0.1 Previous Work
Stern has extensive experience in statistical work involving all empirical parts of this project. He has much experience dealing with large primary and secondary data sets (e.g., Holt, Merwin, and Stern 1996, 1999, 2000; Bearse, Gurmu, Rapaport, and Stern 1999; Fitzgerald, Shaunesey, and Stern 2000), count models such as the model specified for health professional supply (Stern 1993; Gurmu, Rilstone, and Stern 1999; Bearse, Gurmu, Rapaport, and Stern 1999; and Fitzgerald, Shaunesey, and Stern 1999), semiparametric estimation such as the model specified for outcomes (Gurmu, Rilstone, and Stern 1994; Mukarjee and Stern 1994; Stern 1996; and Gurmu, Rilstone, and Stern 1999; Pepper and Stern 2000), estimation of production functions such as the model specified for outcomes (Shipe 1992; and Horewitz 1997), and dealing with missing values (Stern 1991; and Lavy, Palumbo, and Stern 1998).

0.2 Hypotheses of Interest
1) Community characteristic \( k \) affects the supply of health professionals for lots of \( k \)’s.
2) Near-by communities compete with each other for health professionals.
3) Ruralness and the proportion of the population that is minority have significant effects on the supply of health professionals.
4) Different types of health professionals are substitutes for each other to some degree but not perfect substitutes.
5) The availability of health professionals in near-by communities has a significant positive effect on outcomes of interest.
6) There is no such thing as a single critical value for identifying areas with health professional shortages. Outcomes depend continuously on the mix of health professionals, to some degree, health professionals are substitutes, and the marginal effect of each health professional type varies across types.

0.3 A Model and Estimation Procedure for the Number of Health Professionals
Consider a model of numbers of health professionals of a particular type where \( h_{ij} \) is the number of professionals of type \( i \) in area \( j \). We can model

\[ h_{ij} \sim \text{Poisson}(\lambda_{ij}) \]

with

\[ \lambda_{ij} = \exp \{ \alpha_i + X_j \beta_i + u_{ij} \} \]

for \( i = 1, 2, \ldots, I \) and \( j = 1, 2, \ldots, J \). Let \( u_j = (u_{1j}, u_{2j}, \ldots, u_{IJ})' \), \( u = (u_1', u_2', \ldots, u_J') \), and assume, for example, that

\[ u \sim N(0, \Omega). \]

---

1 Shipe and Horewitz wrote dissertations under Stern’s supervision.
The $u$ vector, frequently called the unobserved heterogeneity, is capturing the effect of unobserved effects on the number of health professionals and can allow for various types of correlation across the unobserved effects. Consider various forms for $\Omega$, the covariance matrix of $u$:

Example 1: Assume

$$u_{ij} = e_j + v_{ij}$$

(1)

with

$$e_j \sim iidN(0, \sigma^2_e),$$
$$v_{ij} \sim iidN(0, \sigma^2_v);$$

i.e., that the unobserved heterogeneity has two components, an area-specific component, $e_j$, and an idiosyncratic component, $v_{ij}$. This specification allows for the number of health professionals to be correlated across types within the same area. Then

$$\Omega = \sigma^2_e (I_J \otimes \iota_K') + \sigma^2_v (I_J \otimes I_I)$$

where $I_K$ is a $K \times K$ identity matrix and $\iota_K$ is a $K$–vector of 1’s.

Example 2: Assume again that equation (1) holds, but that

$$e_j \sim N(0, \sigma^2_e),$$
$$Ee_j e_k = \sigma^2_e \rho(d(j,k))$$
$$v_{ij} \sim iidN(0, \sigma^2_v)$$

where $d(j,k)$ is a measure of distance between areas $j$ and $k$ and $\rho(\bullet)$ is a correlation function decreasing in absolute value as $d$ increases with $\rho[0] = 1$ and $\rho[d] = 0$ for all $d \geq D$. Define

$$R = \begin{bmatrix}
\rho[d(1,1)] & \rho[d(1,2)] & \cdots & \rho[d(1,J)] \\
\rho[d(1,2)] & \rho[d(2,2)] & \cdots & \rho[d(2,J)] \\
\vdots & \vdots & \ddots & \vdots \\
\rho[d(1,J)] & \rho[d(2,J)] & \cdots & \rho[d(J,J)]
\end{bmatrix}$$

Then

$$\Omega = \sigma^2_e (R \otimes \iota_K' \iota_K) + \sigma^2_v (I_J \otimes I_I).$$

In general, the parameters to estimate are $\theta = (\alpha, \beta, \Omega)$. The log likelihood function is

$$L = \log \int \prod_{j=1}^J \prod_{i=1}^I \exp \left\{ -e^{\alpha_i + X_i \beta_i + u_{ij}} + h_{ij} [\alpha_i + X_i \beta_i + u_{ij}] \right\} \phi(u | \Omega) du$$

(3)
where \( \phi(\bullet | \Omega) \) is the joint normal density of \( u \). Equation (3) can be evaluated numerically when \( \Omega \) has a small number of factors, and, in general, it can be simulated.

This procedure gives us consistent estimates of the effects of the effects of community characteristics, \( \beta \), on health professional supply and of the covariance among various unobserved effects, \( \Omega \), on health professional supply. With these estimates, we can test if community characteristics such as ________ and ________ have an effect on supply. We can also measure how much different areas compete against each other as measured by the structure of the covariance matrix of the unobserved heterogeneity. Finally, we can use the results to predict the number of health professionals in areas where it is missing in the data. This is of value in the next part of the project.

### 0.4 A Model and Estimation Procedure for Outcomes As a Function of the Mix of Health Professionals

Consider some measure of outcome, such as ___________ or ________________, denoted as \( y_{ik} \), the outcome measure \( k \) in area \( i \). Assume that \( y_{ik} \) is a function, \( f \), of \( M \) area characteristics, \( Z_i \), and the supply of health professionals, \( h_i = (h_{i1}, h_{i2}, \ldots, h_{iJ})' \) as described in the previous section. Then, in general there will be \( M + J \) dimensions to the function relating community characteristics and the supply of health professionals to the outcome of interest. On the one hand, especially since we are interested in substitution and complementarity effects across health professionals of different types, it is important to make as few assumptions as possible about \( f \). On the other hand, if \( M + J \) is large, then it will be infeasible to get estimates of the function \( f \) with any precision without imposing more structure on the problem. We propose first a compromise where we assume that

\[
y_{ik} = Z_i \gamma_k + g(h_i) + e_{ik}. \tag{4}
\]

This structure assumes that the effect of community characteristics can be captured in a linear index, \( Z_i \gamma_k \), while the effect of the supply of health professionals must be captured in a general function, \( g(\bullet) \), that should be estimated nonparametrically. Now the problem is of dimension \( J + 1 \), and yet we are still able to control for the effects of community characteristics.

The next issue to deal with is that a consumer is not limited to receiving medical care in his own area. It is important to take into account the availability of medical resources in areas close to the home of the consumer. We can adjust the model in equation (4) to account for such an effect as

\[
y_{ik} = Z_i \gamma_k + \sum_{m=1}^{I} \phi[d(i, m)] g(h_m) + e_{ik} \tag{5}
\]

where, as is \( \rho(\bullet) \) in equation (2), \( \phi(\bullet) \) is a weighting function of distance \( d(i, m) \) between are areas \( i \) and \( m \). This model still allows area \( i \)'s community char-
acteristics to affect the outcome, and it lets the quality of medical resources available in all near areas $m, g(h_m)$, to affect the outcome.

The parameters to estimate in equation (5), $\theta_y$, are $\gamma_k$, parameters implicit in the weighting function $\phi(\bullet)$, and parameters implicit in the medical quality function $g(\bullet)$. The most promising way to estimate $\theta_y$ is to specify $\phi(\bullet)$ and $g(\bullet)$ using one of the flexible functional forms available in the literature [cites] and then estimate $\theta_y$ using nonlinear least squares (NLS) or maximum likelihood estimation (MLE). NLS has the advantage that no significant assumptions (other than exogeneity of regressors) are necessary about the errors. The advantage of MLE is that it provides more efficient estimates conditional on the correct specification of the distribution of the errors. Both can be tried easily.

The last methodological problem to address is that we expect many of the observations on health professionals to be missing. Given our estimates of the parameters in the supply equation in the previous section, we can construct the density of any missing observation conditional on observed variables from the same and adjacent areas. Then we can simulate from that density and use in an adjusted version of equation (5) as described in Lavy, Palumbo and Stern (1998).

The estimates from this procedure give us information about how the mix of health professionals affects each of the outcomes of interest. It allows us to construct useful cutoff points to identify areas where there are critical shortages, and it provides information about the most effective way to reduce the shortage. It also provides information about the substitutability and/or complementarity of different types of health professionals for each other. Finally, it provides information about how the availability of health professionals in neighboring areas affects each of the outcomes of interest.

0.5 References

References


5


