Microeconometric Essays in Education

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Economics

By

Michael D. Barker, M.A.

Washington, DC
November 20, 2015

Microeconometric Essays in Education

Michael D. Barker, M.A.

Dissertation Advisor: Francis Vella, Ph.D.

Abstract

Chapter 1 addresses a long-standing problem in econometrics, that of identifying social interactions or peer effects. The estimation strategy is based on an estimator developed by Klein and Vella (2010) that uses information from the conditional distribution of error terms. This strategy exploits the excess variance in outcomes induced by social interactions and has the potential to identify the endogenous and exogenous channels. I demonstrate the proposed estimator using both simulated data and actual data from Project STAR, the classroom-size reduction experiment.

Chapter 2 evaluates the impacts of a conditional transfer program designed to increase primary school participation in Cambodia. The absolute impact of the program is estimated, as well as the relative performance of food and cash transfers. Results indicate that both transfer modalities decrease the probability of primary school dropout. No adverse impact from the receipt of cash transfers was observed.

Chapter 3 applies a structural econometric approach to the Cambodian conditional transfer program, developing and estimating a dynamic discrete choice model. With further development, this model will enable the simulation of a wide range of counterfactual policy alternatives. This work contributes to the underutilized set of tools of structural estimation, by providing further evidence of its complementarity with experimental program evaluation.

Index words: Education, Peer Effects, Social Interactions, RCT, CCT, Cambodia, Structural Estimation, Dynamic Discrete Choice

Contents

Chapter 1

Estimating Peer Effects in Project Star

1.1 Introduction

> Perhaps the strongest evidence for social interactions is now, as it always has been, the prima facie evidence on the high degree of stratification in the U.S. by income, education, race, and other characteristics across neighborhoods and schools, and the high variance across areas and schools that this strong sorting implies.(Moffitt, 2001)

Social interactions, or peer effects, occur when an individual's outcome is affected by the outcomes or attributes of those around him. Empirical estimation of social interactions has proved difficult due to a fundamental identification problem known as the reflection problem, formalized by Charles Manski in 1993. Manski focused on the linear-in-means model, where individual outcomes, group outcomes, and group independent characteristics are all potentially linear functions of individual characteristics. Manski begins from the observation that individual and group outcomes are frequently highly correlated. He defines three distinct causal channels that could explain this excess correlation: endogenous social interaction effects, exogenous social interaction effects, and correlated effects.

Analysis of social interactions frequently focuses on endogenous social interaction effects. Endogenous effects occur when the decision or outcome of an individual is influenced by the decisions or outcomes of the individual's social group. This channel is distinct from exogenous social interactions, which occur when individuals are influenced by the shared exogenous characteristics of their social groups. Endogenous and exogenous effects, though distinct, can both be described as a type of social interaction effect. In both cases, the members of an individual's group are impacting the outcome of the individual. The same description cannot be applied to correlated effects, which refers to excess correlation in individual outcomes that is not caused by the characteristics or outcomes of other group members. Rather, individuals within a group may make similar decisions simply because they face a similar decision problem. Correlated effects could be driven by shared environmental characteristics or by unobservable similarities between group members induced by an endogenous group-formation process. When not properly accounted for, correlated effects can introduce spurious correlations in outcomes, that may be incorrectly interpreted as evidence of social interactions.

In the absence of correlated effects, the basic linear-in-means model can identify the presence of social interaction effects, but cannot separately identify exogenous and endogenous effects. Instead, a single parameter is identified that is a composite of endogenous and exogenous effects. In many cases, simple identification of the presence of social interaction effects is interesting and useful. The potential to draw policy implications, though, is greatly expanded if endogenous and exogenous effects can be separately identified. In the presence of endogenous social interaction effects, changes in individual outcomes impact the entire group. The same is not true of exogenous social interaction effects, which operate through the fixed, exogenous characteristics of the group.

The potential to separate the separate channels of social interactions in a given empirical application depends critically on the richness of the data being analyzed. Identification is simplified if the analyst has *a priori* knowledge of the full social network, including the whether or not a connection exists between each pair of individuals. Given full knowledge of the social network, and excluding certain special case network structures, separate identification of endogenous and exogenous effects is possible, even in the presence of correlated effects. Identification in this case was established by Bramoullé, Djebbari, and Fortin (2009). Although explicit data on social networks are still rare,[1] it is possible in some cases to infer social connections between agents from other characteristics. Connections between homeowners, for example, may be inferred from the physical distance between their homes. Relationships between firms may be inferred from the volume of their transactions.

Unfortunately, there is a limited range of empirical applications in which it is possible to infer social connections between individual agents. In the field of education, for example, it may be impossible to determine the precise network structure of social connections between individual students from available data. Unless information on friendships is explicitly collected, the best estimate of a student's sphere of social influence may be the student's classroom. Using classroom membership as a proxy for social connections leads to a special case social network, where individuals are partitioned into mutually exclusive groups. The group interaction network structure is one of the special case network structures where the identification result of Bramoullé et al. (2009) does not hold.

Despite the difficulty in estimating social interaction effects in this context, this paper will focus on the estimation of social interaction effects on educational outcomes using a model of group interactions. There are many reasons to focus on the group interactions model in the context of education. First, it is a setting where peer interactions are thought to play a significant role in individual outcomes. Second,

[1] A notable exception is The National Longitudinal Study of Adolescent to Adult Health (Add Health), which asks adolescents to identify their friends explicitly.

there are significant policy implications if social interaction effects can be accurately identified. The group interaction structure mirrors the structure of the classroom. This makes estimation results immediately relevant to the decision problem faced by school administrators, who must evaluate policy options in the classroom context. The identification of social interactions may inform decisions on ability tracking and school choice policies. Ability tracking alters the allocation of students within schools, while school choice policies impact the allocation of students between schools. The ultimate impact of these policies will depend on both the presence and the nature of social interaction effects. Finally, the approach can be implemented using existing administrative data. The only information required on the structure of social networks is classroom membership.

More broadly, the estimator can be deployed to estimate other models of group interactions. The group interaction network structure may be a special case within the broader set of all possible network structures, but it is highly relevant for applied work. Without *a priori* knowledge of the social network structure, group membership may be the only reasonable indicator of social influence.

I propose an estimation approach that will allow the separate identification of endogenous and exogenous peer effects in a setting of high policy relevance using readily available data. The strategy is based on an estimator developed by Klein and Vella (2010, hereafter KV). In theory, only moderate restrictions on the covariance structure are required to achieve identification, though in practice additional restrictions are made to facilitate estimation. This estimation strategy will enable the estimation of social interactions in a wide field of non-experimental settings that have previously been inaccessible to rigorous empirical evaluation.

1.1.1 Literature Review

Peer Effects in Education

Previous work in the field of education has focused on identifying the presence of any type of peer effects, rather than the separate identification of endogenous and exogenous effects. These studies rule out correlated effects from selection into groups by focusing on applications where students are randomly assigned to groups through an exogenous administrative process.

Sacerdote (2001) studies peer effects in the context of roommate pairs at Dartmouth University. Freshman roommates are randomly assigned, conditional on a small set of indicated preferences. He estimates freshman year GPA as a function of own academic performance in high school and roommate's academic performance in high school. Peer effects are measured by the coefficient on roommate's academic performance in high school, but the exact nature of the peer effect, whether endogenous or exogenous, cannot be determined. [2]

Sacerdote finds limited evidence of peer effects on academic outcomes. In particular, he finds that having a roommate in the top quartile of academic performance in high school increases freshman GPA by 0.047 points. Sacerdote finds more robust impacts on social outcomes. He finds that the probability of joining a fraternity after freshman year is 8 percentage points higher for students whose roommates joined a fraternity.

Peer effects in a setting of group interactions are explored by Carrell, Fullerton, and West (2009). They studied first-year students in the US Air Force Academy, who are randomly assigned to peer-groups defined by membership in a particular squadron.

[2] Sacerdote (2000) proposes an approach to separately identify these effects in the working paper version, but considers the required variance restrictions too onerous.

They found that the average peer verbal SAT score had a significant positive impact on freshman GPA.

Similarly, Lyle (2007, 2009) estimates peer effects amongst first-year students at the United States Military Academy at West Point. Lyle, however, found very little evidence for the existence of peer effects at West Point. His analysis produced some evidence of peer effects in non-academic outcomes, such as choice of major and re-enlistment, but he found no evidence thate peer effects influence academic outcomes.

Identification of a composite peer effect parameter in a particular educational environment is a valuable contribution, especially given our limited understanding of peer effects and the high political salience of educational outcomes. Clear evidence of peer effects in education increases our understanding of how learning happens. The policy relevance of such evidence, however, is limited. The single composite parameter does not distinguish between peer effects driven by exogenous characteristics of students and those driven by the endogenous choices students make in a particular environment. If an individual student improves his or her outcome through an increase in unobservable effort, it is impossible to predict whether that change will impact the entire class.

The policy implications, then, are limited to changes in the composition of classrooms. A fully linear model, moreover, can only inform questions of equity. Classrooms may be reconfigured to increase or decrease the variation between classrooms, but aggregate outcomes cannot be changed. Carrell, Sacerdote, and West (2013) circumvent the limitations of the fully linear model by looking for non-linear impacts of peer effects. They divide students into terciles based on expected academic performance, and allow for different peer effects within each group. Based on these results, the authors were able to design a classroom assignment procedure that produced a Pareto improvement in expected outcomes. In this groundbreaking study, the authors were able to implement their classroom assignment procedure in a subsequent cohort of first-year students. The results, unfortunately, were the opposite of what was expected. Low-ability students, who were to have benefited most from the new assignment procedure, actually had worse outcomes than under random assignment.

These results illustrate the fragility of peer effect estimates, and their dependence on a particular social environment. Changes in group-composition or the group assignment process can alter the nature of peer effects. Estimates should be regarded as context specific, and should be applied to other settings with caution. This is particularly true when peer effects are estimated as a single composite parameter.

The divergence of actual outcomes from predicted outcomes could be attributed in part to the lack of separate identification between endogenous and exogenous peer effects. It is plausible that endogenous and exogenous peer effects could react differently to changes in group composition. Exogenous peer effects are those that can be attributed to the mean exogenous characteristics of a group. By definition, individuals are not able to change their exogenous characteristics. In theory, then, exogenous peer effects should be stable after resorting individuals into groups.

Endogenous peer effects are the result of choices that are made by individuals within a given peer group. They are frequently motivated by concepts such as social conformity and unobservable effort. Endogenous peer effects are likely to be sensitive to changes in group composition. Given the sensitivity of endogenous peer effects, clear identification in a given environment will not reliably predict the nature of endogenous peer effects after a change in group composition. It will, however, provide separate estimates of effects that are likely to be robust to changes in group composition versus those that are likely to be influenced by group composition.

Separate identification of endogenous and exogenous effects, then, has the potential to improve predictions of impacts from changes in group composition. It also has implications for policies that target individual behavior or performance within groups. Endogenous peer effects imply a multiplicative effect, whereby improving the outcome of an individual student has a positive impact on all students in a classroom. Likewise, a decrease in a single student's performance will decrease the performance of all students in the classroom. Exogenous peer effects do not produce this multiplicative impact. The presence of endogenous peer effects, then, has implications for policies such as targeted tutoring and classroom aids, that only directly impact specific students within a classroom. In the presence of endogenous peer effects, targeted interventions may actually improve the outcomes of all students.

An example where endogenous and exogenous peer effects have different policy implications can be found in the work of Carrell, Malmstrom, and West (2008), who study academic cheating in college. Like previous studies, they analyze data from students in American military service academies. The data were collected from anonymous surveys mailed to the graduates of each academy, that asked graduates to self-report cheating behavior, both in high-school and in college. They consider the probability that a respondent reported cheating in college as a function of the prevalence of reported cheating among the respondent's classmates.

Two measures of peer cheating are considered. The first is based on reported cheating in high school, and the second is based on reported cheating in college. Given that the outcome variable of interest is cheating in college, the prevalence of cheating in high school is an exogenous characteristic of the peer group, while the prevalence of cheating in college is an endogenous characteristic. Both measures of peer cheating produce large and significant estimates of the impact of peer cheating on self-cheating. Using a linear probability model, the estimated coefficient on the proportion of peers that cheated in high school is 0.342. This is larger than the coefficient on own cheating in high school, 0.132. The endogenous specification also estimates peer effects using a linear probability model. The coefficient on the proportion of peers who cheated in college is 0.746.

The authors use the terms endogenous and exogenous to describe these two specifications, but the terms can only accurately be applied to the nature of the peer cheating variables. The specifications do not provide separate estimates of the endogenous and exogenous channels of peer effects identified by Manski. Instead, each specification estimates a single, composite parameter composed of both endogenous and exogenous peer effects. As noted by the authors:

> Absent a more refined, specific, and well-justified theoretical model, we cannot say whether peers affect own college cheating through simultaneous peer college cheating, or predetermined peer characteristics, or some combination of each. In the former specification, our reduced form coefficients represent the correlation of pretreatment peer characteristics on college cheating, whether the effect is direct or indirect through peer college cheating. In the latter specification, we restrict the entire peer effect to occur through peer college cheating. For ease of exposition, we refer to these approaches as an exogenous peer effect and an endogenous peer effect. (p.183)

In this application, identification of the causal channel of the estimated peer effects would clarify the root causes of academic cheating. This information, in turn, could inform policy recommendations aimed at reducing academic cheating. The presence of endogenous peer effects suggest that there would be large returns to interventions to reduce cheating, such as increased enforcement or the implementation of an honor code. Exogenous peer effects, meanwhile, do not respond to interventions targeted at contemporaneous behavior. Instead, they are driven by previous behavior. Reducing the exogenous peer effects of cheating would require an intervention at the point of student admission, such as increased screening of applicants with any history of cheating.

Separate identification of endogenous and exogenous peer effects, then, can improve the policy relevance of peer effects estimates. More importantly, separate identification is crucial to understand the true nature of peer effects and the production of educational outcomes. Consider, for example, the black-white test score gap. Data show a persistent performance gap between white and black students on standardized tests designed to measure vocabulary, math, and scholastic aptitude. While the existence of the performance gap has been well documented, no consensus has emerged as to its root cause. In many studies, the gap persists even after controlling for individual and family characteristics (the exception is Fryer Levitt, 2004).

After controlling for individual characteristics, societal and environmental explanations must be considered. Card and Rothstein find that the test score gap is higher in cities with higher degrees of segregation. This evidence suggests that students are being influenced by the performance or characteristics of their peers, but cannot identify the exact channel of the interactions. For example, shared economic and social characteristics of peers within a segregated city may act through the channel of exogenous peer effects. If endogenous peer effects are also present, any change to the average outcome of the peer groups will be magnified throughout the group. Endogenous peer effects can magnify differences between racial groups in a segregated city that would otherwise have been small.

According to Jencks and Phillips (1998), closing the black-white test score gap would be one of the most effective ways to eliminate racial inequality. Understanding the nature of the gap and the social forces driving it is the first step in closing it. Given the underdeveloped state of econometric techniques to estimate social interactions, many studies seek only to identify the presence of social interactions. The approach presented here goes further, and attempts to identify the different channels of social interactions. This approach has the potential to inform important societal questions, such as the underlying causes of the black-white test score gap.

Identification

Charles Manski (1993) both formalizes and names the reflection problem in his seminal paper. A second major treatment of the problem was undertaken by Robert Moffitt in 2001. A survey of developments in the identification of social interactions since Manski's initial statement of the problem is provided by Blume, Durlauf, Brock, and Ioannides (2010). Broadly speaking, social interaction estimation strategies can be categorized into two approaches, regression based and variance based (Durlauf and Tanaka, 2008).

Regression based approaches generally take one of two routes to identification. The first introduces nonlinearity to separate group-level characteristics and outcomes from individual outcomes. Binary choice models, for example, specify an individual's choice as a non-linear function of that individual's characteristics, group characteristics, and group outcomes. Identification in this case was established by Brock and Durlauf (2007).

The second approach uses instrumental variables to separate endogenous and exogenous social interactions. Instrumental variables in this context are problematic, as traditional exclusion restrictions are difficult to justify. An excluded variable would have to impact the outcome of an individual's social reference group without directly affecting the individual outcomes of group members. Instead of relying on traditional exclusion restrictions, the IV approach relaxes strict network structure that results from the assumption of group-interactions. Given *a priori* knowledge of the network structure, it is possible to construct instruments using the characteristics and outcomes of friends-of-friends. Friend-of-friend relationships can impact outcomes indirectly through the channel of endogenous peer effects, without directly contributing to the average characteristics of an individual's first-order network connections. This allows the separate identification of endogenous and exogenous peer effects. Under the assumption of group interactions, individuals are completely partitioned, so second-order relationships do not exist.

The relaxation of the group-interaction structure moves the problem into a more general setting, where techniques from spatial econometrics and network economics can be applied. In the context of spatial econometrics, group-interactions can be described as a special case in which the spatial weighting matrix is block-diagonal and the elements of each block are the inverse of group-size. Overviews of the fields of network economics and spatial econometrics can be found in the works of Jackson (2008) and Anselin, Florax, and Rey (2004), respectively. Recent work on identification in the context of network interactions can be found in the work of Bramoullé, Djebbari, and Fortin (2009) and Lee, Liu, and Lin (2010).

Variance based approaches attempt to identify characteristics of the model covariance structure that are indicative of the presence of social interactions. The crucial insight that social interactions will impact the covariance structure was provided by Glaeser, Sacerdote, and Scheinkman (1996). The insight was subsequently developed into an estimation strategy capable of producing point estimates for social interactions (Graham, 2008a). Graham's approach requires random assignment to groups and is not able to separately identify endogenous and exogenous social interactions.

The approach presented here combines elements of the regression and variance-based methods. It is similar to strategies proposed by Krauth (2006) and Sacerdote (2000). Krauth works in a group setting using binary outcome variables, while Sacerdote works with roommate pairs and analyzes a continuous outcome. The applications chosen by these two authors are quite different, but they both impose the restriction of joint normality on the unobservable characteristics of individuals within peer groups. In both cases, the joint normality assumption implies a specific functional relationship between the unobservable characteristics of individuals within each group. This relationship is used as an additional estimation equation, which allows for the separate identification of endogenous and exogenous peer effects. These methods require consistent estimates of individual unobservables, which places significant restrictions on group-level unobservables and inter-group correlation. Similarly, the approach presented here also requires restrictions on group-level unobservables. The assumption of joint normality, however, is not required.

1.2 Empirical Strategy

Like many previous studies, I will use the linear-in-means specification to model social interactions. The linear-in-means model has been widely deployed to estimate social interactions in a wide range of contexts. Microeconomic foundations for the linear-in-means model have recently been developed by Blume et al. (2010). The linear-in-means specification is derived from the Bayes-Nash equilibrium strategy profiles of individuals participating in a noncooperative game of incomplete information. Individual utility functions of this game depend on both individual and group outcomes.

The estimator is based on the work of Klein and Vella (2010, hereafter KV). The KV estimator was designed to identify the parameter on an endogenous regressor in a triangular system of equations without the use of an exclusion restriction. The estimator introduces a non-linear control function, which provides the non-linearity required to overcome the reflection problem. The non-linear control function is estimated using information from the covariance structure. The control function will be estimated with both nonparametric and semiparametric methods (Ichimura, 1993), which allows the required covariance assumptions to be relaxed relative to those required by previous variance-based estimators.

After a full exposition of the estimation strategy, including Monte Carlo simulation results, I will demonstrate the strategy using data from the class-size reduction experiment, Project STAR (Word et al., 1990). The data has an experimental component that has been used in previous analysis (Krueger, 1999; Krueger and Whitmore, 2000; Whitmore, 2005; Graham, 2008a; Boozer and Cacciola, 2001). The current analysis most closely follows the work of Graham (2008a). Graham's approach takes advantage of the random assignment of the small classroom treatment, so can be considered an unbiased estimation of social interactions. Graham develops a novel estimation strategy in that his primary equation estimates a conditional variance, rather than a conditional mean. He constructs an instrumental variable using the assumption that the random small-class assignment will alter the within-classroom conditional variance of student unobservables, but will not impact other classroom characteristics that contribute to student outcomes.

The same IV approach cannot be applied to the model below, as the primary equation estimates the conditional mean of student outcomes. In the model below, the endogenous variable is the class average outcome. It is not plausible that the small class-size treatment could effect the average classroom outcome without effecting individual student outcomes directly. Indeed, any potential exclusion restriction is problematic for this reason.

The randomized treatment variable is not used directly to achieve identification in the current approach. The particular application was chosen because plausible estimates of peer effects already exist, against which the current estimates can be compared. To this end, this analysis will consider the same student sample and outcome variables used by Graham. The use of experimental data to validate the construction of models that do not require experimental design has been undertaken with other data from other randomized control trials, e.g. (Todd and Wolpin, 2006).

1.2.1 The Model

I will estimate a linear-in-means model of the form below. For notational simplicity, I will designate expected group-level variables using overline notation, e.g. $\bar{Z}_g \equiv \mathbb{E}\left[Z_{gi} \mid i \in g\right]$. Expectations are replaced with simple group-means when the equation is estimated.

$$Y_{gi} = \bar{Y}_g \beta + \bar{X}_g' \gamma + X_{gi}' \eta + \alpha_g + \epsilon_{gi} \tag{1.1}$$

Where, for individual i in classroom g:

Y_{gi} Individual outcome

$\bar{Y}_g$ Expected value of individual outcome for group g

X_{gi} Individual exogenous characteristics

$\bar{X}_g$ Expected value of individual exogenous characteristics for group g

α_g Classroom unobservable

ϵ_{gi} Individual unobservable

Manski's three components of social interactions are captured by:

β Endogenous effect

γ Exogenous or contextual effect

$\alpha_g + \bar{\epsilon}_g$ Correlated effect

An equation for the endogenous variable, $\bar{Y}_g$, can be found by taking the expected value of equation (1) over groups and solving for $\bar{Y}_g$.

$$\bar{Y}_g = \frac{\bar{X}_g'(\gamma + \eta)}{1 - \beta} + \frac{\alpha_g + \bar{\epsilon}_g}{1 - \beta} \tag{1.2}$$

In order to consider equations (1) and (2) as a single triangular system, I impose the assumption that every group is in social equilibrium. Within each group, all members have the same expectation about the average outcome of that group. Furthermore, the shared expectation is correct (Graham and Hahn, 2005). Given this assumption, expectations of group-level variables may be replaced by the group-means of individual-level variables. The triangular system described by (1) and (2) will have a unique, self-consistent solution for $\beta \in (-1, 1)$ (Durlauf and Tanaka, 2008).

Considering equations (1) and (2) as a triangular system, there is not yet any identification problem. Instead, these equations present an endogeneity problem due to correlated error terms. Equation (1) can be estimated, after replacing expectations with corresponding group-averages, but the coefficients will be biased, as $corr\left(\alpha_g + \epsilon_{gi}, \frac{\alpha_g + \bar{\epsilon}_g}{1-\beta}\right) \neq 0$. This correlation between unobservable components implies that OLS estimation of equation (1) will result in biased coefficients, as $\mathbb{E}\left[\alpha_g + \epsilon_{gi} \mid \bar{Y}_g, \bar{X}_g, X_{gi}\right] \neq 0$. This endogeneity problem becomes an identification problem when $\frac{\alpha_g + \bar{\epsilon}_g}{1-\beta}$ is purged from the model using linear methods, such as two-stage least squares.

The approach I propose here avoids the reflection problem. Instead of purging the model of the endogenous error term, this method will construct a non-linear control function to account for the non-zero expected error term in the individual equation.

1.2.2 Identification

Following the design of the KV estimator, I will re-write the model as a triangular system. This is only a simplification of the notation used above. Using the composite parameters and unobservables defined below, equations (1) and (2) can be written as follows:

$$\begin{aligned}
&\pi \equiv \frac{(\gamma+\eta)}{1-\beta}\,;\; u_{gi} \equiv \alpha_g + \epsilon_{gi}\,;\; v_g \equiv \frac{\alpha+\bar{\epsilon}_g}{1-\beta} \\
&Y_{gi} = \bar{Y}_g\beta + \bar{X}_g^{\prime}\gamma + X_{gi}^{\prime}\eta + u_{gi} \\
&\bar{Y}_g = \bar{X}_g^{\prime}\pi + v_g
\end{aligned} \tag{1.3}$$

The relationship between group and individual unobservables implies that $corr(u_{gi}, v_g) \neq 0$. Like many two-stage estimators, the KV estimator uses an estimate of the error term from the secondary equation to construct a control function that is introduced in the primary equation. The control function captures the linear relationship between the two error terms using an estimate of $\mathbb{E}\left[u_{gi} \mid \bar{Y}_g, \bar{X}_g, X_{gi}\right]$, or, equivalently, $\mathbb{E}\left[u_{gi} \mid v_g, \bar{X}_g, X_{gi}\right]$. Including this expression in the primary equation, controls for the endogenous component of the error term, restoring the orthogonality conditions that are required for OLS regression. If we do not make any distributional assumptions on the error structure, we could estimate the linear conditional relationship between u_{gi} and v_g by solving the following minimization problem, where $\mathbf{X} = \left(\bar{X}_g, X_{gi}\right)$:

$$\begin{aligned}
A(\mathbf{X}) &= arg\min_A \mathbb{E}\left[u_{gi} - Av_g \mid \mathbf{X}\right]^2 \\
&= \frac{cov\left(u_{gi}, v_g \mid \mathbf{X}\right)}{var\left(v_g \mid \mathbf{X}\right)}
\end{aligned} \tag{1.4}$$

Then:

$$A(\mathbf{X})\, v_g = \mathbb{E}\left[u_{gi} \mid v_g, \bar{X}_g, X_{gi}\right] \tag{1.5}$$

Control function estimators frequently assume independence between the error terms and $\mathbf{X}$, so that the distribution of the true error terms, u_{gi} and v_g do not depend on $\mathbf{X}$. The conditional coefficient, $A(\mathbf{X})$, then simplifies to a constant coefficient, $a = \frac{cov(u,v)}{Var(v)}$, and $\mathbb{E}\left[u_{gi} \mid \bar{Y}_g, \bar{X}_g, X_{gi}\right] = av_{gi}$. In practice, av is estimated by including an estimate of v in the primary regression. In the absence of an exclusion restriction, however, the estimate, $\hat{v}$, is a linear function of $\bar{Y}_g$ and $\bar{X}_g$, so the regressors of the controlled regression are not of full rank. Exclusion restrictions are difficult to justify in the context of group-interactions. An excluded variable would have to impact the outcome of an individual's social reference group without directly affecting the individual outcomes of group members. Instead, this approach does not impose the assumption that the distribution of the error terms is independent of $\mathbf{X}$, and instead estimates $A(\mathbf{X})\, v_g$ directly. Inclusion of the control function in the primary equation leads to the final, controlled regression equation.

$$Y_{gi} = \bar{Y}_g\beta + \bar{X}_g'\gamma + X_{gi}'\eta + A(\mathbf{X})v_g + e_{gi} \tag{1.6}$$

where $e_{gi} \equiv u_{gi} - A(\mathbf{X})v_g$ and $corr(e_{gi}, v_i \mid \mathbf{X}) = 0$.

The specific control function estimator proposed by KV was constructed from an error structure composed of two correlated random variables, u and v. Each random variable appears in exactly one equation of the triangular system. The resulting control function is not consistent with the current social interaction model, where the error term of each equation is a composite of group and individual components. A new control function must be derived to reflect the error structure of the social interaction model. Derivation of the control function will generally follow the KV approach, where the nonlinear structure of the control function is modeled under the assumption of multiplicative heteroscedasticity. The KV approach models the conditional variance of the equation-specific error terms, u and v. For the social interaction model, conditional variance functions will be modeled separately for group and individual error components, α and ε. Group and individual conditional variance functions are defined in 1.7 below. Each error component, then, can be written as the product of a homoscedastic term and a scaling function that depends on observable characteristics. Finally, the error term from each equation can be restated in terms of group and individual components.

$$\begin{aligned} S^2_{\epsilon gi} &\equiv Var(\epsilon_{gi}|\mathbf{X}) & S^2_{\alpha g} &\equiv Var(\alpha_g|\mathbf{X}) \\ \epsilon_{gi} &= S_\epsilon \epsilon^*_{gi} & \alpha_g &= S_\alpha \alpha^*_g \\ u_{gi} &= S_\alpha \alpha^*_g + S_\epsilon \epsilon^*_{gi} & v_g &= \frac{1}{(1-\beta)} \frac{1}{n_g} \sum_{i \in g} \left(S_\alpha \alpha^*_g + S_\epsilon \epsilon^*_{gi} \right) \end{aligned} \tag{1.7}$$

Derivation of the control function for this error structure requires several assumptions.

$$\begin{aligned} \mathbb{E}(\epsilon^*_{gi}|\mathbf{X}) &= 0 \\ \mathbb{E}(\alpha^*_g|\mathbf{X}) &= 0 \\ corr\left(\epsilon^*_{gi}, \alpha^*_g \mid X\right) &= 0 \\ corr\left(\epsilon^*_{gi}, \epsilon^*_{gj} \mid X\right) &= \delta \end{aligned} \tag{1.8}$$

The first two assumptions are standard strict exogeneity conditions. The third assumption requires that the group and individual error components are uncorrelated. Finally, the last restriction requires that any correlation of individual unobservable characteristics within groups is constant, after adjusting for individual heteroscedasticity. The full derivation of the control function is given in the appendix. The final control function for this error structure is:

$$A\left(\mathbf{X}\right)v_g = \left(1-\beta\right)\frac{\left[S_{\alpha g}^2+\frac{1}{n_g}\left(S_{\epsilon gi}^2+\delta S_{\epsilon gi}\sum_{j\neq i}S_{\epsilon gj}\right)\right]}{\left[S_{\alpha g}^2+\frac{1}{n_g^2}\sum_{i\in g}\left(S_{\epsilon gi}^2+\delta S_{\epsilon gi}\sum_{j\neq i}S_{\epsilon gj}\right)\right]}v_g \tag{1.9}$$

It is not immediately clear whether it is possible to estimate this control function without additional assumptions. For the purpose of discussion, I will assume that it is possible to estimate the components of the control function. I will return to the issue of feasibility in the next section.

Suppose it is possible to produce consistent estimates of the following two functions:

$$\begin{aligned} f_{gi} &= S_{\alpha g}^2+\frac{1}{n_g}\left(S_{\epsilon gi}^2+\delta S_{\epsilon gi}\sum_{j\neq i}S_{\epsilon gj}\right) \\ h_g &= S_{\alpha g}^2+\frac{1}{n_g^2}\sum_{i\in g}\left(S_{\epsilon gi}^2+\delta S_{\epsilon gi}\sum_{j\neq i}S_{\epsilon gj}\right) \end{aligned} \tag{1.10}$$

Then, including an estimate of $\frac{f_{gi}}{h_g}v_g$ as a generated regressor would yield an estimated coefficient, $\rho=(1-\beta)$. An estimate of the group-level unobservable, $\hat{v_g}$, can be obtained through an OLS regression on the secondary equation in (3). Given consistent estimates of the non-linear functions, f_{gi} and h_g, then, the final controlled regression would be estimated as follows:

$$Y_{gi} = \bar{Y_g}\beta + \bar{X}_g'\gamma + X_{gi}'\eta + \rho\frac{\hat{f_{gi}}}{\hat{h_g}}\hat{v_g} + e_{gi} \tag{1.11}$$

All coefficients in this regression are identified, including separate estimates of endogenous and exogenous social interactions. Furthermore, two estimates of endogenous social interactions are provided, one directly from β, and a second from ρ, which is a known transformation of β. Specifically, $\rho=(1-\beta)$, as direct inspection of the control function reveals. Of course, this strategy relies on consistent estimates of the non-linear components, f_{gi} and h_g. I will consider that question in the next section.

1.2.3 Simplified Error Structure

Estimation

Construction of the control function requires estimation of two complex non-linear functions. Furthermore, the performance of this estimator depends on the degree of heteroscedasticity in the data generating process. In order to demonstrate the concept of this estimator, I will present simulation evidence using a simplified data generating process. Specifically, I will assume that both the variance of group-level unobservables and the correlation between individual unobservables within groups are equal to zero ($S_{\alpha g}^2 = 0, \delta = 0$). This simplified error structure will demonstrate the key role that individual-level heteroscedasticity plays in this estimator.

For this simulation, I will assume that outcomes are generated according to the following system of equations:

$$
\begin{aligned}
Y_{gi} &= \bar{Y}_g \beta + \bar{X}_g' \gamma + X_{gi}' \eta + u_{gi} \\
\bar{Y}_g &= \bar{X}_g' \pi + v_g
\end{aligned}
$$

Given the simplifying assumptions above, the error terms can be written as follows:

$$
\begin{aligned}
u_{gi} =& S_{\epsilon gi} \epsilon^* \\
v_g =& \frac{1}{(1-\beta)} \frac{1}{n_g} \sum_{j \in g} S_{\epsilon gj} \epsilon_{gj}^*
\end{aligned}
$$

The simplified control function for this system is:

$$A(\mathbf{X})\, v_g = (1-\beta) \frac{S_{\epsilon gi}^2}{\frac{1}{n_g}\sum_{j\in g} S_{\epsilon gj}^2} v_g$$

The control function contains only one unknown, non-linear component, $S_{\epsilon gi}$. Estimation of this component will be simplified if we consider the within transformation of the data.

$$\widetilde{Y}_{gi} = \widetilde{X}_{gi}^{'}\eta + \tilde{u}_{gi}$$

Where for any $z_{gi} \in (Y_{gi}, X_{gi}, u_{gi}) : \tilde{z}_{gi} = z_{gi} - \frac{1}{n_g}\sum_{j\in g} z_{gj}$. The within transformation eliminates the endogenous regressor, $\bar{Y}_g$, enabling the consistent estimation of $\tilde{u}_{gi}$. Furthermore, estimates of $\tilde{u}_{gi}$ provide a straightforward way to estimate $S_{\epsilon gi}^2$.

$$\begin{aligned}
var\left(u_{gi} \mid X_{gi}\right) &= S_{\epsilon gi}^2 \\
var\left(\tilde{u}_{gi} \mid X_{gi}\right) &\approx \frac{n_g - 1}{n_g} S_{\epsilon gi}^2 \\
var\left(\sqrt{\frac{n_g}{n_g - 1}}\tilde{u}_{gi} \mid X_{gi}\right) &\approx S_{\epsilon gi}^2
\end{aligned}$$

$S_{\epsilon gi}^2$ can be estimated semiparametrically, imposing the constraint that X_{gi} enters the conditional variance function as a linear function, with the coefficient on the first element constrained to 1. The index parameters are then estimated using semiparametric least squares (Ichimura, 1993). Estimates for $S_{\epsilon gi}^2$ are simply the predicted values from the semiparametric least squares regression.

The control function also requires an estimate of v_g. This estimate cannot be recovered from the within transformation, since all group means have been purged from the data. Instead, v_g will be estimated as the residual from the between group regression.

$$\bar{Y_g} = \bar{X}_g'\pi + v_g$$

The coefficient estimates from this regression will be composites of the true structural parameters, (β, γ, η). Specifically, $\pi = \frac{(\gamma+\eta)}{1-\beta}$. The residuals, however, will be consistent estimates of v_g. The generated regressor, then, will be: $\frac{\hat{S}^2_{\epsilon gi}}{\frac{1}{n_g}\sum_{j\in g}\hat{S}^2_{\epsilon gj}}\hat{v}_g$. The coefficient estimate on this regressor will be $\hat{\rho} = \widehat{1-\beta}$.

The form of this estimator is very similar to the social interactions estimator developed by Graham, though the two approaches differ in motivation and interpretation. Graham's approach identified the effect of social interactions through a ratio of within variation to between variation. In the current application, the control function also takes the form of a ratio of within to between variation. This is only due, however, to the assumptions of zero variance in group-level unobservables and zero within-group correlation of unobservables. The goal of the control function estimator is to construct a ratio of covariance to variance, as all regression coefficients are estimated. In the absence of group-level variance and within group correlation, the covariance between u_{ig} and v_g reduces to the variance of u_{ig} times the scalar, $\frac{1}{1-\beta}$

Simulation

Simulations of the control function estimator were conducted with the following data generating process:

$$
\begin{aligned}
X_{1gi} &\sim N(0,2)\\
X_{2gi} &\sim N(0,2)\\
\varepsilon^*_{gi} &\sim N(0,2)\\
\varepsilon_{gi} &= \exp\left(0.3 * X_{1gi} + 0.6 * X_{2gi}\right) * \varepsilon^*_{gi}\\
Y_{gi} &= 0.2 * \bar{Y}_g + X_{1gi} + 2 * X_{2gi} + 0.5 * \bar{X}_{1g} + 0.5 * \bar{X}_{2g} + \varepsilon_{gi}
\end{aligned}
$$

The simulation results are from 50 replications. The number of groups in each replication was fixed at 500. Group size, however, varied uniformly from 16-30 members, for an average sample size of 11,480. Variation in group size is not required for this estimator. Group sizes vary in this simulation simply to reflect the typical structure of classroom groups. Simulation results are presented in table 1. The standard deviation over the 50 replications is given in parentheses below each parameter estimate.

Two approaches are taken in estimating the individual conditional variance function. The first approach restricts the conditioning set to individual-level covariates only. This is a natural approach for the current simulation, where the conditional variance function is known to be a function of the individual-level covariates, X_{1gi} and X_{2gi}. The second approach estimates the variance function conditional on both individual and the group-level covariates. The included group-level covariates are the group means and variances of individual characteristics, as well as class size. This approach is meant to reflect the approach that might be taken in an empirical application, where the true conditioning set is not known. Indeed, this is the approach that I use in the empirical application presented in section 1.3.

Estimates from the first approach, using only individual-level covariates, are presented in column (3) of table 1.1. In most cases, the parameter estimates are close to their true values. Aside from the γ parameters, which represent exogenous peer

Table 1.1: Simulation results

Parameter	True (1)	OLS (2)	Ind (3)	Grp (4)
β	0.2	1	0.191	0.153
		(<0.000)	(0.041)	(0.0614)
$\rho = 1 - \beta$	0.8		0.809	0.847
			(0.041)	(0.0614)
γ_1	0.5	-1.018	-0.176	0.566
		(0.076)	(0.055)	(0.119)
γ_2	0.5	-2.027	-0.358	0.493
		(0.137)	(0.114)	(0.188)
η_1	1	1.018	1.001	0.998
		(0.076)	(0.049)	(0.0106)
η_2	2	2.027	2.000	1.988
		(0.137)	(0.076)	(0.0171)
$\frac{\theta_2}{\theta_1}$	2		1.92	
			(0.505)	

effects, these estimates do not reject the null hypothesis that each underlying parameter is equal to its true value. The γ parameter estimates, however, exhibit significant bias. This may be due to the orthogonality of $\hat{v}_g$ and $\bar{X}_g$ that is a result of estimating $\hat{v}_g$ from the between equation, combined with the lack of specific group-level heteroscedasticity. In this model, the only group-level heteroscedasticity is generated by averaging over individual heteroscedasticity. Additional group-level heteroscedasticity may improve the estimates of γ in the controlled regression.

Column (4) contains the estimates from the second approach, conditioning on both individual and group-level covariates. The estimates of exogenous peer effects, represented by the γ parameter, are improved by the addition of group-level covariates. The estimates of endogenous peer effects, represented by the β paramenter, however, are biased in the second approach.

Coefficient estimates are heavily reliant on individual heteroscedasticity, as this is the only source of non-linearity enabling the separate identification of exogenous and endogenous peer effects. Heteroscedasticity of the within residuals was tested for each replication. Specifically, I tested the null hypothesis of constant variance against the alternative that the variance of the within residual is an unrestricted function of a linear combination of X_1 and X_2. I used a version of the Breusch-Pagan test for heteroscedasticity that does not require normally distributed errors (Breusch and Pagan, 1979; Koenker, 1981). The degree of heteroscedasticity in this model is very high. The mean of the test statistic (distributed chi-squared with two degrees of freedom) over the 50 replications was 313. Even with this high level of heteroscedasticity, it is difficult to estimate both endogenous and exogenous peer effects in the same equation.

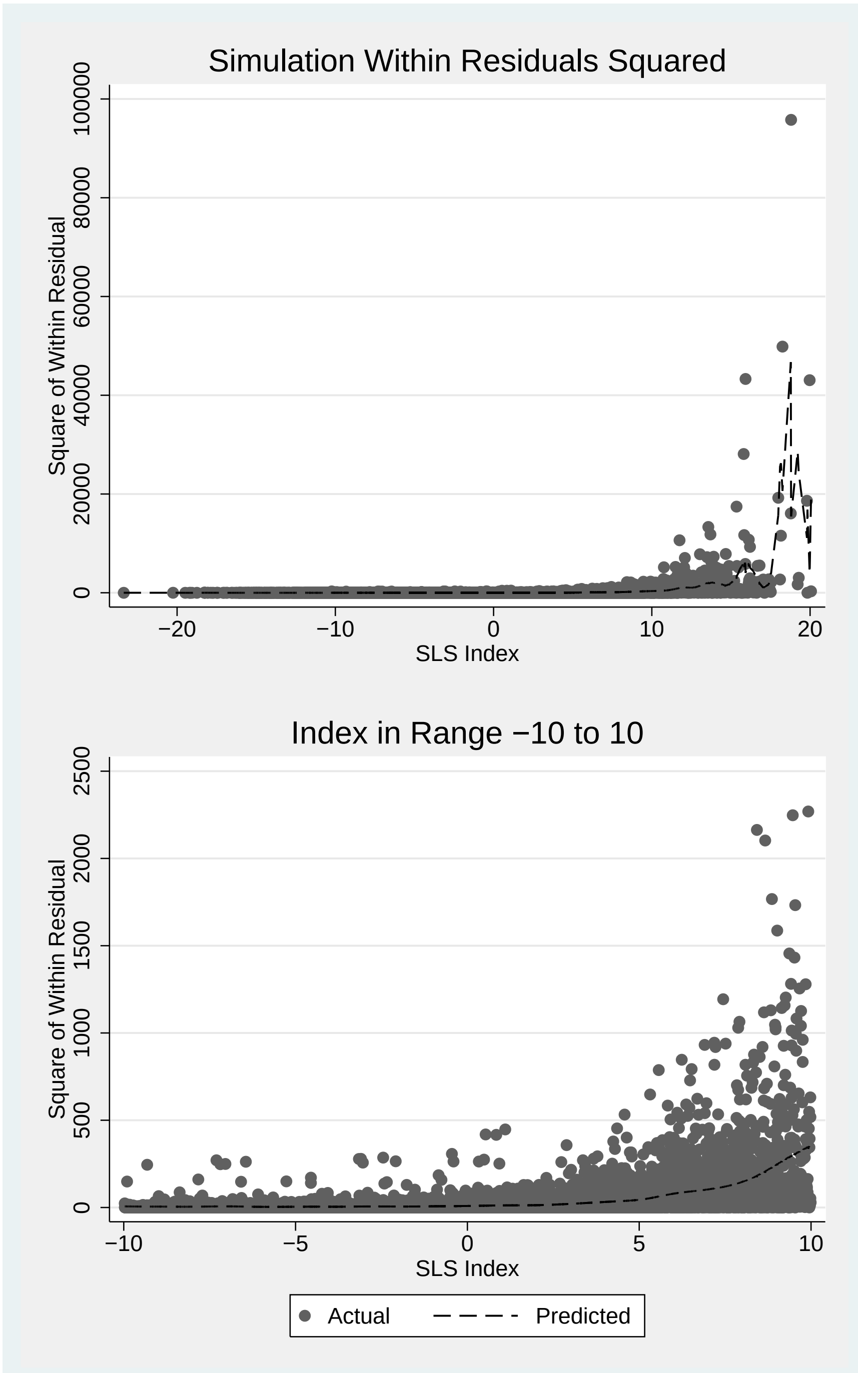
Figure 1.1 shows the observed residuals (squared) and the predicted conditional variance for the within residuals. These graphs are based on the data generated from the first replication only. Graphs based on data from subsequent replications are similar. The top graph shows all residuals and estimated values in the simulation. The bottom graph shows only those observations where the estimated semiparametric index value is between -10 and 10. The second graph simply shows more detail on the behavior of the conditional variance function over these mid-range values.

1.3 Application

1.3.1 Data Description

I will demonstrate the application of this estimation technique by re-estimating the effect of social interactions in the same context as Graham (2008a).

Figure 1.1: Conditional Variance Estimation



I will use data from Project STAR, the class-size reduction experiment conducted in Tennessee in 1985 (Word et al., 1990). The program design randomly assigned incoming students to classrooms within schools. While school choice is still endogenous in this design, classroom assignment is random. In order to evaluate the effect of smaller class sizes, incoming students were assigned to one of three types of classroom. Classroom types are small (13 to 17 students), regular (22 to 25 students), or regular with a full-time aide.

In essence, classrooms were first randomly assigned to a treatment group of control, small-class treatment, or full-time aide treatment. Classroom assignment determined the number of seats available in each classroom. Students were then randomly assigned to a type of classroom. If a school had more than one classroom type, it is not clear how students were assigned to classrooms within each type. All participating schools have at least three kindergarten classrooms, so have at least one classroom of each type. Many schools (48 of 79) have more than three kindergarten classrooms, so have more than one class for some type. Because students were randomly assigned to classroom types rather than classrooms, the possibility exists for sorting or matching within classroom types. Despite this possibility, Graham presents results showing that the assumption of random assignment to classrooms is plausible.

For the preliminary analysis, I analyze the same sample used by Graham. Reconstruction of the sample was simple, as Graham generously provided his Stata files as supplementary material (Graham, 2008b). This sample consists of 6,172 incoming kindergarten students who are divided into 317 classrooms across 79 schools. The final analysis sample consists of 5,643 students who had non-missing math and reading scores.

Graham's estimation strategy took full advantage of the experimental nature of this data set. His strategy relied on both the randomized assignment classrooms to treatment groups and of students to classroom type. First, test score variance was decomposed into within-group and between-group components. Given a random allocation of students to classrooms, the within-group variance of student-test scores predicts one component of between-group variation. The other component of between-group variation comes from variation in teacher characteristics. Furthermore, the impact of within-group variation on between-group variation is magnified in the presence of peer effects.

Graham argues that the small-class treatment affects the variance of student characteristics but does not affect the variance of teacher characteristics. In estimating between-group variance as a function of within-group variance, then, the small-class treatment can be used to construct an instrumental variable for within-group variance. It is this exogenous change in within-group variance that is used to identify the peer effect coefficient.

The approach pursued here also relies on the random allocation of students to classrooms. The small-classroom treatment variable, however, cannot be used to construct an IV, as in Graham's analysis. In the current model, individual test scores are estimated as a function of classroom average test scores. It is not plausible that the small-classroom treatment, or any classroom intervention, could impact classroom average test scores without affecting individual test scores. In fact, this is a general problem in estimating peer effects in a setting of group-interactions. It is difficult to imagine any plausible exclusion restriction, because the key endogenous variable is simply an average of individual outcomes. There is no way to change the group average outcome except by changing individual outcomes within the group.

1.3.2 Model Specification

For the preliminary estimation, I assume the same simplified error structure described above. This requires zero correlation between individual unobservables and zero variance of teacher and classroom unobservables.

The first assumption is plausible in the current application. Where students have been randomly assigned to classroom types. The same assumption is required for the estimator developed by Graham and deployed on this data. Some potential remains for correlation of individual unobservables. Large schools may have more than one of each type of classroom. Students assigned to that classroom type could be sorted into classrooms of the same type. This is unlikely in the analysis data, which is limited to results from kindergarten. It is unlikely that student unobservables were known to each other or to administrators before enrollment in kindergarten.

The second assumption, of zero variance in teacher and classroom unobservable characteristics, is likely overly strong for the present application. This requires that all variance in group unobservables is due solely to variance in the mean of individual unobservables. In fact, teacher and classroom characteristics are likely to vary between groups, increasing the correlation of outcomes within classrooms. Because this excess correlation will not be modeled in the control function, it will likely be included in the endogenous or exogenous social interaction coefficients, leading to excessively high estimates of peer effects.

Although these estimates are based on unrealistic assumptions on unobservables, they will still be useful in evaluating the performance of this estimator. In future work, I plan to relax the assumptions on unobservables, allowing for more flexible and realistic specifications. The current estimates will provide a baseline, so it will be possible to see how the coefficient estimates change as the assumptions on the error term are relaxed.

The preliminary estimation will be conducted according to the following model:

$$
\begin{aligned}
Y_{gi} &= \bar{Y}_g\beta + \bar{X}_g'\gamma_1 + Z_g'\gamma_2 + X_{gi}'\eta + u_{gi} \\
\bar{Y}_g &= \bar{X}_g'\pi_1 + Z_g'\pi_2 + v_g \\
Var\left(\mathbf{U}_g \mid \mathbf{X}_g\right) &= \begin{bmatrix} S_{ug1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & S_{ugn_g}^2 \end{bmatrix}
\end{aligned}
$$

where Y_{gi} are individual test scores for individual i in class g. Both math and reading test scores are analyzed separately. Test scores were normalized to have zero mean and variance one. X_{gi} are individual observable characteristics for individual i in class g. All individual characteristics that are present in the data set are used. The first three variables are indicators for black, female, and low socioeconomic status. The final individual-level variable is age in quarters. Classroom averages of these variables are included in the group-level variables, $\bar{X}_g$. Finally, teacher characteristics are included in Z_g, including years of teaching experience, advanced degree status, and race. The classrooms small-classroom assignment status is also included as a classroom level control variable. Summary statistics for these variables are presented in tables 1.2 and 1.3.

Semiparametric estimation of conditional variance

Like the second approach of the simulation, individual conditional variance is estimated with individual and group-level covariates. This is done partially out of necessity. Semiparametric least squares requires at least one continuous independent variable. The individual-level variables available in this data set are binary or discrete.

Table 1.2: Student Characterisitics

Variable	Mean	Std. Dev.	Min.	Max.
Reading Score	0.0001	1.0002	-3.8494	6.0292
Math Score	0.0045	0.9997	-3.4596	2.9431
Age (years)	5.3882	0.3472	3.75	7.5
Low SES	0.4813	0.4997	0	1
Female	0.4861	0.4999	0	1
Black	0.3247	0.4683	0	1
N	5643			

Table 1.3: Teacher Characteristics

Variable	Mean	Std. Dev.	Min.	Max.
Teacher Black	0.161	0.368	0	1
Teacher Master's Deg.	0.35	0.478	0	1
Teacher Experience (years)	9.202	5.809	0	27
N	317			

The inclusion of class averages provides continuous regressors for semiparametric least squares. Including additional regressors also adds flexibility to the specification of the conditional variance function. Including classroom-level regressors in the conditional variance function allows group characteristics to influence individual variances. This indirectly allows for some variance in group-level unobservables, despite the assumption of zero variance in classroom unobservables.

It is possible that group-level covariates do not impact individual conditional variance, but this is not a concern in this application. The control function only requires the expected value of the conditional variance. Individual parameters of the index are not relevant. The inclusion of extraneous regressors in the conditional variance function may introduce problems in estimating and interpreting individual index parameters, but it will not interfere with the calculation of expected values of the dependent variable. In the same spirit, the within classroom variance of each individual-level regressor is included in the semiparametric conditional variance estimation.

The final regressor in the conditional variance estimation is class size. Class size is used in the conditional variance estimation in two distinct ways. First, class size is used as a scale factor to adjust for the change in variance caused by the within transformation. After the within transformation, conditional variance is reduced because of the subtraction of classroom means: $var\left(\tilde{u}_{gi} \mid X\right) \approx \frac{n_g - 1}{n_g} S^2_{\epsilon gi}$. The within residuals, then, are scaled up, so that the original conditional variance function can be estimated. This is done by using $\left(\frac{n_g}{n_g - 1}\right) \tilde{u}^2_{gi}$ as the dependent variable for conditional variance estimation. Second, class size may impact the distribution of individual unobservables independently of its impact as a statistical artifact of the within transformation. Smaller classrooms, for example, may result in more pressure to conform, constraining the distribution of individual unobservables. Class size is included as an explanatory variable in the conditional variance function to address these types of impacts.

The presence of heteroscedasticity in the distribution of residuals from the within equation is an important component of this estimation. As in the simulations, I tested for heteroscedasticity using a version of the Breusch-Pagan test that does not require normally distributed errors (Breusch and Pagan, 1979; Koenker, 1981). For the conditioning variables, I included the variables that are included in the semiparametric estimation of conditional variance. These variables include class size, three teacher characteristics, four individual characteristics, and classroom-level means and variances of each individual characteristics. The test statistics, both distributed chi-squared with 16 degrees of freedom, are 35.8 for math-score residuals and 69.9 for reading score

Table 1.4: Breusch-Pagan Test Results

Estimator	Math	Read
Semiparametric	35.61	70.46
chi2(16)	(0.0033)	(<0.000)
Nonparametric	27.11	74.54
chi2(15)	(0.0279)	(<0.000)

P-values in parentheses

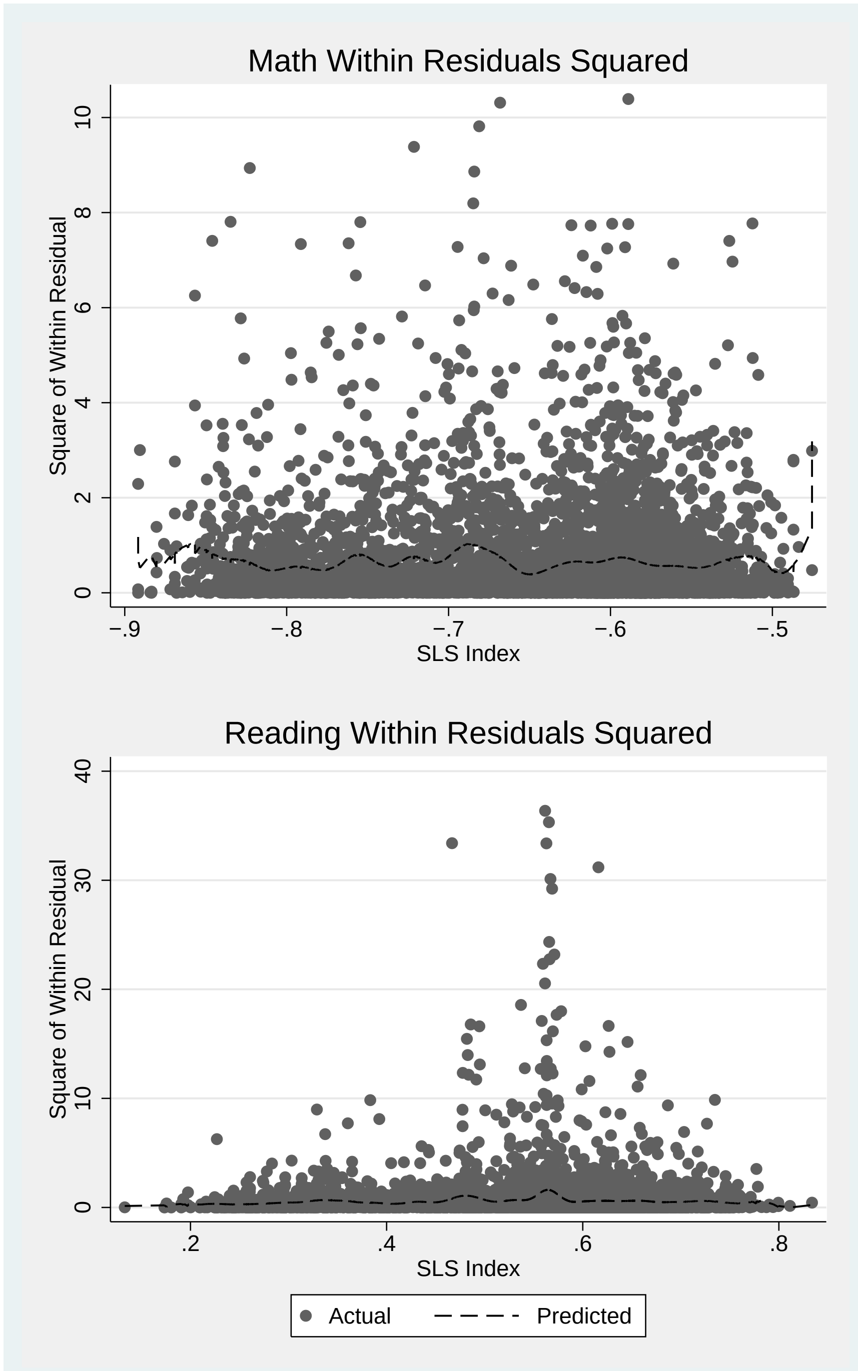
residuals. In both cases, the null hypothesis of constant variance can be rejected at the one-percent level, but the test statistics are an order of magnitude smaller than the average test-statistic in the simulated data.

Nonparametric estimation of conditional variance

Given the limited support of the individual-level covariates, the conditional variance function can also be estimated nonparametrically. Nonparametric estimates are essentially cell-averages of the dependent variable for cells defined by all unique combinations of the values of the independent variables. Coefficients are estimated with OLS. As in the semiparametric case, the dependent variable is the square of the residual from the within regression equation. The independent variables are all individual-level covariates. The majority of individual-level covariates, low SES, female, and black, are binary, which helps keep the number of cells manageable. Age in quarters, though, is not binary. Rather than estimate an additional set of cell-means for every value of age in quarters, I will use a binary measure of age. This measure will separate older and younger students using the median age of 21 quarters as a cut-off.

Implementation of the nonparametric estimation of conditional variance consisted of an OLS regression of the square of within residuals on the fully interacted co-

Figure 1.2: Semiparametric Conditional Variance Estimation



variates. Sixteen coefficients were estimated, representing the cell-averages for each combination of the four binary covariates. These estimates are presented in the appendix.

The 16 interaction terms were also used to test for heteroscedasticity of the within residuals using the same Breusch-Pagan test described above. The test statistic for the math residuals was 27.11, chi2(15), for a p-value of 0.0279. The test statistic for the reading residuals was 74.54, chi2(15), for a p-value less than 0.001. Given the limited support of the individual-level covariates, nonparametric estimation is an alternative approach to modeling conditional variance.

1.3.3 Results

Results are presented for math and reading test scores in tables 1.5 and 1.6, respectively. Interpretation of results from math and reading scores are similar. I will focus on the results from math scores for the purpose of discussion.

Endogenous Peer Effects

Endogenous peer effects refer to the impact of average peer outcomes on an individual's outcome. The estimate for endogenous peer effects is $\beta_{SP} = 0.67$ using semi-parametric estimation of conditional variance, and $\beta_{NP} = 0.451$ using nonparametric estimation.

The endogenous peer effect coefficient and the control function coefficient very nearly sum to one, as predicted by the econometric model. This restriction is not enforced within the estimation equations. It may be possible to produce more precise estimates by enforcing this equality restriction, but that extension is not pursued here.

Table 1.5: Math Results

VARIABLES	(1) OLS	(2) SPCTRL	(3) NPCTRL
Ave. Math Score	1.010***	0.672***	0.451***
	(0.0195)	(0.108)	(0.154)
Control Fxn		0.334***	0.553***
		(0.105)	(0.151)
Low SES	-0.422***	-0.420***	-0.419***
	(0.0260)	(0.0260)	(0.0260)
Age (quarters)	0.0705***	0.0711***	0.0713***
	(0.00775)	(0.00775)	(0.00775)
Female	0.133***	0.134***	0.134***
	(0.0214)	(0.0213)	(0.0213)
Black	-0.377***	-0.373***	-0.381***
	(0.0457)	(0.0457)	(0.0457)
Ave. Low SES	0.412***	0.244***	0.133
	(0.0606)	(0.0805)	(0.0973)
Ave. Age (quarters)	-0.0622*	-0.0259	-0.00188
	(0.0318)	(0.0338)	(0.0358)
Ave. Female	-0.128	0.0925	0.238*
	(0.0957)	(0.118)	(0.138)
Ave. Black	0.385***	0.347***	0.333***
	(0.0634)	(0.0645)	(0.0649)
Teacher Black	0.0155	0.0454	0.0650*
	(0.0335)	(0.0348)	(0.0361)
Teacher Master's Deg.	-0.00698	-0.0241	-0.0352
	(0.0228)	(0.0234)	(0.0241)
Teacher Experience (years)	-0.000546	0.00357	0.00627**
	(0.00190)	(0.00230)	(0.00266)
Constant	-0.170	-1.013	-1.565**
	(0.672)	(0.722)	(0.772)
Observations	5,643	5,643	5,643
R-squared	0.393	0.394	0.395

Standard errors in parentheses

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

Table 1.6: Reading Results

VARIABLES	(1) OLS	(2) SPCTRL	(3) NPCTRL
Ave. Reading Score	0.999***	0.855***	0.598***
	(0.0207)	(0.0679)	(0.0827)
Control Fxn		0.144**	0.401***
		(0.0647)	(0.0801)
Low SES	-0.459***	-0.461***	-0.466***
	(0.0265)	(0.0265)	(0.0265)
Age (quarters)	0.0482***	0.0490***	0.0490***
	(0.00791)	(0.00792)	(0.00790)
Female	0.160***	0.160***	0.161***
	(0.0218)	(0.0218)	(0.0217)
Black	-0.248***	-0.245***	-0.253***
	(0.0466)	(0.0466)	(0.0466)
Ave. Low SES	0.455***	0.369***	0.216***
	(0.0623)	(0.0734)	(0.0784)
Ave. Age (quarters)	-0.0446	-0.0557*	-0.0741**
	(0.0323)	(0.0327)	(0.0328)
Ave. Female	-0.158	-0.0709	0.0814
	(0.0975)	(0.105)	(0.108)
Ave. Black	0.251***	0.231***	0.209***
	(0.0647)	(0.0653)	(0.0651)
Teacher Black	0.00165	0.00889	0.0217
	(0.0341)	(0.0343)	(0.0343)
Teacher Master's Deg.	-0.00103	-0.00820	-0.0210
	(0.0233)	(0.0235)	(0.0236)
Teacher Experience (years)	-0.000324	0.00159	0.00500**
	(0.00195)	(0.00213)	(0.00222)
Constant	-0.0731	0.141	0.524
	(0.683)	(0.690)	(0.692)
Observations	5,643	5,643	5,643
R-squared	0.369	0.370	0.372

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

The estimates for endogenous peer effects, particularly those from the nonparametric control function specification, are very close to those estimated by Graham. Graham's approach, like many in the peer-effects literature, does not separately identify endogenous and exogenous peer effects. Instead, Graham estimates a composite parameter composed of endogenous and exogenous peer effects. Graham notes on page 646:

> Note that γ_0 may be a composite function of multiple "structural" parameters. In Manski (1993) it depends on the strength of what he terms "exogenous" and "endogenous" social effects. Distinguishing between these two types of social interactions is not a goal of this paper.

It is not immediately clear, though, what role exogenous peer effects would play in Graham's model, as no exogenous individual or group characteristics are included. Given the similarity in estimates, it is worth discussion, even if the two approaches are not fully compatible.

Graham estimates a peer effect coefficient that is a transformation of the coefficient estimate in the current model. Graham uses γ to represent the peer effect coefficient, while I use β. Comparing the two structural models provides the equivalence mapping $\beta = 1 - \frac{1}{\gamma}$. Graham ultimately estimates γ^2. To get a comparable estimate, I will transform Graham's estimates by $1 - \frac{1}{\sqrt{\gamma^2}}$. Graham presents two estimates of social interactions from the math equation: $\gamma_R^2 = 3.47$ from a restricted equation, and $\gamma_U^2 = 2.33$ from a more flexible specification. The comparable numbers for the current approach are $(1 - \frac{1}{\gamma_R}) = 0.46$ and $(1 - \frac{1}{\gamma_U}) = 0.34$. The results from the nonparametric control function estimation, $\beta_{NP} = 0.451$, are very close to Graham's restricted estimate. The results from the reading equation are similar. Graham's restricted estimate of γ^2 was 5.28, giving the transformed equivalent value of $(1 - \frac{1}{\gamma_R}) = 0.56$. The nonparametric control function from the reading equation is $\beta_{NP} = 0.598$. In both the math and reading equations, estimates using the semiparametric control function overstated the magnitude of endogenous peer effects, relative to Graham's estimates.

These estimates, though comparable with other estimates in the literature, may seem implausibly large. Interpreting the linear relationship between classroom average and individual math test scores suggest that a one unit change in a the classroom average score results in a 0.45 unit change in individual test scores. Almost 50% of exogenous changes in classroom average outcomes are passed on to individual outcomes. This interpretation does not take into account, though, that changes in classroom averages should carry more weight than changes in individual outcomes of similar magnitudes. In estimating the impact of a change in the classroom average score, the size of the class should be considered.

Graham incorporates this subtlety into his interpretation of the impacts of peer effects by altering the size of the change in classroom average outcome that he considers. Rather than considering the impact of a one unit change in mean classroom outcome, Graham considers the impact of one standard deviation change in the mean classroom outcome. The size of a one standard deviation change in the mean classroom outcome, however, depends on the size of the classroom. One standard deviation in the mean classroom outcome is smaller for large classrooms than for small classrooms. The standard deviation of classroom means is related to the standard deviation of individual means by the scaling factor of $\frac{1}{\sqrt{n_g}}$. Graham's implied impacts are described by a more reasonable range of 0.22 to 0.18.

This interpretation is appropriate in the context of Graham's estimator, which is focused on variance relationships. But it does not change the linear relationship between individual outcomes and classroom averages, which are both are measured in the same units. The linear relationship is apparent in the structural equation, where individual outcomes are a linear function of average classroom outcomes. Given this structure, there are more appropriate ways to incorporate class size into an intuitive understanding of the the endogenous peer-effect coefficient. Specifically, consider a one unit change in the math score of a single class member, and its effect on other class members. Class size varies between 12 and 28, so a one unit change in an individual student's math test score would have a first-order impact on the score of all other classroom members by between 0.04 and 0.02 units. Of course, this increase would have additional impacts through the magnification effect of endogenous peer effects. The total impact, $\frac{1}{n_g}\frac{\beta}{(1-\beta)}$ would be between 0.07 units for 12 student classes and 0.03 units for 28 student classes. All of these interpretations are based on the same (or very similar) coefficient estimates. The differences are only in interpretation. This discussion shows that the magnitude of endogenous peer effects is better understood when considered in the context of reasonable changes in classroom average outcomes.

Exogenous Peer Effects

Exogenous peer effects refer to the impact of average peer group characteristics on individual outcomes. In this model, the exogenous peer effects are represented by the coefficient estimates on the classroom averages of age in quarters and the indicator variables for low socioeconomic status, female, and black.

The OLS estimates of exogenous peer effects strongly suggest the presence of identification problems. In every case, the estimated coefficient on classroom average characteristics is close to -1 times the estimated coefficient on the corresponding individual observable. This relationship is especially apparent in the simulation results in table 1.1. As described in section 1.3.3 below, I consider the coefficient estimates on individual characteristics to be unbiased, and interpret this unlikely relationship as evidence of bias in the coefficient estimates on classroom averages.

Introduction of the control function appears to reduce the bias in estimates of exogenous peer effects in all cases except the coefficient on classroom race, defined as the classroom proportion of black students. This is likely due to the highly segregated nature of classroom composition. Approximately 55% of classrooms are entirely black or entirely white.[3] It may not be possible to identify the effect of classroom race composition, due to the high degree of overlap between the individual variable and the classroom average variable.

Individual and Teacher Characteristics

As predicted by the econometric model, the coefficients on individual observables are identified in every specification in which they appear. The coefficient estimates are stable across the OLS specification, the controlled regression, and the within regression. The estimates from the within regression have been purged of group-level influence, including endogenous and exogenous peer effects and classroom unobservables. The within regression provides the strongest evidence that the estimated coefficients on individual observables are in fact unbiased, rather than simply equally biased across specifications.

In the both of the semiparametric control function specifications, the estimated coefficients on teacher characteristics were not found to be statistically significantly different from zero. This is consistent with other research in the field, which finds that observable teacher characteristics have little impact on student outcomes (Rockoff, 2004; Aaronson et al., 2007). The same research, however, finds large impacts from unobservable teacher characteristics. Rockoff (2004) finds significant impacts of teacher experience on student test scores after controlling for teacher unobservables.

[3] Less than 1% of students in this sample are identified as a race other than white or black.

While this evidence is consistent with the current estimates, it also raises questions about the treatment of unobservable teacher characteristics. The restrictions imposed on teacher unobservables may be increasing the correlation of student results within classrooms, creating upwardly biased estimates of endogenous peer effects.

In both of the nonparametric control function specifications, positive effects are estimated for teacher experience. Furthermore, the estimates of endogenous peer effects in these models are smaller than those in the semiparametric specification. This evidence suggests a different interpretation from the previous one. As discussed previously, correlated effects drive similarities in outcomes due to shared environmental characteristics, such as teacher quality. The lack of impact attributed to teacher observable characteristics combined with the larger estimates of endogenous peer effects suggest that the semiparametric specification may be producing biased estimates. The true effect of shared teacher characteristics is being misattributed as excess correlation in individual outcomes and increasing the estimates of endogenous peer effects. It is difficult to know the true meaning of these coefficients, as they are based on implausible assumptions on the distribution of unobservable teacher characteristics.

1.3.4 Robustness Checks

I conducted several specification tests to explore the dependence of estimation results on the identifying assumptions. For this exercise, I only consider endogenous peer effects. The full control function, again, is:

$$A(\mathbf{X})\, v_g = \frac{\frac{1}{(1-\beta)}\left[S_{\alpha g}^2 + \frac{1}{n_g}\left(S_{\epsilon gi}^2 + \delta S_{\epsilon gi}\sum_{j\neq i} S_{\epsilon gj}\right)\right]}{\frac{1}{(1-\beta)^2}\left[S_{\alpha g}^2 + \frac{1}{n_g^2}\sum_{i\in g}\left(S_{\epsilon gi}^2 + \delta S_{\epsilon gi}\sum_{j\neq i} S_{\epsilon gj}\right)\right]} v_g$$

The following two assumptions were made for identification:

$$S_{\alpha g}^{2} \equiv Var(\alpha_g|\mathbf{X}) = 0$$

$$\delta = \mathbb{E}\left[\epsilon^*_{gi}, \epsilon^*_{gj} \mid \mathbf{X}\right] = 0$$

Instead of restricting these parameters to zero, I restrict them to be constant, $S_{\alpha g}^{2} = \sigma_\alpha^2$ and $\delta_{gij} = \delta$. I then recalculate the control function and reestimate the primary equation, using a range of plausible values for σ_α^2 and δ. For both parameters, a natural lower-bound exists at zero. The choice of a plausible upper bound is less clear. Both of these parameters are components of group-level variance. So one potential upper bound is the estimated variance of the between residual, $\hat{v_g}$. This would provide an upper bound of 0.29 for the math equation, and 0.26 for the reading equation. A more conservative bound for both equations is 1, which is the variance of both standardized dependent variables. I use the more conservative upper bound, letting parameter values range between 0 and 1. On each graph, the variance of the group-level residual is marked on the x-axis with a vertical line, to show the less conservative upper bound.

These simulations cover a range of values of σ_α^2 and δ. I will present results for a range of trial values for σ_α^2, and only three fixed values of δ. Restricting one parameter is done in the interest of clarity. Delta is chosen as the fixed parameter, because restrictions on the correlation of error terms within groups is both more common in empirical work and more plausible theoretically.

Estimates of individual conditional variance are not effected by changing assumptions on $S_{\alpha g}^{2}$ and δ_{gij}. These components are purged from the data by the within transformation before the conditional variance functions are estimated. Thus there is no need to re-estimate conditional variance. Results are presented using both semi-parametric and nonparametric conditional variance estimates.

Considering first the math estimates, figure 1.3, the maximum estimate of β occurs at $\sigma_\alpha^2 = 0$ and $\delta = 0$, suggesting an upper bound on peer effects for any constant values of $S_{\alpha g}^2$ and δ_{gij}. Furthermore, estimates of β are well-behaved across the range of parameter values, with an apparent minimum close to zero.

Reading estimates, figure 1.4, also appear to provide an upper bound for β, but are not well behaved in general. Estimates are large and negative for a large range of test values. While it is theoretically possible for peer effects to be negative, they are generally thought to be positive in most settings. Graham, in fact, estimates γ^2, so cannot distinguish between negative and positive values of the original γ parameter. He reasonably suggests that social interactions are likely to be positive in the current application. The $\beta - \sigma_\alpha^2$ curves for each value of δ appear to be linear over this range. They are in fact convex, though they have much less curvature than the math estimate curves.

It is not clear why the reading estimates behave poorly. The Breusch-Pagan test statistic on the within reading residuals was higher than that for the within math residuals. Considering figure 1.2, it appears that there are more high-variance residuals, but they are concentrated in fewer individuals, relative to math residuals. These results are consistent with Graham's estimates of social interactions. His estimates using reading test scores were larger in magnitude, but less precise than estimates from math scores.

1.4 Conclusion

The linear-in-means specification, while consistent with microeconomic decision models, is essentially a reduced-form approach. There is no underlying behavioral model, and no primitive preference parameters are identified. As such, the estimation results

Figure 1.3: Math Estimates as Identifying Assumptions Change

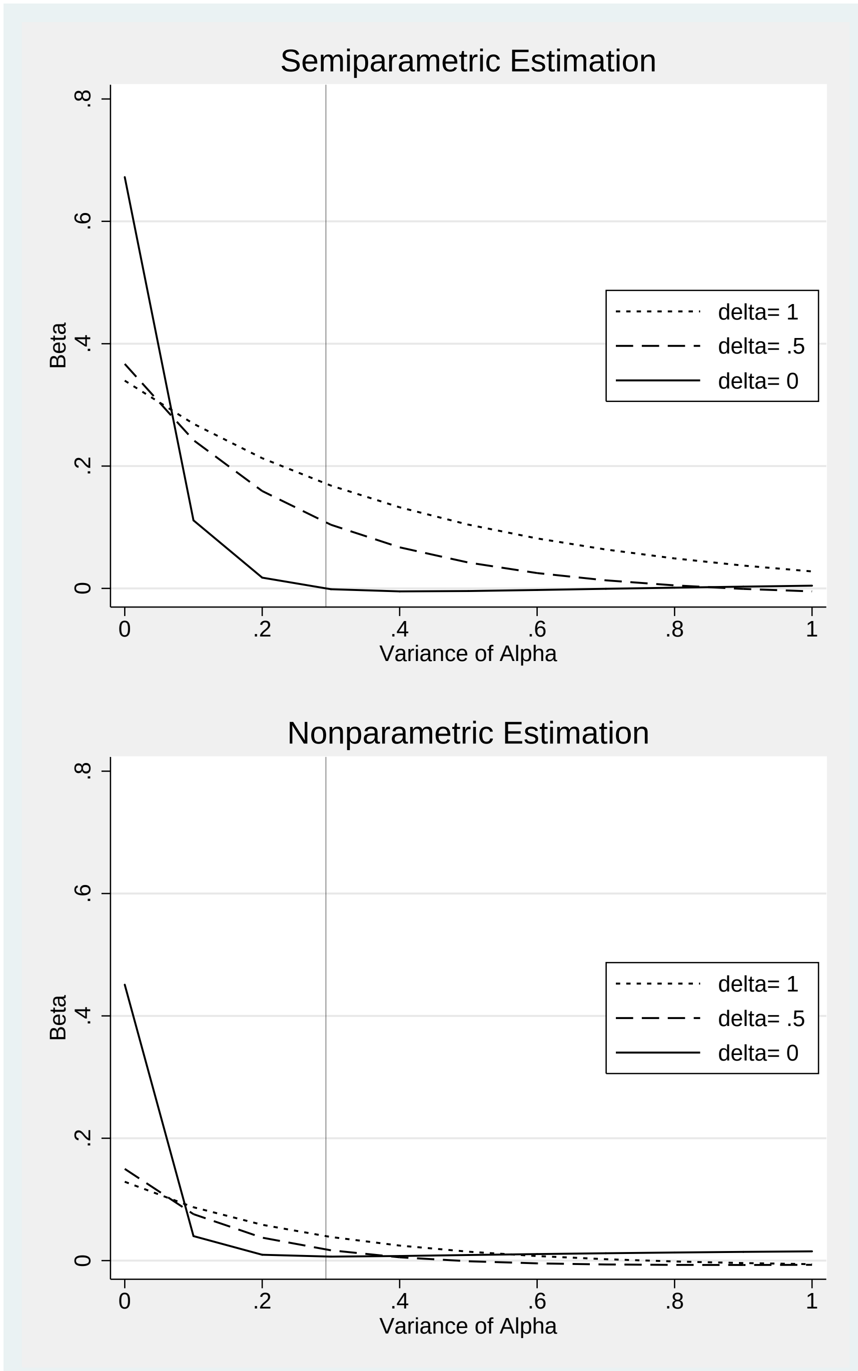
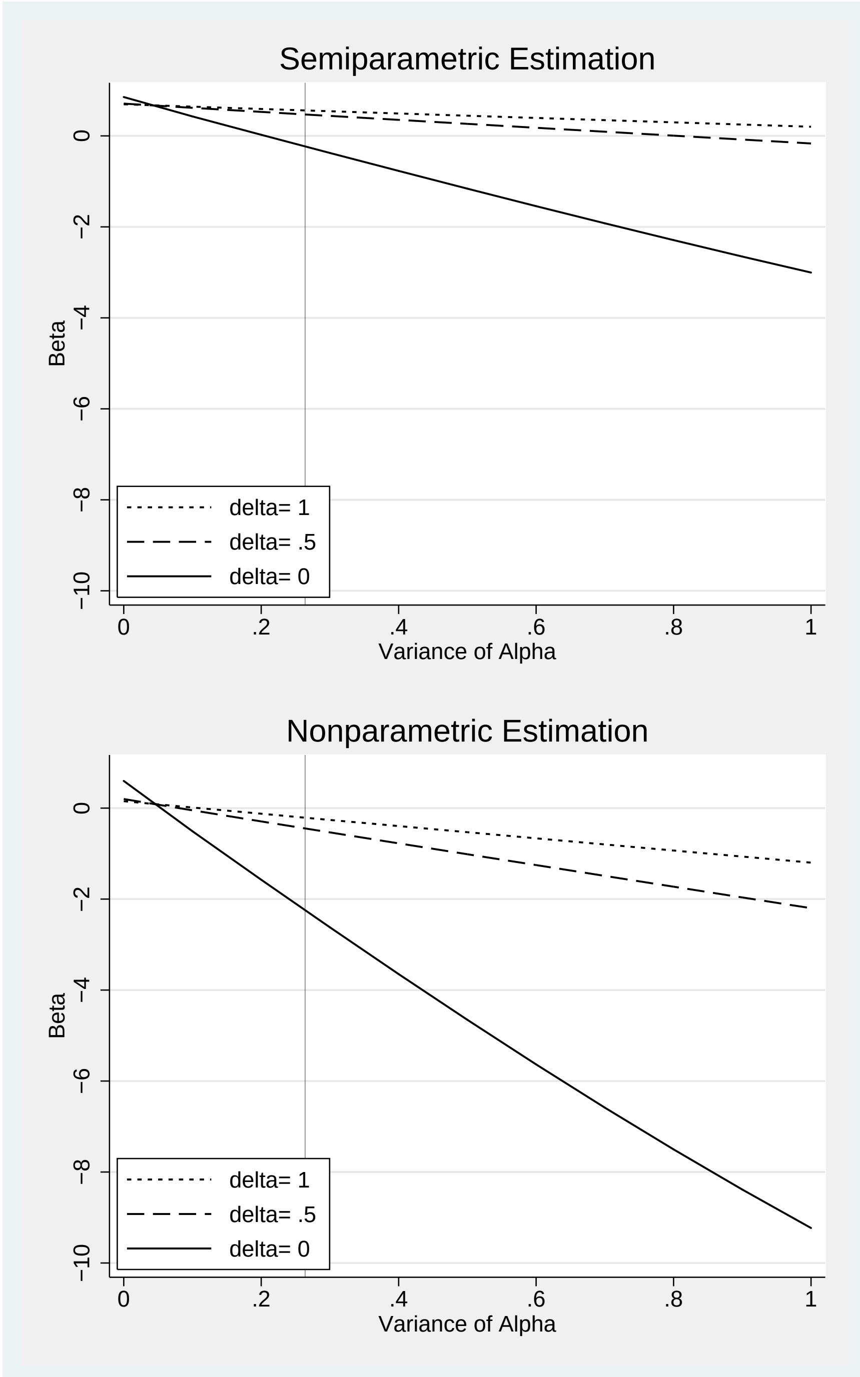


Figure 1.4: Reading Estimate as Identifying Assumptions Change



are informative on the nature of social interactions in a particular setting. This is consistent with the current understanding of peer effects. As noted by Epple and Romano (2010), "One would not expect the magnitude of peer effects to be the same in all classroom settings and all grade levels. . . . Even in seemingly similar academic settings, peer group influences can operate quite differently as evidenced by the contrast in findings for the US Military Academy (Lyle, 2007, 2009) and for the US Air Force Academy (Carrell, Fullerton, and West, 2009)."

The strength of the linear-in-means model, then, is not the generalizability of its results, but rather the generalizability of its application. The approach described here does not require an experimental setting or detailed information on network structure. When social interactions are relevant to a policy decision, such as whether to track students by ability, they can be estimated directly. The main weakness of this approach in its current form, is the required assumption of zero variance in group unobservables. This assumption is difficult to justify theoretically in the context of primary education, where peer groups are defined by classroom membership. The characteristics of a group include the teacher assigned to that classroom, so any variation in unobservable teacher characteristics would violate the assumption. Despite this burdensome assumption, the approach proposed here is worth pursuing.

There is some evidence that the required assumption is not overly restrictive in practice. The inclusion of group-level variables in SLS estimation may give the resulting conditional variance function flexibility to account for group-level variance. The estimates of endogenous peer effects generated in this manner are consistent with those generated by Graham, where identification in based on the randomized assignment of classroom treatments. Graham's approach still imposes some restrictions on classroom unobservables, but they are not as onerous as those required by the current approach. Specifically, Graham requires that the distribution of teacher effectiveness is equal for small and large classrooms, up to an additive shift.

Other applications exist where the identifying assumptions are not difficult to justify theoretically. Restricting the variance of group level unobservables to zero could be justified in any model where the fixed environmental characteristics are not specific to each group. In other words, where group and environmental unobservables do not overlap perfectly. Group-specific unobservables, then, are only aggregates of individual unobservables, so the required identifying assumption would not be violated by variation in other group-specific unobservables. A model of this type would require each peer-influence group to be defined by something other than classroom membership.

A model of small group interactions could fulfill these requirements. For example, Sacerdote (2000) studies peer effects between roommate pairs. Other examples may be found in the context of secondary and post-secondary education, where students are not taught by a single teacher. In the United States, students in the later years of primary school frequently have a primary home-room teacher and subject-specific teachers for certain specific subjects, such as Math and English. In this setting, peer effects in subject-specific classrooms could be evaluated using peer-influence groups defined by home-room classroom membership. Carrell, Fullerton, and West analyze peer effects in a similar setting in the US Air Force Academy, where peer-influence groups are defined by squadrons, rather than classrooms. Applications also exist outside of the educational context. A potential application from the field of development economics comes from microfinance borrowing groups, where groups do not interact in unique, fixed environments.

Alternatively, it may be possible to relax the identifying assumptions by using richer data. For example, classroom level data with repeated observations on teacher and classroom outcomes would allow teacher and classroom unobservables to be identified. These components could either be purged from the model or incorporated explicitly into the control function. The difficulty in this approach is finding data where both teachers and students are effectively randomly assigned to classrooms.

A second approach to relaxing the identifying assumptions could be pursued through a simultaneous estimation procedure. It may be possible extend the current approach to allow variance in group-level unobservables and/or within-group correlation. In this case, group-level components would be estimated simultaneously in the primary (second-stage) regression. This is the approach taken in the classic KV estimator. The conditional variance function of the primary equation error term is estimated simultaneously with the coefficients from the primary equation. In the current application, this approach would require significant group-level heteroscedasticity. This approach has an additional payoff, in that it has the potential to identify group effects caused by correlated unobservables. Exploration of this approach is left for future research.

Chapter 2

Cash vs Food as Educational Conditional Transfer with Deon Filmer and Jamele Rigolini of the World Bank

2.1 Introduction

> From better health to increased wealth, education is the catalyst of a better future for millions of children, youth and adults. No country has ever climbed the socioeconomic development ladder without steady investments in education.
>
> Irina Bokova, Director-General of UNESCO (UNICEF, 2010)

Education has long been seen as a cornerstone of both economic and human development. A clear consensus exists within the international community on the importance of education, as reflected in the second Millennium Development Goal, which calls for universal primary education (UN, 2014b). International initiatives such as the Global Partnership for Education and the Education for All Initiative advocate investing in education as a highly effective means of poverty reduction, which is critical to both economic and human development (UN, 2014a; EFA, 2014).

Investment in education has positive returns both at the national and individual level. Since the late 1950s, a large body of literature has debated and estimated the individual returns to investment in education (see for instance Griliches, 1970; Mincer, 1974; Rauch, 1993; Heckman and Klenow, 1997; Acemoglu and Angrist, 2000; Card, 2001; Bils and Klenow, 2000). In their review of this work, Psacharopoulos and Patrinos (2004) conclude that "it is established beyond any reasonable doubt that there are tangible and measurable returns to investment in education." Their compilation of available estimates across 98 countries using a range of methodologies finds that the average rate of return to an additional year of schooling is 10 percent, with higher returns in low and middle-income countries and in Latin America and Sub-Saharan Africa (Psacharopoulos and Patrinos, 2004).

Given these high potential returns, many households in developing countries appear to underinvest in education. This is a problem of considerable scale: the Global Initiative on Out of School Children estimates that a total of 58 million children between the ages of 6 and 11 are not in school, and finds almost no progress in reducing this number since 2007 (GMR and UIS, 2014). Many explanations have been proposed to explain this economic puzzle, including credit constraints and lack of information on potential returns.

Demand-side educational interventions aim to reduce educational underinvestment by providing immediate incentives for educational investment. Most frequently, incentives take the form of food or cash transfers that are made conditional on school participation. Conditional transfer programs provide short-term returns to investments in education. Immediate returns may motivate an activity, in this case school attendance, that would not otherwise be considered productive. Alternatively, the transfers may be used to defray costs that could be preventing educational participation, including the opportunity cost of the student's time. Whether framed as an increase in the benefits of education or a cost-reduction intervention, the end result is the same. Conditional transfers increase the attractiveness of educational participation relative to other potential activities.

This paper will analyze the results of a pilot program evaluating the impacts of cash and food transfers in the context of an educational conditional transfer program.

Specifically, the paper will focus on the educational impacts of the program, which is the focus of the program's conditionality. The pilot program was conducted in the context of an existing Food for Education (FFE) program in Cambodia operated jointly by the World Food Programme (WFP) and the Cambodian Ministry of Education, Youth, and Sport (MoEYS). The results of this evaluation are relevant on many levels. Locally, results will inform the formation of educational policy in Cambodia. Data collected on both educational participation and learning outcomes will provide valuable insight on the performance of the existing primary school scholarship program. Internationally, the results will influence the future direction of WFP school feeding programs in other low-income countries. As of 2009, WFP supported school feeding programs in 70 countries, reaching approximately 22 million children (Bundy et al., 2009). Even more broadly, this evaluation will contribute to the growing literature on the design aspects of conditional transfer programs. Although other randomized trials have looked separately at the role of cash transfers and of food transfers on education, no robust comparative analysis of the two transfer modalities on educational outcomes has been carried out to date. This evaluation, therefore, is the first to provide sound evidence on the relative performance of food and cash in the context of a educational conditional transfer program. Results indicate that, while both transfer types increase school participation, transfers in cash are more effective than equivalently valued food transfers.

A change to cash assistance has the potential to reduce the logistical burden of in-kind transfers without sacrificing educational objectives. This evaluation is motivated by the fact that the actual impact of conditional transfers on educational outcomes will depend on choices made within beneficiary households, which cannot be reliably predicted, as well as local conditions including school infrastructure, quality of education, physical access, and food availability. The impacts of both food and cash assistance must, therefore, be evaluated jointly before undertaking any large-scale or permanent change in school feeding programs.

2.1.1 Literature Review

The most prevalent type of conditional transfer program distributes food to students through the education system. These programs, known collectively as school feeding programs or food for education (FFE) programs, include subsidized school meals, in-school meal programs, and take-home rations. FFEs are attractive development programs, as they have the potential to address both educational and nutritional objectives by tying food supplementation to regular school attendance. Traditional in-school meal programs have an implicit conditionality, whereby students do not receive a meal if they do not attend school. Benefits are distributed daily, conditional on attendance. Food may also be distributed to students in the form of take-home rations. In the take-home ration format, benefits are distributed less frequently based on an explicitly defined condition on attendance. By distributing transfers in food, school feeding programs can target objectives such as increased school enrollment and attendance, reduced student drop out, and improved educational outcomes, as well as increased household food security. The versatility of school feeding programs has contributed to their widespread adoption in a wide range of policy and economic settings. As noted in a joint publication from the World Bank and the World Food Programme (WFP), there are few countries that do not have a school feeding program (Bundy et al., 2009).

Despite their wide-spread adoption, school feeding programs have been criticized as being less cost effective than alternative programs that focus directly on either educational or nutritional outcomes. According to these criticisms, the multiple objectives of school feeding programs lead to high logistical costs, driven largely by the necessity of large-scale food distribution (Kakwani et al., 2005). A change to cash transfers has the potential to reduce the logistical burden of direct food distribution. This is precisely the approach taken by CCT programs. Direct cash transfers were pioneered in 1997 with the establishment of PROGRESA in Mexico. Since that time, CCT programs have been adopted throughout Latin America, as well as in several countries in Africa and Southeast Asia (Schady and Fiszbein, 2009).

Previous evaluations of CCTs and FFEs have demonstrated that both can be effective at increasing school participation. CCT programs were frequently introduced with some element of randomization, enabling robust estimates of program impacts. There is a large body evidence on program impacts in a variety of environments. Evaluations of several CCT programs in Latin America found that CCTs generally had positive impacts on school enrollment and attendance, especially in settings with low initial enrollment (Schady and Fiszbein, 2009; Shari et al., 2013). In Cambodia, scholarship programs at the secondary school level have been shown to have large impacts on school participation (Filmer and Schady, 2006, 2009).

A review of the evidence available on FFE programs in 2008 found a large body of literature, but few high-quality experimental evaluations with the capacity to produce reliable, causal estimates of program effects (Adelman et al., 2008). A recent evaluation of the WFP school feeding program in Cambodia, for example, used quasi-experimental methods, comparing changes between program and non-program schools over time. The evaluation found that both school meals and food scholarships increased enrollment and decreased dropout. Schools that participated in the school feeding program between 2002 and 2009 saw their enrollment over the entire period increase by 6.1% more than schools that were not included in the program. In grades 2-5, drop-out rates were estimated to decrease by 1 to 3 percentage points as a result of the program. Causal interpretation of these estimates is questionable, though, as program schools were not randomly selected. Program schools were selected according to a specific set of criteria, so are not directly comparable to non-program schools (Nielsen et al., 2010).

The majority of experimental evidence on FFE programs comes from a series of pilot programs undertaken by the World Food Programme. These pilot programs evaluated the relative performance of in-school meals and take-home rations in FFE programs in Burkina Faso, Lao PDR, and Uganda. In Burkina Faso, take-home rations increased school enrollment and improved the nutritional status of young children (Alderman et al., 2009). In Uganda, the FFE program did not increase enrollment, but did increase school attendance (Alderman et al., 2010). No impacts were found in the Lao PDR program, but this is likely due to constraints in the design of the pilot program, which allowed for randomization at the district level only and drew the control sample from a different province (Alderman et al., 2011).

Within each type of program, research has been conducted to evaluate program design characteristics, with the ultimate goal of fulfilling program objectives in the most cost-effective manner. The previously mentioned WFP programs, for example, were designed to compare the relative cost-effectiveness and performance of take-home rations and in-school meals. CCT programs, similarly, have experimented to identify optimal targeting criteria, conditionality parameters, and transfer amount (Baird et al., 2014; Schady and Fiszbein, 2009; Filmer and Schady, 2011). CCT and FFE programs share many of the same design characteristics, including targeting criteria, conditionality parameters, and transfer value. The only irreconcilable difference between the two program types is the modality of the transfer, food or cash. Despite their similarities, these two types of conditional transfer program have remained clearly divided by their respective transfer types.

Previous evaluations of food and cash transfers have been undertaken, but only in the context of unconditional transfer programs. Programs evaluated in this literature focus directly on the provision of a basic social safety net without conditions on recipient behaviors. These tend to look at marginal propensity to consume food, the "stickiness" of food assistance, and, occasionally, cost of delivery. Two evaluations conducted through United States Food Stamp Program in the 1980s replaced traditional food stamps with direct cash payments. In Puerto Rico, the evaluation found that the value of food consumption did not change in response to cash payments (Fraker et al., 1986; Moffitt, 1989). Additional trial programs were conducted in Alabama, Washington, and California. Evaluation of these trials found that the change to cash payments reduced the value of household food consumption, and that the impact was greater when the transfer amount was larger in proportion to the household food budget (Fraker et al., 1995).

Several evaluations conducted in low and middle income countries compared both the quantity and value of food consumption under each transfer regime. Specifically, the program evaluations took place in Mexico (Skoufias et al., 2008; Leroy et al., 2010), Sri Lanka (Mohiddin et al., 2007), and Ecuador (Hidrobo et al., 2014). Results from these studies were broadly similar. Both treatments increased the value of food consumption by similar amounts, but had different impacts on the types of food consumed. Cash recipients consumed fewer calories than food households, but had a more diverse diet.

These studies provide some guidance about the possible response of household consumption to food or cash transfers, but do not inform the relative performance of each transfer type as an incentive for educational investment. To the best of my knowledge, there is only one completed study comparing the impacts and costs of cash and in-kind assistance in the context of an educational scholarship program.

The study took advantage of a policy change in Bangladesh, where a school feeding program was completely replaced by a conditional cash transfer program. Both types of transfer were found to have similar, positive effects on school enrollment. The cash transfer, however, was not found to have any effect on food consumption. Instead, parents reported using the cash transfer to meet educational and other household expenses (Ahmed, 2005).

This result is not consistent with the previous evidence drawn from the comparison of unconditional cash and food transfers, which found that both transfer types resulted in similar changes to household food consumption. This contradictory result may simply be a result of the limitations of the study design. The survey did not collect full information on household expenditure, but rather asked specifically about the money received in the transfer. It is not clear if the result reflects an actual increase in educational expenditure, or if the same expenditure would have been made from different sources. Another source of uncertainty comes from quasi-experimental design of the evaluation, which compares the two transfer types across distinct time periods. This design does not allow for the separate identification of the causal effects of the program from contextual effects that may have changed between the two time periods.

Assuming that the result is reliable, it suggests that framing effects may influence the use of cash transfers received in the context of a conditional transfer program. Arguments for the use of cash transfers over food transfers have frequently been based on cost-effectiveness. But it may be possible that cash transfers may actually result in greater educational investments than food transfers. This result, along with its questionable reliability, illustrate the importance of rigorous quantitative evaluation when considering the choice of food or cash payments in the context of conditional transfer programs.

2.1.2 Expected Outcomes

Economic theory suggests that the impact of an in-kind transfer that is inframarginal with respect to the existing household budget will be indistinguishable from the impact of a equivalently-valued cash transfer. Given an inframarginal in-kind transfer, the household will simply forgo planned purchases of that good, freeing up an equivalent amount of cash in the household budget. The in-kind transfer considered here consists of 10 kg of rice per month, with a cash value of 5 USD. This transfer is inframarginal to almost all households in the evaluation survey. Over 99% of households consumed more than 5 USD of rice per month according to the baseline survey. The food scholarship can also be compared to the average household rice consumption of all Cambodian households. The 2008 Comprehensive Food Security and Vulnerability Analysis (CFSVA) estimated monthly rice consumption at 62 kg per household. The food scholarship, then, represents 16% of average monthly household rice consumption (Santacroce, 2008).

The potential for differential impacts from food and cash transfers exists in the current context, despite the relatively small size of the rice transfer relative to baseline rice consumption. This is due to the nature of household production and consumption in rural Cambodia, where a significant portion of the food consumed is not purchased with cash, but is derived from household production. Amongst poor, rural households, the average value of monthly food consumption is greater than average monthly income. According to the 2010 Cambodia Socio-Economic Survey (CSES), average monthly disposable income for the poorest, rural quintile, was approximately 30 USD per month, while the average value of monthly food consumption was 73 USD per month (CSE, 2013). Households that rely on subsistence farming may not be able to substitute in-kind transfers for cash within the household by altering their purchasing behavior. Instead, the conversion of food to cash requires a market transaction with an agent outside the household. The resulting transaction costs should lower the value of an in-kind transfer relative to a cash transfer.

If food and cash transfers are not perfectly substitutable within the household, the potential exists for differential impacts. The current consensus on cash vs. in-kind assistance is that, while cash transfers are generally preferred by beneficiaries, in-kind assistance results in superior outcomes in regard to specific, measurable objectives. This conclusion is supported by evidence drawn from evaluations comparing food and cash transfers in the context of nutritional programs. Food transfers have resulted in larger changes in nutritional indicators. Cash transfers provide households flexibility to address their needs, including, but not limited to, nutrition (Hidrobo et al., 2014).

Of course, in-kind transfers will produce larger impacts only when the type of good can be targeted to the outcome of interest. The larger impacts of food transfers can be attributed to the fact that food is a direct input for nutritional outcomes. When considering changes in educational investment and learning outcomes, there is no reason to believe that food transfers will be superior. In fact, there are several reasons why cash transfers may result in superior educational outcomes.

First, cash transfers are generally preferred to equivalently valued in-kind transfers. Cash gives households the flexibility to fulfill their most pressing needs, be they nutritional, medical, or recreational. Cash transfers will provide a greater incentives to meet the conditions of the program. Second, if children are working for wages, cash transfers may be a more effective in offsetting the opportunity cost of continued education. As discussed previously, in an environment of subsistence farming, a food transfer does not necessarily increase the amount of cash available in the household budget. Cash transfers may offset some of the foregone cash wages that a child could earn through wage labor instead of school attendance.

Third, the requirements for improving learning outcomes are complex. There is no single input that can improve educational outcomes as reliably as food can improve nutritional outcomes. Cash transfers can enable a household to provide the diverse resources required for successful school participation. These may include school fees, books, transportation, and uniforms. Finally, framing effects may encourage recipients to invest the transfer in education. Even if a household were able to costlessly convert a food transfer to cash, it is unlikely that any framing effects would transfer to these new funds.

These potential causes of differential impacts between food and cash transfers can be reduced to two underlying causes. The first two reasons are driven by the way a household values each transfer type. Differences in valuation, whatever the cause, are likely to induce different impacts on whatever outcomes are linked to the conditionality of the program. A more highly valued transfer will induce more eligible households to satisfy the conditions for receipt. The third and fourth reasons are related to the way a household spends each transfer type. Differences in disposal have the potential to cause different outcomes that are not directly linked to the conditionality of the program, including educational expenditure and learning outcomes.

2.2 Program Design

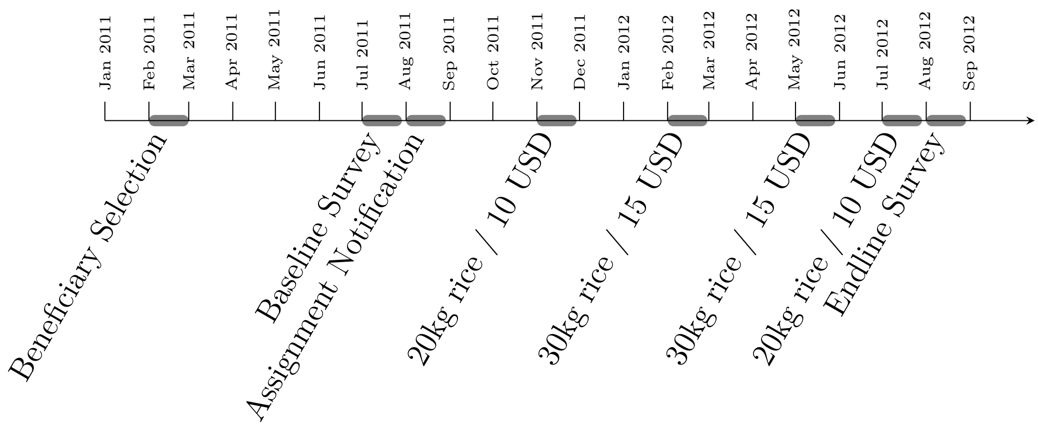
Cambodia has made rapid progress in expanding primary education. According to the 2012 Education for All Global Monitoring Report, the net primary school enrollment rate was 96% in 2010, up from 87% in 1999. This improvement coincides with the establishment of a school feeding program for primary school students. In partnership with the Ministry of Education, Youth, and Sport (MoEYS), the World Food Programme launched its first school feeding program in 2001.

The pilot program evaluated here was implemented as an extension of the existing WFP school feeding program, which distributes take-home rations. The pilot was designed with dual goals to expand the existing FFE program and to explore the possibility of primary school scholarships in the form of direct cash transfers. Cash or food payments were made directly to families, conditional on their child's continued attendance and enrollment. In considering both food and cash transfers, this pilot program bridged the divide between FFE and CCT programs.

Eligible students were given a food or cash transfer or scholarship, conditional on school enrollment and an attendance rate of at least 80%. The food scholarship was unchanged: a take-home ration, consisting of 10 kg of unfortified rice per month. The cash scholarship of 5 USD per month was chosen to match the value of the existing food transfer. While the quantities of rice and cash were chosen to be of roughly equal value, the price of rice fluctuated throughout the evaluation period, so the two transfers were never precisely equivalent. Over the course of the year, the total value of the rice scholarship was less than 10% more than the value of the cash scholarship. The estimate of average monthly household food consumption from the baseline survey was 122 USD. A value transfer of 5 USD, then, represents about 4.1% of monthly food consumption.

Apart from the unavoidable differences in valuation, the two transfer types were designed to be equivalent in all other aspects, including beneficiary targeting and scholarship distribution. Distributions for both types of scholarship followed the same process and schedule. Both scholarships accrued monthly, but were distributed at 2 or 3 month intervals for the duration of the 10 month school year. A timeline of the distribution schedule can be found in figure 2.1. Scholarships were distributed to school clusters, consisting of 1-3 schools, one of which served as the distribution point for all schools in the cluster.

Figure 2.1: Evaluation Timeline



Eligibility for the program was determined in a two-step process, targeting first schools, then students. Targeting followed the same procedure used by WFP for its existing school feeding programs. WFP selects the schools where it initiates school feeding programs according to three main criteria. Schools are more likely to be selected if they serve poor and food insecure communities, have poor education performance, or are not located close to any main road. School targeting is not carried out regularly. Once a program exists in a school, it is permanent. New schools are identified only when the school feeding program is expanded.

Beneficiary targeting is carried out each academic year. Beneficiaries are selected from students in grades 4 - 6, from poor and disadvantaged households. A household may be classified as poor or disadvantaged through one of two procedures. First, a household may be identified by the Identification of Poor Households Programme (ID-Poor). This government program uses a standard procedure to identify poor households on a national level. Identified households are issued an ID Poor card, which is used as a basis for targeting by both governmental and non-governmental social protection agencies.

The second procedure uses an alternative set of WFP targeting criteria to establish eligibility. The criteria were designed to target households that are raising orphans, have chronically ill or disabled household heads, own no farm land, and that have many young children. Students were selected through a participatory process in which students and teachers compiled a list of potential participants which was then reviewed by a committee of school officials, teachers, and community leaders. The proposed student list from the school and community was then checked by the provincial and district education officials before final confirmation. School and beneficiary targeting was carried out for all schools prior to cash or food assignment (WFP, 2011).

2.2.1 Study Design

Schools selected for the pilot program included schools where WFP already had an existing school feeding program as well as new schools, where were no school feeding program was active. The existing school sample was composed of all schools that had an existing take-home ration program, but no other FFE component, such as in-school meals. As discussed later, the analysis does not rely on the direct comparison of outcomes between new and existing schools.

Randomization into assignment groups was conducted at the level of school clusters, stratified by province. Each school cluster contains 1-3 schools, with one of the schools serving as the common scholarship distribution point for the cluster. Randomization was carried out separately for existing program schools and new program schools. Schools in the new sample were randomly assigned to one of three treatment groups, food transfer, cash transfer, or control. Existing program schools were

Table 2.1: Baseline Sample

	Program Status		
	Existing	New	Total
Clusters			
Ctrl	0	41	41
Food	53	23	76
Cash	46	26	72
Total	99	90	189
Schools			
Ctrl	0	67	67
Food	136	37	173
Cash	134	44	178
Total	270	148	418
Households			
Ctrl	0	690	690
Food	1,318	390	1,708
Cash	1,303	456	1,759
Total	2,621	1,536	4,157

assigned to one of two treatment groups, food transfer or cash transfer. For existing program schools, the changes required for the pilot program were minimal. Those in the food treatment group simply continued the operation of their existing program, while those in the cash treatment group changed only the modality of the transfer. The program design did not allow for a control group for existing program schools, so the results from existing schools will only be used to consider the relative performance of cash and food transfers.

The allocation of these households, clusters, and schools between new and existing program status and treatment assignment can be seen in table 2.1. In total, 418 schools took place in the evaluation. A sample of 10 students was drawn from each school for the baseline survey, which was conducted in July and August of 2011. Students were drawn from the beneficiary list of each school, stratified by grade. The total baseline sample was 4,157 households. Endline data collection for this evaluation took place in July and August of 2012. The endline survey completed follow-up interviews with 3,845 of the original beneficiary households.

The primary respondent of the interview was generally the beneficiary's primary caregiver. One module of the survey instrument, however, was administered directly to the beneficiary student. This module consisted of three cognitive tests, described below, intended to measure learning outcomes. In some cases, although the household interview was completed, the beneficiary was not present. The sample for the cognitive testing module, then, is smaller than that of the full household sample.

2.2.2 Indicators

The first immediate goal of the program was to increase access to education. We first looked at several indicators related to the conditionality of the program. The first condition of the program is enrollment in school. Due to the time frame of this evaluation, however, it was not possible to measure the program's impact on beneficiary enrollment. Enrollment is established at the beginning of the school year, and is required for students to be eligible for the program. All beneficiaries, then, are enrolled initially. The time frame for the evaluation was one academic year, so data was not collected on beneficiary re-enrollment the following year.

The relevant condition for beneficiary access to education, then, is attendance. Beneficiaries must maintain a monthly enrollment rate of 80% for continued eligibility. A beneficiary could fail to meet the attendance requirement if he or she has positive but insufficient attendance, or if he or she drops out completely. In our sample, weekly attendance was well above the required minimum level for students who have not completely stopped attending classes. This analysis, therefore, will focus on student dropout, though an indicator for average weekly attendance is also analyzed. A beneficiary is assumed to have dropped out of school if he or she did not attend school for a single day in June, the final month of the school year.

In addition to dropout and attendance, we consider two indicators related to participation in primary school that are not tied to scholarship receipt. The first is the number of days per week when any school work was done at home. The second indicator is annual educational expenditures, which includes expenditures on books, school supplies, school fees, transportation to and from school, and gifts or donations to the school. These two indicators reflect educational investments in time and money beyond those which are required by the conditionality of the scholarship program.

The ultimate goal of education is increased cognitive performance and knowledge. The cash and food scholarship program has the potential to impact learning outcomes directly through the conditionality of the program, which was designed to increase access to education. The potential also exists for the scholarship to impact cognitive performance through its use as a generic value transfer. The increased wealth of the household could enable increased investments in the educational and nutritional status of primary aged-children. We look for impacts on these educational outcomes using an array of cognitive tests. Three tests were administered to the beneficiary child: digit span, Raven's Progressive Matrices, and a local standardized math test.

The digit span test measures working memory. To administer the test, the enumerator reads a short string of digits to the respondent, who is asked to repeat the numbers. Upon each successful repetition, the test continues with a longer string of digits, until the respondent either reaches the end of the test or can no longer remember the digits. The test is repeated a second time with backwards recall, where the respondent is asked to repeat the string of digits in reverse order. Scores on the digit span test range from 0-16.

Raven's Progressive Matrices is designed to measure general cognitive ability. The Raven's test is non-verbal, and based on visual pattern matching. We administered a subset of the full Raven's test, consisting of the 12 questions of set A and the first 6 questions from set B. The range of possible scores for this test is from 0-18. The questions are multiple choices with six possible responses, so a score based on purely random guessing should be 3.

The math exam is an excerpt from a Cambodian standardized math test for grade 6. This is the only test that directly tests material that is part of the Cambodian primary school curriculum. The test consists of 13 multiple choice questions, each with 4 response choices. A score based on random guessing for this test should be 3.25.

2.3 Data

2.3.1 Randomization

The baseline characteristics of beneficiary households are compared across treatment groups in table 2.2. Households are compared across randomly assigned treatment categories, but only within categories defined by the status of their school feeding program at baseline. Households from existing program schools are compared to other households from existing program schools only. The same is true for households from new program schools, who were not covered by any school feeding program at baseline.

The first three columns compare the characteristics of beneficiary households from existing program schools. Columns (1) and (2) present average values from the food treatment group and the cash treatment group, respectively. Column (3) gives the F-test statistic from testing the null hypothesis that the two assignment groups share the same mean. The p-value from the same test is in parentheses. Two differences are

Table 2.2: Baseline Characteristics By Treatment Group

	Existing Schools			New Schools			
Variable	Food	Cash	Fstat	Ctrl	Food	Cash	Fstat
Num. Adults	2.69	2.64	0.503	2.80	2.74	2.68	1.039
	(0.05)	(0.05)	(0.480)	(0.05)	(0.06)	(0.06)	(0.358)
Num. Children	3.07	3.06	0.007	3.35	3.38	3.42	0.244
	(0.08)	(0.08)	(0.932)	(0.07)	(0.09)	(0.08)	(0.784)
HoH Female	0.34	0.31	1.172	0.38	0.32	0.34	1.547
	(0.02)	(0.02)	(0.282)	(0.03)	(0.03)	(0.03)	(0.218)
Expenditure PC	396.60	388.82	0.549	444.61	423.58	423.87	0.679
	(7.93)	(6.95)	(0.460)	(15.34)	(13.97)	(13.11)	(0.510)
Food Exp. PC	248.82	246.04	0.148	270.76	256.58	266.36	1.121
	(5.20)	(5.07)	(0.702)	(6.61)	(7.10)	(8.52)	(0.331)
Own Ag. Land	0.57	0.68	4.945**	0.66	0.68	0.69	0.292
	(0.04)	(0.04)	(0.028)	(0.03)	(0.05)	(0.04)	(0.748)
Own Village Land	0.83	0.90	5.932**	0.82	0.81	0.85	0.754
	(0.02)	(0.01)	(0.017)	(0.02)	(0.04)	(0.02)	(0.473)
Age in years	11.53	11.57	0.093	12.11	11.76	11.70	5.259***
	(0.09)	(0.09)	(0.761)	(0.10)	(0.17)	(0.09)	(0.007)
Female	0.56	0.55	0.040	0.57	0.62	0.56	1.875
	(0.02)	(0.01)	(0.842)	(0.02)	(0.02)	(0.03)	(0.159)
N	1318	1303	.	690	390	456	.

[1] Significance levels: * < 10% ** < 5% *** < 1%
[2] Standard errors in parentheses

apparent between food and cash groups from existing schools. Households that were assigned to the cash treatment are significantly more likely to own both residential and agricultural land, relative to households in the food scholarship group.

The group mean characteristics of households from new schools are presented in columns (4)-(6). Column (7) gives the F-statistic and p-value from testing the null hypothesis that all three groups share the same mean. The null hypothesis is rejected at the 10% level for only one variable. Beneficiaries from control households were roughly 4-5 months older than beneficiaries from food or cash households.

As expected, there are some differences between treatment groups, but this does not imply that the randomization was compromised. Random assignment of school clusters was conducted in Washington D.C. by the principal investigators of this evaluation. Schools were not notified of their assignment status until after baseline data collection was completed. There was no obvious opportunity for selection based on knowledge of future treatment. Instead these differences are treated as statistical outliers.

2.3.2 Attrition

Of the initial baseline analysis sample of 4,157 households, 3,845 were located and re-interviewed, for 7.5% attrition rate. While the attrition rate is relatively low, attrition does not seem to be random. Table 2.3 compares several key household and beneficiary characteristics between attrited and non-attrited households. On average, attrited households had fewer adult members, with an average of 2.58 adults versus 2.7 adults for households that participated in the endline survey. Attrited households also reported lower annual expenditure per capita by approximately 21 USD per year. These differences are significant at the 10% level. Attrited households were less likely to own land, both residential and agricultural. Finally, beneficiaries from attrited households were, on average, 3 months older than those from non-attrited households. Many households in Cambodia migrate for work, particularly from locations near the Thai border. This explanation is consistent with the characteristics of attrited households, which tend to have smaller families, older children, and are less tied to land holdings.

Table 2.3: Baseline Characteristics by Attrition Status

Variable	InSample	Attrited	Diff
Num. Adults	2.70	2.59	-0.112^{*}
	(0.03)	(0.06)	(0.067)
Num. Children	3.18	3.20	0.017
	(0.04)	(0.10)	(0.094)
HoH Female	0.34	0.34	0.003
	(0.01)	(0.03)	(0.027)
Expenditure PC	409.21	388.49	-20.716^{*}
	(4.97)	(12.29)	(12.565)
Food Exp. PC	253.96	257.75	3.794
	(2.87)	(7.98)	(7.866)
Own Ag. Land	0.65	0.53	-0.125^{***}
	(0.02)	(0.04)	(0.037)
Own Village Land	0.86	0.74	-0.122^{***}
	(0.01)	(0.04)	(0.034)
Age in years	11.66	11.90	0.238^{**}
	(0.05)	(0.12)	(0.121)
Female	0.56	0.60	0.041
	(0.01)	(0.03)	(0.031)
N	3845	312	4157

[1] Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$
[2] Standard errors in parentheses

Non-random attrition will not bias the estimates of program impact, as long as attrition is not related to treatment assignment. Analysis of the attrition rates and baseline characteristics of attrited households suggests that attrition is not related to treatment assignment. These results are presented in tables 2.4 and 2.5.

Table 2.4: Household and Beneficiary Attrition Rates By Treatment Group

Variable	Existing Schools Food	Existing Schools Cash	Fstat	Ctrl	New Schools Food	New Schools Cash	Fstat
Household Attrition	0.08	0.06	1.074	0.08	0.08	0.08	0.102
	(0.01)	(0.01)	(0.303)	(0.01)	(0.02)	(0.01)	(0.903)
N	1318	1303	.	690	390	456	.
Beneficiary Attrition	0.17	0.15	0.579	0.16	0.16	0.13	0.768
	(0.02)	(0.01)	(0.449)	(0.01)	(0.02)	(0.02)	(0.467)
N	1253	1227	.	656	367	428	.

[1] Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$
[2] Standard errors in parentheses

The attrition rate is not significantly different between analysis groups. Furthermore, with a few exceptions, baseline characteristics of attrited households appear balanced across treatment groups. Amongst attrited households, the most concerning difference is in land ownership of the households from existing program schools. Attrited households from existing food schools were significantly less likely to own either residential or agricultural land relative to attrited households from existing cash schools. This difference is also observed in the full baseline sample. The difference amongst attrited households may simply be a reflection of the difference in the initial sample, rather than a product of differential attrition between the two treatment groups. Although not necessary for identification, baseline characteristics that are related to household attrition are included as control variables in the final analysis. This is done to control for nonrandom selection effects and to reduce the variance of program impact estimates.

Table 2.5: Baseline Characteristics of Attrited Households By Treatment Group

Attrited Households

	Existing Schools				New Schools		
Variable	Food	Cash	Fstat	Ctrl	Food	Cash	Fstat
Num. Adults	2.70	2.64	0.112	2.62	2.24	2.40	2.425*
	(0.12)	(0.13)	(0.740)	(0.12)	(0.13)	(0.10)	(0.097)
Num. Children	3.07	2.96	0.148	3.43	3.82	3.23	2.394*
	(0.19)	(0.19)	(0.701)	(0.24)	(0.15)	(0.26)	(0.100)
HoH Female	0.31	0.30	0.029	0.43	0.36	0.37	0.297
	(0.03)	(0.05)	(0.866)	(0.07)	(0.09)	(0.07)	(0.744)
Expenditure PC	367.52	400.84	1.118	382.79	386.63	433.37	0.837
	(17.85)	(26.29)	(0.294)	(28.67)	(51.90)	(29.97)	(0.438)
Food Exp. PC	258.59	265.53	0.136	241.94	244.15	273.32	0.517
	(14.78)	(11.92)	(0.714)	(19.27)	(28.75)	(26.28)	(0.599)
Own Ag. Land	0.33	0.57	6.139**	0.60	0.73	0.71	0.689
	(0.07)	(0.07)	(0.016)	(0.07)	(0.12)	(0.08)	(0.506)
Own Village Land	0.63	0.81	4.042**	0.77	0.85	0.74	0.678
	(0.08)	(0.05)	(0.049)	(0.07)	(0.06)	(0.08)	(0.512)
Age in years	11.92	11.74	0.332	12.34	11.73	11.74	1.074
	(0.21)	(0.23)	(0.566)	(0.30)	(0.47)	(0.38)	(0.348)
Female	0.68	0.57	1.815	0.57	0.48	0.60	0.563
	(0.04)	(0.07)	(0.183)	(0.07)	(0.07)	(0.10)	(0.573)
N	107	84	.	53	33	35	.

[1] Significance levels: * < 10% ** < 5% *** < 1%

[2] Standard errors in parentheses

When households were interviewed in the endline survey, the beneficiary was not always available. Beneficiary attrition, then, is higher than household attrition at 16.9%. This figure includes the 7.6% household attrition rate, leaving 9.3% of baseline households that were re-interviewed without the beneficiary being present. As with household attrition, beneficiary attrition appears to be non-random. Table 2.6 compares household and beneficiary characteristics between attrited and non-attrited beneficiaries. On average, attrited beneficiaries were close to one year (0.88) older than non-attrited beneficiaries. As might be expected given the age difference, attrited beneficiaries were also taller, weighed more, and performed better on cognitive tests.

Beneficiary attrition, although non-random, does not appear to be related to assignment. Neither attrition rates, nor the characteristics of attrited beneficiaries are significantly different between treatment groups, with one exception. The beneficiary attrition rates and characteristics of attrited beneficiaries are compared in tables 2.4 and 2.7, respectively. Beneficiaries from cash households scored lower on the Raven's Progressive Matrices test, relative to attrited beneficiaries from food households. As beneficiary attrition is related to age, beneficiary age is included as a control variable in all analysis of individual-level indicators to reduce variance and control for nonrandom selection effects.

Beneficiary-specific attrition has the potential to influence estimates of program impacts on cognitive testing outcomes, but it will not influence estimates of the primary educational outcomes, including dropout and attendance. Interview questions regarding the beneficiary's enrollment status and school attendance are answered by the primary caregiver, so the beneficiary does not need to be present. These measures could be influenced by household attrition, but not by attrition of the beneficiary alone.

Table 2.6: Baseline Characteristics by Beneficiary Attrition Status

Variable	InSample	Attrited	Diff
Num. Adults	2.68	2.71	0.027
	(0.03)	(0.05)	(0.053)
Num. Children	3.18	3.24	0.067
	(0.04)	(0.08)	(0.076)
HoH Female	0.34	0.33	−0.010
	(0.01)	(0.02)	(0.020)
Expenditure PC	409.49	404.54	−4.951
	(4.97)	(8.63)	(8.898)
Food Exp. PC	256.32	257.70	1.380
	(2.97)	(5.72)	(5.495)
Own Ag. Land	0.65	0.57	-0.087^{***}
	(0.02)	(0.03)	(0.025)
Own Village Land	0.86	0.81	-0.048^{***}
	(0.01)	(0.02)	(0.018)
Age in years	11.51	12.40	0.892^{***}
	(0.05)	(0.08)	(0.081)
Female	0.57	0.56	−0.008
	(0.01)	(0.02)	(0.022)
N	3307	624	3931

[1] Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$
[2] Standard errors in parentheses

Table 2.7: Baseline Characteristics By Treatment Group

Attrited Beneficiaries

	Existing Schools				New Schools		
Variable	Food	Cash	Fstat	Ctrl	Food	Cash	Fstat
Num. Adults	2.71	2.78	0.292	2.66	2.70	2.58	0.168
	(0.09)	(0.09)	(0.590)	(0.11)	(0.17)	(0.15)	(0.846)
Num. Children	3.03	3.30	1.568	3.35	3.57	3.32	0.610
	(0.14)	(0.17)	(0.214)	(0.17)	(0.16)	(0.25)	(0.546)
HoH Female	0.32	0.29	0.608	0.44	0.32	0.30	2.274
	(0.03)	(0.03)	(0.438)	(0.05)	(0.07)	(0.05)	(0.110)
Expenditure PC	393.19	388.61	0.065	431.36	420.45	434.98	0.036
	(11.46)	(13.95)	(0.799)	(22.96)	(46.02)	(31.80)	(0.965)
Food Exp. PC	252.53	255.27	0.043	266.18	254.15	273.64	0.208
	(9.69)	(9.03)	(0.836)	(14.68)	(21.38)	(23.32)	(0.813)
Own Ag. Land	0.45	0.59	3.467*	0.60	0.73	0.70	1.291
	(0.06)	(0.05)	(0.066)	(0.05)	(0.08)	(0.08)	(0.281)
Own Village Land	0.76	0.87	4.290**	0.80	0.83	0.82	0.189
	(0.05)	(0.03)	(0.041)	(0.04)	(0.05)	(0.06)	(0.828)
Age in years	12.36	12.36	0.000	12.71	12.43	12.09	1.605
	(0.13)	(0.16)	(0.993)	(0.22)	(0.29)	(0.27)	(0.208)
Female	0.56	0.56	0.017	0.56	0.53	0.58	0.149
	(0.03)	(0.04)	(0.895)	(0.05)	(0.06)	(0.06)	(0.862)
N	215	189	.	103	60	57	.

[1] Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$

[2] Standard errors in parentheses

Table 2.8: Compliance

	Reported			
Assignment	None	Food	Cash	Total
Ctrl	621	15	1	637
Food	121	1,420	27	1,568
Cash	99	28	1,513	1,640
Total	841	1,463	1,541	3,845

2.3.3 Compliance

Respondents were asked about scholarships received by each primary school student in the household. This was done to confirm that treatment households were receiving the correct scholarship and that control households were not receiving alternative scholarships. Specifically, respondents were asked if they had received a scholarship payment at any point in the school year from WFP, then asked if they had received a scholarship payment from any other source at any point in the year. Respondents were asked about the entire year, so all students from treatment schools should have received at least one distribution, unless they dropped-out before the first distribution.

Compliance with the assigned treatment was high, with 92.4% of households reporting receipt of their assigned scholarship type (or no scholarship). A household is considered to have received a scholarship if they report receiving at least one distribution over the duration of the 2011/2012 academic year. Table 2.8 presents a cross-tabulation of assigned and reported treatment. The majority of deviation from assigned treatment came from eligible treatment households that reported not receiving a scholarship of any type.

In total 220 eligible households reported not receiving any scholarship distributions. There are several potential reasons for these discrepancies. First, a beneficiary could drop out of school before the first scholarship distribution. In 105 of the 220 households, the beneficiary had dropped out at some point during the school year, though we do not know the exact month of dropout. Discrepancies could also be a result of incomplete knowledge within the household. The survey respondent may have been unaware of the beneficiary's participation in the WFP scholarship program, or could have known of the scholarship program under a different organization, such as MoEYS. Of the 115 cases of non-compliance where the beneficiary did not dropout, 16 reported receipt of a food or cash scholarship from a different source. An inconsistency in reported treatment could also reflect an actual deviation in treatment, where the beneficiary was eligible, but did not receive the assigned treatment. Finally, these discrepancies could be the result of enumeration or data entry errors. There is no evidence to suggest such errors in distribution or data collection, but neither can they be completely ruled out. Aside from the two schools with completely new beneficiary lists, there is no discernible pattern of non-compliance. Because non-compliance is low and seemingly random, only intent-to-treat estimates will be discussed in the analysis that follows.

2.4 Analysis

2.4.1 Econometric Approach

The econometric approach is motivated by the Rubin Causal Model (RCM) (Rubin, 1974; Imbens and Rubin, 2008; Holland, 1986). Originally postulated for analysis of data from clinical trials, the RCM was formally linked to the structural equation approach favored in economics by Angrist et al. (1996). The parameter of interest is the Average Treatment Effect (ATE) of the conditional transfer. A nonparametric estimate of this parameter can be obtained by differencing the average outcomes of households across randomly assigned treatment groups.

We consider several different outcome variables, but each can be modeled within the same framework. Specifically, define the two potential outcomes, (Y_0, Y_1) and treatment $d \in 0, 1$. For any individual, we observe a single outcome, $Y = dY_1 + (1 - d)Y_0$. The ATE can be defined as:

$$\begin{aligned} ATE =& E(Y_1 - Y_0 | d = 1) \\ =& E(Y_1 | d = 1) - E(Y_0 | d = 1) \end{aligned}$$

The second term, $E(Y_0|d = 1)$ is not observed in the data. Given effective randomization, however, a consistent estimate of this term can be obtained from the control group mean. $E(Y_0|d = 1) = E(Y_0|d = 0)$. Replacing this expression above, we get the following estimator for ATE.

$$ATE = E(Y_1|d = 1) - E(Y_0|d = 0)$$

The use of this simple, completely nonparametric estimator of treatment effects demonstrates the core strength of randomized program evaluation. Randomization provides statistically robust estimates of program impacts, free from any assumptions on an underlying economic model. For example, there are no distributional assumptions required on unobservables, nor is there required an assumption of additive separability of unobservables. (Heckman, 1996; Blundell and Costa-Dias, 2007) Of course, the separation of program evaluation results from an underlying economic framework is also a weakness. Without an economic framework, the results of randomized program evaluations cannot inform economic policy outside of the narrowly defined experimental setting. This limitation of randomized evaluations is addressed further in the next chapter, where I model the household decision problem in a dynamic discrete choice framework. The use of this model greatly expands the realm of policy questions that can be addressed using the data collected in this evaluation.

The full sample, composed of new and existing schools, was restricted to analyze different aspects of the program. First, the absolute impact of either type of transfer program was estimated. Only the 148 new program schools were randomized into cash, food, and control groups, while the 270 existing program schools were randomized only between food and cash. The control group, therefore, provides a valid counterfactual for the other new program schools, but cannot be compared to existing program schools. In estimating the absolute program impact, the sample was restricted to new program schools only. Using this reduced sample preserves the validity of the comparison between control and treatment schools, enabling a causal interpretation of any differences in outcomes.

The second aspect of the program to be analyzed is the relative impact of a cash scholarship compared to a food scholarship. In order to maximize sample size for this analysis, treatment schools were pooled across new and existing program schools. Control schools are excluded from this analysis. New and existing schools are not directly comparable, but assignment to food or cash treatment is effectively randomized across the pooled treatment group. This randomization enables estimation of the impacts of a cash scholarship relative to those of a food scholarship.

Several alternative specifications were evaluated, in order to investigate the robustness of the results obtained through the simple mean-comparison estimates. First, program impacts were estimated using OLS regression with control variables. The additional control variables have the potential to reduce the variance of the conditional mean estimate for each treatment group. The cost is that we must impose assumptions of linearity and additive separability of unobservable components. The resulting regression equations for the pooled estimates of absolute and relative program impact are as follows:

$$y_{is} = \delta_1 T_s + X_{is}\beta + \epsilon_{is} \quad for\ s \in NewSchools \tag{2.1}$$

$$y_{is} = \delta_2 C_s + X_{is}\beta + \epsilon_{is} \quad for\ s \in TreatmentSchools \tag{2.2}$$

where y_{is} is the outcome of interest for student i in school s, and X_{is} is a vector of control variables including both individual and household characteristics. In the first equation, T_s is a dummy indicating whether the school is part of the treatment group, and the baseline category is new control schools. The absolute treatment effect of the program is given by δ_1. In the second equation, C_s is a dummy indicating whether the school is part of the cash scholarship group, and the baseline category is food scholarship schools. The relative effect of cash vs. food treatment is given by δ_2.

The primary outcome variable of interest is a binary variable indicating whether the beneficiary dropped-out, or stopped attending school before the end of the school year. Due to the binary nature of this variable, probit regressions were also estimated, both with and without control variables. The probit specification requires an additional assumption relative to the OLS specifications, that the unobservable component is normally distributed.

Several secondary indicators were also analyzed, related to beneficiary educational participation and performance. For these indicators program impacts were estimated using both the nonparametric difference-in-means estimator as well as the OLS specification with control variables. Estimates are only presented from pooled specifications. The absolute impact of the treatment was estimated with a pooled treatment group of both cash and food recipient schools, while the relative impact was estimated after pooling new and existing schools.

The same set of household and individual-level variables were included in all specifications with control variables. Household-level controls include variables that were significantly different between analysis groups in the baseline data, variables related to attrition, indicators for disasters such as flood and drought, and province dummies. Indicators for vitamin A supplementation and deworming treatment are also included, and indicate that someone in the household received the treatment. Baseline values were used for almost all control variables, so they could not be affected by the treatment. The only exceptions are the disaster indicators, which are taken from the endline data. Natural disasters are plausibly unrelated to treatment, and the endline responses cover the relevant time period. Individual-level controls include an indicator for female, and indicator for recent diarrhea, and a flexible specification of age in months including square and cubic terms. The full set of household control variables, excluding province dummies and individual controls, are presented in table 2.9.

2.4.2 Results

School Participation

The mean dropout rate for each treatment group is given in table 2.10. The nonparametric treatment effect estimates are given in the first column of table 2.11. Several different estimates were calculated for absolute and relative program impacts with and without pooling. In the first column, the estimate is simply the estimated difference in means between the relevant treatment and control groups. In subsequent columns, the program impact coefficient from each specification is reported. For OLS specifications, estimates correspond to δ_1 or δ_2 in the specification equations above. For probit specifications, the program impact estimate is the average marginal effect

Table 2.9: Control Variables

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Age (months)	3,845	155.6	21.40	64	238
Female	3,845	0.561	0.496	0	1
Num. Adults	3,845	2.702	1.252	0	9
Num. Children	3,845	3.182	1.470	0	16
HoH Female	3,845	0.336	0.473	0	2
Food Exp. PC	3,845	254.0	123.6	32.61	760.9
Own Village Land	3,845	0.859	0.355	0	5
Own Ag. Land	3,845	0.651	0.477	0	1
Expenditure PC	3,845	409.2	236.5	78.29	4,516
Vit. A Treatment	3,845	0.336	0.472	0	1
Deworm Treatment	3,845	0.699	0.458	0	1
Flood	3,845	0.337	0.473	0	1
Drought	3,845	0.188	0.390	0	1
Other Disaster	3,845	0.106	0.307	0	1

of the treatment coefficient. Full regression results from the OLS and probit specifications are given in the appendix.

Estimates of program impact on beneficiary dropout are broadly similar across the specifications. Estimates from the first column indicate that the absolute impact of the scholarship program reduced the beneficiary dropout rate by 4.3 percentage points. The dropout rate for targeted children in the control group was 9.4%, so this represents a 46% reduction in primary school dropout. Analysis of the relative impact of food and cash scholarships suggests that cash scholarships are more effective than food scholarships at reducing beneficiary dropout. Relative to a food scholarship, receipt of a cash scholarship reduces beneficiary dropout by 1.9 percentage points.

Table 2.10: Mean Dropout Rate

VARIABLES	(1) New	(2) Existing
Ctrl	0.0942	
	(0.0169)	
Food	0.0560	0.0892
	(0.0112)	(0.0107)
Cash	0.0475	0.0673
	(0.00974)	(0.00818)
Observations	1,415	2,430

Robust standard errors in parentheses

Table 2.11: Treatment Effect Estimates

VARIABLES	(1) NP	(2) OLS	(3) Probit	(4) Probit w Ctrls	(5) N
Food	-0.0382*	-0.0311*	-0.0406*	-0.0357*	994
	(0.0203)	(0.0186)	(0.0229)	(0.0202)	
Cash	-0.0467**	-0.0403**	-0.0498**	-0.0470**	1,058
	(0.0195)	(0.0177)	(0.0222)	(0.0192)	
Treat	-0.0428**	-0.0365**	-0.0422**	-0.0386**	1,415
	(0.0184)	(0.0170)	(0.0181)	(0.0161)	
Cash New	-0.00852	-0.00720	-0.00848	-0.00378	681
	(0.0149)	(0.0127)	(0.0148)	(0.0142)	
Cash Existing	-0.0219	-0.0256**	-0.0219	-0.0245**	2,430
	(0.0135)	(0.0113)	(0.0135)	(0.0111)	
Cash Pooled	-0.0194*	-0.0217**	-0.0194*	-0.0208**	3,208
	(0.0110)	(0.00926)	(0.0110)	(0.00915)	

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

The dropout rate for food schools was 8.2%, so this 1.9 percentage point reduction represents a 23% reduction in primary school dropout.

Educational Inputs

The results on educational inputs and outcomes are presented in a condensed format. For each indicator, we consider both the absolute and relative impact of the scholarship program. The absolute impact is the impact of receiving any scholarship, as opposed to receiving no scholarship. The absolute impact is estimated using new program schools only, where cash and food scholarship schools are pooled in a single treatment group. Columns (1) and (2) present the mean of each indicator for new control schools and new treatment schools from the endline survey. Column (3) presents the difference-in-difference estimates of the absolute impact of the scholarship program, which is represented by δ_1 in the first analysis equation. The standard set of control variables is used to produce the estimates in column (3).

The relative impact of the program is the impact of receiving a cash scholarship, as opposed to receiving a food scholarship. The relative impact of the program is estimated using both new and existing program schools. New and existing schools are pooled into cash and food analysis groups. New control schools are excluded from this analysis. Columns (4) and (5) present the mean levels of indicators for food and cash scholarship schools, respectively, from the endline survey. Column (6) presents the difference-in-difference estimates of the relative impact of the scholarship program, including control variables. This estimate corresponds to δ_2 in the second analysis equation.

The sample for the three remaining indicators, average attendance, days with homework, and educational expenditures, is smaller than the sample used to estimate beneficiary dropout. This is largely due to the fact that the remaining estimates are conditional upon beneficiary school attendance. This complicates interpretation of these estimates, as discussed below. The drop in sample size is actually larger than a decrease based only on beneficiary dropout. This is due to non-response on the beneficiary questionnaire. This non-response is plausibly random with respect to the treatment, so should not bias the estimates of program impact.

Non-response due to beneficiary dropout, however, is problematic, as dropout is clearly related to treatment assignment. These results must be interpreted with caution. This analysis is predicated on the assumption that outcomes of treatment and control groups can be compared to obtain valid, unbiased estimates of the effect of the program. Restricting the sample to non-dropouts has altered the composition of the groups, which will likely introduce bias in the resulting estimates.

While the exact nature of the bias cannot be known, it is possible to hypothesize about the direction of the bias under reasonable assumptions. Programs that incentivize enrollment have the potential to lower average educational outcomes of enrolled students, as they induce marginal students to enroll in school (Schuck and Zeckhauser, 2007). These students, who would not have chosen to enroll without the incentive, may be less motivated or less able in the classroom, lowering the average performance of the class. The resulting estimates are a combination of the true impact of the program and the negative selection effect, producing a downward bias in the estimates of program impact.

Despite this caveat, these estimates are informative for two reasons. First, given the assumptions above, these estimates of program impacts are a lower bound for the true impact of the program. It is plausible that the true impact of the program on an individual student is greater or equal to the estimates presented here. Second, while these estimates are not unbiased estimates of the effect of the scholarship, they do reflect the actual, observed outcomes in the two groups of schools. The biased estimates provide the best prediction of future average outcomes if the program is expanded to schools facing similar conditions.

Table 2.12: Educational Investment

	Absolute			Relative		
Variable	Ctrl	Treat	DinD	Food	Cash	DinD
Attendance	5.81	5.86	0.067	5.90	5.89	−0.013
	(0.04)	(0.02)	(0.044)	(0.01)	(0.01)	(0.027)
Home Work Days	3.58	3.75	0.352**	3.60	3.63	0.078
	(0.09)	(0.08)	(0.173)	(0.06)	(0.07)	(0.120)
Education Exp.	36.98	37.88	2.740	40.36	41.33	3.052
	(1.73)	(1.96)	(3.212)	(1.66)	(1.37)	(2.609)
N	543	700	1243	1384	1475	2859

[1] Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$
[2] Standard errors in parentheses

Despite the potential negative bias, estimates of the absolute impact of the program on conditional educational outcomes are generally positive or not significant. For those beneficiaries that completed the academic year, students in treatment schools did school work at home an additional 0.35 days per week, compared to students in control schools. There was no difference in educational expenditure or attendance. Attendance was well above the required rate of 80%, even in control schools. Primary schools in Cambodia are in session 6 days per week, so the control group average attendance of 5.8 days per week represents an attendance rate of 96.8%.

Turning to the relative impact of the program, the outcomes of the cash scholarship group were not statistically different from those of the food scholarship group, conditional on non-dropout. Of course, these estimates are subject to the same selection bias as the estimates of the absolute impact. It cannot be determined from these estimates whether the program truly had no effect on these outcomes, or if any impact was negated by selection effects. Again, it is possible to hypothesize about the probable direction of any unobservable selection bias.

Given the larger impact of the cash scholarship, it is likely that the marginal student induced to complete the academic year in all cash scholarship schools was of lower ability than the marginal student across all food scholarship schools. As in the previous discussion of treatment vs. control schools, the cash scholarship schools should have a lower average ability than the food scholarship schools. The estimates of the relative impact of cash relative to food, then, are biased downwards. The estimates here are a lower bound for the relative impact of a cash scholarship relative to a food scholarship. While there is no evidence that cash scholarship recipients make more investments in education, there is some evidence that cash recipients do not invest less in education than food scholarship recipients.

Learning Outcomes

These indicators are intended to measure the program's impact on learning outcomes. These outcomes are generally associated with education, though any effect measured here could be a result of the additional education induced by the scholarship program or improved nutrition. Educational outcomes are measured using several cognitive tests: digit span, Raven's Progressive Matrices, and a local standardized math test. The math exam was only administered in the endline survey, so estimates for the math test are single-difference estimates. It is important to note that these tests were administered to all beneficiaries in the sample. Testing was not conditional on school enrollment, so these estimates do not suffer from the selection bias described in the previous section.

Estimates of the effect of the program on educational outcomes were not significantly different from zero. This is true both for the absolute impact of any scholarship

Table 2.13: Educational Outcomes

	Absolute			Relative		
Variable	Ctrl	Treat	DinD	Food	Cash	DinD
Digitspan	5.76	5.77	0.117	5.85	5.87	−0.022
	(0.06)	(0.05)	(0.101)	(0.05)	(0.05)	(0.073)
Ravens	11.42	11.28	−0.152	11.32	11.42	0.147
	(0.14)	(0.15)	(0.204)	(0.18)	(0.13)	(0.173)
Math Test (0-13)	6.92	6.81	−0.110	5.96	6.15	0.190
	(0.28)	(0.28)	(0.393)	(0.15)	(0.18)	(0.233)
Testing Index	0.13	0.08	−0.028	−0.05	0.02	0.069
	(0.08)	(0.09)	(0.098)	(0.08)	(0.06)	(0.070)
N	553	678	1231	1345	1409	2754

[1] Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$

[2] Standard errors in parentheses

and the relative impact of a cash scholarship versus a food scholarship. Thus, despite the increase in attendance and enrollment, scholarship recipients did not improve their performance on these cognitive and knowledge tests.

There are several possible underlying reasons. It could be due to the nature of the tests administered. Raven's Progressive Matrices, in particular, is designed to measure general cognitive ability, rather than specific knowledge taught in classrooms. Alternatively, it could be that the quality of teaching is such that the impact of a year of treatment does not reflect on educational outcomes. This explanation is most relevant for the local math test, which covers material from the Cambodian primary school curriculum. Of the three tests, this one is most likely to be sensitive to changes in schooling.

Table 2.14: Treatment Effects - by Sex

	Male			Female		
VARIABLES	NP	OLS	N	NP	OLS	N
	(1)	(2)	(3)	(4)	(5)	(6)
Food	-0.0396	-0.0306	408	-0.0366	-0.0333	586
	(0.0286)	(0.0262)		(0.0228)	(0.0212)	
Cash	-0.0557*	-0.0392	458	-0.0398*	-0.0420**	600
	(0.0285)	(0.0253)		(0.0204)	(0.0195)	
Treat	-0.0489*	-0.0357	593	-0.0383**	-0.0390**	822
	(0.0259)	(0.0237)		(0.0194)	(0.0183)	
Cash New	-0.0160	-0.00308	320	-0.00321	-0.00953	458
	(0.0242)	(0.0219)		(0.0190)	(0.0176)	
Cash Existing	-0.0292*	-0.0310**	1,096	-0.0162	-0.0225	1,334
	(0.0157)	(0.0147)		(0.0181)	(0.0152)	
Cash Pooled	-0.0274**	-0.0253**	1,416	-0.0131	-0.0213*	1,792
	(0.0136)	(0.0125)		(0.0145)	(0.0125)	

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

2.4.3 Heterogeneous Impacts

Differential treatment effects were estimated for male and female beneficiary subgroups. The treatment effect coefficients from nonparametric and OLS specifications are given in table 2.14. Differential effects were also estimated in a single regression, with results given in 2.15. Additional indicator and interaction terms were included to allow treatment effects to vary for females, depending on the presence of a young child child in the household (5 years or younger). This specification is motivated by the idea that older female children may participate in the childcare of younger children. Secondary indicators by sex and presence of a dependent child are included in the appendix.

Table 2.15: Heterogeneous Treatment Effects

VARIABLES	(1) Absolute	(2) Relative
Treat	-0.0489*	
	(0.0260)	
FemTreat	0.0157	
	(0.0280)	
FemDepTreat	-0.0132	
	(0.0297)	
FemDep	0.0141	0.0116
	(0.0243)	(0.0203)
Female	-0.0137	-0.00424
	(0.0215)	(0.0177)
Cash		-0.0274**
		(0.0136)
FemCash		0.0161
		(0.0223)
FemDepCash		-0.00348
		(0.0273)
Constant	0.0989***	0.0812***
	(0.0229)	(0.0110)
Observations	1,415	3,208
R-squared	0.007	0.002

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2.5 Conclusion

2.5.1 Absolute Impact

The scholarship program had a positive impact on school participation. Beneficiaries dropped out of school at a lower rate in treatment schools relative to control schools. In absolute terms, the impact was modest; the dropout rate was decreased by almost 4 percentage points. Relative to the dropout rate of control households, though, this is a fairly large impact. Given the counterfactual dropout rate of 9.4%, the WFP food and cash scholarship reduced the dropout rate by 41%.

Considering educational indicators for beneficiaries that did not drop out of school, estimates show that beneficiaries who received a scholarship had higher levels of educational investment, in terms time spent on homework, compared to beneficiaries who did not receive a scholarship. While these estimates may suffer from selection bias, as previously discussed, they suggest that the scholarship enabled households to increase educational investments beyond the minimum conditions required to receive the scholarship. The increased investments in education did not, however, benefit learning outcomes as measured by the three tests administered.

This is consistent with the existing evidence on CCT programs. While effective at increasing school participation, CCTs generally fail to produce improvement in learning outcomes (Shari et al., 2013; Schady and Fiszbein, 2009; Baird et al., 2014). Of course, the lack of evidence could be due to the difficulty in measuring learning outcomes. Leaving that possibility aside, the lack of impact on learning outcomes suggests that children induced to participate in school did not gain from their participation. This possibility raises questions about the assumptions that motivate the implementation of conditional transfer programs. Conditional transfer programs are predicated on the assumption that students induced to participate in school will benefit from their participation. Furthermore, the choice of non-participation is sub-optimal in the long-run. If prolonged schooling does not improve learning outcomes, it may be the case that households are acting optimally by withdrawing their children from primary school.

It is always possible that, due to the particular characteristics of the school and the student, some children may not accrue any benefit from additional schooling. This is not to say that there are children who are incapable of learning, or that there are schools that are incapable of providing any enrichment. In any school, there will always be students who benefit from instruction, no matter how rudimentary. Similarly, any student is capable of learning given appropriate instruction. Some students will require more intensive instruction in order to improve their learning outcomes, and some schools will not have the resources to provide intensive or specialized instruction. For any student whose instructional needs outstrip the school's resources, school participation may not convey any benefits. School participation will unavoidably impose some cost, even if only the opportunity cost of the student's time. For students with zero potential benefit, non-participation will always be the optimal choice.

Conditional transfer programs will only benefit learning outcomes if they reach students who have positive potential benefits from attending their local school. For these students, the potential benefits of attendance, though positive, may be outweighed by the cost of school attendance. The cost of attendance may be driven by school fees, transportation costs, or opportunity costs.

Non-participation may be driven either by the high opportunity cost of attending school or the insufficient quality of classroom instruction, conditional on student ability. Conditional transfer programs will only provide benefits in terms of educational outcomes in the first case. In the second case, resources would be better spent on improving the quality of classroom instruction. The two causes, however, are difficult to separate. Data from conditional transfer programs can provide evidence on the root cause of non-participation. The data collected here are valuable as evidence on the cost of attendance and the quality of classroom instruction, even if the conditional transfer program does not itself produce improvements in educational outcomes. In the current setting, the results indicate that educational concerns would be better addressed through supply-side interventions designed to improve the quality of classroom instruction.

2.5.2 Relative Impact

The cash scholarship had a greater impact on primary school dropout than the food scholarship. While the exact reason for this differential impact is not known, there are several plausible explanations. In section 2.1.2, four potential reasons were proposed why cash and food scholarships may have different impacts on educational outcomes. Two of the reasons are based on how the household might spend the transfer, whether on educational investments or other expenses. The other two reasons are based on how the household values each type of transfer.

Primary school is free in Cambodia, so it is likely that the differential impact of cash transfers is driven by the way the household values the transfer, rather than how it spends the transfer. There is no particular expense that must be met in order to enroll a child in primary school, so it is unlikely that greater spending flexibility would promote increased participation. Indeed, if the household's valuation of the two scholarship types were equal, there should be no difference in participation rates. In that case, any differential impacts would be driven by the way that the transfers were spent. For example, the flexibility of cash could enable diverse educational investments, leading to an increase in learning outcomes.

This reasoning suggests that cash transfers are valued at a premium relative to food transfers. One explanation for the cash-valuation premium is simply the inconvenience of transporting large amounts of rice. As described earlier, accrued scholarship transfers are distributed every 2 or 3 months. Food scholarship recipients, then, must transport 20 to 30 kg of rice at each distribution. Beneficiary households could incur significant transportation costs to transport this rice from the distribution point.

Alternatively, households may have a particular need for cash. Recipients of food transfers may face some monetary or non-monetary cost to exchange rice for cash. Results from the full evaluation report on this program suggest that a significant portion of the transfer payment was used to repay debt. This was true for both food and cash transfer recipients. The fact that food households were able to repay debt in similar amounts as cash households suggests that they were able to effectively convert the value of their food transfer to cash. The differential valuation may be driven by the inconvenience of enacting this transaction in order to repay debt.

Food transfers, while valued by recipient households, also impose logistical costs. The logistical costs of food transportation are also paid by the distributing agency, WFP. The impact of the transfers, both in terms of educational outcomes and the ultimate disposition of the transfer, was statistically indistinguishable between food and cash transfers. The only statistically significant difference between recipients of the two transfer types was linked to the valuation of the scholarship, and is likely driven by the logistical costs of large food transfers. In the context of this program, transfers impose unnecessary costs on both the donor and recipient, with no benefit in terms of educational outcomes. These results support the use of cash transfers over food transfers in context of the current program. This preference is likely to carry over to any conditional transfer program with functioning markets, where food is readily available.

Chapter 3

Structural Estimation of Household Decisions

3.1 Introduction

The previous chapter estimated the causal impact of a conditional transfer program in an experimental framework. The outcomes of the different treatment arms were compared directly, without appealing to an underlying economic model. The results have a strong claim to a causal interpretation from a statistical vantage point, but do not illuminate the underlying economic decision. Nor do the results provide estimates on the potential impacts of policy options that were not considered in the experiment.

At its core, the decision of whether and which children to enroll in school is made simultaneously at the level of the household. When children reach an age where they can be productive contributors to the household budget, the decision to enroll a child in school involves a trade-off between current household income and the future human capital of the child. Looking at the experimental results from a single evaluation misses several important components of this decision, including the relative values that a household assigns to different activities of children and whether certain household activities are substituted between children. This chapter will therefore consider the household decision problem, rather than focusing on the reduced-form estimates of the program impact on individual children. This will allow the results of the program evaluation to provide insight into the underlying household decision regarding child time allocation. A deeper understanding of the household decision problem, considering not only the targeted child's time allocation but also that of his or her siblings, can answer questions that cannot be addressed by a basic experimental evaluation, such as the nature of substitution between children.

The comparison of cash and food transfer modalities highlights the importance of analyzing the impact of other - potentially more effective - policy interventions on schooling. This is difficult to achieve strictly within the framework of reduced-form models estimated using data from randomized control trials (RCTs). Social experiments can provide statistically robust estimates of program impacts, but are costly to implement and can provide relatively little variation in treatment (Todd and Wolpin, 2006). Evaluation of alternative policies within the RCT framework requires the deployment of a new social experiment.

In addition to providing a deeper understanding of household dynamics, structural estimation provides a way to estimate the impact of policy alternatives without actually implementing and evaluating each one. Once a structural model has been developed, a wide array of policy alternatives can be incorporated into the model. The potential impacts of each alternative can be explored with much less time and cost through simulated outcomes.

Structural estimation provides a method to simulate counterfactual outcomes, but the predictions from a model can be difficult to verify. Structural models rely on functional form and distributional assumptions to identify parameters. Identification generally cannot be formally proven. Combining structural estimation with experimental data provides an approach to validate the predictions of a structural model. Some portion of the treatment sample can be separated and not used to estimate the structural model. The structural model can then be used to simulate the actual treatment, and the resulting simulated treatment effects compared to the experimental estimates.

In this chapter, I develop and estimate a structural decision model to analyze the decisions made by households in the context of the program evaluated in the previous chapter. Instead of estimating the effects of conditional transfers solely on the schooling of the targeted child, this model investigates how a decrease in the cost of schooling for the targeted child affects household decisions on the overall time allocation of children in the household. This model will combine the results of the reduced-form approach from the previous chapter with structural estimation of a dynamic discrete choice model, with the goal of modeling the effects of counterfactual policy alternatives for increasing schooling. The model is estimated using baseline data only, and the final simulated treatment effects are compared to the experimental estimates from the endline data.

Structural estimation is not without cost. The development of a robust and flexible structural model requires a substantial investment in research time. Additionally, solving and estimating the parameters of such a model can impose significant computational burdens. I will first explore the feasibility of estimating the model in a dynamic discrete choice framework. Specifically, I will consider the model developed by Todd and Wolpin (2006) to analyze a similar household decision problem faced by households in Mexico with the introduction of the PROGRESA conditional cash transfer program.

The formal model for the household decision problem in the Cambodian context is defined in section 3.2. In section 3.3, I modify some elements of this model to produce a basic exploratory model, based on the model from Todd and Wolpin (2006). I will simulate outcomes using the parameters estimated by Todd and Wolpin (2006), but using data from the current Cambodian evaluation. Comparison of both observed and simulated data will inform whether this type of model is appropriate in the current context. I will then attempt to recover the parameter estimates from outcomes simulated from Cambodian data. This exercise will inform the feasibility of developing and estimating a dynamic discrete choice model designed specifically for the Cambodian setting.

In section 3.4, I return to the initial decision problem proposed to describe the specific context of rural Cambodia. I introduce additional modifications intended to ease the computational burden of model estimation. I proceed to estimate the structural model and simulate outcomes using baseline data from the randomized control trial of the Cambodia conditional transfer program. This estimation and simulation based on baseline data provides evidence that the model predicts activity choices that are broadly similar to those observed in the data.

After simulating baseline activity choices, decisions are simulated again after including each type of scholarship transfer. Estimates of the program impact are calculated based on these simulated decisions and compared with the experimental estimates from the reduced-form approach presented in the previous chapter. Based on these results, I conclude that the model shows potential to capture the relevant factors of the household decision problem, though some additional work is required. Specifically, the treatment effect estimates for the food conditional transfer calculated from simulated outcomes are similar in direction and magnitude as the experimental estimates from the RCT evaluation. The simulated outcomes for the cash conditional transfer, however, show no impact from receiving the transfer, which is not consistent with the experimental estimates from the RCT evaluation. Upon completion of this work, I intend to use the final structural model to estimate the effectiveness of alternative policy interventions that were not tested in the initial RCT evaluation. Specifically, I intend to investigate the relative efficacy of expanded scholarship eligibility of children within each household against an increased scholarship amount for a single beneficiary child.

3.1.1 Literature Review

As highlighted in the previous chapter, many households in developing countries appear to underinvest in education, in spite of evidence of strong positive returns both at the national and individual level. One widespread explanation for this is that parents face a trade-off between sending their children to school (which has upfront costs and delayed economic benefits) and employing them in labor-intensive jobs (of which the economic benefits are upfront, but the costs - in terms of foregone human capital – are delayed). This apparent substitution effect may have lasting impacts on the fight against extreme poverty: as expressed by the International Labor Organization, the concern is that "child labor perpetuates poverty across generations by keeping children of the poor out of school and limiting their prospects for upward social mobility" (ILO, 2015).

A wide economic literature investigates the determinants of household decisions regarding the time allocation of children (see for instance Basu and Van (1998); Baland and Robinson (2000); Emerson and Souza (2007, 2008); Bacolod and Ranjan (2008); Dammert (2010); Attanasio et al. (2010)). Much of the work in this field relies on a combination of case studies and economic modeling, with little use of experimental evidence.

Ravallion and Wodon (2000) use a quasi-experimental approach, based on a targeted enrollment subsidy in Bangladesh, to test for substitution effects between child labor and schooling. They use reduced-form analysis motivated by an economic decision model to demonstrate that the enrollment subsidy increases schooling without necessarily reducing child labor. Their findings suggest that the subsidy increases school attendance of the targeted child and reduces the incidence of child labor, but only by a fraction of the increase in the school enrollment rate. This is partly due to the fact that the school day in rural Bangladesh is structured on a part-time basis, allowing sufficient time for children to engage substantially in both schooling and labor.

Ravallion and Wodon (2000) look at the trade-off between activities of a single child. A second question in household decisions concerns the trade-off of activities between children. Ferreira et al. (2009) investigate this question using regression discontinuity approach based on a secondary school scholarship program in Cambodia. The authors found that the scholarship increased the probability that the scholarship recipient would enroll in school by roughly 20 percentage points. Furthermore, there was no impact on the enrollment probabilities of ineligible children within the household.

This analysis highlights the importance of evaluating the effects of conditional transfers not just in terms of the activity conditioned upon, but also in terms of impacts on potentially competing activities. Accordingly, the current chapter will consider the effect of the Cambodia conditional transfer program from a household decision perspective, to enable a more nuanced understanding of its impacts on the time allocation of the targeted children and their siblings. This allows the analysis to contribute to a wider discussion regarding the linkages between education, poverty, child labor, and growth.

The second main innovation of this paper with respect to traditional program evaluation methods for conditional education transfers is that it applies evidence from a randomized control trial to analysis of policy alternatives, by combining the reduced-form program evaluation approach with structural estimation. This follows the approach of Todd and Wolpin (2006), who put the combination of program evaluation and structural estimation into practice in addressing a similar question. They use data from a randomized control trial of the PROGRESA school subsidy program in Mexico to build a structural estimation model that estimates the efficiency of alternative counterfactual policies, which could not be tested using a reduced-form approach given that they were not implemented as part of the experiment. They further compare the reduced-form estimation of PROGRESA's treatment effect on household decisions of child time allocation (provided by the randomized control trial) to estimates of the program's impact as provided by an out-of-sample prediction from their structural model. This allows them to validate their structural model and its forecasts of alternative policy choices. Similar approaches are followed by Attanasio et al. (2010), in another evaluation of the PROGRESA program, and by Bourguignon et al. (2003), in their analysis of Brazil's Bolsa Escola program.

Combining program evaluation approaches with structural estimation models in forecasting the impacts of counterfactual policies is a relatively recent but growing practice. The benefits of this hybrid approach are discussed at length by Heckman (2010), who argues that comparing structural estimates with experimental estimates "extend[s] the interpretability and range of policy questions" that policy analysis can answer, essentially by reintroducing economic choice theory into the logic of program evaluation. As Todd and Wolpin (2006) explain, social programs are usually short (so that long-term effects cannot be readily estimated from a reduced-form approach), costly, and have scarce variation in the range of treatments. Randomized control trials can therefore answer only a very limited range of questions based on what policy variables can be randomly assigned; yet they have the benefit of being more transparent and easier to replicate than structural estimation models. Structural estimation models, on the other hand, are capable of taking into account both ex-ante and ex-post returns by jointly modeling outcome and choice equations, and they can estimate impacts of policies not actually carried out in the experimental setting. As Heckman (2010) aptly explains, combining these two approaches can help move beyond estimating the "effects of causes", the primary aim of program evaluation, to modeling the underlying economic decisions, or the "causes of effects."

The methodology of this analysis comes from the field of dynamic discrete choice models. This approach relies heavily on the pioneering work developed and described in Rust (1996, 1987). A recent survey of this literature can be found in Aguirregabiria and Mira (2010).

3.2 Methodology

3.2.1 Program and Data Description

This analysis will use data from a recent randomized program evaluation in Cambodia. The details of the program evaluation are covered in the previous chapter. Here, I will review only the aspects of the program that are relevant to the current structural analysis.

The program under evaluation is an on-going Food for Education program conducted by the World Food Programme. Eligible students receive 10 kg of rice per month, conditional on continued enrollment and participation in primary school. Eligibility is determined for specific students, not for the entire household. In order to be eligible, a student must be enrolled in grades 4-6, and come from an economically disadvantaged home.

The evaluation aimed to provide information both on the effectiveness of the existing food transfer scheme as well as an alternative scheme of cash transfers. The cash transfer scheme was designed to be identical to the existing food transfer program in all aspects except the modality of the transfer. The transfers were of similar values (50 USD per year), distributed on the same time line, and followed the same operational procedures. Beneficiary selection was carried out simultaneously before treatment assignment.

The baseline evaluation population consisted of students attending one of two types of schools. The first group of schools, which I will refer to as 'existing' schools, had an existing WFP food for education program in place. The second group of schools, which I will refer to as 'new' schools, did not have a WFP food for education program, or any other type of cash or food conditional transfer or scholarship program. Randomized treatment was assigned to schools, rather than students. Existing schools were assigned to either continue their existing program or change to the cash transfer scheme. New schools were assigned to one of three treatment groups: control, food transfer, or cash transfer.

Data collection for the evaluation was conducted via two rounds of household surveys. In each round, approximately 10 households were interviewed from each school, or 4,000 households in total. In creating the estimation sample, I follow several of the sample restrictions introduced by Todd and Wolpin (2006), to create a relatively homogeneous evaluation sample. Specifically, only nuclear households are included, defined by the presence of a female less than 50 years of age who is identified as the spouse of the household head.

Unlike Todd and Wolpin (2006), I will not limit the sample to landless households, for both practical and theoretical reasons. Practically, this restriction would reduce the Cambodian estimation sample too much. Home production is prevalent in Cambodia, even amongst the poorest households. Excluding home production in this context is theoretically, as well as practically, untenable. The inclusion of land-holding households requires that I estimate home production for each household, as well as wage income.

Table 3.1: Baseline Summary Statistics

VARIABLES	Cambodia Sample N	mean	sd	PROGRESA Sample mean	sd
Wife's Age	2,079	39.54	7.414	30.5	8.1
Husband's Age	2,079	41.79	8.363	34.4	9.5
Number of kids born	2,079	3.812	1.628	3.01	1.92
High Grade for Ages 7 to 11	2,637	2.390	1.381	2.39	1.41
High Grade for Ages 12 to 15	1,918	4.434	1.480	5.79	1.76
High Grade for Ages 16+	5,863	3.223	2.929	6.60	2.81
High Grade for Males 16+	3,008	3.716	3.028	6.64	2.82
High Grade for Females 16+	2,855	2.703	2.727	6.56	2.81

After restricting the sample to nuclear households, my estimation sample contains 1,962 households. I restrict the sample further based on computational considerations. As discussed below, the computational requirements increase with the number of potential children in each household. I limit households to a maximum of four children, and remove all households with more than four children from the sample. Removing households with 5 or more children excludes almost 30% of the sample. This is done in order to make the problem computationally tractable, and may introduce selection bias. In future work, I intend to increase the maximum allowable household size to address this issue. My final estimation sample, then, contains 1,398 households. This is comparable to the size of the Todd and Wolpin (2006) estimation sample, which consisted of 1,316 households. Baseline summary statistics for both estimation samples are presented in table 3.1.

3.2.2 Decision Model

In this model, a household makes decisions on the joint time allocation of existing children, simultaneously with reproductive decisions for potential children, in order to maximize expected utility. The time frame covered by the model includes mothers beginning at age 30 until age 60. The central focus of this model is the trade-off between current consumption and educational achievement. The household decision problem is defined in table 3.2.

Feasible Activity Decisions

The discrete choice model considers a married couple making time allocation decisions for members of the household. In each discrete time period, the couple chooses their own primary activity as well the primary activity of each child. The father's activity choices are wage labor and home production. The mother has the same choices, with the additional option of pregnancy. When a mother chooses pregnancy, I assume that she engages in home production to some degree, but wage labor is precluded. The possible activities for children are staying at home, working for wages, and attending school. There are age restrictions for both parents and children for various activities. Pregnancy is only an option when the mother is less than 45 years of age. The reproductive period covers half of the full model time span, which includes mothers from age 30 to age 60. Children may attend school from age 6 until age 17. Children may engage in wage labor at ages 12 and above. This age restriction is based on the observed primary activities within the data, where only one child was reported working below age 12. Children are able to remain at home at any age, and it is the only option for children less than 6 years old. Children remaining at home may contribute, either

Table 3.2: Household Decision Problem

$$\max_{D_1^{T-1}} \quad \mathbb{E}_1 \left[\sum_{t=1}^{T-1} \beta^t U(y_t, p_t, X_t, D_t) \; + \; \beta^T TV(X_T) \;\middle|\; X_1 \right] \tag{3.1}$$

$$\begin{aligned} D_1^T &= \{D_1, \ldots, D_t, \ldots, D_T\} \\ D_t &= \{d_{1,t}, \ldots, d_{n_t,t}\} \\ d_{i,t} &\in \{w, s, h, b\} \\ X_t &= \{x_{1,t}, \ldots, x_{n_t,t}\} \end{aligned}$$

$$U(y_t, p_t, X_t, D_t) = u_1\left(\frac{y_t}{n_t}, \frac{p_t}{n_t}\right) + u_2(X_t, D_t) \tag{3.2}$$

$$u_1\left(\frac{y_t}{n_t}, \frac{p_t}{n_t}\right) = \frac{1}{1-\theta_y}\left(\frac{y_t}{n_t}\right)^{1-\theta_y} + \frac{1}{1-\theta_p}\left(\frac{p_t}{n_t}\right)^{1-\theta_p}$$

$$u_2(X_t, D_t) = \theta_U' L_U(X_t, D_t)$$

$$TV(X_T) = \theta_{TV}' L_{TV}(X_T) \tag{3.3}$$

$$y_t(X_t, D_t) = \sum_{i=1}^{n_t} \mathbb{I}(d_{i,t} = w) y_{i,t}(x_{i,t}) \tag{3.4}$$

$$log(y_{i,t}(x_{i,t})) \sim N(\mu(x_{i,t}), \sigma(x_{i,t})) \tag{3.5}$$

$$log(p(X_t, D_t)) \sim N(\mu(X_t, D_t), \sigma(X_t, D_t)) \tag{3.6}$$

Variable Definitions

β^t	: discount factor
$x_{i,t}$	: sex, age, and education of household member i
$d_{i,t}$	: primary activity of household member i
w	: work for wages
s	: school
h	: home production or leisure
b	: pregnancy (or baby)
y_t	: household wage income
$y_{i,t}$	: wage of household member i
p_t	: household home production
n_t	: number of household members
$\pi(X_{t+1}\|X_t, D_t)$	: conditional transition probability

positively or negatively, to home production. Expected contributions are estimated empirically.

Income Realization

The actual values of wages and home production for each household are realized at the end of each time period, after household activities have been decided. The distributions of both wages and home production depend on the characteristics of household members engaging in the activity. Wages are realized independently for each household member, depending on sex, age, and education. Wage income for the household is simply the sum of the wages of all working household members. Home production, unlike individual wages, depends on the combined labor input of multiple household members. Each household realizes a single value for home production, depending on the joint characteristics of all household members whose main activity is home production. I assume that individual wages and the value of home production follow log-normal distributions with distributional parameters determined by the characteristics of the household members participating in the relevant activity. The distributions for wages and home production are defined in equations 3.5 and 3.6.

In estimating the parameters for the home production distribution, the characteristics of all household members at home are considered. This is done with the expectation that very young children may reduce production efficiency, while older children may increase production efficiency. The parameters of the home production distribution are estimated nonparametrically. I enforced some restrictions on the estimation of home production in order to reduce the dimensionality of the problem. Specifically, the contribution of very young children does not depend on gender. For older children, only the first male and the first female are considered. Children in the lower-middle age range, from 6-12, are considered neutral to home production, as they may make marginal contributions but will also require oversight. Children in this age range, therefore, do not enter the home production function.

The timing of income realization and consumption in this model differs from that of many dynamic models. As described below, all income is consumed in the same period in which it is realized. The household has no ability to save or otherwise accumulate assets from wage earnings or home production. Exogenous shocks to income in the current period have an impact on current utility, but they do not carry forward to the value function of the next period. This formulation disallows a characteristic of intertemporal choice that is the focus of many dynamic models, that is the way that current period shocks influence future state transitions and decisions. This model, instead, focuses on the accumulation of wealth in the form of human capital. Restricting the impact of stochastic income shocks simplifies the dynamics of the model, which focuses the analysis on the central intertemporal problem of educational achievement.

Household Utility

The household utility function and terminal value function are defined in equations 3.2 and 3.3. The utility function is composed of two additively separable functions. The first component, u_1, is defined on the stochastic components of household income, wages and home production. The second component, u_2, is defined on the characteristics and decisions of household members. The household is assumed to have full knowledge of u_2 when choosing the primary activity of each member. Finally, the terminal value function is defined on the characteristics of household members in the terminal period.

Before describing the components of household utility in detail, it is useful to review the motivation of the model and how that informs the structure of the utility and terminal value functions. The central feature of the household decision problem is the trade-off between current consumption and educational achievement. The relationship between current consumption and the choice of primary activity should operate through the wage and home production equations. A child attending school must sacrifice labor for wages or for home production. Immediate benefits from educational achievement may be realized in current consumption, if household members with more education earn higher wages. The primary impact of educational achievement, however, should be realized in the terminal value function. Parents may wish to sacrifice current consumption in order to send their children to school for purely altruistic reasons, so their children will succeed in life. Alternatively, parents may simply expect to share in the enhanced earnings of their children. In either case, these benefits will accrue only after the children reach adulthood, which is outside of the time frame considered here.

The primary opposing forces of the model, then, act through current period consumption, $u_1\left(\frac{y_t}{n_t}, \frac{p_t}{n_t}\right)$, and terminal period state variables, $TV(X_T)$. The remaining component of household utility, $u_2(X_t, D_t)$, captures other costs and benefits accruing from the current choice of activities that are not directly related to consumption. One term in this function, for example, captures the benefit of having an older female child at home when there is also a young child at home who may need supervision or care. The essential characteristic of these factors is that they should be temporary in nature. These factors should influence the timing of activities, rather than the long-run payoffs which are captured in the terminal value function.

In modeling household consumption, I make the simplifying assumption that the household derives utility directly from its production of cash wages and the value of home production. The model does not permit households to save between time periods or to convert assets between holdings in food and holdings in cash. Of course, this assumption precludes many of the complex strategies that households pursue in order to maximize and smooth consumption, including conversion between cash and food in market transactions. Indeed, it is plausible that the observed preference for cash transfers over food transfers in the current evaluation is related to a household's ability to transfer and store value in each of the two modalities.

Rather than modeling household savings behavior directly, I allow the two sources of income to be valued differently in the household utility function. Specifically, the household is assumed to exhibit constant relative risk aversion (CRRA) in both income from wages and income from home production. A separate relative risk aversion coefficient is estimated for each type of income. The different risk aversion coefficients are intended to reflect the relative strength of each type of income as a financial instrument, rather than an indication of the household's preference for consumption from each source.

For example, if cash income permits greater flexibility for the household in market transactions and savings, a household might derive greater utility from an unexpectedly large realization of wage income. This would imply that the household is less risk averse in cash income, relative to income from home production. Viewed only as a financial instrument, home production is both less storable and less fungible than cash. The fruits of home production can be converted to cash, but this implies a transaction cost of some type. A household may derive less utility from an unexpectedly large realization of home production, and will exhibit a greater degree of risk aversion to income from home production.

Considering the specific context of this evaluation, it is also plausible that rice would be the preferred instrument for savings. Households in rural Cambodia very likely do not have access to formal savings institutions. This characteristic, shared with rural areas in other developing countries, can make it difficult to save in cash. Whether these difficulties stem from community interference or intrahousehold commitment problems, they can be ameliorated through the use of an outside commitment mechanism (Ashraf et al., 2006). Accumulating savings in nonperishable food, such as rice, could serve as such a commitment device.

Cambodia's unique history and economic environment also influence preferences for saving in cash. Many contemporary Cambodians lived through the radical upheaval and instability imposed by the Khmer Rouge. The Khmer Rouge abolished all private property and market transactions. The regime effectively wiped-out cash savings through the elimination of the national currency, going so far as to physically demolish the Cambodian central bank. Consequently, Cambodia had no national currency from 1975 until 1979 (Jackson, 1989; Prasso, 2001). The Cambodian Riel was reintroduced in 1980, but has yet to gain wide acceptance. As of 2010, the Cambodian economy was 80% dollarized (Duma, 2011). Given this complex legacy, it is impossible to predict how rural households may value excess income in cash wages relative to home production. Estimating separate CRRA coefficients for each type of income will allow the model to capture some of these complexities while maintaining tractability.

The second component of household utility function is implemented as a linear combination of a vector of indicators, defined by $L_U(X_t, D_t)$, and the parameter vector, θ_U. The indicators defined by L_U are binary or count variables reflecting certain combinations of state variables and decisions that could modify the cost of a particular decision. As mentioned above, one term indicates the presence of a young child and an older female whose main activity is home production. Other terms represent the count of females attending primary and secondary school, and the count of any children attending school in seventh grade, the first year of secondary school.

The final component, $TV(X_T)$, is constructed in a similar fashion, as a linear combination of indicators and parameters. Indicators in the terminal value function are based on terminal state variables only. Activity choices, which last a single-period, should not effect the final continuation payoff, which covers an arbitrarily large number of future time periods. The household does not choose activities in the terminal period. The linear terms of the terminal value function reflect the distribution of educational achievement of male children and female children within the household.

State Transition

The final component of the household decision model that must be defined is the nature of transitions between states. Given the current definitions, household transitions are largely deterministic. Each child will progress one year in age each time period. Children who attended school advance one grade each time period. The only stochastic state transition occurs when a new child is born. Pregnancies and births are deterministically chosen by the household, but the resulting child's sex is stochastic. Each new child has a 50% probability of being of each sex. As discussed in sections 3.3 and 3.4, the assumption of deterministic progression through ages and states will be modified in order to accommodate different estimation strategies.

3.2.3 Model Solution

The household decision problem can be simplified by transforming it into a recursive Bellman equation. In recursive form, the optimal decision in each state can be solved sequentially. It is not necessary to solve for the full sequence of decisions from 1 to $T-1$ simultaneously. The forward-looking household value function and recursive Bellman equation are defined below for periods $t \in \{1, \dots, T-1\}$. The value function for period T, or terminal value function, has already been defined in equation 3.3.

$$V\left(X_t\right) = \max_{D_t^{T-1}} \; \mathbb{E}\left[\sum_{\tau=t}^{T-1} \; \beta^{\tau-t} U(y_\tau, p_\tau, X_\tau, D_\tau) \; + \; \beta^T TV(X_T) \;\middle|\; X_t\right]$$

$$V\left(X_t\right) = \max_{D_t} \; \mathbb{E}\left[U(y_t, p_t, X_t, D_t)|X_t, D_t\right] + \beta \, \mathbb{E}\left[V\left(X_{t+1}\right)|X_t, D_t\right] \tag{3.7}$$

As discussed in section 3.2.2, the household realizes and consumes wage income and household production in the same period. The decision over primary activities is made in a state of uncertainty with respect to current period income. Expected values must be calculated for both current-period utility and for the continuation value. Expectations over current-period utility are taken over income from wages and home production. Expectations over the continuation value are taken over transition probabilities. I have made the assumption that income realizations and transition probabilities are independent, after conditioning on the current period decision. This enables the separation of current-period expected utility from the expected continuation value in equation 3.7 and equation 3.8 below.

Solving the maximization problem requires evaluation of the value function at each potential decision available to the household. These values are calculated using the choice-dependent value function, defined in equation 3.8. The choice-dependent value function returns the total discounted expected utility flowing from a specific decision in the current period, followed by optimal decisions in all following time periods.

$$V\left(X_t, D_t\right) = \mathbb{E}\left[U(y_t, p_t, X_t, D_t)|X_t, D_t\right] + \beta \, \mathbb{E}\left[V\left(X_{t+1}\right)|X_t, D_t\right] \tag{3.8}$$

Evaluating equation 3.8 requires evaluation of both expectation operators. Expected utility is calculated over two continuous variables, which requires a double integral over wages and home production. Recall from equation 3.2 that the utility function is additively separable into $u_1\left(\frac{y_t}{n_t}, \frac{p_t}{n_t}\right)$, defined on the stochastic variables, and $u_2(X_t, D_t)$, defined on deterministic variables. This property of the utility function enables further separation of the expected utility function.

$$\begin{aligned}
&\mathbb{E}\left[U(y_t, p_t, X_t, D_t)|X_t, D_t\right] \\
&\qquad = \mathbb{E}\left[u_1\left(\frac{y_t}{n_t}, \frac{p_t}{n_t}\right)|X_t, D_t\right] + u_2(X_t, D_t) \\
&\qquad = \int_{p_t}\int_{y_t} u_1\left(\frac{y_t}{n_t}, \frac{p_t}{n_t}\right) f(y_t|X_t, D_t) g(p_t|X_t, D_t) + u_2(X_t, D_t) \qquad (3.9)
\end{aligned}$$

The expected continuation value is calculated over discrete transition states, and can be calculated as a probability-weighted sum. Conditional transition probabilities for each state, $\pi(X_{t+1}|X_t, D_t)$, represent the probability that a household will reach state X_{t+1}, given their state and decision taken in period t. These probabilities may reflect the uncertainty that a child may fail if enrolled in school or the sex of a new child. The nature of transition probabilities differ between the models presented in sections 3.3 and 3.4. Discussion of specific details is deferred to each respective section.

$$\mathbb{E}\left[V\left(X_{t+1}\right)|X_t, D_t\right] = \sum_{X_{t+1}} V\left(X_{t+1}\right)\pi(X_{t+1}|X_t, D_t) \qquad (3.10)$$

After each choice-dependent value function is evaluated, a smoothed value function can be constructed by supposing the existence of an unobservable stochastic factor that enters each choice-dependent value function. Letting $j \in \{1, \ldots, J\}$ index the feasible decisions of a household, unobservable components are represented by ε^j, and are assumed to follow a type-1 extreme value distribution. Unobservable factors are assumed to be known to the household decision maker but unobservable by the econometrician. Integrating over these unobservables results in the log-sum formula given in equation 3.11 below.

$$
\begin{aligned}
V_t^{\varepsilon}\left(X_t\right) &= \max_{j\in\{1,\ldots,J\}} \; V_t\left(X_t, D_t^j\right) + \varepsilon^j \\
\mathbb{E}\left[V_t^{\varepsilon}\left(X_t\right)\right] &= \log\left(\sum_{j=1}^{J} \exp\left(V_t^{\varepsilon}\left(X_t, D_t^j\right)\right)\right)
\end{aligned}
\tag{3.11}
$$

Maximum Likelihood Estimation

Given a set of parameter values, the expected value function can be calculated through backwards-induction. The expected value function can then be used, together with the observed data, to evaluate a multinomial logit likelihood function. Maximum likelihood estimation proceeds by iteratively calculating these two functions until the optimal parameters have been found. Following Rust (2000, p. 3-4), I use a combination of the BHHH and BFGS optimization algorithms when estimating the model parameters.

Numerical Considerations

Solving for the expected value function for this model requires evaluation of the expected values in equations 3.9 and 3.10. The expectation operator in equation 3.10 is taken over discrete outcomes, such the sex of a new baby. The expected value is simply a probability-weighted sum of each possible discrete outcome. This approach is not possible, however, for the continuous variables in equation 3.9. These continuous stochastic variables are realized each period, conditional on the household's current-period decision. The expected value over each of these variables is taken each period by numeric integration using Gauss-Hermite quadrature.

Typically, calculating the full expected value function is done through backwards induction through the state space beginning from the terminal period. The state space in this model is extremely large. Consider the state space defined by the three characteristics of a single child: sex, age, and education. Considering children aged 0 to 18, with 0 to 12 years of education, yields $2 \times 19 \times 13 = 494$ possible states. A k-child household requires 494^k states to represent all possible combinations of child characteristics. Representing all two-child states requires $244,036$ states, and three children requires over 100 million. Four children requires almost 60 billion states, and 5 children almost 30 trillion. This huge state space makes traditional backwards induction computationally infeasible for any reasonable household size.

Given the computational challenges involved in estimating this model, I pursue a strategy of model validation through the replication of existing results from the work of Todd and Wolpin (2006). Replication results are presented next, in section 3.3. Replicating the model of Todd and Wolpin produced some promising results, but did not result in a feasible estimation strategy for the current problem. I develop a feasible and robust estimation strategy and present preliminary results in section 3.4.

3.3 Model Validation

The model defined in section 3.2.2 draws heavily on the model developed by Todd and Wolpin (2006, hereafter TW). Few modifications of the current model are required in order reconcile the two models. In the discussion that follows, I will describe modifications that must be made to the current model in order to make it consistent with TW model. In doing so, I do not intend to convey the idea that the current model was developed independently. The previous work by TW has served as a guide and inspiration to the work described here. I proceed by describing differences between the formal definitions of the two models, followed by the details of my implementation approach to the TW model.

3.3.1 TW Model Definition

The modifications described below are undertaken in order to investigate the numerical properties of the model, and to determine whether a model of this type is appropriate for use with the current data. The resulting model is not intended to reflect the actual economic decision faced by households in rural Cambodia. Rather it is only intended to be numerically and computationally similar to the final model. As such, I will not discuss the economic implications of each modification. The results of this exercise will be useful in building a model that both accurately reflects the household decision problem and is computationally feasible to estimate.

The first modification is to eliminate home production as a source of income, as well as a potential choice of primary activity for parents. TW assume that the contribution of the married couple to household income is exogenous and stochastic. As in the current model, households cannot save or borrow. Household consumption is equal to household wage income, y_t, which is the sum of parental income and the earnings of working children.

A second difference concerns the treatment of adult children. In the TW model, children achieve independence at age 16, and separate from the household. In the current model, adult children remain in the household and participate in household earnings and consumption. In addition, the household derives value from the educational status of adult children in the terminal period. These differences in the maximum age of children have implications for the maximum level of educational achievement. In both models, children are allowed to attend school beginning at age 6. By the age of 16, a student can have attained a maximum level of education of 10th grade. The TW model allows for stochastic grade progression. There is some probability that children attending school will fail to progress to the next grade.

The next two differences concern the unobservable shocks and household types, and require additional changes to reconcile the two models. The first change is a simplification from the unobservable types estimated in the TW model, which allows parameters to vary over three different household types. For the purpose of simulating decisions, I use only the parameter estimates from the most prevalent household type. The second change is an alteration to the general scheme of unobservable shocks. TW assume that most unobservable shocks are drawn from a normal distribution. I, instead, follow the approach developed by Rust, assuming that unobservable shocks are drawn from a type-1 extreme value distribution.

The utility function developed by TW is broadly consistent with the one defined here, in equation 3.2, but does have three terms that are inconsistent with the current model. It is composed of a CRRA component defined on current wages, along with a linear combination of terms defined by the current state and decision. The additive linear terms incorporate features including utility derived from the number of children who have graduated from different grades, the psychic cost of sending a child to school who is behind in grade, and the value of having an older girl at home together with very young children. The terms that do not fit the current formulation are all interactions between household consumption and the characteristics of children in the household, such as the total number of children and their average grade attainment. A complete list of utility function terms can be found in table 3.5.

Auxiliary equations for parent wages, child wages, and probability of grade failure are estimated independently, before the primary likelihood maximization. Log wages are estimated using linear regression and failure probabilities are modeled in a logit specification. Age, sex, and educational achievement are included as independent variables. In order to avoid endogenous selection issues in estimating the wage equations, TW include an additional exogenous variable that is excluded from the estimation of the household decision parameters. Specifically, they include the distance from the household to the nearest provincial capital, which was calculated using GPS coordinates of each household. They impose the exclusion restriction that a household's distance from a major city impacts the potential wages that a household member can earn, but will not otherwise affect the household member's choice of main activity.

In validating the use of a discrete choice model in the current context, two major tasks were undertaken. First, decisions were simulated for all households in the estimation sample using the TW equation specifications and parameter estimates. Equations and parameters were reproduced both for the main discrete choice model and the auxiliary wage and grade progression equations. In replicating the wage equations I also computed household distances to major cities using using GPS data. I chose provincial capitals as the relevant major cities, and used the GPS coordinates defined by the GeoNames database (Wick, 2015). In considering the performance of the discrete choice model, I compare the simulated decisions against those produced by the TW model, rather than the observed decisions in the Cambodian data. There is no reason why the simulated outcomes would reflect the actual decisions from the Cambodian data, as they were produced using parameter estimates from a completely different setting.

In the second task of model validation, these simulated decisions were used to re-estimate the parameters of the model. The simulated data were used as if they were the observed decisions, and parameter values were recovered through maximum likelihood estimation.

3.3.2 Computation

Simulation of decisions using existing parameter estimates still requires solving the expected value function of the dynamic discrete choice model. The numerical challenges, as described in section 3.2.3, must be addressed. The solution implemented by TW estimates the value function at a subset of the state-space for each time period, then uses a flexible regression specification to predict expected values for states that were not sampled. This procedure is described in further detail in Keane and Wolpin (1994).

I do not pursue the regression approach used by TW, but rather attempt to estimate the full expected value function at all discrete points in the state space and approximating expected values for continuous variables only. The approach I pursue for this exploratory work considers the nature of the state space and transition process of the model in order to reduce the size of the state space that must be estimated. In the following discussion, I use the TW model definitions when considering the structure of the state space. Specifically, the maximum age of children in the household is 16, and the maximum educational achievement is 10th grade.

The calculations in section 3.2.3 are motivated by a simple data structure, where each characteristic of each child is treated as independent. The size of this naive solution could be reduced significantly by eliminating state space components that are impossible to realize. For example, a 5 year-old child cannot have obtained 10 years of schooling. Other aspects of the problem can be simplified by considering the structure of the transition matrix. Once certain characteristics are known about a household, the number of transition states is greatly reduced. For example, a child's sex is determined at birth and will not change. A child's age will change each year, but in a deterministic manner. Each child will age one year per year.

Once the age and sex profile of children in a household is established, the household has started on a deterministic transition path that necessarily excludes almost all other paths. Only the education levels of children in the household are affected by their actions. Conditional on the sex and age profile, then, and assuming a maximum educational achievement of 10th grade, a child may be in a maximum of 11 different states. These states may be further bounded by the child's current age. Assuming that children are not allowed to skip grades or enroll before the age of 5, the number of achievable educational states is limited by the number of years a child has been old enough to attend school. Formally, this upper bound is $\min(\max(age - 5, 0), 10)$.

In light of these considerations, the state space problem can be reduced by first considering the number of sex-age profiles, then considering the possible educational states. The number of possible sex-age profiles for a single child is $17 \times 2 = 34$. Excluding the possibility of identical twins, a household of k children would conform to one of $\binom{34}{k}$ profiles, and would progress through a maximum of 11^k educational states. Under this scheme, the complete state space of all four-child households would be of size $46,376 \times 14,641 \approx 6.8 \times 10^8$.

The state space has been reduced from roughly 60 billion to less than one billion, but it is still not computationally feasible. Another computational savings can be realized by recognizing that the full set of $46,376$ deterministic profiles may not be relevant for a specific empirical problem. Estimation of the dynamic discrete choice model is ultimately based on the actual observations in any data set. Any deterministic profile that is not present or reachable from the observed household states are irrelevant to the estimation of the model. The maximum size of the relevant state space of four-child households, then, is given by $N \times 14,641$, where N is the number of observations in the estimation sample. In the current sample of 1,376 households, the relevant state space is roughly 20 million states.

The initial solution presented here solves the expected value functions of every household independently. I limit household size to three children, and the maximum level of education to grade 10, so the maximum size of the relevant state space is roughly 1.8 million. In practice, the relevant state space is smaller for many households. If children have already reached a certain level of education, for example, any lower levels of education do not have to be considered.

This limitation is taken into account when performing the backwards-induction solution. A terminal state space is constructed for each household, based on their current state and the feasible decisions they will face throughout their lifetime. This household-specific terminal state space is greatly reduced from the set of all potential terminal states. The household-specific backwards-induction solution is still computationally intensive, but it is sufficient for the limited purpose of model validation undertaken with respect to the TW model.

One major drawback to the household-specific expected value computation is the irregularity of the state space and transition probability matrix. Each observed household in the estimation sample may have a unique state space and transition probability matrix. This makes it difficult to reduce computation time through precalculation of transition probabilities. It also complicates the indexing and storage of each household's expected value function.

Ideally, the characteristics of each household should map to an index in a storage array, where the expected value for that state is stored. Because the each household occupies such a small area of the full state space, however, it would not be efficient to create a single, global indexing scheme. Similarly, a single transition matrix, giving zero probability to impossible transitions, would not be efficient in terms of memory allocation.

In implementing this approach I favored flexibility over efficiency. The model is defined by rules governing the allowable activities of individual household members, and the associated transition probabilities for each household member. The decision space and transition probabilities for the household are constructed according to the individual definitions. The backwards-induction algorithm only considers those states that are reached with positive probability. Rather than precomputing and storing a potentially very large transition probability matrix, transition states and probabilities are constructed from individual transition probabilities and decision rules at each state. Expected values for each state are stored in a hash table, indexed by the household state vector.

The flexibility of this implementation was useful for exploratory work, and for comparing results with those from the TW model. The model was quite costly, however, in terms of computation. The most obvious cost is the requirement to calculate separate value functions for each household. There is also a significant amount of computational overhead in recalculating transition states and probabilities each time they are needed for a specific state and decision. Finally, there were costs incurred by the choice of data storage structure. The hash table did result in a smaller memory footprint, as only the expected values from positive-probability states were stored. It also required some additional computational overhead to transition between state vectors and the internal storage index. Additionally, there are significant slowdowns if the initial memory allocation is insufficient, resulting in storage collisions and potentially requiring the table to be rebuilt.

3.3.3 Simulated Activities

Results from the initial simulation are presented in table 3.3, along with the observed distribution of child activities. These results can be compared to the observed and simulated outcomes from Todd and Wolpin (2006), presented in table 3.4.

There are several reasons why these simulated outcomes may not match those from Todd and Wolpin (2006) precisely. First, these outcomes are conditional on the observable characteristics of the households in the estimation sample. There will always be differences in observable characteristics, though the results in table 3.1 do

Table 3.3: Observed and Simulated Activities by Age and Sex

	Actual			Predicted		
Age	School	Work	Home	School	Work	Home
Boys						
6	0.792	.	0.208	0.979	.	0.021
7	0.936	.	0.064	0.991	.	0.009
8	0.954	.	0.046	1.000	.	.
9	0.952	.	0.048	0.985	.	0.015
10	0.968	.	0.032	0.986	.	0.014
11	0.953	.	0.047	0.943	.	0.057
12	0.898	.	0.102	0.781	0.164	0.055
13	0.889	0.062	0.049	0.643	0.204	0.153
14	0.615	0.108	0.277	0.407	0.444	0.148
15	0.400	0.318	0.282	0.415	0.492	0.092
Girls						
6	0.874	.	0.126	0.986	.	0.014
7	0.939	.	0.061	0.991	.	0.009
8	0.981	.	0.019	0.965	.	0.035
9	0.995	.	0.005	0.987	.	0.013
10	0.982	.	0.018	0.941	.	0.059
11	0.985	.	0.015	0.918	.	0.082
12	0.905	0.029	0.067	0.767	0.068	0.165
13	0.795	0.096	0.108	0.667	0.105	0.229
14	0.549	0.216	0.235	0.590	0.205	0.205
15	0.382	0.273	0.345	0.392	0.392	0.216

suggest that the differences are not large. Furthermore, household size was limited to five individuals in this simulation. This was done to reduce the computational burden. Given the large potential differences between the data sources, simulated outcomes from the current model are quite close to those from the original TW paper.

Table 3.4: Observed and Simulated Activities by Age and Sex(Todd and Wolpin, 2006)

	Actual			Predicted		
Age	School	Work	Home	School	Work	Home
Boys						
6	0.933	.	0.066	0.923	.	0.077
7	0.981	.	0.019	0.980	.	0.020
8	0.987	.	0.013	0.980	.	0.020
9	0.994	.	0.006	0.979	.	0.021
10	0.982	.	0.018	0.974	.	0.026
11	0.977	.	0.023	0.964	.	0.036
12	0.885	0.021	0.094	0.846	0.039	0.115
13	0.780	0.084	0.136	0.736	0.078	0.186
14	0.677	0.157	0.166	0.619	0.191	0.190
15	0.490	0.276	0.235	0.520	0.251	0.229
Girls						
6	0.965	.	0.035	0.942	.	0.058
7	0.976	.	0.024	0.968	.	0.032
8	0.989	.	0.011	0.976	.	0.024
9	0.991	.	0.009	0.975	.	0.025
10	0.979	.	0.021	0.970	.	0.030
11	0.969	.	0.031	0.948	.	0.052
12	0.896	0.007	0.097	0.854	0.020	0.126
13	0.726	0.028	0.245	0.676	0.025	0.299
14	0.582	0.089	0.329	0.566	0.092	0.342
15	0.419	0.123	0.458	0.402	0.157	0.442

3.3.4 Parameter Recovery

Results from the parameter recovery exercise are presented in table 3.5. Maximum likelihood estimation is more computationally intensive than producing simulated outcomes for known parameters. The parameter recovery exercise, therefore, was conducted after limiting households to a maximum of four members. New outcome data was simulated for four-member households, and the MLE was conducted on the resulting data set. The first column of table 3.5 gives the parameter values estimated by TW. These are the true parameters for the simulated data.

MLE was conducted using several different constraint sets. In table 3.5, parameters with a dash in place of a standard error estimate were constrained to their true values in that specification. Starting values for freely estimated parameters were generally set at 0, not their true values. The only exception is the CRRA parameter, which was given a starting value of 1 to avoid division by zero.

With each of the three MLE specifications, constraints were relaxed, allowing the free estimation of additional parameters. The first specification estimates only the first four parameters, which are all in terms that include the continuous consumption variable. All other parameters were constrained to their true values, which is indicated by the dash in place of each standard error estimate.

The second specification estimates most of the remaining linear parameters from the utility function. Only parameters related to family size and pregnancy are constrained to their true values. These parameters were problematic to estimate in the current exploratory model. This is likely due to the limited variation in household size, which is restricted to four members. Most households already have four members, so there is little variation to estimate these parameters. The problematic nature of the family size and pregnancy parameters is apparent in the final specification, which does not impose any constraints. Many of the family size and pregnancy parameters are estimated with zero standard errors.

Table 3.5: Parameter Recovery

Parameters	True Values	Estimated		
		(1)	(2)	(3)
Crra coefficient	.87	.88	.88	.87
	(.019)	(.011)	(.026)	(.036)
Consumption	-.0014	.034	.028	.14
* Children net of current birth	(.02)	(.073)	(.092)	(.17)
Consumption	-.11	-.25	-.34	-.77
* Children age 12-15	(.023)	(.29)	(.36)	(1.1)
Consumption	.0017	.0014	.0029	.0025
* Average grade attainment	(.0092)	(.00036)	(.0015)	(.0019)
Children under 3	474	474	474	2.2
	(7,368)	-	-	(1.3e+06)
Children under 3 squared	-812	-812	-812	2.2
	(1,570)	-	-	(0)
Number of children	434	434	434	66
	(1,135)	-	-	(0)
Number of children squared	1,111	1,111	1,111	197
	(67)	-	-	(46,832)

Standard Errors in Parentheses *Continued on next page*

Estimates with no s.e. were constrained to true values

Table 3.5 – *Continued from previous page*

Parameters	True Values	Estimated		
		(1)	(2)	(3)
Number of children	135	135	8.8	87
* Average grade attainment	(32)	-	(78)	(76)
Children who have completed	12	12	-14	5
grade 6 or higher	(36)	-	(22)	(22)
Children who have completed	147	147	105	120
grade 9 or higher	(43)	-	(70)	(67)
Average grade attainment	276	276	235	190
	(88)	-	(52)	(53)
Pregnancy in first year marriage	-31,834	-31,834	-31,834	2.6e-11
	(2,194)	-	-	(0)
Pregnancy at age 20-24	-1,126	-1,126	-1,126	4.2e-12
	(1,111)	-	-	(0)
Pregnancy at age 25-29	-3,072	-3,072	-3,072	13
	(1,869)	-	-	(0)
Pregnancy at age 30-34	-24,414	-24,414	-24,414	-12
	(3,209)	-	-	(1.4e+06)
Pregnancy at age 35-39	-27,204	-27,204	-27,204	11
	(3,905)	-	-	(0)

Standard Errors in Parentheses *Continued on next page*

Estimates with no s.e. were constrained to true values

Table 3.5 – *Continued from previous page*

Parameters	True Values	Estimated		
		(1)	(2)	(3)
Pregnancy at age 40-43	-59,672	-59,672	-59,672	-11
	(8,182)	-	-	(153975)
Current pregnancy	-37,002	-37,002	-37,002	1.0e-11
* Previous pregnancy	(2,128)	-	-	(0)
Children age 12 attending school	187	187	275	182
	(206)	-	(348)	(344)
Children attending secondary school	.13	.13	.13	.13
* Distance to secondary school	(.0076)	-	-	-
Children attending grade 10	1,061	1,061	-35	26
	(250)	-	(2,914)	(2,851)
Boys attending grade 10	674	674	173,047	19
	(316)	-	(6,314)	(10,388)
Boys age 12-15, attending school	-79	-79	-297	-116
in grade 6 or higher	(246)	-	(504)	(489)
Girls age 12-15, attending school	64	64	97	27
in grade 6 or higher	(274)	-	(506)	(501)
Children age 12-15, attending school	84	84	-4.9	76
and not behind	(188)	-	(878)	(816)

Standard Errors in Parentheses *Continued on next page*

Estimates with no s.e. were constrained to true values

Table 3.5 – *Continued from previous page*

Parameters	True Values	Estimated		
		(1)	(2)	(3)
Children age 12-15, attending school	-266	-266	-156	-248
and behind 1 or more years	(209)	-	(715)	(675)
Children age 12-15, attending school	-32	-32	-28	5.1
and behind 2 or more years	(171)	-	(314)	(310)
Children age 12-15, attending school	-11	-11	-214	-212
and behind 3 or more years	(187)	-	(229)	(232)
Boys age 12-15, attending school	-156	-156	-93	70
and not behind	(202)	-	(1,298)	(2,507)
Boys age 12-15, attending school	32	32	-73	-196
and behind 1 or more years	(198)	-	(601)	(585)
Boys age 12-15, attending school	-47	-47	71	50
and behind 2 or more years	(195)	-	(390)	(382)
Boys age 12-15, attending school	-139	-139	-179	-174
and behind 3 or more years	(230)	-	(324)	(319)
Children age 12-15, at home and	475	475	539	411
behind 1 or more years in school	(327)	-	(2,272)	(2,157)
Children age 12-15, at home and	83	83	355	285
behind 2 or more years in school	(351)	-	(783)	(787)

Standard Errors in Parentheses

Continued on next page

Estimates with no s.e. were constrained to true values

Table 3.5 – *Continued from previous page*

Parameters	True Values	Estimated		
		(1)	(2)	(3)
Children age 12-15, at home and	284	284	332	259
behind 3 or more years in school	(255)	-	(509)	(511)
Boys age 12-15, at home and	-641	-641	-33	-66
behind 1 or more years in school	(414)	-	(7,219)	(3,479)
Boys age 12-15, at home and	-87	-87	-38	-67
behind 2 or more years in school	(453)	-	(1,090)	(1,094)
Boys age 12-15, at home and	-265	-265	-146	-124
behind 3 or more years in school	(327)	-	(760)	(759)
Boy age 6 at home	1,422	1,422	14	8.9
	(1,723)	-	(3,108)	(1,730)
Boy age 7 at home	644	644	-12	-11
	(3.4e+06)	-	(29,070)	(12,188)
Boys age 8-11 at home	808	808	3	-25
	(4,326)	-	(721)	(718)
Boys age 12-15 at home	1,065	1,065	9.9	-48
	(555)	-	(6,875)	(2,722)
Girl age 6 at home	1,676	1,676	-28	-6.8
	(494)	-	(6,277)	(13,093)

Standard Errors in Parentheses

Continued on next page

Estimates with no s.e. were constrained to true values

Table 3.5 – *Continued from previous page*

Parameters	True Values	Estimated		
		(1)	(2)	(3)
Girl age 7 at home	1,072	1,072	-48	-16
	(1,281)	-	(4,747)	(7,860)
Girls age 8-11 at home	1,179	1,179	140	163
	(564)	-	(723)	(715)
Girls age 12-15 at home	969	969	641	512
	(499)	-	(2,174)	(2,063)
Girls age 12-15 at home	3.3	3.3	18	-13
* Number of children under 6	(423)	-	(1,394)	(1,310)
Girls age 14-15 at home	287	287	-24,599	-22
* Number of children under 6	(571)	-	(0)	(2,800)
Converged		1	0	0
Log Likelihood		-563.4	-808.7	-565.3
Iterations		18	50	50
Total Time (hours)		4.2	18.3	22.5
Ave Time (minutes)		14	22	27

Standard Errors in Parentheses

Estimates with no s.e. were constrained to true values

3.4 Estimation

The parameter recovery exercise using the TW model demonstrated that the current data and model are consistent with a dynamic discrete choice estimation framework. Estimation of model parameters, however, will require a more computationally efficient expression of the model. I develop and estimate such a model in this section.

The next section describes the modifications that were made specifically for computational purposes. These changes necessitate further changes to the specifications for wage earnings, home production, household utility, and terminal value. The final specification for each of these equations is described is section 3.4.2.

3.4.1 Computation

The computational time requirements for the exploratory model are high, but not prohibitive. The parameter recovery model, however, is limited to four household members: two parents and two children. This limitation greatly reduces the ability of the model to capture dynamics between the activities of children within the household. Ideally, I would like to increase the maximum household size to at least six members. To do so, I will have to reduce time required to compute the backwards induction solution.

There are two major approaches I follow in addressing the computational time requirements of the backwards induction solution method. First, I recoded the solution algorithm in C, a lower level programming language, which greatly increased the speed of the computation, regardless of the complexity of the calculation. Second, I reduced the dimensionality of the problem by reducing the number of future states that must be considered.

Parent Characteristics

Several changes to the model were made in order to reduce the size of the state space. First, parental characteristics were largely eliminated from the state vector. Given the initial sample restriction to nuclear households, many parental characteristics are of limited importance to the model. For example, parental education and father's age only enter in the parental wage equation. The only parental characteristic that is critical to the model is the age of the mother, which determines the mother's fertility status. The mother's age is also the primary time index of the model, determining the number of periods remaining before the terminal decision period. Given that mother's age is captured in the time index of the model, parental characteristics can effectively be eliminated from the state space.

The decision-space of parents was also reduced to a single component of the decision vector. The joint activities of parents are indexed from 1-5, representing different combinations of wage work, home production, and pregnancy. Only the combination of activities is differentiated, the activity of each parent is not represented individually. In theory, this implies that the parents are completely substitutable. In practice, this only has implications for the split decision Work/Home. This decision combination could represent the mother at home and the father working, or the father working and the mother at home. Any parent at home is assumed to be working in home production.

Binning

A second, larger reduction of the state space was accomplished through binning child characteristics. Both age and educational attainment were condensed into four categories, which are listed in table 3.7. Age bins were chosen to be of equal size while

Table 3.6: Parent Activity Decisions

Index	Activity
1	Work/Work
2	Home/Home
3	Work/Home
4	Preg/Work
5	Preg/Home

respecting the previously defined age-restrictions on the choice of a child's primary activity. The age restrictions set the minimum ages for school attendance and wage work at 6 and 12 years, respectively. In the previous model, the maximum age for school attendance was 15, while in the current model children are allowed to attend school until age 17. School attendance by older children is consistent with the observed data from Cambodia.

Transition probabilities between age-categories were constructed based on the number of years covered by each bin. Each age category covers 6 years, so a child will remain in the same age category with probability $5/6$, and will progress to the next age category with probability $1/6$. Over any 6 year time span a child is expected to advance by 1 age category. Once a child reaches the final age category, he or she will remain in that age category with probability 1.

Unlike age categories, education categories were not chosen to be of uniform size. Instead, education categories were chosen to reflect particular states or milestones in the educational system. Educational categories reflect educational attainment, or completed grades. For example, a child who has educational attainment of 6th grade and whose main activity is education is actually enrolled in 7th grade, the first level of secondary education.

The lowest category includes children with no education up to completion of 2nd grade. The next category includes children who have completed grades 3-5. If they are attending school, the children in this category will be enrolled in grades 4-6, and thus may be eligible for the primary school scholarship being evaluated here. Completion of grade 6 is retained as a unique category, as it represents the educational milestone of primary school completion. Continued school participation once a child has reached this category indicates some attendance in secondary school. The final educational category reflects the completion of one or more grades in secondary school.

Transition probabilities between educational categories are constructed in the same manner as those for age categories. The probability of transition to the next level is simply the inverse of the number of years covered by the current category. Of course, this transition is dependent on school attendance, but there is no probability of failing, if the child attends. A child whose main activity is not school attendance will remain in his or her current education category with probability 1. Because the third category covers a single year, a child attending school after completing primary school will progress to the final educational category with probability 1. Any child reaching the final educational category will remain there with probability 1, regardless of his or her main activity. All child characteristics and transition probabilities are listed in table 3.7.

Probabilistic transitions between age and education categories introduce the undesirable possibility that a child's education could out-pace his or her age. It is unrealistic that a primary school aged child would be in secondary school, for example. Education level should be limited by age, but doing so directly would distort the transition probabilities for education. If a child could only advance in education after advancing in age, progression through the grades would be slowed. Furthermore, if there was no probability of advancement, the family would have no incentive to send

Table 3.7: Child Characteristic Categories

Sex

Index	Sex	Transition Probabilities	
		Remain	Advance
0	Male	1	0
1	Female	1	0

Age

Index	Ages (Years)	Transition Probabilities	
		Remain	Advance
0	0-5	0.167	0.833
1	6-11	0.167	0.833
2	12-17	0.167	0.833
3	18+	1	0

Education

Index	Highest Grade Completed	Conditional Transition Probabilities	
		Remain	Advance
0	0-2	0.333	0.667
1	3-5	0.333	0.667
2	6	0	1
3	7+	1	0

the child to school until after the child advanced in age. To address this problem without distorting educational transitions, I introduce the idea of the effective level of education. Effective education is limited by the child's current age category. This number is used to calculate the potential wage earnings of the child and the terminal value continuation payoff. The child's actual education level is allowed to progress independently of age, and is used to calculate the household's index value and transition probabilities.

The number of potential states for a single child based on sex, age, and eduction, is $2 \times 4 \times 4 = 32$. Any household of four children can be fully represented in $32^4 = 1,048,576$ states. Unlike the previous implementation of the model, the issue of many deterministic and exclusive age-sex profiles is not present. Age transition is no longer deterministic, so households will be able to transition between various age-profiles. When computing the expected value function, it will no longer be the case that many states are unreachable and irrelevant to the households in the estimation sample. This model does retain deterministic sex-profiles, as children cannot transition between states, but, with only $2^4 = 16$ profiles, it is likely that each one will be relevant.

There are still efficiency gains to be had from careful consideration of the state space. In the binned implementation, the problem is not irrelevant states, but redundant states. With probabilistic age transitions, the same essential household can be represented in different ways. For example, a household may be composed of two boys, both without any education, and in the same age category. One of the boys may progress in age, while all other characteristics remain unchanged. This household could be represented equally well by ordering the older boy first or second. All permutations of the same set of children are essentially the same state. Any household with k children, then, can be represented as a multiset of cardinality k drawn from the set of integers 0 to 31. The size of the state space for all such households is given by the multiset coefficient, $\left(\!\binom{32}{k}\!\right)$. In this context, a multiset of cardinality k is simply a k-combination drawn with replacement, and the multiset coefficient is the standard binomial coefficient adjusted for replacement: $\left(\!\binom{n}{k}\!\right) = \binom{n+k-1}{k}$. The size of the state space of all households with 4 children is $\left(\!\binom{32}{4}\!\right) = 52,360$.

Indexing

Unlike the household-specific approach to calculating the expected value function, every possible 4-child household will be represented as one of these 52,360 states. Rather than storing these in a hash table, it is worthwhile to develop an indexing scheme linking every possible household state to an integer representation between 0 and 52,359. The first step of this procedure is to convert each child-state into an integer. This is accomplished by mapping the age and education categories to integers 0-3, and the sex categories to 0-1. The characteristics are ordered beginning with sex, followed by age and education.

Any 3-digit combination can be treated as an base-4 integer, ranging from 000 to 133. Converted to base-10, these integers are 0-31, providing a unique integer to represent any combination of characteristics. By placing the 2-category characteristic, sex, in the left-most position, all combinations of child characteristics can be represented as digits of a single base (4). This indexing scheme can be extended to include characteristics with varying numbers of categories. In that case, the characteristics of each child would be encoded as a mixed radix number, in which the base of each digit varies according to the number of categories for the relevant characteristic(Knuth, 1981).

After the individual child characteristics are mapped to the integers 0-31, each household can be considered an unordered multiset. Before mapping multisets to integers, each multiset is sorted from left to right in descending order using a simple bubble sort algorithm.[1] Once the multisets are sorted, they can be mapped to the set of integers from 0 to 52,359. The approach I use for mapping multisets is based on the combinatorial number system. The combinatorial number system defines a bijection between k-combinations and non-negative integers, for a fixed k (Buckles and Lybanon, 1977; Knuth, 2005). In a standard positional numeral system, the value of each digit is equal to the digit multiplied by its base raised to a power, p. The power, p, is determined by that digit's position, starting at zero and increasing to the left. For example, consider the digit 4, in the number 345, in the base-10 number system. The digit is in the 2nd position, so $p = 1$ (because powers begin at zero), and the value of that digit in that position is 4×10^1. In a combinatorial number system, digits must be sorted descending from left to right. The value of a digit, d, is equal to the number of combinations of length l that can be drawn from a set that contains d elements. For each digit, l is determined by the digit's position, starting at one and increasing to the left to a maximum of k, the number's base. For example, the four digit combinatorial number, (28, 15, 3, 1), can be converted to its corresponding integer as: $\binom{28}{4} + \binom{15}{3} + \binom{3}{2} + \binom{1}{1} = 13,109$

Combinatorial number systems are traditionally defined based on combinations taken without replacement, but the system can be extended to apply to combinations taken with replacement, or multisets. Extending the combinatorial number system to multisets is straight-forward, and simply requires replacing binomial coefficients with multiset coefficients in the expression above. Ordering multisets according to their

[1]The bubble sort or sinking sort algorithm is inefficient compared to many other sorting algorithms(Knuth, 1998). Bubble sort is very efficient, however, when the list to be sorted is already sorted. It is a reasonable choice in the current application, as the list under consideration is only four elements long, and will frequently not require any sorting. The combinations being sorted are primarily generated as transitions from sorted combinations. Because sex is the dominant digit when converting from base-4, reordering will only be possible when there are multiple children of the same sex and age category in the same household.

natural number equivalents produces a lexicographical ordering of multisets, where lexicographical ordering is determined by comparing multiset elements in descending order (left to right under the current sorting regime).

This discussion has focused on mapping integer values to households with a fixed number of children. The state space, however, must include households of any size up to a maximum number of children. If the maximum number of children is 4, the size of the state space must be: $\left(\!\binom{32}{1}\!\right) + \left(\!\binom{32}{2}\!\right) + \left(\!\binom{32}{3}\!\right) + \left(\!\binom{32}{4}\!\right) = 58,904$ The indexing scheme must also allow for households of varying size, to positively map households to indices. This is accomplished with a simple index offset for households with more than one child. For single child households, no offset is required, and the indices range from 0 to 31. The indices of two-child households begin at $\left(\!\binom{32}{1}\!\right) = 32$. The indices of three-child households begin at $\left(\!\binom{32}{1}\!\right) + \left(\!\binom{32}{2}\!\right) = 560$. Finally, the indices of four-child households begin at $\left(\!\binom{32}{1}\!\right) + \left(\!\binom{32}{2}\!\right) + \left(\!\binom{32}{3}\!\right) = 5,984$.

To summarize, household state vectors are indexed according to the following steps. First three characteristics of each child, which are stored as integers from 0-3, are combined into a single three-digit number, in base-4. These numbers, ranging from 000-133, are converted into base-10 integers ranging from 0-31. For each household of k children, base-10 integers are treated as a k element multiset drawn from the set of integers 0-31. Multisets are themselves converted to integers using a modified combinatorial number system as described above. Finally, the indices of households with more than one child, or $k > 1$, are offset by the count of all households of size $k - 1$ or smaller.

Transition Matrix

Careful mapping and indexing of the state space has reduced the dimension of the space from 1,048,576 states to 58,904. This is a significant reduction in computational requirements. The next aspect of the model to consider is the mapping of states to allowable decisions, and from decisions to transition states.

The expected value calculation for a single state requires iteration through each possible decision that could be taken from that state, as well as through every potential transition state following each decision. Rules regarding the possible decisions for each state, as well as transition probabilities for each decision, are defined for individual household members, rather than for the household as a unit. A single expected value calculation, then, requires iteration through the members of the household to determine allowable decisions, followed by iteration through all allowable decisions. For each decision, transition states and probabilities must be calculated, again by iterating through each household member and aggregating transition probabilities. In the previous implementation of this model, feasible decisions and transition probabilities were calculated as needed throughout the backwards induction solution algorithm. These calculations add to the computational overhead of the model solution.

The alternative to recalculating transition probabilities for every state is to precalculate and store them in a transition probability matrix. Because transition probabilities are linked to the primary activity of each household member, a complete transition probability matrix must map from the state-decision space to the state space. The maximum size of the decision space for any state includes 5 potential parental activity combinations, and 3 potential activities for each child. This results in $5 \times 3^4 = 405$ potential decisions, and $405 \times 58,904 = 23,856,120$ state-decision combinations. A complete transition matrix, then, would have dimensions of $23,856,120 \times 58,904$, and would contain 1.4×10^{12} cells. From most states, the full range of 405 potential decisions is not available. Children may be too young or too old to attend school, for example. Construction and computation using the full transition matrix would require the storage of, and computation over, a large number of zero-probability transitions.

As discussed in Rust (2000), a sparse transition matrix provides opportunities for efficiency gains relative to standard matrix multiplication routines. The transition matrix in Rust's bus engine replacement problem takes the form of a well-structured band matrix. From any initial state, all potential transition states are represented in a contiguous and bounded block of neighboring cells within the same row of the transition matrix. The maximum size of the required contiguous block determines the bandwidth, b, of the band matrix. All non-zero transitions, then, lie within b columns of the diagonal of the transition matrix. Rust takes advantage of this banded structure by partitioning the transition matrix into diagonal and off-diagonal blocks, where the elements of the off-diagonal blocks are mostly zero and all relevant transition information is contained in the diagonal blocks. Subsequent computation can be performed using the diagonal blocks only.

The transition matrix in the current problem does not have any regular pattern of sparsity. Matrix indices in this problem reflect a mixture of characteristics and position in the state vector. There is little or no sense in which states that have neighboring index values are "close" in terms of child-characteristics. From any given state and decision the index values of transition states will generally not be contained in a contiguous, dense interval. There is no simple or elegant solution to exploit the sparse structure of the transition matrix of this problem. But neither can the transition matrix be used in its sparse form, due to the excessive memory requirements. Storing transition probabilities as double-precision floating point numbers would require $8 \times 23,856,120 \times 58,904 \approx 10^{13}$ bytes, or 10,000 Gb of memory.

Instead, I implement a brute force approach to precalculate the full transition matrix and store the information in a dense array. The brute force approach uses individual decision rules to iterate through all feasible decisions of each state. For each state-decision combination, I construct the full $1 \times 58,904$ row vector containing the transition probabilities for next-period states. Each of these row vectors represents one row of the full transition matrix. The majority of entries in the row vector will be zero. Instead of storing the entire row vector, only the transition probability and state-index number are stored for those transition states with positive probability. Transition probabilities and state-index numbers are stored in two three-dimensional arrays, indexed indexed according to initial state, decision number, and transition state number. Dense storage of transition information is achieved by allowing the sizes of the second and third dimensions to vary, allocating only enough memory to store transition information for those states with positive transition probabilities.

The first dimension of each three-dimensional array is 58,904, to match the number of states. The size of the second dimension varies by state, and is equal to the number of allowable decisions from that state. The size of the third dimension varies by state and decision, and is equal to the number of positive-probability transition states that may be reached from each state-decision combination. For a given state, the decision index may run from 0 to 10, while in another state it may run from 0 to 28. Similarly, the first allowable decision from a given state may map to 4 transition states, while the second decision maps to 10 transition states.

Two three-dimensional arrays are utilized to store transition information. The first stores all positive transition probabilities, and the second stores the index values of each corresponding transition state. Two additional arrays store the dimension bounds. The first dimension-storage array is one-dimensional $(58,904 \times 1)$, and stores the number of allowable decisions for each state. This array provides the size of the second-dimension for each state. The second dimension-storage array is indexed by state and decision number. This array stores the number of transition states for each state-decision combination, and provides the size of the third dimension for each state-decision combination.

Using variable-dimensional arrays to store transition data enables the precalculation of all transition probabilities, while avoiding the excess memory consumption and computation implied by a large and sparse transition matrix. Considering feasible decisions only reduces the number of state-decision combinations under consideration from 23,856,120 to 1,411,754. Similarly, the number of transition states under consideration is reduced from 1.4×10^{12} to 3.8×10^{7}. The memory requirements to store transition probabilities in double precision and transition state indexes as integers are approximately 300 Mb and 150 Mb, respectively.

Construction of the dense transition arrays is not trivial, as the variable dimensions are not known ahead of time. Memory must be allocated dynamically according to the storage requirements of each state and decision. Each storage array is initialized as a one-dimensional $(58,904 \times 1)$ array of pointers. For each state, the number of feasible decisions is calculated based on the characteristics of household members. This calculation determines the size of the second dimension for that state. The dimension is stored for later use, and memory is allocated to store state-decision data. These data will include the number of transition states in a two-dimensional array, as well as additional pointers that will be used in the construction of three-dimensional arrays.

For each decision, transition probabilities are calculated and stored for all transition states. The number of positive probability transitions are counted, which determines the size of the third-dimension for that state-decision. Memory is allocated and linked to the corresponding pointer from the second dimension. Transition probabilities and corresponding transition state indices are stored in the memory allocated to the third-dimension. After iterating through all states, decisions, and transition states, the arrays storing transition data are written to disk. These transition data are read from the disk whenever the model is solved through backwards induction.

The process of writing and reading the dense transition matrices to and from the disk is necessary in order to use them in future calculations, but it also serves a useful secondary purpose. Because the array dimensions are not known ahead of time, memory allocation is performed in a piecemeal fashion. Small blocks of memory are allocated as needed as the algorithm iterates through states and decisions. The data for each array is likely inefficiently stored, as a fragmented collection of small memory blocks. When the data are written to disk, each array is written sequentially, as if it were stored in a single, continuous block of memory. Data are read back into memory in the same sequential fashion, defragmenting the memory for each array in the process. Continuous blocks of memory are allocated to store the data from each array before the data are read in sequentially from the disk. After all the data have been read into memory, the pointer offsets for two and three dimensional arrays must be set. The second dimension pointers are updated using pointer arithmetic based on the running sum of the number of decisions available in each state. Similarly, the third-dimension pointers are updated using pointer arithmetic based on the running sum of the number of transition states following each state-decision.

Precalculation of the transition matrices greatly speeds the backwards induction solution of the model. One component of the savings is the reduction of conversion between state indices and state vectors. For each time period, the backwards induction algorithm traverses through indices, while all transition states and probabilities are calculated based on long-form state vectors. Calculation on the fly requires repeated conversions from index numbers to state vectors and from state vectors to index numbers. These conversions can be completely eliminated from the backwards induction solution by precalculating the expected utility for each state-decision combination. For a given set of parameters, expected utilities are calculated and stored in a two-dimensional array, indexed by state and decision. The size of the second-dimension varies according to the number of allowable decisions in each state, and the indexing scheme is identical to the state-decision indexing of the transition matrices.

With state-dependent utilities precalculated and stored, the backwards induction solution can be calculated entirely on precalculated information in the transition arrays and the choice-dependent utility array. The expected value at any state is calculated using the log-sum formula, which requires a summation over the set of feasible decisions from each state. This can be accomplished by looping over the second dimension of the utility array and the transition arrays. For each state-decision, the choice-dependent continuation value requires the calculation of a weighted sum of the expected values of all possible transition states. This can be accomplished by looping over the third dimension of the transition arrays, which give the probability and index number of each potential transition state. The backwards induction algorithm has already computed the expected value for every state in the transition time period, so the values can be easily retrieved from the model solution using the index number of each transition state.

3.4.2 Estimation Results

Auxiliary estimates for wages and home production are calculated before the main optimization routine. These are treated as stochastic variables that enter the household utility function in the main decision model. The expected value of the household utility function is calculated using Gaussian quadrature to estimate the integrals over total household wages and home production. This approach requires estimates of the distributional parameters for wages and home production, which are assumed to follow log-normal distributions.

Distributional parameters are estimated conditional on the characteristics of household members participating in each activity. I pursue a nonparametric ap-

Table 3.8: Parent Wage Estimates

	Log USD per year	
Activity	Sample Mean	Sample Variance
Work/Work	6.30	0.976
Home/Home	.	.
Work/Home	5.74	1.344
Preg/Work	6.01	1.175
Preg/Home	.	.

proach to estimating the mean and variance of each conditional distribution. Cells are defined by the combinations of characteristics of individuals participating in each activity. This approach is enabled by the current binned implementation of the model, which has reduced the number of combinations of individual characteristics. For each cell, the mean and variance are estimated by the sample mean and sample variance of log wages or log home production.

Wages are estimated separately for parents and children. As discussed previously, parents' characteristics have been largely eliminated from the model, so cells are defined by parental activities only. Wage earnings are only realized if at least one parent is working. Estimates are presented in table 3.8, with missing values in the empty cells where neither parent is working.

Child wages are estimated independently for each combination of individual child characteristics. The model assumes that children can work for wages beginning at age 12, so wage parameters are estimated for the older two age categories only. Considered together with sex and education, the characteristics of working children define sixteen cells. The sample mean and variance of each cell are reported in table 3.9.

Table 3.9: Child Wage Estimates

Sex	Age	Log USD per year Education	Sample Mean	Sample Variance
Male	12-17 Years	Grade 0-2	5.58	1.083
Male	12-17 Years	Scholarship Elig (3-5)	5.66	0.953
Male	12-17 Years	Complete Primary (6)	5.75	1.258
Male	12-17 Years	Secondary (7+)	5.69	0.944
Male	18-24 Years	Grade 0-2	5.91	0.920
Male	18-24 Years	Scholarship Elig (3-5)	6.03	1.184
Male	18-24 Years	Complete Primary (6)	6.26	0.981
Male	18-24 Years	Secondary (7+)	5.96	1.085
Female	12-17 Years	Grade 0-2	5.54	1.386
Female	12-17 Years	Scholarship Elig (3-5)	5.59	1.284
Female	12-17 Years	Complete Primary (6)	5.59	0.730
Female	12-17 Years	Secondary (7+)	5.66	1.204
Female	18-24 Years	Grade 0-2	5.86	0.932
Female	18-24 Years	Scholarship Elig (3-5)	5.97	1.029
Female	18-24 Years	Complete Primary (6)	6.27	0.667
Female	18-24 Years	Secondary (7+)	6.00	0.919

Like the wage parameter estimates, the distributional parameters for home production are also estimated using nonparametric cell means and variances. The approach taken for home production, however, differs in two important ways. First, the value of home production is realized for the entire household, conditional on the specific combination of household members who are primarily at home. Notice that this includes older children, who are assumed to be engaged in productive labor, as well as younger children, whose presence at home may reduce the overall output of the household. Young children at home require care from older household members, taking time away from productive labor. Considering the full profile of all household members introduces a problem of dimensionality, as this would require estimates for every state in the full state space.

Instead, I collapse the set of possible profiles to reflect only the most relevant aspects of the household members who are primarily at home. Specifically, cells are defined by the number of parents at home, the presence of one or more working age son, the presence of one or more working age daughter, and the presence of one or more young child. Working age children are age 12 or older, and young children are age 11 or younger. This combination of characteristics defines $3 \times 2 \times 2 \times 2 = 24$ possible profiles for home production. One of these profiles is never observed in the data, so the distributional parameters cannot be estimated. Instead, I assume that this combination of household members at home, having never been observed as the preferred choice in the data, is particularly unproductive. I replace the sample mean and variance with -1 and 1, respectively.

The second major difference between home production and wage labor, is that I assume household members engage in some home production, even if it is not their primary activity. For example, a woman who is pregnant or working for wages will still work at home, but possibly with reduced output relative to a woman whose primary activity is home production. Households receive home production, then, even if no individual is engaged in home production as his or her primary activity. The sample mean and variance for each cell are reported in table 3.10.

Parameter estimates for the primary model equations are reported in the first column of table 3.11. These parameters define the household utility function and the household terminal value function. In choosing the specifications for these functions, I endeavored to remain consistent with the central focus of the household decision problem, the trade-off between current consumption and human capital accumulation. Given this focus, the primary mechanisms of the model should act through

Table 3.10: Value of Home Production

				Log of Market Value of Annual Rice Harvest (Valued at 1,000 USD per ton)	
Parents (Num)	Any Boy Age 12-24	Any Girl Age 12-24	Any Child Age 0-11	Sample Mean	Sample Variance
0	0	0	0	6.44	0.569
0	0	0	1	6.48	0.563
0	0	1	0	6.91	0.362
0	0	1	1	6.81	0.206
0	1	0	0	6.21	1.051
0	1	0	1	6.92	0.493
0	1	1	0	.	.
0	1	1	1	6.21	1.000
1	0	0	0	6.59	0.758
1	0	0	1	6.61	0.661
1	0	1	0	6.52	1.004
1	0	1	1	6.82	0.248
1	1	0	0	6.77	0.593
1	1	0	1	6.61	0.709
1	1	1	0	6.52	0.206
1	1	1	1	6.38	0.551
2	0	0	0	6.77	0.561
2	0	0	1	6.76	0.579
2	0	1	0	6.73	0.342
2	0	1	1	6.74	0.616
2	1	0	0	6.83	0.611
2	1	0	1	6.95	0.501
2	1	1	0	6.86	0.762
2	1	1	1	6.69	0.473

current production and terminal educational achievement. This consideration motivated relatively parsimonious specifications for the household utility and terminal value functions.

Table 3.11: Parameter Estimates

	Estimates (1)	Recovery (2)
Utility Function		
Cash CRRA Coeff	1.349***	1.343***
	(0.014)	(0.013)
Rice CRRA Coeff	1.733***	1.938***
	(0.072)	(0.208)
Female Primary	0.927***	0.900***
	(0.113)	(0.114)
Female Secondary	0.927***	0.900***
	(0.113)	(0.114)
Female and Child at Home	-0.109	-0.089
	(0.166)	(0.164)
Male and Child at Home	-0.315	-0.176
	(0.163)	(0.152)
Terminal Value Function		
One Daughter	-24.579***	-26.011***
	(1.556)	(1.551)
One Son	-28.676***	-27.323***
	(1.481)	(1.382)
Additional Children (Count)	-28.587***	-29.839***
	(0.740)	(0.780)
One Daughter Primary	4.602*	5.765**
	(1.889)	(2.009)
One Son Primary	4.923**	2.618
	(1.736)	(1.716)
Additional Children Primary	10.730***	11.494***
	(0.826)	(0.860)
One Daughter Secondary	5.456**	5.476**
	(1.943)	(1.987)
One Son Secondary	10.007***	12.753***
	(1.790)	(1.891)
N	1367	1367

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

3.4.3 Simulation and Recovery

The exercise of decision simulation and parameter recovery is undertaken in order to evaluate the accuracy and consistency of the model. First, the estimated parameters are used to calculate choice probabilities over all feasible decisions for each household. The full set of choice probabilities for a given household will partition the unit interval. Each household draws a number on the unit interval from a uniform random distribution, which determines the simulated decision for that household. Simulated decisions are compared to the decisions observed in the data, as an indication of the accuracy of the model. I compare simulated to observed activities across sex and age groups in tables 3.12, and across sex and education groups in table 3.13.

Overall, the simulated decisions are consistent with observed decisions. There are some indications, however, that suggest ways the household decision model could be improved. Consider, for example, the proportion of children attending school after they have passed grade 7. The model underpredicts the number of children who continue attending school after grade 7. This inaccuracy may be due to the aggregation of grade achievements above grade 7. In the binned model, once a child has achieved grade 7, he or she is already at the highest level of educational achievement. Indeed, it is surprising that the model ever predicts that children will continue attending school after passing grade 7. The fact that it does may be attributable to the specification of the utility function, where households gain current utility when female children attend secondary school. Alternatively, it could be caused by the distinction between effective and actual education described in section 3.4.1. If a younger child advanced to secondary school at a unrealistically young age, the household may be making decisions based on the child's effective, rather than actual, level of education.

Table 3.12: Simulation Results by Age and Sex

	Observed			Simulated		
Age	School	Work	Home	School	Work	Home
Male						
0-5 Years	.	.	1.000	.	.	1.000
6-11 Years	0.765	.	0.235	0.744	.	0.256
12-17 Years	0.752	0.099	0.148	0.692	0.155	0.153
18-24 Years	.	0.538	0.462	.	0.503	0.497
Female						
0-5 Years	.	.	1.000	.	.	1.000
6-11 Years	0.821	.	0.179	0.830	.	0.170
12-17 Years	0.770	0.097	0.132	0.766	0.116	0.119
18-24 Years	.	0.583	0.417	.	0.518	0.482

The final step in this exercise requires reestimating the dynamic discrete choice model using the simulated, rather than observed, decisions. Estimation of model parameters based on simulated data follows the same process as estimation based on observed data, the primary difference being that the "true" parameters are known in advance. While successful recovery of the "true" parameters does not prove that a model is correct, failure to do so may indicate the presence of identification problems. Recovered parameters from this exercise are reported in the second column of table 3.11.

Table 3.13: Simulation Results by Education and Sex

	Observed			Simulated		
Education	School	Work	Home	School	Work	Home
Male						
Grade 0-2	0.369	0.023	0.608	0.419	0.035	0.547
Scholarship Elig (3-5)	0.788	0.066	0.146	0.709	0.085	0.206
Complete Primary (6)	0.547	0.283	0.170	0.547	0.245	0.208
Secondary (7+)	0.373	0.331	0.296	0.155	0.373	0.472
Female						
Grade 0-2	0.400	0.037	0.563	0.455	0.023	0.522
Scholarship Elig (3-5)	0.817	0.052	0.131	0.805	0.051	0.144
Complete Primary (6)	0.591	0.174	0.235	0.461	0.304	0.235
Secondary (7+)	0.432	0.308	0.260	0.363	0.301	0.336

3.4.4 Treatment Simulation

Treatment simulations were conducted using the baseline data only. Decisions were simulated for the control group exactly as they were in the previous section. In order to simulate scholarship receipt for the food and cash treatment groups, the expected utility function was modified to include additional consumption valued at 50 USD per year, for the first child attending school in grades 4-6. The additional consumption was added to home production or wage earnings, for the food and cash scholarship treatments, respectively. In both cases, the scholarship accrued to the household, and was not linked to attendance of a particular child. As long as at least one child attended school in the relevant grades, the household received the scholarship, and did not receive additional transfers if additional children attended.

In the treatment simulation, all children in the scholarship grades are equally eligible to receive the scholarship. There is no specific, identified beneficiary child. This aspect of the model is not consistent with the functioning of the scholarship program in reality, but is unavoidable given the current implementation of the model. There is no child characteristic that positively identifies scholarship eligibility, so the beneficiary cannot be tracked based on individual characteristics. Identification at the household level, by the child's ordering of the household state vector, for example, is also impossible in the current implementation. Consideration of the ordering of children in the household state vector was ruled-out when redundant permutations of the household state vector were eliminated from the state space. Given this limitation of the current household identification scheme, I consider the impact of the scholarship treatment on all children in the relevant educational category, rather than on the identified beneficiaries alone.

Simulated treatment effects can be compared informally to the experimental treatment effects from the previous chapter. There are several important differences between the two sets of estimates, though, that make formal statistical comparison impossible. First, as previously discussed, the current estimates are based on all children within the appropriate range of educational achievement, rather than the specific program beneficiary. Second, the current estimates are based on data from the baseline survey only. Experimental estimates are based on results from both survey rounds.

Finally, the experimental estimates are based on a sample, identified beneficiaries, who are initially enrolled in school by definition. The scholarship's impact is to reduce beneficiary dropout over the course of the year, rather than to increase school enrollment. Negative experimental treatment effect estimates, then, represent increased school participation. The estimation sample for simulated treatment effects includes identified beneficiaries, but also contains non-beneficiary siblings. Simulated treatment effects have the potential to impact both school enrollment and completion of the academic year. The outcome variable of interest is school attendance,

Table 3.14: Treatment Simulation

	(1)	(2)	(3)
	Attend School		Dropout
Treatment Group	Mean	Diff	RCT
Control	0.733***		
	(0.0102)		
Food	0.769***	0.0367***	-0.0382*
	(0.00971)	(0.0141)	(0.0203)
Cash	0.721***	-0.0117	-0.0467**
	(0.0103)	(0.0145)	(0.0195)
Observations	5,646		
R-squared	0.742		

RCT estimates are from table 11, column (1) of the previous chapter.
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

rather than school dropout. Increased school participation is indicated by a positive treatment effect estimate.

Recognizing these limitations, it is informative to compare the experimental treatment effect estimates from the previous chapter to the estimates based on simulated data. Both sets of estimates are reported in table 3.14. The reported experimental estimates are those from the nonparametric, single-difference specification reported in the first column of table 2.11 in the previous chapter. This specification was chosen because it is methodologically similar to the simulated treatment effect estimates from this chapter.

The simulated treatment effect estimate of the food scholarship is consistent with the experimental estimate from the previous chapter. In fact, after allowing for the difference in sign, the two point estimates are extremely close. The simulation-based estimate indicates that the food scholarship increased school enrollment by 3.67 percentage points, while the experimental estimates indicates a causal impact that decreased the dropout rate by 3.82 percentage points. These estimates are significant at the 1% and 10% levels, respectively. Of course, the fact that the two point estimates are very close should not be over-interpreted, given the many differences discussed above. It is encouraging, though, that the two estimates are generally consistent, especially as they are estimated on two completely separate data sets. The simulated estimates are based on baseline data only[2], while the experimental estimates are based on endline data only.

While the treatment effect estimates of the food scholarship are encouraging, the estimates for the cash scholarship indicate the need for additional work. Introduction of the cash scholarship seemed to have no impact on the simulated activity decisions. This is inconsistent with the experimental estimates, which indicated a larger impact from the cash scholarship than from the food scholarship. Given the clear impact of the food scholarship on simulated outcomes, it is likely that the poor performance of the simulated cash scholarship is caused by some cash-specific component of the model, rather than the overall structure of the model. Specifically, this reasoning indicates that a problem may exist either in the cash component of the household utility function or in the wage income specification. Additional work is required to tune the components that model wage earnings and consumption from cash.

[2]The distributional parameters for home production were estimated using endline data, as the relevant question was not asked in the baseline survey.

3.4.5 Conclusion

This chapter uses structural estimation to model the impact of a decrease in the cost of schooling on the household decisions regarding the overall time allocation of children in the household. The approach developed here begins from the work of Todd and Wolpin (2006), whose model I replicate and consider for the current application. Following this exploratory exercise, I modify the household decision model both to increase computational speed and to accurately model the specific context of rural Cambodia. After estimating the full dynamic discrete choice model, I use the results to estimate treatment effects for both food and cash scholarships. I find that the estimated treatment effect of the food scholarship matches the reduced-form experimental estimates very well, while that of the cash scholarship indicates the need for additional tuning of the model. This work adds to the growing but underutilized set of tools for structural estimation, by exploring its synergies with program evaluation methods. Furthermore, it sheds light on the nature of household decision-making on child time allocation (specifically the trade-off between schooling and labor) in a new setting (rural Cambodia). In order to further enhance practitioners' understanding of how this relationship is affected by different policy combinations, I plan to continue this research, with the goal of using the validated structural model to estimate the effectiveness of alternative policy interventions that could not be tested in practice.

Appendix A

Peer Effects Appendix

A.1 Auxiliary Equation Estimates

Table A.1: Within Transformation

VARIABLES	(1) Math	(2) Read
Low SES	-0.422***	-0.460***
	(0.0260)	(0.0265)
Age (quarters)	0.0705***	0.0482***
	(0.00775)	(0.00790)
Female	0.133***	0.160***
	(0.0213)	(0.0218)
Black	-0.374***	-0.248***
	(0.0457)	(0.0466)
Constant	0.00462	0.000213
	(0.0104)	(0.0106)
Observations	5,639	5,639
R-squared	0.081	0.074
SP BP test	35.61	70.46
df	16	16
pval	0.0033	0.0000
NP BP test	27.11	74.54
df	15	15
pval	0.0279	0.0000

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

Table A.2: Between Transformation

VARIABLES	(1) Math	(2) Read
Ave. Low SES	-0.505***	-0.611***
	(0.161)	(0.153)
Ave. Age (quarters)	0.115	-0.0712
	(0.0910)	(0.0864)
Ave. Female	0.666**	0.600**
	(0.270)	(0.257)
Ave. Black	-0.0958	-0.115
	(0.131)	(0.124)
Teacher Black	0.103	0.0498
	(0.102)	(0.0968)
Teacher Master's Deg.	-0.0574	-0.0497
	(0.0687)	(0.0652)
Teacher Experience (years)	0.0121**	0.0133**
	(0.00565)	(0.00536)
Constant	-2.622	1.482
	(1.980)	(1.880)
Observations	5,643	5,643
R-squared	0.128	0.163
Number of classK	317	317

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

Table A.3: Semiparametric Conditional Variance Estimation

VARIABLES	(1) Math	(2) Read
Var. Age (quarters)	0.0103***	-0.0181***
	(0.00214)	(0.00275)
Var. Female	0.0532	-1.098***
	(0.0848)	(0.199)
Var. Black	0.0697***	-0.0808***
	(0.0230)	(0.0229)
Ave. Low SES	0.00671	-0.288***
	(0.00923)	(0.0158)
Ave. Age (quarters)	-0.0302***	0.0552***
	(0.00612)	(0.00887)
Ave. Female	0.112***	-0.0811***
	(0.0152)	(0.0221)
Ave. Black	-0.0122	0.0226*
	(0.00828)	(0.0127)
Low SES	0.00198	0.0887***
	(0.00372)	(0.00734)
Age (quarters)	-0.0113***	-0.0171***
	(0.00107)	(0.00184)
Female	0.00820***	0.0343***
	(0.00312)	(0.00361)
Black	0.0264***	-0.0254**
	(0.00599)	(0.0108)
Teacher Black	-0.0140***	0.0233***
	(0.00420)	(0.00657)
Teacher Master's Deg.	-0.0174***	0.0197***
	(0.00340)	(0.00412)
Teacher Experience (years)	0.000831***	-0.00128***
	(0.000288)	(0.000450)
Class Size	-0.00166***	-0.000715
	(0.000383)	(0.000717)
Observations	5,643	5,643

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

Table A.4: Nonparametric Conditional Variance Estimation

Indicator Variables				(1)	(2)
Low SES	Black	Female	Older	Math	Read
0	0	0	0	0.597***	0.671***
				(0.0413)	(0.0760)
0	0	0	1	0.713***	0.971***
				(0.0397)	(0.0729)
0	0	1	0	0.571***	0.528***
				(0.108)	(0.199)
0	0	1	1	0.558***	0.538**
				(0.117)	(0.216)
0	1	0	0	0.627***	0.832***
				(0.0421)	(0.0775)
0	1	0	1	0.705***	0.948***
				(0.0425)	(0.0782)
0	1	1	0	0.589***	0.528***
				(0.109)	(0.200)
0	1	1	1	0.828***	1.317***
				(0.116)	(0.213)
1	0	0	0	0.608***	0.473***
				(0.0600)	(0.110)
1	0	0	1	0.793***	0.668***
				(0.0577)	(0.106)
1	0	1	0	0.544***	0.417***
				(0.0524)	(0.0964)
1	0	1	1	0.540***	0.375***
				(0.0569)	(0.105)
1	1	0	0	0.606***	0.452***
				(0.0587)	(0.108)
1	1	0	1	0.718***	0.588***
				(0.0632)	(0.116)
1	1	1	0	0.563***	0.439***
				(0.0511)	(0.0940)
1	1	1	1	0.648***	0.515***
				(0.0588)	(0.108)
Observations				5,643	5,643
R-squared				0.274	0.118

Standard errors in parentheses, *** $p<0.01$

A.2 Klein and Vella Error Structure

Klein and Vella (2010) introduce a class of error structures for which the control function can be consistently estimated. While the control function for the current application, social interactions in grouped data, does not take the same functional form,the conditions for consistent application are similar.

Estimation of the Klein and Vella control function requires an error structure that can be written as follows:

$$S^2_{ugi} \equiv Var(u_{gi}|\mathbf{X}) \qquad S^2_{vg} \equiv Var(v_g|\mathbf{X})$$

$$u_{gi} = S_{ugi}u^*_{gi} \qquad v_g = S_{vg}v^*_g$$

This is a decomposition of each error term into an unscaled, homoscedastic component (u^*_{gi} , v^*_g), and a scaling function that depends on $\mathbf{X}$. Consistent estimation requires that the following conditional mean and conditional covariance restrictions are satisfied:

$$E(u^*_{gi}|\mathbf{X}) = 0$$

$$E(v^*_g|\mathbf{X}) = 0$$

$$\rho \equiv E(u^*_{gi}v^*_g|\mathbf{X}) = E(u^*_{gi}v^*_i)$$

Given the error structure above, the non-linear control function, $A(\mathbf{X})\, v_g$, reduces as follows:

$$
\begin{aligned}
\mathbb{E}\left[u_{gi} \mid v_g, \mathbf{X}\right] &= A\left(\mathbf{X}\right) v_g \\
&= \frac{cov\left(u_{gi}, v_g \mid \mathbf{X}\right)}{var\left(v_g \mid \mathbf{X}\right)} v_g \\
&= \frac{\mathbb{E}(S_{ugi}u^*_{gi}S_{vg}v^*_g|\mathbf{X})}{S^2_{vg}} \\
&= \rho\frac{S_{ugi}S_{vg}\mathbb{E}(u^*_{gi}v^*_g|\mathbf{X})}{S^2_{vg}} v_g \\
&= \rho[\frac{S_{ugi}}{S_{vg}}]v_g
\end{aligned}
$$

A.3 Derivation of Peer Effects Control Function

The heteroscedastic error structure of the group interactions model is assumed to take the following form:

$$
\begin{aligned}
S^2_{\epsilon gi} \equiv& Var(\epsilon_{gi}|\mathbf{X}) & S^2_{\alpha g} \equiv& Var(\alpha_g|\mathbf{X}) \\
\epsilon_{gi} =& S_\epsilon \epsilon^*_{gi} & \alpha_g =& S_\alpha \alpha^*_g \\
u_{gi} =& S_\alpha \alpha^*_g + S_\epsilon \epsilon^*_{gi} & v_g =& \frac{1}{(1-\beta)}\frac{1}{n_g}\sum_{i\in g}\left(S_\alpha \alpha^*_g + S_\epsilon \epsilon^*_{gi}\right)
\end{aligned}
$$

The following assumptions are required on the error structure:

$$
\begin{aligned}
\mathbb{E}(\alpha^*_g|\mathbf{X}) &= 0 \\
corr\left(\epsilon^*_{gi}, \alpha^*_g \mid X\right) &= 0 \\
corr\left(\epsilon^*_{gi}, \epsilon^*_{gj} \mid X\right) &= \delta
\end{aligned}
$$

The covariance matrix of unobservables for a single group or classroom then takes the form:

$$\mathbf{U}_g = \left(u_{g1}, \ldots, u_{gn_g}\right)'$$

$$Var\left(\mathbf{U}_g \mid \mathbf{X}_g\right) = S_\alpha^2 1_{n_g} 1_{n_g}' + \begin{bmatrix} S_{\epsilon g1}^2 & \cdots & \delta S_{eg1} S_{egn_g} \\ \vdots & \ddots & \vdots \\ \delta S_{egn_g} S_{eg1} & \cdots & S_{\epsilon gn_g}^2 \end{bmatrix}$$

The control function is derived from the solution to the following minimization problem:

$$\begin{aligned} A\left(\mathbf{X}\right) &= arg \min_A \mathbb{E}\left[u_{gi} - Av_g \mid \mathbf{X}\right]^2 \\ &= \frac{cov\left(u_{gi}, v_g \mid \mathbf{X}\right)}{var\left(v_g \mid \mathbf{X}\right)} \end{aligned}$$

The control function is:

$$
\begin{aligned}
A(\mathbf{X})\, v_g &= \mathbb{E}\left[u_{gi} \mid v_g, \mathbf{X}\right] \\
&= \frac{cov\left(u_{gi}, v_g \mid \mathbf{X}\right)}{var\left(v_g \mid \mathbf{X}\right)} v_g
\end{aligned}
$$

Taking each component separately:

$$
\begin{aligned}
&cov\left(u_{gi}, v_g \mid \mathbf{X}\right) \\
&\quad = \mathbb{E}\left[\left(S_\alpha \alpha_g^* + S_\epsilon \epsilon_{gi}^*\right)\left(\frac{1}{(1-\beta)} S_\alpha \alpha_g^* + \frac{1}{(1-\beta)} \frac{1}{n_g} \sum_{i \in g} S_\epsilon \epsilon_{gi}^*\right) \Bigg| \mathbf{X}\right] \\
&\quad = \frac{1}{(1-\beta)}\left[S_{\alpha g}^2 + \frac{1}{n_g}\left(S_{\epsilon gi}^2 + S_{\epsilon gi} \sum_{j \neq i} S_{\epsilon gj} \mathbb{E}\left[\epsilon_{gi}^*, \epsilon_{gj}^* \mid \mathbf{X}\right]\right)\right] \\
&\quad = \frac{1}{(1-\beta)}\left[S_{\alpha g}^2 + \frac{1}{n_g}\left(S_{\epsilon gi}^2 + \delta S_{\epsilon gi} \sum_{j \neq i} S_{\epsilon gj}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&var\left(v_g \mid \mathbf{X}\right) \\
&\quad = \mathbb{E}\left[\left(\frac{1}{(1-\beta)} \frac{1}{n_g} \sum_{i \in g}\left(S_\alpha \alpha_g^* + S_\epsilon \epsilon_{gi}^*\right)\right)^2 \Bigg| \mathbf{X}\right] \\
&\quad = \frac{1}{(1-\beta)^2}\left[S_{\alpha g}^2 + \frac{1}{n_g^2} \sum_{i \in g}\left(S_{\epsilon gi}^2 + S_{\epsilon gi} \sum_{j \neq i} S_{\epsilon gj} \mathbb{E}\left[\epsilon_{gi}^*, \epsilon_{gj}^* \mid \mathbf{X}\right]\right)\right] \\
&\quad = \frac{1}{(1-\beta)^2}\left[S_{\alpha g}^2 + \frac{1}{n_g^2} \sum_{i \in g}\left(S_{\epsilon gi}^2 + \delta S_{\epsilon gi} \sum_{j \neq i} S_{\epsilon gj}\right)\right]
\end{aligned}
$$

The full control function, then, is:

$$
A(\mathbf{X})\, v_g = (1-\beta) \frac{\left[S_{\alpha g}^2 + \frac{1}{n_g}\left(S_{\epsilon gi}^2 + \delta S_{\epsilon gi} \sum_{j \neq i} S_{\epsilon gj}\right)\right]}{\left[S_{\alpha g}^2 + \frac{1}{n_g^2} \sum_{i \in g}\left(S_{\epsilon gi}^2 + \delta S_{\epsilon gi} \sum_{j \neq i} S_{\epsilon gj}\right)\right]} v_g
$$

Simplified Control Function

The simplified control function is derived under the assumption that both the variance of group-level unobservables and the correlation between individual unobservables within groups are equal to zero ($S^2_{\alpha g} = 0, \delta = 0$). The error structure, then, can be described as follows:

$$
\begin{aligned}
u_{gi} =& S_{\epsilon gi}\epsilon^* \\
v_g =& \frac{1}{(1-\beta)}\frac{1}{n_g}\sum_{j\in g} S_{\epsilon gj}\epsilon^*_{gj} \\
Var\left(\mathbf{U}_g \mid \mathbf{X}_g\right) =& \begin{bmatrix} S^2_{ug1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & S^2_{ugn_g} \end{bmatrix} \\
cov\left(u_{gi}, v_g \mid X\right) =& \frac{1}{(1-\beta)}\frac{1}{n_g}S^2_{\epsilon gi} \\
var\left(v_g \mid X\right) =& \frac{1}{(1-\beta)^2}\frac{1}{n_g^2}\sum_{j\in g} S^2_{\epsilon gj}
\end{aligned}
$$

The control function for this system is:

$$
A\left(\mathbf{X}\right) v_g = (1-\beta)\,\frac{S^2_{\epsilon gi}}{\frac{1}{n_g}\sum_{j\in g} S^2_{\epsilon gj}} v_g
$$

Appendix B

Cash vs Food Appendix

B.1 Treatment Effects Regressions

Table B.1: Nonparametric Difference in Expected Outcomes

	Absolute Impact			Relative Impact of Cash		
	Food	Cash	Treat	New	Existing	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Dropout	Dropout	Dropout	Dropout	Dropout	Dropout
Policy	-0.0382*	-0.0467**	-0.0428**	-0.00852	-0.0219	-0.0194*
	(0.0203)	(0.0195)	(0.0184)	(0.0149)	(0.0135)	(0.0110)
Constant	0.0942***	0.0942***	0.0942***	0.0560***	0.0892***	0.0816***
	(0.0169)	(0.0169)	(0.0169)	(0.0113)	(0.0107)	(0.00881)
Observations	994	1,058	1,415	778	2,430	3,208
R-squared	0.005	0.007	0.007	0.000	0.002	0.001

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table B.2: Probit Coefficients

	Absolute Impact			Relative Impact of Cash		
	Food	Cash	Treat	New	Existing	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Dropout	Dropout	Dropout	Dropout	Dropout	Dropout
Policy	-0.274*	-0.354**	-0.316***	-0.0805	-0.151*	-0.142*
	(0.142)	(0.141)	(0.122)	(0.140)	(0.0914)	(0.0795)
Constant	-1.315***	-1.315***	-1.315***	-1.589***	-1.346***	-1.394***
	(0.101)	(0.101)	(0.101)	(0.0997)	(0.0663)	(0.0583)
Observations	994	1,058	1,415	778	2,430	3,208

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table B.3: Probit Coefficients Average Marginal Effects

	Absolute Impact			Relative Impact of Cash		
	Food	Cash	Treat	New	Existing	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Dropout	Dropout	Dropout	Dropout	Dropout	Dropout
Policy	-0.0406*	-0.0498**	-0.0422**	-0.00848	-0.0219	-0.0194*
	(0.0229)	(0.0222)	(0.0181)	(0.0148)	(0.0135)	(0.0110)
Observations	994	1,058	1,415	778	2,430	3,208

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

Table B.4: OLS Regression with Controls

	Absolute Impact			Relative Impact of Cash		
	Food	Cash	Treat	New	Existing	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Dropout	Dropout	Dropout	Dropout	Dropout	Dropout
Policy	-0.0311*	-0.0403**	-0.0365**	-0.00720	-0.0256**	-0.0217**
	(0.0186)	(0.0177)	(0.0170)	(0.0127)	(0.0113)	(0.00926)
Age (months)	-0.0739**	-0.0417	-0.0527**	-0.0600*	-0.00939	-0.0159
	(0.0348)	(0.0265)	(0.0249)	(0.0312)	(0.0168)	(0.0159)
Age Squared	0.000464**	0.000258	0.000332**	0.000388*	4.10e-05	8.78e-05
	(0.000227)	(0.000176)	(0.000165)	(0.000206)	(0.000113)	(0.000107)
Age Cubed	-9.17e-07*	-4.86e-07	-6.51e-07*	-7.94e-07*	-5.17e-09	-1.17e-07
	(4.86e-07)	(3.87e-07)	(3.60e-07)	(4.48e-07)	(2.47e-07)	(2.34e-07)
Female	-0.00536	-0.00168	-0.00306	-0.00433	0.0205*	0.0139
	(0.0157)	(0.0153)	(0.0127)	(0.0147)	(0.0105)	(0.00876)

Robust standard errors in parentheses *Continued on next page*

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

Table B.4 – *Continued from previous page*

	(1)	(2)	(3)	(4)	(5)	(6)
Num. Adults	-0.00524	-0.00715	-0.00616	-0.00504	-0.00491	-0.00473
	(0.00717)	(0.00718)	(0.00601)	(0.00777)	(0.00359)	(0.00330)
Num. Children	0.00525	0.00481	0.00452	0.00226	0.00365	0.00375
	(0.00686)	(0.00613)	(0.00525)	(0.00617)	(0.00342)	(0.00303)
HoH Female	-0.00458	-0.0137	-0.0125	-0.0204	0.00471	-0.00121
	(0.0177)	(0.0164)	(0.0133)	(0.0125)	(0.0120)	(0.00961)
Food Exp. PC	0.000115	3.18e-05	7.77e-05	7.05e-05	2.88e-05	4.34e-05
	(7.68e-05)	(7.12e-05)	(6.39e-05)	(9.37e-05)	(6.16e-05)	(5.21e-05)
Own Village Land	0.00492	0.00740	0.00477	0.00276	0.00378	0.00436
	(0.0222)	(0.0213)	(0.0181)	(0.0238)	(0.0144)	(0.0122)
Own Ag. Land	-0.0138	-0.0221	-0.0132	-0.000252	-0.00350	-0.00468
	(0.0197)	(0.0180)	(0.0159)	(0.0218)	(0.0110)	(0.0101)
Expenditure PC	-3.49e-05	-4.05e-05	-3.21e-05	1.06e-06	-2.86e-05	-2.30e-05
	(2.76e-05)	(3.01e-05)	(2.47e-05)	(3.73e-05)	(2.42e-05)	(2.04e-05)

Robust standard errors in parentheses *Continued on next page*

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

Table B.4 – *Continued from previous page*

	(1)	(2)	(3)	(4)	(5)	(6)
Vit. A Treatment	0.000440	0.0146	0.00595	0.00247	0.00387	0.00286
	(0.0168)	(0.0156)	(0.0136)	(0.0189)	(0.0130)	(0.0107)
Deworm Treatment	0.00378	-0.0354*	-0.0200	-0.0276	-0.0136	-0.0166
	(0.0203)	(0.0196)	(0.0166)	(0.0203)	(0.0141)	(0.0117)
Flood	0.0188	0.0121	0.0182	0.0234	-0.00529	-0.000742
	(0.0222)	(0.0218)	(0.0182)	(0.0210)	(0.0117)	(0.0102)
Drought	0.0295	0.00937	0.0190	0.0182	0.00645	0.00763
	(0.0224)	(0.0210)	(0.0169)	(0.0199)	(0.0135)	(0.0113)
Other Disaster	0.0333	0.00980	0.0148	-0.00412	-0.00894	-0.00855
	(0.0309)	(0.0281)	(0.0227)	(0.0220)	(0.0149)	(0.0125)
Constant	3.774**	2.186*	2.700**	2.983*	0.571	0.872
	(1.747)	(1.294)	(1.234)	(1.557)	(0.826)	(0.784)
Observations	994	1,058	1,415	778	2,430	3,208
R-squared	0.074	0.081	0.076	0.080	0.082	0.078

Robust standard errors in parentheses

Continued on next page

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

Table B.5: Probit Coefficients with Controls

	Absolute Impact			Relative Impact of Cash		
	Food	Cash	Treat	New	Existing	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Dropout	Dropout	Dropout	Dropout	Dropout	Dropout
Policy	-0.276*	-0.388***	-0.336***	-0.0397	-0.192**	-0.175**
	(0.144)	(0.140)	(0.125)	(0.149)	(0.0851)	(0.0754)
Age (months)	-0.292	-0.208	-0.232	0.536	-0.0630	-0.0868
	(0.351)	(0.338)	(0.353)	(1.248)	(0.144)	(0.148)
Age Squared	0.00208	0.00155	0.00177	-0.00217	0.000541	0.000740
	(0.00209)	(0.00199)	(0.00208)	(0.00720)	(0.000872)	(0.000897)
Age Cubed	-4.49e-06	-3.37e-06	-3.93e-06	2.67e-06	-1.12e-06	-1.62e-06
	(4.12e-06)	(3.87e-06)	(4.07e-06)	(1.38e-05)	(1.74e-06)	(1.79e-06)
Female	-0.0529	0.00910	-0.0140	-0.0669	0.163**	0.126*
	(0.115)	(0.118)	(0.105)	(0.161)	(0.0798)	(0.0723)

Robust standard errors in parentheses

Continued on next page

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

Table B.5 – *Continued from previous page*

	(1)	(2)	(3)	(4)	(5)	(6)
Num. Adults	-0.0233	-0.0386	-0.0332	-0.0456	-0.0361	-0.0343
	(0.0590)	(0.0608)	(0.0527)	(0.0805)	(0.0273)	(0.0267)
Num. Children	0.0354	0.0300	0.0313	0.0185	0.0288	0.0294
	(0.0462)	(0.0449)	(0.0401)	(0.0577)	(0.0230)	(0.0222)
HoH Female	0.00333	-0.0817	-0.0786	-0.234	0.0543	0.0105
	(0.135)	(0.134)	(0.117)	(0.156)	(0.0904)	(0.0803)
Food Exp. PC	0.00118*	0.000547	0.000867	0.000941	0.000269	0.000383
	(0.000640)	(0.000705)	(0.000592)	(0.000996)	(0.000527)	(0.000449)
Own Village Land	0.00450	0.0319	0.00438	0.0247	0.0214	0.0300
	(0.181)	(0.187)	(0.165)	(0.286)	(0.122)	(0.112)
Own Ag. Land	-0.110	-0.215	-0.146	-0.0469	-0.0279	-0.0419
	(0.146)	(0.137)	(0.131)	(0.231)	(0.0789)	(0.0777)
Expenditure PC	-0.000462	-0.000504	-0.000389	9.99e-06	-0.000298	-0.000212
	(0.000387)	(0.000450)	(0.000343)	(0.000473)	(0.000272)	(0.000228)

Robust standard errors in parentheses

Continued on next page

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

Table B.5 – *Continued from previous page*

	(1)	(2)	(3)	(4)	(5)	(6)
Vit. A Treatment	0.0232	0.148	0.0715	0.0161	0.0260	0.0266
	(0.128)	(0.130)	(0.119)	(0.206)	(0.0991)	(0.0874)
Deworm Treatment	-0.00103	-0.299**	-0.186	-0.302	-0.129	-0.162*
	(0.142)	(0.140)	(0.124)	(0.188)	(0.101)	(0.0883)
Flood	0.200	0.150	0.207	0.294	-0.0300	0.000241
	(0.155)	(0.163)	(0.141)	(0.194)	(0.0881)	(0.0798)
Drought	0.229	0.0787	0.178	0.269	0.0153	0.0369
	(0.172)	(0.177)	(0.155)	(0.252)	(0.112)	(0.105)
Other Disaster	0.227	0.0719	0.106	-0.134	-0.0472	-0.0544
	(0.192)	(0.208)	(0.176)	(0.285)	(0.115)	(0.107)
Constant	10.10	6.158	6.647	-42.90	-0.460	0.283
	(19.45)	(18.94)	(19.74)	(71.82)	(7.800)	(8.060)
Observations	994	1,058	1,415	681	2,430	3,208

Robust standard errors in parentheses

Continued on next page

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

Table B.6: Probit Average Marginal Effects

	Absolute Impact			Relative Impact of Cash		
	Food	Cash	Treat	New	Existing	Pooled
	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Dropout	Dropout	Dropout	Dropout	Dropout	Dropout
Policy	-0.0357*	-0.0470**	-0.0386**	-0.00378	-0.0245**	-0.0208**
	(0.0202)	(0.0192)	(0.0161)	(0.0142)	(0.0111)	(0.00915)
Age (months)	-0.0377	-0.0252	-0.0267	0.0511	-0.00803	-0.0103
	(0.0459)	(0.0414)	(0.0409)	(0.119)	(0.0183)	(0.0175)
Age Squared	0.000269	0.000188	0.000203	-0.000207	6.89e-05	8.77e-05
	(0.000274)	(0.000244)	(0.000242)	(0.000687)	(0.000111)	(0.000106)
Age Cubed	-5.80e-07	-4.09e-07	-4.52e-07	2.55e-07	-1.43e-07	-1.92e-07
	(5.40e-07)	(4.76e-07)	(4.73e-07)	(1.31e-06)	(2.22e-07)	(2.12e-07)
Female	-0.00683	0.00110	-0.00161	-0.00637	0.0208**	0.0150*
	(0.0149)	(0.0142)	(0.0121)	(0.0154)	(0.0102)	(0.00860)

Robust standard errors in parentheses

Continued on next page

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

Table B.6 – *Continued from previous page*

	(1)	(2)	(3)	(4)	(5)	(6)
Num. Adults	-0.00301	-0.00467	-0.00382	-0.00434	-0.00461	-0.00407
	(0.00754)	(0.00720)	(0.00597)	(0.00767)	(0.00348)	(0.00316)
Num. Children	0.00457	0.00363	0.00360	0.00176	0.00368	0.00349
	(0.00599)	(0.00543)	(0.00461)	(0.00545)	(0.00296)	(0.00265)
HoH Female	0.000431	-0.00990	-0.00904	-0.0223	0.00692	0.00124
	(0.0174)	(0.0162)	(0.0135)	(0.0150)	(0.0115)	(0.00952)
Food Exp. PC	0.000153*	6.63e-05	9.98e-05	8.97e-05	3.42e-05	4.54e-05
	(8.89e-05)	(8.76e-05)	(7.10e-05)	(9.54e-05)	(6.70e-05)	(5.30e-05)
Own Village Land	0.000582	0.00387	0.000504	0.00235	0.00273	0.00355
	(0.0234)	(0.0227)	(0.0190)	(0.0273)	(0.0156)	(0.0133)
Own Ag. Land	-0.0142	-0.0260	-0.0168	-0.00447	-0.00356	-0.00497
	(0.0186)	(0.0162)	(0.0149)	(0.0220)	(0.0101)	(0.00922)
Expenditure PC	-5.97e-05	-6.11e-05	-4.47e-05	9.52e-07	-3.79e-05	-2.51e-05
	(5.18e-05)	(5.62e-05)	(4.05e-05)	(4.51e-05)	(3.45e-05)	(2.70e-05)

Robust standard errors in parentheses

Continued on next page

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

Table B.6 – *Continued from previous page*

	(1)	(2)	(3)	(4)	(5)	(6)
Vit. A Treatment	0.00300	0.0180	0.00823	0.00153	0.00331	0.00315
	(0.0164)	(0.0155)	(0.0135)	(0.0196)	(0.0126)	(0.0103)
Deworm Treatment	-0.000134	-0.0362**	-0.0214	-0.0288	-0.0164	-0.0192*
	(0.0183)	(0.0161)	(0.0140)	(0.0179)	(0.0129)	(0.0106)
Flood	0.0258	0.0182	0.0238	0.0280	-0.00382	2.86e-05
	(0.0202)	(0.0198)	(0.0163)	(0.0187)	(0.0112)	(0.00947)
Drought	0.0295	0.00954	0.0205	0.0256	0.00195	0.00438
	(0.0219)	(0.0214)	(0.0176)	(0.0238)	(0.0143)	(0.0124)
Other Disaster	0.0293	0.00872	0.0122	-0.0128	-0.00602	-0.00645
	(0.0244)	(0.0251)	(0.0202)	(0.0273)	(0.0147)	(0.0127)
Observations	994	1,058	1,415	681	2,430	3,208

Robust standard errors in parentheses

Continued on next page

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

B.2 Heterogeneous Treatment Effects

Table B.7: Educational Investment-Females

Variable	Absolute Ctrl	Absolute Treat	Absolute DinD	Relative Food	Relative Cash	Relative DinD
Attendance	5.84	5.86	0.004	5.90	5.89	-0.034
	(0.03)	(0.03)	(0.055)	(0.02)	(0.02)	(0.030)
Home Work Days	3.60	3.79	0.394^{**}	3.74	3.74	-0.026
	(0.10)	(0.09)	(0.175)	(0.08)	(0.09)	(0.128)
Education Exp.	38.08	37.35	0.799	40.03	41.83	4.855^{*}
	(1.92)	(1.95)	(3.941)	(1.62)	(1.45)	(2.544)
N	313	410	723	770	817	1587

[1] Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$
[2] Standard errors in parentheses

Table B.8: Educational Investment-Males

Variable	Absolute Ctrl	Absolute Treat	Absolute DinD	Relative Food	Relative Cash	Relative DinD
Attendance	5.76	5.87	0.118^{*}	5.91	5.89	0.008
	(0.06)	(0.03)	(0.070)	(0.01)	(0.01)	(0.036)
Home Work Days	3.55	3.70	0.361^{**}	3.42	3.49	0.174
	(0.12)	(0.12)	(0.177)	(0.08)	(0.08)	(0.145)
Education Exp.	35.47	38.63	5.610	40.77	40.71	-0.237
	(2.09)	(2.86)	(4.440)	(2.07)	(1.80)	(3.007)
N	230	290	520	614	658	1272

[1] Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$
[2] Standard errors in parentheses

Table B.9: Educational Outcomes-Females

Variable	Absolute Ctrl	Absolute Treat	Absolute DinD	Relative Food	Relative Cash	Relative DinD
Digitspan	5.78	5.70	0.021	5.75	5.84	−0.075
	(0.07)	(0.07)	(0.122)	(0.05)	(0.05)	(0.093)
Ravens	11.38	10.82	−0.581**	10.93	11.06	0.183
	(0.17)	(0.18)	(0.241)	(0.18)	(0.14)	(0.185)
Math Test (0-13)	6.97	6.78	−0.181	6.08	6.02	−0.017
	(0.33)	(0.29)	(0.316)	(0.18)	(0.19)	(0.201)
Testing Index	0.14	−0.04	−0.171*	−0.14	−0.09	−0.042
	(0.09)	(0.10)	(0.093)	(0.08)	(0.06)	(0.070)
N	315	404	719	765	795	1560

[1] Significance levels: * < 10% ** < 5% *** < 1%
[2] Standard errors in parentheses

Table B.10: Educational Outcomes-Males

Variable	Absolute Ctrl	Absolute Treat	Absolute DinD	Relative Food	Relative Cash	Relative DinD
Digitspan	5.73	5.87	0.253**	5.97	5.92	0.050
	(0.09)	(0.08)	(0.120)	(0.07)	(0.06)	(0.080)
Ravens	11.47	11.95	0.458	11.83	11.88	0.248
	(0.21)	(0.22)	(0.284)	(0.25)	(0.16)	(0.219)
Math Test (0-13)	6.84	6.85	0.046	5.81	6.32	0.450**
	(0.30)	(0.33)	(0.349)	(0.16)	(0.21)	(0.204)
Testing Index	0.11	0.26	0.150	0.07	0.16	0.233***
	(0.10)	(0.11)	(0.106)	(0.10)	(0.08)	(0.072)
N	238	274	512	580	614	1194

[1] Significance levels: * < 10% ** < 5% *** < 1%
[2] Standard errors in parentheses

Table B.11: Educational Investment

Variable	Absolute Ctrl	Absolute Treat	Absolute DinD	Relative Food	Relative Cash	Relative DinD
Attendance	5.82	5.84	0.034	5.89	5.89	0.012
	(0.05)	(0.04)	(0.063)	(0.03)	(0.02)	(0.042)
Home Work Days	3.40	3.90	0.702^{***}	3.81	3.76	−0.195
	(0.12)	(0.12)	(0.244)	(0.10)	(0.11)	(0.178)
Education Exp.	33.59	35.29	3.073	38.73	39.94	4.833
	(2.81)	(2.25)	(5.518)	(2.09)	(1.94)	(3.435)
N	123	187	310	333	336	669

[1] Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$
[2] Standard errors in parentheses

Table B.12: Educational Outcomes

Variable	Absolute Ctrl	Absolute Treat	Absolute DinD	Relative Food	Relative Cash	Relative DinD
Digitspan	5.77	5.59	0.034	5.73	5.81	0.082
	(0.09)	(0.11)	(0.179)	(0.09)	(0.07)	(0.137)
Ravens	11.52	10.78	-0.691^{*}	11.07	10.91	−0.117
	(0.31)	(0.28)	(0.375)	(0.23)	(0.20)	(0.254)
Math Test (0-13)	6.80	6.53	−0.310	6.06	5.82	−0.209
	(0.42)	(0.34)	(0.462)	(0.21)	(0.20)	(0.235)
Testing Index	0.13	−0.15	−0.244	−0.13	−0.18	−0.070
	(0.14)	(0.14)	(0.164)	(0.11)	(0.08)	(0.103)
N	117	182	299	327	335	662

[1] Significance levels: $^{*} < 10\%$ $^{**} < 5\%$ $^{***} < 1\%$
[2] Standard errors in parentheses

Bibliography

(2013). Supplementary notes, commenting the results of the Cambodia Socio-Economic Survey , CSES 2012. Technical report, National Institute of Statistics, Royal Government of Cambodia, Phnom Penh.

Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and Student Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25(1):95–135.

Acemoglu, D. and Angrist, J. D. (2000). How Large are Human-Capital Externalities? Evidence from Compulsory Schooling Laws. *NBER Macroeconomics Annual*, 15:9–59.

Adelman, S. W., Gilligan, D. O., and Lehrer, K. (2008). How Effective are Food for Education Programs? A Critical Assessment of the Evidence from Developing Countries.

Aguirregabiria, V. and Mira, P. (2010). Dynamic Discrete Choice Structural Models: A survey. *Structural Models of Optimization Behavior in Labor, Aging, and Health*, 156(1):38–67.

Ahmed, A. U. (2005). Comparing Food and Cash Incentives for Schooling in Bangladesh.

Alderman, H., Buttenheim, A., and Friedman, J. (2011). Impact Evaluation of School Feeding Programs in Lao PDR.

Alderman, H., de Walque, D., and Kazianga, H. (2009). Educational and Health Impacts of Two School Feeding Schemes Evidence from a Randomized Trial in Rural Burkina Faso.

Alderman, H., Gilligan, D. O., and Lehrer, K. (2010). The Impact of Food for Education Programs on School Participation in Northern Uganda.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, 91(434):444–455.

Anselin, L., Florax, R. J. G. M., and Rey, S. J., editors (2004). *Advances in spatial econometrics: methodology, tools and applications*. Number XII in Advances in Spatial Sciences. Springer.

Ashraf, N., Karlan, D. S., and Yin, W. (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *The Quarterly Journal of Economics*, (May).

Attanasio, O., Fitzsimons, E., Gomez, A., Gutierrez, M. I., Meghir, C., and Mesnard, A. (2010). Children's Schooling and Work in the Presence of a Conditional Cash Transfer Program in Rural Colombia. *Economic Development and Cultural Change*, 58(2):181–210.

Bacolod, M. P. and Ranjan, P. (2008). Why Children Work, Attend School, or Stay Idle: The Roles of Ability and Household Wealth. *Economic Development and Cultural Change*, 56(4):791–828.

Baird, S., Ferreira, F. H., Özler, B., and Woolcock, M. (2014). Conditional, unconditional and everything in between: a systematic review of the effects of cash transfer programmes on schooling outcomes. *Journal of Development Effectiveness*, 6(December):1–43.

Baland, J.-M. and Robinson, J. A. (2000). Is Child Labor Inefficient? *Journal of Political Economy*, 108(4):663–679.

Basu, K. and Van, P. H. (1998). The Economics of Child Labor: Reply. *The American Economic Review*, 89(3):412–427.

Bils, M. and Klenow, P. J. (2000). Does schooling cause growth? *American Economic Review*, 90(5):1160–1183.

Blume, L. E., Durlauf, S. N., Brock, W. A., and Ioannides, Y. M. (2010). Identification of Social Interactions. In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics*, chapter 18, pages Pages 853–964. Elsevier, 1 edition.

Blundell, R. and Costa-Dias, M. (2007). Alternative Approaches to Evaluation in Empirical Microeconomics.

Boozer, M. A. and Cacciola, S. E. (2001). Inside the 'Black Box' of Project Star: Estimation of Peer Effects using Experimental Data.

Bourguignon, F., Ferreira, F. H. G., and Leite, P. G. (2003). Conditional Cash Transfers, Schooling, and Child Labor: Micro-Simulating Brazil's Bolsa Escola Program. *The World Bank Economic Review*, 17(2):229–254.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55.

Breusch, T. and Pagan, A. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 47(5):1287–1294.

Brock, W. A. and Durlauf, S. N. (2007). Identification of binary choice models with social interactions. *Journal of Econometrics*, 140(1):52–75.

Buckles, B. P. and Lybanon, M. (1977). Algorithm 515: Generation of a Vector from the Lexicographical Index [G6]. *ACM Transactions on Mathematical Software*, 3(2):180–182.

Bundy, D., Burbano, C., Grosh, M., Gelli, A., Jukes, M., and Lesley, D. (2009). *Rethinking School Feeding: Social Safety Nets, Child Development, and the Education Sector*. The World Bank, Washington, DC.

Card, D. (2001). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica*, 69(5):1127–1160.

Carrell, S. E., Fullerton, R. L., and West, J. E. (2009). Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, 27(3):439–464.

Carrell, S. E., Malmstrom, F. V., and West, J. E. (2008). Peer Effects in Academic Cheating. *The Journal of Human Resources*, 43(1):173–207.

Carrell, S. E., Sacerdote, B. I., and West, J. E. (2013). From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica*, 81(3):855–882.

Dammert, A. C. (2010). Siblings, child labor, and schooling in Nicaragua and Guatemala. *Journal of Population Economics*, 23(1):199–224.

Duma, N. (2011). Dollarization in Cambodia : Causes and Policy Implications.

Durlauf, S. N. and Tanaka, H. (2008). Understanding regression versus variance tests for social interactions. *Economic Inquiry*, 46(1):25–28.

EFA (2014). Education for All Global Monitoring Report 2013/4 Teaching and Learning: Achieving Quality for All. Technical report, United Nations Educational, Scientific and Cultural Organization (UNESCO), Paris, France.

Emerson, P. M. and Souza, A. P. (2007). Child Labor, School Attendance, and Intrahousehold Gender Bias in Brazil. *The World Bank Economic Review*, 21(2):301–316.

Emerson, P. M. and Souza, A. P. (2008). Birth Order, Child Labor, and School Attendance in Brazil. *World Development*, 36(9):1647–1664.

Epple, D. and Romano, R. E. (2010). Peer Effects in Education: A Survey of the Theory and Evidence. In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics; Vol. 1B*, chapter 20, pages 1053–1163. Elsevier, Amsterdam.

Ferreira, F. H. G., Filmer, D., and Schady, N. (2009). Own and Sibling Effects of Conditional Cash Transfer Programs: Theory and Evidence from Cambodia.

Filmer, D. and Schady, N. (2006). Getting Girls Into School: Evidence from a Scholarship Program in Cambodia.

Filmer, D. and Schady, N. (2009). School Enrollment, Selection and Test Scores.

Filmer, D. and Schady, N. (2011). Does more cash in conditional cash transfer programs always lead to larger impacts on school attendance? *Journal of Development Economics*, 96(1):150–157.

Fraker, T., Devaney, B., and Cavin, E. (1986). An Evaluation of the Effect of Cashing Out Food Stamps on Food Expenditures. *The American Economic Review*, 76(2):230–234.

Fraker, T. M., Martini, A. P., and Ohls, J. C. (1995). The Effect of Food Stamp Cashout on Food Expenditures: An Assessment of the Findings from Four Demonstrations. *The Journal of Human Resources*, 30(4):633–649.

Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996). Crime and Social Interactions. *Quarterly Journal of Economics*, (111):507–548.

GMR and UIS (2014). Progress in getting all children to school stalls but some countries show the way forward. Technical report, Education for All Global Monitoring Report (GMR) and UNESCO Institute for Statistics (UIS), Paris.

Graham, B. S. (2008a). Identifying social interactions through conditional variance restrictions. *Econometrica*, 76(3):643–660.

Graham, B. S. (2008b). Supplement to 'Identifying Through Conditional Variance Restrictions': Additional Empirical Results, Specification Tests, and Power Calculations. *Econometrica Supplemental Material*, 76.

Graham, B. S. and Hahn, J. (2005). Identification and estimation of the linear-in-means model of social interactions. *Economics Letters*, 88(1):1–6.

Griliches, Z. (1970). Education, Income, and Human Capital. In Hansen, W. L., editor, *Education, Income, and Human Capital*, chapter Notes on t, pages 71–127. NBER, Cambridge, MA.

Heckman, J. J. (1996). Randomization as an Instrumental Variable. *The Review of Economics and Statistics*, 78(2):336–340.

Heckman, J. J. (2010). Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature*, 48(2):356–398.

Heckman, J. J. and Klenow, P. J. (1997). Human Capital Policy.

Hidrobo, M., Hoddinott, J., Peterman, A., Margolies, A., and Moreira, V. (2014). Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador. *Journal of Development Economics*, 107:144–156.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396):945–960.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, 58(1-2):71–120.

ILO (2015). International Labour Standards on Child Labour.

Imbens, G. W. and Rubin, D. B. (2008). Rubin Causal Model. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics*. 2 edition.

Jackson, K. D. (1989). The Ideology of Total Revolution. In Jackson, K. D., editor, *Cambodia, 1975-1978: Rendezvous with Death*, chapter 2, page 49. Princeton University Press.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press, 1 edition.

Kakwani, N., Soares, F., and Son, H. H. (2005). Conditional Cash Transfers in African Countries. Technical report.

Keane, M. P. and Wolpin, K. I. (1994). The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence. *The Review of Economics and Statistics*, 76(4):648–672.

Klein, R. and Vella, F. (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics*, 154(2):154–164.

Knuth, D. E. (1981). Radix Conversion. In *The Art of Computer Programming: Volume 2: Seminumerical Algorithms*, chapter 4.4, pages 300–313. Addison-Wesley, 2nd edition.

Knuth, D. E. (1998). Sorting by Exchanging. In *The Art of Computer Programming: Volume 3: Sorting and Searching*, chapter 5.2.2, pages 106–110. Addison-Wesley, 2nd edition.

Knuth, D. E. (2005). Generating all Combinations and Partitions. In *The Art of Computer Programming: Volume 4, Fascicle 3*, chapter 7.2.1.3. Addison-Wesley, 1st edition.

Koenker, R. (1981). A Note on Studentizing a Test for Heteroscedasticity. *Journal of econometrics*, 17(1):107 – 112.

Krauth, B. (2006). Simulation-based estimation of peer effects. *Journal of Econometrics*, 133(1):243–271.

Krueger, A. B. (1999). Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, (May):497–532.

Krueger, A. B. and Whitmore, D. M. (2000). The Effect of Attending a Small Class in the Early Grades on College-Test Taking and Middle School Test Results: Evidence from Project Star.

Lee, L.-f., Liu, X., and Lin, X. (2010). Specification and estimation of social interaction models with network structures. *Econometrics Journal*, 13(2):145–176.

Leroy, J. L., Gadsden, P., Rodriguez-Ramirez, S., and de Cossio, T. G. (2010). Cash and In-Kind Transfers in Poor Rural Communities in Mexico Increase Household Fruit, Vegetable, and Micronutrient Consumption but also Lead to Excess Energy Consumption. *The Journal of Nutrition*, 140:612–617.

Lyle, B. D. S. (2009). The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point. *American Economic Journal: Applied Economics*, 1(4):69–84.

Lyle, D. S. (2007). Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point. *Review of Economics and Statistics*, 89(2):289–299.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies Ltd.*, 60(3):531–542.

Mincer, J. A. (1974). *Schooling, Experience, and Earnings.* NBER, Cambridge, MA.

Moffitt, R. (1989). Estimating the value of an in-kind transfer: The case of food stamps. *Econometrica: Journal of the Econometric Society*, 57(2):385–409.

Moffitt, R. A. (2001). Policy Interventions, Low-Level Equilibria, and Social Interactions. In Durlauf, S. N. and Young, H. P., editors, *Social Dynamics*, chapter 3, pages 45–82. Brookings Institution Press and MIT Press, Washington, DC and Cambridge, MA.

Mohiddin, L., Sharma, M., and Haller, A. (2007). Comparing Cash and Food Transfers: Findings from a Pilot Project in Sri Lanka. *Field Exchange*, (30):19–21.

Nielsen, N. S., Godden, K., Leguene, P., Ruegenberg, D., and Rudiger, J. (2010). WFP Cambodia School Feeding 2000-2010: A Mixed Method Impact Evaluation. Technical report, World Food Programme, Office of Evaluation, Rome.

Prasso, S. T. (2001). The Riel Value Of Money: How The World's Only Attempt To Abolish Money Has Hindered Cambodia's Economic Development. *Asia Pacific Issues*, (49).

Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to Investment in Education: A Further Update. *Education Economics*, 12(2):111–134.

Rauch, J. E. (1993). Productivity Gains from Geographic Concentration of Human Capital: Evidence rom the Cities.

Ravallion, M. and Wodon, Q. (2000). Does Child Labour Displace Schooling? Evidence on Behavioral Responses to an Enrollment Subsidy. *The Economic Journal*, 110(462):C158–C175.

Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *The American Economic Review*, 94(2):247–252.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5):688–701.

Rust, J. (1987). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, 55(5):pp.999–1033.

Rust, J. (1996). Numerical Dynamic Programming in Economics. In Amman, H. M., Kendrick, D. A., and Rust, J., editors, *Handbook of Computational Economics*, number 1, chapter 14, pages 620–722. Elsevier B.V., 1 edition.

Rust, J. (2000). Nested Fixed Point Algorithm Documentation Manual. (October):1–43.

Sacerdote, B. (2000). Peer Effects with Random Assignment: Results for Dartmouth Roommates.

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, 116(2):681–704.

Santacroce, P. (2008). Kingdom of Cambodia: Comprehensive Food Security and Vulnerability Analysis (CFSVA). Technical report, World Food Programme, VAM Food Security Analysis, Rome.

Schady, N. and Fiszbein, A. (2009). *Conditional Cash Transfers: Reducing Present and Future Poverty*. The World Bank, Washingon, DC.

Schuck, P. H. and Zeckhauser, R. (2007). *Targeting in Social Programs: Avoiding Bad Bets, Removing Bad Apples (Google eBook)*. Brookings Institution Press.

Shari, K., Howard, W., and Ella, C. (2013). Quality education for all children? What works in education in developing countries.

Skoufias, E., Unar, M., and González-Cossio, T. (2008). The Impacts of Cash and In-Kind Transfers on Consumption and Labor Supply: Experimental Evidence from Rural Mexico.

Todd, P. E. and Wolpin, K. I. (2006). Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility. *The American Economic Review*, 96(5):1384–1417.

UN (2014a). Resources for Speakers on Global Issues Education for All (EFA).

UN (2014b). The Millennium Development Goals Report. Technical report, United Nations, New York, NY.

UNICEF (2010). World leaders and experts meet to raise the profile of education on the development agenda [Press release]. Retrieved from http://www.unicef.org/media/media_56121.html.

WFP (2011). Food and Cash Scholarship Operational Guideline. Technical report.

Whitmore, D. (2005). Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment. *American Economic Review Papers and Proceedings*, 95(2):199–203.

Wick, M. (2015). GeoNames Database. Retrieved from http://www.geonames.org/.

Word, E., Johnston, J., Bain, H. P., DeWayne Fulton, B., Zaharias, J. B., Achilles, C. M., Lintz, M. N., Folger, J., and Breda, C. (1990). The State of Tennessee's Student/Teacher Achievement Ratio (STAR) Project: Technical Report 1985-1990. Technical report.