11 A Switching Simultaneous Equations Model of Physician Behaviour in Ontario

Dale J. Poirier

11.1 Introduction

This study provides an econometric model of physician behavior utilizing a switching regression framework.[1] Briefly it assumes that various measures of physician behavior (e.g., hours worked, number of patients, quantity of labor employed) are determined in a simultaneous equations framework belonging to one of two regimes. Assignment of a particular physician to one of these two regimes is based on the physician's choice to "opt-in" or "opt-out" of the Ontario Health Insurance Plan (OHIP).[2] The switching equation intended to explain this discrete binary choice depends on various exogenous variables (e.g., number of dependents, school of graduation) and an additive stochastic term.

Recognition of this discrete choice aspect of the model suggests its numerous connections with the limited dependent variable, LDV, literature. The eclectic nature of our model will become obvious as an estimation scheme is devised. Because the econometric machinery needed in this estimation is substantial, it is not possible to fully discuss the underlying theory of physician behavior for which this model provides a statistical framework. (See Wolfson and Tuohy 1980 for a detailed discussion.) We instead discuss this model's econometric implications. However, we provide in this section brief overview of the behavioral model employed by Wolfson et al. (1980) which should help to motivate the econometric discussion in section 11.2.

This study describes the econometric model developed by the author for use by Wolfson and Tuohy (1980). While thanks are owed to Alan Wolfson, Carolyn Tuohy, and Stewart Iglesias for their comments, the author accepts the sole responsibility for the results reported here.

1. For surveys dealing with switching regression models, see Goldfeld and Quandt (1973, 1976), Lee and Trost (1978), Maddala and Nelson (1975) and Poirier (1976, chapter 7).

2. For readers who are not familiar with OHIP a brief description may be helpful. OHIP is a comprehensive government-sponsored health insurance plan for the residents of Ontario. It provides a wide scope of benefits for medical and hospital services plus additional benefits for the services of certain other health practitioners. The plan pays 90 percent of the Ontario Medical Association (OMA) schedule of fees for all physicians' services that are medically required. Those physicians who have opted-in submit their fee billings directly to OHIP and accept the plan's allowance as full payment. Those physicians who have opted-out usually bill their patients directly and also submit a claim to OHIP on behalf of the patient. OHIP then reimburses the patient based on 90 percent of the OMA Schedule of Fees, and the patient supplements this amount accordingly in paying the physician. Physicians who have opted-out generally charge higher fees; however, they lose the advantage of having payments guaranteed.

The major goal of the study undertaken by Wolfson and Tuohy (1980) was to obtain the understanding of physician practice behavior required for planning and policy development in the physician sector. Insofar as some of the variables that are influential determinants of practice behavior are instruments of public policy, or susceptible to policy actions, planning agencies can both forecast and effect change in physician behavior by adjusting these determining variables in an appropriate fashion.

Briefly the underlying theoretical model hypothesizes that a physician's utility is positively related to net income, the amount of leisure time, the extent to which there exists excess demand for services, the quality of care provided, the extent to which the practice can be independent (free from scrutiny and control by either government or peers), and the degree to which peer group standards are met. These variables are in turn related to various instruments under the physician's control. For example, the physician has some control over three basic classes of discretionary services: (1) services provided by the physician, (2) services provided by other parts of the system but complimentary to the physician's own services in that they generate more services for the physician such as laboratory tests, and (3) services provided by others that substitute for those of the physician such as referrals. The physician also has control over auxiliary inputs such as labor (e.g., nurses) or capital (e.g., waiting rooms) inputs. Further the physician has control over the number of hours and speed at which to work and most important whether or not to opt-out of OHIP.

As will be seen in the following sections, the physician's option decision will play a central role in the econometric model that follows. This is convenient for policy purposes because the option decision is of great concern to politicians. Politically the existence of the option has both advantages and disadvantages for government. To the extent that government incurs the dissatisfaction of voters faced with out-of-pocket costs for medical care under a government health insurance program, it is politically costly. To the extent, however, that it eases government-professional relationships, it has political advantages.

11.2 Econometric Model

The purpose of this section is to lay out a general switching simultaneous equations model. So as to simplify partly the elaborate notation that must follow, only the case of two regimes will be considered—although extensions to more than two regimes can be easily incorporated. Furthermore, while the physician behavior context of our model will be dealt with throughout, we will abstract in sections 11.2 and 11.3 from many of the pertinent problems (such as nonrandom sampling) covered later in section 11.4. This decision does not reflect their secondary importance but rather only the fact that their econometric implications will be difficult to grasp until the basic underlying model has been specified.

To begin, for a sample of size T, consider the following quantities for the tth physician. Let J_t be an observed dichotomous variable denoting sample separation into two groups or regimes: physicians who have opted-in ($J_t = 1$) and physicians who have opted-out ($J_t = 2$). Let $\mathbf{z}_t$ be a $1 \times m$ vector of fixed exogenous variables, let $\boldsymbol{\alpha}$ be a $m \times 1$ vector of unknown coefficients, and suppose

$$\begin{aligned} J_t &= 1, \quad \text{iff } \mathbf{z}_t\boldsymbol{\alpha} < u_t, \\ J_t &= 2, \quad \text{iff } \mathbf{z}_t\boldsymbol{\alpha} \geq u_t, \end{aligned} \tag{11.1}$$

where $u_t \sim$ i.i.d. $N(0, 1)$ for $t = 1, 2, \ldots, T$. Note that in terms of the binary variable

$$I_t = \begin{cases} 0, & \text{iff } J_t = 1, \\ 1, & \text{otherwise,} \end{cases} \tag{11.2}$$

specification (11.1) implies the familiar probit model

$$\begin{aligned} \text{Prob}(I_t = 0) &= 1 - \Phi(\mathbf{z}_t\boldsymbol{\alpha}), \\ \text{Prob}(I_t = 1) &= \Phi(\mathbf{z}_t\boldsymbol{\alpha}), \end{aligned} \tag{11.3}$$

where $\Phi(\cdot)$ denotes the standard normal distribution function.

Based on (11.1) define the index sets

$$\mathbf{S}_j = \{t \mid J_t = j\}, \tag{11.4}$$

$j = 1, 2$, containing T_1 and $T_2 = T - T_1$ observations, respectively, and suppose that the overall model has the following structure

$$\begin{aligned} &\text{Regime 1:} \quad \mathbf{y}_{1t}\mathbf{B}_1 + \mathbf{x}_{1t}\boldsymbol{\Gamma}_1 + \boldsymbol{\varepsilon}_{1t} = \mathbf{0}, \\ &\text{Regime 2:} \quad \mathbf{y}_{2t}\mathbf{B}_2 + \mathbf{x}_{2t}\boldsymbol{\Gamma}_2 + \boldsymbol{\varepsilon}_{2t} = \mathbf{0}, \end{aligned} \tag{11.5}$$

where for $j = 1, 2$, $\mathbf{y}_{jt}$ is a $1 \times G_j$ vector corresponding to the tth observation on the G_j endogenous variables in regime j, $\mathbf{x}_{jt}$ is a $1 \times K_j$ vector corresponding to the tth observation on the K_j exogenous variables in regime j, $\mathbf{B}_j$ and $\boldsymbol{\Gamma}_j$ are $G_j \times G_j$ and $K_j \times G_j$ matrices of unknown coefficients, with all diagonal elements of $\mathbf{B}_j$ equal to minus one, and

$$\varepsilon_{jt} \sim \text{i.i.d. } N(\mathbf{0}, \mathbf{\Sigma}_j). \tag{11.6}$$

In matrix notation (11.5) can alternatively be written as

$$\mathbf{Y}_j\mathbf{B}_j + \mathbf{X}_j\mathbf{\Gamma}_j + \boldsymbol{\varepsilon}_j = 0, \tag{11.7}$$

$j = 1, 2$, where $\mathbf{Y}_j$ is a $T_j \times G_j$ matrix with row t equal to $\mathbf{y}_{jt}$, $\mathbf{X}_j$ is a $T_j \times K_j$ matrix with row t equal to $\mathbf{x}_{jt}$, and $\boldsymbol{\varepsilon}_j$ is a $T_j \times G_j$ matrix with row t equal to $\boldsymbol{\varepsilon}_{jt}$. Imposing any exclusion restrictions, the g th equation of (11.7) can be concisely expressed as

$$\mathbf{y}_{jg} = \mathbf{Y}_{jg}\boldsymbol{\beta}_{jg} + \mathbf{X}_{jg}\boldsymbol{\gamma}_{jg} + \boldsymbol{\varepsilon}_{jg}, \tag{11.8}$$

where, for $j = 1, 2$, $\mathbf{y}_{jg}$ and $\boldsymbol{\varepsilon}_{jg}$ are the g th columns of $\mathbf{Y}_j$ and $\boldsymbol{\varepsilon}_j$, respectively, $\boldsymbol{\beta}_{jg}$ is the g th column of $\mathbf{B}_j$ omitting the g th element which has been normalized to equal minus one and any zero elements, $\mathbf{Y}_{jg}$ is a matrix consisting of the columns of $\mathbf{Y}_j$ corresponding to the elements of $\boldsymbol{\beta}_{jg}$, $\boldsymbol{\gamma}_{jg}$ is the g th column of $\mathbf{\Gamma}_j$ omitting any zero elements, and $\mathbf{X}_{jg}$ is a matrix consisting of the columns of $\mathbf{X}_j$ corresponding to the elements of $\boldsymbol{\gamma}_{jg}$.

Assuming that $\mathbf{B}_j (j = 1, 2)$ is nonsingular, the reduced form corresponding to (11.5) is

$$\text{Regime 1:} \quad \mathbf{y}_{1t} = \mathbf{x}_{1t}\mathbf{\Pi}_1 + \mathbf{v}_{1t}, \tag{11.9}$$

$$\text{Regime 2:} \quad \mathbf{y}_{2t} = \mathbf{x}_{2t}\mathbf{\Pi}_2 + \mathbf{v}_{2t}, \tag{11.10}$$

or

$$\mathbf{Y}_j = \mathbf{X}_j\mathbf{\Pi}_j + \mathbf{V}_j \tag{11.11}$$

for $j = 1, 2$, where

$$\mathbf{\Pi}_j = -\mathbf{\Gamma}_j\mathbf{B}_j^{-1}, \tag{11.12}$$

$$\mathbf{v}_{jt} = -\boldsymbol{\varepsilon}_{jt}\mathbf{B}_j^{-1}, \tag{11.13}$$

$$\mathbf{V}_j = -\boldsymbol{\varepsilon}_j\mathbf{B}_j^{-1}. \tag{11.14}$$

If for $j = 1, 2$ the joint distribution of the structural error $\boldsymbol{\varepsilon}_{jt}$ and the switching error u_t is given by[3]

$$(\boldsymbol{\varepsilon}_{jt}, u_t)' \sim \text{i.i.d.} N(\mathbf{0}, \mathbf{C}_j), \tag{11.15}$$

3. As Lee (1976) has noted, $E(\varepsilon_{1t}\varepsilon'_{2t})$ is not identifiable.

where

$$\mathbf{C}_j = \begin{bmatrix} \mathbf{\Sigma}_j & \mathbf{\Sigma}_{ju} \\ \mathbf{\Sigma}'_{ju} & 1 \end{bmatrix}, \tag{11.16}$$

then it follows from (11.13) that the joint distribution of the reduced form error $\mathbf{v}_{jt}$ and the switching error u_t is given by

$$(\mathbf{v}_{jt}, u_j)' \sim \text{i.i.d.}\, N(\mathbf{0}, \mathbf{\Omega}_j), \tag{11.17}$$

where

$$\mathbf{\Omega}_j = \begin{bmatrix} (\mathbf{B}_j^{-1})'\mathbf{\Sigma}_j\mathbf{B}_j^{-1} & (\mathbf{B}_j^{-1})'\mathbf{\Sigma}_{ju} \\ \mathbf{\Sigma}'_{ju}\mathbf{B}_j^{-1} & 1 \end{bmatrix}. \tag{11.18}$$

The model just outlined posits that there exists two regimes, describing the determination of various measures y_{jt} of a physician's behavior under option choice j, and a binary random variable J_t, indicating the physician's actual option choice. Given exogenous variables x_{1t}, x_{2t}, and z_t, it is possible in theory to observe both y_{1t} and y_{2t} as well as J_t. However, in practice only those measures of a physician's behavior corresponding to a physician's actual choice are observed: y_{1t} is observed iff $J_t = 1$, and y_{2t} is observed iff $J_t = 2$. Thus there is a sample selectivity problem in the observed data if the unobserved determinant u_t of the option decision is correlated with the unobserved determinant ε_{jt} of the physician's behavior as measured by y_{jt}. To emphasize this partial observability, we will hereafter add whenever appropriate the condition $t \in S_j$ to denote that observation t belongs to regime j.

Since identification logically precedes estimation, it is appropriate to discuss it here before the estimation of the model described by (11.1) to (11.18). Goldfeld and Quandt (1973, p. 482) have noted that identification of the model can be achieved if the structural equations in each regime are identifiable, and if each equation in any one regime satisfies the same a priori restrictions as the corresponding equation in the other regime. However, as Lee (1979, p. 989) has noted, when sample separation information is available, this latter condition is not necessary. Thus in the following sections we will assume that a sufficient number of restrictions have been imposed on $\mathbf{B}_j$ and $\mathbf{\Gamma}_j$ $(j = 1, 2)$ in each regime to insure identifiability.

11.3 Estimation

The g th reduced form equation of regime j in (11.9) or (11.10) can be written as

$$y_{jgt} = \mathbf{x}_{jt}\boldsymbol{\pi}_{jg} + v_{jgt}, \quad t \in \mathbf{S}_j, \tag{11.19}$$

where, for $j = 1, 2$, y_{jgt} and v_{jgt} are the g th elements of $\mathbf{y}_{jt}$ and $\mathbf{v}_{jt}$, respectively, and $\boldsymbol{\pi}_{jg}$ is the g th column of $\boldsymbol{\Pi}_j$. Using well-known properties of the truncated normal distribution, it is straightforward to derive the following properties for v_{jgt}:

$$E(v_{jgt} \mid J_t = j) = \omega_{jg,G_j+1}\,\delta_{jt}, \quad t \in \mathbf{S}_j, \tag{11.20}$$

$$E(v_{jgt}^2 \mid J_t = j) = \omega_{jgg} + (\omega_{jg,G_j+1})^2\,(\mathbf{z}_t\boldsymbol{\alpha})\delta_{jt}, \quad t \in \mathbf{S}_j, \tag{11.21}$$

where

$$\delta_{1t} = \phi(\mathbf{z}_t\boldsymbol{\alpha})[1 - \Phi(\mathbf{z}_t\boldsymbol{\alpha})]^{-1}, \quad t \in \mathbf{S}_1, \tag{11.22}$$

$$\delta_{2t} = -\,\phi(\mathbf{z}_t\boldsymbol{\alpha})[\Phi(\mathbf{z}_t\boldsymbol{\alpha})]^{-1}, \quad t \in \mathbf{S}_2, \tag{11.23}$$

and where ω_{jgi} is the element in the g th row and i th column of $\boldsymbol{\Omega}_j$, and $\phi(\cdot)$ denotes the standard normal density.

If $\omega_{jg,G_j+1} = 0$ (u_t and ε_{jgt} are independent), and we have enough sample observations for regime j, then, in the absence of any other complications, each regime can be estimated separately by one of the usual techniques such as two-stage least squares (2SLS). However, if $\omega_{jg,G_j+1} \neq 0$, then it can be seen from (11.20) that application of ordinary least squares, OLS, to (11.19), as is done in the first stage of 2SLS, will lead to inconsistent estimators.[4]

As a first step toward developing a consistent estimation scheme, define

$$\eta_{jgt} \equiv v_{jgt} - E(v_{jgt} \mid J_t = j), \quad t \in \mathbf{S}_j, \tag{11.24}$$

for $j = 1, 2$, and rewrite (11.19) as

$$y_{jgt} = \mathbf{x}_{jt}\boldsymbol{\pi}_{jg} + \omega_{jg,G_j+1}\,\delta_{jt} + \eta_{jgt}, \quad t \in \mathbf{S}_j. \tag{11.25}$$

Since by construction $E(\eta_{jgt} \mid J_t = j) = 0$, consistent estimators of π_{jg,G_j+1} can be obtained by applying OLS to (11.25), provided $\boldsymbol{\alpha}$ is known.

4. Note that, if (11.20) were nonzero but identical for all $t \in \mathbf{S}_j$, then only the estimator of the intercept in (11.19) would be biased. However, if the nonzero value of (11.20) varies across observations, then the estimators of all reduced form coefficients in (11.19) are biased.

The bias and inconsistency of the OLS estimator based on (11.19) arises from a specification error—namely, the exclusion of the sample selectivity regressor δ_{jt}. (The sign of this bias cannot in general be determined.) On an intuitive level the sample selectivity regression δ_{jt} is probably best thought of as an omitted variable. The sign of this omitted variable is determined solely by the observed option decision. Opted-in ($J_t = 1$) physicians have positive δ_{jt}'s and opted-out ($J_t = 2$) physicians have negative δ_{jt}'s. The magnitude (absolute value) of δ_{jt} is determined by how surprising is the observed option decision given the individual's probit score $\mathbf{z}_t\boldsymbol{\alpha}$.

While the implications for estimation of ignoring the sample selectivity regressor are fairly straightforward, it should also be pointed out that the sample selectivity regressor plays a crucial role in the interpretation of parameters in the model. To see this, suppose we wish to determine the impact, conditional on $J_t = j$, of an exogenous variable on the expected value of the g th endogenous variable in in the reduced form of regime j, that is, on

$$E(y_{jgt} \mid J_t = j) = \mathbf{x}_{jgt}\boldsymbol{\pi}_{jg} + \omega_{jg,G_j+1}\,\delta_{jt}. \tag{11.26}$$

Let the exogenous variable in question be the first regressor in $\mathbf{x}_{jgt}$, and denote it by $x_{jgt}^{(1)}$. Then

$$\frac{\partial[E(y_{jgt} \mid J_t = j)]}{\partial x_{jgt}^{(1)}} = \pi_{jg}^{(1)} + \omega_{jg,G_j+1}\,[\delta_{jt}(\delta_{jt} - \mathbf{z}_t\boldsymbol{\alpha})]\frac{\partial(\mathbf{z}_t\boldsymbol{\alpha})}{\partial x_{jgt}^{(1)}}, \tag{11.27}$$

where $\pi_{jg}^{(1)}$ is the first element in $\boldsymbol{\pi}_{jg}$, and the bracketed part of the second term is second term is $\partial\delta_{jt}/\partial x_{jgt}^{(1)}$. If $x_{jgt}^{(1)}$ is also an exogenous variable in the option decision, $[\partial(\mathbf{z}_t\boldsymbol{\alpha})/\partial x_{jgt}^{(1)}] \neq 0$, and if sample selectivity is present, $\omega_{jg,G_j+1} \neq 0$, then $\pi_{jg}^{(1)}$ cannot be interpreted as the impact of $x_{jgt}^{(1)}$ on $E(y_{jgt} \mid J_t = j)$; rather $\pi_{jg}^{(1)}$ gives the impact unconditional on the impact that works itself through the option decision.

Fortunately consistent estimators are still obtainable, even if $\boldsymbol{\alpha}$ is unknown. Heckman (1976) and Lee (1976) have shown that if $\hat{\boldsymbol{\alpha}}$ is a consistent estimator of $\boldsymbol{\alpha}$ (see section 11.4), then OLS applied to

$$\begin{aligned} y_{jgt} = {} & \mathbf{x}_{jt}\boldsymbol{\pi}_{jg} + \omega_{jg,G_j+1}\,\hat{\delta}_{jt} \\ & + [\eta_{jgt} + \omega_{jg,G_j+1}(\delta_{jt} - \hat{\delta}_{jt})], \quad t \in \mathbf{S}_j, \end{aligned} \tag{11.28}$$

where

$$\hat{\delta}_{1t} = \phi(\mathbf{z}_t\hat{\boldsymbol{\alpha}})[1 - \Phi(\mathbf{z}_t\hat{\boldsymbol{\alpha}})]^{-1}, \quad t \in \mathbf{S}_1, \tag{11.29}$$

$$\hat{\delta}_{2t} = -\phi(\mathbf{z}_t\hat{\alpha})[\Phi(\mathbf{z}_t\hat{\alpha})]^{-1}, \quad t \in \mathbf{S}_2, \tag{11.30}$$

will result in consistent estimators of π_{jg} and ω_{jg,G_j+1}. More specifically, defining the $T_j \times (K_j + 1)$ matrix,

$$W_j = [\mathbf{X}_j \mid \hat{\boldsymbol{\delta}}_j], \tag{11.31}$$

$j = 1, 2$, where

$$\hat{\boldsymbol{\delta}}_j = [\hat{\boldsymbol{\delta}}_{j1}, \hat{\boldsymbol{\delta}}_{j2}, \ldots, \hat{\boldsymbol{\delta}}_{jT_j}]'; \tag{11.32}$$

then the OLS estimator of (11.28) is given by

$$\hat{\mathbf{\Pi}}_{jg} \equiv \left[\frac{\hat{\boldsymbol{\pi}}_{jg}}{\hat{\omega}_{ig,G_j+1}}\right] = (\mathbf{W}_j'\mathbf{W}_j)^{-1}\mathbf{W}_j'\mathbf{y}_{ig}. \tag{11.33}$$

Based on the results of Lee, Maddala, and Trost (1980), it can be shown that the asymptotic distribution of $\sqrt{T_j}(\hat{\mathbf{\Pi}}_{jg} - \mathbf{\Pi}_{jg})$ is normal with mean zero and covariance matrix $T_j\mathbf{T}_j$, where

$$\begin{aligned}\mathbf{T}_j = \omega_{jgg}(\mathbf{W}_j'\mathbf{W}_j)^{-1} - \omega_{jg,G_j+1}^2 [\mathbf{A}_j\mathbf{W}_j(\mathbf{W}_j'\mathbf{W}_j)^{-1}]' \\ (\mathbf{A}_j^{-1} - \mathbf{Z}_j\mathbf{\Xi}\mathbf{Z}_j')[\mathbf{A}_j\mathbf{W}_j(\mathbf{W}_j'\mathbf{W}_j)^{-1}],\end{aligned} \tag{11.34}$$

and where $\mathbf{\Xi}$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\alpha}}$ to be defined implicitly by (11.52). In (11.34) the $T \times m$ matrix of explanatory variables in the switching equation, namely,

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_T \end{bmatrix} = \begin{bmatrix} \mathbf{Z}_1 \\ \hline \mathbf{Z}_2 \end{bmatrix}, \tag{11.35}$$

has been ordered and partitioned into two matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$, such that the first T_1 observations correspond to regime 1 and the last T_2 observations correspond to regime 2. This ordering is also utilized in defining the diagonal matrices

$$A_1 = \begin{bmatrix} \delta_{11}(\delta_{11} - \mathbf{z}_1\boldsymbol{\alpha}) & 0 & \cdots & 0 \\ 0 & \delta_{12}(\delta_{12} - \mathbf{z}_2\boldsymbol{\alpha}) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & \delta_{1T_1}(\delta_{1T_1} - \mathbf{z}_{T_1}\boldsymbol{\alpha}) \end{bmatrix}, \tag{11.36}$$

$$A_2 = \begin{bmatrix} \delta_{2,T_1+1}(\delta_{2,T_1+1} - \mathbf{z}_{T_1+1}\boldsymbol{\alpha}) & 0 & \cdots & 0 \\ 0 & \delta_{2,T_1+2}(\delta_{2,T_1+2} - \mathbf{z}_{T_1+2}\boldsymbol{\alpha}) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \delta_{2,T}(\delta_{2,T} - \mathbf{z}_T\boldsymbol{\alpha}) \end{bmatrix} \tag{11.37}$$

The unknown parameters in covariance matrix (11.34) may be estimated consistently by $\hat{\boldsymbol{\alpha}}$, $\hat{\omega}_{jg,G_j+1}$ given in (11.33), and

$$\hat{\omega}_{jgg} = T_j^{-1} \sum_{t \in S_j} [\hat{\eta}^2_{jgt} - (\hat{\omega}_{jg,G_j+1})^2(\mathbf{z}_t\hat{\boldsymbol{\alpha}} - \hat{\delta}_{jt})\hat{\delta}_{jt}], \tag{11.38}$$

where

$$\hat{\eta}_{jgt} = y_{jgt} - \mathbf{x}_{jgt}\hat{\boldsymbol{\pi}}_{jg} - \hat{\omega}_{jg,G_j+1}\hat{\delta}_{jgt}, \quad t \in S_j, \tag{11.39}$$

are the OLS residuals from (11.28).[5]

It is important to note that while (11.28) may be estimated by OLS, the usual covariance for OLS would be inappropriate, because it ignores the fact that $\hat{\boldsymbol{\alpha}}$ rather than $\boldsymbol{\alpha}$ is used in constructing the sample selectivity regressor and because, even if $\boldsymbol{\alpha}$ were known, the residuals are heteroscedastic. Ignoring this first aspect implies the incorrect covariance matrix

$$\omega_{jgg}(\mathbf{W}_j'\mathbf{W}_j)^{-1} - \omega^2_{jg,G_j+1}[\mathbf{W}_j(\mathbf{W}_j'\mathbf{W}_j)^{-1}]'\mathbf{A}_j[\mathbf{W}_j(\mathbf{W}_j'\mathbf{W}_j)^{-1}], \tag{11.40}$$

which, as Lee, Maddala, and Trost (1980) note, underestimates all the variances. Ignoring the second aspect (the heteroscedasticity) would involve additional underestimation resulting from neglect of the second term in (11.40).

To obtain consistent estimators of the parameters in the gth ($g = 1, 2, \ldots, G_j$) structural equation in regime j ($j = 1, 2$), we can proceed as follows. Reconsider (11.8) which expresses the gth structural equation

5. Since the disturbances in (11.28) are heteroscedastic, it may be desirable on efficiency grounds to perform a two-step weighted least squares, WLS, procedure to (11.28). Similarly, since all the disturbances in (11.28) depend on $\hat{\boldsymbol{\alpha}}$, they are also serially correlated. See also Lee, Maddala, and Trost (1980) on this point.

(assumed to be identified) in regime j after imposition of all zero restrictions:

$$y_{jg} = \mathbf{Y}_{jg}\boldsymbol{\beta}_{jg} + \mathbf{X}_{jg}\gamma_{jg} + \boldsymbol{\varepsilon}_{jg}. \tag{11.41}$$

In terms of the notation just introduced, $\mathbf{Y}_{jg}$ is $T_j \times (G_{jg}^{\Delta} - 1)$, $\boldsymbol{\beta}_{jg}$ is $(G_{jg}^{\Delta} - 1) \times 1$, $\mathbf{X}_{jg}$ is $T_j \times K_{jg}^*$, and γ_{jg} is $K_{jg}^* \times 1$. Analogous to (11.28), it is easy to show that

$$E(\varepsilon_{jg} \mid J_t = j, t \in S_j) = c_{jg,G_j+1}\boldsymbol{\delta}_j, \tag{11.42}$$

$j = 1, 2$, where c_{jg,G_j+1} is the element in the g th row and $(G_j + 1)$th column of $\mathbf{C}_j$. As in the case of reduced form estimation (11.42) indicates that (11.41) suffers from sample selectivity.

To correct this sample selectivity, define

$$\psi_{jg} = \varepsilon_{jg} - E(\varepsilon_{jg} \mid J_t = j, t \in \mathbf{S}_j), \tag{11.43}$$

and rewrite (11.41) as

$$y_{jg} = Y_{jg}\beta_{jg} + X_{jg}\gamma_{jg} + c_{jg,G_j+1}\hat{\delta}_j + [\psi_{jg} + c_{jg,G_j+1}(\delta_j - \hat{\delta}_j)]. \tag{11.44}$$

While the inclusion of $\hat{\delta}_j$ in (11.44) removes the sample selectivity present in (11.41), (11.44) still suffers from simultaneous equations bias. However, this simultaneous equations bias in (11.44) can be handled in the usual fashion by two-stage least squares provided that instruments for the endogenous variables $\mathbf{Y}_{jg}$ are obtained from selectivity-free reduced form equation (11.28) rather than (11.19). In words, application of conventional 2SLS to the original structural equations augmented by the appropriate sample selectivity regressors will provide consistent estimators of the structural parameters. Alternatively the entire procedure may be viewed as 2SLS applied to an augmented system in which the estimated sample selectivity regressor appears in each structural equation. For further discussion, see Lee, Maddala, and Trost (1980).

From a computational standpoint, 2SLS applied to (11.54) can be equivalently viewed as OLS applied to

$$\begin{aligned} \mathbf{y}_{jg} = \hat{\mathbf{Y}}_{jg}\boldsymbol{\beta}_{jg} + \mathbf{X}_{jg}\gamma_{jg} + c_{jg,G_j+1}\hat{\boldsymbol{\delta}}_j + [\psi_{jg} + c_{jg,G_j+1}(\boldsymbol{\delta}_j - \hat{\boldsymbol{\delta}}_j) \\ + (\mathbf{Y}_{jg} - \mathbf{Y}_{jg})\boldsymbol{\beta}_{jg}], \end{aligned} \tag{11.45}$$

where $\hat{\mathbf{Y}}_{jg}$ is a $T_j \times (G_{jg}^{\Delta} - 1)$ matrix of predicted values of included endogenous variables obtained from reduced form (11.28). Defining the $T_j \times (G_{jg}^{\Delta} + K_{jg}^{*})$ matrix

$$\mathbf{D}_{jg} = [\hat{\mathbf{Y}}_{jg} \mid \mathbf{X}_{jg} \mid \hat{\boldsymbol{\delta}}_j], \tag{11.46}$$

the OLS estimator of (11.45) is defined by

$$\hat{\boldsymbol{\lambda}}_{jg} \equiv \begin{bmatrix} \hat{\boldsymbol{\beta}}_{jg} \\ \hat{\boldsymbol{\gamma}}_{jg} \\ \hline \hat{c}_{jg,G_j+1} \end{bmatrix} = (\mathbf{D}'_{jg}\mathbf{D}_{jg})^{-1}\mathbf{D}'_{jg}\mathbf{y}_{jg}. \tag{11.47}$$

Based on arguments similar to those employed previously, it can be shown that the asymptotic distribution of $T_j^{1/2}(\hat{\boldsymbol{\lambda}}_{jg} - \boldsymbol{\lambda}_{jg})$ is normal with mean zero and covariance matrix $T_j\boldsymbol{\Lambda}_{jg}$, where

$$\boldsymbol{\Lambda}_{jg} = c_{jgg}(\mathbf{D}'_{jg}\mathbf{D}_{jg})^{-1} - c_{jg,G_j+1}[\mathbf{A}_j\mathbf{D}_{jg}(\mathbf{D}'_{jg}\mathbf{D}_{jg})^{-1}]' \\ (\mathbf{A}_j^{-1} - \mathbf{Z}_j\boldsymbol{\Xi}\mathbf{Z}'_j)[\mathbf{A}_j\mathbf{D}_{jg}(\mathbf{D}'_{jg}\mathbf{D}_{jg})^{-1}]. \tag{11.48}$$

As in the case of estimating the reduced form, the usual OLS procedure for constructing the covariance matrix will neglect the second term (11.48). A consistent estimator of the disturbance variance in structural equation g of regime j is given by

$$\hat{c}_{jgg} = T_j^{-1} \sum_{t \in \mathbf{S}_j} [\hat{\psi}_{jgt}^2 - (\hat{c}_{jg,G_{jg}+1}^2)\hat{\delta}_{jt}(\mathbf{z}_t\hat{\boldsymbol{\alpha}} - \hat{\delta}_{jt})]. \tag{11.49}$$

11.4 Estimation of the Switching (Option) Equation

The assumption in (11.1) that $u_t \sim N(0, 1)$ for $t = 1, 2, \ldots, T$ implies that a physician's decision to opt-in or opt-out can be described by the familiar probit model. Associated with this decision is the observed price charged by the physician. This price variable can be described as follows.

Consider two physicians performing an identical service, and suppose one has opted-in and the other has opted-out. The opted-in physician will bill the patient according to OHIP reimbursements which specify a rate equaling 0.9 times the Ontario Medical Association, OMA, scheduled fee for the service in question. The opted-out physician will bill at a rate in excess of 0.9 times the OMA fee. Thus the probability density for the relative price of this typical service will have a spike at 0.9 for opted-in physicians and a continuous positive segment over the range greater than

0.9 for opted-out physicians. This density is of course reminiscent of the classic model of Tobin (1958).

It would, however, seem overly restrictive to assume that the option and pricing decisions can be treated identically.[6] Thus as numerous authors (e.g., Cragg 1971) have suggested, we will treat them separately. However, we will not require the error terms in the option equation and the price equation to be statistically independent. In fact we expect them to be highly correlated, because both likely capture similar attitudinal characteristics.

Another connection with the limited dependent variable literature lies not with the model's structure but rather with the data-gathering process employed. In the qualitative choice literature exogenous sampling is distinguished from choice-based or, more appropriately here, regime-based sampling. In the present context exogenous sampling is one in which physicians are drawn and their choice behavior (e.g., option-decision) is observed. In contrast regime-based sampling is one in which a preassigned number of opted-in physicians and a preassigned number of opted-out physicians are selected, and their behavioral characteristics are observed. The statistical relevance of regime-based sampling is that it renders the usual ML estimator inconsistent. As an alternative to ML, Manski and Lerman (1977) and Manski and McFadden, chapter 1, have suggested a computationally simple weighted exogenous sampling maximum likelihood, WESML, estimator. In terms of the marginal likelihood for the option equation, the WESML of $\boldsymbol{\alpha}$ may be obtained by maximizing

$$L^*(\alpha) = \sum_{t=1}^{T} \xi_1 I_t \ln[\Phi(-\mathbf{z}_t\boldsymbol{\alpha})] + \xi_2(1 - I_t)\ln[\Phi(\mathbf{z}_t\boldsymbol{\alpha})]. \tag{11.50}$$

The weights in (11.50) are given by

$$\xi_j = \frac{Q(j)}{H(j)}, \tag{11.51}$$

$j = 1, 2$, where $Q(j)$ is the fraction of physicians in the population selecting option j and $H(j)$ is the analogous fraction for the regime-based sample.

6. If they were treated identically, then the switching equation would be analogous to the model considered by Tobin (1958). Lee, Maddala, and Trost (1980) have considered such a switching simultaneous equation in the case of random sampling, and they have outlined a two-step estimation procedure similar to that discussed in the text. Asymptotic covariance matrix (11.48) is, however, inappropriate, and it is no longer true that the usual 2SLS covariance matrix is biased downward.

The WESML estimator $\hat{\boldsymbol{\alpha}}$ is consistent and asymptotically normal; however, in general it is not asymptotically efficient. The asymptotic covariance matrix of $T^{1/2}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ is given by $T\Xi$, where Ξ can be consistently estimated by

$$\hat{\Xi} = \left[\sum_{t=1}^{T} \left\{\hat{\phi}_t(\mathbf{z}_t\hat{\boldsymbol{\alpha}})(\xi_1 - \xi_2) - \frac{\hat{\phi}_t^2(\xi_1(1-\hat{\Phi}_t) + \xi_2\hat{\Phi}_t)}{\hat{\Phi}_t(1-\hat{\Phi}_t)}\right\}\mathbf{z}_t'\mathbf{z}_t\right]^{-1}$$
$$\cdot\left[\sum_{t=1}^{T} \left\{\frac{\hat{\phi}_t^2(\xi_1^2(1-\hat{\Phi}_t) + \xi_2^2\hat{\Phi}_t)}{\hat{\Phi}_t(1-\hat{\Phi}_t)}\right\}\mathbf{z}_t'\mathbf{z}_t\right]$$
$$\cdot\left[\sum_{t=1}^{T} \left\{\hat{\phi}_t(\mathbf{z}_t\hat{\boldsymbol{\alpha}})(\xi_1 - \xi_2) - \frac{\hat{\phi}_t^2(\xi_1(1-\hat{\Phi}_t) + \xi_2\hat{\Phi}_t)}{\hat{\Phi}_t(1-\hat{\Phi}_t)}\right\}\mathbf{z}_t'\mathbf{z}_t\right]^{-1}, \quad (11.52)$$

where $\hat{\Phi}_t = \Phi(\mathbf{z}_t\hat{\boldsymbol{\alpha}})$ and $\hat{\phi}_t = \phi(\mathbf{z}_t\hat{\boldsymbol{\alpha}})$ for $t = 1, 2, \ldots, T$.

In the model specification considered here, the switching or option equation drives the model in the sense that the two regimes depend in a recursive fashion on the option equation. While we have permitted correlation between the disturbance term in the option equation and the disturbance terms in the structural equations of each regime, we have omitted any of the endogenous variables in each regime from appearing in the option equation itself. Although this specification seems reasonable given the stickiness and infrequency of the option decision, more general specifications may be employed. See Lee (1979) for a discussion.

11.5 Empirical Results

In Wolfson and Tuohy (1980) the model presently under discussion is estimated, using a survey sample of 309 Ontario physicians. Given the size of the model—a total of 26 structural equations involving 549 coefficients in addition to the option equation—the present discussion will involve only a few selected issues. Specifically section 11.6 will provide a brief discussion of the option equation, and section 11.7 will provide a similarly brief discussion of one particular structural equation appearing in both regimes, namely, the equation describing the extent to which the physician generates medical services by referring patients to other physicians (measured in terms of the dollar value of the referred services). However, before proceeding on to a discussion of these equations, a brief overview of the entire model will be given.

In the sample of $T = 309$ physicians, 82 physicians (0.2654 percent) opted-out of OHIP. The opted-out regime for these $T_2 = 82$ physicians consists of $G_2 = 9$ structural equations involving $K_2 = 43$ exogenous variables. The nine endogenous variables are (1) the total OHIP billings for the physician during the period May 1975 to January 1976, (2) the total discrete patient load for the physician during the five-month period September 1975 to January 1976, (3) the average number of days between the time a patient requests services and the time of an encounter with the physician, (4) the total hours worked per week by the physician on the provision of patient care (exclusive of teaching, administration, and research), (5) the salaries paid to office employees, (6) the total annual office expenses less salaries, (7) the dollar value of laboratory investigations and radiological tests and ordered from other physicians during the period May 1975 to January 1976, (8) the dollar value of medical services arising from referral of patients to other physicians during the period May 1975 to January 1976, and (9) a price index reflecting the average price charged by the physician relative to the price charged by an opted-in physician. In all cases these structural equations are extremely overidentified—the average degree of overidentification being approximately twenty.

For the remaining $T_1 = 227$ physicians, the opted-in regime is divided into two subregimes: one for the 128 general practitioners and one for the 99 specialists contained in the opted-in sample.[7] Both of the subregimes contain a structural equation for each of the first eight endogenous variables considered in the opted-out regime. In addition the subregime for opted-in general practitioners contains an additional structural equation describing the extent to which each general practitioner's patients use the services of other physicians on a nonreferral basis. The subregime for general practitioners involves 33 exogenous variables and the subregime for specialists involves the same 33 exogenous variables plus 6 specialist dummies. Hence there are a total of $G_1 = 17$ endogenous variables and $K_1 = 72$ exogenous variables in the overall opted-in regime. As in the opted-out regime all equations in the opted-in regime are highly overidentified.

7. The small sample size precluded a similar division for opted-out physicians.

11.6 Estimated Option Equation

In the theoretical model developed in Wolfson and Tuohy (1980) the option equation plays a central role. The overall model is predicated on the belief that physicians who opt-out of OHIP are inherently different in terms of practice behavior from those who opt-in. The option decision is viewed as disjoint from the rest of the model in the sense that it does not depend on contemporaneous characteristics of the physician's practice. Rather it depends on basic characteristics of the individual and the community in which he practices. It is not, however, fully recursive from the rest of the model because the unobserved elements that help determine the option decision are allowed to be correlated with unobserved components of the various other endogenous variables in the model.

Since the overall population proportion of Ontario physicians who opt-out is comparatively small, opted-out physicians were oversampled to guarantee that they would appear in sufficient numbers in the final sample. Specifically in terms of (11.51) the appropriate weights are

$$\xi_1 = 0.4435 \quad \text{and} \quad \xi_2 = 1.201. \tag{11.53}$$

In the results that follow both weighted and unweighted maximum likelihood estimates will be presented.

The central role of the option decision arises from a methodological position that underscores the sociopolitical factors created by government intrusion into medical care. The existence of the opting-out provision has had a major symbolic and strategic importance for organized medicine in Ontario. It symbolizes the independence of physicians, both individually and as a self-governing profession, from government, and it also enhances the bargaining power of organized medicine in negotiating the schedule of payments to physicians under the government plan.

In general the local medical community, which in many areas is coincident with the hospital medical staff association, is politically, socially, and economically cohesive. This may not be so much a matter of the exercise of political, social, and economic sanctions (although certainly these sanctions exist in the form of the granting of hospital privileges, the referral of patients, the scheduling of operating room time, and social amenities) as it is a matter of the existence of a community of shared expertise and experience within which political and economic behavior is shaped.[8]

As can be seen from table 11.1, the option decision depends on $m = 26$ variables $z_1, z_2, \ldots, z_{26}$. These twenty-six variables can be conveniently broken down into eight broad groups: family characteristics, community characteristics, teaching duties, foreign background, political orientation, attitudes toward government, specialty effects, and a catch-all miscellaneous classification. Leaving more detailed definitions to Wolfson and Tuohy (1980), it will simply be noted here that the descriptions in table 11.1 are for the most part self-explanatory. Variables z_2, z_3, z_7 through z_{22}, and z_{24} are all dummies defined in an obvious fashion. Throughout the subsequent analysis one, two, and three asterisks will be used to denote asymptotic significance at the 10, 5, and 1 percent levels, respectively.

Overall the results contained in table 11.1 are quite satisfactory. All coefficients have the sign expected a priori, and as can be seen from table 11.2, the joint effects of groups of variables are for the most part highly significant. Leaving a detailed discussion to Wolfson and Tuohy (1980), we will dwell on only one economic aspect here, namely, the crucial role played by z_6 in the option decision. The results strongly suggest that the percent of opted-out physicians in the same specialty in the community is a major determinant in a physician's own option decision. In fact this emulation of peers is so strong that it swamps most other community and specialty effects. While attitudinal characteristics do play a role, the impact of the local market for the physician's services, as reflected in z_6, is most interesting.

The most interesting econometric aspect of the estimated option equation is the effect of weighting to correct for regime-based sampling. The estimated standard errors of coefficients from WESML are uniformly smaller than their unweighted counterparts. As a result the effects of variables tend to look more significant when considering the weighted as opposed to unweighted probit estimates. While of course the inconsistency of unweighted probit estimates makes any inference based on unweighted

8. Economically this local cohesiveness facilitates the establishment and maintenance of common price structures for opted-out physicians. In most professional markets, including medicine, professionals have treated the practitioner who charges prices lower than those of his peers as suspect of offering services of a quality lower than the community standard as well. Price-cutters in medicine may find themselves in conflict with local agencies of peer review, such as hospital staff committees, who have the effective ability to award, refuse, or curtail hospital privileges.

Table 11.1
Switching (option) equation

Number	Variable descriptions	Weighted probit: Estimated coefficient	Weighted probit: Estimated Standard deviation	Unweighted probit: Estimated coefficient	Unweighted probit: Estimated standard deviation
	Family characteristics				
1	number of dependents	0.08422*	0.05070	0.08794	0.06029
2	spouse employed	0.05768	0.2149	0.1012	0.2417
	Community characteristics				
3	urban	0.8568***	0.2947	0.9793**	0.4566
4	average income per capita	0.1923	0.8731	0.2821	1.100
5	physicians per capita	−0.1983	0.2252	−0.1400	0.2689
6	percent opted-out in the physicians own speciality	0.0358***	0.007447	0.03775***	0.007530
	Teaching duties				
7	part-time	0.5646***	0.2160	0.6380***	0.2456
8	full-time	−0.2692	0.4209	−0.1458	0.4472
	Foreign background				
9	born in England	0.4601	0.3238	0.4689	0.3891
10	born outside of Canada or England	−0.1852	0.2731	−0.1964	0.3318
11	Ontario medical education	0.4940*	0.2685	0.5308*	0.3136
	Political orientation				
12	liberal	−0.3983*	0.2190	−0.4155*	0.2517
13	conservative	0.3612	0.2286	0.3893	0.2692

Table 11.1
(continued)

Number	Variable descriptions	Weighted probit: Estimated coefficient	Weighted probit: Estimated Standard deviation	Unweighted probit: Estimated coefficient	Unweighted probit: Estimated standard deviation
	Attitude toward government				
14	more power among physicians	−0.02671	0.2474	−0.02823	0.2834
15	less power in the medical review committee	0.6619***	0.2325	0.7431***	0.2781
	Specialties				
16	family practice	0.2009	0.2782	0.2079	0.3943
17	obstetrics and gynecology	0.5502	0.5072	0.3868	0.5847
18	pediatrics	−0.3473	0.4777	−0.3756	0.6296
19	anaesthesia	−0.6880*	0.3930	−0.5737	0.4755
20	psychiatry	0.1872	0.6305	0.2908	0.6938
21	surgery	0.5684	0.4531	0.4630	0.5572
22	medical specialties	1.029*	0.6023	0.8461	0.7340
	Miscellaneous				
23	years of practice in present locality	0.01656*	0.009207	0.01880	0.01211
24	member of a medical advisory committee	0.04963	0.1820	0.05657	0.2215
25	billings of physician's specialty relative to the billings of all physicians	−3.973**	1.829	−3.355	2.107
26	constant term	−0.3745	1.798	−0.7800	2.128

Note: * Significant at 10 percent level; ** significant at 5 percent level; *** significant at 1 percent level.

Table 11.2
Tests of joint effects

Test description	Degrees of freedom	Weighted probit	Unweighted probit
Family characteristics			
$\alpha_1 = \alpha_2 = 0$	2	2.773*	2.216
Community characteristics			
$\alpha_3 = \alpha_4 = 0$ (demographic)	2	10.61***	5.700**
$\alpha_5 = \alpha_6 = 0$ (medical)	2	24.19***	25.13***
$\alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$ (total)	4	38.70***	33.79***
Teaching duties			
$\alpha_7 = \alpha_8 = 0$	2	7.992**	7.604**
Foreign background			
$\alpha_9 = \alpha_{10} = 0$ (birth)	2	4.373	2.893
$\alpha_9 = \alpha_{10} = \alpha_{11} = 0$ (total)	3	7.327*	5.379
Political orientation			
$\alpha_{12} = \alpha_{13} = 0$	2	7.566**	6.401**
Attitude toward government			
$\alpha_{14} = \alpha_{15} = 0$	2	8.333**	7.227**
Specialties			
$\alpha_{17} = \alpha_{18} \ldots = \alpha_{22} = 0$ (non-GP's)	6	11.23*	7.300
$\alpha_{16} = \alpha_{17} = \ldots = \alpha_{22} = 0$ (total)	7	11.89	7.500
Total			
$\alpha_1 = \alpha_2 = \ldots = \alpha_{26} = 0$	26	85.04***	85.57***

Note: *Significant at 10 percent level; ** significant at 5 percent level; ***significant at 1 percent level.

estimates tenuous at best, it is interesting to note that, unless one was overly attached to magical significance levels, then one would be led to basically the same economic conclusions using the unweighted as opposed to the weighted estimates.[9]

Table 11.3 suggests, however, that predictions of opting-out probabilities do differ substantially between the weighted and unweighted estimates. Not surprisingly, the unweighted estimates, which ignore the oversampling of opted-out physicians, yield higher predicted probabilities of opting-out than the weighted estimates. Based on a critical predicted probability of 0.5, the weighted estimates predict observed options slightly more accurately than the unweighted estimates.

Finally, the sample selectivity regressors implied by the weighted and unweighted coefficient estimates have a correlation of 0.9755 and fairly similar partial correlations with respect to other variables. Thus whether weighted or unweighted sample selectivity regressors are used in the analysis of either regime is of little practical consequence.

11.7 Estimated Referral Equation

In any publicly financed medical scheme an important consideration for policy purposes is the degree to which the medical profession generates medical services from within through referrals for consultations with other physicians. Specifically, it is of interest to determine whether various characteristics of the physician, his community, and his practice are systematically related to the referrals he generates. In addition there also exists the question of whether these systematic determinants differ in their impact on opted-in and opted-out physicians.

To assess the determinants of physician referrals the following equations were estimated. The equations in both regimes share numerous mutual determinants. As in the option equation family and community character-

9. The marginal effect of z_{ti} on the probability of the tth physician opting-out is given by

$$\frac{\partial[\text{Prob}(I_t = 1)]}{\partial z_{ti}} = \phi(\mathbf{z}_t\boldsymbol{\alpha})\alpha_i.$$

This effect may be estimated for the average physician by $\phi(\bar{\mathbf{z}}\hat{\boldsymbol{\alpha}}) \cdot \hat{\alpha}_i$. From table 11.3 it is seen that $\phi(\bar{\mathbf{z}}\hat{\boldsymbol{\alpha}}) = 0.1383$ in the case of the weighted estimates, and $\phi(\bar{\mathbf{z}}\hat{\boldsymbol{\alpha}}) = 0.2521$ in the case of unweighted estimates. As a result these marginal effects tend to be overestimated in the case of unweighted probit, since the unweighted estimates of the α_i's tend to be larger in absolute value.

Table 11.3
Descriptive statistics

Statistic	Weighted probit	Unweighted probit
Proportion of successful predictions	0.8317	0.8285
Weighted proportion of successful predictions	0.5854	0.6199
Population proportion opting-out	0.1177	0.1177
Predicted probability of opting-out at mean index	0.07272	0.1690
Mean predicted probability of opting-out	0.1736	0.2658
Density evaluated at mean index	0.1383	0.2521

istics, teaching duties, political orientation, and specialty variables reappear. The only new variables included in these groups are the number of hospitals at which the physician has privileges and the average referral rate of the physician's peer group. In addition three other groups of variables are employed: three endogenous variables, three attitudinal variables, and two practice characteristics.[10]

Table 11.4 contains the results of estimating the referral equation for opted-in general practitioners. (The results for opted-in specialists have been omitted for the sake of brevity.) Both OLS and 2SLS estimates are provided. In the case of 2SLS two estimated standard deviations are given: one is uncorrected for heteroscedasticity and the estimated nature of the

10. The three attitudinal variables are dummy variables referring to the physician's practice behavior. The first equals one if the physician takes into account the cost of treatment, besides the patients medical condition, when recommending treatment. Similarly the second equals unity if the physician takes into account the patient's wishes when recommending treatment. The third equal unity if the physician feels harassed by problem patients. The first practice characteristic variable is a dummy variable equaling unity if the physician belongs to a multispecialty group. The second is a scaled variable so that larger values indicate that the physican provides, relatively speaking, more complex services. In other words, a large value indicates that the physician's practice is concentrated in services that are themselves concentrated among relatively few physicians.

Table 11.4
Estimated referral equation for opted-in general practitioners

Number	Variable descriptions	2SLS Coefficient	2SLS Standard deviation (uncorrected)	2SLS Standard deviation (corrected)	OLS Coefficient	OLS Standard deviation
	Endogenous variables					
1	patient load	3.485***	0.7342	0.8204**	2.669***	0.3321
2	labor employed	0.5312	0.09448	0.9150	0.09997**	0.0423
3	overhead	−0.02303	0.07798	0.07489	−0.04109	0.0250
	Family characteristics					
4	number of dependents	−68.05	157.1	160.8	−34.23	129.7
5	spouse employed	654.5	570.5	688.9	682.7	479.6
	Community characteristics					
6	urban	513.4	627.9	2579.0	565.0	533.5
7	average income per capita	−0.3419	0.2465	0.6024	−0.2872	0.2060
8	physicians per capita	−2.065	674.4	907.5	−143.2	568.1
9	hospital privileges	198.6	244.8	237.9	193.7	205.6
10	average referrals of others in the physicians own speciality	9.208***	2.956	3.463**	9.534***	2.417
	Teaching duties					
11	part-time	−510.3	1056.0	2561.0	−606.8	861.0
	Political orientation					
12	liberal	442.2	571.0	1062.0	237.6	472.2
13	conservative	22.01	942.7	1017.0	−30.66	754.6

Table 11.4
(continued)

		2SLS			OLS	
Number	Variable descriptions	Coefficient	Standard deviation (uncorrected)	Standard deviation (corrected)	Coefficient	Standard deviation
	Practice attitudes toward					
14	cost of treatment	170.9	531.3	970.4	320.5	434.3
15	patients' wishes	−399.9	530.8	683.2	−504.0	437.5
16	problem patients	656.1	569.7	605.6	492.5	459.4
	Practice characteristics					
17	member of a multispecialty group practice	−755.5	971.5	1320.0	−1174.0	770.2
18	complexity	−2815.0*	1678.0	2333.0	−1700.0	1255.0
	Specialty					
19	family practice	−2129.0***	783.4	908.0**	−1792.0***	625.3
	Miscellaneous					
20	sample selectivity regressor	5391.0	3416.0	47077.0	4703.0*	2842.0
21	constant term	−357.3	2231.0	7843.0	−509.4	1893.7

Note: * Significant at 10 percent level; ** significant at 5 percent level; *** significant at 1 percent level.

sample selectivity regressor, and the other is corrected for both. Joint hypothesis tests based on Wald's method are given in table 11.5. Asterisks on the corrected standard deviations indicate the significance of the coefficients when the corrected standard deviations are used.

Basically the results indicate that the principle determinants of referrals for an opted-in general practitioner are (not surprisingly) the patient load, the referral behavior of peers, and whether the physician has a certificate from the College of Family Physicians. The overall fit of the equation is indicated by the 2SLS $R^2 = 0.4255$.

Surprisingly there is little evidence of sample selectivity. Given the corrected variance estimator $\hat{c}_{1gg} = 58913.0$, it is seen that the estimated correlation between u_t and ε_{1gt} is $5391/(58913.0)^{1/2} = 2.221$, which exceeds unity by a substantial margin.[11] Hence the point estimate for $\hat{c}_{1g,G_1+1}$ is both implausible and insignificant.

Tables 11.6 and 11.7 contain estimates analogous to tables 11.4 and 11.5, but for opted-out physicians. The results indicate that the principle determinants of referrals for an opted-out physician are family characteristics, the referral behavior of his peers, and his political orientation. The strong negative effect of the number of dependents is consistent with the hypothesis that an opted-out physician with family dependents will be hesitant to make referrals and run the risk of losing patients. The nature of the employed spouse effect is somewhat unclear and does not correspond to the a priori expected sign. The negative political orientation effects indicate that opted-out physicians who consider themselves to be more liberal or more conservative than the Ontario Medical Association are also less likely to make referrals. For opted-out physicians $R^2 = 0.6922$.

Once again there is little evidence of sample selectivity. Given the corrected variance estimator $\hat{c}_{2gg} = 23274.5$, the estimated correlation between u_t and ε_{2gt} is $-35.07/(23274.5)^{1/2} = -0.02299$. Taking into account the estimated nature of the sample selectivity regressor has a negligible effect on the estimated standard deviations.

Formal testing of identical behavior across the regimes is of little consequence here since the regimes were specified to be different on a priori

11. The fact that the estimated correlation exceeds unity explains why the corrected standard deviations in table 11.4 are sometimes smaller than the uncorrected standard deviations.

Table 11.5
Test of joint effects in referral equation (opted-in general practitioners)

Test description	Degrees of freedom	2SLS (corrected)	2SLS (uncorrected)	OLS
Endogenous variables				
$\beta_{1g1} = \beta_{1g2} = \beta_{1g3} = 0$	3	23.48***	30.72***	91.35***
Family characteristics				
$\gamma_{1g1} = \gamma_{1g2} = 0$	2	0.9851	1.578	2.140
Community characteristics				
$\gamma_{1g3} = \gamma_{1g4} = 0$ (demographic)	2	0.8214	2.335	2.717
$\gamma_{1g5} = \gamma_{1g6} = \gamma_{1g7} = 0$ (medical)	3	11.84***	12.67***	19.25***
$\gamma_{1g2} = \ldots = \gamma_{1g7} = 0$ (total)	5	14.96***	16.53***	27.28***
Political orientation				
$\gamma_{1g9} = \gamma_{1g,10} = 0$	2	0.1880	0.6013	0.2594
Practice attitudes				
$\gamma_{1g,11} = \gamma_{1g,12} = \gamma_{1g,13} = 0$	3	1.969	1.961	2.430
Practice characteristics				
$\gamma_{1g,14} = \gamma_{1g,15} = 0$	2	5.400*	4.936*	5.331*
Total				
$\beta_{1g1} = \ldots = \beta_{1g3} = \gamma_{1g1} = \ldots = \gamma_{1g1,17} = 0$	20	88.87***	94.80***	176.1***

Note: * Significant at 10 percent level; ** significant at 5 percent level; *** significant at 1 percent level.

Table 11.6
Estimated referral equation for opted-out physicians

Number	Variable descriptions	2SLS Coefficient	2SLS Standard deviation (uncorrected)	2SLS Standard deviation (corrected)	OLS Coefficient	OLS Standard deviation
	Endogenous variables					
1	patient load	0.1393	0.7337	0.7340	0.1836	0.4553
2	labor employed	0.05294	0.06784	0.06791	0.1322***	0.03123
3	overhead	0.03548	0.03817	0.03820	−0.004030	0.01785
	Family characteristics					
4	number of dependents	−319.4***	122.3	122.6***	−276.7***	104.0
5	spouse employed	−862.9*	482.6	483.6*	−1005.0**	413.4
	Community characteristics					
6	urban	1068.0	1372.0	1382.0	1099.1	1197.0
7	average income per capita	0.1841	0.2728	0.3147	0.09507	0.2171
8	physicians per capita	−837.4	603.1	603.6	−597.0	500.9
9	average referrals of others in the physicians own specialty	0.7752***	0.1530	0.1532**	0.7298***	0.1255
10	percent opted-out in physicians own specialty	−31.88	20.71	20.89	−25.51**	17.62
	Teaching duties					
11	part-time	−759.7	502.3	504.4	−664.0	422.6
12	full-time	−1679.0*	992.6	993.9	−1168.0	789.0

Table 11.6
(continued)

Number	Variable descriptions	2SLS Coefficient	2SLS Standard deviation (uncorrected)	2SLS Standard deviation (corrected)	OLS Coefficient	OLS Standard deviation
	Political orientation					
13	liberal	−1391.0**	502.3	572.7**	−1257.0**	494.1
14	conservative	−1393.0***	501.9	502.5***	−1311.0***	436.1
	Practice attitudes toward:					
15	cost of treatment	53.28	504.7	505.2	−203.0	406.4
16	patients' wishes	355.4	581.5	581.8	492.9	499.5
17	problem patients	−464.4	461.4	461.8	−551.4	393.2
	Practice characteristics					
18	member of multispecialty group practice	−283.3	1088.0	1088.0	−166.5	953.0
19	complexity	−1950.0	1220.0	1221.0	−2185.0	1000.0
	Specialties					
20	family practice	1609.0	979.3	983.8	1452.0*	855.2
21	obstrics and gynecology	595.0	981.4	983.7	412.4	842.8
22	pediatrics	−154.9	1369.0	1370.0	174.0	1174.0
23	anaesthesia	−78.06	1495.0	1496.0	573.9	1228.0
24	psychiatry	−1275.0	1054.0	1057.0	−588.9	831.1
25	surgery	−240.6	1100.0	1101.0	−87.93	928.9
26	medical specialties	−984.1	1029.0	1030.0	−538.1	863.6
	Miscellaneous					
27	sample selectivity regressor	−35.07	706.9	1023.0	136.0	615.4
28	constant term	3835.0	3265.0	3792.0	388.6	2728.0

Note: * Significant at 10 percent level; ** significant at 5 percent level; *** significant at 1 percent level.

Table 11.7
Test of joint effects in referral equation for opted-out physicians

Test description	Degrees of freedom	2SLS (corrected)	2SLS (uncorrected)	OLS
Endogenous variables				
$\beta_{2g1} = \beta_{2g2} = \beta_{2g3} = 0$	3	5.117	5.120	30.94***
Family characteristics				
$\gamma_{2g1} = \gamma_{2g2} = 0$	2	9.518***	9.578***	11.87***
Community characteristics				
$\gamma_{2g3} = \gamma_{2g4} = 0$ (demographic)	2	1.278	1.707	1.564
$\gamma_{2g5} = \gamma_{2g6} = \gamma_{2g7} = 0$ (medical)	3	29.88***	30.03***	38.97***
$\gamma_{2g3} = \ldots = \gamma_{2g7} = 0$ (total)	5	32.25***	32.64***	41.29***
Teaching duties				
$\gamma_{2g8} = \gamma_{2g9} = 0$	2			
Political orientation				
$\gamma_{2g,10} = \gamma_{2g,11} = 0$	2	11.14***	11.18***	12.86***
Practice attitudes				
$\gamma_{2g,12} = \gamma_{2g,13} = \gamma_{2g,14} = 0$	3	1.397	1.398	2.633
Practice characteristics				
$\gamma_{2g,15} = \gamma_{2g,16} = 0$	2	2.552	2.554	4.839
Specialties				
$\gamma_{2g,18} = \ldots = \gamma_{2g,23} = 0$	6	3.949	3.957	2.685
$\gamma_{2g,17} = \ldots = \gamma_{2g,23} = 0$	7	8.668	8.736	8.274
Total				
$\beta_{2g1} = \ldots = \beta_{2g3} = \gamma_{2g1} = \ldots = \gamma_{2g,24} = 0$	27	167.0***	184.4***	263.4***

Note: * Significant at 10 percent level; ** significant at 5 percent level; *** significant at 1 percent level.

grounds.[12] However, one similarity stands out, namely, the predominant influence of referral practices of the physician's peer group. Such peer group effects are not isolated to the referral equation alone but rather form one of the basic founding blocks upon which the model of Wolfson and Tuohy (1980) is built.

11.8 Concluding Remarks

The development of the econometric model of physician behavior reported here and in Wolfson and Tuohy (1980) necessitated addressing three major econometric issues: (1) the regime-based nature of the sampling procedure, (2) the switching role played by the option equation, and (3) simultaneous equations bias. Here a few concluding remarks are offered concerning each of these issues.

With respect to (1) the overall impression one obtains from tables 11.1 through 11.3 is that the WESML estimator yields tangibly better results than the unweighted estimator. However, the choice between the sample selectivity regressors of the two methods had little effect on estimation in the individual regimes.

With respect to (2) the results of the overall model indicate that the sample selectivity regressor has little explanatory power in any of the structural equations. No matter whether one considers OLS estimates of the 26 reduced form equations, or either OLS or 2SLS estimates of the 26 structural equations, there are never more than 3 out of 26 sample selectivity regressors individually significant at the 10 percent significance level. The practical implication of this result is that the unobserved variables on which the option decision depends appear to be uncorrelated with the unobserved components of the various structural equations, and hence the two regimes may be analyzed separately from the option equation without fear of serious sample selectivity.

12. Such a priori specifications are of two types. First are theoretically motivated specifications such as belonging to the opted-out regime of the percent of opted-out physicians in an individual's peer group but not to the opted-in regime. This specification is premised on the grounds that the behavior of opted-out physicians is highly dependent on the market situation they face. Second are sample-motivated specifications such general practitioners and specialists combined in the opted-out regime due to small sample sizes. Despite such a priori specifications, equality across regimes of remaining coefficients may be tested by combining the two regimes in a manner reminiscent of the *D*-method of Goldfeld and Quandt (1972) and Lee (1976).

Finally, with respect to (3) the acknowledgment of simultaneity by the use of 2SLS led to somewhat different results than arose from the use of OLS; however, for the most part the differences were slight. In particular the weak explanatory power of the sample selectivity regressor persisted under both 2SLS and OLS.

References

Cragg, J. G. 1971. Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica.* 39: 829–844.

Goldfeld, S. M., and R. E. Quandt. 1972. *Nonlinear Methods in Econometrics* Amsterdam: North-Holland.

Goldfeld, S. M., and R. E. Quandt. 1973. The Estimation of Structural Shifts by Switching Regressions. *Annals of Economic and Social Measurement.* 2: 475–485.

Goldfeld, S. M., and R. E. Quandt. 1976. Techniques for Estimating Ssitching Regressions. In *Studies in Nonlinear Estimation,* ed. S. M. Goldfeld and R. E. Quandt. Cambridge, Mass.: Ballinger, chapter 1.

Heckman, J. J. 1976. Simultaneous Equation Models with Continuous and Discrete Endogenous Variables and Structural Shifts. In *Studies in Nonlinear Estimation,* ed. S. M. Goldfeld and R. E. Quandt. Cambridge, Mass.: Ballinger, pp. 235–272.

Lee, L. F. 1976. Two Stage Estimation of Limited Dependent Variable Models. Ph.D dissertation. Department of Economics, University of Rochester.

Lee, L. F. 1979. Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables. *Econometrica.* 977–996.

Lee, L. F., and R. P. Trost. 1978. Estimation of Some Limited Dependent Variable Models with Application to Housing Demand. *Journal of Econometrics.* 357–382.

Lee, L. F., G. S. Maddala, and R. P. Trost. 1980. Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Model with Selectivity. *Econometrica.* 48: 491–503.

Maddala, G. S., and F. D. Nelson. 1975. Switching Regression Models with Exogenous and Endogenous Switching. *Proceedings of the Business and Economics Statistics Section, American Statistical Association,* pp. 423–426.

Manski, C. F., and S. Lerman. 1977. The Estimation of Choice Probabilities from Choice Based Samples. *Econometrica.* 45: 1977–1988.

Poirier, D. J. 1976. *The Econometrics of Structural Change* Amsterdam: North-Holland.

Tobin, J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica.* 26: 24–36.

Wolfson, A. D., and C. Tuohy. 1980. *Opting-Out of Medicare; Private Medical Markets within the Ontario Programme.* Toronto: University of Toronto Press.