A_2, \ldots, A_n)$ solves*

$$
\begin{aligned}
\max_{\{M, h_1, h_2, ... h_n, h_t, h_s\}} U &= [M^\rho + (h_t - h_s)^\rho + (T - h_t)^\rho]^{\frac{1}{\rho}} \\
s.t. \quad (i) \quad M &= \sum_{j=1}^{n} \alpha_j A_j + y. \\
(ii) \quad A_j &= b_j h_j^\delta \epsilon, \quad j = 1, 2, ..., n. \\
(iii) \quad h_s &= \sum_{j=1}^{n} h_j.
\end{aligned}
$$

then medical services can be aggregated in $q < n$ groups of services. The aggregate service vector is

[32] See, for example, Deaton and Muellbauer (1980).

$(\sum_{j\in\Theta_1}\alpha_j^0 A_j, \sum_{j\in\Theta}\alpha_j^0 A_j, ..., \sum_{j\in\Theta_q}\alpha_j^0 A_j)$ *and the associated price vector is* $(\theta_1, \theta_2, ..., \theta_q)$.

Proof: The indirect utility function is $V(w, y) = \left[(wh_s^\delta\epsilon + y)^\rho + 2^{1-\rho}(T - h_s)^\rho\right]^{\frac{1}{\rho}}$, where $w = \left[\sum_{j=1}^n (b_j\alpha_j)^{\frac{1}{1-\delta}}\right]^{1-\delta}$. The expenditure function, $e(w, u^0)$, is the amount of non-labour income needed to set to $V(w, e(w, u^0)) = u^0$. This gives:

$$
\begin{aligned}
\left[(wh_s^\delta\epsilon + e(w, u^0))^\rho + 2^{1-\rho}(T - h_s)^\rho\right]^{\frac{1}{\rho}} &= u^0 \quad \text{or} \\
e(w, u^0) &= \left[(u^0)^\rho - 2^{1-\rho}(T - h_s)^\rho\right]^{1/\rho} - wh_s^\delta\epsilon.
\end{aligned}
$$

Applying Shephard's Lemma, the appropriate composite service is the derivative of $e(w, u^0)$ with respect to θ_k (conditional on h_s). We have:

$$
-\frac{de}{d\theta_k} = \frac{dw}{d\theta_k} h_s^\delta\epsilon. \tag{38}
$$

The derivative of w with respect to θ_k is

$$
\begin{aligned}
\frac{dw}{d\theta_k} &= \frac{d}{d\theta_k}\left(\sum_{j\in\Theta_1}(b_j\theta_1\alpha_j^0)^{\frac{1}{1-\delta}} + \sum_{j\in\Theta_2}(b_j\theta_2\alpha_j^0)^{\frac{1}{1-\delta}} + ... + \sum_{j\in\Theta_k}(b_j\theta_k\alpha_j^0)^{\frac{1}{1-\delta}} + ... + \sum_{j\in\Theta_q}(b_j\theta_q\alpha_j^0)^{\frac{1}{1-\delta}}\right)^{1-\delta} \\
&= \sum_{j\in\Theta_k} b_j\alpha_j^0 \left(\frac{(b_j\alpha_j)^{\frac{1}{1-\delta}}}{\Delta}\right)^\delta,
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta &= \sum_{j\in\Theta_1}(b_j\theta_1\alpha_j^0)^{\frac{1}{1-\delta}} + \sum_{j\in\Theta_2}(b_j\theta_2\alpha_j^0)^{\frac{1}{1-\delta}} + ... + \sum_{j\in\Theta_k}(b_j\theta_k\alpha_j^0)^{\frac{1}{1-\delta}} + ... + \sum_{j\in\Theta_q}(b_j\theta_q\alpha_j^0)^{\frac{1}{1-\delta}} \\
&= \sum_{j\in\Theta_1}(b_j\alpha_j)^{\frac{1}{1-\delta}} + \sum_{j\in\Theta_2}(b_j\alpha_j)^{\frac{1}{1-\delta}} + ... + \sum_{j\in\Theta_k}(b_j\alpha_j)^{\frac{1}{1-\delta}} + ... + \sum_{j\in\Theta_q}(b_j\alpha_j)^{\frac{1}{1-\delta}}.
\end{aligned}
$$

Substituting into (38), we have:

$$
-\frac{de}{d\theta_k} = \sum_{j\in\Theta_k} b_j\alpha_j^0 \left(\frac{(b_j\alpha_j)^{\frac{1}{1-\delta}}}{\Delta}\right)^\delta h_s^\delta\epsilon.
$$

The optimal allocation of hours across services implies

$$
h_j = \frac{(b_j\alpha_j)^{\frac{1}{1-\delta}}}{\Delta} h_s.
$$

Hence,

$$\begin{aligned}\frac{de}{d\theta_k} &= \sum_{j\in\Theta_k} \alpha_j^0 b_j h_j^\delta \epsilon \\ &= \sum_{j\in\Theta_k} \alpha_j^0 A_j.\end{aligned}$$

The composite service is total revenue from the services in Θ_k during period t, evaluated at base-period prices. The price of the composite service is θ, the percent change in prices over time.

9.1 Aggregation over services with different b's

Let $\mathcal{A}_k$ denote a group of services $A_j, j = \{1, 2, \ldots, J_{\mathcal{A}_k}\}$, within which relative prices are constant across services. Then, in any period t,

$$\alpha_{j,t} = \theta_t \alpha_{j,0} \quad \{j : A_j \in \mathcal{A}_k\},$$

from which it follows that:

$$\alpha_{j,t} = \psi_j \alpha_{1_k,t} \quad \forall t \quad \text{and} \quad j \in \{2, \ldots, J_{\mathcal{A}_k}\}, \quad \text{where} \quad \psi_j = \frac{\alpha_{j,0}}{\alpha_{1_k,0}},$$

which is constant over time.

The earnings of physician k in period t are then given by:

$$\begin{aligned}\mathcal{E}_{k,t} &= \left[\sum_k \sum_{A_j\in\mathcal{A}_k} \left(\alpha_{j,t} b_j\right)^{\frac{1}{1-\delta}}\right]^{1-\delta} h_s^\delta \epsilon \\ &= \left[\sum_k \sum_{A_j\in\mathcal{A}_k} \alpha_{1_k,t}^{\frac{1}{1-\delta}} \left(b_1{}^{\frac{1}{1-\delta}} + \sum_{j=2}^{J} \psi_j{}^{\frac{1}{1-\delta}} b_j{}^{\frac{1}{1-\delta}}\right)\right]^{1-\delta} h_s^\delta \epsilon_t \\ &= \left\{\sum_k \sum_{A_j\in\mathcal{A}_k} \theta_{k,t}^{\frac{1}{1-\delta}} \left[\left(\alpha_{1_k,0} b_1\right)^{\frac{1}{1-\delta}} + \sum_{j=2}^{J} \left(\alpha_{1_k,0}\psi_j b_j\right)^{\frac{1}{1-\delta}}\right]\right\}^{1-\delta} h_s^\delta \epsilon_t \\ &= \left\{\sum_k \sum_{A_j\in\mathcal{A}_k} \theta_{k,t}^{\frac{1}{1-\delta}} \left[\left(\alpha_{1_k,0} b_{1_k}\right)^{\frac{1}{1-\delta}} + \sum_{j=2}^{J} \left(\alpha_{j_k,0} b_{j_k}\right)^{\frac{1}{1-\delta}}\right]\right\}^{1-\delta} h_s^\delta \epsilon_t \\ &= \left[\sum_k \theta_{k,t}^{\frac{1}{1-\delta}} \tilde{b}_k^{\frac{1}{1-\delta}}\right]^{1-\delta} h_s^\delta \epsilon_t,\end{aligned}$$

where $\alpha_{j_k,0}$ denotes the price of the j^{th} service of group k in the base period 0. In the presence of heterogenous b_js within the aggregated commodity, we estimate a composite parameter

$$\tilde{b}_k = \sum_{j=2}^{J} \left(\alpha_{j_k,0} b_{j_k}\right)^{\frac{1}{1-\delta}},$$

which is constant over time because the constant bs are weighted by base-level prices through assumption 4. The aggregate service k is given by the volume of services provided within group k, weighted at base level prices $\alpha_{j_k,0}$. The price of the aggregate service $\theta_{k,t}$ is the percentage change in prices of the services in group k, relative to the base period $t = 0$.

Appendix A4: Aggregation and Variable Construction

We aggregate services through the composite-commodity theorem. Our data cover a period during which the Québec government changed the relative prices paid to physicians for the completion of medical services. To aggregate services, we considered the (geometric) average price increase of each service between the years 2000 and 2002, rounded to the nearest 5%. This provides six groups of services, whose prices increased by 0, 5, 10, 15, 20 and 25 percent.[33]

Let α_j^t be the nominal price of service j in year t, for t =1996, 1997, 1998, 1999, 2000, 2001, 2002. Since prices are constant between 1996 and 2000, we treat 2000 as the base year. We calculate θ based on the geometric average growth rate of the price of service j between $t = 2000$ and $t = 2002$. Denote this geometric average by λ, then

$$\lambda_j = Round_{0.05}\left[\left(\frac{\alpha_j^{2002}}{\alpha_j^{2000}}\right)^{0.5} - 1\right]$$

Where $Round_{.05}$ denotes the rounding operator. All services with the same λ were aggregated into the same group. If there are $m > 2$ services for with the same λ, their composite service volume – provided by physician i – is calculated as, $\sum_{j=1}^{m} \alpha_j^{2000} A_{ij}^t$, where A_{ij}^t is the number of services j performed by physician i at time t. The nominal price of this composite service is then $\theta = \lambda + 1$. We then convert nominal prices to real prices for each period, by dividing by a price index. [34]

[33] The derivation of the theorem within our context as well as the construction of the data is given in Shearer, Somé, and Fortin (2018).

[34] We use the price index of health care services: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/econ161f-eng.htm.

Appendix A5: Data

The first group of specialists, which we denote G_2, provided 2 services. It has, in turn, two subgroups. G_{12} is made up of physicians who suppled services 1 and 2. It contains specialities Endocrinology , Otorhinolaryngology, Gastroenterology, and Cardiology. G_{13} is made up of neurologists who supplied services 1 and 3. Earnings for specialist s in G_2 are calculated as

$$E_s = \alpha_1 A_{1s} + \alpha_{2'} A_{2's},$$

where $\alpha_{2'} = \mathbb{1}_{G_{12}}(s)\alpha_2 + \mathbb{1}_{G_{13}}(s)\alpha_3$ and $A_{2's} = \mathbb{1}_{G_{12}}(s)A_{2s} + \mathbb{1}_{G_{13}}(s)A_{3s}$ with $\mathbb{1}_{G_{ij}}(s) = 1$ if the specialist s belongs to the subgroup G_{ij}; 0 otherwise. A_{js} is the observed quantity of service $j = 1, 2, 3$ provided by specialist s and α_j the fee paid for service j.

For physicians providing 3 services, we have $G_3 = G_{123} \cup G_{125} \cup G_{126}$ where G_{123} , G_{125}, G_{126} are 3 disjoint subsets. G_{123} contains physicians who offered services 1, 2 and 3. It is made up of General surgeons and dermatologists .

The subgroup G_{125} contains physicians who provided services 1, 2 and 5. It is made up of pediatricians. G_{126} represents physicians who offered services 1, 2 and 6. It is made up of internal medicine physicians. Earnings for each specialist s in this case is computed as

$$E_s = \alpha_1 A_{1s} + \alpha_2 A_{2s} + \alpha_{3'} A_{3's},$$

where

$$\begin{aligned} \alpha_{3'} &= \alpha_3 \mathbb{1}_{G_{123}}(s) + \alpha_5 \mathbb{1}_{G_{125}}(s) + \alpha_6 \mathbb{1}_{G_{126}}(s) \\ A_{3's} &= A_3 \mathbb{1}_{G_{123}}(s) + A_{5s} \mathbb{1}_{G_{125}}(s) + A_{6s} \mathbb{1}_{G_{126}}(s), \end{aligned}$$

with $\mathbb{1}_{G_{12k}}(s) = 1$ if s belongs to the subgroup G_{12k} $(k = 3, 5, 6)$ and 0 otherwise; A_{js} is the observed quantity of service $j = 1, 2, 3, 5, 6$ provided by specialist s and α_j the fee of service j.

The last case we can find in data is the one in which each specialist supplies 4 services. We denote this group of physicians, G_4. It includes two separate subgroups. G_{1234} contains specialists who provided services 1, 2, 3, and 4. It contains physicians who specialize in Obstetrics and Gynecology. Physicians in the second subgroup G_{1245} provided services 1, 2, 4 and 5. In this set we find only Orthopedic surgeons. Finally, $G_4 = G_{1234} \cup G_{1245}$ and $G_{1234} \cap G_{1245} = \emptyset$. We calculate physician's earnings for this group as

$$E_s = \alpha_1 A_{1s} + \alpha_2 A_{2s} + \alpha_4 A_{4s} + \alpha_{4'} A_{4's},$$

where

$$\begin{aligned}
\alpha_{4'} &= \alpha_3 \mathbb{1}_{G_{1234}}(s) + \alpha_5 \mathbb{1}_{G_{1245}}(s) \\
A_{4's} &= A_{3s} \mathbb{1}_{G_{1234}}(s) + A_{5s} \mathbb{1}_{G_{1245}}(s),
\end{aligned}$$

with $\mathbb{1}_{G_{124k}}(s) = 1$ if s belongs to the subgroup G_{124k} $(k = 3, 5)$ and 0 otherwise; A_{js} is the observed quantity of service $j = 1, 2, 3, 4, 5$ provided by specialist s and α_j the fee of service j.

Table 9: Personal income tax structure in Québec 1996-2002

Year	Bracket Lower Bound	Federal Rate	Provincial Rate	Combined Rates
1996	0	0.17	0.16	0.33
1996	7000	0.17	0.19	0.36
1996	14000	0.17	0.21	0.38
1996	23000	0.17	0.23	0.4
1996	29590	0.26	0.23	0.49
1996	50000	0.26	0.24	0.5
1996	59180	0.29	0.24	0.53
1997	0	0.17	0.16	0.33
1997	7000	0.17	0.19	0.36
1997	14000	0.17	0.21	0.38
1997	23000	0.17	0.23	0.4
1997	29590	0.26	0.23	0.49
1997	50000	0.26	0.24	0.5
1997	59180	0.29	0.24	0.53
1998	0	0.17	0.2	0.37
1998	25000	0.17	0.23	0.4
1998	29590	0.26	0.23	0.49
1998	50000	0.26	0.26	0.52
1998	59180	0.29	0.26	0.55
1999	0	0.17	0.2	0.37
1999	25000	0.17	0.23	0.4
1999	29590	0.26	0.23	0.49
1999	50000	0.26	0.26	0.52
1999	59180	0.29	0.26	0.55
2000	0	0.17	0.19	0.36
2000	26000	0.17	0.225	0.395
2000	30004	0.25	0.225	0.475
2000	52000	0.25	0.25	0.5
2000	60009	0.29	0.25	0.54
2001	0	0.16	0.17	0.33
2001	26000	0.16	0.2125	0.3725
2001	30754	0.22	0.2125	0.4325
2001	52000	0.22	0.245	0.465
2001	61509	0.26	0.245	0.505
2001	100000	0.29	0.245	0.535
2002	0	0.16	0.16	0.32
2002	26700	0.16	0.2	0.36
2002	31677	0.22	0.2	0.42
2002	53405	0.22	0.24	0.46
2002	63354	0.26	0.24	0.5
2002	103000	0.29	0.24	0.53

Source: Milligan (2016)