Econometric considerations when using the net benefit regression framework to conduct cost-effectiveness analysis

Abstract

This chapter considers the analysis of a cost-effectiveness dataset from an econometrics perspective. We link cost-effectiveness analysis to the net benefit regression framework and explore insights and opportunities from econometrics and their practical implications. As an empirical illustration, we compare various econometric techniques using a cost-effectiveness dataset from a published study. The chapter concludes with a discussion about implications for applied practitioners and future research directions.

Jeffrey S. Hoch, PhD
Department of Public Health Sciences
University of California, Davis
Davis, California USA 95817

Pierre Chaussé, PhD
Department of Economics
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1

Corresponding Author:
Jeffrey S. Hoch, PhD,
Email: jshoch@ucdavis.edu

1.0 Introduction

Healthcare costs are an important consideration for policy advisors and decision makers. Costly innovations are arriving at an increasing rate, and there is concern about how to spend limited healthcare budgets. Most countries throughout the world use a health technology assessment process to help inform their healthcare funding decisions. Economic evidence is an important part of this, and cost-effectiveness analysis is one of the most popular economic evaluation techniques used to inform healthcare spending decisions.

Two types of cost-effectiveness analysis involve creating decision models (using data from multiple sources) or estimating statistical models (using data from a single dataset). Statistical cost-effectiveness analysis has many intriguing features for those interested in theoretical and applied econometrics. The analyst generally has a small dataset of N study participants of whom n_1 received a new treatment (or intervention) and n_0 received usual care (where $N = n_1 + n_0$). At a minimum, each observation i provides a data triplet of cost (c_i), outcome or effectiveness (e_i) and a treatment indicator (tx_i). A sample consists of two sets $\{(c_i, e_i, tx_i = 1) : i = 1 \text{ to } n_1\}$ and $\{(c_i, e_i, tx_i = 0) : i = 1 \text{ to } n_0\}$. Typically, c_i and e_i are assumed to be jointly distributed, potentially correlated dependent variables. Many times the data come from a randomized controlled trial where it is typical to assume that covariates ($\mathbf{X}_i$) are not associated with the treatment allocation. For cost-effectiveness analysis, the analyst must produce functions of the estimates of $E(c \mid tx = 1)$, $E(e \mid tx = 1)$, $E(c \mid tx = 0)$ and $E(e \mid tx = 0)$. The key econometric questions involve 1) what is the best way to obtain the estimates of the functions of these moments and 2) how best should their uncertainty be characterized.

We explore these questions in this chapter. After providing additional background on statistical cost-effectiveness, we consider questions about estimation and uncertainty next and then subsequently illustrate findings with an empirical example. We conclude with a discussion about implications for applied practitioners and future research directions.

2.0 Background

2.1 The incremental cost-effectiveness ratio (ICER)

Most health technology assessment processes throughout the world require partial results from a constrained optimization problem. The academic rationale for this appears to be related to viewing the fixed healthcare budget as a constraint (i.e., the amount of money that can be spent is limited) and viewing the objective in healthcare to be maximizing the population's health. Thus, when considering which healthcare treatments to reimburse, a healthcare payer is assumed to face the following problem:

> Choose the optimal levels of funding (i.e., δ going from 0 – 100%) of M Treatments (i.e., τ_i for i = 1 to M), assuming the τ_i's have health outcomes of τ_i^o and costs of τ_i^c with an objective of maximizing $\Sigma\delta_i\tau_i^o$ within a fixed budget of Ξ (i.e., $\Sigma\delta_i\tau_i^c \leq \Xi$).

Weinstein and Zeckhauser (1973) showed the optimal decision rule is equivalent to funding new treatments or interventions when the ratio of the extra cost (ΔC) to the extra health outcome or effectiveness (ΔE) is less than a willingness to pay threshold (λ). In other words, decision makers should fund a new treatment if $\Delta C/\Delta E < \lambda$ (when $\Delta E > 0$). The incremental cost-effectiveness ratio ($\Delta C/\Delta E$) has played a major role in cost-effectiveness analysis based on this stylized version of how decision makers are assumed to behave. Nevertheless, some methodological and practical challenges attend the use of the incremental cost-effectiveness ratio (ICER).

While the goal of cost-effectiveness analysis is to understand the trade-off between ΔC and ΔE, results from Weinstein and Zeckhauser (1973) appear to impose that this tradeoff should be estimated as a ratio (the ICER); however, ratios can be challenging to estimate. For example, if one denotes the population expected values of cost and effectiveness for $tx_i = 0$ and 1 as μ_{c0}, μ_{e0}, μ_{c1} and μ_{e1}, respectively, then the population ICER statistic is defined as $R \equiv (\mu_{c1} - \mu_{c0}) / (\mu_{e1} - \mu_{e0})$ or simply $\mu_{\Delta C} / \mu_{\Delta E}$. With a cost-effectiveness dataset, one can replace the population parameters with their sample analogues (i.e., replace population cost and effectiveness expectations with the sample cost and effectiveness averages). However, the common estimate of the ICER, the ratio of the differences in the sample means of cost and effectiveness

(1) $$\hat{R} = \frac{\bar{c}_1 - \bar{c}_0}{\bar{e}_1 - \bar{e}_0} = \frac{\hat{\mu}_{\Delta C}}{\hat{\mu}_{\Delta E}}$$

is biased. That is, $E(\hat{R}) \neq R$ and this divergence is inversely related to the *unknown parameter* $\mu_{\Delta E}$. In addition, the 95% confidence interval for (1) is not trivial to compute. A parametric solution is sometimes available through Fieller's theorem which involves solving

(2) $$\frac{\hat{\mu}_{\Delta C}^2 + R^2\hat{\mu}_{\Delta E}^2 - 2R\hat{\mu}_{\Delta E}^2\hat{\mu}_{\Delta C}}{R^2\widehat{Var}(\hat{\mu}_{\Delta E}) + \widehat{Var}(\hat{\mu}_{\Delta C}) - 2R\widehat{Cov}(\hat{\mu}_{\Delta E}, \hat{\mu}_{\Delta C})} = z_{\alpha/2}^2$$

for R, where $z_{\alpha/2}$ is the $\alpha/2$ percentile of the standard normal cumulative distribution function. This equation can be a source of difficulties for applied practitioners who cannot always expect to be able to calculate an upper and lower 95% confidence interval (sometimes because of calculations errors, sometimes because of imaginary roots and sometimes because of both). Bootstrapping can serve as an alternative approach. However, Siani et al (2000) show that Fieller's method can perform better than bootstrap methods that become unstable or even inapplicable when the difference between average effects approaches zero statistically (i.e., $\Delta E \to 0$). Severens et al (1999) state that both the Fieller and bootstrap methods lead to "unsatisfactory results" when the difference in effectiveness is approximately zero (i.e., $\Delta E \approx 0$).[1]

---

[1] A natural alternative is to consider using a Taylor series approximation of the variance of a function of two random variables (often termed the Delta method) to estimate the variance of the ICER. Success in this endeavor allows one to use standard parametric assumptions to produce a confidence interval of the form $\hat{R} \pm z_{\alpha/2} \, var(\hat{R})^{1/2}$. Briggs and Fenn (1998) note that "a high coefficient of variation for the denominator of the ratio (i.e. a non-negligible probability of observing a zero value) means that the sampling distribution of the ICER is likely to be non-normal and that the Taylor series will give a poor estimate of variance (Armitage and Berry, 1994)." Moreover, van Hout et al note the ratio of two normal distributed variables (e.g., $\Delta C/\Delta E$) has neither a finite mean nor a finite variance, and one of the consequences is that using a Taylor approximation to calculate 95% confidence limits is formally incorrect.

Last but not least, there is the delicate issue of what to do about λ. Decision makers are, in theory, considering whether to fund a new treatment based on whether $\Delta C/\Delta E < \lambda$, where λ represents the willingness to pay for an additional health outcome or unit of effectiveness. If $\hat{R} < \lambda$ when $\Delta E > 0$, then the new treatment or intervention is described as "cost-effective". Estimates of ΔC and ΔE come from the data; an actual number for λ requires a value judgment from the decision maker. Providing an estimate for R without putting it into context in relation to λ seems incomplete and does not facilitate researchers making policy recommendations. While it is difficult to comment on whether $R < \lambda$ without formally considering λ in the analysis, λ is generally unknown and $\hat{R}$ is a biased estimate with a tricky confidence interval. The incremental net benefit approach represents an attractive alternative (Stinnett and Mullahy, 1998; Tambour et al., 1998).

2.2 The incremental net benefit (INB)

The incremental net benefit (INB) addresses two statistical problems with conducting estimation of and inference on the ICER (i.e., that $\hat{R}$ is a biased estimate of the ICER and that 95% confidence intervals are often difficult to construct or interpret[2]). When using the INB for cost-effectiveness analysis, both estimation and inference are greatly simplified, since the INB estimate

(3) $$\hat{B} = \lambda(\bar{e}_1 - \bar{e}_0) - (\bar{c}_1 - \bar{c}_0) = \lambda \cdot \hat{\mu}_{\Delta E} - \hat{\mu}_{\Delta C}$$

is a linear function. As $E(\hat{B}) = B \equiv \lambda \cdot \mu_{\Delta E} - \mu_{\Delta C}$, $\hat{B}$ made from sample means is unbiased.[3] In addition, the 95% CIs can be made in the standard way as $\hat{B} \pm z_{\alpha/2}\sqrt{\widehat{Var}(\hat{B})}$ where

(4) $$Var(\hat{B}) = \lambda^2 Var(\hat{\mu}_{\Delta E}) + Var(\hat{\mu}_{\Delta C}) - 2\lambda \cdot Cov(\hat{\mu}_{\Delta E}, \hat{\mu}_{\Delta C}).$$

$\widehat{Var}(\hat{B})$ can be calculated by using the sample estimates for variances (Var) and covariance (Cov) in equation (4). Alternatively, $\hat{B}$ and the associated 95% CI can be obtained directly from net benefit regression, as discussed shortly. Of course, λ must be specified; however, this is also true for any decision based on the ICER because treatment is only deemed "cost-effective" if $R < \lambda$. Thus, one cannot avoid specifying λ, which plays an implicit role in the ICER approach and an explicit role in the INB approach.

Because of the tautology that $\hat{B} > 0$ whenever $\hat{R} < \lambda$, both their estimates and uncertainty are intimately connected. When the willingness to pay value λ is set equal to $\hat{R}$, then $\hat{B} = 0$. Also, the upper and lower 95% CIs for the INB are related to the Fieller 95% CI for the ICER. A graph of $\hat{B}$ by λ has a y-intercept equal to $-\Delta C$, a slope of ΔE and an x-intercept of $\hat{R}$ (see Figure 1). The addition to the graph of 95% CIs for the INB can illustrate, at their x-intercepts, the lower and upper 95% Fieller CIs for the ICER (see Figure 1).

[2] This especially true when either 1) dealing with ICERs < 0 or 2) characterizing uncertainty when ΔC and/or ΔE are not significantly different from 0.

[3] $E(\hat{B}) = E(\lambda\,\Delta E - \Delta C) = E[\lambda(\bar{e}_1 - \bar{e}_0) - (\bar{c}_1 - \bar{c}_0)] = \lambda E[(\bar{e}_1 - \bar{e}_0)] - E[(\bar{c}_1 - \bar{c}_0)] = \lambda \cdot \hat{\mu}_{\Delta E} - \hat{\mu}_{\Delta C}$.

2.3 The net benefit regression framework (NBRF)

The NBRF is a regression framework for the net benefit approach (Hoch et al, 2002). Under this framework, each subject's net benefit, nb_i, is defined as $nb_i \equiv e_i \cdot \lambda - c_i$ using the observed data on e_i and c_i, the effectiveness and cost data for person *i*. If $nb_i > 0$, then the benefits (in $) outweigh the costs (in $) for person *i*. If E(B | tx = 1) > E(B | tx = 0), then the net benefits from new treatment outweigh the net benefits from usual care, overall. This comparison can be made with sample data by comparing $\sum_{i=1\ to\ n_1} \mathrm{nb}_i / n_1$ to $\sum_{i=1\ to\ n_0} \mathrm{nb}_i / n_0$. The NBRF places this comparison in a regression framework.

In its simplest form, the NBRF involves fitting the following regression model $nb_i = \beta_0 + \beta_{tx}\, tx_i + \varepsilon_i$ where tx_i and ε_i are the i^{th} person's treatment indicator and stochastic error term, respectively. The regression is typically fit several times, each time with a different λ value (e.g., a small, medium and large value and Figure 1 shows which values of λ make intuitive sense to consider when illustrating INB). The regression can be enhanced with a vector of subject characteristics ($\mathbf{X}_i$) to improve the efficiency of the β estimates (as illustrated in Section 3). In addition, interaction terms (e.g., $\mathbf{X}_i \times tx_i$) or stratification can be used to test for patient subgroups for whom the cost-effectiveness of a treatment varies with membership. When it was originally proposed, Hoch et al (2002) suggested Ordinary Least Squares (OLS) to produce β estimates. The OLS estimate of β_{tx} equals the difference in the average NB for tx = 1 and 0. This difference is the INB since $\hat{\beta}^{OLS}_{tx} = (\lambda\bar{e}_1 - \bar{c}_1) - (\lambda\bar{e}_0 - \bar{c}_0)$ = λΔE - ΔC ≡ INB. If $\hat{\beta}^{OLS}_{tx}$ is > 0, the new treatment is cost-effective; if $\hat{\beta}^{OLS}_{tx} < 0$, the new treatment is not cost-effective. Thus, with a simple OLS regression, one can assess a new treatment or intervention's cost-effectiveness through the INB's estimate and uncertainty, as indicated by what the data tell us about β_{tx}.

In the NBRF, one can separate the regression equation $nb_i = \beta_0 + \beta_{tx}\, tx_i + \varepsilon_i$ into cost and effectiveness parts; e.g., $c_i = \alpha_0 + \alpha_{tx}\, tx_i + v_i$ and $e_i = \xi_0 + \xi_{tx}\, tx_i + u_i$. It is possible to verify that $\hat{B} = \lambda \cdot \hat{\xi}_{tx} - \hat{\alpha}_{tx} = \hat{\beta}_{tx}$ and $\hat{R} = \hat{\alpha}_{tx} / \hat{\xi}_{tx}$. The NBRF allows the exploration of a system of equations (i.e., one for c_i and one for e_i). While the NBRF solves many problems, questions still remain:

- What is the best way to estimate the INB (e.g., are more complex methods needed or helpful)?
- Should the c_i and the e_i equations be estimated as a system of simultaneous equations or as a single net benefit regression equation?
- What are the best methods to use for estimation and uncertainty?

3.0 Methods

3.1 Criticisms of the net benefit regression framework

Two critical issues related to using OLS to estimate the INB with a net benefit regression approach are that 1) the distribution of ε may not be well suited for OLS[4] and 2) not all covariates in **X** may be in both the cost and effectiveness equations. This section considers these issues in an econometric framework.

[4] Common concerns include skewness and/or heteroscedasticity.

We assume a cost-effectiveness dataset composed of a sample of cost and effectiveness data for patients receiving either new treatment (tx = 1) or usual care (tx = 0), drawn from a data generating process with a general bivariate distribution

$$\begin{pmatrix} c_{i,tx} \\ e_{i,tx} \end{pmatrix} \sim \left(\begin{pmatrix} \mu_{c_{tx}} \\ \mu_{e_{tx}} \end{pmatrix} , \begin{pmatrix} \sigma^2_{c_{tx}} & \sigma_{c_{tx}e_{tx}} \\ \sigma_{c_{tx}e_{tx}} & \sigma^2_{e_{tx}} \end{pmatrix} \right)$$

where i = 1 to n_{tx}.[5] The regression equations of interest can be presented in a general way:

(5.1) $c_i = \alpha_0 + \alpha_{tx}\, tx_i + \boldsymbol{\alpha_x}\, \mathbf{X}_i^c + v_i$

(5.2) $e_i = \xi_0 + \xi_{tx}\, tx_i + \boldsymbol{\xi_x}\, \mathbf{X}_i^e + u_i$

(5.3) $nb_i\,(\lambda) = \beta_0 + \beta_{tx}\, tx_i + \boldsymbol{\beta_x}\, \mathbf{X}_i + \varepsilon_i$.

To emphasizes the fact that nb_i is a function of λ, equation (5.3) is written as $nb_i\,(\lambda)$. The vector of covariates in the cost equation $\mathbf{X}_i^c$ may differ from the vector of covariates in the effectiveness equation $\mathbf{X}_i^e$.[6] Note that the vector of covariates may contain interaction terms (e.g., the product of a patient characteristic x_i and the treatment indicator tx_i).

Willan et al. (2004) proposed the use of a system of seemingly unrelated regression (SUR) equations to estimate the coefficients in equations (5.1) and (5.2) as it does not require that the set of independent variables for costs and effectiveness be the same (i.e., it allows $\mathbf{X}_i^c \neq \mathbf{X}_i^e$).[7] However, they also observed that it was possible for interaction term estimates of α, ξ or both to be not statistically significant but the additional test of the hypothesis $\lambda \cdot \xi - \alpha = 0$ is required to determine if there is a significant interaction between the variable in question and the treatment group. In addition, they noted that if the covariates for cost and effectiveness in equations (5.1) and (5.2) are the same (i.e., $\mathbf{X}_i^c = \mathbf{X}_i^e$), then SUR estimates and uncertainty measures for equations (5.1) and (5.2) match those of OLS. And, these are related to OLS estimates of (5.3) in the form of $\hat{\beta} = \lambda \cdot \hat{\xi} - \hat{\alpha}$. Lastly, they observed that if $\mathbf{X}_i^c \neq \mathbf{X}_i^e$, efficiency gains over OLS are possible.

In their empirical example, Willan et al. (2004) explore the case of $\mathbf{X}_i^c \neq \mathbf{X}_i^e = 0$, and from a histogram of the residuals in the cost equation, they found evidence of skewing. They addressed concerns about the residual's distribution not being well suited for OLS by conducting simulations showing the robustness of OLS in the presence of right-skewing, in particular for cost data that are log-normal.

---

[5] There will be n_0 participants receiving usual care and n_1 receiving new treatment with the total sample size being equal to $N = n_1 + n_0$.

[6] The covariate tx_i is always in both the cost and effectiveness regression equations, and the covariate vectors $\mathbf{X}_i$, $\mathbf{X}_i^c$ and $\mathbf{X}_i^e$ contain any other covariates.

[7] The claim is that estimating equation (5.3) is consistent with assuming that each variable in $\mathbf{X}$ is also in equations (5.1) and (5.2), and with a variable x_c assumed to be in $\mathbf{X}^c$ but not in $\mathbf{X}^e$, piecewise estimates of INB like $\lambda \cdot \hat{\xi}_{tx} - \hat{\alpha}_{tx}$ will be more accurate than the single equation estimate of β_{tx}, unless $\xi_{x\_c} = 0$.

3.1 Critique of seemingly unrelated regression (SUR) in CEA

It is possible that SUR is not the optimal choice for estimation and uncertainty either in general or in the case that Willan et al (2004) consider:

(6.1) $$c_i = \alpha_0 + \alpha_{tx}\, tx_i + \boldsymbol{\alpha}_{\mathbf{x}}\, \mathbf{X}_i^{Conly} + v_i$$

(6.2) $$e_i = \xi_0 + \xi_{tx}\, tx_i + u_i$$

In addition, it is not clear how $\mathbf{X}_i^{Conly}$, the variables that appear in the cost equation but not the effectiveness equation, should be considered.[8] For example, they could be constrained in a simultaneous system of equations involving a net benefit regression equation like

(6.3) $$nb_i(\lambda) = \beta_0 + \beta_{tx}\, tx_i + \boldsymbol{\beta}_{\mathbf{x}}\, \mathbf{X}_i + \varepsilon_i.$$

Given the intimate relationship between OLS, SUR, and Generalized Method of Moments (GMM), it is natural to consider GMM in this scenario. While SUR may be better than OLS because it imposes a particular data structure, GMM may be better than SUR because it does so in an optimal fashion.

3.2 An overview of Generalized Method of Moments (GMM)

GMM is a family of methods for which SUR and OLS are special cases. We briefly illustrate this by considering a three-equation regression system of the form:

$$\mathbf{Y} = \mathbf{Z}\theta + \omega,$$

where

$$\mathbf{Y} = \begin{pmatrix} \mathbf{c} \\ \mathbf{e} \\ \mathbf{nb}(\lambda) \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{X}^c & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}^e & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X} \end{pmatrix}, \quad \theta = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\xi} \\ \boldsymbol{\beta} \end{pmatrix}, \quad \omega = \begin{pmatrix} \mathbf{v} \\ \mathbf{u} \\ \boldsymbol{\varepsilon} \end{pmatrix}$$

and $\mathbf{0}$ represents a matrix of appropriate dimensions in which all elements are zero. By construction, $nb_i(\lambda) = \lambda e_i - c_i$, which implies that $\mathbf{X}_i$ is the union of $\mathbf{X}^c_i$ and $\mathbf{X}^e_i$. The SUR assumption requires $\mathbf{X}_i$ to be orthogonal to the error term in each equation. The following moment conditions are therefore satisfied *by assumption*:

$$E[g_i(\theta)] \equiv E\begin{bmatrix} v_i \mathbf{X}_i \\ u_i \mathbf{X}_i \\ \varepsilon_i \mathbf{X}_i \end{bmatrix} = 0.$$

Because of the way $nb_i(\lambda)$ is defined, $\varepsilon_i = \lambda u_i - v_i$. The third vector of moment conditions is collinear with the first two. It is impossible to estimate the model without imposing restrictions on the parameters. An easy way to do so is to estimate the system composed of any two equation and obtain the third using the linear relationship between the three dependent variables (i.e., c, e and nb).

[8] More general is the case where vectors like $\mathbf{X}_i^{Conly}$ and $\mathbf{X}_i^{eonly}$ exist and there are some variables that only appear in the cost equation and some that only appear in the effectiveness equation.

To introduce the GMM method, we will consider the first two equations,[9] redefining the above matrices to be

$$\mathrm{Y} = \begin{pmatrix}\mathbf{c}\\ \mathbf{e}\end{pmatrix}, \qquad \mathrm{Z} = \begin{pmatrix}\mathbf{X}^{c} & 0\\ 0 & \mathbf{X}^{e}\end{pmatrix}, \qquad \boldsymbol{\theta} = \begin{pmatrix}\boldsymbol{\alpha}\\ \boldsymbol{\xi}\end{pmatrix}, \qquad \boldsymbol{\omega} = \begin{pmatrix}\mathbf{v}\\ \mathbf{u}\end{pmatrix}.$$

Also, let $g_i(\boldsymbol{\theta}) = \boldsymbol{\omega}_i \otimes \mathbf{X}_i = \{v_i\mathbf{X}'_i, u_i\mathbf{X}'_i\}'$ be the moment function (where $\otimes$ is the Kronecker product). Then the GMM estimator is defined as the vector $\hat{\theta}$ that makes the sample moment $g_n(\boldsymbol{\theta}) = (1/n)\ \Sigma_{i=1 \text{ to } n}\ g_i(\boldsymbol{\theta})$ as close as possible to its population value which is zero by the SUR assumption. More precisely, the GMM estimator is defined as

$$\hat{\theta}(\widehat{\mathbf{W}}) = \underset{\theta}{\operatorname{argmin}}\ g_n(\boldsymbol{\theta})'\widehat{\mathbf{W}}g_n(\boldsymbol{\theta}),$$

where $\widehat{\mathbf{W}}$ is a possibly random weighting matrix that converges to a non-random and positive definite matrix **W**. If the SUR assumption and other regularity conditions are satisfied, the GMM estimator is consistent for all $\widehat{\mathbf{W}}$.

However, the choice of $\widehat{\mathbf{W}}$ impacts the efficiency of the estimate. The efficiency bound is reached when $\widehat{\mathbf{W}}$ converges to the inverse of the asymptotic variance of $\sqrt{n}\, g_n(\boldsymbol{\theta})$, which we define as **S**. Since the estimation of the optimal **W** depends on $\boldsymbol{\theta}$, the GMM estimator is often obtained in two steps. First, we obtain a consistent vector of estimates $\tilde{\boldsymbol{\theta}}$ using a fixed **W** (e.g., the identity matrix), then we compute an estimate of the covariance matrix of the sample moments, $\hat{\mathbf{S}}(\tilde{\boldsymbol{\theta}})$. The efficient GMM estimate is then obtained by replacing $\widehat{\mathbf{W}}$ by $\hat{\mathbf{S}}(\tilde{\boldsymbol{\theta}})^{-1}$ to arrive at

$$\hat{\boldsymbol{\theta}}\left[\hat{\mathbf{S}}(\tilde{\boldsymbol{\theta}})^{-1}\right] = \underset{\theta}{\operatorname{argmin}}\ g_n(\boldsymbol{\theta})'\hat{\mathbf{S}}(\tilde{\boldsymbol{\theta}})^{-1}g_n(\boldsymbol{\theta}).$$

The way we compute $\hat{\mathbf{S}}(\tilde{\boldsymbol{\theta}})$ depends on assumptions about the variance $\boldsymbol{\omega}_i$ which we define as an $n \times 2$ matrix with the i^{th} row being $\{v_i, u_i\}$. If we assume conditional homoscedasticity, implying that

$$E(\boldsymbol{\omega}_i\boldsymbol{\omega}_i' \mid \mathbf{X}_i) = \begin{bmatrix} Var(v_i) & Cov(v_i,u_i)\\ Cov(v_i,u_i) & Var(u_i)\end{bmatrix} \equiv \boldsymbol{\Omega},$$

---

[9] To simplify exposition, the covariate tx_i is assumed to be part of the covariate vectors **X**, $\mathbf{X}^c$ and $\mathbf{X}^e$. The vector of parameters $\boldsymbol{\theta}$ can be estimated by OLS, SUR or GMM. When the vector of covariates in the cost equation is the same as those in the effectiveness equation ($\mathbf{X}^c = \mathbf{X}^e$), there is no gain from joint estimation (Fiebig, 2001). In our case study, we face a case where $\mathbf{X}^c \neq \mathbf{X}^e$ suggesting the use of SUR on two fronts: first to gain efficiency in estimation by combining information on the different equations, and second to impose and/or test restrictions that involve parameters in different equations (Moon and Perron, 2006). However, efficient estimators propagate misspecification and inconsistencies across equations, so if any equation is misspecified, then the entire coefficient vector will be inconsistently estimated by efficient methods (Moon and Perron, 2006). In this sense, equation-by-equation OLS provides some degree of robustness since it is not affected by misspecification in other equations in the system (Moon and Perron, 2006). Although efficient GMM is not better than SUR regarding the propagation of misspecification across equations, it does provide additional benefits.

then $\hat{\mathbf{S}}(\tilde{\boldsymbol{\theta}}) = \tilde{\Omega} \otimes [\mathbf{X'X}/n]$, with $\tilde{\Omega} = \tilde{\omega}'\tilde{\omega}/n$ and the GMM estimates are identical to those using SUR. However, if we are not willing to make such a restrictive assumption, we can use the following heteroscedasticity consistent estimator:

$$\hat{\mathbf{S}}(\tilde{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}\tilde{v}_i^2\mathbf{X}_i\mathbf{X}_i' & \tilde{v}_i\tilde{u}_i\mathbf{X}_i\mathbf{X}_i' \\ \tilde{v}_i\tilde{u}_i\mathbf{X}_i\mathbf{X}_i' & \tilde{u}_i^2\mathbf{X}_i\mathbf{X}_i'\end{pmatrix}.$$

In order to compare SUR and GMM, it helps to compare the covariance matrix of the GMM versus the efficient GMM estimator.[10] In general, we have

$$\sqrt{n}\,[\hat{\boldsymbol{\theta}}(\widehat{\mathbf{W}}) - \boldsymbol{\theta}] \rightarrow \mathrm{N}[0, \mathrm{V}(\mathbf{W})]$$

where

$$\mathrm{V}(\mathbf{W}) = [\mathbf{G'WG}]^{-1}\mathbf{G'WSWG}[\mathbf{G'WG}]^{-1}$$

with

$$\mathbf{G} = \begin{pmatrix}\mathrm{E}(\mathbf{X}_i^c\mathbf{X}_i') & 0 \\ 0 & \mathrm{E}(\mathbf{X}_i^e\mathbf{X}_i')\end{pmatrix}.$$

If $\widehat{\mathbf{W}}$ converges to $\mathbf{S}^{-1}$ then $\mathrm{V}(\mathbf{W}) = \mathrm{V}(\mathbf{S}^{-1}) = [\mathbf{G'S}^{-1}\mathbf{G}]^{-1}$.

3.3 The relationship between GMM, SUR and OLS

SUR corresponds to efficient GMM under the assumption that

$$\mathrm{E}(\omega_i\omega_i' \mid \mathbf{X}_i) = \begin{bmatrix} Var(v_i) & Cov(v_i,u_i) \\ Cov(v_i,u_i) & Var(u_i)\end{bmatrix} \equiv \boldsymbol{\Omega}.$$

In the presence of heteroscedasticity, SUR is less efficient than efficient GMM. Furthermore, in that case, the SUR covariance matrix should be a consistent estimate of V(**W**), where

$$\mathbf{W} = \boldsymbol{\Omega}^{-1} \otimes \mathrm{E}(\mathbf{X}_i\mathbf{X}'_i)^{-1} \neq \mathbf{S}^{-1}.$$[11]

There is also a relationship between OLS and GMM. In the simplest case in which $\mathbf{X}^e = \mathbf{X}^c = \mathbf{X}$, the model is just (or exactly) identified. The weighting matrix plays *no* role since $\hat{\boldsymbol{\theta}}$ is simply the solution to the linear system of equations $g_n(\boldsymbol{\theta}) = 0$. It is easy to show that the solution is identical to the equation by equation OLS estimate, since the linear system of equations $g_n(\boldsymbol{\theta}) = 0$ are the OLS first order conditions, in this case. As such, there is no gain from using GMM or SUR; they only differ by the choice of $\widehat{\mathbf{W}}$ which no longer affects the solution.

However, a key point is that we can use the GMM setup to test restrictions involving parameters from different equations. In fact, we can show that the asymptotic variance of the GMM estimator in the simple case of $\mathbf{X}^e = \mathbf{X}^c = \mathbf{X}$ is $\mathrm{V}(\mathbf{W}) = \mathbf{G}^{-1}\mathbf{S}\mathbf{G}^{-1}$.[12] It is therefore possible to obtain confidence intervals for $\hat{B}$

[10] For complete coverage of GMM for systems of equations, see Hayashi (2000).
[11] These are asymptotic results. In finite samples, the relative efficiency of GMM over SUR or OLS is unclear.
[12] We kept the argument **W** in the V() even though it plays no role in the just identified case.

even though the coefficients come from different equations. Again, **S** can be estimated by either a non-robust or a robust covariance matrix. It is also possible to show that when setting

$$\hat{\mathbf{W}} = \hat{\boldsymbol{\Omega}}^{-1} \otimes (\mathbf{X}'_i\mathbf{X}_i/n),$$

where $\hat{\boldsymbol{\Omega}}$ is diagonal, then GMM is identical to equation by equation OLS. In other words, OLS equation by equation is GMM with the SUR, homoscedasticity and no correlation between the error terms assumptions. As a corollary, by not imposing that $\hat{\boldsymbol{\Omega}}$ be diagonal, but instead assuming

$$\mathbf{S} = \boldsymbol{\Omega} \otimes \mathrm{E}(\mathbf{X}_i\mathbf{X}_i')$$

with $\boldsymbol{\Omega}$ being diagonal, then SUR and efficient GMM are asymptotically identical to equation by equation OLS.

A final important issue to be aware of regarding the estimation of a system of equations as a whole is that in order for the efficient GMM or SUR procedures to produce consistent estimates, all equations must be correctly specified. Suppose, for example, that one regressor in $\mathbf{X}^c$ but not in $\mathbf{X}^e$ was correlated with u_i (the error term for the effect equation) but not with v_i (the error term for the cost equation). Then, one of the moment conditions $E(u_i\ \mathbf{X}_i) = 0$ would not be satisfied. As a result, the equation by equation OLS estimates would be consistent, but the efficient GMM or SUR would not. In fact, the violation of one moment condition in one equation can contaminate all equations. OLS is therefore more robust to misspecification (Moon and Perron, 2006).

4.0 Case study

This section describes the empirical example and motivates the methods we use.

4.1 Background on the study

The Program in Assertive Community Treatment is one of the most studied models of care for persons with severe and persistent mental illnesses (SPMI) (Stein and Test, 1980; Olfson, 1990; Burns and Santos, 1995; Scott and Dixon 1995). Lehman et al. (1999) found that an assertive community treatment (ACT) program, relative to usual community services, reduced days homeless for homeless persons with SPMI in Baltimore, Maryland (USA). The study's rationale was that by providing potentially more expensive but coordinated, community-based care through the ACT program, homeless persons with SPMI would spend more days in stable community housing with savings realized by shifting the patterns of care from higher cost crisis-oriented inpatient and emergency services to lower cost, ongoing ambulatory services. The results suggest that in the city of Baltimore, ACT was effective in achieving important outcomes warranting an examination of the cost-effect trade-off. Lehman et al. (1999) conducted an economic evaluation of the ACT program as it was implemented. An analysis of the cost-effectiveness dataset by Hoch et al (2002) used net benefit regression. The same dataset was analyzed by Willan et al (2004) using SUR. The analysis that we report next presents OLS, SUR and GMM results.

4.2 Background on the data

Direct treatment costs across the one year intervention period were examined from the perspective of the state mental health authority. Housing status was chosen as the main effectiveness. A day of stable housing was defined as living in a non-institutionalized setting not intended to serve the homeless (e.g., independent housing, living with family, etc.). Subjects randomized to the comparison usual care condition had access to services usually available to homeless persons in the city of Baltimore. Lehman et al. (1999) offer more detail about the study. Cost-effectiveness analyses of these data have used the complete data on 73 participants randomly assigned to the ACT program (tx = 1) and 72 randomly assigned to usual care services (tx = 0).

An unusual feature of the sample data is that while the two treatment groups appeared comparable with respect to most covariates (e.g., age and Global Assessment of Functioning), there was a greater than expected percentage of African Americans *not* randomized to the innovative ACT treatment ($p < 0.01$). This observation serves as the point of departure for various modeling strategies. Both Hoch et al. (2002) as well as Hoch and Blume (2008) addressed the imbalance between race and treatment allocation using net benefit regressions of the form

(6.3′) $$nb_i(\lambda) = \beta_0 + \beta_{tx}\, tx_i + \beta_{Black}\, Black_i + \beta_{Black\_tx}\, Black_i \times tx_i + \varepsilon_i.$$

The indicator for race ($Black_i$) was 1 for African Americans and 0 otherwise, and the indicator for randomization group (tx_i) was 1 for the ACT group and 0 otherwise.

In contrast to the net benefit regression modeling strategy, Willan et al. (2004) made use of the simultaneous equation nature of the net benefit regression framework to focus on the cost and effectiveness regression equations (6.1) and (6.2). This is justified by referring to Altman (1985) in explaining that the confounding effect of a baseline covariate has more to do with the magnitude of the between group difference and the magnitude of its effect on the outcome variable, rather than with the statistical significance of the between-group difference; consequently, a regression model was used to examine the effects of covariates suspected of affecting the outcome. As a result, they employed a modeling strategy of the form

(6.1′) $$c_i = \alpha_0 + \alpha_{tx}\, tx_i + \alpha_{Black}\, Black_i + \alpha_{Black\_tx}\, Black_i \times tx_i + v_i$$

(6.2′) $$e_i = \xi_0 + \xi_{tx}\, tx_i + u_i$$

estimating the coefficients using a system of seemingly unrelated regression equations (SUR). The differing covariate structure (i.e,. $\mathbf{X}_i^c \neq \mathbf{X}_i^e = 0$) is suggested by OLS results in Table A.

The results for the effectiveness regression equation exhibit non-significant coefficient estimates for the $Black_i$ and $Black_i \times tx_i$ variables. However, in the cost regression equation, the estimates are statistically significant. Willan et al. (2004) note that because the coefficient for race and its interaction were significant for cost, there is some evidence for concluding that ACT's impact on cost depends on race; the implication being that cost-effectiveness likely varies by race.

Two additional challenges are the presence of heteroscedasticity and the skewed nature of the data. Breusch-Pagan / Cook-Weisberg tests for heteroscedasticity using the results in Table A show mixed evidence. For the effectiveness equation, the null hypothesis of constant variance cannot be rejected at conventional levels ($\chi^2_{(3)} = 1.11$, $P = 0.29$); however, for the cost equation homoscedasticity is rejected ($\chi^2_{(3)} = 14.35$, $P < 0.001$). The skewness of the data is illustrated in Figure 2, where various transformations to achieve normality are compared. The cost data seem non-normal, as is common.

4.3 Estimation strategies

In this section we provide the rationale behind the estimation strategies used to analyze the data. The first two strategies we consider involve equation by equation estimation by OLS of the cost, effectiveness and net benefit regressions with all of the same covariates (i.e., $\mathbf{X}_i^c = \mathbf{X}_i^e$) and allowing for different covariates (i.e., $\mathbf{X}_i^c \neq \mathbf{X}_i^e$). We call these strategies OLS – I and OLS – II.

The OLS strategies are included to provide a yardstick to compare other estimation and uncertainty procedures. Given that there is some evidence of heteroscedasticity, we use the "robust" option to produce White-corrected standard errors in the presence of heteroscedasticity (MacKinnon and White, 1985; Davidson and MacKinnon, 2004). We also consider simultaneous equation estimation of the cost, effectiveness and net benefit regressions (where possible), both with all the same covariates (i.e., $\mathbf{X}_i^c = \mathbf{X}_i^e$) and without (i.e., $\mathbf{X}_i^c \neq \mathbf{X}_i^e$). We denote these strategies SUR – I and SUR – II when we use SUR to produce estimates, and we label these strategies GMM – I and GMM – II when we use GMM to produce estimates. The SUR methods allow for the potential correlation of the regression equation residual terms. The GMM estimation allows for strategies that incorporate the restrictions in a potentially more optimal manner.

Regression equations for effectiveness, cost and net benefit cannot be estimated jointly all together using a simple simultaneous method as their covariance matrix of errors is singular. In other words, with estimates for the parameters for two of the regression equations, one can produce the remaining estimates of the third. As such we report three types of estimates of B, the incremental net benefit. The first comes from calculating the estimate as a function of the ΔE and ΔC estimates from the separate effectiveness and cost regressions (i.e., $\hat{B} = \lambda\, \widehat{\Delta E} - \widehat{\Delta C}$). The second and third estimates come from estimating the net benefit regression simultaneously with either the effectiveness regression equation or the cost regression equation. All analyses were done in R; the OLS and SUR results were verified in Stata. Having described the different methods used to produce estimates and characterize their uncertainty, we now present results.

5.0 Results

This section describes the results from the estimation strategies described in the previous section.

5.1 Equation by equation estimation

The regression results are presented in Tables 1 and 2. Table 1 illustrates the OLS results using robust standard errors.[13] With equation by equation estimation of the cost, effectiveness and net benefit regressions with all of the same covariates (i.e., $\mathbf{X}_i^c = \mathbf{X}_i^e$), it is clear the results seem to differ by race. For the effectiveness equation, the coefficients on the Black_i and $\text{Black}_i \times \text{tx}_i$ variables are not statistically significant. The ACT treatment indicator indicates an increase in stable housing by 98 days. By introducing different covariates for the cost and effectiveness regression equations (i.e., removing the Black_i and $\text{Black}_i \times \text{tx}_i$ variables from the effectiveness regression equation), the ΔE estimate becomes approximately 53 days of stable housing for the same estimated cost savings. The 98 days estimate is for White individuals only and the 53 days estimate is for all individuals. While the coefficient on the treatment indicator variable is statistically significant in both the OLS-I (i.e., $\mathbf{X}_i^c = \mathbf{X}_i^e$) and the OLS-II (i.e., $\mathbf{X}_i^c \neq \mathbf{X}_i^e$) specifications,

[13] The robust standard error is of type HC1, which is the default in Stata.

the estimate of ΔC is only significant at $P < 0.10$. While the ΔE estimate in the more parsimonious specification (OLS-II) is smaller (52.66 days vs. 98.10 days), its statistical significance is greater ($P < 0.01$ vs. $P < 0.05$). From the net benefit regression results of the fuller specification (OLS-I), there is evidence that the estimates of cost-effectiveness do not achieve statistical significance at conventional levels.

5.2 Simultaneous equations estimation

Table 2 shows the results from simultaneous equations estimation both with and without imposing the restrictions (i.e., $\mathbf{X}_i^c \neq \mathbf{X}_i^e$). SUR-I shows the results with $\mathbf{X}_i^c = \mathbf{X}_i^e$. As expected, SUR results with a full specification match those of OLS with a full specification (i.e., SUR-I estimates match OLS-I estimates). However, the statistical significance of the SUR estimates is much greater. This is because of the homoscedasticity assumption, if the errors are heteroskedastic (as our initial test results suggest), then the reported standard errors (assuming constant variance) will be wrong. The SUR-II specification has $\mathbf{X}_i^c \neq \mathbf{X}_i^e$. The effectiveness results for SUR-II match those of the OLS-II specification; however, the cost results for SUR-II do not match those for the OLS-I, OLS-II or SUR-I specifications. While the net benefit regression results are the same no matter how they are derived in the SUR-I setting, they appear to differ in the SUR-II setting.

When the SUR-II coefficients from the cost and effectiveness regression equations (6.1′) and (6.2′) are added to compute the net benefit regression estimates for the column labeled NB(λ=\$10)^a, the coefficient on the ACT treatment indicator is \$51,361. This matches the coefficient on the ACT treatment indicator in a net benefit regression equation when it is jointly estimated with the effectiveness regression equation in the column labeled NB(λ=\$10)^e. However, this coefficient is \$63,729 when the cost and the net benefit regression equations are jointly estimated as shown in the column labeled NB(λ=\$10)^c. The results are identical to SUR-I. When the cost and the net benefit equations are estimated by SUR, the model is no longer over-identified. The results are therefore like OLS.

To summarize, the SUR-II results show that when the effectiveness and net benefit regression equations are estimated together, they produce INB estimates consistent with those from jointly estimated effectiveness and cost regression equations. However, when the cost and net benefit regression equations are estimated together, the SUR-II results match the SUR-I results. In other words, the restrictions (i.e., $\mathbf{X}_i^c \neq \mathbf{X}_i^e$) are not imposed. In our specific case the covariates in the effectiveness regression equation are a proper subset of those in the cost regression equation. However, if the effectiveness equation included a covariate (e.g., age_i) that was not in the cost equation, e.g.,

(6.1′′) $$c_i = \alpha_0 + \alpha_{tx}\, tx_i + \alpha_{Black}\, Black_i + \alpha_{Black\_tx}\, Black_i \times tx_i + v_i$$

(6.2′′) $$e_i = \xi_0 + \xi_{tx}\, tx_i + \xi_{age}\, age_i + u_i$$

(6.3′′) $$nb_i(\lambda) = \beta_0 + \beta_{tx}\, tx_i + \beta_{age}\, age_i + \beta_{Black}\, Black_i + \beta_{Black\_tx}\, Black_i \times tx_i + \varepsilon_i.$$

then jointly estimating the net benefit regression equation (6.3′′) with either equation (6.1′′) or (6.2′′) using SUR would not be an option. To estimate the β coefficients in (6.3′′), what is needed is a way to impose coefficient restrictions like $\beta_{age} = \xi_{age}$, $\beta_{Black\_tx} = \alpha_{Black\_tx}$ and $\beta_{tx} = \alpha_{tx} + \xi_{tx}$ in a situation where residual terms may be non-normally distributed and exhibiting heteroscedasticity of an unknown form.

The GMM methods address these challenges. As noted above, the GMM – I model produces the same results as a the SUR – I model. There are differences in the GMM – II scenario when using efficient GMM without assuming homoscedasticity. Both the estimates and the uncertainty measures differ slightly. This is illustrated in Figure 4 which shows SUR – II and GMM – II results. The difference in findings is most pronounced when it comes to the ICER's estimate and uncertainty. This is evident by the different x-intercepts for the estimate and 95% CI lines. In both cases, the ICER estimate is negative and this presents special challenges, especially for characterizing uncertainty.

5.3 Characterizing uncertainty

Figure 3 shows graphs of the INB estimate by willingness to pay value (λ) from net benefit regressions stratified by race. The solid line is the INB estimate (i.e., $\hat{\beta}_{tx}$). It has a positive slope and a negative x-intercept. This means that the estimate for $\Delta E > 0$, the estimate for $\Delta C < 0$ and the estimate for the ICER < 0. In these situations, it is considered good practice ***not*** to report an ICER (Stinnett and Mullahy, 1998). It is ok to compute estimates of ΔC and ΔE, but the ratio loses key mathematical properties (e.g., transitivity) when it is negative. If one wants to report an estimate of the cost-effectiveness, the INB is a ready alternative. For ICER analyses, in these situations, the main focus switches to characterizing the ICER's statistical uncertainty. As noted in section 2.1, Fieller's theorem sometimes provides a way to express confidence intervals. We use the relationship between Fieller's theorem and the INB illustrated in Figure 1 to show that for African American participants, it is impossible to compute a 95% CI using Fieller's method. The graph to the left in Figure 3 shows the upper and lower confidence bounds for the INB estimate (as dashed lines). It is clear that neither of the dashed lines intersects the x-axis. Consequently, there is no upper and no lower confidence bound produced when using Fieller's theorem. For Caucasian subjects, one of the confidence bounds is negative, again raising concerns about negative ICERs. Once more, the INB appears useful. When studying a new intervention's cost-effectiveness, the INB estimate can have meaning whether it is positive (or negative), indicating the degree to which the extra benefits outweigh (or are outweighed) by the extra costs. In addition, the uncertainty measures for the INB (e.g., the 95% CI) can be used to characterize uncertainty for the ICER as well.

6.0 Discussion

The key question is whether it is better to estimate the INB piecemeal using $\widehat{\Delta E}$ and $\widehat{\Delta C}$ from separate effectiveness and cost regression equations (possibly estimated jointly) to compute $\hat{B} = \lambda\,\widehat{\Delta E} - \widehat{\Delta C}$, or whether it is better to include a net benefit regression in a system of simultaneous equations with a focus on estimating β_{tx}. The goal is to produce an estimate and characterize uncertainty. Methods that optimize and facilitate both tasks are of great value to applied practitioners. This chapter demonstrates the cases when joint estimation represents a potential improvement over independent estimation and when it does not. There are gains to estimating as a system of equations only when the independent variables for the cost equation differ from those of the effectiveness equation (i.e., $\mathbf{X}_i^c \neq \mathbf{X}_i^e$). In these situations, SUR is limited compared to GMM in terms of how to estimate a system of equations. Another key is that incremental net benefit can be estimated directly by jointly estimating net benefit and either cost or effectiveness equations when $\mathbf{X}_i^c < \mathbf{X}_i$ or $\mathbf{X}_i^e < \mathbf{X}_i$, respectively. The identical results for SUR – II reported in Table 2 in the columns NB(λ=\$10)[a] and NB(λ=\$10)[e] can be explained by the fact that the moment conditions implied by the Effect-Cost system are identical to the conditions implied by the Effect-NB system.

Our analysis adds to a sparse literature on cross-sectional applications of GMM (Wooldridge, 2001), with an example of how GMM methods can be useful in statistical cost-effectiveness analysis. With our case study, we have focused on the evaluation of a healthcare intervention. The economic evaluation of other types of interventions or programs (e.g., those for education or the environment) can be produced using the methods we have described. The major advantages about using a GMM strategy in the NBRF is that it produces options for studying data sets with less savory characteristics (e.g., randomization failures, heteroskedastic errors, coefficient restrictions, etc.). This flexibility suggests many promising directions for future research. For example, what is the connection between our findings and those related to strategies for observational (non-randomized) data, covariate specification and joint estimation with non-linear links (e.g., see Mantopoulos et al., 2016)? In addition, future research could explore extensions of these methods into longitudinal or hierarchical data analysis settings. GMM may play an important role in the analysis of cost-effectiveness data in the future.

References

Altman DG. Comparability of randomised groups. The Statistician 1985; 34: 125–136.

Armitage, P. and Berry, G. Statistical methods in medical research, 3rd edition. Oxford: Blackwell Scientific Publications, 1994.

Burns BJ, Santos AB. Assertive community treatment: an update of randomized trials. Psychiat Serv 1995; 46: 669–675.

Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. Health Econ. 1998;7(8):723-40.

Crépon B, Mairesse J. The Chamberlain Approach to Panel Data: An Overview and Some Simulations, in Mátyás L and Sevestre P. eds. The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice, 3rd Edition, Springer-Verlag Berlin Heidelberg, 2008, 113 - 183.

Davidson R, MacKinnon JG (2004): Econometric Theory and Methods. Oxford University Press, New York.

Hayashi F (2000): Econometrics. Princeton University Press, New Jersey.

Hoch JS, Briggs AH, Willan AR. Something old, something new, something borrowed, something blue: a framework for the marriage of health econometrics and cost-effectiveness analysis. Health Econ. 2002 Jul;11(5):415-30.

Hoch JS, Blume JD. Measuring and illustrating statistical evidence in a cost-effectiveness analysis. J Health Econ. 2008 Mar;27(2):476-95.

Fiebig, DG. (2001): Seemingly Unrelated Regression, in Baltagi, B. eds, A Companion to Theoretical Econometrics, Backwell Publishers, 101-121.

Lehman AF, Dixon LB, Hoch JS et al. Cost-effectiveness of assertive community treatment for homeless persons with severe mental illness. Br J Psychiatry 1999; 174: 346–352.

MacKinnon, JG, White H (1985). "Some Heteroskedastic-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties". Journal of Econometrics. 29 (29): 305–325.

Mantopoulos T, Mitchell PM, Welton NJ, McManus R, Andronis L. “Choice of statistical model for cost-effectiveness analysis and covariate adjustment: empirical application of prominent models and assessment of their results.” Eur J Health Econ. 2016 Nov;17(8):927-938.

Moon HR and Perron B. (2006): Seemingly unrelated regressions, in Durlauf SN and Blume LE. eds, The New Palgrave Dictionary of Economics, 2nd Edition, 1-9.

Olfson M. Assertive community treatment: an evaluation of the experimental evidence. Hospital Community Psychiatry 1990; 41: 634–641.

Scott JE, Dixon LB. Assertive community treatment and case management for schizophrenia. Schizophrenia Bull 1995; 21: 657–668.

Severens JL, De Boo TM, Konst EM. Uncertainty of incremental cost-effectiveness ratios. A comparison of Fieller and bootstrap confidence intervals. Int J Technol Assess Health Care. 1999 Summer;15(3):608-14.

Siani C, de Peretti C, Moatti JP. Revisiting methods for calculating confidence region for ICERs. Are Fieller's and bootstrap methods really equivalent? Mimeo, 2000.

Stein LI, Test MA. Alternative to mental hospital treatment: I. Conceptual model, treatment program, and clinical evaluation. Arch. Gen Psychiatry 1980; 37: 392–397.

Stinnett AA, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. Med Decis Making. 1998; 18(2 Suppl): S68-80.

Tambour M, Zethraeus N, Johannesson M. A note on confidence intervals in cost-effectiveness analysis. Int J Technol Assess Health Care. 1998; 14(3): 467-71.

Willan AR, Briggs AH, Hoch JS. Regression methods for covariate adjustment and subgroup analysis for non-censored cost-effectiveness data. Health Econ. 2004 May;13(5):461-75.

Wooldridge JM, "Applications of Generalized Method of Moments Estimation" Jo Econ Persp; 15(4) (Autumn, 2001): 87-100.

Figure 1: Illustration of the relationship between the estimates and uncertainty for the Incremental Cost Effectiveness Ratio (ICER) and the Incremental Net Benefit (INB) as a function of Willingness to Pay (λ)

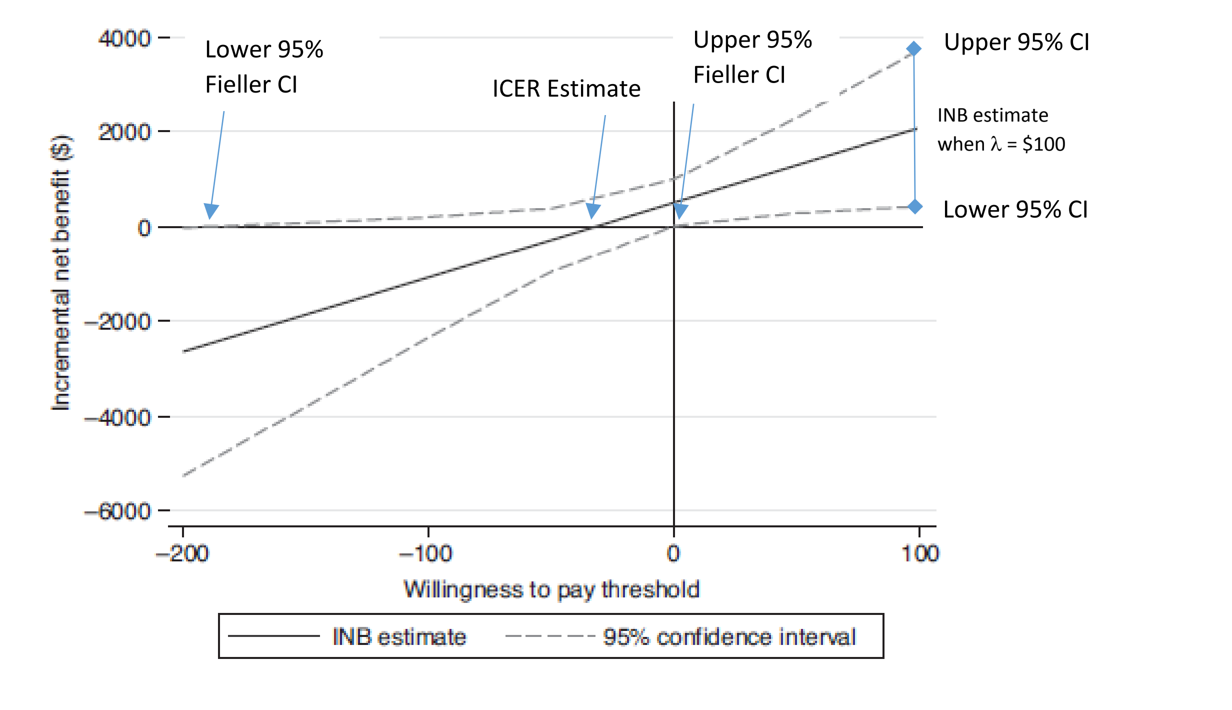


Figure 2: Histogram of the Cost and Effectiveness data after various transformations (including no transformation indicated by Identity)

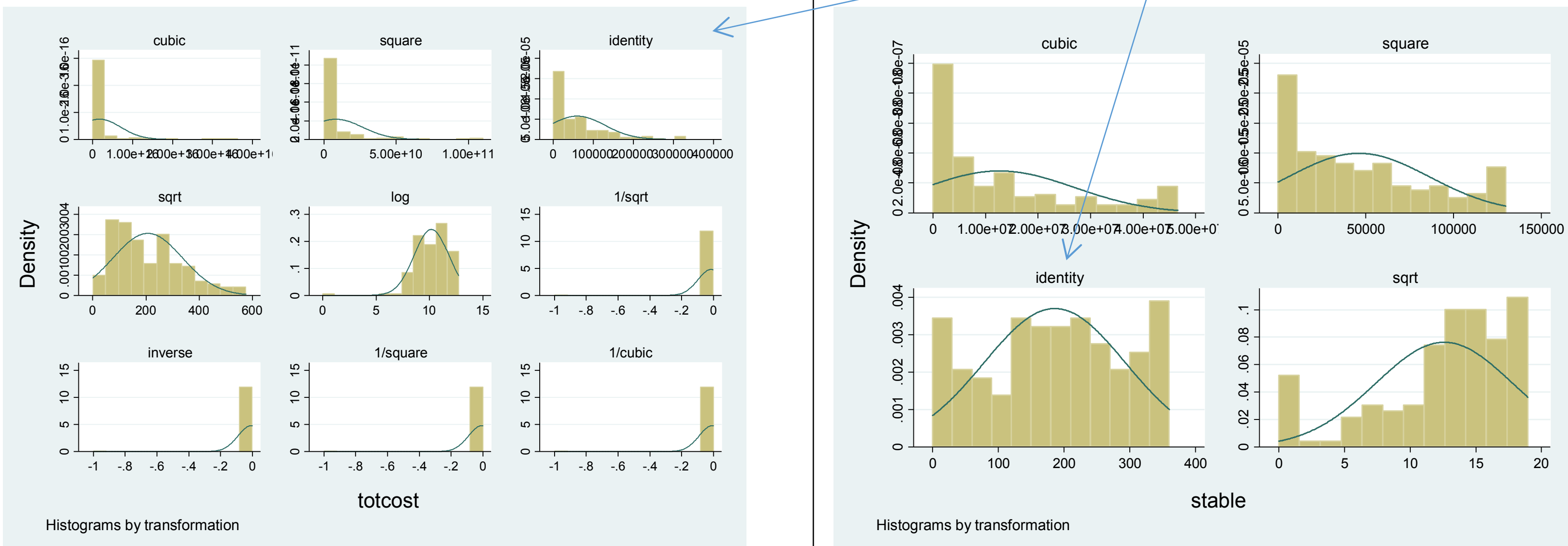


Table A: OLS regression results for the effectiveness and cost equations

VARIABLES	(1) Effectiveness (stable housing days)	(2) Cost (US $)
Tx	98.10***	-62,748***
	(26.65 - 169.55)	(-109,269 - -16,227)
Black	31.92	-53,809**
	(-33.57 - 97.40)	(-96,446 - -11,171)
Black × tx	-62.48	57,676**
	(-144.8 - 19.81)	(4,092 - 111,260)
Constant	132.65***	112,239***
	(72.87 - 192.43)	(73,317 - 151,162)
N	145	145
Adjusted R^2	0.056	0.035
F test p-value	0.0110	0.0469

95% CI in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 1

			Dependent variable			
			Effectiveness (stable housing days)	Cost (US \$)	NB(λ=\$10)[reg]	
	OLS – I[r]	TX	98.10**	-62,748*	63,729*	
	OLS – I[r]	Black	31.92	-53,809*	54,128*	
	OLS – I[r]	Black*TX	-62.48	57,676*	-58,301*	
Equation by Equation Estimation	OLS – I[r]	Constant	132.65***	112,239***	-110,913***	
	OLS – II[r]	TX	52.66***	-62,748*	—	
	OLS – II[r]	Black	—	-53,809*	—	
	OLS – II[r]	Black*TX	—	57,676*	—	
	OLS – II[r]	Constant	159.25***	112,239***	—	

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: [r] = “robust” option used to produce White-corrected standard errors in the presence of heteroscedasticity. The robust standard error used is of type HC1, which is the default in Stata.

[reg] = coefficients from a net benefit regression, $nb_i(\lambda) = \beta_0 + \beta_{tx}\, tx_i + \boldsymbol{\beta}_{\mathbf{x}} \mathbf{X}_i + \varepsilon_i$. Results using robust standard errors of type HC0, HC1, HC2 and HC3 are available from the authors.

.

Table 2

			Dependent variable					
			Effect	Cost	NB(λ=$10)[a]	NB(λ=$10)[e]	NB(λ=$10)[c]	
Simultaneous system of equations estimation	SUR – I[i]	TX	98.10***	-62,748***	63,729***	63,729***	63,729***	
	SUR – I[i]	Black	31.92	-53,809**	54,128**	54,128**	54,128**	
	SUR – I[i]	Black*TX	-62.48	57,676**	-58,301*	-58,301**	-58,301**	
	SUR – I[i]	Constant	132.65***	112,239***	-110,913***	-110,913***	-110,913***	
	SUR – II[i]	TX	52.66***	-50,835**	51,361***	51,361**	63,729***	
	SUR – II[i]	Black	—	-45,441**	45,441**	45,441**	54,128**	
	SUR – II[i]	Black*TX	—	41,294*	-41,294*	-41,294*	-58,301**	
	SUR – II[i]	Constant	159.25***	105,266***	-103,674***	-103,674***	-110,913***	
	GMM – I	TX	Same results as SUR – I					
	GMM – I	Black						
	GMM – I	Black*TX						
	GMM – I	Constant						
	GMM – II	TX	56.74***	-46,1445*	46,712*			
	GMM – II	Black	—	-39,072	39,072			
	GMM – II	Black*TX	—	36,180	-36,180			
	GMM – II	Constant	160.77***	98,372***	-96,764***			

*** p<0.01, ** p<0.05, * p<0.1

Note: [i] = "isure" option specifying iteration over the estimated disturbance covariance matrix and parameter estimates until the parameter estimates converge. Under seemingly unrelated regression (SUR), this iteration converges to the maximum likelihood results. However, these SUR estimates have been calculated under the assumption of homoscedasticity. [a] = addition of the effectiveness regression and cost regression coefficients (e.g., $\lambda \cdot \hat{\xi} - \hat{\alpha}$). [e] = coefficients for both the effectiveness and the net benefit regression equations simultaneously estimated together. [c] = coefficients for both the cost and the net benefit regression equations simultaneously estimated together.

Figure 3: Incremental net benefit by willingness to pay graphs by race

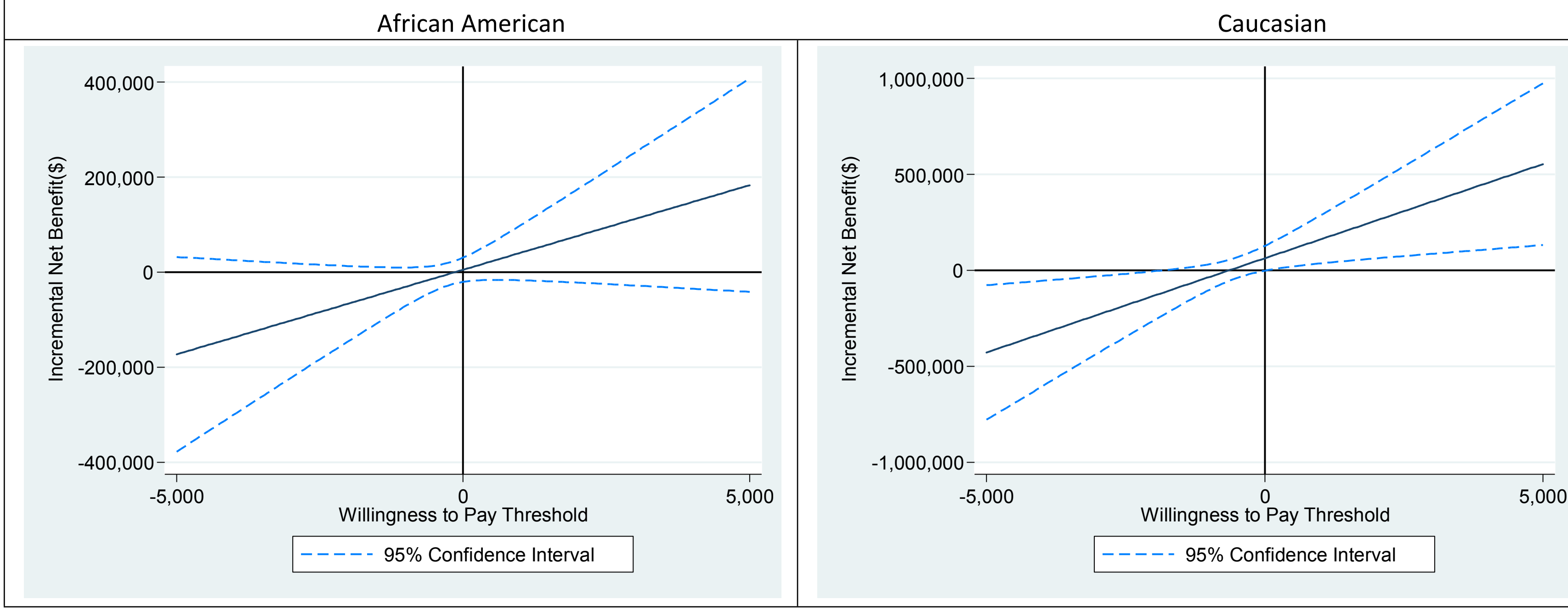
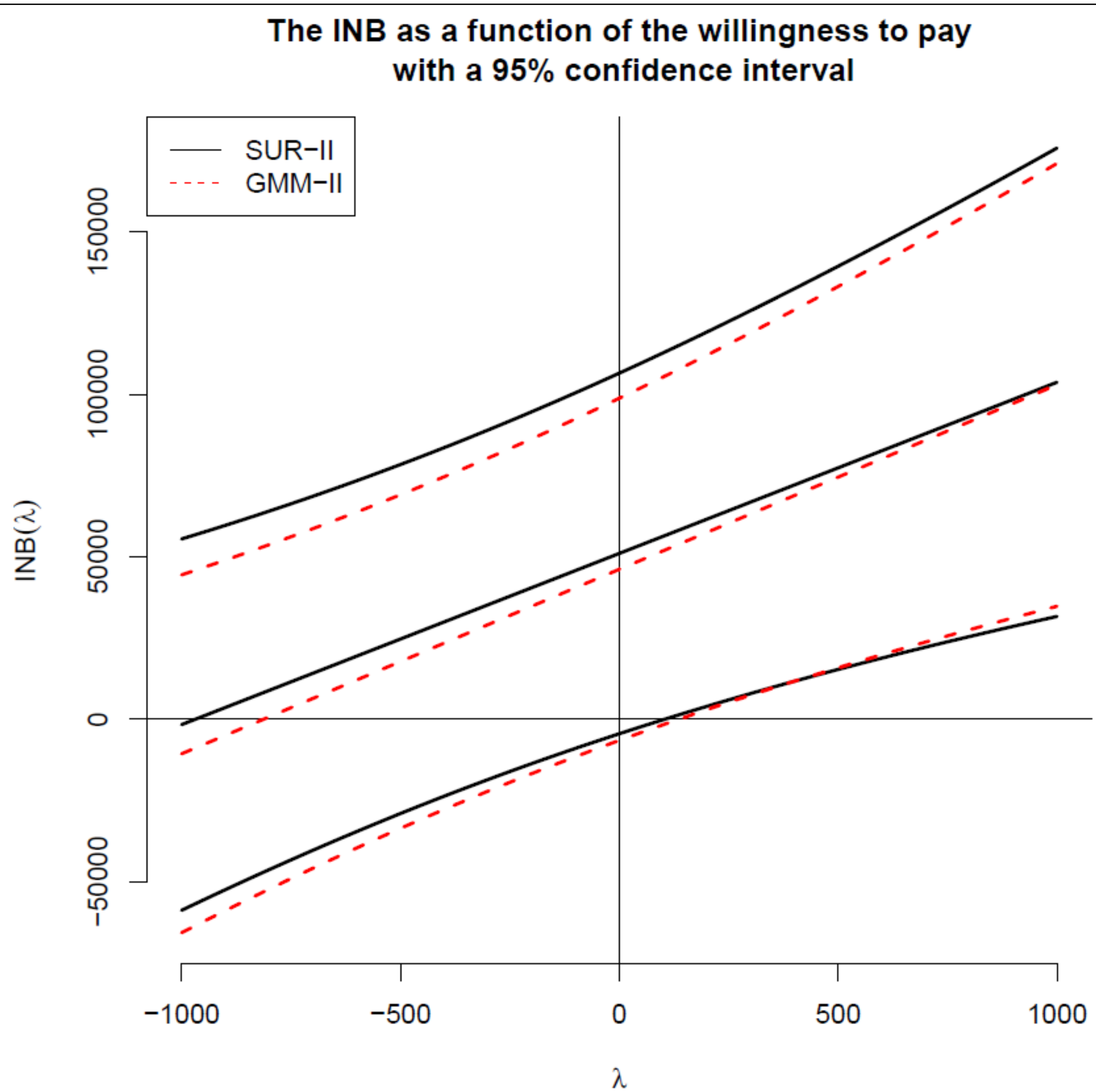


Figure 4


The INB as a function of the willingness to pay
with a 95% confidence interval
SUR−II
GMM−II
INB(λ)
−50000
0
50000
100000
150000
−1000
−500
0
500
1000
λ