CONSIDERATIONS OF EFFICIENCY IN POLICY EVALUATION:

AN APPLICATION TO CHILD WELFARE POLICY

By

Ashley Provencher

Submitted to the

Faculty of the College of Arts and Sciences

of American University

in Partial Fulfillment of

the Requirements for the Degree

of Doctor of Philosophy

In

Economics

Chair:

Mary E. Hansen, Ph.D.

Amos Golan, Ph.D.

Dean of the College of Arts and Sciences

Alison Jacknowitz, Ph.D.

July 21, 2011

Date

Walter Park, Ph.D.

2011

American University

Washington, D.C. 20016

CONSIDERATIONS OF EFFICIENCY IN POLICY EVALUATION:

AN APPLICATION TO CHILD WELFARE POLICY

BY

Ashley Provencher

ABSTRACT

Amartya Sen's capability framework emphasizes that human well-being depends both on the resources available to individuals and on the unique processes by which individuals convert resources into well-being. Although the capability approach has inspired many scholars and policymakers, its usefulness for policy evaluation remains in question. Further, there is no consensus about how to capture the conversion process in an evaluation of policy. In this dissertation, I argue that it is the explicit treatment of the conversion process that makes the capability framework highly useful for the evaluation of policy that aims to improve human well-being. I further show that a well-known econometric technique, stochastic frontier analysis, captures the most salient aspects of the conversion process. Finally, I use the stochastic frontier analysis to evaluate the effectiveness and efficiency of outpatient mental health services provided to children who come into contact with child protective services in the United States.

ACKNOWLEDGMENTS

I thank my family and Mr. Richard and Mrs. Valerie Roberts, who always inspire me. I also thank my friends and colleagues at Simmons College.

I am grateful for the constructive comments, enriching discussion, and continued guidance of my committee members: Professor Mary Hansen, Professor Walter Park, Professor Alison Jacknowitz, and Professor Amos Golan.

This document includes data from the National Survey on Child and Adolescent Well-Being, which was developed under contract with the Administration on Children, Youth, and Families, U.S. Department of Health and Human Services (ACYF/DHHS). The data have been provided by the National Data Archive on Child Abuse and Neglect. The information and opinions expressed herein reflect solely the position of the author(s). Nothing herein should be construed to indicate the support or endorsement of its content by ACYF/DHHS.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF ILLUSTRATIONS

Figure

CHAPTER 1

THEORETICAL DIFFERENCES BETWEEN THE NEOCLASSICAL ECONOMIC FRAMEWORK AND THE CAPABILITY FRAMEWORK

The capability framework, first introduced by Amartya Sen, is a conceptual framework to assess individual well-being. Well-being is defined in terms of the array of opportunities available to a person to actually achieve a physical, emotional, and mental state, i.e. a state of being, given her available resources and abilities to use those resources.

The capability framework redirects economists to evaluate the impact of policies that aim to improve an individual's well-being. Using this framework, policymakers can directly evaluate a policy either on its ability to improve an individual's actual state of being (her functioning) or her ability to achieve a state of being (her capability). The capability framework has already affected policy evaluation in the areas of health (for example, see Anand and Dolan 2005; Hopper 2007; Ruger 2004a; Ruger 2004b; Sen 2002; Verkerk, Busschbach, and Karssing 2001), disability (for example, see Burchardt 2004; Kuklys 2005; Lelli 2005; Nussbaum 2006; Terzi 2005; Zaidi and Burchardt 2005), and education (for example, see Unterhalter 2003; Walker 2005; Walker 2006).

The capability framework is an alternative economic framework to the more widely used neoclassical framework to conduct evaluation of policies that aim to improve individuals' states of being. A policy evaluation using the neoclassical framework instead measures the effect of a policy as its ability to alter the consumption behavior of an individual with resources valued at market prices and serving as a proxy for her well-being (see the founding works of Hicks 1932; Hicks 1934; Hobson 1925; Roll 1938; Stigler 1941; and Veblen 1900). But not all policies affect an individual's consumption behavior. Some policies may only aim to influence an individual's state of being and have no intended effect on an individual's market behavior.

For instance, the Adoption and Safe Families Act of 1997 explicitly charges the U.S. child welfare system with improving the well-being of children in foster care (Wulczyn, Barth, Yuan, Harden, and Landsverk 2005). While it is possible to evaluate the policy based on resulting changes in the behaviors of the state and its agents, such as foster parents and adoptive parents, to measure the effect of the public policy, a more explicit valuation would estimate the well-being of a foster child. In the cases where a policy aims to improve an individual's well-being, the capability framework can be used to evaluate policies based on their stated mission rather than relying on instrumental proxies to measure the effect of a policy.

Differences in the conceptualization of well-being between the capability framework and neoclassical framework can result in different measured effects of a policy and policy prescriptions. Until now, the literature on policy outcome evaluation has not thoroughly reviewed these differences, nor has the capability literature emphasized these differences. As a result, researchers are skeptical as to whether the capability framework offers anything for policy outcome evaluation different from the neoclassical framework (see Sugden (1993); Ysander (1993); Srinivasan (1994); and Roemer (1996)).

In what follows, I survey policy evaluations conducted by researchers using each framework and compare the findings of these studies in order to demonstrate the divergence in policy prescriptions that results from the application of each framework. I show that in some instances the theoretical frameworks have different implications for the most effective policy response.

A Brief Review of the Neoclassical Framework

The objective of policy outcome evaluation is to measure the impact of economic and social policies on the welfare of individuals and society. The neoclassical framework, the more widely used framework for policy outcome evaluation, assesses a policy based on its measured effect on individual or social utility. In this section, after a brief review of the neoclassical framework, I highlight the fact that the neoclassical framework does not capture the ability of individual to transform resources into well-being.

Policy evaluation is a subset of literature within the neoclassical paradigm that is concerned with the measurement of the effectiveness and efficiency of policies (see Head (2008) and Rossi and Williams (1972)). This literature focuses on three types of evaluation: outcome evaluation, cost-benefit evaluation, and process evaluation. While each of these strands of literature exhibits some similarity to the capability framework, none fully capture the theoretical underpinnings of the capability framework. Most obviously, the policy evaluation literature assumes an individual efficiently utilizes resources in order to maximize utility subject to their resource constraint but efficient use of resources is not assumed in the capability framework.

In the neoclassical framework, the individual *i* is presumed to make decisions by maximizing her utility *u* subject to her budget constraint. Her budget constraint is dependent on her exogenous income *m*, the prices of goods *p*, and her consumption of goods *x*:

$$max\, u_i = u(x) \quad s.t.\ px = m_i \quad where\ i = 1, \ldots, n. \qquad (5)$$

For simplicity, assuming there is perfect competition in the marketplace and any good can be purchased in the market place, then prices are the same for all individuals and all market and non-market goods have some market price. Note that monetary resources are the primary factor constraining an individual's utility maximization, other than the individual's utility function, which describes her preference rankings.[1]

Assuming local nonsatiation, the individual will choose to consume the bundle of goods *x** that maximizes her utility and fully expends her monetary resources given a level of prices and income, *px*=m*. Now, the consumer's maximization problem is:

$$v_i = v(p, m_i) = max\, u(x) \ \ s.t.\ px = m \qquad (6)$$

where the indirect utility function *v* measures the maximum utility achievable at given prices *p* and income *m*.

---

1. It is conceivable that a different preference ranking could result in a different level of achieved utility given the same market prices and income. Conceptually, the individual would maximize a different utility function subject to the same market prices on goods and services and the same income. Of course, any new ranking must still conform to the assumptions of revealed preference theory.

Within the neoclassical framework, the indirect utility function can be used to estimate the effect of a policy as the change in utility before and after the imposition of a policy. The effect of the policy π on individual welfare can be estimated as[2,3]:

$$d\mathrm{v_i} = \frac{\partial \mathrm{v}}{\partial \mathrm{m_i}}\frac{\partial \mathrm{m_i}}{\partial \pi}. \qquad (7)$$

A policy increases individual well-being if she is able to achieve a higher level of utility after the policy is imposed. (See table 2 for a summary of measured policy effects.)

Table 1. Measured Policy Effects by Framework

Framework	Measured effect
Neoclassical thought	$v_i = v(p, m_i) = \max u(x)\ \ s.t.\ \ px = m$ $dv_i = \frac{\partial v}{\partial m_i}\frac{\partial m_i}{\partial \pi}$
Capability framework	$b_i = f_i(c(x_i)\mid z_i, z_s, z_e)\ ,\ f_i(\cdot) \in F_i\ and\ \forall\ x_i \in X_i$ $B_i = \{b_i \mid b_i = f_i(x_i), \quad \forall\ f_i(\cdot) \in F_i\ and\ \forall\ x_i \in X_i\}$ $db_i = \frac{\partial f_i}{\partial x}\frac{\partial x}{\partial \pi}d\pi + \frac{\partial f_i}{\partial z}\frac{\partial z}{\partial \pi}d\pi$

Conceptually, the measured change in consumption of a good for a price change correlates with an examination of the shape of the demand curve for that good, or its elasticity. For instance, suppose the stated goal of a policy is to increase the consumption of a normal good *x*, denoted Q_x. The government will subsidize the cost of *x* to effectively lower the price P_x. In figure 1, this is shown as a counterclockwise pivot of

---

2. Four assumptions are necessary to estimate the effect of a policy change: the social utility function is differentiable; the social utility function is continuous; the social utility function is separable; and the social utility function has ratio scale measurability.

3. This discussion follows Kuklys (2005, 13-14).

the budget constraint such that relatively more of good *x* can be consumed at each level of consumption of all other goods (AOG). If the quantity of *x* consumed increases from *x* to *x'*, the subsidy is evaluated as successful. The subsidy is evaluated as unsuccessful if people consume the same amount or less of *x* and consume more AOG. The subsidy increases the welfare of the individual regardless of whether the subsidy is evaluated as successful or unsuccessful. That is, we could conclude that the subsidy was not effective at meeting its goal even if consumers have higher utility because of the income effect. Moreover, the neoclassical framework does not consider the effect of the subsidy on the opportunities for the individual to achieve higher levels of utility (i.e. the size of the area between the budget constraints in figure 1).

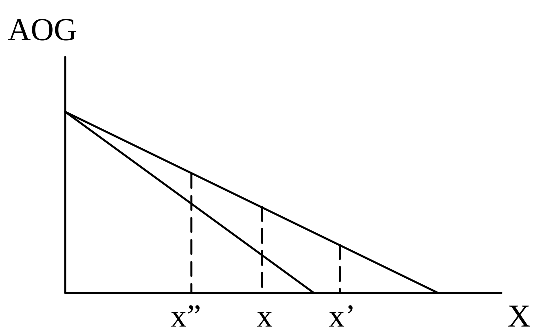

Figure 1. Effect of a Policy Change in the Neoclassical Framework.

Social welfare *W* can be measured using the social utility function *G*. The social utility function *G* is dependent on the consumption of goods of each of its members, i.e. the utility of each individual, where individual indirect utilities are weighted:

$$W = G\big(v_1(p, m_1), \ldots, v_n(p, m_n)\big). \quad (8)$$

The effect of a policy on social welfare *W* therefore is measured as the change in income of each individual:

$$dW = \sum_{i=1}^{n} \frac{\partial W}{\partial G} \frac{\partial G}{\partial m_i} dm_i = \sum_{i=1}^{n} \beta_i \, (m_i) dm_i \quad (9)$$

where $\beta_i(m_i)$ is the marginal social utility of income m_i. If marginal social utility of income is constant and equal across individuals so $\beta_i(m_i)=1$, then:

$$dW = \sum_{i=1}^{n} dm_i. \quad (10)$$

The effect of a policy change on social welfare is equal to the sum of the partial derivatives with respect to the social utility function.

Sen's Critique of the Neoclassical Framework

Sen has argued that the neoclassical framework incompletely models individual and social welfare. Using three primary critiques, Sen highlights the limitations of the neoclassical framework to assess individual and social welfare: its emphasis on the consequence of an act and disregard for the motivation of an act; its aggregation of individual welfare rankings to acquire a social welfare ranking; and its reliance on monetary resources to gauge individual welfare. Below, I present Sen's critique and explain how his conceptual framework, the capability framework, attempts to address each of these criticisms.

First Criticism: Fails to Consider the Reason for Choice

The neoclassical framework relies heavily on assumptions about preferences and choice so that they may disregard any considerations of why an individual chooses a particular element from a menu of options.[4] In doing so, the neoclassical framework underscores the consequences of choice without considering the deliberation for choice. But the process by which preferences develop may also significantly affect the level of utility gained from a consumption bundle. For instance, an individual's history of consumption may influence the marginal level of happiness she experiences from consuming the next bundle solely because extensive exposure to certain influences or situations has altered her preferences. Such preferences are called adaptive preferences.[5]

The neoclassical framework relies on the Weak Axiom of Revealed Preference (WARP) and Strong Axiom of Revealed Preference (SARP) to guarantee that preferences are revealed in a given choice situation (see appendix A). Unfortunately these assumptions are not strong enough to make inferences across all choice situations. In order to more completely model choice behavior, two additional assumptions are necessary to derive implications from choice functions (Sen 1970, 16-20): contraction consistency and expansion consistency.

4. Appendix A lists the familiar assumptions of the neoclassical framework.

5. The neoclassical framework considers adaptive expectations in macroeconomic models (see Fisher (1930)). However, models of individual choice behavior do not account for adaptive expectations or adaptive preferences. It is important to note that expectations and preferences are not synonymous. Preferences are individual tastes for outcomes while expectations influence an individual's predictions of future outcomes.

Contraction consistency asserts that the removal of any element from a set does not affect an individual's preferences over the remaining elements. Consider an element *x* that is contained in two sets, *S* and *T*, where set *S* is contained within set *T*. It must be true that element *x* will be chosen from set *S* if the element *x* is chosen from set *T*. Suppose *x* is chosen from one choice set *S*. Then if *x* is also an element in a choice set *T* and *T* is a proper subset of *S* (every element in *T* is also in *S*), *x* must be chosen from set *T*.

Expansion consistency assumes that expansion of the choice set does not affect the ranking of elements within the original set. Consider two elements *x* and *y* that are contained in two sets *S* and *T* where set *S* is contained within set *T*. If element *x* is chosen over element *y* in set *S*, then the element *x* must also be chosen over element *y* in set *T*. Moreover, if an element *x* is chosen from every set in a particular class, it must also be chosen from their union.

While contraction consistency and expansion consistency can be assumed to more completely model choice behavior, such assumptions may be too strict.[6] There are three reasons why these assumptions may fail. First, the assumptions of contraction and expansion consistency may not hold if the decision rule relies on the position of the element in some ordering of the elements. Let person *i* choose between taking the last apple *y* or having nothing *x*. Person *i* decides not to take the last apple as she feels this would be rude. However, if the basket had two apples *y* and *z*, then she would have chosen *y*.

6. Empirical studies in behavioral economics have demonstrated the failure of these properties (Johnson and Mathews 2001; Kalai, Rubinstein, and Spiegler 2002; Miguel, Ryan, and Amaya-Amaya 2005; Sippel 1997).

Second, these properties will be contradicted if the information influencing the decision rule changes because of the addition or exclusion of elements. Let person *j* choose between a large slice of cake *x* and a smaller slice *y*. Person *j*, adhering to socially acceptable behaviors, does not want to appear greedy by taking the largest slice so he chooses *y*. His choice would still be consistent with his principle not to appear greedy if he chose *x* from the set *{x,y,z}* and *z* is an even larger slice than *x*.

Finally, the decider must have the ability to reject alternatives. For instance, an individual might select to be nourished or malnourished. We cannot deduce if the individual selected to fast or was starved of nutrition in the case of malnourishment. All three failures can be averted if we observe some external object of choice to infer how an individual will choose. These failures further demonstrate that an individual's decision rule is determined by her motivation for the choice. In order to determine if the individual has made an irrational choice or an inconsistent choice, we must consider the context and the individual's motivation (Sen 1993; Sen 1995; Sen 1997b; Sen 1999b). We might judge some observed choice to be irrational using standard axioms of choice but no longer denote this as irrational once we understand the individual's reason for the choice. Sen argues, for this reason, that the standard axioms of rational choice theory incompletely capture real choices. A more complete framework for welfare would consider both an individual's level of well-being and the process by which she achieves well-being.

Dowding (2002) argues that the standard axioms of revealed preference are theoretically justified and are used to make inferences within the modeling process. Revealed preference analysis allows researchers to measure marginal elasticities and changes thereof without considering the motivation behind individual choices (Dowding 2002, 276-277). The axioms allow us to identify testable hypotheses. Empirically, researchers acknowledge that an individual is unlikely to be faced with exactly the same choice set under precisely the same conditions. If she were, she would choose the same consumption bundle. Preferences are assumed to be explained by the correlations between the structural variables in aggregate data analysis (Dowding 2002, 279).

Still, revealed preference theory evaluates each individual's circumstances in terms of her own system of values but never identifies those values. Even with consistency of choice, revealed preference theory is problematic since choice is rooted in each individual's values. Consider an individual who chooses to fast because of religious faith and an individual who is anorexic. Both individuals choose not to eat but we know nothing about each individual's value for food or value of not eating.

To improve on the neoclassical framework, the capability framework directly accounts for the individual's process to utilize resources and achieve a state of well-being. The capability framework models an individual's utilization of resources, emphasizing that the individual's abilities can be limited by personal factors or external factors such as her environment or social norms. (The individual's model in the capability framework is presented later.)

Second Criticism: Summed Individual Welfare Functions Does Not Equate With Social Value

Sen argues that it is permissible to identify affordable consumption bundles given a particular amount of resources but it is unacceptable to assume that individuals will achieve the same level of utility from the consumed bundles. An individual with seemingly consistent choice behavior over similar menus may not have congruent utilities even if we can ensure the consistency of choice over menus. This has critical implications for how we understand demand functions.

Demand functions are determined by individual utility achievements for particular bundles of goods. But individuals with the same demand function do not necessarily have the same interpersonal utility for any given commodity bundle (Sen 1997a, 392). For instance, “even if a person who is disabled or ill or depressed happens to have the same demand function as another who is not disadvantaged in this way, it would be quite absurd to assume that she is having exactly the same utility or well-being from a given commodity bundle as the other can get from it” (Sen 1997a, 392). Assuming individual utilities are comparable effectively ignores fundamental characteristics, such as health status, which may affect the well-being of an individual.[7]

Market demand is the result of the aggregation of individual demand functions, and therefore individual preferences. Individual demand curves represent an individual’s demand for a particular good at a given market price. The derivation of an individual demand curve is contingent on the amount of marginal utility that the individual receives from consuming a particular amount of the good, her budget constraint, and market prices. An individual maximizes her utility subject to her budget constraint. In the case of

---

7. A neoclassical economist might posit $u(c, A) = Ax\alpha$, where c is consumption and A is technical efficiency of consumption, where A is a function of good health (not disabled) and the environment to get at a similar analysis. But this analysis would be limited to the level of utility achieved given the maximization problem. In this regard, utility is comparable to functioning achievement. The capability framework, however, measures individual well-being as the set of functionings available to the individual given her resources, environment, and conversion efficiency, similar to the consumption space in the neoclassical framework.

two goods (*x* and *y*), the individual maximizes her utility when the marginal rate of substitution between *x* and *y* is equal to the price ratio of the two goods. The marginal valuation of utility does not inform us of the absolute valuation that an individual places on a particular good. Utils have no intrinsic value; they are ordinal. Additionally, the neoclassical framework does not consider the individual's motivation for such a ranking. Aggregating independent individual values may not, and often will not, produce a meaningful understanding of the social value of a particular bundle.

Moreover, the sum-ranking of individual utilities ignores personal distribution of utility (Sen 1973, 16). As previously noted, the amount of resources necessary to achieve a given level of utility may vary across individuals. Failing to account for these distributional differences also ignores differences in the opportunities available for each individual as a result of these resources. The capability framework stresses this view by assuming individuals have different abilities to utilize resources. Differences in individuals' abilities are shown to lead to differences in individuals' opportunities.

Third Criticism: Monetary Resources Do Not Completely Measure Well-Being

Even if the assumptions of individual preferences and values are overcome, Sen does not believe that measuring resource consumption is sufficient to gauge an individual's level of well-being. How the individual uses resources and what the individual achieves from using these resources are more indicative of her well-being.

Individuals can use various combinations of resources in order to achieve the same state of being. Monetary resources only provide individuals with the means to achieve functionings and have no intrinsic value (Sen 1997a, 393). Sen identifies five reasons why the conversion of monetary resources into states of being is not consistent across individuals: personal heterogeneities in personal characteristics; environmental differences; variations in social climate; variation between communities; and intra-household distribution of resources (Sen 1997a, 385-386). These heterogeneities are why monetary resources are not a sufficient indicator of well-being.

The capability framework incorporates heterogeneity in the conversion process through conversion functions and conversion factors. Both influence an individual's level of functioning and thereby her capability set. The capability framework therefore directly addresses how variation in the conversion of resources may result in different levels of well-being across individuals.

An appropriate framework for policy outcome evaluation would directly measure achieved states of being especially given the complexities by which they are achieved. Some people may require more monetary resources than others in order to achieve the same level of well-being. For instance, individuals may require different amounts of income to buy sufficient goods to achieve the same level of social functioning as 'appearing in public without shame' in the capability of 'taking part in the life of the community', particularly in cross-country analysis where countries may have different levels of income (Sen 1992, 115).

The capability framework does not completely dismantle or discount income measures as proxies for well-being. Indeed, in some cases, income proxies may be highly associated with functioning achievements. The capability framework instead offers economists a broader framework in which to consider the impact of policies on the well-being of individuals using direct measures of well-being.

The Capability Framework

I describe the theoretical structure of the capability framework in this section, and derive theoretically the effect of a policy on the well-being of an individual. The predicted effect highlights the framework's fundamental assumption that an individual's well-being is related to her ability to utilize resources.

The capability framework conceptualizes well-being as a multidimensional outcome that relates to an individual's freedom to achieve a state of being. In choosing a state of being to achieve, an individual chooses among functioning vectors. Functionings are the various achievements that an individual values and has the opportunity to achieve (Sen 1992), such as good health, adequate shelter, or sufficient nourishment. An individual's level of functioning is her actual state of being. (For the reader's convenience, table 1 gives definitions of the key terms that appear throughout this dissertation.) An individual achieves a level of functioning as the result of utilizing resources. The capability framework does not assume that an individual efficiently utilizes resources in order to achieve her best level of functioning, only that she values the achieved level of the functioning vector

Table 2. Definitions of Key Terms

Term	Definition
Functioning	A state of being that an individual values and has the opportunity to achieve (Sen 1999a, 75)
Capability set	An individual's set of feasible functioning vectors; This set describes an individual's opportunities to achieve well-being (Sen 1992, 40)
Agency	An individual's ability to act on behalf of what she values (Sen 1985, 206)

Following Sen (1985), let x_i be a vector of commodities possessed by individual i and f and c be conversion functions that describe the transformation of the commodity vector into a vector of functionings b_i. The conversion function c describes the ability of individuals to utilize resources x given the characteristics of those resources. Since the characteristics of resources are the same for all people, the conversion function c is not unique for individual i. The conversion function f describes the particular abilities of individual i to convert the commodities into some level of functioning b:

$$b_i = f_i(c(x_i)|z_i, z_s, z_e)\ , f_i(\cdot) \in F_i\ and\ \forall\ x_i \in X_i \qquad (1)$$

The function $f_i(\cdot)$ is a member of the set F_i which contains all possible ways a person might transform the given commodities. Conversion functions may vary across individual if each individual has a different technology to convert her resources into functionings. An individual's education or physical attributes (e.g. disability) may impact her technology.

Conversion factors also limit the conversion function f. Conversion factors may be attributed to the individual z_i, society z_s, or the environment z_e. Individual conversion factors include gender, age, race, and physical disabilities. Social conversion factors include property rights, population density, and institutional norms. Environmental influences include climate, pollution, and geography.

At any given point of time, an individual achieves some level of well-being which can be described by a particular functioning vector. The capability framework does not assume that an individual chooses the functioning vector which is good for her, only that she values the functioning vector.

Sen asserts the act of choosing has intrinsic value. The simple act of being able to choose between two or more goods or outcomes enhances an individual's well-being. But, the absence of choice may not necessarily reduce individual well-being. Consider three scenarios in which a high school student applied to universities for admittance and receives a notification from each school regarding her application. She prefers most to attend Princeton University. In the first scenario, she receives an acceptance letter only from Princeton University. In the second scenario, she receives an acceptance letter from Princeton University and Harvard University. In the final scenario, she receives an acceptance letter from Princeton University, Harvard University, and Yale University. Using the neoclassical framework, the student's utility is the same in all three scenarios – she prefers to attend and she received a letter of acceptance from Princeton University. The neoclassical framework only measures her utility from the outcome she chooses. The neoclassical framework does not explicitly consider the menu of options available to the individual in determining her utility. The capability framework, in contrast, argues that the menu of options available matters and the act of choosing has intrinsic value (see Pattanaik and Xu (1990)). In this example, the student has the greatest freedom of choice in the third scenario since she received letters of acceptance from more schools. Also, the individual only chooses in the second and third scenario so her well-being should be higher in these scenarios compared to her well-being in the first scenario.

Each individual's set of feasible functioning vectors is called her capability set. A capability set describes the individual's opportunities to achieve well-being. These opportunities can be alternatively thought of as the possible states of being that an individual can achieve (Sen 1992, 40).

For a given vector of commodities $\overline{x_i}$, the set of feasible functionings A_i is determined by the set of conversion functions F_i:

$$A_i = \{b_i | b_i = f_i(\overline{x_i}),\ for\ any\ f_i(\cdot) \in F_i\}. \quad (3)$$

Denoting an individual's budget set by X_i, the set of feasible functionings is then given by

$$B_i(X_i) = \{b_i |\ b_i = f_i(c(x_i)|z_i, z_s, z_e)\ \forall\ f_i \in F_i\ and\ \forall\ x_i \in\ X_i\} \quad (4)$$

where the set B_i reflects the capabilities of the i^{th} individual. Capabilities depend on the individual's command over commodities (the set X_i) and her ability to transform commodities into functionings (the set F_i).

The capability framework recognizes that an individual's choice of functioning from the capability set is influence by her ability to act on behalf of what she values. The capability framework refers to this ability as the individual's agency (Sen 1985). Agency is dependent on individual circumstances, interpersonal relations, social conditions, contexts, and arrangements, and political and civil rights. Any of these factors may reduce or enhance an individual's agency. This might occur when people are unable to exert agency when they are alienated from their behavior[8], coerced into a situation, submissive, or passive (Ryan and Deci 2004).

Note that agency need not advance well-being. Agency will only advance well-being if the goals that the individual thinks are important are tied to higher levels of

8. An individual is alienated from her behavior if she behaves in a way that she feels she has to instead of behaving in a way that she wants to. In most cases, the individual has considered a given behavior and has rejected it because she deems it undesirable or not worthy of pursuing. The individual may continue to desire the behavior but act differently despite her best efforts. For instance, an individual who is addicted to drugs but no longer wishes to consume drugs will alienate herself from her behavior if and when she consumes them (Frankfurt 1971, 17).

well-being. Indeed, some goals may reduce well-being but improve agency (once the goal is achieved). Because of this, there are ambiguities and potential conflicts in the formulation of agency. Individuals may or may not actually value that which is socially valued, or they may value things that are detrimental to others. For instance, the capability framework can be used to understand why an individual commits violent criminal acts. An individual who murders another person may value his own life over the victim's and may exercise his ability to commit the act regardless if social values dictate that murder is wrong.

A policy also may influence an individual's choice of a functioning from her capability set. The effect of a policy π on an individual's well-being can be measured by the evaluation of b_i (following from equation (3) and (4)) [9] :

$$db_i = \frac{\partial f_i}{\partial x}\frac{\partial x}{\partial \pi}d\pi + \frac{\partial f_i}{\partial z}\frac{\partial z}{\partial \pi}d\pi \qquad (11)$$

where the first term measures the impact on goods and the second term measures the impact of the policy on conversion factors.

As is shown in the next section, Sen addressed each of his criticisms of the neoclassical framework in the developing the capability framework. The capability framework models well-being differently from the neoclassical framework, and so the predicted effect of policy on an individual's well-being also diverges across frameworks.

9. Sen does not explicitly assume any functional form of the conversion function. However, in empirical analyses, researchers often explicitly or implicitly assume some functional form such as continuous or differentiable.

Comparing the Predicted Effect of a Policy across Frameworks

Sen's first critique of the neoclassical framework is its inability to consider an individual's reason for choice. In contrast, the capability framework considers the process to achieve a level of well-being, in addition to the individual's level of well-being, using functionings and conversion functions. This is an important difference between the frameworks that contributes to differences in their predicted effect of a policy. In the context of policy outcome evaluation, both an individual's level of well-being and conversion function are determinants of the effect of a policy (see equation (11)). The conversion function can be used to determine all possible levels of functionings that an individual could achieve given her resources and conversion factors. These possible levels of well-being comprise the individual's capability set. The conversion function captures and individual's deliberation for choice indirectly since any decision rules will influence and individual's possible levels of well-being. However, the capability framework does not specify a method for identifying which level of functioning an individual will achieve given her capability set. The framework only goes so far as to require that the individual values the level. Since individuals may have different values, there is no guarantee that any two individual with the same capability sets will achieve the same level of functioning.

Moreover, since capability sets do not depend on individual preference or valuation (individual preference only influences which functioning an individual chooses from her capability set), interpersonal comparisons are possible using capability sets. Sen addresses his second criticism of the neoclassical framework by directly accounting for distributional differences in resources across individuals, and the ability of the distribution to affect individual well-being. Failing to account for distributional differences of resources and abilities to use resources across individuals also ignores differences in their opportunities to achieve some level of well-being as a result of resources. The individual conversion function and conversion factors (see equation (11)) capture these distributional differences.

Sen's capability framework does not rely soley on monetary resources to understand the effect of policies on individual well-being. The capability framework highlights conversion factors as a contributing factor for individual well-being. A policy has the ability to influence an individual's well-being without influencing her level of resources by altering the individual's conversion factors. (Shown in equation (11) as the marginal effect of a policy π on an individual's conversion factors z.)

For instance, one goal of paternity leave policies is to encourage fathers to stay at home with their children, altering the social norm that mothers should stay at home and care for children while fathers remain in the workforce as the bread winner for the family. Paternity leave policies have the ability to affect the conversion factors of fathers and mothers in their ability to achieve well-being. Fathers might be happier if they develop stronger bond with children through increased interaction; or mothers might gain increased confidence by spending more time in the workforce and contributing a larger percent of monetary resources to the household income. Even still, if mothers and fathers do not alter the time they spend at home, the enactment of paternity leave policies may improve the well-being of parents by removing the social stigmas associated with gender roles and parental child care. Fathers might feel less social pressure to work and mothers may feel less social guilt about returning to the workforce.

Both the capability framework and neoclassical framework aim to model an individual's use of resources to achieve an outcome. The frameworks differ in how they conceptualize the process by which an individual utilize resources and the way factors influence an individual's utilization of resources. The measured effect of policies that alter resource availability or resource utilization will differ across the frameworks because of differences in their conceptualization. This is particularly true for policy outcome evaluation of welfare policies and human development policies.

While the consumption of goods is a determinant of the effect of a policy in both the capability framework and neoclassical framework, the measured effect of a policy using the capability framework explicitly accounts for personal heterogeneity in the conversion of the goods into well-being. In the capability framework, two individuals with the same resources need not achieve the same functioning nor have the same capability set. Individual conversion functions and conversion factors may result in different levels of functionings and capability sets.

Consider providing the same education to a woman and a man in a country where gender norms stipulate that a woman cannot participate in the labor market. Given the same resources (e.g. education), the man and woman might achieve different levels of functioning and therefore have different capabilities in the capability space of economic well-being. In this example, only the man has the freedom to utilize his education to improve his economic well-being through increased labor market wages.

Policy evaluation will have similar conclusions of the effect of a policy regardless of whether the capability framework or the neoclassical framework is used when that policy has no effect on the individual's conversion factors and when the individual's conversion function can be described as an indirect utility function. In these cases, the individual's ability to utilize resources will not have changed and her she will value both utilizing resources efficiently and maximizing her utility so that she achieves the highest number of utils given her budget constraint.

Results of a policy evaluation differ across frameworks because it is often the case that policies do in fact alter an individual's ability to utilize resources and thereby alter her conversion factors. Also, an individual's conversion function and utility function may differ across contexts. For instance, an individual might choose to fast for religious reasons even though food is abundant. She would choose not to utilize resources efficiently in order to achieve some valued outcome in this case.

Researchers might posit that the capability framework does not offer anything different theoretically that cannot already be captured in the neoclassical framework; they argue that the capability framework simply provides a more complex framework in which to evaluate the effect of polices on individuals. However, the variability in results across empirical applications seem to suggest that the complexity of the capability framework better captures the intricacies of an individual's use of resources, achieved level of well-being, and the constraints on her abilities that persist and may in fact contribute to the effectiveness of a policy.

A Survey of Empirical Work Operationalizing the Capability Framework

Few studies that operationalize the capability framework clearly explain the divergence in policy prescriptions in comparison to the prescriptions of the neoclassical framework. In what follows, I survey the literature which operationalizes the capability framework in five domains of research. In each domain of research, I show that the estimated effect of a policy and the policy prescription that follows from the empirical analysis differ across framework. The differences in the estimated policy effects are attributable to the impact of the policy on the conversion function and conversion factors as shown in equation (11).

Macroeconomic Policies: Human Development

Sen (1985) first demonstrated the strengths of the capability framework by comparing traditional macroeconomic measures of well-being to functioning achievements across countries. Sen (1985) used 1980 data on the functionings of life expectancy, infant mortality, and child death rates among Brazil, China, India, Mexico, and Sri Lanka. He ranked countries according to their achievement of each functioning and their gross national product (GNP). According to Sen's theory, developed economies should have high life expectancies and low infant mortality and child death rates. While Brazil and Mexico had higher levels of GNP per capita – approximately seven times higher – than the other countries, both countries performed poorly across the functionings. Life expectancy, infant mortality, and child death rates were best in Sri Lanka, more favorable in China compared to India, and better in Mexico compared to Brazil. Country rankings based on GNP per capita ultimately were different from the rankings based on functionings. GNP per capita did not completely convey the level of development for all countries. While both measures provide a sufficient indicator of production in formal markets, they do not explain resource distribution, inequality, informal market activity[10], social conditions, government strength, or diversification across sectors. Information on these particular dimensions would present a more accurate account of the quality of life within a country and, consequently, its level of human development.

The human development index (HDI), established by the United Nations Development Project (UNDP), was the first measure formally adopted that is based on the capability framework. The HDI attempts to measure a country's actual level of development in comparison to its potential level of development. The index underscores the multidimensional nature of human development as it is a composite index that measures a country's average achievements in three dimensions of human development: health, knowledge, and income. Since 1990, the UNDP has released an annual report that measures a HDI for each country and ranks the countries according to this measure and their gross domestic product (GDP) per capita. Positional rankings of country rankings varied depending on whether countries are ranked using GDP per capita or HDI. The set of measured functionings changes each year.[11]

10. Informal markets can be a large portion of an economy, particularly in low-income countries. (See Pratap and Quintin (2006) for further discussion.) Gross domestic product (GDP) and gross national product (GNP) only measure formal market activity. Thus in countries where informal markets comprise a substantial amount of market activity, GDP per capita and GNP per capita fail to capture a country's level of human development and economic development.

11. The UNDP has since created three additional indices: the Gender-related Development Index, the Gender Empowerment Measure, and the Human Poverty Index. These indices utilize similar estimation techniques but consider different sets of functionings.

The HDI and the work of Sen (1985) demonstrate that income proxies do not fully capture a country's level of human development. Other indicators of a country's human development might produce a different relative ranking of a country's human development. A country's success in one dimension of human development does not always correlate with a country's success in a second dimension of human development. These results ultimately demonstrate that, given a similar set of resources, countries can utilize those resources differently to achieve various levels of human development across indicators.

Microeconomic Policies: Income

There is an expanding literature that uses micro-data and finds evidence to support the conclusion that individuals who are economically poor are not necessarily functionally poor (Balestrino 1996; Klasen 2000; Phipps 2002; Qizilbash 2002; Reddy, Visaria, and Asali 2006; Ruggeri Laderchi 1997; Ruggeri Laderchi 1999; Ruggeri Laderchi, Saith, and Stewart 2003). Few studies describe the relevance of these conclusions in a policy context other than to suggest that monetary measures of deprivation do not correlate with measures of well-being. An exception is Qizilbash (2002), who examined how policy implications differ for each conceptual approach. Using 1996 Census data from South Africa, Qizilbash (2002) examined the monetary-poor and functioning-poor populations. His results support findings from previous literature that individuals who are economically poor are not necessarily functionally poor. Qizilbash (2002) notes that the government distributes public funds to South African provinces based on economic deprivation, not functioning deprivation. Scarce public funds may be misdirected or inefficiently distributed to those households with the lowest welfare. In a similar conclusion, Balestrino (1996) argues that in-kind transfers may be more effective at fighting deprivation compared to cash transfers, especially for functioning poor individuals.

Equivalence scales are used to control for possible heterogeneity in consumption across individuals in the neoclassical framework (see Blundell and Lewbel (1991); Muellbauer (1977); Pollak and Wales (1979); Pollak and Wales (1992); and Slesnick (1998)). These scales are applied to individuals or groups of individuals to adjust for possible economies of scale in the process of converting resources into well-being so that observational units are comparable. But equivalence scales do not fully capture heterogeneity across individuals in their conversion of resources into well-being. Equivalence scales merely capture (dis)economies of scale in the conversion process of resources into well-being; the scale factor is assumed to be the same across individuals. This implies that there is only heterogeneity in the conversion process based on the magnitude of resources and well-being and not in the ability to convert resources across individuals with the same level of resources.

Lelli (2005) assesses household welfare each of the methods of the capability framework and the neoclassical framework, and then contrasts her results across the frameworks. Using household data from Italy and Belgium she estimates the needs of a household, proxied by household income and a set of controls including household size and composition, age, gender, area of residence, type of occupation and sector, level of educational attainment, and marital status. Lelli (2005) uses equivalence scales in an effort to capture demographic differences in preferences across households in the process of converting resources into well-being. Equivalence scales are computed as the amount of income necessary to guarantee individual *h* has identical fulfillment as individual *r* on a given dimension of well-being. For each functionings vector, the income level Y_h^* is defined as:

$$\{Y_h^* | f_r^m(Y_r, \pi_r) = f_h^m(Y_h^*, \pi_h)\}. \quad (12)$$

The equivalence coefficient is defined as:

$$m_h = \frac{Y_h^*}{Y_r}. \quad (13)$$

This method assumes that income is positively related to functioning achievement.

Lelli (2005) finds that the variation in household income only partly explains variability in functioning achievements. Thus, income transfers are insufficient to compensate individuals for low functioning achievement.

In a related study, Schokkaert and Van Ootegem (1990) study unemployment compensation policies in Belgium. The authors measure functionings using factor analysis. Factor analysis assumes that observed variables are linear combinations of some common underlying dimensions, called factors (see also Balestrino and Sciclone 2000; Lelli 2001). In the context of the capability framework, the factors represent functionings. Factors serve as predictors in deriving the observed variables. The factors are assumed to be uncorrelated with each other. The factor loadings are easily interpreted as regression weights and correlation coefficients (Lelli 2001).

This approach assumes that that the observed variables are dependent on one or more latent variables. Schokkaert and Van Ootegem (1990) model the vector of observed variables (y, that is dimension $k \times 1$) as determined by a vector of latent variables called factor scores (f, that is dimension $m \times 1$ where $m<k$), a coefficient matrix called factor loadings (Λ, that is dimension $k \times m$), and a matrix of residuals ε:

$$y = \Lambda f + \varepsilon. \quad (14)$$

From this equation (and knowing y and Λ), f can be estimated using a least squares technique.

Schokkaert and Van Ootegem (1990) conclude that the effects of unemployment surpass mere income-loss as exhibited by functioning deprivation in social, psychological, and physical well-being functionings. Monetary transfers to people who are unemployed has the ability to impact an individual's utilization function and conversion factors. Any evaluation of this policy therefore should measure the effect of such transfers to the unemployed on their conversion function and conversion factors.

Previous studies also have tested empirically whether additional monetary resources might be distributed in order to equalize the well-being of individuals. Kuklys (2005) developed a theoretical model for capability wherein she estimated capability sets in an effort to show that the opportunities available to individuals matters. She applies her model in the context of disabled and non-disabled individuals (see also Zaidi and Burchardt (2005) and Mitra (2006)). The capability set is modeled as:

$$Q_i(Y_i) = h(Y_i | z_i, z_s, z_e) \quad (15)$$

where the capability set Q for individual i given her disposable income Y is determined by the individual's conversion function h. Kuklys (2005) assumes: the effect of conversion factors (non-monetary constraints) on the functioning can be expressed as an effect on the capability set; the characteristics of goods are the same for every individual (so $c(x)=x$); all goods affecting welfare are marketable; more income results in more capability; and the conversion function is monotonic.

To estimate her theoretical model, Kuklys (2005) inverts the model:

$$Y_i | z_i, z_s, z_e = h^{-1}(Q_i). \quad (16)$$

The right-hand side of the equation defines a monotonic transformation of the capability set. The left-hand side is readily estimated.

Empirically, she estimated equivalent household income as the income necessary for household h to achieve the same level of income satisfaction as the reference household r using equivalence scales

$$\frac{y^h}{y^r} = exp\left\{\frac{1}{\hat{\beta}_1} * \hat{\beta}_2 (z^r - z^h)\right\} \quad (17)$$

where y is household income and z is an indicator of disability. Note this required an additional assumption that overall household utility is additively separable in utility derived from consumption of goods and utility derived from other sources.

Her results suggest that a disabled individual in Britain requires 56 percent more income than a non-disabled individual in order to achieve the same level of income satisfaction. Furthermore, a disabled individual has a capability set that is 36 percent smaller than the capability set of a healthy individual.

Mitra (2006) discusses how disability can be measured in the capability framework to more completely identify its economic causes and consequences. The capability framework constructs two distinct states of disability: potential disability and actual disability. An individual is actually disabled if she cannot do or be the things that she values doing or being. In contrast, other models of disability only allow for individuals to be disabled or not to be disabled. The capability framework also accounts for personal heterogeneities in resource availability, characteristics, and environment. These factors influence an individual's actual and potential disability. Other models of disability using the neoclassical framework often fail to account for such personal heterogeneities.

Microeconomic Policies: Happiness

In another comparison of results using the capability framework and the neoclassical framework, Anand and van Hees (2006) examined the relation between capabilities and happiness u_i, as defined in traditional utilitarian economics. They modeled:

$$u_i = h_i(b_i) \quad (18)$$

where the happiness or utility u of individual i is a function of their functionings b and a utility function h that relates functions to happiness and varies between individuals. The study considered seven dimensions: happiness, sense of achievement, health, intellectual stimulation, social relation, environment, and personal projects. Their empirical estimation relied on ordinal logistic regression models, ordered logit models, and Spearman rank correlations. Anand and van Hees (2006) concluded that higher income levels are associated with lower capability satisfactions. The authors suggest that this may indicate a tradeoff between objective improvement and subjective dissatisfaction. The authors also found evidence that individuals refer to their own capabilities in order to assess the distribution of opportunities within society in all dimensions considered except health and the environment.

Microeconomic Policies: Education

Terzi (2005) used the capability framework to contrast educational policies for teaching children with special needs. Terzi (2005) redefines disability and special needs in the context of the capability framework to consider justice within the educational setting.

There is a policy debate in the education literature regarding whether it is more developmentally beneficial for children with special needs to learn alongside children without special needs, and whether this co-learning is detrimental to the development of children without special needs. This debate is referenced in the education literature as the “dilemma of difference.” Proponents of segregated teaching argue that teaching children with and without special needs in the same classroom diverts resources (e.g. teachers time) away from children without special needs in order to accommodate children with special needs.

Terzi (2005) does not provide an empirical examination of this issue and therefore does not operationalize the capability framework to make any definitive responses to the dilemma of differences. Even still, Terzi (2005) offers a thoughtful discussion of how individual capabilities may be limited by those of others. Terzi (2005) argues the focus of education policy should be directed to the capabilities afforded to the students across learning environments. In this regard, educating students with special needs alongside children without special needs may stunt the learning of children without special needs.

In another application in the education literature, Kelly (2010) identifies the theoretical gains of operationalizing the capability framework over the neoclassical framework in the context of school choice policy in England. Suppose parents choose a school for their child to attend that presumably will increase the child's well-being.[12] In the context of the capability framework, simply increasing the quantity of schools available for the child to enroll in, is unlikely to improve her well-being unless it affects her motivation for school choice, e.g. her expectations for success (Kelly 2010, 327).

Kelly (2010) uses Sen's concept of adaptive preferences to argue that students in public schools with few resources may develop low-expectations of themselves and their achievements. Adaptive preferences suggest that people might adapt to unfavorable circumstances so any preference based valuation will be distorted.

Lorgelly, Lawson, Fenwick, and Briggs (2010) argued that adaptive preferences may be a particularly difficult issue for operationalizing the capability framework in the context of policy outcome evaluation. The public, as they argue, may value a policy outcome based on its initial outcome while those most affected by the policy might consider its stream of benefits overtime (Menzel, Dolan, Richardson, and Olsen 2002). The issues related to adaptive preferences and policy outcome evaluation can be avoided

12. Kelly (2010) notes that this model is hypothesized under complete information about each school but in reality parents face much uncertainty in their decision of which school to enroll their child (Kelly 2010, 319). Parents do not know which school offers the best quality education for their child nor is it clear how the child will respond to the particular schooling.

if experts are left to determine the valuation of policy outcomes (Coast, Smith, and Lorgelly 2008). Low expectations of the children may deter the children from utilizing more advanced (e.g. higher-quality) schools. Indeed, differences in expectations may help to explain why the best students at the worst schools are most likely to take advantage of school vouchers and increase their choice set so that they might select higher-quality schools (Kelly 2010, 327). Under-challenged students at the low-performing schools may maximize their utility and be "happy" according to the neoclassical framework, but she may not maximize their capability in the context of the capability framework.

The capability framework emphasizes an individual's values as central to her choice behavior. Using the capability framework in the context of school choice, we may explain the scenario of a family who does not value education and therefore does not select high-quality schools for their child, and transfers their low expectations of school achievement to their child. The child's low academic achievement will persist if she never maximizes her potential well-being. This result is not possible in the neoclassical framework – we would only observe that the student maximizes her utility given her preferences (for low academic achievement).

In the context of school choice, the effect of vouchers should be assessed based on the child's potential achievements, not their actual achievements. The student ultimately might achieve a higher functioning as a result of the change in her valuation of education.

Assessing the policy effect of vouchers on the child and her family's capabilities highlights the interconnected relationship between policy evaluation and policy formation. The measured effect of policies in the capability framework has the potential to reveal key areas of improvements for policymakers to enact change. For instance, policymakers might target particular conversion factors or aim to alter individual utilization functions in an effort to affect the individual's level of functioning and capability.

In the case of school vouchers, the measured effect of vouchers on the capabilities of children and their families would suggest that school vouchers should be distributed to families for which the child experiences the greatest increase in her opportunities for academic achievement. This includes correcting the child and family's low expectations of academic success and low valuation of education – the effects on the child and family's utilization function and conversion factors. The increase in opportunities would increase the likelihood of a change in the student's expectations such that she would adopt higher expectations of her academic achievements and thereby alter her valuation of academic achievement.[13]

Increasing the choice set to include inferior schools does not necessarily increase the child's well-being of functioning achievement. If choice is a probabilistic outcome and inferior schools are added to the set of schools for possible enrollment, then there is an increased likelihood that you might enroll your child in an inferior school (Kelly 2010, 330). Regardless of the framework for analysis, it may not be the interest of the parent to select an alternative school for their child (i.e. utilize the voucher) out of fear of selecting an inferior school. Thus, functioning and utility would remain the same. However, the capability framework would capture this change in the potential to achieve some

13. In this regard, capabilities are endogenous. The child must be placed in the school system given her potential academic achievement and the possible effects on her valuations, not her initial capabilities

academic level. The inferior school would impact the child's capability set – by increasing the number of possible low-income achievements.

Microeconomic Policies: Social Norms

The capability framework also considers social norms as intrinsically important in the effect of policies. For instance, Lewis and Giullari (2005) examine the gender division of labor between market production and household production, particularly unpaid and paid caring labor. Lewis and Giullari (2005) argue that capability deprivation will persist across gender lines unless additional policies are enacted to correct for such deprivation. Such policies would equalize unequal power relations in the choice of employment, particularly in the selection of which gender is to perform caring labor (which is typically unpaid and women predominantly perform). Olson (2002) argued for similar changes in the evolution of social and institutional policy related to caring labor and gender norms. These studies use the capability framework as a means to explain deprivation within society and to develop policies which will alleviate this deprivation.

In a related work, Alkire (2002) conducted a cost-benefit analysis of three community-level development projects in Pakistan on the capabilities of community members using the neoclassical framework and also using the capability framework. The projects included goat rearing, female literacy classes, and rose garland production. Monetary valuations were utilized to calculate the cost-benefit analysis using the neoclassical framework. Qualitative data was acquired on functionings and capabilities before and after the projects were implemented through reflection group discussions. Participants relayed stories of personal experiences and were asked to rank the three functionings that were most affected by each project. Finally, each project was ranked subsequently relative to their ability to reduce poverty. The ordinal rankings of the projects varied across the frameworks with discrepancies among the rankings of two of the three projects. Using the neoclassical framework, the goat rearing project was the most beneficial and the literacy courses for women were more costly than beneficial (because female labor markets are small or non-existent in the respective region). In contrast, the capability framework ranked the literacy classes for women as the most beneficial of the three development projects. The literacy classes improved the confidence and self-esteem among the women, effectively altering the conversion factors of women.

Current Weaknesses with Operationalizing the Capability Framework

While Sen has extensively described the capability framework, he has not identified how capability sets should be valued and compared, or which functionings are relevant. Researchers are left to debate these questions. Indeed, these gaps in the literature underscore why researchers struggle to operationalize the approach. Researchers attempting to operationalize the capability framework often cite three specific problems: whether to measure functionings or capabilities, which capabilities are relevant, and how can functionings be accounted for in a multidimensional framework of capabilities.

It is unclear from the literature whether capabilities or functionings are the most relevant measure of well-being. Some studies focus on capabilities while others only measure functioning achievement, and there is no explicit discussion of why a researcher measured one and not another. Research that focuses on an individual's capabilities instead of her functionings supports the belief that each person should have equal opportunities to pursue a good life but requires the individual be responsible for their own choices. This guarantees that each individual has equal opportunities to pursue a good life but not everyone achieves a good life. In contrast, studies that measure functioning achievement consider an individual's actual achievement of well-being. However, the capability framework assumes that functioning achievement occurs because the individual has reason to value the outcome. Thus, studies that measure functioning achievements must assume that the individual has reason to value the outcomes.

Thus far, empirical applications which examine individual functionings have dominated the literature. Fewer studies attempt to measure individual capabilities. The measurement of capabilities has been developed more slowly because of data limitations and measurement constraints on how to accurately operationalize capabilities.

There is also disagreement in the literature as to which capabilities are relevant for measuring well-being. At the extremes, Vallentyne (2005) argued that all capabilities are important and should be included in analysis. Other scholars argue that it is sufficient to examine a single capability. These studies are partial capability analyses (Robeyns 2006, 366). Of course there are scholars in the middle of this spectrum who argue that a set of capabilities is sufficient. Nussbaum (2000) goes so far as to develop a well-defined list of ten capabilities which she believes must be considered together: life, bodily health, bodily integrity, senses, imagination and thought, emotion, practical reason, affiliation, other species, play, and control over one's environment (see also Nussbaum 2003).

Sen has yet to provide a definitive list of capabilities and is unlikely to provide such a list in the future (Sen 1993; Sen 2004). Researchers often criticize Sen's resistance to supply a definitive list of capabilities (Nussbaum 1988, 176; Qizilbash, 1998, 54; Williams 1987, 96) while others criticize Sen for placing too much value on some capabilities (Sugden 1993, 1952-3). Sen argued that the relevant capabilities for a particular application are application specific, and may not be relevant for other applications. It is up to the researcher to determine which capabilities are relevant for their empirical application.

The multidimensional nature of capabilities requires some consideration of how best to include various functionings in a measure of well-being. At one extreme, each functioning can be equally weighted. The HDI from the UNDP is a prominent example of this method. The HDI examines three functionings: educational achievement, life expectancy, and economic standard of living. Each functioning receives an equal weight and are aggregated to construct a measure of overall well-being. The functioning for educational attainment is composed of two variables with different weights: literacy (with a weight of two-thirds) and school enrollment (with a weight of one-third).

Researchers often utilize multiple weighting systems in order to validate their conclusions. Other researchers rely on the variance of indicators of functionings in order to measure weights for each functionings. It is argued that this method abstracts from explicit value judgments (Kuklys 2005). For instance, researchers statistically derive weights when they utilize factor analysis. Weights (i.e. the variance of variables) are measured according to the degree of information which the variables provide about the functioning. This issue, though, is not unique to the capability framework. The neoclassical framework grapples with these same issues, employing exchange values to weigh commodities. This consistency has allowed researchers to operationalize the neoclassical framework, even if its fundamental assumptions are contentious. Sen (1996, 397) asserts that the selection and weighting of capabilities depends on personal value judgments. Weights should be socially determined, either by the relevant group or all of society.

Still, the weaknesses of the capability framework as discussed in the literature can be overcome. Researchers should rely on the intended goal of a policy to guide their analysis of capabilities or functionings, and which dimension of well-being the policy targets. In the next chapters, I provide researchers with a new empirical tool to consider both functionings and capabilities. The empirical tool also directly models an individual's conversion process, which has received little attention in the literature until recently.

Conclusion

The capability framework offers researchers an alternative framework to evaluate policies. The capability framework allows researchers to evaluate policies based on their stated goal and measured effect on the being of individuals. While the capability framework and the neoclassical framework may produce similar results in cases where policies only aim to affect an individual's available resources, results diverge when we consider how an individual uses those resources. Since policymakers adopt both policies that provide resources to individuals and policies that target the ability of individual's to utilize resources, it important that the model used for policy outcome evaluation adequately capture the effects of these policies.

CHAPTER 2

MEASURING CONVERSION EFFICIENCY
IN THE CAPABILTY APPROACH:
A PARAMETRIC APPROACH

Chapter one showed how the capability framework offers policy analysts an alternative conceptual framework for measuring the effect of public policies on the well-being of individuals, where "well-being" is defined as an individual's set of possible states of being, or her "capability." The process by which individuals transform resources into achieved (actual) states of being is called the "conversion function." Thus, in the capability framework, both resources and conversion functions are determinants of an individual's achieved state of being, which is called her "functioning." Indeed, a key element of the capability framework is that it explicitly notes that it is possible for an individual to achieve some state of being that is below her potential state of being. If an individual achieves a state of being below her potential, this is indicative of some inefficiency in her transformation of resources into well-being.

Conversion efficiency describes an individual's ability to transform resources into a level of well-being using a minimum amount of resources. A fundamental assumption of the capability framework is that individuals have different abilities to use resources. Thus, a policy that provides individual with similar resources may not necessarily result in similar levels of well-being among individuals. An accurate evaluation of the effect of a policy that provides resources to individuals would take into account individual conversion efficiency.

Theoretically, the effect of a policy π on the functioning b of individual i is attributed to its impact on the utilization function *f* and conversion factors *z* (which can be related to the individual *i*, society *s*, or the environment *e*):

$$db_i = \frac{\partial f_i}{\partial x}\frac{\partial x}{\partial \pi}d\pi + \frac{\partial f_i}{\partial z}\frac{\partial z}{\partial \pi}d\pi \qquad (11)$$

Empirically, any policy evaluation operationalizing the capability framework should measure the effect of a policy on the individual's utilization function and conversion factors. It also should capture the individual's conversion efficiency.

Relatively few empirical studies in the capability literature measure conversion efficiency relative to those studies that estimate functioning or capability. The scant volume of literature is likely the result of empirical difficulties in identifying an adequate empirical technique that concurrently measures functionings, capabilities, and conversion efficiency. Moreover, while empirical methods exist that could measure conversion efficiency, these methods were not explicitly developed to analyze conversion efficiency and so must be adapted to appropriately capture the conceptual underpinnings. For instance, conversion efficiency does not assume that individuals are efficient in their transformation of resources into well-being. Most existing empirical methods, however, assume efficiency in the transformation of resources into well-being.

Even within the capability literature there is no consensus about what is the best method to measure conversion efficiency in the production of well-being. Previous studies that measure conversion efficiency employ simultaneous equations or data envelope analysis. But neither of these techniques adequately account for possible individual heterogeneity in the conversion of resources into states of being. Simultaneous equations and data envelope analysis methods instead assume homogeneity in the conversion of resources into states of being.

I adapt an empirical technique from the industrial organization literature, stochastic frontier analysis, to measure any deviation between achieved states of being and potential states of being. This technique has the ability to account for individual heterogeneity in the conversion process.

This chapter begins with a survey of the empirical literature that operationalizes the capability framework and measures or accounts for conversion efficiency. Then, I discuss how stochastic frontier analysis (SFA) provides researchers with the necessary empirical tools to model an individual's production of well-being and measure conversion efficiency.

Current Techniques to Measure Conversion Efficiency: Overview and Critique

Conversion efficiency is a measure of the efficiency with which individuals convert their resources into achieved functionings. It captures the difference in functioning achievement between individuals that does not come from differences in their resources. Namely, conversion efficiency captures differences in functioning achievement across individuals that are caused by differences in personal z_i, social z_s, and environmental z_e conversion factors.

As presented in chapter one, the conversion function *f* maps the resources into the space of functionings. An individual *i* converts resources *x* from all available resources *X* into some level of functioning achievement *b*.

$$b_i = f_i(c(x_i)|z_i, z_s, z_e) \; \forall \; f_i \in F_i \; and \; \forall \; x_i \in \; X_i \qquad (1)$$

There are two steps in the conversion of resources into a state of being. First, a conversion function *c(•)* maps the vector of resources into the space of characteristics such that $c=c(x_i)$ (Lancaster 1966). Conversion functions for the space of characteristics are similar for all individuals. This is solely due to the fact that the characteristics of a resource do not vary across individuals. Second, an individual benefits from the characteristics of a resource via her own conversion function *f* from her set of all possible conversion functions *F*. In other words, there may be individual heterogeneities in the utilization of those resources given their characteristics.

The set of all feasible functioning vectors *b* for individual *i*, comprise her capability set *B*.

$$B_i(X_i) = \{b_i |\; b_i = f_i(c(x_i)|z_i, z_s, z_e) \; \forall \; f_i \in F_i \; and \; \forall \; x_i \in \; X_i\} \qquad (4)$$

Recall, a functioning is a state of being. The capability set therefore describes an individual's possible states of being.

Individuals with the same resources may not necessarily achieve the same state of being and may have a different set of possible states of being since individuals can derive different benefits from the characteristics of resources. For instance, consider the provision of free health care for children through a public health insurance program. The access to health care will provide different health benefits for children with different preferences, values, and health conditions. Some children may not change their use of health care even if the health care is freely available through public health insurance. Free health care will not impact the health of these children. Other children who increase their use of health services may exhibit higher levels of health functioning (assuming a positive relationship between health and health service use). Regardless of the effects on functioning, the provision of free health care may impact a child’s capability for good health by improving the child’s potential health outcomes. Consider further, two children, one with a disability and the other without a disability and their capability to achieve good health. The capability of a child who is disabled and in need of relatively more medical care may expand relatively more than the capability for a child who is not disabled. Free health care is welfare enhancing if it expands a child’s capability regardless if the child’s level of health changes.

Survey of Empirical Methods to Measure Conversion Efficiency

In the literature, there are two primary approaches to measure conversion efficiency. Both require estimation of an individual’s conversion function. The first approach focuses on the coefficient estimates of the conversion factors. The second approach, and more widely used, examines the estimated residuals of the conversion function to estimate conversion efficiency.

Method One: Estimated Coefficients on Conversion Factors

One strand of empirical literature on the measurement of conversion efficiency focuses on conversion factors (i.e. z_i, z_s, or z_e). Estimated coefficients on conversion factors measure the rate of transforming resources into functionings. Chiappero Martinetti and Salardi (2008) first used this approach to examine which population subgroups (by age and gender) are more efficient in converting resources into functionings. The authors estimated conversion functions for ‘being healthy,’ ‘being educated,’ and ‘living in a healthy and safe environment.’ The coefficient estimates on each conversion factor were compared across population subgroups to understand differences in conversion efficiency.

Similarly, Hasan (2009) estimated conversion efficiency using coefficient estimates on conversion factors. His model, however, accounted for the interdependency between conversion efficiency, functioning, and capability. Hasan (2009) estimated a three-stage least squares (3SLS) model:

$$capability = f(functioning, freedom) \qquad (19)$$

$$functioning = g(conversion\ efficiency)$$

$$conversion\ efficiency = h(constraints, resources)$$

His analysis supports the theoretical hypothesis that efficient use of resources results in higher states of being, and that freedom is positively related to capabilities.

Both Chiappero Martinetti and Salardi (2008) and Hasan (2009) assume that an individual maximizes her functioning achievement given her resources and conversion factors but this assumption may be too strong. The capability framework allows for an individual to achieve some level of functioning achievement that is within her capability but does not efficiently utilize resources. Consider an individual's capability for being well-nourished. If the individual chooses to fast for religious reasons, then the individual chooses to achieve some level of nourishment below what is possible given the food available to her. In contrast, an individual who is starving because she lacks the means to purchase food will achieve a similar low level of nourishment but this may maximize her capability. In modeling the functioning achievement of 'being nourished' for the individual who fasts and the individual who starves, the approaches of Chiappero Martinetti and Salardi (2008) and Hasan (2009) treat both individuals the same without knowledge of whether they had voluntarily decided to stop consuming food. Both individuals would have the same estimated conversion efficiencies. In reality, the individual who is starving more efficiently transforms resources into a level of functioning achievement compared to the conversion efficiency of an individual who chooses to fast and achieves some level of functioning below that which is possible given her available resources.

Method Two: Residuals of the Conversion Function

A second approach to measuring conversion efficiency examines the residuals of a conversion function for functioning achievement. Sen (1985) first utilized this method to link life expectancy to national income estimates. For instance, Sri Lanka's life expectancy was approximately 20 times higher than the country's actual income. The residuals were interpreted as estimates of country-level inefficiency. Other reports that rely on residuals include the World Health Organization's World Health Report for 1999 (Jamison, Creese, and Prentice 1999), the World Banks' World Development Report (1993), and the United Nations Development Program's Human Development Reports (see also Kakwani 1993; Moore, Leavy, Houtzager, and White 2000; Wang, Jamison, Bos, Preker, and Peabody 1999).

More recently, this approach has been formalized with the use of data envelope analysis (DEA). This method is a nonparametric technique that has been used in the capability literature to simultaneously measure capability, functioning, and conversion efficiency (Binder and Broekel 2008; Binder and Broekel 2010; Ramos 2008; Ramos and Silber 2005).

Deutsch, Silber, and Yacouel (2001) were the first to use DEA to estimate functioning achievement. They estimate a Malmquist index that depends on the measure of inefficiency to examine measures of technical inefficiency. Efficiency is a relative measure. The best individual achieves a level of functioning achievement given her resources that is far below her minimum level of resources necessary to achieve this level of functioning achievement. For this analysis, they assume that all individuals have the same level of functioning achievement. The individual who best utilizes her resources will achieve an actual level of functioning achievement that surpasses what she is predicted to achieve. Every other individual's residual is compared to the individual with the best (i.e. smallest) residual.

Deutsch, Silber, and Yacouel (2001) use time-survey data from Israel's Central Bureau of Statistics that was collected from 1992 to 1993. They estimate a vector of well-being for each household as a function of resources (e.g. own or rent the household, number of rooms per household member, number of cars per household member, number of televisions per household member, and household income) and individual characteristics (e.g. age, gender, ethnic origin, period of immigration, marital status, educational status, and area of residence). The authors found that conversion efficiency decreases with age at a decreasing rate. However, at age 57, conversion efficiency reaches its minimum value. Individuals who are older than 57 have an increasing ability to transform resources into some quality of life. Among conversion factors, an individual's education and geography do not affect the conversion efficiency index.

Ramos and Silber (2005) utilize this technique to compare estimates of human development acquired through four conceptual frameworks of well-being: Sen's (1985) capability framework, Narayan, Chambers, Shah, and Petesch's (2000) dimensions of well-being, Cummins (1996) domains of life satisfaction, and Allardt's (1993) comparative Scandinavian welfare study. The authors examine whether the estimates of well-being for each framework will produce similar measures of well-being given that each framework focuses on different dimensions of well-being. Ramos and Silber (2005) use efficiency analysis to estimate indices of individual functioning in a single dimension and to estimate an aggregate index of functioning across dimensions for each individual. The aggregate index attempts to capture an individual's level of well-being. Using data from the 1997 British Household Panel Survey, they find that the indices of well-being are highly correlated across approaches. All indices also are weakly correlated with an income measure of well-being (equivalent income). These results suggest the conceptual frameworks of well-being, which advocate for a multidimensional measure of well-being, are similar enough that estimates are robust across the frameworks. However, there is value-gained in using these multidimensional approaches beyond the usual income measures.

Binder and Broekel (2008) develop a measure of conversion efficiency to capture the efficiency with which individual resources are transformed into achieved functioning. Using a nonparametric efficiency procedure, the authors estimate an order-m efficiency method. Broekel and Binder (2008; 2010) model the capability frontier as a societal optimum which can be reached with a given bundle of resources. They assume that some individuals in their sample have already reached the capability frontier. In order to conduct their analysis, the authors must assume that everyone has the same preferences for the functionings considered. Else, it would be possible to prefer some other functioning over one that was measured and therefore shift her resources toward her preferred functioning instead of the measured functioning. Estimated inefficiency would therefore be biased.

Binder and Broekel (2008) consider six functionings: being happy, being educated, being healthy, being well-sheltered, being nourished, and having satisfying social relations. Using a single wave of data from the 2006 British Household Panel Survey, the authors estimate an ordinary least squares (OLS) model to evaluate the effect of conversion factors (age, age-squared, gender, employment status, marital status, geography, and disability status) on an individual's functionings. They find that nearly 50 percent of individuals are not efficient as the best 50 percent of individuals. Separating the sample into efficient and inefficient individuals, the average inefficient individual achieves about 20 percent less functioning than an efficient individual with the same resources. Conversion efficiency is positively affected by getting older, being self-employed, being married, having no health problems, and living in the London area. It is negatively affected by unemployment, separation, divorce or widowed marital status, and physical disability.

Binder and Broekel (2010) extend this analysis to estimate conversion efficiency over time using the same nonparametric order-m approach and data from the 1991 to 2006 British Household Panel Survey. They estimate a Malmquist index to measure the change in an individual's conversion efficiency across two time periods. This index is a measure of the movement of an individual's functioning relative to the benchmark frontier and demonstrates how an individual might decrease or increase her conversion efficiency relative to the order-m best practice individuals. Their findings corroborate Binder and Broekel (2008). Fewer than 30 percent of individuals were efficient in converting resources into functionings during the sample period. Age, education and selfemployment increase an individual's conversion efficiency. Living in London, being disabled, and being separated, divorced, widowed, or married are negatively associated with individual's conversion efficiency.

Use of DEA requires researchers to evaluate individual efficiency with respect to the maximum functioning achievement observed in the data given a certain level of resources. However, DEA does not consider that the "best" person may not be the ideal. For instance, Binder and Broekel (2008; 2010) assume that all individuals prefer to achieve a level of functioning that maximizes her capability. But some individuals may choose to achieve some functioning below their potential. In which case, estimation results might identify a subset of individuals who have persistent inefficiencies.

While relative social comparisons may matter, the ideal may be an individual who "best" utilizes resources given her capabilities and values, not society's capabilities and values. Researchers (Broekel and Binder 2008; Broekel and Binder 2010) using DEA to operationalize the capability framework have argued that this simplification to an individual is necessary since it is difficult to derive a theoretical maximal functioning achievement for society given a certain level of resources.

A New Technique to Measure Conversion Efficiency

I offer a new empirical approach, stochastic frontier analysis (SFA), to estimate the conversion efficiency of an individual's conversion function that improves on existing estimation methods. Unlike other parametric techniques (such as those used in Chiappero Martinetti and Salardi (2008) and Hasan (2009)), SFA allows for the decomposition of the error term into a random component and a measure of technological inefficiency. Thus the level of conversion efficiency (i.e. the measure of technological inefficiency) and the rate of conversion efficiency (i.e. the estimated coefficient on a conversion factor) can be estimated distinctly.

SFA also offers researchers an advantage in operationalizing the capability framework beyond the contributions of DEA. DEA is a non-parametric technique requiring that at least one observation in the sample lie along the capability frontier. Thus DEA ignores the possible influence of measurement errors and other sources of noise. All deviations from the frontier are assumed to be a consequence of technical inefficiency. But random error is possible in the context of the production of functionings. Variability in the transformation of resources into functioning achievement is subject to a myriad of possible errors, particularly measurement error if incomplete proxies are used to control for conversion factors.

Frontier analysis is typically used to model a firm's transformation of a set of inputs x into a set of outputs y. A firm is inefficient when it produces some output below its potential, i.e. below its frontier. This could occur if the firm does not use all available knowledge. A firm below its frontier could, for instance, better organize its machines or labor.

The capability frontier describes the most efficient transformation of resources into functioning. Individuals who are most efficient in transforming their resources into a functioning lie along the frontier. This conceptualization of the capability frontier assumes that resources are necessary for an individual to increase her level of functioning achievement. An individual's distance from the capability frontier is interpreted as a measure of how inefficient the individual is in converting her resources into achieved functioning, i.e. her level of conversion efficiency. Estimated coefficients on conversion factors measure rates of conversion for any particular conversion factor.

General Approach for Distance Functions

In studies employing distance functions, input functions are estimated to measure individual functionings and output functions are estimated to measure overall well-being as the composition of individual functionings (see Ramos 2008).

Production is a process of transformation of a set of inputs $x \in X$ into a set of outputs $y \in Y$. The input set *L(y)* consists of all input vectors *x* which can produce the output vector *y*:

$$L(y) = \{x: x\ can\ produce\ y\} \quad (20)$$

The production function is the isoquant *IQ(y)*:

$$IQ(y) = \{x: x \in L(y)\ and\ \rho x \notin L(y) if\ 0 \leq \rho < 1\} \quad (21)$$

where ρ is a scalar.

The input set *L(y)* is the area bounded from below by the isoquant *IQ(y)* in figure 2. The isoquant depicts the minimum amongst these input combinations for each proportion of inputs. The input vector $A=(x_{1A}, x_{2A})$ is inside the input set. It can be proportionally contracted to point $B=(x_{1B}, x_{2B})$. Point *B* lies along the isoquant *IQ(y)*. It

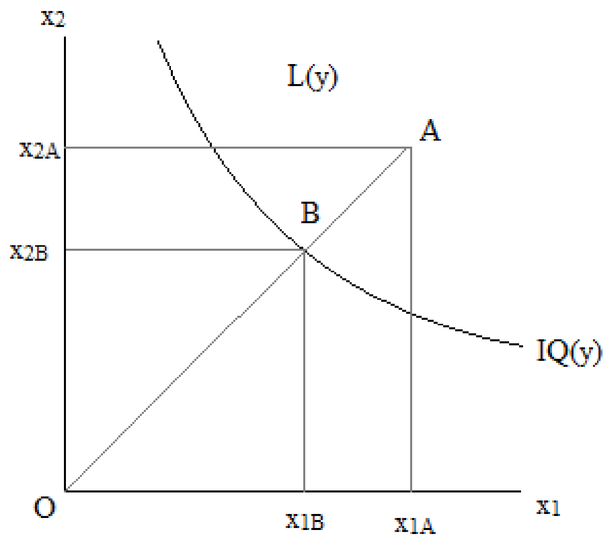

Figure 2. Input Distance Functions.

cannot be contracted proportionally any longer without changing the output vector *y* or the production technology. Conversion efficiency changes when the individual increases her level of functioning achievement without acquiring new or additional resources.

The input distance function $d_i(x,y)$ scales the input vector:

$$d_i(x, y) = Max\{\rho: (x/\rho) \in L(y)\} \qquad (22)$$

The input distance function has four properties. First, the input distance function is increasing in *x* and decreasing in *y*. Second, it is linearly homogenous in *x*. Third, if *x* belongs to the input set of *y* (i.e. $x \in L(y)$), then $d_i(x,y) \geq 1$. Finally, the input distance function is equal to unity if *x* belongs to the frontier of the input set (the isoquant of *y*).

In figure 2, the distance function for point *A* measures the distance (ray *OA*) between this point and the *IQ(y)*, as the inverse of the factor by which the production of all input quantities could be reduced while still remaining above the feasible isoquant for a given output vector. The distance function of the individual producing output set *y* using the input levels defined by point *A* equals the ratio $(OA/OB)=\rho$. At point *B*, $D_i(x,y)=1$; this is true for any point along the isoquant *IQ(y)*. If *x* belongs to *L(y)*, $D_i(x,y)\geq 1$.

The output distance function is the complement to the input distance function. The output set *P(x)* is the set of all outcome vectors *y* which can be produced using the input vector *x*:

$$P(x) = \{y: x \; can \; produce \; y\} \quad (23)$$

The production function is the production possibility frontier *PPF(x)*:

$$PPF(x) = \{y: y \in P(x) \; and \; \delta y \notin P(x) if \; \delta > 1\} \quad (24)$$

where δ is a scalar.

Consider an outcome vector *x* in figure 3. The output set *P(x)* is the various output combinations (y_1, y_2) that could be produced given input vector *x*. The output set corresponds to the area bounded by the two axes and the production possibility frontier *PPF(x)* which depicts the maximum amongst these output combinations. Or, it is the maximum amount of one of the outputs (e.g. y_1) that could be produced for a given amount of the other output (e.g. y_2) and the input vector *x*. The output vector A=(y_{1A}, y_{2A}) is inside the output set. It can be expanded proportionally to point B=(y_{1B}, y_{2B}), which lies

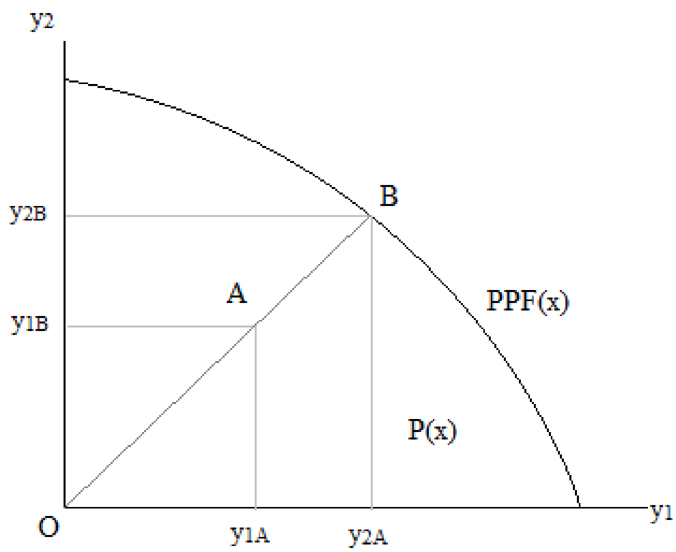

Figure 3. Output Distance Functions.

on the *PPF(x)*. The vector cannot be expanded any longer without changing the input vector *x* or the production technology.

The output distance function $d_o(x,y)$ scales the output vector:

$$d_o(x,y) = Min\{\delta: (y/\delta) \in P(x)\}. \quad (25)$$

The output distance function also requires four properties. First, the output distance function is increasing in *y* and decreasing in *x*. Second, it is linearly homogenous in *y*. Third, if *y* belongs to the production possibility set of *x* (i.e. $y \in P(x)$) then $d_o(x,y) \leq 1$. Finally, the output distance function is equal to unity if *y* belongs to the frontier of the production possibility set (the production possibility frontier of *x*).

In figure 3, the output distance function for point *A* measures the distance (ray *OB*) between this point and the *PPF(x)*, as the inverse of the factor by which the production of all output quantities could be increased while still remaining within the feasible production possibility set for a given input vector. The distance function of the individual using input vector *x* to produce the output levels defined by point *A* equals the ratio $(OA/OB)=\delta$ (where δ is a scalar), whereas the distance function value at point *B* is 1. $D_o(x,y)\leq 1$ if *y* belongs to *P(x)* and $D_o(x,y)=1$ if *y* lies on the *PPF(x)*.

The Capability Framework and Stochastic Frontier Analysis

Stochastic frontier analysis is a parametric estimation of distance functions. For a cross section of firms, the stochastic frontier is modeled as a production function:

$$y_i = a + \beta x_i + t z_i + v_i - u_i \quad (26)$$

where output *y* for firm *i* is determined by a vector of inputs *x* conditional on characteristics *z* of the firm that are time invariant, random error *v*, and a measure of the firm's inefficiency *u*; *β* and *t* are the parameters to be estimated. Time invariant characteristics of the firm may include the type of industry in which the firm operates. Inefficiency, again, is understood to be a measure of the firm's inability to convert inputs into output.

In the context of the capability framework, the capability frontier is modeled as a production function for functioning achievement:

$$B_i = a + \beta x_i + t z_i + v_i - u_i \quad (27)$$

where potential functioning achievement *B* for individual *i* is determined by a vector of resources *x* conditional on conversion factors *z* of the individual that are time invariant, random error *v*, and a measure of the individual's inefficiency *u*. β is the estimated effect of resources on an individual's potential functioning achievement and *t* is the estimated effect of conversion factors (i.e. rate of conversion efficiency).

Two assumptions are necessary. First, v_i are independent and identically distributed and normally distributed with mean zero and standard deviation σ_v^2. Second, u_i are independent of *x* and *v*. Ravallion (2005, 281) notes that this assumption may be a justified in the assessment of social policies when the inefficiency is unknown to the producer and so could not affect input choices.

The second assumption can be relaxed if the stochastic frontier is estimated with fixed effects:

$$B_{it} = a + \beta x_{it} + t z_i + v_{it} - u_{it} \quad (28)$$

for $u_i \geq 0; i=1,\ldots,N, t=1,\ldots,T$. This model controls for the time period *t* in which individual *i* is observed. Conversion efficiency is time varying since individuals may acquire new technologies, i.e. learn better ways to utilize market and nonmarket goods.

A few additional assumptions are necessary. First, v_{it} are independent and identically distributed, and normally distributed with mean zero and standard deviation σ_v^2. Second, x_{it} and v_{js} are independent for $t,s=1,\ldots,T$, $i,j=1,\ldots,N$.

Heterogeneity may not be fully captured in the above specification. A time invariant, individual specific constant term, a_i is commonly included. If

$a_i = a - u_i$ so that $a_i \leq a$ for all i, then the above equation can be rewritten as:

$$B_{it} = a_i + \beta x_{it} + tz_i + v_{it} \quad (29)$$

where u_i is lower-bounded by zero and is upper-bounded by a_i. This specification is a stochastic frontier estimated using fixed effects, which can account for unobservable characteristics of the individual that are constant over time in the individual-effects term a_i. The model also accounts for variation over time in the individual's ability to convert resources into higher levels of functioning achievement in the inefficiency term u_{it}.

The residual now is decomposed into random noise, a measure of inefficiency, and individual-effects. But there is an identification problem since individual-effects are time invariant and inefficiency can be time invariant. It is necessary to distinguish between heterogeneity across individuals that is unrelated to the inefficiency and the inefficiency itself in order to avoid mistakenly measuring heterogeneity as inefficiency. Greene (2005) suggests modeling inefficiency as time variant and refers to this specification as the "true" fixed effects model:

$$B_{it} = a_i + \beta_i x_{it} + tz_i + v_{it} - u_{it}. \quad (30)$$

Here the estimated coefficients on the resources are specific to the individual. They capture how the individual performs relative to her own standards. This is important since it captures differences in technologies across individuals. Intuitively, this captures individual heterogeneity in the rate of conversion of resources into functioning achievement. Inefficiency u_{it} is measured relative to its most efficient level of functioning production.

Distributional assumptions about the inefficiency term are necessary to produce efficiency estimates. This assumption is often cited as the primary weakness of SFA. The half-normal distribution is the default distributional assumption, suggesting that relatively more individuals achieve functionings close to their capability frontier while fewer individuals achieve some functioning far from their capability frontier. Different distributional assumptions (e.g. gamma, exponential) can be assumed in order to check for the robustness of results.

Conclusion

SFA provides researchers with the empirical tools to estimate capabilities, functionings, and conversion efficiency. This method offers several advantages over the methods currently used in the capability literature to operationalize the approach such as simultaneous equations and data envelope analysis. First, researchers can explicitly and distinctly estimate rates of conversion efficiency and levels of inefficiency. Second, SFA accounts for possible random error in the transformation of resources and conversion factors into states of being.

In the next chapter, I demonstrate how SFA can be used to operationalize the capability framework and conduct an evaluation of the child welfare system in the United States.

CHAPTER 3

EFFECTIVENESS AND EFFICIENCY OF OUTPATIENT SERVICES FOR CHILDREN WHO COME INTO CONTACT WITH STATE CHILD WELFARE SYSTEMS

In this chapter, I demonstrate how to use stochastic frontier analysis to operationalize the capability framework for policy outcome evaluation with an empirical application to the U.S. child welfare system. I measure the effect of the provision of mental health services on the ability of these services to reduce the future prevalence of mental health problems for children receiving mental health services. This application shows how the measured effect of a policy is the result of its influence on both resources and conversion factors (as predicted in equation (11)).

My empirical work contains two innovations. First, I consider the efficiency of mental health outcomes as an alternative assessment of whether state child welfare systems improve child well-being. For this analysis, I model a child's production of mental health outcomes from the use of mental health services as a stochastic frontier model following the discussion in chapter two. The stochastic frontier model allows me to distinguish between a child's achieved level of mental health and her potential level of mental health. Differences between a child's achieved and potential mental health are indicative of some inefficiency of mental health services to affect her mental health.

Secondly, I explicitly control for state effects in the ability of mental health services to improve the mental health of children. I use recently collected data, the National Survey of Child and Adolescent Well-Being (NSCAW), to facilitate across state comparisons of child welfare systems. NSCAW data have been used to explore the importance of integration of services at the local level (Konrad 1996; Kusserow 1991), but the role of differences between states has not been addressed. State effects capture state-specific conversion factors. Decentralization of child welfare policy to states has resulted in different child welfare policies on the provision of child welfare services across states. Differences in social norms, geography, and other state-specific characteristics are also captured in state conversion factors.

When I omit state effects, I find that children who received outpatient services are correlated with a 9 to 13 percent more mental health problems than children who did not receive these services. It is not clear from this analysis whether mental health problems are more prevalent for children who receive outpatient services or if outpatient services result in greater mental health problems for these children. On average, mental health services enhance efficiency conversion. When I include state-effects, I find evidence that simply offering mental health services is not sufficient evidence that the services will improve the mental health of children. State effects are not significant predictors of whether a child will receive a service but are strong indicators in a child’s production of mental health. Thus, assessing the ability of state child welfare systems to improve the well-being of children requires an assessment of the delivery and quality of services provided.

In this chapter, I first explain the policy context in which children come into contact with their state child welfare system and differences in how these systems provide mental health services to children. Next, I explain the theoretical application of how children utilize mental health services which are provided by state child welfare systems to achieve some mental health outcome. The empirical estimation is complicated by the fact that children are not randomly assigned to mental health services. I provide a discussion of how propensity score matching can be used to adjust for selection into mental health services. I proceed with a description of the data. I then discuss results from the propensity score matching technique and stochastic frontier analysis. I conclude with how these findings relate to current work in the child welfare literature.

Background on Child Welfare Policy in the United States

There is a critical need for states to understand how best to provide mental health services to children who come into contact with their state child welfare system. Each year, states spend more than $25 billion on child protective services (Bess, Scarcella, Jantz, Russell, and Geen 2002). Nearly half come from federal funds; however, the share of child welfare costs paid by states increased until the onset of the recession in December 2007 (DeVooght, Allen, and Geen 2008). Little evidence exists to guide states in the effective distribution of their increasingly scarce resources. The findings of this research will provide much-needed information to states by evaluating how well child welfare systems have served children since 1998.

Since the enactment of the Adoption and Safe Families Act of 1997, it has been the explicit goal of the U.S. child welfare system to ensure safety and promote permanency for children to improve child well-being (Wulczyn et al. 2005). While no formal definition of well-being has been adopted, the researcher related to child welfare has examined whether the services provided to children reduce their risk of long-term outcomes. Previous research suggests that former foster children have a higher propensity to drop out of school, receive welfare, experience substance abuse problems, commit crimes, or enter the homeless population, and these long-term outcomes are positively correlated with mental health problems (Burt, Aron, Douglas, Valente, Lee, and Iwen 1999; Clausen, Landsverk, Ganger, Chadwick, and Litrownik 1998; Doyle 2007; Dworsky and Courtney 2001; Vinnerljung, Sundell, Lofholm, and Humlesjo 2006). Between 50 to 75 percent of children entering foster care exhibit problematic behaviors that warrant mental health care (Leslie, Hurlburt, Landsverk, Barth, and Slymen, 2004; Burns, Phillips, Wagner, Barth, Kolko, Campbell, and Landsverk 2004). In comparison, mental health problems occur in about 20 percent of all children (Costello, Angold, Burns, Stangl, Tweed, Erkanli, and Worthman 1996; Kataoka, Zhang, and Wells 2002). Improving the well-being of children who come into contact with state child welfare agencies thus involves reducing the prevalence of mental health problems among children who come into contact with a state child welfare system.

Understanding the Extent of Variability across CPS Agencies

Studies that do not account for differences in state policies and procedures to provide services to children and even differences in their assessments of the needs of the children, do not accurately measure the effect of services. Child welfare services are decentralized to state governments, county governments, and child protective service

(CPS) agencies, and are not administered at the federal level.[14] As such, there is great variability across states in their delivery of services to children who come into contact with their state child welfare system. Previous studies that assess the ability of the U.S. child welfare system to improve the well-being of children fail to address the institutional framework in which these services are provided.

In 2001, the U.S. Department of Health and Human Services (HHS) in partnership with state governments conducted the first Child and Family Services Review (CFSR).[15] The CFSR was conducted in response to a 1994 Congressional mandate that HHS identify the strengths and weaknesses of state child welfare programs and determine the extent to which states comply with federal mandates. In 2007, Final Reports and Program Improvement Plans from all 50 states and the District of Columbia were compiled and reviewed. The reports of the CFSR highlight the variability in mental health provision for and assessment of children who come into contact with child welfare across states. The findings of these reports also underscore the urgent need for reform of child welfare policies to ensure children entering foster care receive effective services and that states efficiently provide these services to children and their families.

The reports of the CFSR found that 16 states required a mental health screening or assessment for some children at or near entry into foster care (McCarthy, Van Buren, and Irvine 2007, 2). Among states that issued such a screening, there is considerable variation

14. The ideal data set would allow for agency-level analysis since policies and procedures to provide mental health services to children can vary across agencies. NSCAW data survey 97 local CPS agencies. Small sample sizes do not allow for agency-level analysis for this dissertation. In the analysis sample, there is an average of 11 children per CPS agency; more than 40 percent of CPS agencies in the analysis sample have fewer than five children. I instead use state level data as the next best alternative.

15. The second round of state reviews began in 2010.

in which children receive the assessment. Only Connecticut required that all children entering foster care receive a mental health assessment (McCarthy et al. 2007, 2). In contrast, Minnesota assessed a wider range of children who came into contact with the child welfare system, including both children in child welfare, those in the juvenile justice system, and child abuse victims who remain at home. Child specific determinants of whether or not a child receives an assessment include age, abuse history, or whether or not his or her parental rights have been terminated.

There are also important differences in who conducts the health assessment. For instance, CPS workers in Alabama and California receive training to identify children with mental health needs, but the methods are not consistent across states and may not be implemented consistently across CPS workers within a state (U.S. Department of Health and Human Services 2008). In Connecticut, the mental health assessment for children entering the child welfare system must be administered by a licensed Master of Social Work. Whereas in Washington, CPS agencies employ specially trained staff to implement the standardized instruments to assess the mental health needs of children entering foster care (Washington State Department of Social and Health Services 2011). The inconsistencies in mental health assessment may also contribute to children not receiving mental health services or the failure of CPS agencies to identify children in need of services.

Even the timeframe over which the assessment is conducted varies substantially across states. One state requires the child be initially assessed within 24 to 48 hours upon entering foster care, with monthly follow-ups. Most other states have a longer timeframe, such as within 30 to 60 days after entering care, if any.

There is considerable variability across states in their procedures to identify a child's need for mental health services, to deliver mental health services, and to monitor the effect of these services (McCarthy et al. 2007). Researchers can take advantage of this variation at the state level to account for which methods are most effective, i.e. result in the best mental health outcomes of children. However, recent studies using national data have failed to exploit this variation and so have failed to provide state policymakers with any meaningful results about how their policies fare relative to those of other states.[16] This study is the first to use national data and control for state effects in an effort to utilize state variation as an explanatory factor of the mental health outcomes of children.

Methodological Framework

The Production of Mental Health

Upon referral by CPS, mental health providers offer services to a child with the aim to improve a child's level of mental health. Figure 4 illustrates the conversion process by which a child's use of mental health services (S) results in a level of mental health (Y). Mental health providers offer services with the aim to improve a child's level of mental health. A child's CPS worker and caregiver choose among vectors which describe a combination of mental health outcomes and mental health service consumption. The set of feasible mental health outcomes describes the child's capability set—illustrated as the mental health capability frontier and the shaded region below it. Both y_1 and y_2 describe possible levels of functionings. A child may produce her

16. For example, Farmer, Mustillo, Wagner, Burns, Kolko, Barth, and Leslie (2010), Raghavan, Inoue, Ettner, Hamilton, and Landsverk (2010), Ringeisen, Casanueva, Urato, and Stambaugh (2009), Shipman and Taussig (2009), and Southerland, Casanueva, and Ringeisen (2009) use NSCAW data that look at the delivery of mental health services but do not address state differences.

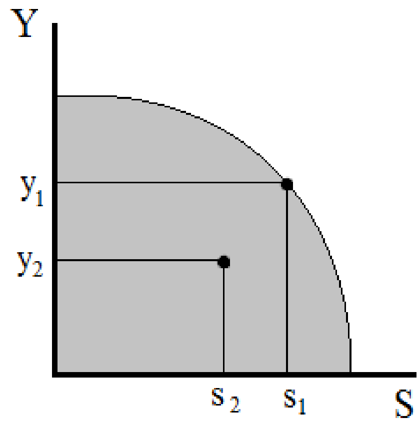

Figure 4. The Mental Health Capability Frontier

maximum level of mental health for the quantity of mental health services received, so that she reaches a point such as y_1, when services s_1 are received. Alternatively, a child may not reach the frontier. In addition to random variation, some mental health services may inefficiently affect the child's level of mental health. A child may achieve a level of mental health (i.e. a functioning) below her capability frontier. A child will achieve a level of mental health below her frontier such as y_2 when she receives services s_2. Such an inefficient point can be achieved for a number of reasons, including conversion factors or individual abilities to utilize resources captured via her conversion function.

The stochastic frontier framework allows for the separation of child-specific inefficiencies from the estimated effectiveness of services and conversion factors. Stochastic frontier analysis (SFA) is typically used to model a firm's transformation of a set of inputs into a set of outputs (see the seminal work of Aigner, Lovell, and Schmidt 1977). I use the SFA to estimate mental health capability frontiers of children who came into contact with their state child welfare system and to measure whether health outcomes are more efficient for children who received mental health services and children who did not receive services.

As its name indicates, the stochastic frontier takes into account the stochastic nature of the transformation of mental health services into measurable outcomes, i.e. functionings. The conversion function describes how individual heterogeneity may affect how a child internalizes an outpatient service and utilizes the treatment in future behaviors. Inefficiency exists when a child achieves a level of mental health below her capability frontier. Levels of functioning below the capability frontier might occur for two possible reasons: idiosyncratic effects and productive inefficiency. Idiosyncratic effects, also termed random error, are events external to the child that may affect her ability achieve her maximum level of functioning. For instance, physical abuse from a caregiver may impose mental trauma on the child. Productive inefficiency is related to an individual's conversion function and conversion factors. The capability framework does not assume that an individual wants to utilize her resources efficiently, so it is possible that an individual chooses to achieve some level of well-being (i.e. level of functioning) below her capability frontier. Using stochastic frontier analysis instead of OLS, the error term is decomposed into a random component and a measure of technological inefficiency.

The stochastic frontier specified as:

$$Y_i = \alpha + T_i\beta_1 + X_i\beta_2 + v_i - u_i \qquad (31)$$

where the level of mental health functioning Y for a child i $(i=1,\ldots,N)$ is determined by a vector of mental health treatment T, conditional on conversion factors related to characteristics of the child X, a two-sided stochastic term that accounts for statistical noise v and a nonnegative stochastic term representing the child's inefficiency u.

Inefficiency *u* represents the proportion by which *Y* falls short of the goal. It is the proportional or percentage of inefficiency. β_1 captures the effect of the resources, i.e. mental health services; β_2 captures the rate of conversion efficiency of the individual's conversion factors.

Standard assumptions of the stochastic terms are $E[v_i]=0$ for all i, $E[v_iv_j]=0$ for all i and j $(i\neq j)$, $E[v_i^2] = \sigma_v^2$, $E[u_i]>0$, $E[u_iu_j]=0$ for all i and j $(i\neq j)$, and $E[u_i^2] = \sigma_v^2$. The stochastic terms v_i and u_i are assumed to be uncorrelated; v_i is normally distributed; and u_i is half-normally distributed. The density function for $\varepsilon_i \equiv v_i - u_i = lny_i - \alpha + T_i\beta_1 + X_i\beta_2$ is:

$$f(\varepsilon_i) = \frac{2}{\sigma}\varphi(\frac{\varepsilon_i}{\sigma})\Phi(\frac{-\varepsilon_i\lambda}{\sigma}) \quad (32)$$

where $\sigma_i^2 = \sigma_{vi}^2 + \sigma_{ui}^2$, $\lambda_i = \sigma_{ui}/\sigma_{vi}$, φ is the standard normal density function, and Φ is the standard normal cumulative distribution function (Kumbhakar and Lovell 2000). The log-likelihood function for the production frontier model is:

$$log\, L_i = \frac{1}{2} log\left(\frac{2}{\pi}\right) - log\sigma - \frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2 + log\Phi(\frac{-\varepsilon_i\lambda}{\sigma}). \quad (33)$$

Child specific technical efficiency is the ratio of observed mental health to the corresponding stochastic frontier mental health (when there is no inefficiency so $u_i=0$). Since mental health outcome is measured in its natural log form, technical efficiency is specified as (Battese and Coelli 1988):

$$TE_i = \frac{y_i}{exp\,(log\, Y_i = \alpha + \beta_1 T_i + v_i)} = \frac{exp\,(log\, Y_i = \alpha + \beta_1 T_i + v_i - u_i)}{exp\,(log\, Y_i = \alpha + \beta_1 T_i + v_i)} = exp\,(-u_i). \quad (34)$$

I measure u_i to predict child-specific technical efficiency. An estimate of ε_i is obtained after I estimate the frontier. The estimate of ε_i is then used to estimate the inefficiency component u_i using the conditional mean function $E[u_i|\varepsilon_i]$ following Jondrow, Materov, Lovell, and Schmidt (1982):

$$E[u_i|\varepsilon_i] = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\varphi(\frac{\varepsilon_{i,\lambda_i}}{\sigma_i})}{1-\Phi(\frac{\varepsilon_{i,\lambda_i}}{\sigma_i})} - \frac{\varepsilon_{i,\lambda_i}}{\sigma_i}\right]. \quad (35)$$

The estimator is the expected value of the inefficiency term given an observation on the sum of inefficiency and the child specific heterogeneity.

<u>Empirical Model: Stochastic Frontier Analysis</u>

I use a log-linear functional form for the stochastic frontier model, estimating:

$$Y_{i,t+1} = T_i\beta_1 + X_{i,t}\beta_2 + I_{i,t}\beta_3 + D_{i,t}\beta_4 + Y_{i,t}\beta_5 + \varepsilon_{i,t+1} \quad (36)$$

where Y_i is the natural log of the mental health of child *i* in period *t+1* and T_i is an indicator variable for children who received outpatient services over the sampling period.[17] The predictors are measured at period *t*, where X_i are characteristics of the child, I_i are factors of the availability of services, and D_i is an indicator for the child's state of residence. $\beta_1, \beta_2, \beta_3$, and β_4 are parameters to be estimated.

The effect of services β_1 is interpreted as the estimated effect of service receipt on mental health at time *t+1*, net of mental health at baseline.

17. Data are log-transformed to correct for the non-normal distribution of the mental health measure.

Coefficient estimates on state indicators β_4 estimate state effects. State effects capture state policies, state social programs including Medicaid, geographical differences, and state social norms among other state specific factors.

Current mental health in model (36) adjusts for non-clinical factors which persist at period *t* and *t+1* and have a consistent effect on current and future levels of mental health regardless of service receipt (Stahmer, Leslie, Hurlburt, Barth, Webb, Landsverk, and Zhang 2005). There are two drawbacks to predicting future mental health as a function of current mental health. First, including lagged mental health fails to take into account random variability in baseline mental health and therefore creates a biased correlation between baseline mental health and the error term. Second, unobservable factors still may bias estimates if they have a differential effect on current and future mental health.

Correcting for the Nonrandom Provision of Outpatient Services to Children

The empirical application in this dissertation requires an adjustment for the nonrandom distribution of mental health services to children. The ideal data for program evaluation would be generated by random assignment of individuals to the treatment. However, children are non-randomly assigned to outpatient services so their unobservable characteristics are likely non-normally distributed across groups. In the absence of random assignment, there may be some unobservable characteristics of the child which influence her receipt of mental health services.

The process of how a child comes into contact with their state child welfare system and their possible receipt of mental health services is not always similar across states. Most children come into contact with CPS as the result of a report of suspected child abuse or neglect. Reports can be filed by any person. In 2006, most reports were filed by teachers (17 percent), lawyers or police officers (16 percent), and social services staff (10 percent) (Child Welfare Information Gateway 2008). More than half of all reports were made by people who came into contact with the child as a consequence of his or her professional position. As of June 2007, there were 18 states with an explicit state law requiring any individual to report suspected child abuse or neglect (Child Welfare Information Gateway 2008).

Once a report is filed, CPS workers screen reports in order to determine whether the reported abuse or neglect qualifies as maltreatment according to the state's legal definition. States define types of abuse or neglect differently. For instance, according to Child Welfare Information Gateway (2008), most states generally define neglect as a caregiver's failure to provide food, clothing, shelter, medical care, or supervision. But, 24 states and the District of Columbia also include a caregiver's failure to educate a child as a form of neglect.

CPS workers only conduct an investigation of the suspected maltreatment if there is sufficient evidence to support an investigation and the reported abuse or neglect falls within the state legal definition of maltreatment. In 2006, an estimated 3.3 million reports of child abuse and neglect were received by state child welfare agencies. Of these, approximately 62 percent received a follow-up investigation (U.S. Department of Health and Human Services 2008). Still, not all cases that undergo investigation actually determine that an act of abuse or neglect occurred. In 2006, 60 percent of the cases that were investigated ultimately were substantiated.

Children may receive mental health services regardless of whether the reported abuse or neglect is substantiated or at any time during the child's contact with the child welfare system. In some cases, CPS workers might refer the child to receive mental health services while in other cases, mental health services might be court ordered. Much of this variation in mental health service provision can be traced to preferences of the CPS worker and policies of the CPS agency. As a result, children who receive mental health services and children who do not are likely to differ in observable and unobservable factors, including history of maltreatment, parental cooperation with CPS, or their socioeconomic factors (Shlonsky 2007).

I cannot simultaneously observe a child both receiving services and not receiving services, or randomly assign children to services. I must rely on statistical methods to adjust for selection bias in who receives services.

I model whether a child receives outpatient services as a propensity p_i^* that depends on observable characteristics of the child and her environment, w_i:

$$p_i^* = w_i\alpha + e_i \quad (37)$$

where α is a vector of parameters and e_i is a random error. If any of the determinants of a child's receipt of mental health services w_i also affect her production of mental health but are not included in the frontier model, then the treatment indicator variable in the frontier model is correlated with the error term ε_i (in the frontier model). Estimators of β are biased if they do not account for the endogeneity of the nonrandom distribution of mental health services. Also, the model suffers from omitted variable bias if the selection model is incorrectly specified. Omitted variable bias will occur when variables that affect or are correlated with mental health are omitted from the selection model.

To my knowledge Mayen, Balagtas, and Alexander (2010) is the only study to apply PSM in productivity analysis to address selection bias.[18] They examine whether technical efficiency varies across organic and conventional dairy farms. Mayen et al. (2010) use PSM to address the farmers' selection into organic or conventional farming. I adopt the methodology of Mayen et al. (2010) to utilize PSM and stochastic frontier analysis to examine the effect of outpatient services on the mental health of children who come into contact with their state child welfare system.

Children are divided into two groups: children who received mental health services and children who did not receive services. Then, children from each group are matched on their observable characteristics. The difference between the matched children's mental health is the measured effect of the treatment.

The effect of mental health services on the mental health of children is defined as $E(Y_1 - Y_0|T = 1) = E(Y_1|Z, T = 1) - E(Y_0|Z, T = 1)$ where Y_0 is the mental health of a child who does not receive mental health services ($T=0$), Y_1 is the mental health of a child who does receive mental health services ($T=1$), and Z is a vector of conditioning variables consisting of any x variables from the frontier model and any w variables from the propensity model. $E(Y_1 - Y_0|Z, T = 1)$ is referred to as the average treatment effect on the treated (ATT).

18. Greene (2010) provides a method to correct for sample selection with the stochastic frontier model that builds on Heckman's (1976; 1979) sample selection model. Heckman's sample selection model is an alternative method to correct for sample selection but is not pursued in this paper. It is an area for future work.

We do not observe $E(Y_0|Z, T = 1)$ but we do observe the mental health of children who do not receive the services, $E(Y_0|Z, T = 0)$. It is possible therefore to calculate:

$$E(Y_1|Z, T = 1) - E(Y_0|Z, T = 0) = \delta. \quad (38)$$

The difference between δ and ATT can be calculated as:

$$\delta = E(Y_1|Z, T = 1) - E(Y_0|Z, T = 1) + E(Y_0|Z, T = 1) - E(Y_0|Z, T = 0) \quad (39)$$

$$\delta = ATT + E(Y_0|Z, T = 1) - E(Y_0|Z, T = 0)$$

$$\delta = ATT + \Omega$$

where Ω measures selection bias.

Predicted ATE ($\widehat{ATE}$) estimates the difference in outcome for between the treated and untreated groups:

$$\widehat{ATE} = E(Y|T = 1) - E(Y|T = 0). \quad (40)$$

Following Mayen et al. (2010), I first estimate a probability model for the receipt of mental health services. The estimates measure the probability or propensity score of receiving mental health services for each child. Second, each child who received mental health services is matched to a child who did not receive mental health services with a similar propensity score.

Tests of PSM Assumptions

In order to conduct propensity score matching, two assumptions must be met: conditional independence and common support condition. Conditional independence assures that the potential outcomes are independent of the treatment status after controlling for observable covariates. Intuitively, the treatment assignment is similar to a random assignment of treatment if the data satisfy conditional independence. The common support condition guarantees that every observation has a positive probability of being both treated and untreated.

Together the conditional independence and conditional support assumptions guarantee that the distribution of the covariates somewhat overlap. Overlap ensures that the regression does not completely extrapolate from the behavior of one group to predict the behavior of the other. In the absence of overlapping covariates, the PSM method would predict the effect of mental health services on the mental health of children who do not and will not receive services.

Unobservable variables may affect the child's receipt of mental health services. Such unobservable variables are not directly controlled for and so limit the results of the PSM approach. I assume that the distributions of unobservable factors are the same for all children, regardless of whether or not they receive mental health services. PSM is valid if unobservable factors that affect a child's receipt of mental health services are independent of her level of mental health (see Imbens (2004)). I conduct formal tests of the endogeneity of the treatment dummy variable in the frontier model to provide some empirical evidence that the PSM approach is in fact eliminating selection bias. However, the extent that results are limited by unobservable variables is an empirical question that can only be determined with better data. With better data though, such an issue of selection bias may not persist.

Empirical Model: Propensity Score Matching

The propensity score for each child is the estimated as the probability that a child *i* will receive mental health services over the sampling period. Propensity scores are estimated using the following probit specification:

$$Pr(o_i = 1) = w_i\alpha + e_i \quad (41)$$

where *o* is a binary outcome for whether the child received mental health services (*o=1*) or did not receive services (*o=0*). w_i is a vector of child-specific characteristics, the availability of services, and institutional factors, α is a vector of parameters, and *e* is an error term.

I use caliper matching without replacement as my matching algorithm. Each child who receives services is matched with children who have a propensity score within a specified range but did not receive services. Robustness checks are presented for various matching methods (see appendix C). A subsample is created of only children who were matched (Dehejia and Wahba 2002).

Data

I use data on children who came into contact with state CPS in the 46 states and the District of Columbia surveyed in the NSCAW.[19] NSCAW is the first national

19. Four states were excluded from the NSCAW study because they have a state law that requires that the caregiver of any child selected for the study be first contacted by the CPS agency staff rather than

probability sample of this population.[20] The first cohort of NSCAW includes two samples: 5,501 children involved in CPS investigations and 727 children in long-term foster care (LTFC). The observations in NSCAW are drawn without replacement from a two-stage stratified sample. At the first stage, there are nine strata. The first eight are the states with the largest number of CPS caseloads. The ninth stratum consists of the remaining states. Primary sampling units (PSUs) were formed and selected within these strata. In most cases, PSUs correspond to a county or a group of adjacent counties. In more densely populated places, a PSU may be a single CPS agency.[21]

Children in the CPS and LTFC samples have distinctly different histories with the child welfare system. Children in the long-term foster care sample had been in long-term care for at least one year prior to baseline interviews. In contrast, the sampling frame for the CPS sample includes all children who come into contact with their state child welfare system. Some of these children never received services and most were never removed from their guardians. My sample includes children from both the CPS and LTFC samples to include a broader range of children with different histories with the child welfare system.

NSCAW is a panel. I use data collected at three points in time for each child: between November 1999 and April 2001 (baseline); 18 months after baseline interviews

---

by NSCAW personnel. The target population subsequently is “all children in the U.S. who are subjects of child abuse or neglect investigations (or assessments) conducted by CPS and who live in states not requiring agency first contact,” (National Data Archive on Child Abuse and Neglect 2008, 2-13).

20. I use sample probability weights in my model estimation to appropriately account for the survey structure and estimate standard errors.

21. For this analysis, estimation cannot simultaneously control for a child’s state of residence and CPS agency. The small sample size limits child-level analysis using CPS agency indicators.

(wave 2); and 36 months later (wave 3).[22] NSCAW data therefore allows for a child-level longitudinal analysis of children who received services between the investigation and initial interview and those who did not (Webb, Dowd, Harden, Ladsverk, and Testa 2010). Over the sampling period, children may enter or exit the child welfare system, or they may come back into contact with the system after leaving it. Likewise, children may receive mental health services at any point in time over the sampling period or never receive services.

At each wave, current caregivers were asked about the child's mental health service receipt using the Child and Adolescent Services Assessment (CASA), the Child Health Questionnaire from National Evaluation of Family Support Programs, Brief Global Health Inventory, and NSCAW developed questions on services (Ascher, Farmer, Burns, and Angold 1996).[23] CASA was developed as a survey instrument for children ages 8 to 18 (Ascher et al. 1996). Survey questions about the child's receipt of mental health services, however, were administered to children ages 6 to 18. There is no specific reason cited in the NSCAW data manual to explain why particular age cut-offs were imposed. To ensure a consistent group of mental health services is developmentally appropriate for all children in my study, I restrict the sample to the 2,482 children 6 years old or older at the time of the initial interview and 18 years old or younger at wave 3.[24]

---

22. NSCAW data include five waves data for the CPS sample but only four waves for the LTFC sample. To construct the largest sample size, only data collected at baseline, wave 2, and wave 3 were used for this analysis. Of the 1075 children in the analysis sample, 167 children are in the LTFC sample.

23. Provider assessment of youths' mental health problems and provider knowledge of available services have been shown to be more significant determinants of service provision compared to youths' self-reported mental health (see Stiffman, Hadley-Ives, Dore, Polgar, Horvath, Striley, and Elze (2000)).

24. See appendix B for a discussion of the sample creation.

Measures and Descriptive Statistics

Functioning: Mental Health

I use the Child Behavior Checklist (CBCL) as the main indicator of a child's mental health outcome. The CBCL is a widely-used and well-established measure of emotional and behavioral problems in children and their need for mental health treatment (Achenbach 1991).[25] On the CBCL, caregivers are queried about the frequency with which children exhibit 113 problem behaviors. This is a composite measure that results from the individual queried questions. Children are scored based on the individual questions. I consider a child's aggregate score.

Following Achenbach (1991), a child with an aggregate CBCL score of 64 or above is identified as clinical. I restrict the sample to the 1,075 children who are clinical at baseline since mental health problems are most prevalent among these children at baseline, when they come into contact with the child welfare system.

Table 3 shows descriptive statistics of the CBCL scores for the sample at the start and end of the sampling period. Descriptive statistics reported here are representative of the total sample of children age 6 or older at baseline and age 18 or younger at wave 3.[26] The mean CBCL score of all children age 6 to 18 years declined from 60 to 58 from

25. In this paper, I examine the standardized CBCL scores, not the raw CBCL scores of children. Raw CBCL scores are standardized using T scores. Z-scores were calculated using population mean 60.1 and standard deviation 12.7 (National Data Archive on Child Abuse and Neglect 2008, 5-33). Henceforth, reference to CBCL scores implies standardized CBCL scores.

26. I do not correct for the complex survey design in sample statistics (National Data Archive on Child Abuse and Neglect 2008). Instead, I discuss the sample as given and use a matching method to simulate a randomized subsample of children.

Table 3. Mental Health Outcomes of Children Ages 6 to 18 Years Old by Clinical Status: Analysis Sample

	Mean		SD	Min	Max
Total Sample					
CBCL score at baseline	60.016	***	12.558	23	94
CBCL score at wave 3	57.510		12.777	23	94
Children who were clinical at baseline					
CBCL score at baseline	71.681	***	5.758	68	94
CBCL score at wave 3	64.214		10.796	24	94

Note: The total sample includes 2,482 children of which 1,075 children were clinical at baseline. Standard deviation (SD), minimum (Min), and maximum (Max) are reported. Statistical significance is reported for differences in the means of CBCL scores across waves for each sample.

*p < .05 **p < .01 ***p < .001

baseline to wave 3 (t=10.407, p<0.001) (see table 3). The mean CBCL score also declined for children who were clinical at baseline from 72 to 64 (t=23.396, p<.001). The decline in CBCL scores was greatest for children who were clinical at baseline compared to the decline in CBCL scores for the entire sample.

Resources: Outpatient Services

I consider outpatient services to be the resources which a child converts in order to achieve a level of mental health functioning. Outpatient services are services that may have been received from an outpatient drug or alcohol clinic, a mental health or community center, a private professional, or a non-psychiatric doctor. These services aim to help children with emotional, behavioral, learning, attentional, or substance abuse problems. At each wave, current caregivers are asked about the child's outpatient service receipt using the CASA, the Child Health Questionnaire from National Evaluation of

Family Support Programs, Brief Global Health Inventory, and NSCAW developed questions on services (Ascher et al. 1996).

I construct a binary indicator for whether or not a child received outpatient services any time during the sampling period. For this analysis, I exclude children who ever received outpatient services before the sampling period[27]. The final group of children who received outpatient services during the sampling period includes 1061 children. The time frame restrictions on when a child received outpatient services attempts to capture only children who received outpatient services during the observational period in order to isolate the effect of services over the observational period. Also, treatment at any time during the sampling period might affect a child's mental health frontier at wave 3. This analysis attempts to capture whether or not outpatient services actually have the intended lasting effects on the mental health of children who receive these services.

Table 4 describes the use of outpatient services by children in the sample over the sampling period. At baseline, 475 sampled children received outpatient services. More children had received outpatient services at wave 2 (approximately 18 months later), at 614 children. Fewer children received services at wave 3 compared to wave 2 but more than at baseline at 623 children. While the number of sampled children who received outpatient services at any given wave are statistically different, the number of children who moved in and out of treatment between waves is not. For instance, the number of children who received services at baseline (193 children) was unchanged at wave 2 (170

27. This restriction does not exclude any children from the sampling who were not already excluded due to partial missing data or age restrictions.

Table 4. Statistics for the Use of Outpatient Services Over the Sampling Period Among Children Ages 6 to 18 years old at Initial Contact: Analysis Sample

	Total sample		Children who were clinical at baseline	
Received services	N	PCT	N	PCT
At least once	1061	42.748	660	61.395
Never	1421	57.252	415	38.605
Baseline	475	44.769	301	45.606
Wave 2	614	57.870	399	60.455
Wave 3	623	58.718	424	64.242
Baseline only	193	18.190	107	16.212
Wave 2 only	170	16.023	88	13.333
Wave 3 only	180	16.965	100	15.152
Baseline and wave 2	75	7.069	41	6.212
Baseline and wave 3	74	6.975	54	8.182
Wave 2 and wave 3	236	22.243	171	25.909
Across all waves	133	12.535	99	15.000

Note: Percent (PCT). Sample size (N).

children) and wave 3 (180 children) ($p>.10$). Of the 1061 sampled children who received services during the sampling period, 133 children received services at every wave.

The trend in outpatient service use over the sampling period is generally consistent for the subset of children who are clinical at baseline (see table 5). Of all children who received outpatient services, 46 percent of the children received services at baseline. This is smaller than the 60 percent of children who received services at wave 2 and the 64 percent of children who received services at wave 3. A greater percent of children received outpatient services at any given wave if they were clinical compared to the total sample.

The steady flow of children in and out of service over the sampling period provides further support for defining the group of children who received outpatient services as all children who received these services at any point over the sampling period.

Functionings and Resources: CBCL Scores and Outpatient Services

Among children who were clinical at baseline, the mean CBCL scores declined for both children who received outpatient services and children who did not receive these services suggesting that mental health problems are less prevalent among both groups over the sampling period (see table 5). The total CBCL score for children who were clinical at baseline and received outpatient services at some point over the sampling

Table 5. Mental Health Outcomes of Children Ages 6 to 18 Years Old Who Were Clinical at Initial Contact by Receipt of Outpatient Services: Analysis Sample

	Mean	SD	Min	Max	
Received outpatient services					
CBCL score at baseline	72.624	6.009	64	94	***
CBCL score at wave 3	66.476	10.424	26	94	
Did not receive outpatient services					
CBCL score at baseline	70.181	4.987	64	89	***
CBCL score at wave 3	60.617	10.407	24	85	

Note: Of the 1,075 children who were clinical at baseline, 660 received outpatient services and 415 children did not receive services. Standard deviation (SD), minimum (Min), and maximum (Max) are reported.

*$p < .05$ **$p < .01$ ***$p < .001$

period declined from 73 to 66 while the decline was greatest for children who did not receive services, from 70 to 61 (F(1,884)=8.59, p<.001).

The prevalence of mental health problems declines for both children who received outpatient services and children who did not, but the decline is greatest for children who did not receive services. The CBCL score for children who received services declines from 73 to 66 while the CBCL score for children who did not receive services had a lower CBCL score at baseline and at wave 3 with a decline from 70 to 61.

Figure 5 illustrates the distribution of total CBCL scores for children by receipt of outpatient services at the start and end of the sampling period. A majority of the children who received services at some point during the sampling period were clinical at baseline (55 percent) and wave 3 (52 percent). A smaller percent of children who never received outpatient services during the sampling period were clinical at baseline (27 percent) or wave 3 (17 percent).

At baseline, 41 percent of children in the sample had a CBCL score in the clinical

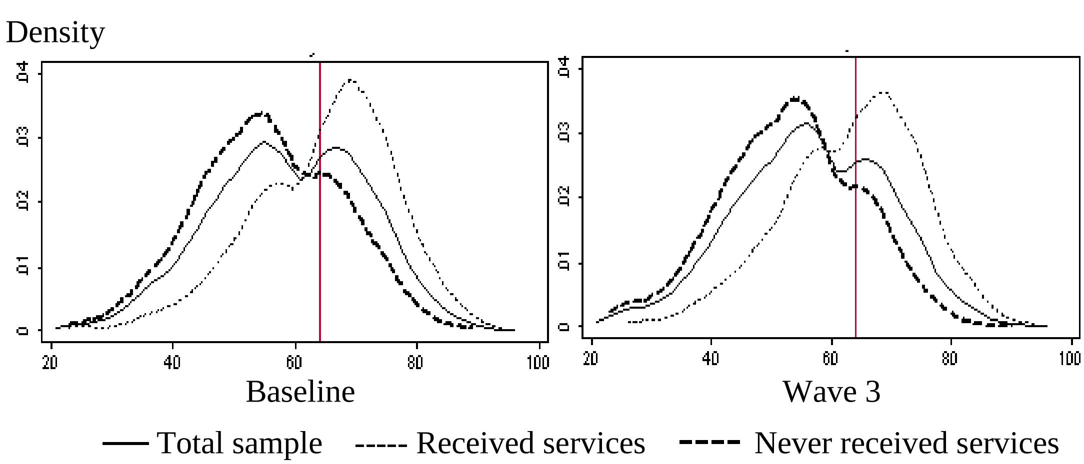

Figure 5. Distribution of Child Behavior Checklist (CBCL) Scores by Receipt of Outpatient Services and Wave

range (children with a CBCL score to the right of the vertical line); 62 percent of children who received services were clinical compared to 29 percent of children who did not receive services. A smaller percent of the sample had a CBCL score in the clinical range at wave 3, at 35 percent. Even still, a greater percent of children who received outpatient services were clinical at wave 3 compared to children who did not receive services at 54 percent and 21 percent.

The striking differences in CBCL scores over the sampling period for children who received outpatient services and those who did not suggests that the receipt of services may be a strong predictor of child's total CBCL score. Also, the variability in the decline of these scores for each group further suggests that the relationship between outpatient services and the child's CBCL score maybe more complex than generally understood.

Conversion Factors

For conversion factors, I control for characteristics of the child such as age, sex, race, or history of abuse that may influence their use of outpatient services. I also control for the child's history with the child welfare system by controlling for whether or not the child is adopted, the number of times the child has been in OOH care, if the child resides in OOH care at baseline and the type of OOH care. Other conversion factors include the availability of outpatient services and the institutional structure. To get at service availability, I control for the child's type of health insurance coverage and the number of local community health centers. Institutional factors include the child's state of residence and who initially reported the child abuse or neglect to the CPS agency. Table 6 presents descriptive statistics for the sample of children.

Table 6. Sample Characteristics for Children Age 6 to 18 Years Old at Initial Contact: Analysis Sample

	Total sample		Clinical at baseline			
	Mean	SD	Mean	SD	Min	Max
Age	10.077	2.709	10.240	2.705	6	15
Race						
White	.441	.497	.485	.500	0	1
Black	.328	.470	.305	.461	0	1
Other	.230	.421	.210	.408	0	1
Female	.530	.499	.489	.500	0	1
Type of abuse						
Physical	.299	.458	.315	.465	0	1
Sexual	.179	.383	.203	.402	0	1
Emotional	.138	.345	.144	.351	0	1
Child is adopted	.020	.141	.032	.175	0	1
Type of OOH placement						
Foster care	.161	.367	.196	.397	0	1
Kinship care	.120	.326	.096	.294	0	1
Group home	.025	.157	.037	.189	0	1
Residential care	.027	.161	.045	.207	0	1
Other	.009	.094	.009	.096	0	1
Number of OOH placements	1.398	1.808	1.658	1.923	0	11
Child's health insurance coverage						
Medicaid	.716	.451	.767	.423	0	1
Private insurer	.189	.391	.152	.359	0	1

Table 6 continued.

	Total sample		Clinical at baseline			
	Mean	SD	Mean	SD	Min	Max
Child's health insurance coverage						
None	.095	.293	.081	.273	0	1
Number of local community health centers	1.688	4.000	1.702	4.176	0	37
Who contacted CPS?						
Teacher or school staff	.411	.492	.421	.494	0	1
Doctor	.197	.398	.222	.416	0	1

Note: The total sample includes 2,482 children of which 1,075 children were clinical at baseline. Standard deviation (SD), minimum (Min), and maximum (Max) are reported. Out of home (OOH) care. Child protective services (CPS).

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Conversion Factors: Characteristics of the Child

I control for child-specific conversion factors such as age, race, sex, and history of abuse. Research suggests that a child's race, age, and history of abuse are strong predictors of a child's receipt of mental health services (Garland and Besinger 1997; Garland, Lau, McCabe, Hough, and Landsverk 2005; Hurlburt, Leslie, Landsverk, Barth, Burns, Gibbons, Slymen, and Zhang 2004; Landsverk, Garland, and Leslie 2002; Leslie, Landsverk, Ezzet-Lofstrom, Tschann, Slymen, and Garland 2000; Leslie, Hurlburt, James, Landsverk, Slymen, and Zhang 2005). Older children and those that have been sexually abused are more likely to use outpatient services. Black children are less likely, compared to White children, to receive these services.

Among sampled children, the average child was 10 years old at baseline. The sample was 44 percent White, 33 percent Black, and 23 percent were some other race.

Slightly more than 50 percent of the children were girls. Most children had been physically abused (30 percent) compared to the percent who had been sexually abused (18 percent) or emotionally abused (14 percent).[28]

Conversion Factor: The Child's History with the Child Welfare System

A child's history with the child welfare system may influence her ability to utilize mental health services. For instance, caregivers of children with a prior history with the child welfare system may better understand how to navigate the system to acquire mental health services. Also, if a caseworker has worked previously with a child, the caseworker may use any knowledge from the previous interaction to influence her decision of whether to refer the child for mental health services.

Previous research has found that children in OOH care are more likely to use outpatient services (Leslie et al. 2005). For instance, Leslie et al. (2005) find that children placed in OOH care are two to three times more likely to receive outpatient services compared to children who remain at home. Findings in the literature are inconsistent as to whether OOH placement improves or worsens mental health outcomes (Berger, Bruch, Johnson, James, and Rubin 2009; Berzin 2008; Courtney 2000; Doyle 2007; McDonald, Westerfelt, and Piliavin 1996; Waldfogel 2000).

I consider four OOH living arrangements: kinship care, family foster care, group home, or residential treatment facility. Kinship caregivers include relatives of the child either by birth or marriage but who are not parents of the child. Group homes include

28. Children may have experienced more than one type of abuse.

orphanages. At initial contact, 16 percent of the children lived in foster care and 12 percent lived in kinship care. Fewer children lived in a group home (3 percent) or residential care facility (3 percent).

I also control for the number of times a child has lived in OOH care as this may be a contributing factor of a child's mental health. Newton, Litrownik, and Landsverk (2000) find that multiple placement changes significantly increases a child's risk of mental health problems at the end of one year. The extent to which the multiple placements cause greater need for mental health care, or whether multiple placements simply capture the high propensity for foster parents to ask for the child to be moved is not known (Rubin, O'Reilly, Luan, and Localio 2007). On average, children in this sample have been in OOH care once by the time of baseline interviews. Children in the sample who have been in OOH care at least once have experienced an average of three OOH placements prior to baseline interviews.

Children who have been adopted may have been in OOH placement at some point prior to the adoption. Relatively few sampled children were in adoptive care at baseline at 64 children (2 percent). Yet these children may have had a history with child welfare systems which could influence their utilization of outpatient services.

Conversion Factor: Availability of Outpatient Services

When a child does not receive a service, it is not known whether any particular mental health service is available or offered but its use declined.[29] I am, however, able to control for ability to pay by observing the child's health insurance status and the possible availability of outpatient services by observing the number of local community health centers. The availability of outpatient services may alter a child's ability to convert these services into a level of mental health functioning.

A large percent of sampled children may face financial barriers to access outpatient services. Approximately 10 percent of children in the sample were uninsured at baseline. Seventy-two percent of the children in the sample were covered by Medicaid or some other state-funded health insurance, and 19 percent were covered by private health insurance[30].

Many CPS agencies work with local community health centers and other service providers to offer a wide array of mental health services (Kolko, Herschell, Costello, and Kolko 2009; McCarthy et al. 2007). Thus the degree of coordination between CPS agencies and these service providers can also be a limiting factor in the availability of services for children and families who come into contact with their state child welfare system. At baseline, children had an average of two community health centers in their community, though there was great variation in the distribution of local community

---

29. Studies consistently document that many children who are in need of mental health services do not receive them (Hurlburt et al. 2004; see also Burns et al. 2004; Leslie et al. 2000; Leslie et al. 2005; Stahmer et al. 2005).

30. Caregivers were only able to denote one form of health insurance coverage for each child.

health centers. Some caregivers reported no local community health center nearby while others reported more than 20 facilities nearby.

Conversion Factor: Institutional Structure

In order to control for variation in state child welfare programs, I include state dummy variables. State effects may capture how the abilities of children to convert resources into mental health functionings are limited across states. NSCAW data allow me to control for the eight states in which the largest number of CPS caseloads were filed. (The remaining 38 states comprise a separate sampling unit.) NSCAW data is the first national data set that allows for such state-level analysis.

In the PSM model, I also control for reports of child abuse or neglect that were filed by teachers or other school staff and doctors with CPS agencies. These individuals are mandated reporters but their report also proxies as a measure of the severity of abuse or neglect. Doctors who filed a report for possible abuse or neglect likely observed substantial evidence of abuse or neglect given the possible avenues by which they interacted with children. This is also true for teachers and other school staff who spend a substantial amount of time with children each week. Indeed, teachers or other school staff filed 41 percent of all initial reports of sampled children. Approximately 20 percent of reports were filed by doctors (see table 6).

Results and Discussion

I use propensity score matching (PSM) to identify a subsample of children who differ only in whether or not they received outpatient services. I then run the stochastic frontier model on the subsample of children to model the conversion of mental health services into mental health outcomes given the child's conversion factors and level of efficiency in converting resources into outcomes.

Propensity Score Matching Analysis

I begin by evaluating whether there is reason to be concerned with selection into outpatient services. Before correcting for selection into services, there are differences in the means for continuous variables and the proportions for binary variables of children who received outpatient services and children who did not (see table 7). First, children who received outpatient services had higher CBCL scores at baseline, were less likely to be Black, more likely to be male, and more likely to have been physically abused. These children had been in a greater number of OOH arrangements prior to baseline interviews, and were more likely to be adopted or to currently reside in foster care, a group home, or residential care. Children who received outpatient services also lived near fewer community health centers, and were more likely to be covered by Medicaid and less likely to be covered by private health insurance. The PSM procedure creates a matched sample in which the mean covariates do not differ for children who received outpatient services from those who did not.

I conduct a Durbin-Wu-Hausman (DWH) test of the endogeneity of the treatment dummy variable (i.e. the indicator variable that a child received outpatient services at least once during the sampling period). The DWH procedure is to test for endogenous regressors. First, I estimate a probit model (41) for whether or not a child will receive outpatient services using the entire sample. I then estimate the stochastic frontier model (36) with both the treatment indicator variable and the residuals from the probit model. I

Table 7. Sample Means of Covariates for the Total Sample and PSM Sub-Sample by Receipt of Outpatient Services.

	Total sample			PSM sample				
				Selection model without states			Selection model with states	
Received outpatient services?	Yes	No		Yes	No		Yes	No
Baseline CBCL (ln)	4.282	4.249	***	4.257	4.264		4.259	4.265
Age (ln)	2.299	2.274		2.268	2.283		2.267	2.267
Age-squared (ln)	5.359	5.251		5.217	5.291		5.220	5.218
Race								
Black	.255	.386	***	.353	.276	*	.325	.299
Other	.209	.212		.191	.234		.231	.199
Female	.445	.559	***	.502	.452		.486	.481
Type of abuse								
Physical	.335	.284	*	.311	.293		.297	.307
Sexual	.217	.181		.198	.192		.192	.183
Emotional	.156	.125		.148	.146		.157	.133
Child is adopted	.042	.014	**	.018	.025		.017	.021
Type of OOH placement								
Foster care	.247	.116	***	.184	.176		.147	.183
Kinship care	.097	.094		.078	.096		.122	.100
Group home	.059	.002	***	.000	.004		.004	.004
Residential care	.067	.010	***	.025	.017		.021	.017
Other	.012	.005		.000	.008		.007	.008
Number of OOH placements (ln)	.898	.466	***	.574	.628		.596	.629
Health insurance								
Medicaid	.814	.694	***	.749	.715		.734	.751
Private insurer	.130	.186	**	.173	.184		.171	.178

Table 7 continued.

Received outpatient services?	Total sample Yes	Total sample No		PSM sample: Selection model without states Yes	Selection model without states No	Selection model with states Yes	Selection model with states No
Number of local community health centers (ln)	.598	.689	**	.658	.642	.645	.606
Who contacted CPS?							
Teacher or school staff	.403	.451		.417	.435	.406	.440
Doctor	.239	.195	*	.223	.234	.203	.207
State (binary)							
California	.098	.094				.115	.091
Florida	.032	.048				.035	.033
Illinois	.058	.048				.066	.058
Michigan	.064	.058				.084	.066
New York	.048	.070				.077	.054
Ohio	.079	.080				.073	.083
Pennsylvania	.071	.046	*			.052	.054
Texas	.086	.092				.063	.058

Note: The outcome is the CBCL scores (ln) at wave 3. The matching algorithm is caliper (.005) without replacement. Standard deviation (SD), minimum (Min), and maximum (Max) are reported. Child Behavior Checklist (CBCL). Out of home (OOH) care. Child protective services (CPS).

*$p < .05$ **$p < .01$ ***$p < .001$

reject the null hypothesis that the treatment indicator is exogenous in the frontier model ($\chi^2(1)=3.09$, p=.079). I conclude it is necessary to correct for the endogeneity of a child's receipt of outpatient services.

In order to examine state effects on a child's propensity to receive outpatient services, I estimate the probit model (41) using two specifications: with state indicators and without state indicators. The motivation for including state-effects in the selection model is to capture differences in the delivery of outpatient services to children across states. This is justified given the extent of variation in mental health assessment of children when they enter the system and differences in state policies in which children receive services that have been documented in the literature, particularly in the reports of the CFSR.

I first estimate the probit model (41) without state indicator variables. The model correctly predicts a child's receipt of outpatient services for 71 percent of the sample. The probit model (41) without state indicators accurately predicts that a child receives outpatient services for 80 percent (0.80=529/660) of all children who actually received services (see table 8). The model is less successful at correctly predicting for children who did not receive outpatient services with only 57 percent (0.57=236/415) correct

Table 8. Prediction Table of Probit Estimates for the Total Sample Using the PSM Model with State Effects

		Actual receipt of services			
		Without state effects		With state effects	
Estimated receipt of services	Total	Received	Did not receive	Received	Did not receive
Received	708	529	179	528	180
Did not receive	367	131	236	132	235
Total	1075	660	415	660	415

Note: An estimate is classified as predicting a receipt of outpatient services if the propensity score is greater than or equal to 0.5. The number of observations is reported.

cases. I conduct Pearson goodness-of-fit chi-square test to examine the overall fit of the model. The Pearson chi-square statistic of 574.97 (DF=552, p=.241) suggests a good overall fit of the model.

When I include state effects, the probit model correctly predicts a child's receipt of outpatient services for children who received services in 80 percent of the sampled cases (where 0.80=528/660). The probit model incorrectly predicts a child's receipt of outpatient services for children who did not receive services for approximately 43 percent of the sample (where 0.43=180/415), and incorrectly predicts a child did not receive services when she in fact did receive a service for approximately 20 percent of the cases (where 0.20=132/660). The probit model with state effects correctly predicts that a child did not receive outpatient services for children who did not receive services at 64 percent (where 0.64=235/367). The probit model correctly predicts whether or not a child received outpatient services in 71 percent of all the cases. The Pearson chi-square statistic of 574.83 (DF=545, p=.182) suggests a good overall fit of the model.

After creating the PSM subsample and running the same probit model on the PSM subsample, the model without state effects only correctly predicts 54 percent of all cases while the model with state effects only correctly predicts 47 percent of all cases. This suggests that PSM created a subsample of observations that differ only in their receipt of outpatient services. Controlling for state effects helps to remove some of the observable differences in the children such that the probit model has more difficulty accurately predicting which children receive outpatient services.

Marginal effects from the probit model (41) for the specifications excluding and including state indicators are presented in table 9. This table shows the marginal effects

Table 9. Estimated Marginal Effects for the PSM Model with States and Predicting a Child's Receipt of Outpatient Services

	Child A	Child B	Child C
Baseline CBCL score	1.432 ***	1.312 ***	1.219 ***
	(.234)	(.218)	(.209)
Age (ln)	1.097	1.005	.934
	(1.052)	(.963)	(.893)
Age-squared (ln)	-.255	-.233	-.217
	(.234)	(.215)	(.199)
Race			
Black	-.215 ***	-.239 ***	-.232 ***
	(.040)	(.039)	(.038)
Other	-.067	-.058	-.054
	(.047)	(.039)	(.035)
Female	-.079 **	-.068 **	-.062 **
	(.035)	(.030)	(.028)
Type of abuse			
Physical	.055	.048	.048
	(.038)	(.034)	(.034)
Sexual	.079 *	.077 *	.072 *
	(.043)	(.045)	(.042)
Emotional	.020	.019	.018
	(.049)	(.047)	(.044)
Child is adopted	.205 **	.226 **	.219 **
	(.085)	(.108)	(.109)
Type of OOH placement			
Foster care	.101 *	.096 *	.091 *
	(.057)	(.056)	(.054)

Table 9. continued.

	Child A		Child B		Child C	
Kinship care	.003		.002		.002	
	(.065)		(.059)		(.055)	
Group home	.368	***	.495	***	.505	***
	(.066)		(.111)		(.124)	
Residential care	.281	***	.335	***	.331	***
	(.074)		(.111)		(.117)	
Other	.122		.124		.118	
	(.175)		(.197)		(.192)	
Number of OOH placements (ln)	.206	***	.188	***	.175	***
	(.038)		(.031)		(.029)	
Health insurance						
Medicaid	.146	**	.119	**	.108	**
	(.062)		(.048)		(.042)	
Private insurer	.106		.106		.100	
	(.066)		(.071)		(.069)	
Number of local community health centers (ln)	-.054	**	-.049	**	-.046	**
	(.026)		(.024)		(.022)	
Who contacted CPS?						
Teacher or school staff	-.037		-.035		-.033	
	(.034)		(.032)		(.030)	
Doctor	.067	*	.065		.061	
	(.040)		(.040)		(.038)	
State						
California	.035		.033		.031	
	(.062)		(.059)		(.056)	

Table 9 continued.

	Child A	Child B	Child C
Florida	-.044	-.038	-.035
	(.090)	(.078)	(.071)
Illinois	.056	.054	.051
	(.075)	(.074)	(.070)
Michigan	-.057	-.050	-.046
	(.073)	(.062)	(.056)
New York	-.058	-.050	-.046
	(.073)	(.062)	(.056)
Ohio	-.048	-.042	-.039
	(.067)	(.058)	(.053)
Pennsylvania	.058	.055	.052
	(.073)	(.073)	(.069)
Texas	.012	.011	.010
	(.067)	(.062)	(.058)
Receipt of outpatient services	.572	.324	.276

Note: Standard errors in parenthesis. Child Behavior Checklist (CBCL). Out of home (OOH) care. Child protective services (CPS).

Marginal effects were calculated for three children: A, B, and C. All three children are 10 year old Black males with a baseline CBCL score of 74, and live in the ninth strata (one of the other 38 states) near two community health centers. The children have Medicaid health insurance. Their initial report of abuse or neglect was filed by a teacher. The three children differ in their current and previous interaction with the system. Child A was physically abused and placed in foster care at baseline. This was his first OOH placement. Child B was also physically abused but was not removed from his home. He has never been in OOH care. Finally, child C did not suffer any abuse, lives at home, and has never been in OOH care.

**p* < .05 ***p* < .01 ****p* < .001

for three children who might come into contact with the child welfare system. All three children have a baseline total CBCL score of 72 and are 10 years old (the sample means). I further condition that the children are Black males who live in the ninth strata (one of the other 38 states), near one community health center, and have health insurance through Medicaid. The initial report of abuse or neglect was filed by a teacher. The three children differ in their current and previous interaction with the system. The first child (A) was physically abused but placed in foster care at baseline; this was his first OOH placement. The second child (B) was physically abused, lives at home, and has never been placed in OOH care. The third child (C) did not suffer any abuse (physical, sexual, or emotional) and lives at home. He has never been in OOH care.

For all three children, the child's baseline level of mental health, whether the child is Black, male, had been sexually abused, lived in foster care, a group home, or residential care, if the child was adopted, had Medicaid health insurance, the number of local community health centers nearby, and whether a doctor contacted CPS were strong predictors of whether a child received outpatient services.

Black children are 21 to 24 percent less likely to have received outpatient services with children in OOH care being more likely to receive services. These results support other findings in the literature that Black children disproportionately do not use mental health services. It is not clear from these results if child welfare systems disproportionately distribute services to children by race, although other work has suggested this (Garland, Landsverk, and Lau 2003).

A child's history of sexual abuse was also a significant predictor of whether the child would receive outpatient services at a 7 to 8 percent probability. This was consistent across children, suggesting that a child's recent placement in foster care does not significantly alter the probability that the child will receive outpatient services.

Children in adoptive care are 21 to 23 percent more likely to receive outpatient services compared to children not in OOH care (children B and C) but only 18 percent more likely to receive outpatient services compared to child (A) in foster care.

Children in group homes and residential care facilities were more likely to receive outpatient services across all three children. This is unsurprising since these facilities are likely to provide children with mental health services. Children who lived in group homes were 37 to 51 percent more likely to receive outpatient services compared to the child living in residential care at 28 to 34 percent more likely to receive services compared to children living at home (children B and C). The difference in probabilities across children A, B, and C provide additional evidence that children in OOH care are more likely to receive outpatient services.

I use the probit estimates for the model with state indicators to generate a propensity score for each child. Then, I create a subsample of children who did not receive outpatient services that I match to children who did receive services by selecting for each child with treatment the child without treatment with a propensity score within .005 to that of the treated child without replacement.[31] Propensity scores for all children who did not receive services are included before matching. After matching, however, only the propensity scores for a subset of children who did not receive services are considered – those children who were matched to children who did receive services.

Figure 6 shows kernel density estimates for the distribution of propensity scores for each selection model and treatment group before and after matching. As expected,

31. See appendix C for a discussion of matching algorithms and the robustness of results across matching techniques.

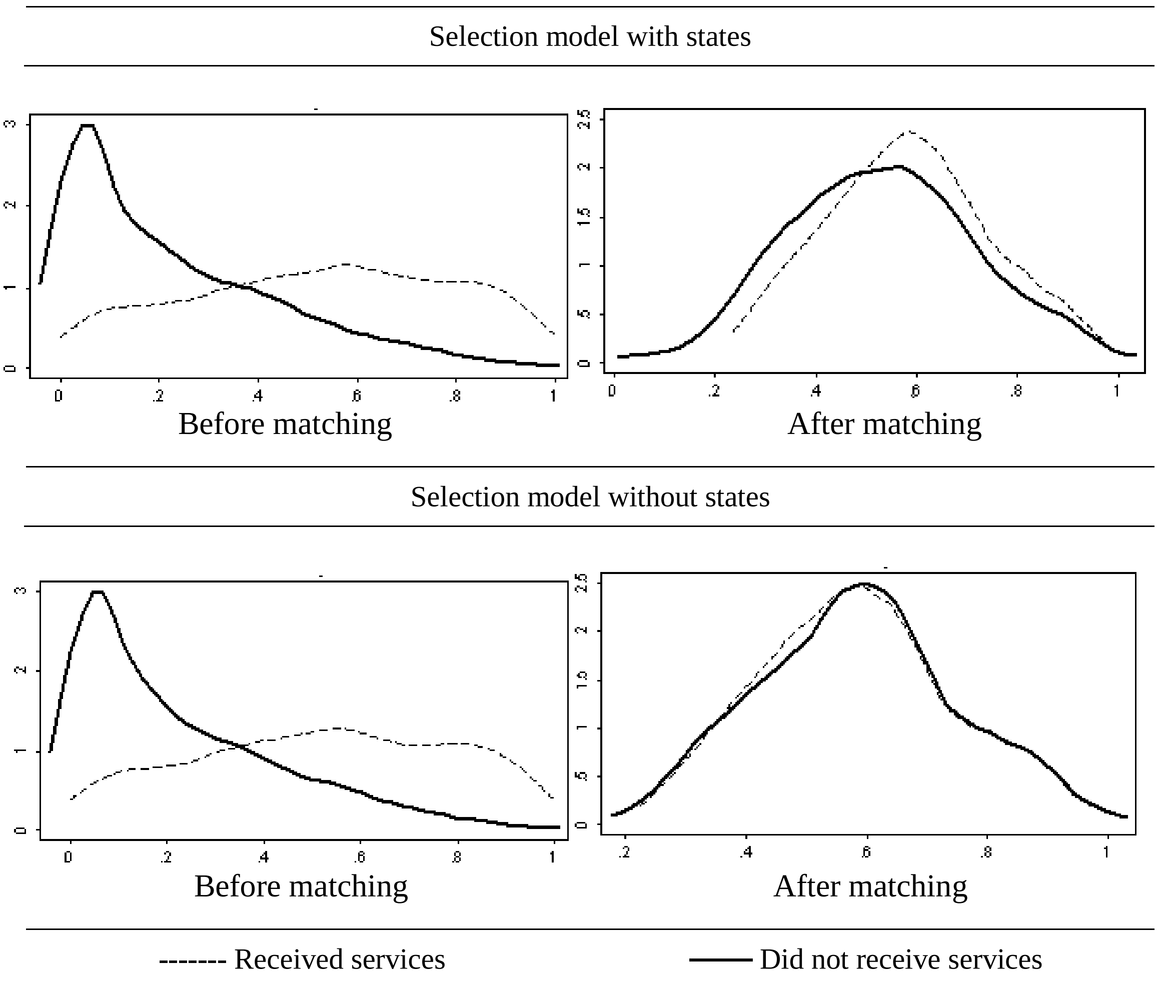

Figure 6. Distribution of Propensity Scores Before and After Matching by Selection Model

children who received outpatient services have a higher predicted propensity to receive services. Before matching, the propensity score distribution for children who received services is skewed left and the propensity score distribution for children who did not receive services is skewed right. After matching, the distribution for the matched subsample more closely resembles that for children who received services.

I also perform the Kolmogorov-Smirnov (KS) test for equality of distributions. I fail to reject the null hypothesis that the distribution of propensity scores for the treated and control groups are drawn from the same distribution (KS statistic=0.028, $p>.10$). This demonstrates that the matched sample satisfies the common support condition.

The matched sample also satisfies the assumption of conditional independence. Table 7 presents the mean of the covariates in the selection model before and after the children are matched. Before matching, the mean of the covariates differ between the children who received outpatient services and the children who did not receive services. After matching, however, the means of the covariates are equal ($p>.10$) across groups, suggesting that the samples are balanced.

The resulting subsample of matched children consists of 575 children (286 children who received outpatient services and 289 children who did not receive outpatient services) for the selection model with state indicators. By construction, the matched groups of children who did and did not receive outpatient services more closely resemble each other in their propensity to receive outpatient services. I again conduct a DWH test for endogeneity of the treatment dummy variable when the stochastic frontier model (36) is estimated over the PSM subsample. Now I cannot reject the null hypothesis that the treatment indicator variable is exogenous ($\chi^2(1)=0.53$, $p=.469$). This suggests that the PSM approach successfully generated a subsample of children for which receipt of outpatient services seems to be randomly assigned.

Stochastic Frontier Analysis

I estimate the stochastic frontier model (36) on the PSM subsample of 575 children in order to estimate a child's mental health capability frontier. The stochastic frontier model describes how a child converts outpatient services into a mental health outcome in the future (at wave 3), controlling for the child's conversion factors including her baseline characteristics, history with the child welfare system, availability of services, and institutional factors.

Effectiveness and Efficiency of Services

Table 10 shows estimates for the stochastic frontier model (36) on the PSM subsample.[32] I estimate a 12 to 13 percent higher CBCL score at wave 3 for children who receive outpatient services at some point during the sampling period ($p<.001$). This suggests that even after controlling for conversion factors and baseline CBCL scores, mental health problems are more prevalent for children who receive outpatient services compared to children who did not receive services.

At first glance, this result might seem to suggest that outpatient services are not effective at reducing the future prevalence of mental health problems for children. However, this result might signal that mental health problems really are a nested problem. For instance, in the capability framework, it is possible outpatient services actually exacerbate a child's mental health problems. A child who is continually told that she has mental health problems and should be treated differently may develop further psychological problems as a result of dealing with the social stigma and isolation due to her mental health. Alternatively, a child who is told that she has mental health problems may develop adaptive expectations wherein she believes it is acceptable behavior for her to exhibit problematic behaviors because others expect her to behave this way. Adaptive

32. I estimated an ordinary least squares (OLS) regression as a variant of the stochastic frontier model on the PSM sample in order to demonstrate empirical differences from using SFA instead of OLS (see appendix D).

Table 10. Covariate Estimates for the Stochastic Frontier: Children Ages 6 to 18 and Clinical at Initial Contact.

	PSM sample				Total sample			
	Without states		With states		Without states		With states	
Received outpatient services	.097	***	.094	***	.096	***	.098	***
	(.020)		(.020)		(.020)		(.020)	
CBCL score at baseline (ln)	1.013	***	.965	***	.822	***	.768	***
	(.164)		(.153)		(.121)		(.117)	
Age (ln)	1.181	**	1.118	*	.009		-.024	
	(.648)		(.637)		(.513)		(.509)	
Age-squared (ln)	-.261	**	-.247	*	-.001		.005	
	(.147)		(.144)		(.115)		(.114)	
Race								
Black	.011		.013		.025		.021	
	(.028)		(.028)		(.018)		(.018)	
Other	-.017		-.023		-.050	**	-.045	*
	(.031)		(.031)		(.025)		(.023)	
Female	.021		.026		-.009		-.005	
	(.024)		(.023)		(.017)		(.017)	
Type of abuse								
Physical	-.043	*	-.045	*	-.034	*	-.033	*
	(.024)		(.023)		(.019)		(.018)	
Sexual	.005		.005		.019		.019	
	(.025)		(.024)		(.017)		(.018)	
Emotional	-.028		-.034		.002		-.005	
	(.032)		(.031)		(.034)		(.032)	
Child is adopted	.085		.082		-.049		-.060	
	(.058)		(.062)		(.065)		(.061)	

Table 10 continued.

	PSM sample				Total sample			
	Without states		With states		Without states		With states	
Type of OOH placement								
Foster care	-.111		-.111		-.018		-.019	
	(.094)		(.092)		(.040)		(.040)	
Kinship care	-.100		-.108	*	-.124	***	-.118	***
	(.063)		(.061)		(.039)		(.040)	
Group home	.066		.098		-.048		-.050	
	(.135)		(.139)		(.059)		(.049)	
Residential care	-.152		-.142		.017		.028	
	(.135)		(.145)		(.048)		(.049)	
Other	-.147	***	-.150	***	-.039		-.043	
	(.048)		(.054)		(.041)		(.043)	
Number of OOH placements (ln)	-.019		-.022		.004		-.002	
	(.028)		(.027)		(.018)		(.019)	
Health insurance								
Medicaid	-.035		-.031		-.016		-.030	
	(.030)		(.032)		(.030)		(.032)	
Private insurer	-.090	**	-.097	**	-.018		-.040	
	(.036)		(.038)		(.032)		(.034)	
Number of local community health centers (ln)	-.004		-.022		-.017		-.025	
	(.019)		(.020)		(.015)		(.016)	
State								
California			.026				-.012	
			(.040)				(.032)	

Table 10 continued.

	PSM sample		Total sample	
	Without states	With states	Without states	With states
Florida		.031		.022
		(.038)		(.028)
Illinois		.014		.052
		(.080)		(.051)
Michigan		.065 **		.012
		(.033)		(.029)
New York		.003		-.130 **
		(.049)		(.055)
Ohio		.182 **		.115 ***
		(.048)		(.041)
Pennsylvania		-.012		-.026
		(.048)		(.027)
Texas		.045		-.034
		(.044)		(.033)
Constant	-1.485	-1.221	.603	.901
Variance of *v*	.157	.154	.153	.150
Variance of *u*	.0001	.0001	.0001	.0001
λ	.001	.001	.001	.001
Wald χ^2	160.23 ***	179.34 ***	164.96 ***	207.14 ***
Log pseudolikelihood	110,925	116,638	234,758	244,573
N	546	546	1029	1029

Note: Standard errors in parenthesis. Child Behavior Checklist (CBCL). Out of home (OOH) care. Child protective services (CPS).

Predicting a child's CBCL score at wave 3 (ln).

*$p < .05$ **$p < .01$ ***$p < .001$

expectations may work in a similar way for caregivers and caseworkers. Caregivers and caseworkers may perceive the child's behavior as more problematic given their repeated exposure to the child and developed expectation that the child will continue to exhibit problematic behaviors. Since caregiver responses are provided for the CBCL scores of children, the higher CBCL score at wave 3 may at least somewhat capture the adaptive expectations of caregivers, in addition to those of the child.

To compare the actual mental health of children by receipt of service, I estimate the average efficiency of mental health outcomes for children by their receipt of outpatient services (see table 11). This analysis examines the child's level of functioning given her resources and conversion factors, and relative to her capability. Again, the PSM technique produced a subsample of children who differ only in their receipt of outpatient services. My results suggest that children are efficient on average regardless of whether or not they receive outpatient services. I conduct a Kruskall-Wallis test of equality of populations test to determine if the average efficiency of mental health outcomes for children who receive outpatient services is statistically different from those mental health outcomes of children who did not receive services. I conclude that the average efficiencies of mental health outcomes of children by receipt of outpatient services are different (for the frontier model without state indicators $\chi^2(1)=4.824$; $p=0.028$) (for the stochastic frontier model with state indicators $\chi^2(1)=5.768$; $p=0.016$). Children who receive services have mental health outcomes that are approximately 0.01 percent more efficient than those of children who do not receive services.

Table 11. Sample Means and Standard Errors of Efficiency

	Received services		Did not receive services		Difference in means
	Mean	SE	Mean	SE	
PSM subsample					
Without states	99.991	3.94e-8	99.991	4.26e-8	**
With states	99.991	3.29e-8	99.991	3.83e-7	**
Total sample					
Without states	99.993	4.10e-8	99.993	4.33e-8	
With states	99.993	1.58e-8	99.993	2.31e-7	**

Note: Standard error (SE).

*$p < .05$ **$p < .01$ ***$p < .001$

While the estimated average technical efficiency of mental health outcomes differs for children by whether they received outpatient services, the magnitude of the difference in the means is quite small at 0.01 percent.

These findings suggest that outpatient services are enhance efficiency conversion. On average, children who receive outpatient services exhibit mental health behaviors that are indicative of their actual mental health. Controlling for the child's state of residence marginally improves average technical efficiency estimates for groups of children by receipt of service.

I also estimate average technical efficiencies for children who did and did not receive outpatient services using the estimated stochastic frontiers on the total sample (i.e. the sample without propensity score matching) in order to demonstrate the empirical difference if I had failed to control for selection into outpatient services (table 11). The estimated average technical efficiency for children who receive outpatient services in the total sample is less than the average technical efficiency for children who received services using the PSM subsample. Failing to control for the nonrandom distribution of services results in a larger estimated average technical efficiency for children who did not receive outpatient services. The difference in the average technical efficiency for each group is smaller for the total sample compared to the PSM subsample. Moreover, average technical efficiency does not differ significantly for children in the total sample if the model excludes state indicators ($\chi^2(1)=.191$; $p=.662$). However, controlling for state indicators results in a significantly different average efficiency for the children by receipt of outpatient services ($\chi^2(1)=4.439$; $p=.035$). Thus, failing to account for the selection bias in the delivery of outpatient services and state effects results in different estimated average technical efficiencies.

State Effects and the Mental Health of Children

State effects capture differences in conversion factors across states. Recall that conversion factors influence the child's ability to utilize resources and achieve some level of well-being. Thus, state effects represent state-specific conversion factors that influence a child's ability to utilize mental health services. For instance, state conversion factors may vary as a result of differences in state child welfare policies or social norms. Child welfare policies are decentralized to state governments so state conversion factors attempt to capture the variation in child welfare policies across states, among other state-specific factors.

Two of the state effects are statistically significant in the frontier model (36): Michigan and Ohio. In both states, the estimated CBCL score at wave 3 is higher than the CBCL scores in the remaining 38 states (i.e. the ninth strata). Children in Michigan have CBCL scores that are 7 percent higher and children in Ohio have CBCL scores that are 18 percent higher. These results capture differences in the mental health outcomes of children solely due to state effects.

The findings from this analysis are among the first to provide motivation for future work on the systematic differences in how state CPS agencies deliver services and the ability of service providers to meet the mental health needs of children who come into contact with state child welfare agencies. Recall, state effects were not strong predictors of whether a child would receive outpatient services (in the PSM model). Combined with the finding that state effects matter in the frontier model suggests that merely knowing that a state provides a particular service is insufficient information to determine the prevalence of mental health problems for a child in the future. Thus, researchers should investigate how services are delivered to children within states, either from CPS agencies or from health practitioners.

A model of service delivery that describes the quality of a service would provide greater insight into the effectiveness of services. The current literature, in contrast, merely describes whether or not a state provides services. The CFSR reports document differences in mental health assessment and provision of services (McCarthy et al. 2007). However, most children in OOH placement or adoptive care receive services through their state Medicaid system since these children are eligible for Medicaid coverage. As a result, the issues that plague the Medicaid system, such as a shortage of physicians and specialized practitioners to provide care for Medicaid patients, directly impact these children. Moreover, Medicaid programs are state operated similar to the child welfare system such that federal monies heavily subsidize state programs but states have the ability to adopt various policies and procedures to deliver health services. A more nuanced control for state Medicaid procedures and coverage would help to decompose state effects into a child welfare component, a Medicaid component, and other state effects.

Other Conversion Factors

In addition to whether the child receives outpatient services and the child's prevalence of mental health problems at initial contact, and without controlling for state effects, only the child's placement in kinship care at baseline is a strong predictor of the child's future CBCL scores. Mental health problems are less prevalent for children in kinship care at wave 3 with a 12 percent decline in their CBCL scores. Controlling for state effects, however, improves the overall fit of the model (Wald χ^2= 179.34, p<.001) and alters the set of significant predictors.

Controlling for state effects, a child's age, placement in residential care, and the number of local community health centers are strong predictors of a child's CBCL scores at wave 3. These predictors are in addition to whether the child receives outpatient services and her CBCL score at baseline. Estimates support previous findings in the literature that older children have relatively poor mental health, and that there are diminishing marginal effects to age. Children in residential treatment facilities are estimated to have a CBCL score that is 16 percent greater than children not in these facilities, all else equal. Also, children who live near relatively more community health centers are estimated to have lower CBCL scores at wave 3. A one percent increase in the number of local community health centers at initial contact is associated with children having a 5 percent lower CBCL score at wave 3.

Extensions of the Analysis and Robustness of Results

Additional PSM models were estimated on various subsamples, with different outcome variables, and using different matching algorithms to check for the robustness of the PSM application for this data sample and for its use in the context of child welfare and the effect of outpatient services.

While the final estimates were estimated over the subsample of children who were clinical at baseline, I also estimated the PSM model (41) over three other subsamples of the data. First, I estimated the PSM model (41) over the entire sample of children age 6 or older at baseline and age 18 or younger at wave 3. The PSM model fails to balance on the covariates for this sample. I then estimated the PSM model (41) over the sample of children who had a CBCL score in the middle 50th percentile of the distribution (a CBCL score between 51 and 69) at baseline. The children still differ substantially on the covariates and the PSM subsample failed to balance. My final extension on the sample examined children with a CBCL score in the upper 50th percentile (60 or above) at baseline. This subsample does balance and pass the common support condition. However, using the median CBCL score to condition the subsample seemed as arbitrary as using the clinical range (64 or above). I ultimately use the measure consistent with the literature, the clinical range. Still, these extensions demonstrate the need to condition on the child's initial level of mental health to appropriately implement the PSM approach. It seems likely that children are sorted into outpatient services based on the prevalence of mental health problems that they exhibited at initial contact with their state child welfare system.

I also considered alternative outcome measures for a child's mental health to test the robustness of the total CBCL score. As previously mentioned, the CBCL is a composite measure. The measure can be decomposed into individual components or additional composite measures that consider only a subset of mental health problems. The two most widely used composite measures of the CBCL measure a child's internalizing behavior (internalizing CBCL) and her externalizing behavior (externalizing CBCL). I run the PSM model (41) on the child's internalizing behavior score and externalizing behavior score instead of the child's aggregate CBCL score to determine if covariates differ across models.

Both extensions satisfy the balancing property and common support condition of PSM, and the set of significant covariates do not vary substantially across models. A child's race, history of abuse, and OOH placement are significant predictors in the original model using the aggregate CBCL score which are still significant in the alternative specifications. For the model using a child's internalizing score as the outcome of interest, whether a doctor filed the initial report and the child's state of residence (for two of the eight indicators) are significant predictors of whether a child received outpatient services. Using a child's externalizing score as the outcome of interest, a child's medical insurance coverage (Medicaid) and gender are also significant predictors in addition to the original set of predictors. A child's history of abuse does not significantly predict a child's propensity to receive outpatient services when her externalizing score is the outcome of interest. The extension of the PSM model (41) to outcome measures provides evidence that the PSM method sorting children into matched groups that at the very least produce similar estimates across measures.

In addition to the extension for the PSM model, I extended the stochastic frontier model to include interaction terms between a child's state of residence and her receipt of outpatient services (see table 12). In addition to the state effects for Ohio and Michigan, California also has a statistically significant conversion factor. I find that children in Michigan who received outpatient services are correlated with higher CBCL scores at wave 3 than children who did not receive these services. But, this was not the case in Ohio or Texas, where children who received outpatient services are correlated with lower CBCL scores than children who did not. The extension of the stochastic frontier model to include interactions provides further evidence that state conversion factors contribute to the effect of outpatient services on the capability and functioning of children. It is not clear from this analysis why conversion factors are significant only for these three states. Moreover, because of the generalized collection of child welfare policies and procedures reported in the CFSR reports, no additional insights can be gained for understanding why state conversion factors are significant only in Ohio, Michigan, and California.

Limitations

The present study has several limitations. First, the estimates presented here calculate the marginal effect of the average child but this cannot inform us about the effect of treatment on children at the margin of treatment. The current study focuses on

Table 12. Covariate Estimates for the Stochastic Frontier Model: Children Ages 6 to 18 and Clinical at Initial Contact.

	PSM sample	
Received outpatient services	.127	***
	(.026)	
CBCL score at baseline (ln)	.954	***
	(.155)	
Age (ln)	.990	
	(.605)	
Age-squared (ln)	-.219	
	(.138)	
Race		
Black	-.0004	
	(.029)	
Other	-.021	
	(.027)	
Female	.019	
	(.024)	
Type of abuse		
Physical	-.049	**
	(.022)	
Sexual	-.009	
	(.026)	
Emotional	-.032	
	(.032)	
Child is adopted	.012	
	(.068)	

Table 12 continued.

	PSM sample	
Type of OOH placement		
Foster care	-.118	
	(.090)	
Kinship care	-.101	
	(.062)	
Group home	.096	
	(.155)	
Residential care	-.150	
	(.136)	
Other	-.170	***
	(.063)	
Number of OOH placements (ln)	-.017	
	(.026)	
Health insurance		
Medicaid	-.041	
	(.034)	
Private insurer	-.109	***
	(.040)	
Number of local community health centers (ln)	-.022	
	(.019)	
State		
California	.034	
	(.049)	
Florida	.048	
	(.069)	

Table 12 continued.

	PSM sample	
Illinois	.011	
	(.133)	
Michigan	.127	***
	(.037)	
New York	.017	
	(.036)	
Ohio	.245	***
	(.051)	
Pennsylvania	-.009	
	(.060)	
Texas	.111	*
	(.067)	
Interactions		
California*Received services	-.021	
	(.068)	
Florida*Received services	-.034	
	(.081)	
Illinois*Received services	.007	
	(.145)	
Michigan*Received services	-.115	*
	(.063)	
New York*Received services	-.053	
	(.086)	
Ohio*Received services	-.171	***
	(.051)	

Table 12 continued

	PSM sample	
Pennsylvania*Received services	-.006	
	(.085)	
Texas*Received services	-.144	*
	(.085)	
Constant	-1.027	
Variance of *v*	.152	
Variance of *u*	.0001	
λ	.001	
Wald χ^2	256.50	***
Log pseudolikelihood	119,901	
N	546	

Note: Standard errors in parenthesis.

Predicting a child's CBCL score at wave 3 (ln).

*p < .05 **p < .01 ***p < .001

children who have clinical mental health problems when they come into contact with their state child welfare system. These children have an urgent need for mental health services. However, there is a large percent of children who have a CBCL score just below the clinical range. These children are 'at the margin of treatment.' It is unclear which of these children will receive services and how state CPS workers will distribute services to these children.

Second, this analysis is limited to a short time horizon. In fact, a child's consumption of mental health services may fluctuate overtime, especially in response to changes in the child's clinical severity. Children do not move linearly through the continuum of mental health. Some children may regress, for instance if an episode of abuse or neglect occurs again. A more nuanced analysis would consider when the service was first provided to the child, changes in the child's service use, the duration of service use, and the child's mental health outcomes over multiple periods of time. Changes in the child's level of mental health and the child's service consumption may help us to better understand the effectiveness of mental health services.

Third, this analysis is limited to a binary indicator of whether or not a child received outpatient services but this does not capture the intensity with which children received outpatient services or the type of outpatient service she received. A child with relatively poor mental health may receive greater doses of outpatient services (Foster 2003, 1190; Salzer, Bickman, Lambert 1999). The measured benefits of outpatient services on a child's mental health requires some control for the intensity at which children receive these services. Controlling for the dosage of outpatient services received would be an attempt to control for the quality of services.

The current NSCAW data are not complete enough to simultaneously control for a child's state of residence and her dose of outpatient services. The survey includes two questions that measure a child's dose of outpatient services: how many days per week did the child receive outpatient services; and how many minutes was each session? (National Data Archive on Child Abuse and Neglect 2008). NSCAW data on the amount of service use, however, have poor response rates. Missing data reduce the sample size to fewer than one hundred observations and limit the analysis possible when state of residence is also controlled for. This paper argues that differences in state delivery systems of mental health services and child welfare systems are important factors in the effectiveness of these services. As such, I do not control for the dosage of outpatient services received by sampled children so that I might proceed with state-level analysis. I acknowledge that the measured effects of services would be better estimated with more complete data on the dose of outpatient services received instead of using a dichotomous indicator for the receipt of services.

An alternative measure for the quality of care is the quality-adjusted life year (QALY). The QALY is widely used and conceptually based in the neoclassical framework (see Brook and McGlynn (1996), Guyatt, Feeny, and Patrick (1993), and Verkerk, Busschbach, and Karssing (2001)). It measures the number of years of life added to an individual's lifetime as a result of an intervention (presumably health interventions aim to increase the longevity of an individual's life). QALY measures, however, suffers from the same weakness as other quality measures based in the neoclassical framework: the measures do not capture distributional differences in how individuals utilize resources. Working within the capability framework, more detailed data will help to improve estimation of the effect of mental health services on the well-being of children.

Fourth, outpatient services include a wide array of different services and this may reduce the measured effect of individual services on a child's mental health. NSCAW data aggregate outpatient services received at an outpatient drug or alcohol clinic, a mental health or community center, a private professional, or a non-psychiatric doctor. Services received at different outpatient facilities cannot be disaggregated to estimate particular effects with each facility. The ideal dataset would allow for a more nuanced estimation of the effect of separate types of outpatient services on the mental health of a child.

Finally, this study controls for state effects but these do not directly capture differences in state child welfare policies. Instead, state effects capture a broader range of differences between states including social norms, available resources, and policies not related to child welfare. It would be better to include indicator variables for specific policies that are implemented in states to directly measure differences in state child welfare policies and to measure the effect of these policies. For instance, indicators as to whether the child received a mental health assessment upon entering the child welfare system, an indicator for who administered the assessment (e.g. caseworker, mental health professional, etc.), the outcome of the assessment, and a control for the length of time between a mental health assessment was administered and the child received mental health services.

Conclusion

The objective of this dissertation was to explain the theoretical and empirical gains of operationalizing the capability framework in policy outcome evaluation. This application to child welfare policy provides researchers with a concrete example of how the capability framework might be operationalized using stochastic frontier analysis, and demonstrates that the effect of a policy goes beyond its impact on increasing the resources available to individuals. In the case of child welfare policy, the effect of providing mental health services to children who come into contact with their state child welfare program is related to state conversion factors, or state-effects on the ability of children to utilize the mental health services.

The results of this application to child welfare policy demonstrate that policies that aim to affect individual well-being influence both the individual's resources and her ability to utilize those resources. In the context of U.S. child welfare policy, state child welfare systems provide mental health services via state Medicaid programs to reduce the prevalence of the mental health problems of children who come into contact with the system. This goal is a part of the overall federal goal under the Adoption and Safe Families Act to improve the well-being of these children.

The results of the policy outcome evaluation using the capability approach show that the effect of the policy is dependent on its impact on conversion factors. My study is the first to provide evidence that state conversion factors contribute to variation in mental health outcomes of children who come in contact with their state child welfare system. This is my key finding. Previous research on the differences in receipt of services and mental health outcomes of children who come in contact with their state child welfare system has not considered differences between states. Until recently, sufficient data to conduct a longitudinal analysis and control for state of residence has not been available.

While I was only able to control for the eight states with the largest number of CPS caseloads, state effects were significant in two cases. My results provide preliminary evidence that there are differences in the mental health outcomes of children who come into contact with state child welfare programs even after controlling for their observable characteristics. Thus the significant state effects motivate future research to better understand how states differ in their provision of mental health services to children.

The federal government is responsible for ensuring that their laws and regulations, such as the Adoption and Safe Families Act, are fulfilled. While the CFSR reports are a first step to hold states accountable for the well-being of children who come into contact with their state child welfare system, a detailed analysis that links state child welfare policies with the outcomes of children is necessary. The CFSR reports should better document the explicit policies in place to deliver services to children and which providers CPS agencies work with to distribute these services. The CFSR reports have the potential to clearly identify state policies that work, and highlight new policy initiatives across states. Moreover, additional research is necessary to identify which factors contribute to a state's delivery of mental health services and improve mental health outcomes of children in a particular state.

APPENDIX A

ASSUMPTIONS OF THE NEOCLASSICAL FRAMEWORK

The neoclassical framework relies on the assumptions of revealed preference theory (see Samuelson (1947)) and utilitarianism to model and predict individual behavior. Below I review these assumptions and relate them to a policy outcome evaluation using the neoclassical framework.

Fundamental Assumptions

Revealed preference theory begins with our observation of an individual's choice behavior over a choice set. The consumer has some ranking of preferences across all alternatives. Thus, she selects a particular alternative from a menu because she knows that she prefers this element to every other element in the menu. The alternatives that the consumer does not reject from the menu comprise a choice set.

Neoclassical economists assume that the consumer's choice behavior can be characterized by a choice function that depends only on the menu and context. A menu is well-defined non-exhaustive listing of the available, mutually-exclusive alternatives from which an individual chooses. A menu is complete with all alternatives available to the consumer. A context consists of variables which influence the consumer's selection of items for the choice set.

We can observe the choice behavior of an individual across various menus and use the outcomes to determine the probabilistic outcome of her choice given other menus containing a subset of elements. Such inference is possible only after making a few assumptions:

Assumption 1. Preferences are complete, reflexive, and transitive.

Assumption 2. Preferences are menu-independent.

Assumption 3. Preferences are context-independent.

Assumptions 2 and 3 ensure consistency of preferences across menus and contexts. Assumption 2 ensures that preferences do not change because of the particular menu of alternatives that are available to an individual. By assumption 3, an individual's preferences are independent of the variables that influence an individual's choice set from a particular menu, and therefore are consistent across menus.

Individual choice is assumed to reveal individual preference. For instance, a person prefers some bundle of goods x to another bundle y if she chooses x from a set of options which includes y. The observed choice behavior reveals that the chosen good is at least as good as whatever goods the individual has rejected. Individual choice behavior therefore allows economists to deduce individual preferences. The concept does not examine the individual's deliberative process of choice to assert her preference rankings. Only the act of choosing, not the motivation of choice, is necessary to identify an individual's preferences. Assumptions (1-3) allow us to rank an individual's preferences for goods and services and make inferences about her future choice behavior in hypothetical situations.

The consumer chooses an alternative from her choice set. The outcome of this selection is assumed to be menu independent wherein the act of choosing an alternative is independent from the menu from which it is selected. Moreover, the menu provides no additional information to the consumer about the alternatives.

Assumption 4. Choice is menu-independent.

Assuming menu-independence, the bundles in the choice space can be completely ordered. The choice space is the union of the sets of alternatives in the domain. Each bundle in the individual's consumption space can be ranked even if we do not observe directly a pair of bundles. For instance, if a person chooses *x* from a set of options which includes *y*, and the same person chooses *y* from a set of options which includes *z* but not *x*, then we can conclude that the person will choose *x* from a set which includes *z*, even if we never observe such a menu of options. Assuming menu-independence also ensures that the menu does not reveal any additional information about the choices.

A choice function is menu-independent if it satisfies the Weak Axiom of Revealed Preference (WARP). WARP guarantees that preferences that are revealed in a given choice situation are assumed never to be reversed in any other choice situation.

Assumption 5. Weak Axiom of Revealed Preference.

By WARP, if *x* is revealed weakly preferred to *y*, then it is not the case that *y* is revealed strictly preferred to *x*. WARP allows for bundles to be ranked but it is a weak assumption since cyclical preferences can occur. A stronger assumption is necessary to ensure that cyclical preferences do not occur. Thus, neoclassical economists assume the Strong Axiom of Revealed Preference (SARP).

Assumption 6. Strong Axiom of Revealed Preference.

SARP ensures that if *x* is preferred to *y*, then it is not possible to have *y* preferred to *x*, where *x* and *y* are different goods. SARP rules out two outcomes that do not violate WARP. First, two choices cannot both be directly revealed preferred to one another. Second, SARP rules out chains of choices that ultimately lead to two choices that are each revealed preferred to another. The satisfaction of SARP suggests some preference ordering or utility function since there is a ranking of outcomes. SARP, in this regard, identifies an individual's choice behavior and describes what bundles are chosen when other bundles could have been chosen.

Neoclassical economists use utility and utility functions to measure an individual's level of welfare. That is, utility is a measure of an individual's state of being in the neoclassical framework. Utility is an ordinal representation of individual choice behavior, or her achieved level of happiness from consuming a bundle of goods.[33] Inferences about utility follow directly from preference orderings: suppose *x* provides greater utility than *y*, then given the choice of *x* or *y*, the individual will choose *x*. Of course, different combinations of goods may provide an individual with the same level of

33. Ordinal utility permits alternative bundles of goods to be ordered such that an individual may consider one bundle to be worse than, equal to, or better than the other. Utils, however, do not represent a numerical scale of intrinsic meaning.

utility. A utility function describes the various combinations of goods and services that an individual may consume in order to achieve a given level of welfare. The utility function is derived from the individual's preference rankings of goods and services, which are revealed by her choice behavior.

It is important to recall that revealed preference theory does not consider the process of choice. Therefore, we cannot conclude anything more than that the set of bundles described by the utility function provide some level of satisfaction for the individual. Since any consumption bundle is assigned only one level of utility, utility functions are disjoint, and the collection of all utility functions describes the consumption space. Thus, we can identify the level of utility received for any consumption bundle in an individual's consumption space.

Assumption 7. Local non-satiation of consumption.

Assumption 8. Resources are scarce.

The utilitarian framework allows us to associate a metric with a preference ranking. The number of utils assigned to a particular bundle of goods is indicative of the welfare received from the consumption of that bundle. Higher utility implies relatively higher welfare. Since utility is an ordinal concept we cannot determine an absolute level of welfare, only a relative change in welfare. The neoclassical framework assumes non-satiation so consumption of more goods will increase an individual's achieved level of utility. It also assumes that resources are scarce so consumption, or achieved level of utility, is limited to some threshold level. An individual will pursue the highest level of utility possible subject to her resource constraint.

APPENDIX B

THE ANALYSIS SAMPLE

The National Survey of Child and Adolescent Well-being (NSCAW) includes 6,228 children who came in contact with their state child welfare system. In order to ensure that all children in the analysis sample were eligible to receive outpatient services and that they were administered the same Child Behavior Checklist (CBCL) questionnaire, I restrict my sample to the 2,991 children who were younger than age 6 at baseline or older than 18 years old at wave 3. In total, 568 children were excluded from the analysis sample because of missing data: 59 children were missing age data; 33 children were missing CBCL scores at baseline; 469 children were missing CBCL scores at wave 3; and 7 children were missing CBCL scores at both baseline and wave 3. The analysis sample includes 2,482 children.

The 527 children who were excluded from the analysis sample but report a CBCL score at baseline had a mean CBCL score of 57.220 (SD=12.586; and range from 23 to 86) (see table 13). The 57 children who reported a CBCL score at wave 3 had a mean of 56.965 (SD=13.278; and range from 33 to 90). Both groups of children had lower mean CBCL scores than the analysis sample (the analysis sample had a mean CBCL score of 60 at baseline and 58 at wave 3).

The mean CBCL score for children who were clinical at baseline was not different for children in the analysis sample (at 72) and children with some missing data (at 71).

Table 13. Sample Characteristics for Children Age 6 to 18 Years Old at Initial Contact Who Are Excluded from the Analysis Sample Due to Missing Data

	Total		Clinical at baseline			
	Mean	SD	Mean	SD	Min	Max
Age[a]	10.442	2.936	11.224	2.850	6	15
Race						
White	.428	.497	.414	.494	0	1
Black	.264	.470	.266	.443	0	1
Other	.298	.421	.293	.456	0	1
Female	.500	.500	.500	.501	0	1
Type of abuse						
Physical	.313	.464	.374	.485	0	1
Sexual	.181	.386	.203	.403	0	1
Emotional	.150	.357	.189	.393	0	1
Child is adopted	.026	.160	.027	.163	0	1
Type of OOH placement						
Foster care	.127	.367	.126	.333	0	1
Kinship care	.072	.326	.081	.274	0	1
Group home	.021	.157	.041	.198	0	1
Residential care	.023	.161	.041	.198	0	1
Other	.016	.094	.032	.175	0	1
Number of OOH placements	1.571	1.137	1.855	1.476	1	11
Child's health insurance coverage						
Medicaid	.576	.495	.550	.499	0	1
Private insurer	.246	.431	.189	.393	0	1
None	.178	.383	.261	.440	0	1
Number of local community health centers	2.577	3.760	2.532	3.750	0	37

Table 13 continued.

	Total		Clinical at baseline			
	Mean	SD	Mean	SD	Min	Max
Who contacted CPS?						
Teacher or school staff	.410	.492	.441	.498	0	1
Doctor	.174	.380	.198	.400	0	1

Note: There were 568 children excluded from the final sample due to missing data. Of these, 222 children were clinical at baseline. Standard deviation (SD), minimum (Min), and maximum (Max) are reported. Out of home (OOH) care. Child protective services (CPS).

[a] A child's age was reported for 509 children in the total sample and 201 children in the sample of children who were clinical at baseline.

*p < .05 **p < .01 ***p < .001

However, children who were clinical at baseline and excluded from the analysis sample because of missing data had an average CBCL score at wave 3 (at 60) that was lower than that of the analysis sample (at 64).

Fewer children who were excluded from the analysis sample because of missing data received outpatient services over the sampling period. Approximately 25 percent of these children received outpatient services compared to 43 percent of the analysis sample. Among children who were clinical at baseline, 38 percent received outpatient services compared to 61 percent of the analysis sample.

The average age of children excluded from the analysis sample due to missing data was not different from that of the analysis sample at 10 years old. However, the average age of excluded children who were clinical at baseline was slightly older at 11 years old.

Fewer of the excluded children were White or Black and more were some other race, regardless of the child's clinical status at baseline. This contrasts the analysis sample which included mostly White and Black children.

Fewer of the children with missing data were female at 50 percent compared to 53 percent in the analysis sample. The percent of female children who were clinical at baseline was not different (49 percent of the children with missing data and 48 percent of the analysis sample).

More than 30 percent of children with missing data had been physically abused, similar to the analysis sample. Children with missing data and were clinical at baseline were more likely to have experienced physical abuse at 37 percent compared to 30 percent of the analysis sample.

Fewer than 3 percent of all children with missing data had been adopted at baseline, not different from the percent of children in the analysis sample also were adopted at baseline.

Children with missing data had an average of two out-of-home (OOH) placements by the time of baseline interviews, not different from the analysis sample. A smaller percent of children with missing data lived in OOH care at baseline with the majority living in foster care (13 percent of children with missing data) or kinship care (7 percent of children with missing data). This trend was also true for children who were clinical at baseline.

Children with missing data were more likely to be uninsured, regardless of clinical status at baseline, and lived near a larger number of local community health centers compared to children in the analysis sample. The percent of children with missing data who were without health insurance was relatively high at 18 percent for all and 26 percent of children who were clinical at baseline. Only 10 percent of children in the analysis sample were uninsured, and 8 percent of children in the analysis sample were clinical and uninsured.

More than 40 percent of the children in the analysis sample and children with missing data were referred to CPS through a teacher or other school staff member.

APPENDIX C

SELECTING A MATCHING ALGORITHM

FOR THE PROPENSITY SCORE MODEL

Various matching algorithms are available for propensity score matching (PSM) in order to pair children who received outpatient services (i.e. were treated) with children who did not receive outpatient services (i.e. were not treated). Nearest neighbor matching, caliper matching, local linear, and kernel density matching are some of the matching techniques used in the literature. Choosing a matching algorithm from among these techniques involves some tradeoff between bias and variance.

Nearest neighbor matching is the simplest matching algorithm. A child who did not receive outpatient services is matched with a child who did receive outpatient services and who has the closest propensity score. These children will be most similar in observable characteristics and differ only in their receipt of services. However, just because two children are matched as being most similar, does not guarantee that these children are actually similar. For instance, if most children who received outpatient services have high propensity scores but only a few children who did not receive services have high propensity scores, then some of matches will be bad. Some of the children with high propensity scores in the treated group will be matched to children in the untreated group with low propensity scores. Nearest neighbor matching minimizes bias since matched children will be similar on observable characteristics. The tradeoff of this approach is that it does not use information available in other neighbors. Thus, nearest neighbor matching often results in higher variances. To minimize the variance, it is possible to match to a discrete number of neighbors.

Caliper matching uses information available in multiple neighbors to reduce the likelihood of poor matches. Caliper matching requires specification of a maximum propensity score distance by which a match can be made. Any child who did not receive outpatient services with a propensity score within the caliper distance of a propensity score for a child who did receive outpatient services, will be included in the comparison group. Caliper matching is limited in that it is difficult to know *a priori* what caliper distance is reasonable.

Local linear matching and kernel matching are nonparametric matching algorithms. The CBCL score at wave 3 for each child who received outpatient services is compared to a weighted average of the CBCL scores of all the children who did not receive outpatient services. This technique more heavily weights children who did not receive outpatient services and with propensity scores closest to a child who did receive services. The local linear and kernel estimators use more information than other matching methods to match observations. Thus, these nonparametric estimators have smaller variances but can rely on poor matches if propensity scores differ greatly.

The most common variants of these estimating techniques is ‘with replacement’ and ‘without replacement.’ Sampling without replacement restricts one comparison case to serve as the match for only one treated case. Matching without replacement may perform poorly when propensity scores do not overlap or when the control group is small

(Dehejia and Wahba 2002). In contrast, sampling with replacement allows for one comparison case to serve as the match for more than one treated case.

The analysis in this dissertation utilizes caliper matching with a distance of 0.005, without replacement. This matching algorithm was selected after comparing the estimated average treatment effect on the treated (ATT) across matching estimators. Caliper .005 without replacement estimated an ATT near the middle of this range (see table 14). While this paper does not utilize the ATT to capture the effect of outpatient

Table 14. The Average Treatment Effect on the Treated (ATT) Across Matching Methods for the PSM Model

	Without replacement		With replacement	
Matching method	ATT	SE	ATT	SE
None	.095	(.011)		
Caliper (.005)	.093	(.013)	.102	(.017)
Caliper (.001)	.086	(.015)	.089	(.017)
Caliper (.01)	.096	(.013)	.102	(.018)
Caliper (.043261)	.094	(.012)	.102	(.026)
Nearest neighbor	.080	(.013)		
Nearest neighbor (3)			.087	(.026)
Local linear matching	.074	(.024)		
Kernel matching (normal)	.084	(.020)		

Note: Standard error (SE). Standard errors are bootstrapped with 1000 replications.

The outcome variable of interest is a child’s CBCL score (ln) at wave 3. Reported ATT effects are estimated from the PSM model with state effects.

*$p < .05$ **$p < .01$ ***$p < .001$

services on the CBCL scores of children who received these services (the effect of services is measured in the stochastic frontier), the estimates range from 7 to 10 percent, and none are statistically different.[34] Also, this matching algorithm was able to pass the conditional independence and common support assumptions for the PSM technique. For instance, the propensity model did not pass these tests when local linear or kernel estimators were utilized.

34. Note, the estimated ATT for caliper .005 presented in table 13 differs from the estimated coefficient on the binary indicator for whether a child received outpatient services in the stochastic frontier model (table 10) for two reasons. First, some observations of the PSM subsample are dropped from the analysis sample used to estimate the stochastic frontier because of the implementation of probability weights. Sample probability weights were used in accordance with the data manual in order to adjust for oversampling of certain populations (National Data Archive on Child Abuse and Neglect 2008). Second, the functional form differs for the PSM model and the stochastic frontier. The stochastic frontier does not control for who contacted CPS since this has no relevance on the mental health outcome of a child but may influence the child's ability to receive outpatient services. Finally, the stochastic frontier model includes an inefficiency measure that is not captured in the PSM model.

APPENDIX D

A COMPARISON OF ESTIMATES ACROSS STOCHASTIC FRONTIER AND ORDINARY LEAST SQUARES MODELS

The stochastic frontier model is an extension of the more widely used ordinary least squares model (OLS). The OLS model requires fewer assumptions since the residual is not decomposed into a random error and an inefficiency component as in the stochastic frontier model. Instead, the residual only represents random error in the estimation procedure. (The theoretical justification for assuming there is some inefficiency in the transformation of resources into mental health outcomes is discussed in chapter 2 of this dissertation.)

Similar to the application in chapter 3, an OLS model could be used to estimate the effect of outpatient services on the mental health of children. Using the OLS model, it is assumed that mental health outcomes are efficient for all children, regardless of their receipt of mental health services.

I examined whether estimating OLS model on the PSM subsample resulted in different results than estimating a stochastic frontier model with similar observable covariates on the PSM subsample. The estimated coefficients for the model were relatively stable across models, both in significance and magnitude (see table 15).

Table 15. Covariate Estimates for the Ordinary Least Squares Model: Children Ages 6 to 18 and Clinical at Initial Contact.

	PSM sample	
Received outpatient services	.094	***
	(.020)	
CBCL score at baseline (ln)	.965	***
	(.158)	
Age (ln)	1.118	*
	(.654)	
Age-squared (ln)	-.247	*
	(.148)	
Race		
Black	.013	
	(.028)	
Other	-.023	
	(.031)	
Female	.026	
	(.024)	
Type of abuse		
Physical	-.045	*
	(.023)	
Sexual	.005	
	(.025)	
Emotional	-.034	
	(.032)	
Child is adopted	.026	
	(.063)	

Table 15 continued.

	PSM sample	
Type of OOH placement		
Foster care	-.111	
	(.094)	
Kinship care	-.108	*
	(.063)	
Group home	.098	
	(.142)	
Residential care	-.142	
	(.149)	
Other	-.150	***
	(.056)	
Number of OOH placements (ln)	-.022	
	(.028)	
Health insurance		
Medicaid	-.031	
	(.032)	
Private insurer	-.097	**
	(.039)	
Number of local community health centers (ln)	-.022	
	(.020)	
State		
California	.026	
	(.042)	
Florida	.031	
	(.040)	

Table 15 continued.

	PSM sample	
Illinois	.014	
	(.082)	
Michigan	.065	*
	(.034)	
New York	.003	
	(.051)	
Ohio	.182	***
	(.050)	
Pennsylvania	-.012	
	(.050)	
Texas	.045	
	(.046)	
Constant	-1.221	
F(28, 517)	6.08	***
R-squared	.344	
N	546	

Note: Standard errors in parenthesis.

Predicting a child's CBCL score (ln) at wave 3.

**p* < .05 ***p* < .01 ****p* < .001

A child's receipt of outpatient services and baseline CBCL score were significant predictors of the child's CBCL score at wave 3. Children who received outpatient services have an estimated 9 percent higher CBCL score at wave 3 than children who did not receive services over the sampling period.

A child's age was also a significant predictor with older children having higher CBCL scores at wave 3. However, there were diminishing returns to age. Physical abuse also predicts a child's future CBCL score. Children who were physically abused at baseline have an estimated 5 percent lower CBCL score at wave 3 compared to children who were not physically abused. Children who lived in kinship care or some other OOH placement had lower CBCL scores at wave 3 at over 10 percent lower. Improved mental health was also estimated for children who had private health insurance at baseline. Children with private health insurance have an estimated 10 percent lower CBCL score at wave 3. State effects were significant for Michigan and Ohio. Children in these states had higher CBCL scores at wave 3.

REFERENCES

Achenbach, Thomas. *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. Burlington: Department of Psychiatry University of Vermont, 1991.

Aigner, Dennis, C. A. Knox Lovell, and Peter Schmidt. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6, no. 1 (1977): 21-37.

Alkire, Sabina. "Dimensions of Human Development." *World Development* 30, no. 2 (2002): 181-205.

Allardt, Erik. "Having, Loving, Being: An Alternative to the Swedish Model of Welfare Research." In *The Quality of Life*, edited by Martha Nussbaum and Amartya Sen, 88-95. Oxford: Clarendon Press, 1993.

Anand, Paul, and Paul Dolan. "Equity, Capabilities and Health." *Social Science & Medicine* 60, no. 2 (2005): 219-222.

Anand, Paul, and Martin van Hees. "Capabilities and Achievements: An Empirical Study." *Journal of Socio-Economics 35*, no. 2 (2006): 268-284.

Ascher, Barbara, Elizabeth Farmer, Barbara Burns, and Adrian Angold. "The Child and Adolescent Services Assessment (CASA): Description and Psychometrics." *Journal of Emotional and Behavioral Disorders* 4, no. 1 (1996): 12-20.

Balestrino, Alessandro. "A Note on Functionings Poverty in Affluent Societies." *Notizei di Politeai* 12, no. 43/44 (1996): 97-105.

Balestrino, Alessandro, and Nicola Sciclone. "Should We Use Functioning Instead of Income to Measure Wellbeing?" *Rivista Internazionale di Scienze Sociali* 53, no. 1 (2000): 3-22.

Battese, George, and Timothy Coelli. "Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data." *Journal of Econometrics* 38, no. 3 (1988): 387-399.

Berger, Lawrence, Sarah Bruch, Elizabeth Johnson, Sigrid James, and David Rubin. "Estimating the "Impact" of Out-of-Home Placement on Child Well-Being: Approaching the Problem of Selection Bias." *Child Development* 80, no. 6 (2009): 1856-1876.

Berzin, Stephanie. "Difficulties in the Transition to Adulthood: Using Propensity Scoring to Understand What Makes Foster Youth Vulnerable." *Social Service Review* 82, no. 2 (2008): 171–195.

Bess, Roseana, Cynthia Scarcella, Amy Jantz, Victoria Russell, and Rob Geen. "The Cost of Protecting Vulnerable Children III: What Factors Affect States' Fiscal Decisions." *Occasional Paper No. 61*, 2002, Urban Institute http://www.urban.org/uploadedpdf/310596_OP61.pdf (accessed April 1, 2010).

Binder, Martin, and Tom Broekel. "The Neglected Dimension of Well-being: Analyzing the Development of 'Conversion Efficiency' in Great Britain." *Working Paper*, Jena Economic Research Papers (2010): 2010-067.

-----. "Applying a Robust Non-Parametric Efficiency Analysis to Measure Conversion Efficiency in Great Britain." *Working Paper*, SSRN (2008): 1104430. Accepted for publication in *The Journal of Human Development and Capabilities*.

Blundell, Richard, and Arthur Lewbel. "The Information Content of Equivalence Scales." *Journal of Econometrics* 50, no. 1 (1991): 49-68.

Brook, Robert, and Elizabeth McGlynn. "Measuring Quality of Care." *New England Journal of Medicine* 335, no. 13 (1996): 966-970.

Burchardt, Tania. "Capabilities and Disability: The Capabilities Framework and the Social Model of Disability." *Disability and Society* 19, no. 7 (2004): 736-751.

Burns, Barbara, Susan Phillips, H. Ryan Wagner, Richard Barth, David Kolko, Yvonne Campbell, and John Landsverk. "Mental Health Need and Access to Mental Health Services by Youths Involved with Child Welfare: A National Survey." *Journal of the American Academy of Child and Adolescent Psychiatry* 43, no. 8 (2004): 960-970.

Burt, Martha, Laudan Aron, Toby Douglas, Jesse Valente, Edgar Lee, and Britta Iwen. "Homelessness: Programs and the People They Serve: Findings of the National Survey of Homeless Assistance Providers and Clients." *Technical Report*, September 1999, Urban Institute http://www.huduser.org/portal/publications/pdf/home_tech/frontmat.pdf (accessed January 4, 2009).

Chiappero Martinetti, Enrica, and Paola Salardi. "Well-Being Process and Conversion Factors: An Estimation of the Micro-Side of the Well-Being Process." *Working Paper 3/2008*, 2008, Human Development, Capability and Poverty International Research Centre http://www.iusspavia.it/UserFiles/File/HDC-IRC/Chiappero,Salardi%20-%20WP3.pdf (accessed October 3, 2010).

Child Welfare Information Gateway. "Child Abuse and Neglect." *Fact Sheet*, April 2008, U.S. Department of Health and Human Services http://www.childwelfare.gov/pubs/can_info_packet.pdf (accessed January 22, 2011).

Clausen, June, John Landsverk, William Ganger, David Chadwick, and Alan Litrownik. "Mental Health Problems of Children in Foster Care." *Journal of Child and Family Studies* 7, no. 3 (1998): 283-96.

Coast, Joanna, Richard Smith, and Paula Lorgelly. "Should the Capability Approach be Applied in Health Economics?" *Health Economics* 17, no. 6 (2008): 667-670.

Costello, E. Jane, Adrian Angold, Barbara Burns, Danlene Stangl, Dan Tweed, Alaattin Erkanli, and Carol Worthman. "The Great Smoky Mountains Study of Youth: Goals, Design, Methods, and the Prevalence of DSM–III-R Disorders." *Archives of General Psychiatry* 53, no. 12 (1996): 1129–1136.

Courtney, Mark. "Research Needed to Improve the Prospects for Children in Out-of-Home Placement." *Children and Youth Services Review* 22, no. 9/10 (2000): 743-761.

Cummins, Robert. "The Domains of Life Satisfaction: An Attempt to Order Chaos." *Social Indicators Research* 38, no. 3 (1996): 303-332

Dehejia, Rajeev, and Sadek Wahba. "Propensity Score Matching Methods for Nonexperimental Causal Studies." *The Review of Economics and Statistics* 84, no. 1 (2002): 151-161.

Deutsch, Joseph, Jacques Silber, and Nira Yacouel. "On the Measurement of Inequality in the Quality of Life in Isreal." Paper presented at the conference on Justice and Poverty: Examining Sen's Capability Approach, in Cambridge, United Kingdom, June 5-7, 2001.

DeVooght, Kerry, Tiffany Allen, and Rob Geen. "Federal, State, and Local Spending to Address Child Abuse and Neglect in SFY 2006." *Report*, December 2008, Child Trends http://www.childtrends.org/Files//Child_Trends-2009_02_17_FR_CWFinancePaper.pdf (accessed April 1, 2010).

Dowding, Keith. "Revealed Preference and External Reference." *Rationality and Society* 14, no. 3 (2002): 259-284.

Doyle, Joseph. “Child Protection and Child Outcomes: Measuring the Effects of Foster Care.” *American Economic Review*, 97, no. 5 (2007): 1583-1610.

Dworsky, Amy, and Mark Courtney. “Self-Sufficiency of Former Foster Youth in Wisconsin: Analysis of Unemployment Insurance Wage Data and Public Assistance Data.” *Special Report SR81*, August 2001, Institute for Research on Poverty http://aspe.hhs.gov/hsp/fosteryouthWI00 (accessed September 8, 2008).

Farmer, Elizabeth, Sarah Mustillo, H. Ryan Wagner, Barbara Burns, David Kolko, Richard Barth, and Laurel Leslie. “Service Use and Multi-Sector Use for Mental Health Problems by Youth in Contact with Child Welfare.” *Children and Youth Services Review* 32, no. 6 (2010): 815-821.

Fisher, Irving. *The Theory of Interest.* New York: Macmillan, 1930.

Foster, E. Michael. “Propensity Score Matching: An Illustrative Analysis of Dose Response.” *Medical Care* 41, no. 10 (2003): 1183-192.

Frankfurt, Harry. “Freedom of the Will and the Concept of a Person.” *The Journal of Philosophy* 68, no. 1 (1971): 11-25.

Garland, Ann, and Bridgett Besinger. “Racial/Ethnic Differences in Court Referred Pathways to Mental Health Services for Children in Foster Care.” *Children and Youth Services Review* 19, no. 18 (1997): 651–666.

Garland, Ann, John Landsverk, and Anna Lau. “Racial/Ethnic Disparities in Mental Health Service Use among Children in Foster Care.” *Children and Youth Services Review* 25, no. 5-6 (2003): 491-507.

Garland, Ann, Anna Lau, Kristen McCabe, Richard Hough, and John Landsverk. “Racial and Ethnic Differences in Utilization of Mental Health Services Among High-Risk Youths.” *American Journal of Psychiatry* 162 (2005): 1336–1343.

Greene, William. “A Stochastic Frontier Model with Correction for Sample Selection.” *Journal of Productivity Analysis* 34, no. 1 (2010): 15-24.

-----. “Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model.” *Journal of Econometircs* 126, no. 2 (2005): 269-303.

Guyatt, Gordon, David Feeny, and Donald Patrick. “Measuring Health-Related Quality of Life.” *Annals of Internal Medicine* 118, no. 8 (1993): 622-629.

Hasan, Hamid. “Capabilities in Measurement: An Empirical Investigation.” *Working Paper*, Munich Personal RePEc Archive (2009): 1104430.

Head, Brian. “Three Lessons of Evidence Based Policy.” *Australian Journal of Public Administration* 67, no. 1 (2008): 1-11.

Heckman, James. “Sample Selection Bias as a Specification Error.” *Econometrica* 47, no. 1 (1979): 153-161.

-----. “Discrete, Qualitative, and Limited Dependent Variables.” *Annals of Economic and Social Measurement* 4, no. 5 (1976): 475-492.

Hicks, John. “Leon Walras.” *Econometrica* 2, no. 4 (1934): 338-48.

-----. “Marginal Productivity and the Principle of Variation.” *Econometrica, 12*, no. 35 (1932): 79-88.

Hobson, John. “Neo-Classical Economics in Britain.” *Political Science Quarterly 40,* no. 3 (1925): 337-83.

Hopper, Kim. “Rethinking Social Recovery in Schizophrenia: What a Capabilities Approach Might Offer.” *Social Science & Medicine* 65, no. 5 (2007): 868-879.

Hurlburt, Michael, Laurel Leslie, John Landsverk, Richard Barth, Barabara Burns, Robert Gibbons, Donald Slymen, and Jinjin Zhang. “Contextual Predictors of Mental Health Service Use among Children Open to Child Welfare.” *Archives of General Psychiatry* 61, no. 12 (2004): 1217-1224.

Imbens, Guido. “Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review.” *The Review of Economics and Statistics* 86, no. 1 (2004): 4−29.

Jamison, Dean, Andrew Creese, and Thomson Prentice. “The World Health Report: Making a Difference.” *Special Report*, The World Health Organization, 1999 http://www.who.int/whr/1999/en/whr99_en.pdf (accessed December 12, 2010).

Johnson, F. Reed, and Kristy Mathews. “Sources and Effects of Utility-Theoretic Inconsistency in Stated-Preference Surveys.” *American Journal of Agricultural Economics 83*, no. 5 (2001): 1328-1333.

Jondrow, James, Ivan Materov, C. A. Knox Lovell, and Peter Schmidt. “On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model.” *Journal of Econometrics* 19, no. 2/3 (1982): 233-238.

Kakwani, Nanak. “Performance in Living Standards: An International Comparison.” *Journal of Development Economics* 41, no. 2 (1993): 307-36.

Kalai, Gil, Ariel Rubinstein, and Ran Spiegler. “Rationalizing Choice Functions by Multiple Rationales.” *Econometrica* 70, no. 6 (2002): 2481-2488.

Kataoka, Sheryl, Lily Zhang, and Kenneth Wells. “Unmet Need for Mental Health Care Among U.S. Children: Variation by Ethnicity and Insurance Status.” *American Journal of Psychiatry* 159, no. 9 (2002): 1548–1555.

Kelly, Anthony. “School Choice Policy in England: An Adaption of Sen’s Early Work on Capability.” *Journal of School Choice* 4, no. 3 (2010): 317-335.

Klasen, Stephen. “Measuring Poverty and Deprivation in South Africa.” *Review of Income and Wealth* 46, no. 1 (2000): 33-58.

Kolko, David, Amy Herschell, Amanda Costello, and Rachel Kolko. “Child Welfare Recommendations to Improve Mental Health Services for Children who have Experienced Abuse and Neglect: A National Perspective.” *Administration and Policy in Mental Health and Mental Health Services Research* 36, no. 1 (2009): 50-62.

Konrad, Ellen. “A Multidimensional Framework for Conceptualizing Human Services Integration Initiatives.” In *Evaluating Initiatives to Integrate Human Services*, edited by Jules Marquart and Ellen Konrad, 5-19. San Francisco: Jossey-Bass Publishers, 1996.

Kuklys, Wiebke. *Amartya Sen’s Capability Approach – Theoretical Insights and Empirical Applications*. Berlin: Springer, 2005.

Kumbhakar, Subal, and C. A. Knox Lovell. *Stochastic Frontier Analysis.* Cambridge: University Press, 2000.

Kusserow, Richard. “Services Integration: A Twenty-Year Retrospective.” *OEI-01-91-00580,* January 1991, U.S. Department of Health and Human Services http://oig.hhs.gov/oei/reports/oei-01-91-00580.pdf (accessed April 1, 2010).

Lancaster, Kelvin. “A New Approach to Consumer Theory.” *Journal of Political Economy* 74, no. 2 (1966): 132-157.

Landsverk, John, Ann Garland, and Laurel Leslie. “Mental Health Services for Children Reported to Child Protective Services.” In *APSAC Handbook on Child Maltreatment*, edited by John Myers, C. T. Hendrix, Lucy Berliner, Carole Jenny, John Briere, and Theresa Reid, 487-507. Thousand Oaks, California: Sage, 2002.

Lelli, Sara. “Using Functionings to Estimate Equivalence Scales.” *Review of Income and Wealth 51*, no. 2 (2005): 255-284.

-----. “Factor Analysis vs. Fuzzy Sets Theory: Assessing the Influence of Different Techniques on Sen’s Functioning Approach.” *Working Paper,* Katholieke Universiteit Leuven (2001): 01.21.

Leslie, Laurel, Michael Hurlburt, Sigrid James, John Landsverk, David Slymen, and Jinjin Zhang. “Relationship Between Entry Into Child Welfare and Mental Health Service Use.” *Psychiatric Services* 56, no. 8 (2005): 981-987.

Leslie, Laurel, Michael Hurlburt, John Landsverk, Richard Barth, and David Slymen. "Outpatient Mental Health Services for Children in Foster Care: A National Sample.' *Child Abuse and Neglect* 28, no. 6 (2004): 699-714.

Leslie, Laurel, John Landsverk, Roxanne Ezzet-Lofstrom, Jeanne Tschann, David Slymen, and Ann Garland. "Children in Foster Care: Factors Influencing Outpatient Mental Health Service Use." *Child Abuse and Neglect* 24, no. 4 (2000): 465–476.

Lewis, Jane, and Susy Giullari. "The Adult Worker Model Family, Gender Equality and Care: The Search for New Policy Principles and the Possibilities and Problems of a Capabilities Approach." *Economy and Society* 34, no. 1 (2005): 76-104.

Lorgelly, Paula, Kenny Lawson, Elisabeth Fenwick and Andrew Briggs. "Outcome Measurement in Economic Evaluations of Public Health Interventions: A Role for the Capability Approach." *International Journal of Environmental Research and Public Health 7*, no. 5 (2010): 2274-2289.

Mayen, Carlos, Joseph Balagtas, and Corinne Alexander. "Technology Adoption and Technical Efficiency: Organic and Conventional Dairy Farms in the United States." *American Journal of Agricultural Economics* 92, no. 1 (2010): 181-195.

McCarthy, Jan, Erica Van Buren, and Marisa Irvine. "An Analysis of Mental Health Issues in States' Child and Family Service Reviews." National Technical Assistance Center for Children's Mental Health. Georgetown University Center for Child and Human Development. August 2007.

McDonald, Thomas, Alex Westerfelt, and Irving Piliavin. *Assessing the Long-term Effects of Foster Care: A Research Synthesis.* Washington, DC: Child Welfare League of America Press, 1996.

Menzel, Paul, Paul Dolan, Jeff Richardson, and Jan Abel Olsen. "The Role of Adaption to Disability and Disease in Health State Valuation: A Preliminary Normative Analysis." *Social Science Medicine 55*, no. 12 (2002): 2149-2158.

Miguel, Fernando San, Mandy Ryan, and Mabelle Amaya-Amaya. "'Irrational' Stated Preferences: A Quantitative and Qualitative Investigation." *Health Economics 14*, no. 3 (2005): 1099-1050.

Mitra, Sophia. "The Capability Approach and Disability." *Journal of Disability and Policy Studies 16*, no. 4 (2006): 236-247.

Moore, Mick, Jennifer Leavy, Peter Houtzager, and Howard White. "Polity Qualities: How Governance Affects Poverty." *Working Paper*, University of Sussex Institute of Development Studies (2000): 99.

Muellbauer, John. “Testing the Barten Model of Household Consumption Effects and the Cost of Children.” *Economic Journal* 87, no. 347 (1977): 460-487.

Narayan, Deepa, Robert Chambers, Meera Shah, and Patti Petesch. *Crying Out for Change: Voices of the Poor*. New York: Oxford University Press for the World Bank, 2000.

National Data Archive on Child Abuse and Neglect. (2008). National Survey of Child and Adolescent Well-Being (NSCAW): Combined Waves 1-5, Data File User’s Manual Restricted Release Version. Administration on Children and Families. Ithaca: Cornell University.

Newton, Rae, Alan Litrownik, and John Landsverk. “Children and Youth in Foster Care: Disentangling the Relationship Between Problem Behaviors and Number of Placements.” *Child Abuse and Neglect* 24, no. 10 (2000): 1363–1374.

Nussbaum, Martha. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge: Belknap Press, 2006.

-----. “Capabilities as Fundamental Entitlements: Sen and Social Justice.” *Feminist Economics* 9, no. 2-3 (2003): 33-59.

-----. *Women and Human Development: The Capabilities Approach.* Cambridge: Cambridge University Press, 2000.

-----. *Oxford Studies in Ancient Philosophy*. “Nature, Function and Capability: Aristotle on Political Distribution,” Vol. 145.84. Oxford: Oxford University Press, 1988.

Olson, Kevin. “Recognizing Gender, Redistributing Labor.” *Social Politics* 9, no. 3 (2002): 380-410.

Pattanaik, Prasanta, and Yongsheng Xu. “On Ranking Opportunity Sets in Terms of Freedom of Choice.” *Recherches Economiques de Louvain* 56, no. 4 (1990): 383-390.

Phipps, Shelley. “The Well-Being of Young Canadian Children in International Perspective: A Functionings Approach.” *Review of Income and Wealth* 48, no. 4 (2002): 493-515.

Pollak, Robert, and Terence Wales. *Demand System Specification and Estimation.* Oxford: Oxford University Press, 1992.

-----. “Welfare Comparisons and Equivalence Scales.” *American Economic Review* 69, no. 2 (1979): 216-221.

Pratap, Sangeeta, and Erwan Quintin. “Are Labor Markets Segmented in Developing Countries? A Semiparametric Approach.” *European Economic Review* 50, no. 7 (2006): 1817-1841.

Qizilbash, Mozaffar. “A Note on the Measurement of Poverty and Vulnerability in the South African Context.” *Journal of International Development* 14, no. 4 (2002): 757-772.

-----. “The Concept of Well-Being.” *Economics and Philosophy* 14, no. 1 (1998): 51-73.

Raghavan, Ramesh, Megumi Inoue, Susan Ettner, Barton Hamilton, and John Landsverk. “A Preliminary Analysis of the Receipt of Mental Health Services Consistent with National Standards Among Children in the Child Welfare System.” *American Journal of Public Health* 100, no. 4 (2010): 742-749.

Ramos, Xavier. “Using Efficiency Analysis to Measure Individual Well-Being with an Illustration for Catalonia.” In *Quantiative Approaches to Multidimensional Poverty Measurement*, edited by Nanak Kakwani and Jacques Silber, 155-175. Basingstoke: Palgrave Macmillan, 2008.

Ramos, Xavier, and Jacques Silber. “On the Application of Efficiency Analysis to the Study of the Dimensions of Human Development.” *Review of Income and Wealth* 51, no. 2 (2005): 285-309.

Ravallion, Martin. “On Measuring Aggregate ‘Social Efficiency’.” *Economic Development and Cultural Change* 53, no. 2 (2005): 273-292.

Reddy, Sanjay, Sujata Visaria, and Muhammad Asali. “Inter-Country Comparisons of Poverty Based on a Capability Approach: An Empirical Exercise.” *Working Paper*, Q-Squared (2006): 28.

Ringeisen, Heather, Cecilia Casanueva, Matthew Urato, and Leyla Stambaugh. “Mental Health Service Use During the Transition to Adulthood for Adolescents Reported in the Child Welfare System.” *Psychiatric Services* 60, no. 8 (2009): 1084-1091.

Robeyns, Ingrid. “The Capability Approach in Practice.” *The Journal of Political Philosophy* 14, no. 3 (2006): 351-376.

Roemer, John. *Theories of Distributive Justice*. Cambridge: Harvard University Press, 1996.

Roll, Eric. *A History of Economic Thought*. New York, NY: Prentice-Hall, 1938.

Rossi, Peter, and Walter Williams. *Evaluating Social Programs: Theory, Practice and Politics*. New York: Seminar Press, 1972.

Rubin, David, Kevin Downes, Amanda O'Reilly, Robin Mekonnen, Xianqun Luan, and A. Russell Localio. "Impact of Kinship Care on Behavioral Well-Being for Children in Out-of-Home Care." *Archives for Pediatric & Adolescent Medicine* 162, no. 6 (2008): 550-556.

Rubin, David, Amanda O'Reilly, Xianqun Luan, and A. Russell Localio. "The Impact of Placement Stability on Behavioral Well-Being for Children in Foster Care." *Pediatrics* 119, no. 2 (2007): 336-344.

Ruger, Jennifer. "Health and Social Justice." *The Lancet* 364, no. 9439 (2004a): 1075-1080.

-----. "Millennium Development Goals for Health: Building Human Capabilities." *Bulletin of the World Health Organization* 82, no. 12 (2004b): 951-952.

Ruggeri Laderchi, Caterina. "The Many Dimensions of Deprivation in Peru: Theoretical Debates and Empirical Evidence." *Working Paper*, QEH Working Paper Series (1999): 29.

-----. "Poverty and Its Many Dimensions: The Role of Income as an Indicator." *Oxford Development Studies* 25, no. 3 (1997): 345-360.

Ruggeri Laderchi, Caterina, Ruhi Saith, and Frances Stewart. "Does it Matter that We Do Not Agree on the Definition of Poverty? A Comparison of Four Approaches." *Oxford Development Studies* 31, no. 3 (2003): 243-74.

Ryan, Ryan, and Edward Deci. "Autonomy is No Illusion: Self-Determination Theory and Empirical Study of Authenticity, Awareness, and Will." In *Handbook of Experimental Existential Psychology*, edited by J. Greenberg, S. L. Koole and T. Pysczcynski, 449-479. New York: Guilford Press, 2004.

Salzer, Mark, Leonard Bickman, and E. Warren Lambert. "Dose–Effect Relationship in Children's Psychotherapy Services." *Journal of Consulting and Clinical Psychology* 67, no. 2 (1999): 228-238.

Samuelson, Paul A. *Foundations of Economic Analysis*. Princeton: Harvard University Press, 1947.

Schokkaert, Erik, and Luc Van Ootegem. "Sen's Concept of the Living Standard Applied to the Belgian Unemployed." *Recherches Economiques de Louvain* 56, no. 4 (1990): 429-450.

Sen, Amartya. "Capabilities, Lists, and Public Reason: Continuing the Conversation." *Feminist Economics* 10, no. 3 (2004): 77-80.

-----. "Health: Perception versus Observation." *British Medical Journal* 324, no. 7342 (2002): 860-861.

-----. *Development as Freedom*. Oxford: Oxford University Press, 1999a.

-----. “The Possibility of Social Choice.” *American Economic Review* 89, no. 3 (1999b): 349-78.

-----. “From Income Inequality to Economic Inequality.” *Southern Economic Journal* 64, no. 2 (1997a): 383-401.

-----. “Maximization and the Act of Choice.” *Econometrica* 65, no. 4 (1997b): 745-79.

-----. “On the Foundations of Welfare Economics: Utility, Capability and Practical Reason.” In *Ethics, Rationality, and Economic Behaviour*, edited by Francesco Farina, Frank Hahn, and Stefano Vannucci, 50-65. Oxford: Clarendon Press, 1996.

-----. “Is the Idea of Purely Internal Consistency of Choice Bizarre?” In *World, Mind and Ethics: Essays on the Ethical Philosophy of N. Bernard Williams*, edited by James Altham and Ross Harrison, 19-31. Cambridge: Cambridge University Press, 1995.

-----. “Internal Consistency of Choice.” *Econometrica* 61, no. 3 (1993): 495-521.

-----. *Inequality Re-Examined*. Oxford: Clarendon Press, 1992.

-----. *Commodities and Capabilities*. Oxford: Elsevier Science, 1985.

-----. *On Economic Inequality*. Oxford: Clarendon Press, 1973.

-----. *Collective Choice and Social Welfare.* San Francisco, CA: Holden-Day, 1970.

Shipman, Kimberly,and Heather Taussig. “Mental Health Treatment of Child Abuse and Neglect: The Promise of Evidence-Based Practice.” *Pediatric Clinics of North America* 56, no. 2 (2009): 417-428.

Shlonsky, Aron. “Initial Construction of an Actuarial Risk Assessment Measure Using the National Survey of Child and Adolescent Well-Being.” In *Child Protection: Using Research to Improve Policy and Practice*, edited by Ron Haskins, Fred Wulczyn, and Mary Bruce Webb, 62-80. Washington, DC: Brookings Institute Press, 2007

Sippel, Reinhard. “An Experiment on the Pure Theory of Consumer's Behaviour.” *The Economic Journal* 107, no. 444 (1997): 1431-1444.

Slesnick, Daniel. “Empirical Approaches to the Measurement of Welfare.” *Journal of Economic Literature* 36, no. 4 (1998): 2108-2165.

Southerland, Dannia, Cecilia Casanueva, and Heather Ringeisen “Young Adult Outcomes and Mental Health Problems Among Transition Age Youth Investigated for

Maltreatment During Adolescence." *Children and Youth Services Review* 31, no. 9 (2009): 947-956.

Srinivasan, T. N. "Human Development: A New Paradigm or Reinvention of the Wheel." *AEA Papers and Proceedings*. 84, no. 2 (1994): 238-243.

Stahmer, Aubyn, Laurel Leslie, Michael Hurlburt, Richard Barth, Mary Bruce Webb, John Landsverk, and Jinjin Zhang. "Developmental and Behavioral Needs and Service Use for Young Children in Child Welfare." *Pediatrics* 116, no. 4 (2005): 891–900.

Stiffman, Arlene Rubin, Eric Hadley-Ives, Peter Doré, Michael Polgar, Violet Horvath, Catherine Striley, and Diane Elze. "Youths' Access to Mental Health Services: The Role of Providers' Training, Resource Connectivity, and Assessment of Need." *Mental Health Service Research* 2, no. 3 (2000): 141-154.

Stigler, George. *Production and Distribution Theories*. New York: Macmillan, 1941.

Sugden, Robert. "Welfare, Resources, and Capabilities: A Review of Inequality Reexamined by Amartya Sen." *Journal of Economic Literature* 31, no. 4 (1993): 1947-1962.

Terzi, Lorella. "Beyond the Dilemma of Difference: The Capability Approach to Disability and Special Education Needs." *Journal of Philosophy of Education* 39, no. 3 (2005): 443-59.

UNDP. *Human Development Report*. Oxford: Oxford University Press, 1990-2009.

Unterhalter, Elaine. "The Capabilities Approach and Gendered Education: An Examination of South African Complexities." *Theory and Research in Education* 1, no. 1 (2003): 7-22.

U.S. Department of Health and Human Services. "Child Maltreatment 2006." www.acf.hhs.gov/programs/ cb/pubs/cm06/index.htm (accessed April 1, 2008).

Vallentyne, Peter. "Debate: Capabilities Versus Opportunities for Well-Being." *The Journal of Political Philosophy* 13, no. 3 (2005): 359-371.

Veblen, Thorstein. 1900. "Preconceptions of Economic Science." *Quarterly Journal of Economics* 14, no. 2 (1900): 261.

Verkerk, Marian, Jan Busschbach, and Edgar Karssing. "Health-Related Quality of Life Research and the Capability Approach of Amartya Sen." *Quality of Life Research*, *10*, no. 1 (2001): 49-55.

Vinnerljung, Bo, Knut Sundell, Cecilia Andrée Löfholm, and Eva Humlesjö. "Former Stockholm Child Protection Cases as Young Adults: Do Outcomes Differ Between Those That Received Services and Those That Did Not?" *Children and Youth Services Review* 28, no. 1 (2006): 59-77.

Waldfogel, Jane. "Child Welfare Research: How Adequate Are the Data?" *Child and Youth Services Review* 22, no. 9/10 (2000): 705-741.

Walker, Melanie. "Towards a Capability-Based Theory of Social Justice for Education Policy-Making." *Journal of Education Policy* 21, no. 2 (2006): 163-85.

------. *Higher Education Pedagogies: A Capabilities Approach*. Maidenhead: Open University Press, 2005.

Wang, Jia, Dean Jamison, Eduard Bos, Alexander Preker, and John Peabody. *Measuring Country Performance on Health: Selected Indicators for 115 Countries*. Health, Nutrition and Population Series. Washington, D.C.: World Bank, 1999.

Washington State Department of Social and Health Services. "Report to the Legislature: Foster Children - Long Term Needs." Department of Social and Health Services. http://www.dshs.wa.gov/pdf/main/legrep/Leg1003/KidscreenFNL1003.pdf. (accessed January 29, 2011).

Webb, Mary Bruce, Kathryn Dowd, Brenda Jones Harden, John Landsverk, and Mark Testa. "Introduction." In *Child welfare and child well-being: New perspectives from the National Survey of Child and Adolescent Well-Being*, edited by Mary Bruce Webb, Kathryn Dowd, Brenda Jones Harden, John Landsverk, and Mark Testa, xv-xxxi. Oxford: Oxford University Press, 2010.

Williams, Bernard. "The Standard of Living: Interests and Capabilities." In *The Standard of Living*, edited by Amartya Sen, 94-102. Cambridge: Cambridge University Press, 1987.

World Bank. *World Development Report*: *Investing in Health*. New York: Oxford University Press, 1993.

Wulczyn, Fred, Richard Barth, Ying-Ying Yuan, Brenda Harden, and John Landsverk. *Beyond Common Sense: Child Welfare, Child Well-Being, and the Evidence for Policy Reform*. New Brunswick: Aldine Transaction, 2005.

Ysander, Bengt-Christer. "Robert Erikson: Descriptions of Inequality." In *The Quality of Life*, edited by Martha Nussbaum and Amartya Sen. Oxford: Clarendon Press, 1993.

Zaidi, Asghar, and Tania Burchardt. "Comparing Incomes when Needs Differ: Equalization for the Extra Cost of Disability in the UK." *Review of Income and Wealth* 51, no. 1 (2005): 89-114.