Poverty Diagnostics in the Philippines:
Assessing Impacts of Programs through Generalised Linear Models (GLMs)

by

Joel E. Bancolita

A thesis
submitted to the Victoria University of Wellington
in partial fulfilment of the
requirements for the degree of
Master of Science
in Statistics and Operations Research.

Victoria University of Wellington
2011

Abstract

The Philippines is a country where a quarter to one-third of the population is poor. Although the nation has managed to lower poverty incidence in some years, its booming population increases the poor population dramatically. This is why alleviating poverty is a pinnacle program in the country.

In aid of poverty alleviation endeavor, this study focuses on assessing which programs had been effective in alleviating poverty given other family characteristics. Aside from descriptive methods, employing Generalised Linear Models (GLMs) and categorical data analysis are the focus in analysing the effects of existing intervention programs on status of improvement and income of families. In addition, varying effects of programs depending on values of other covariates are also analysed.

Descriptive analysis and modeling are applied on the panel data of families. Intervention programs namely scholarship, Comprehensive Agrarian Reform Program (CARP) and government housing or other housing financing program (GHFP) have been run together with other family characteristics to describe improvement in welfare and income. Interaction effects, between access to intervention programs and other aspects of the family, have been derived to give a richer picture of the phenomenon. The study has come to conclude that the programs are indeed effective in improving lives of families, with some effects varying on some levels of other explanatory variables.

Keywords: Generalised Linear Models, GLM, Linear Regression, Logistic Regression, Logit, Odds ratio, Philippines, Poverty

Acknowledgments

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Poverty reduction has been one of the overarching programs in the Philippines at the national and local level. Ways and means on how to measure and monitor poverty have earned great interest. Income-based poverty indicator and other dimensions of poverty have been used to identify the poor and measure the extent of poverty. In this study, we revisit the concept and status of poverty in the Philippines, programs to address it, and the effects of these programs in improving the lives of the families.

This chapter will discuss research objectives, related literatures and scope of the study. Our objective, iterated in Section 1.1, is to find out whether intervention programs affect the welfare status of families and what characteristics distinguish improved families. This research question, though long-running, is still very relevant today. In addition, studies relating to our objectives will also be tackled in Section 1.2 such as methods in measuring poverty, status of poverty in the Philippines in relation to neighbouring countries, and intervention programs to improve welfare of families. Moreover, definition of improvement and related terms will be discussed in Section 1.3. Improvement will be described through the three concrete intervention programs in our available data namely the *Comprehensive Agrarian Reform Program (CARP)*, scholarship for tertiary education and government housing or other housing financing program (GHFP).

These programs only spanned for 1998 and 1999 in our data which brings us to discuss scope of our study in Section 1.4 due to constraints in our data. In the same section we will also describe the panel data set of the Philippines in 1997-1999 where we will draw our findings from.

In the succeeding chapters, we will discuss methods to address our objectives, answers to them and summary of findings. Chapter 2 will deal on usage of contingency tables and the Generalised Linear Model (GLM) in drawing and interpreting findings that will be discussed in Chapter 3. Lastly, we will summarise these findings and answer our objectives in Chapter 4.

1.1 Objectives

In this paper, the answers to the following main questions are sought:

- Using statistical models, how do intervention programs affect welfare of the families

- What are the characteristics of the families that have improved or have higher incomes

Our analysis will revolve around these two main objectives. Specifically we want to find out

- Does receiving CARP in either year makes helps families to improve across the years? Will this program also help in changing the income pattern of families

- Is scholarship for tertiary education helpful in increasing family income and in pulling them out of poverty?

- Will GHFP put surplus on family income. Will it be a determinant for improvement?

- What characteristics of families distinguish those who improve from those who did not. What separates families who have higher income?

- Are there any interacting effects among those factors that have been identified?

We will answer research questions in the succeeding discussions in the paper, particularly in Chapter 3. In tackling these objectives, it is necessary to revisit related concepts especially poverty. Thus, in the next section, existing literatures relating to poverty in the Philippines will be reviewed.

1.2   Review of related literature

A recent report of the Asian Development Bank (ADB) reveals that the Philippines is one of the 18 economies with more than 10 percent of the population living on less than $1.25 *PPP*[1] a day. In addition, the Philippines ranked 10th in between Turkmenistan and Pakistan [2, p. 65-67]. Although the National Statistics Office (NSO) reported that in 2006, average annual income of Filipino families have increased by 16.8 percent from P148,000 in 2003 to P173,000 in 2006, the country still has been considered as off-track in achieving the $1.25 a day target in *eradicating extreme poverty and hunger* [2]. Furthermore, aside from the 2.4 percentage point increase in poverty incidence among population from 24.4 percent in 2003 to 26.9 percent in 2006 [14], the country has been surpassed by its neighbouring countries such as Thailand, Malaysia, Viet Nam and Cambodia in achieving this target.

Before delving into the Philippines' poverty measures, it is essential to define the term poverty. According to ADB,

---

[1]Purchasing Power Parity (PPP)
[2]Millennium Development Goal (MDG) 1. http://www.un.org.

> " *poverty is a deprivation of essential assets and opportunities to which every human is entitled. Everyone should have access to basic education and primary health services. Poor households have the right to sustain themselves by their labor and be reasonably rewarded, as well as have some protection from external shocks. Beyond income and basic services, individuals and societies are also poor and tend to remain so—if they are not empowered to participate in making the decisions that shape their lives.* "

The traditional and perhaps the most common indicator of deprivation is lack of sufficient income. The establishment of money-metric classification is based on the notion that poor living is a common result of lack of income [21].

1.2.1 Poverty measurement in the Philippines

The international perspective such as in the previous discussion is not much different from the Philippines' official local perspective of poverty since they both pertain to insufficient income [2]. In the Philippines, official poverty statistics is money-metric, that is, the traditional approach is to determine a threshold or poverty line to classify a measurement unit, say family, as poor or not. Analogous to the $1.25 PPP, the Philippines has a set of official thresholds or lines in Philippine Peso (PHP) in identifying the poverty status of families or individuals as shown in Table 1.1.

Threshold	2000	2003	2006
Poverty	11,458	12,309	15,057
Food	7,707	8,149	10,025

Source: NSCB

Table 1.1: Per Capita (*per head*) Poverty and Food Thresholds (in PHP), Philippines

The two poverty lines in Table 1.1 are used in measuring two official

incidence measures in the Philippines—poverty and subsistence. The first one pertaining to the minimum amount to meet basic food and non-food needs while the latter is the minimum amount for food needs [14]. For instance, each family's total annual income will be divided by its family size (to derive the income per capita) and will be compared to the poverty lines to determine its poverty status. These thresholds are based on a long tradition of methodology in the country, the development of which however, has not been simple [20].

The major source of poverty information in the country is the Family Income and Expenditure Survey (FIES) which has been conducted since 1957. The survey covers the whole country and aims to describe family income distribution and spending patterns as well serve as input in updating consumer price indices (CPI). From being conducted every five years from 1957-1975, it has been made more frequent since 1985 by conducting it every three years [8].

Change in frequency of the survey rounds is not the only major change in measurement. The methodology has seen two major changes in 1992 and 2003 mainly in deriving poverty lines. These changes were brought by: (1) change in expenditure measurement (1992); and (2) shifting from regional to provincial prices in valuing "basket of goods". Changes in 1992 were retroactively applied until the 1985 round while the 2003 until 1997. Both changes brought lower poverty lines and consequently lower poverty incidences. For instance, in 2000, the originally published poverty incidence among families was 33.7 percent while the revised due to 2003 methodology is 28.4 percent [20, p. 129].

On the other hand, looking at the poverty incidence is very different in looking at the magnitude. Figures 1.1 and 1.2 shows one paradox of poverty [3]. Due to the high population growth rates in the country, [3] the magnitude of poor population is increasing dramatically in some time

---

[3]NSO projections for 2004 based on the 1995 census was 82.7 and expected to reach 105 million by 2020.

interval although the incidence of poor families is decreasing over time. That is, from 1985 to 2000, the magnitude of poor population increased by at least 4 million even though the poverty incidence was cut by 10.5 percent. In addition, improvement in the situation in rural areas are not quite good compared to that in urban areas [20].

Source of Data: National Statistics Office

Figure 1.1: Poverty Incidence among Families 1985-2000, by Urbanity

Going down further to local level is more effective in combating poverty [16]. Table 1.2 shows the ten poorest provinces in 2006 based on the official poverty incidence while Figure 1.3 shows the location of poorest provinces. From the table, it can be seen that the Autonomous Region in Muslim Mindanao (ARMM) is the region that contains most of the poor provinces. Furthermore, seven out of the ten provinces are located in the

Source of Data: National Statistics Office

Figure 1.2: Magnitude of Poor Individuals 1985-2000

southern Philippines (Mindanao) and this is also shown in the map that majority of the areas are in the south. Tawi-tawi, the poorest province, has increased poverty incidence by 26.5 from 52.4 in 2000 while was not even ranked as one of the poorest in 2003. This demonstrates that the whole poverty picture of the Philippines is not as improved as it looks.

1.2.2 Univariate and multivariate

Other than the headcount and incidence from the income variable given a poverty line, other poverty statistics that are being released using the income variable are intensity of poverty and poverty inequality. These measures can be derived by using the popular Foster-Greer-Thorbecke (FGT)

Province	Region	Poverty Incidence	Rank
Tawi-tawi	ARMM	78.9	1
Zamboanga del Norte	IX	63	2
Maguindanao	ARMM	62	3
Apayao	CAR	57.5	4
Surigao Del Norte	Caraga	53.2	5
Lanao del Sur	ARMM	52.5	6
Northern Samar	VIII	52.2	7
Masbate	V	51	8
Abra	CAR	50.1	9
Misamis Occidental	X	48.8	10

Source: NSCB

Table 1.2: Ten Poorest Provinces in the Philippines, 2006

generalised poverty measure expressed by Equation (1.1)

$$FGT_\alpha = \frac{1}{N} \sum_{i=1}^{N_p} \left(\frac{z - x_i}{z} \right)^\alpha \tag{1.1}$$

where N is the number of measurement units (e.g. families), z is the poverty line, N_p is the number of poor units, x_is are incomes of the units and α is the sensitivity parameter. Known forms of FGT given value of α are the headcount ratio or the poverty incidence (FGT_0), the intensity or poverty gap (FGT_1) and inequality (FGT_2). Poverty gap is the average amount needed to pull everyone up to the poverty line [13]. In 2006, poverty gap and severity were published at 7.7 and 3.1 respectively [14].

Another popular measure is the GINI coefficient (Gini 1912) $\mathcal{G} = A/(A + B)$ which can be derived from the Lorenz curve as shown in Figure 1.4. This measures income inequality by comparing the empirical distribution of income to the theoretical distribution when wealth is distributed equally. The value of $\mathcal{G}$ ranges from 0 (total equality) to 1 (total inequality). Income equality in the country slightly worsened from 0.458 in 2006 to 0.448 in 2009 [14].

Source of basic data: NSCB
Mapped through ArcGIS

Figure 1.3: Map of Poverty Incidence among Families, 2006

Other than cross-sectional univariate analysis of income, cross-sectional time-series analysis of income has been done at the observation unit level. For instance, the Annual Poverty Indicators Survey (APIS) which is conducted to provide poverty indicators in years without the FIES, are available as panel data in 1997, 1998 and 1999. This panel is based on the FIES

Source: The World Bank

Figure 1.4: Lorenz Curve

1997 dataset. Using the extracted panel of families and poverty thresh-olds[4] given in Table 1.3, movements in and out poverty can be analysed as shown in Table 1.4 [18].

1997	1998	1999
11,319	11,935	12,834

Source: NSCB

Table 1.3: Per Capita (*per head*) Poverty Thresholds (in PHP), Philippines

---

[4]It must be noted that measurement of poverty status is not generally based on overall (country) thresholds but on geographical thresholds, i.e. by region and by urbanity. See [18].

Poverty status	1997	1998	1999	Number of Families	Percent
PPP	Poor	Poor	Poor	3563	21.5
PPN	Poor	Poor	Nonpoor	648	3.9
PNP	Poor	Nonpoor	Poor	569	3.4
PNN	Poor	Nonpoor	Nonpoor	522	3.2
NPP	Nonpoor	Poor	Poor	1301	7.9
NPN	Nonpoor	Poor	Nonpoor	1025	6.2
NNP	Nonpoor	Nonpoor	Poor	1154	7.0
NNN	Nonpoor	Nonpoor	Nonpoor	7755	46.9
Total				16537	100

Source of Basic Data: Run from the matched Public Use Files
of the 1997 FIES, and the 1998 and 1999 APIS

Table 1.4: Number and Percent of Families by Status of Poverty, 1997-1999

The table reveals that more than one out of five families are living in poverty through the three years while less than half are living non-poor. Furthermore, 13.3 percent of the panel have moved out poverty in 1999 which is less than the 18.3 percent of families who moved into poverty.

Univariate analysis of income has proven to be useful in analysing headcount poverty, intensity and inequality either cross-sectional or time-series. However, although it is an important means to live a life without deprivation, poverty must be seen ultimately in terms of poor living and income is not the only influence [21]. Besides, multidimensional poverty is widely accepted as the richer concept of poverty [3].

Every Medium-term Philippine Development Plan (MTPDP) since 1992 has referred not just on income poverty targets but human development goals. This shows that the government has already spurred focus not just on income poverty alone [7]. In fact, both the National Anti-Poverty Commission (NAPC) and Department of Interior and Local Government (DILG), have promoted establishment of the 14 core local poverty indicators relating to different dimensions of poverty [6] through adopting a

Community-Based Monitoring System (CBMS) [16]. These variables are shown in Table 1.5.

CBMS indicator	Dimensions of poverty	Core indicators
Survival	Health	Proportion of child deaths (0-5 years old)
		Proportion of women who died due to pregnancy-related causes
	Nutrition	Proportion of malnourished children (0-5 years old)
	Water and sanitation	Proportion of households without access to safe water supply
		Proportion of households without access to sanitary toilet facilities
Security	Shelter	Proportion of households living in makeshift housing
		Proportion of households classified as squatters/informal settlers
	Peace and order	Proportion of persons who were victims of crime
Enabling	Income	Proportion of households with income below the poverty threshold
		Proportion of households with income below the subsistence threshold
		Proportion of households that experienced food shortages
	Employment	Proportion of persons who are unemployed
	Education	Proportion of children 6-12 years old who are not in elementary school
		Proportion of children 13-16 who are not in secondary school

Source: Reyes and Due, *Fighting Poverty with Facts*

Table 1.5: 14 Indicators of Poverty

Moreover, analysis of correlates of poverty as techniques applied on them is not new in the Philippines. For instance, ownership of amenities such as refrigerators, vehicles and computers were identified to be

good distinguishing factors of non-poor families [12]. Models of these correlates have been also explored [7]. Furthermore, aside from existing human development indices, there are existing attempts to establish composite indicator of multidimensional poverty (CPI) due to attractiveness of income of being able to rank observation units [3]. Attempts in the Philippines include simple scoring to establish and index [19] and data reduction techniques such as Principal Components and Multiple Correspondence Analysis (PCA and MCA) [4]. There have been debates on derivation of poverty variable weights for the CPI since techniques such as PCA and MCA yield weights due to the variability of each variable. One might interpret that higher weights mean more important variables to index, while it is not necessarily the case.

Despite these existing undertakings, the literature in the country on multivariate poverty is still lagging compared to univariate income poverty. The identified reasons were: lack of income is seen to be the more pressing problem; and data availability [7]. For instance, two surveys that will be used in this thesis namely FIES and APIS, do not have all the 14 core poverty indicators. CBMS data contains all of the 14 core indicators, however, it is still quite new compared to the two national surveys and not entirely comparable across the country due to decentralised implementation scheme. Income, which will be the indicator of poverty to be used in this study, remains the simplest and official way of measuring poverty and identifying the poor.

1.2.3   Poverty alleviation through intervention programs

Measurement of poverty is one thing, how to alleviate it is another. From 1986 to 2004, each of the four of the country's executive formulated a particular poverty alleviation project namely: *Tulong sa Tao* (Aquino, 1986); *Social Reform Agenda (SRA)* (Ramos, 1992); *Lingap Para sa Mahihirap (Lingap)* (Estrada, 1998); and *Kapit-Bisig Laban sa Kahirapan (KALAHI)* (Arroyo,

2001) [20]. In this study, we have three concrete programs available that aim to help the poor: scholarship program for tertiary education, GHFP, and CARP.

Scholarship for tertiary education and GHFP are the usually implicit programs received by families while CARP is a more specific type of program. CARP stemmed out to address the problem of skewed distribution of land, even landlessness, in the country which dates back to the 400-year colonisation history of the country. In particular, one of the roots of the problem was the establishment of haciendas and emerging encomienda system during Spanish times.

American colonisers in 1930s began to establish land reform measures in the country and formulated appropriate legislations. However, due to circumstances during that time, the bulk of estates for distribution went to American firms, businessmen, and landlords. Marcos administration had the first major attempt at land reform during Martial Law in 1972 and has become comprehensive since the Aquino administration in 1987 for land distribution has covered only rice and corn lands but also private and agricultural land. It then assumed the name and recognises not just farmers but all workers in the land as beneficiaries given that they are landless and willing to cultivate. CARP has been found to have positive effects on income particularly on farmers [17].

1.3   Improvement and other terms defined

We have revisited existing literatures relating to poverty and throughout the paper, we will use the term *improvement* in poverty status or in welfare. We define those who have improved as those families who initially started poor but moved out of poverty in the final year. Before we define how families are classified as poor or non-poor, let us define other terms [9, p. 31] that we require:

- Family. Our observation unit is family which is a group of persons bound by ties of kinship, and live together under the same roof and eat together or share in common the family food.

- Family head. A female or male [5] adult member of the family who is responsible for the organisation of the family. He or she is regarded as such by members of the family.

- Household. Household is an aggregate of persons, generally but not necessarily bound by ties of kinship, who sleep in the same house and have common arrangements for the preparation and consumption of food. This is not necessarily similar to family since a household can be a collection of families.

- Per capita income. Given a family's income, *per capita* or per person income is the quotient of family's income divided by the number of family members.

- Per capita poverty threshold. This refers to the minimum amount needed by a person to meet basic food and non-food needs. This is usually used as annual.

Suppose we are given z_t as the per capita poverty threshold at time t and we let

$$m_{it} = \frac{\text{income of family } i}{\text{\# of members of family } i}$$

as per capita income of family i at time t. Although we acknowledge the multivariate nature of poverty, we adopt the money-metric definition of poverty based on income, which makes our poverty indicator function defined by

$$\mathcal{P}_{it} = \mathbb{I}_{\{m_{it} < z_t\}} = \begin{cases} 1, & \text{poor} \\ 0, & \text{non-poor} \end{cases}$$

---

[5]We will use the term gender for this attribute. It may have a different connotation in other context but here we are referring to biological sex.

and from this, the poverty incidence among n observation units at time t is defined by

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{P}_{it}$$

Finally, if we have $t = 1, 2, \ldots, T$ and we focus on those who are poor at time 1, our improvement indicator function is defined by

$$\mathcal{I}_i = \mathbb{I}_{\{\mathcal{P}_{i1}=0, \mathcal{P}_{iT}=1\}} = \begin{cases} 1, & \text{improved} \\ 0, & \text{did not improve} \end{cases}$$

1.4   Data, scope and limitations

Unlike the study on CARP by Reyes in 2002 [17] which primarily used a targeted 1990-2000 survey of beneficiaries and non-beneficiaries to assess effectiveness of the program, we will utilise the available 1997-1999 panel data giving us $t = 1, 2, 3$. The panel data contain 16,537 longitudinal data of families from matched subset cases from the 1997 Family Income and Expenditure Survey (FIES), and the 1998 and 1999 Annual Poverty Indicators Survey (APIS). This proves to be a limited data set in terms of time factor. This is because panel generation from the National Statistics Office (NSO) has stopped since 2000.

Moreover, one characteristic of the data is that one cross-section—FIES is measured differently compared to the other two (APIS) due to the nature of the survey. The difference is particular in income. Furthermore, there are variables that are not available across the time period. APIS is more comprehensive in poverty indicators, while FIES focuses on income and expenditure. Access to programs such as CARP and scholarship in particular are not present in the 1997 FIES, only in 1998 and 1999 APIS.

The study will also be confined on analysing income as the primary indicator of poverty (as discussed in Section 1.3) due to data constraints, and use available covariates of poverty in the data based on Table 1.5. We

will show this in Chapter 3. It must be mentioned also that there is still no official weights for the subset of cases, which means that our conclusions on finding relationship between improvement and three programs controlling on all possible covariates, will be confined on the matched cases forming the panel data. Hence, there will no projection for the whole population can be done.

Chapter 2

Methodology

In this chapter, methods in analysing dynamics in poverty using several variables will be tackled. Our general setting is we have response variable, specifically improvement attribute, to be described by a set of explanatory variables. Since our variables include mostly categorical responses, these will involve analysis of of contingency tables and identifying underlying associations among the variables. Moreover, we will also discuss techniques in modeling improvement given predictors of welfare to describe associations among variables. In particular, we will focus on using the *Generalised Linear Models (GLM)* in describing these phenomena.

We will review concepts in analysing p-dimensional contingency tables such as cell probabilities and odds ratios in Section 2.1. After that, we will tackle modeling using the GLM in Section 2.2 and model diagnostics in Section 2.3.

2.1   Contingency tables

Presenting variables in tables is one of the simplest and common ways of mining information from the data. In this study, we utilise tables as starting point in digging relationships among variables relating to poverty.

2.1.1 Cell frequencies and probabilities

Consider we have $p = 2$ categorical (or categorised) variables each with I and J factors respectively. Hence, we have $I \times J$ two-dimensional table such as in Table 2.1

X	Y				Total
	y_1	y_2	$\cdots$	y_J	
x_1	n_{11}	n_{12}	$\cdots$	n_{1J}	$n_{1.} = \sum_{j=1}^{J} n_{1j}$
x_2	n_{21}	n_{22}	$\cdots$	n_{2J}	$n_{2.} = \sum_{j=1}^{J} n_{2j}$
$\vdots$	$\cdots$	$\cdots$	$\ddots$	$\vdots$	$\vdots$
x_I	n_{I1}	n_{I2}	$\cdots$	n_{IJ}	$n_{I.} = \sum_{j=1}^{J} n_{Ij}$
Total	$n_{.1} = \sum_{i=1}^{I} n_{i1}$	$n_{.2} = \sum_{i=1}^{I} n_{i2}$	$\cdots$	$n_{.J} = \sum_{i=1}^{I} n_{iJ}$	$n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}$

Table 2.1: General two-dimensional table

where n_{ij} is the frequency in row i and column j while $n_{i.}$ represents total in row i and $n_{.j}$ denotes total in column j. We also call them marginal frequencies. Customarily, we call the column variable with levels y_js as the dependent variable while the row variable with levels x_is as the independent variable.

Now, if we take X and Y as events, denote π_{ij} be the probability of the event x_i and y_j, i.e.

$$\pi_{ij} = P(X = x_i, Y = y_j)$$

Also, let $\pi_{i.}$ and $\pi_{.j}$ be the marginal probabilities of event x_i and y_j respectively which we can represent through

$$\pi_{i.} = \sum_{j=1}^{J} \pi_{ij}, \quad \pi_{.j} = \sum_{i=1}^{I} \pi_{ij}$$

If we have the events and counts in Table 2.1, we can represent the

probabilities through the classical frequency approach, that is

$$
\begin{aligned}
\hat{\pi}_{ij} &= \frac{n_{ij}}{n} \\
\hat{\pi}_{i.} &= P(\widehat{X_i = x_i}) = \sum_{j=1}^{J} \frac{n_{ij}}{n} = \frac{n_{i.}}{n} \\
\hat{\pi}_{.j} &= P(\widehat{Y_j = y_j}) = \sum_{i=1}^{I} \frac{n_{ij}}{n} = \frac{n_{.j}}{n}
\end{aligned}
$$

Hence, we can plug in the estimated probabilities $\hat{\pi}_{ij}$, $\hat{\pi}_{i.}$ and $\hat{\pi}_{.j}$ in one-to-one correspondence with those in Table 2.1. Suppose further that we are interested in estimating probability that one event happened say y_j given another event x_j denoted by $\pi_{j|i}$. By definition of conditional probability, we have

$$
\begin{aligned}
\pi_{j|i} &= P(Y = y_j \mid X = x_i) \\
&= \frac{P(Y = y_j, X = x_i)}{P(X = x_i)} \\
\hat{\pi}_{j|i} &= \frac{n_{ij}}{n} \frac{n}{n_{i.}} = \frac{n_{ij}}{n_{i.}}
\end{aligned}
\tag{2.1}
$$

We will use this concept in analysing improvement of families that we have defined in Section 1.3 and its association with other variables particularly receiving intervention programs. For example, if we cross-tabulate receiving a specific type of program X and improvement of welfare Y, both with two levels, we have the following 2×2 table

Received a program (X)	Improve (Y) (1) Yes	(2) No	Total
(1) Yes	n_{11}	n_{12}	$n_{1.}$
(2) No	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	n

Table 2.2: Two-dimensional table example

Using the quantities in Table 2.2, we have the simplified expression based on Equation (2.1),

$$
\begin{aligned}
\hat{\pi}_{1|1} &= P(Y = \text{Yes} \mid X = \text{Yes}) = \frac{n_{11}}{n_{1.}} \\
\hat{\pi}_{2|1} &= P(Y = \text{No} \mid X = \text{Yes}) = \frac{n_{11}}{n_{1.}} \\
\hat{\pi}_{1|2} &= P(Y = \text{Yes} \mid X = \text{No}) = \frac{n_{21}}{n_{2.}} \\
\hat{\pi}_{2|2} &= P(Y = \text{No} \mid X = \text{No}) = \frac{n_{21}}{n_{2.}}
\end{aligned}
\tag{2.2}
$$

We will use this in formulating the *odds ratio* in the next subsection which will be one of the primary measures in our analysis.

2.1.2 Odds ratio

We define *odds ratio* as the ratio of the *odds* of success of an event to the *odds* of success of another event. The odds of success, on the other hand, is the ratio of success to failure. Specifically, given a probability of success π_i, the odds of success is

$$
\text{odds}_i = \frac{\pi_i}{1 - \pi_i} \in [0, \infty)
$$

when this value is less than 1, we say that the chance of success is less than that of failure.

Now, if we have two events in i and j, the *odds ratio* denoted by θ is defined by

$$
\theta(i, j) = \frac{\text{odds}_i}{\text{odds}_j}
$$

If events in i and j are independent then $\theta(i, j)$ is 1. When $\theta(i, j) > 1$, it means that the odds of success for event i is larger than the odds of success for event j. If we go back to Table 2.2, we can estimate the odds ratio of

row 1 compared to row 2 as

$$\hat{\theta}(1,2) = \frac{\hat{\pi}_{1|1}/\hat{\pi}_{2|1}}{\hat{\pi}_{2|1}/\hat{\pi}_{2|2}} = \frac{\hat{\pi}_{11}\hat{\pi}_{22}}{\hat{\pi}_{21}\hat{\pi}_{12}} = \frac{n_{11}n_{22}}{n_{21}n_{12}} \tag{2.3}$$

We set "success" as improving of a family and compare odds of success of those who received intervention program to the odds of success of those who did not receive the program. This concept of odds ratio will be an integral part of our methodology as well as in our analysis.

Now, we are not only interested in analysing a univariate dependency of a response variable, say improvement. We are more interested in drawing a picture of a multivariate effect on the response variable and analysing them simultaneously. This can be facilitated by building models that will explain improvement given some covariates.

2.2 Generalised Linear Model (GLM)

In modeling, the general setting is there is a response variable Y to be described by a set of say p explanatory variables $(x_1, x_2, \ldots, x_p)$. In this section we will discuss the Generalised Linear Models (GLM) which will be the pinnacle of our analysis. Specifically, the models that will be used are Logistic Regression and Multiple Linear Regression.

2.2.1 Review of GLM

In this study, we are given the values of the predictors x_js, e.g. intervention programs and family characteristics, and we are interested in the expected value of the dependent variable of improvement, Y

$$E(Y|x_1, x_2, \ldots, x_p) = \mu_Y$$

This quantity is aimed to be described by the linear combination of the

explanatory variables (and conventionally including an intercept). Denote $\mathbf{x} = [1 \ x_1 \ x_2 \ \cdots \ x_p]^T$ as the vector of explanatory variables and $\boldsymbol{\beta} = [\beta_0 \ \beta_1 \ \beta_2 \ \cdots \ \beta_p]^T$ as the unknown vector of coefficients, we seek to express the mean of Y using some *link function* $g(\cdot)$ such that

$$g(\mu_Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \tag{2.4}$$

Considering all observations, we have

$$\mathbf{Y}_{n\times 1} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} ; \ \mathbf{X}_{n\times(p+1)} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

and we seek to determine the unknown vector of parameters $\boldsymbol{\beta}_{(p+1)\times 1}$ such that

$$g(\boldsymbol{\mu}_Y) = \mathbf{X}\boldsymbol{\beta}$$

where $\boldsymbol{\mu}_Y = [\mu_{Y_1} \ \mu_{Y_2} \ \cdots \mu_{Y_n}]^T$

In general, GLM has three components that build the structure in the model namely the link function, systematic and random component. In particular, the response variable is determined by the random component while the independent variables in the right hand side are identified by the systematic component. These two are related by using the *link function*[1]. For a multiple linear regression model, the link function is the *identity* link, i.e. $g(\mu_Y) = \mu_Y$ while for a logistic regression model, the link function is the *logit* link, i.e. $g(\mu_Y) = \log\left(\frac{\mu_Y}{1-\mu_Y}\right)$ where $\mu_Y = P(Y = 1)$.

---

[1]For a more detailed discussion, refer to [1, pp. 65-67].

2.2.2 Model building

Estimation of parameters β will be facilitated by **R** through *maximum likelihood estimation (MLE)*. As the term MLE implies, it seeks the optimal value of the parameter β that maximises the chance of getting the values observed. Suppose that we have a $n \times 1$ random response vector **Y** with parameter β and joint density function $f_{\mathbf{Y}}(\mathbf{y}; \beta)$. If we assume that realisations of **Y** are given, the likelihood of β denoted by $L(\beta \mid \mathbf{Y})$ is the same as the density function of **Y**. That is,

$$L(\beta \mid \mathbf{Y}) = f_{\mathbf{Y}}(\mathbf{y}; \beta) \tag{2.5}$$

Now, given values of **Y** the *maximum likelihood estimator* of β is the value $\hat{\beta}$ such that

$$L(\hat{\beta} \mid \mathbf{Y}) \geq L(\beta \mid \mathbf{Y}), \quad \forall \, \beta \in \Omega$$

this is equivalent to satisfying the same condition using *log-likelihood function* $\ell(\beta \mid \mathbf{Y}) = \log L(\beta \mid \mathbf{Y})$, that is

$$\ell(\hat{\beta} \mid \mathbf{Y}) \geq \ell(\beta \mid \mathbf{Y}), \quad \forall \, \beta \in \Omega$$

ML estimator is obtained by differentiating the objective function with respect to the parameter being examined. In practice (particularly in **R**), numerical methods such as Newton-Raphson Method are used instead to come up with the optimal value for $\hat{\beta}$.

2.3 Model diagnostics and selection

After coming up with estimates for the model parameters β_js, their values must be assessed whether they are significant or not. Furthermore, model sufficiency must be also examined to come up with the optimal model. In this section, methods of testing significance of model parameters as well

as basis for model selection will be discussed.

2.3.1 Deviance

Let us assume notation in Section 2.2 where we have p explanatory variables regressed on a response variable. Furthermore, consider two log-likelihood functions: the first one is from a model of interest, e.g. one with a coefficient and an intercept; and the second one is the most complicated model which provides a perfect fit for the the data where each observation has its own parameter (saturated). Denote these log-likelihood by $\ell_{\mathcal{M}}$ and ℓ_s respectively. The following quantity

$$G^2(\mathcal{M}) = 2(\ell_s - \ell_{\mathcal{M}})$$

is called the *deviance* of a GLM which converges in distribution to chi-square with p degrees of freedom (number of coefficients) given that large sample conditions hold. This tests whether all parameters other than those included in model m are zero and will be used in the likelihood ratio test in the next subsection.

2.3.2 Significance tests

We seek to test $\mathcal{H}_0 : \beta_j = 0$ against $\mathcal{H}_1 : \beta_j \neq 0$. In this study, we use the *Wald* and *Likelihood-ratio* tests

(i) *Wald statistic*

$$Z_j = \frac{\hat{\beta}_j}{ASE(\hat{\beta}_j)}$$

Under $\mathcal{H}_0$, Z_j follows the standard normal distribution, i.e. $Z_j \sim \mathcal{N}(0,1)$. Hence, $\hat{\beta}_j$ is significantly different from zero given a level of significance α if $Z_j > Z_{\alpha/2}$

(ii) *Likelihood-ratio test statistic*

Given two models $\mathcal{M}_0$ and $\mathcal{M}_1$ where $\mathcal{M}_0$ is nested in $\mathcal{M}_1$, that is the parameters in $\mathcal{M}_0$ is a subset of the parameters in $\mathcal{M}_1$. In addition, each model has its own deviance—$G^2(\mathcal{M}_0)$ and $G^2(\mathcal{M}_1)$ respectively. The difference between their deviance denoted by

$$G^2(\mathcal{M}_0 \mid \mathcal{M}_1) = G^2(\mathcal{M}_0) - G^2(\mathcal{M}_1)$$

is called the *likelihood-ratio statistic*. This is equivalent to

$$
\begin{aligned}
G^2(\mathcal{M}_0 \mid \mathcal{M}_1) &= 2(\ell_s - \ell_{\mathcal{M}_0}) - 2(\ell_s - \ell_{\mathcal{M}_1}) \\
&= -2(\ell_{\mathcal{M}_0} - \ell_{\mathcal{M}_1})
\end{aligned}
\tag{2.6}
$$

When models $\mathcal{M}_0$ and $\mathcal{M}_1$ provide the same fit, $G^2(\mathcal{M}_0 \mid \mathcal{M}_1)$ has an asymptotic chi-squared distribution with degrees of freedom equal to the difference of their respective residual degrees of freedom, i.e. $df_0 - df_1$.

2.3.3 Variable selection: AIC

A serious challenge in this study is selecting variables for the model. Perhaps, there is no unique way of coming up with an optimal set of variables to make an ideal model especially when dealing with a huge data, however, simple and straightforward models will be adopted.

We start with identifying variables aligned with the framework of the study then we play around selecting which of them are ideal to be included in the model. Basically, *Akaike Information Criterion (AIC)* combined with significance tests in Subsection 2.3.2 will be used to determine the optimal set of regressors.

AIC incorporates goodness of fit as well as penalty in adding further

variables while balancing them. Equation (2.7) shows the formula for AIC

$$AIC = 2\left\{-\log[L(\hat{\boldsymbol{\beta}} \mid \mathbf{Y})] + K\right\} \tag{2.7}$$

$L(\hat{\boldsymbol{\beta}} \mid \mathbf{Y})$ is similar likelihood function in Equation (2.5) while $K = p + 2$ or one more than the parameters in the model. There is no rule of thumb with regards to the value of this figure. The only basis is that the lower the AIC the better the model is. Hence, this works ideally well in comparing different models with different sets of variables. However, although increasing the number of parameters will likely increase the value of the log-likelihood function $\ell(\cdot) = \log L(\cdot)$ and in turn decreases AIC, this is penalised by K.

In this study, AIC will be used in determining whether adding significant parameters in the model improves the fit or just makes the model more complicated without adding more information. Typically, the analyses and model building in Chapter 3 will deal with forward and backward selection of effects.

2.4 Factor interactions

It must be noted that not all variables that are deemed to be insignificant using the significance tests in Section 2.3.2 are indeed unimportant. Instead, care must be done in selecting variables to include particularly when they are highly associated or correlated. When there is high association between included explanatory variables, one can be a near duplicate of the other and some of the coefficients become less important. Hence, there can be some coefficients that are insignificant while in fact they can be associated with the response variable.

Furthermore, one variable may have different effects on the response variable if it will be analysed with respect to the levels of another variable. In this case, we can consider a model with interactions. Although we seek

the most simplistic model that would describe our data, it is essential to capture other phenomena that will give a clearer picture. This sometimes lead us in examining how our variables affect the response variable when they interact with another variable.

The model in Equation (2.4) is just the linear combination of *each* of the explanatory variables regressed on the link function of the mean. It is interesting to see whether a set of variables, particularly among categorical ones, have different effects when they interact. Recall that in linear regression, each effect is looked at holding others constant. For instance, will those who received two different poverty alleviation programs, likely to move out of poverty compared to those who received otherwise.

In Chapter 3, we will tackle adding further interactions between explanatory variables based on significance and AIC. This stepwise inclusion/exclusion of factors is one way of identifying effects of collinearity in the model.

2.5 Model usage

Intepreting the model derived is one of the most important element of this thesis. Generally, we will assess the effect of the independent variables on the response variable by looking at the change in value of the response given unit change in the covariates. Since we will be using model with interaction effects, interpretation must be done using a set of values of the covariates compared to another set of values.

Consider that we have a two-factor interaction model and we have our response variable Y and vector of covariates $\mathbf{x}$ which contains the main and interaction factors. If we are given two sets of values for $\mathbf{x}$ namely $\mathbf{x}_a$ and $\mathbf{x}_b$ from say family a and family b, we can express the change as

$$g(\mu_Y \mid \mathbf{x} = \mathbf{x}_a) - g(\mu_Y \mid \mathbf{x} = \mathbf{x}_b) \tag{2.8}$$

and we can interpret the effect of differences in values of the covariates on the response variable. If we talk about linear regression, then we are just assessing the change in μ_Y, i.e. $\Delta\mu_Y$ due to the identity link. On the other hand, if we are using logistic regression, then we are talking about the logarithm of the odds ratio. Recall the logit link in Subsection 2.2.1, the expression in Equation (2.8) will be

$$
\begin{aligned}
\log\left(\frac{\pi(\mathbf{x}_a)}{1 - \pi(\mathbf{x}_a)}\right) - \log\left(\frac{\pi(\mathbf{x}_b)}{1 - \pi(\mathbf{x}_b)}\right) &= \log\left(\frac{\text{odds}_a}{\text{odds}_b}\right) \\
&= \log[\theta(a, b)] \qquad (2.9)
\end{aligned}
$$

Hence, we will use this to compare different scenarios wherein we can assess whether changes in covariates affect the dynamics in welfare of families.

2.6 Summary

We have discussed the methods and techniques that we will use in analysing whether intervention programs can pull poor families out of poverty. Tabular data will comprise our basic analyses particularly the univariate cases while we will analyse simultaneous effects on improvement of family characteristics and access to programs by using GLMs. We will see application of these in Chapter 3.

Chapter 3

Analysis

Techniques and methods have been discussed in Chapter 2 to analyse several aspects relating to dynamics in poverty among families. The results in applying these techniques will be discussed in this chapter. That is, we will look at how families improve their situation across the period by analysing associations between the improvement variable, as defined in Chapter 1, and other covariates. Hence, we will focus on those who, from 1997, moved out of poverty and identify determining factors of change in family well-being. Section 3.2 will tackle analysis of the dynamics of family through logistic regression while Section 3.3 will focus on how family income changes given several variables using multiple linear regression.

3.1   Situationer

Recall that Table 1.4 showed the magnitude[1] of families by their poverty status across the three-year period. However, the question on what characterizes those who moved out from those who remained in poverty in the last year still remains. We seek to answer this by analysing which of the intervention programs (or *combination* of them) can be proven to be

---

[1]The total in Table 1.4 is 16,537, which is different from the total in the succeeding tables and analyses (16,536) due to a family with missing data.

significant in helping those who have moved out. Furthermore, we will also draw other characteristics of the families that distinguish those who remained in poverty across the period from those who did not.

3.1.1 Access to programs across the period

Section 1.2.3 discussed some of the existing programs relating to housing, agriculture and education. These programs: scholarship assistance for tertiary education; government housing or other housing financing program (GHFP); and Comprehensive Agrarian Reform Program (CARP), are expected to alleviate poverty and help vulnerable groups in the country. At this point, our interest is to see what happened to those who had access to these programs. As clarified in Section 1.4, we can only discuss access to these programs during 1998 and 1999.

- *GHFP*

  Let us begin by looking at access to GHFP. It can be seen in Table 3.1 that access to this program among those families in the panel data declined between 1998 and 1999. In addition, access to this program increased (+0.3%) among those families who are chronically poor (PPP) while decreased (-0.6%) among those who stayed non-poor from 1997-1999. It is also notable that among the 1,723 families who moved into poverty from 1998-1999 (PNP and NNP), there is also a decline in access to this program—from 62 families (3.6%) to 39 families (2.3%).

  Looking closely, it is worth mentioning that the reference period for the survey data is the last 12 months from the date of interview. Thus, the ideal situation is those who are identified as poor last year (whose measured incomes were accrued two years ago), must have received the program in the last 12 months. Now, let us use this information to see the whether those in need indeed received the program.

Poverty status	Total	Availed in 1998		Availed in 1999	
		Magnitude	Proportion	Magnitude	Proportion
Total	16,536	697	4.2	626	3.8
PPP	3,563	76	2.1	86	2.4
PPN	647	20	3.1	16	2.5
PNP	569	18	3.2	10	1.8
PNN	522	10	1.9	14	2.7
NPP	1,301	37	2.8	30	2.3
NPN	1,025	31	3.0	30	2.9
NNP	1,154	44	3.8	29	2.5
NNN	7,755	461	5.9	411	5.3

Source of Basic Data: Run from the matched Public Use Files
of the 1997 FIES, and the 1998 and 1999 APIS

Table 3.1: Number and Percent of Families who acquired House / Lot through GHFP by Status of Poverty, 1997-1999

It can be verified from Table 3.1 that out of the 5,301 (PPP, PPN, PNP and PNN) poor families in 1997, only 124 families received the program in 1998 which translates to 2.3 percent. Moreover, 573 or 5.1 percent of the 11,235 non-poor families (NPP, NPN, NNP, NNN) in the same year—1997 acquired the program in 1998.

Similar finding can be drawn if we shift our focus on status of families during the year 1998. Out of 6,536 poor families in 1998 (PPP, PPN, NPP, and NPN), 164 or 2.5 percent of them received GHFP in 1999 while 533 or 5.3 percent of the 10,000 non-poor families (PNP, PNN, NNP, and NNN) in 1998 became beneficiaries of the program in 1999.

- *CARP*

On the other hand, overall access to CARP also rose from 1998 to 1999 as seen in Table 3.2. In addition, there is also an increase in

beneficiaries among those who had been chronically poor (PPP) and never been poor across the period (NNN), from 1.5 to 1.7 and 1.1 to 1.4 percent respectively. Unlike in GHFP case, there are higher incidence of access to CARP in 1998 among poor families (80 out of 5,301 or 1.5%) compared to non-poor families (131 out of 11,235 or 1.2%) in 1997 while almost equal incidence of beneficiaries in 1999 among those who are poor (86 out of 6,536) and non-poor (125 out of 10,000) in 1998 at 1.3 percent.

Poverty status	Total	Availed in 1998		Availed in 1999	
		Magnitude	Proportion	Magnitude	Proportion
Total	16,536	211	1.3	253	1.5
PPP	3,563	53	1.5	59	1.7
PPN	647	7	1.1	14	2.2
PNP	569	10	1.8	9	1.6
PNN	522	10	1.9	13	2.5
NPP	1,301	15	1.2	15	1.2
NPN	1,025	11	1.1	23	2.2
NNP	1,154	16	1.4	11	1.0
NNN	7,755	89	1.1	109	1.4

Source of Basic Data: Run from the matched Public Use Files
of the 1997 FIES, and the 1998 and 1999 APIS

Table 3.2: Number and Percent of Families who acquired land through CARP by Status of Poverty, 1997-1999

- *Scholarship for tertiary education*

  Looking at Table 3.3, it can be seen that there is a decline in acquisition of tertiary scholarship program from 1998-1999 among panel families. Also, there are decrease in access among poor and non-poor families across the period can be seen (1.4% to 0.8% and 3.2 to 2.6% for PPP and NNN respectively). Furthermore, there is also a disparity in access to this program across poverty groups. Looking closely

in both years 1998 and 1999, there are higher incidences of benefi-
ciaries among non-poor families in 1997 (326 out of 11,235) and 1998
(307 out of 10,000) at 2.9 and 3.1 percent respectively, compared to
1.8 and 1.7 percent among those who are poor (95 out of 5,301 and
114 out of 6,536).

Poverty status	Total	Availed in 1998		Availed in 1999	
		Magnitude	Proportion	Magnitude	Proportion
Total	16,536	421	2.5	311	1.9
PPP	3,563	50	1.4	30	0.8
PPN	647	17	2.6	9	1.4
PNP	569	14	2.5	4	0.7
PNN	522	14	2.7	10	1.9
NPP	1,301	26	2	11	0.8
NPN	1,025	21	2	29	2.8
NNP	1,154	29	2.5	15	1.3
NNN	7,755	250	3.2	203	2.6

Source of Basic Data: Run from the matched Public Use Files
of the 1997 FIES, and the 1998 and 1999 APIS

Table 3.3: Number and Percent of Families who received Scholarship for
Tertiary Education by Status of Poverty, 1997-1999

3.1.2 Access to programs among the poor in 1997

Now, since we are interested in assessing improvement in the the welfare
status of the families given the three-year panel of families, we shift our
focus only on those who are poor in the first year, 1997. Based on our def-
inition of improvement in Section 1.3, we can formulate the improvement
variable here as

$$\text{improve} = \begin{cases} Yes, & \text{if PPN or PNN} \\ No, & \text{if PPP or PNP} \end{cases}$$

and as a result, we have Tables 3.4, 3.5 and 3.6.

Table 3.4 shows the improvement status for all poor families in 1997 (PPP, PPN, PNP, PNN) by access to GHFP. Using the improvement convention, it can be seen that only 22.1 percent of the poor families managed to improve at the end of the period. Now, if we look into improvement given access to GHFP, 24.2 percent of those who accessed the program in 1998 improved which is higher than those who were not able to access the program, 22 percent. Same with those who became beneficiaries of the program in 1999, 23.8 percent of them improved which is slightly higher than 22 percent improvement of those who did not benefit from the program.

		Total	improved		did not improve	
			Magnitude	Proportion	Magnitude	Proportion
Total		5,301	1169	22.1	4132	77.9
Availed	Yes	124	30	24.2	94	75.8
In 1998	No	5,177	1,139	22.0	4,038	78.0
Availed	Yes	126	30	23.8	96	76.2
In 1999	No	5,175	1,139	22.0	4,036	78.0

Source of Basic Data: Run from the matched Public Use Files
of the 1997 FIES, and the 1998 and 1999 APIS

Table 3.4: Number and Percent of Families who acquired House / Lot through GHFP by Status of Improvement, 1997-1999

Compared to Table 3.4, Table 3.5 shows that there is higher improvement given access to CARP in 1999 than in 1998. That is in 1999, there is more than one out four families (28.4%) improved when they access the program compared to about one out of five (21.3%) during 1998. Moreover, in 1998, proportion of those who improved among those who accessed CARP is slightly lower than than of those who were not able to

access CARP (22.1%). On the other hand, there is higher improvement incidence among those who accessed CARP in 1999 than those who did not (21.9%).

		Total	improved		did not improve	
			Magnitude	Proportion	Magnitude	Proportion
Total		5,301	1,169	22.1	4,132	77.9
Availed In 1998	Yes	80	17	21.3	63	78.8
	No	5,221	1,152	22.1	4,069	77.9
Availed In 1999	Yes	95	27	28.4	68	71.6
	No	5,206	1,142	21.9	4064	78.1

Source of Basic Data: Run from the matched Public Use Files
of the 1997 FIES, and the 1998 and 1999 APIS

Table 3.5: Number and Percent of Families who acquired land through CARP by Status of Improvement, 1997-1999

Access to scholarship has more impact compared to GHFP and CARP. Table 3.6 shows that almost one out of three families (32.6%) improved given that they received scholarship in 1998. The 1999 case is slightly higher which is 35.8 percent or about more than one out of three families who received scholarship moved out of poverty. If we look at those who did not access the program, the incidence are almost the same compared to the previous programs.

Given these partial findings, although we can tell that there is an underlying effect of programs on improvement of families, we still seek to tell whether effects of the programs are significant. Furthermore, we are also interested in what set of programs and characteristics of the families differentiate those who have improved or increased in income. In Section 3.2, we will examine those poor families and extract factors that determine their improvement while in Section 3.3, we will tackle what characteristics

		Total	improved		did not improve	
			Magnitude	Proportion	Magnitude	Proportion
Total		5,301	1,169	22.1	4,132	77.9
Availed In 1998	Yes	95	31	32.6	64	67.4
	No	5,206	1,138	21.9	4,068	78.1
Availed In 1999	Yes	53	19	35.8	34	64.2
	No	5,248	1,150	21.9	4,098	78.1

Source of Basic Data: Run from the matched Public Use Files
1997 FIES, and the 1998 and 1999 APIS

Table 3.6: Number and Percent of Families who received Scholarship for Tertiary Education by Status of Improvement, 1997-1999

and variables affect changes in family income.

3.2 Logistic regression among 1997 poor families

In this section we will build a logistic regression model to analyse improvement of families. Let the response variable Y be the indicator function of improvement[2] that we have established previously,

$$Y = \mathbb{I}_{\{PPN \cup PNN\}} = \begin{cases} 1, & \text{improved (if PPN or PNN)} \\ 0, & \text{did not improve (if PPP or PNP)} \end{cases}$$

Now, we let our covariate vector $\mathbf{x}$ include the three intervention programs that spanned for two years and other family characteristics. These explanatory variables are defined in Table 3.7.

Looking at Table 3.7, majority of the variables are implicit except 'hhcond'. The household condition variable 'hhcond' pertains to some of the basic amenities the household must have such as safe water, toilet facility

---

[2]A similar classification is done in [18].

j	x_j	Variable(level)	Description	Value
1	G	gender	family head's gender	1 - female; 0 - male
2	U	urb	Urbanity	1 - urban; 0 - rural
3	A	age99	family head's age	
4	C_1	carp98	acquired agricultural land thru CARP in 1998	1 - yes; 0 - no
5	C_2	carp99	acquired agricultural land thru CARP in 1999	1 - yes; 0 - no
6	F_1	hsfg98	acuired house / lot through GHFP in 1998	1 - yes; 0 - no
7	F_2	hsfg99	acuired house / lot through GHFP in 1999	1 - yes; 0 - no
8	S_1	scholp98	availed scholarship for tertiary education in 1998	1 - yes; 0 - no
9	S_2	scholp99	availed scholarship for tertiary education in 1999	1 - yes; 0 - no
10	H_2	hhcond(2)	fair household condition	1 - if hhcond = 2; 0 - otherwise
11	H_3	hhcond(3)	good household condition	1 - if hhcond = 3; 0 - otherwise
12	E_2	heduc(Coll)	family head's highest educational attainment is tertiary level	1 - if heduc = 'Coll'; 0 - otherwise
13	E_3	heduc(HS Grad)	family head's highest educational attainment is high school graduate	1 - if heduc = 'HS Grad'; 0 - otherwise
14	E_4	heduc(HS)	family head's highest educational attainment is high school level	1 - if heduc = 'HS'; 0 - otherwise
15	E_5	heduc(Elem Grad)	family head's highest educational attainment is elementary graduate	1 - if heduc = 'Elem Grad'; 0 - otherwise
16	E_6	heduc(Elem)	family head's highest educational attainment is elementary level	1 - if heduc = 'Elem'; 0 - otherwise
17	E_7	heduc(NG)	family head did not finish any formal education	1 - if heduc = 'NG'; 0 - otherwise

Table 3.7: Covariates for the Logistic Regression

and secured house or non-makeshift house. Its levels denote the number of amenities a household has with (1) as the special case which includes those households without all the amenities. Since this variable has

three levels (3) 'good', (2) 'fair' and (1) 'bad', we made two dummy variables from it namely H_2 and H_3 with level (1) as the reference level. This can be seen in Table 3.8 In the same way, we have six dummy variables $(E_2, \ldots, E_7)$ for the educational attainment of the family head with college graduate as the reference level which can be seen in Table 3.9.

Level	Description	Dummy variable	Representation
1	bad	(reference)	$\{H_2 = 0, H_3 = 0\}$
2	fair	H_2	$\{H_2 = 1, H_3 = 0\}$
3	good	H_3	$\{H_2 = 0, H_3 = 1\}$

Table 3.8: Household condition variable ('hhcond') transformation

Level	Description	Dummy variable	Representation
Coll Grad	College graduate	(reference)	$\{E_2 = 0, E_3 = 0, E_4 = 0,$ $E_5 = 0, E_6 = 0, E_7 = 0\}$
Coll	College level	E_2	$\{E_2 = 1, E_3 = 0, E_4 = 0,$ $E_5 = 0, E_6 = 0, E_7 = 0\}$
HS Grad	High school graduate	E_3	$\{E_2 = 0, E_3 = 1, E_4 = 0,$ $E_5 = 0, E_6 = 0, E_7 = 0\}$
HS	High school level	E_4	$\{E_2 = 0, E_3 = 0, E_4 = 1,$ $E_5 = 0, E_6 = 0, E_7 = 0\}$
Elem Grad	Elementary graduate	E_5	$\{E_2 = 0, E_3 = 0, E_4 = 0,$ $E_5 = 1, E_6 = 0, E_7 = 0\}$
Elem	Elementary level	E_6	$\{E_2 = 0, E_3 = 0, E_4 = 0,$ $E_5 = 0, E_6 = 1, E_7 = 0\}$
NG	No grade completed	E_7	$\{E_2 = 0, E_3 = 0, E_4 = 0,$ $E_5 = 0, E_6 = 0, E_7 = 1\}$

Table 3.9: Educational attainment of the family head ('heduc') transformation

3.2.1 Estimation of model parameters

Of the $p = 17$ initial explanatory variables, the only continuous variable is age A, the rest are either binary or categorical. Given these covariates $\mathbf{x}$, we begin with the simplest main effects model (without interaction)

$$
\begin{aligned}
\text{logit}[P(Y = 1 \mid \mathbf{x})] &= \log\left(\frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}\right) \\
&= \beta_0 + \beta^G G + \beta^U U + \beta^A A + \beta_1^C C_1 + \beta_2^C C_2 \\
&\quad + \beta_1^F F_1 + \beta_2^F F_2 + \beta_1^S S_1 + \beta_2^S S_2 + \beta_2^H H_2 + \beta_3^H H_3 \\
&\quad + \beta_2^E E_2 + \beta_3^E E_3 + \beta_4^E E_4 + \beta_5^E E_5 + \beta_6^E E_6 + \beta_7^E E_7
\end{aligned}
$$

We will extend this afterwards using forward selection of interactions using AIC and significance tests. Using **R**, we derive table Table 3.10 which shows the estimates for the initial model.

It can be seen in the initial estimates that family characteristics and household condition are significant while the three intervention programs spanned for two years are not significant. However, this does not mean that we have to remove it from the model at this stage since it is possible that one interacts significantly with another covariate. Thus, we will implement a forward selection of covariate interaction instead to see whether there is indeed a set of significant interactions of variables that affect movement in welfare status.

First, we fit a logistic regression model that includes all possible two-factor interactions. Some coefficients have huge standard errors due to the lack of information to evaluate interaction effects. We look at some of the variable pairs, particularly their cell counts, to assess whether including interactions can be sensible. To start with, Table 3.11 shows the frequency of families that received scholarship program in 1999 by educational attainment of the family head.

Notice that there is only one family with family head who did not finish any schooling and received scholarship in 1999. Furthermore, there

x_j	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	Wald Z	$P(\mathcal{Z} > Z)$	
(Intercept)	-1.88	0.241	-7.808	0.000	***
G	0.35	0.116	3.026	0.002	**
U	0.30	0.069	4.254	0.000	***
A	0.02	0.003	8.110	0.000	***
C_1	-0.15	0.295	-0.512	0.609	
C_2	0.38	0.248	1.528	0.127	
F_1	-0.02	0.225	-0.090	0.929	
F_2	0.09	0.225	0.403	0.687	
S_1	0.32	0.231	1.370	0.171	
S_2	0.44	0.297	1.487	0.137	
H_2	-0.50	0.211	-2.355	0.019	*
H_3	-0.07	0.206	-0.329	0.742	
E_2	-0.36	0.118	-3.039	0.002	**
E_3	-0.40	0.098	-4.054	0.000	***
E_4	-0.60	0.151	-3.961	0.000	***
E_5	-0.71	0.136	-5.243	0.000	***
E_6	-0.80	0.125	-6.416	0.000	***
E_7	-0.39	0.147	-2.678	0.007	**

Signif. codes: 0.000 *** 0.001 ** 0.01 * 0.05 . 0.1 1
AIC: 5399.2   Deviance: 5363.2   Df: 5283

Table 3.10: Estimates for the Initial Logistic Regression Model

are only five families with heads who reached high school and received scholarship. These small frequencies signal the need to collapse the categories due to limited information in the existing levels. In the model, we will instead use the variable 'heduc' with the merged categories: no grade completed, elementary and elementary graduate into "NG-Elem"; high school level and high school graduate into "HS"; and college level and college graduate into "Coll". The cell counts are shown in Table 3.12.

This will in turn reduce our dummy variables for the educational attainment to two: E_2 which indicates whether the head reached or graduated from high school and E_3 which is the indicator when the head reached at most elementary. When both E_2 and E_3 are zero then the head reached

Educational attainment	Received scholarship		Total
	Yes	No	
College graduate	15	953	968
College level	13	670	683
High School graduate	11	1429	1440
High School Level	5	394	399
Elementary graduate	3	565	568
Elementary level	5	804	809
No grade completed	1	433	434
Total	53	5248	5301

Table 3.11: Number of families who availed scholarship in 1999 by educational attainment of the head

Educational attainment	Received scholarship		Total
	Yes	No	
Coll	28	1623	1651
HS	16	1823	1839
NG-Elem	9	1802	1811
Total	53	5248	5301

Table 3.12: Number of families who availed scholarship in 1999 by (recoded) educational attainment of the head

or graduated from tertiary.

Moving on, there is also limited information that can be drawn from the variable interaction between 1998 CARP and 1999 GHFP. Table 3.13 shows that there are only three families who received CARP in 1998 then availed GHFP in 1999. Therefore, we do not have enough information to consider $C_1 \times F_2$ interaction for the model.

In the case of 1999 CARP, its cross-tabulation with 1999 scholarship program has limited information as well. Table 3.14 shows that we have limited information on the number of families who received both CARP in 1999 and scholarship for tertiary education in 1999. This means that we can not include this interaction in the model as well.

Received CARP	Received GHFP		Total
	Yes	No	
Yes	3	77	80
No	123	5098	5221
Total	126	5175	5301

Table 3.13: Number of families who received CARP in 1998 and received GHFP in 1999

Received CARP	Received scholarship		Total
	Yes	No	
Yes	1	94	95
No	52	5154	5206
Total	53	5248	5301

Table 3.14: Number of families who availed scholarship and CARP in 1999

The interaction between the 1998 GHFP and scholarship program in 1998 and 1999 also can not be included in the model. Table 3.15 shows that only seven families received both programs in 1998 while only two families who received GHFP in 1998 received scholarship in 1999.

Received GHFP	Received 1998 scholarship		Received 1999 scholarship		Total
	Yes	No	Yes	No	
Yes	7	117	2	122	124
No	88	5089	51	5126	5177
Total	95	5206	53	5248	5301

Table 3.15: Number of families who availed scholarship in 1998-1999 by availment of GHFP in 1998

Household condition's interaction with 1998 and 1999 CARP cannot be also included as suggested by Table 3.16. There are only two families with poor household condition who accessed CARP in 1998. This lack of sufficient information in this subgroup led us to remove this interaction from the model.

Household	Received 1998 CARP		Received 1999 CARP		Total
condition	Yes	No	Yes	No	
bad	2	138	1	139	140
fair	20	138	16	1867	1883
good	58	3220	78	3200	3278
Total	80	5221	95	5206	5301

Table 3.16: Number of families who availed CARP by household condition

Now, after having identified the unsuitable interactions for the model, we now set the the upper scope for the forward model selection. We select the full two-factor interaction of the covariates except the interactions between 1999 CARP and 1999 scholarship ($C_2 \times S_2$), receiving GHFP in 1998 and scholarship in 1998 ($F_1 \times S_1$), and 1999 CARP and dummy variables for household condition ($C_2 \times H_2$ and $C_2 \times H_3$). That is, starting with the simple model (revised coefficients in Table B.1), we will utilise the iterative model selection in **R** through the AIC in determining whether adding further effects will improve model fit. Recall in Chapter 2 that AIC measures the adequacy of the model given the likelihood and number of parameters. Using this concept, the procedure will iterate in adding significant terms until there is no improvement in AIC yielding additional effects in Table 3.17.

It can be seen that each AIC value gradually decreases after an interaction was added. For each step, the added effect is significant based on the likelihood-ratio test $G^2(\mathcal{M}_{k-1} \mid \mathcal{M}_k)$ at $\alpha = 0.15$.

The main effect of 1999 GHFP, F_2, is not significant with p-value 0.7 (refer to Table B.2) and since it has no interaction with any other covariate, we shall remove it from the model as well. Although the p-value for $C_1 \times C_2$ is 0.144 (not significant at $\alpha = 0.10$), we still keep it in the model due to the AIC selection result. This leaves us with the final logistic regression

k	Step	Δdf	$G^2(\mathcal{M}_{k-1}|\mathcal{M}_k)$	df	$G^2(\mathcal{M}_k)$	AIC
1				5287	5382.099	5410.099
2	$+ (E_2 + E_3) \times S_2$	2	7.773	5285	5374.326	5406.326
3	$+ U \times S_2$	1	2.825	5284	5371.502	5405.502
4	$+ G \times S_1$	1	2.341	5283	5369.161	5405.161
5	$+ (E_2 + E_3) \times F_1$	2	4.187	5281	5364.973	5404.973
6	$+ A \times C_1$	1	2.026	5280	5362.948	5404.948
7	$+ (E_2 + E_3) \times U$	2	4.172	5278	5358.775	5404.775
8	$+ C_2 \times C_1$	1	2.126	5277	5356.649	5404.649

Table 3.17: AIC history in adding effects for the Logistic Regression Model

model in Table 3.18 which we can write as in Equation (3.1)

$$
\begin{aligned}
\log\left(\frac{\hat{\pi}(\mathbf{x})}{1 - \hat{\pi}(\mathbf{x})}\right) = {} & -2.14 + 0.32G + 0.46U + 0.02A + 1.25C_1 + 0.24C_2 - 0.63F_1 \\
& -0.5H_2 - 0.06H_3 + 0.19S_1 - 0.32S_2 - 0.4E_3 - 0.21E_2 \\
& +1.9(E_3 \times S_2) - 0.83(E_2 \times S_2) + 1.02(U \times S_2) + 1.15(G \times S_1) \\
& +0.84(F_1 \times E_3) + 1.08(F_1 \times E_2) - 0.04(A \times C_1) \\
& -0.35(U \times E_3) - 0.19(U \times E_2) + 0.94(C_1 \times C_2) \quad (3.1)
\end{aligned}
$$

3.2.2 Final model findings

Now, we examine how do the covariates explain improvement of families. In particular, we analyse the intervention programs that spanned for two years in relation to some characteristics of families. We use the ratio of the odds of improvement given a set of attributes to the odds of improvement given another set of characteristics.

For instance, if we have two families a and b with given values for $\mathbf{x}$ namely $\mathbf{x}_a$ and $\mathbf{x}_b$, we can estimate the odds ratio of the two sets of values using Equation 2.9 as in Equation 3.2

x_j	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	Wald Z	$P(\mathcal{Z} > Z)$	
(Intercept)	-2.14	0.244	-8.756	0.000	***
G	0.32	0.118	2.748	0.006	**
U	0.46	0.117	3.965	0.000	***
A	0.02	0.003	8.387	0.000	***
C_1	1.25	1.159	1.081	0.280	
C_2	0.24	0.283	0.851	0.395	
F_1	-0.63	0.425	-1.488	0.137	
H_2	-0.50	0.211	-2.347	0.019	*
H_3	-0.06	0.206	-0.311	0.756	
S_1	0.19	0.257	0.744	0.457	
S_2	-0.32	0.642	-0.501	0.617	
E_2	-0.21	0.118	-1.765	0.077	.
E_3	-0.40	0.121	-3.287	0.001	**
$E_2 \times S_2$	-0.83	0.79	-1.050	0.294	
$E_3 \times S_2$	1.90	0.868	2.188	0.029	*
$U \times S_2$	1.02	0.696	1.467	0.142	
$G \times S_1$	1.15	0.768	1.504	0.133	
$F_1 \times E_2$	1.08	0.534	2.018	0.044	*
$F_1 \times E_3$	0.84	0.595	1.415	0.157	
$A \times C_1$	-0.04	0.023	-1.542	0.123	
$U \times E_2$	-0.19	0.164	-1.172	0.241	
$U \times E_3$	-0.35	0.174	-2.012	0.044	*
$C_1 \times C_2$	0.94	0.643	1.462	0.144	

Signif. codes: 0.000 *** 0.001 ** 0.01 * 0.05 . 0.1 1
AIC: 5402.8   Deviance: 5356.8   Df: 5278

Table 3.18: Estimates for the Final Logistic Regression Model

$$\hat{\theta}(a, b) = \exp\left\{\mathbf{x}_a^T \hat{\boldsymbol{\beta}} - \mathbf{x}_b^T \hat{\boldsymbol{\beta}}\right\} \qquad (3.2)$$

Using this, we evaluate the model and draw findings with regards to access to intervention programs and family characteristics. We will focus on a set of scenario and *holding the rest of independent variables constant*. We

begin by analysing the impact of GHFP.

- *Receiving 1998 GHFP*

  If a family received 1998 GHFP, the odds of its improvement compared to those who did not can be expressed by

  $$\hat{\theta}\left(\{F_1 = 1\}, \{F_1 = 0\}\right) = \exp\{-0.63 + 0.84E_3 + 1.08E_2\}$$

  given all other independent variables at a fixed level. The odds ratio implies that the program has no impact on those families with heads who reached at least college or tertiary levels with odds ratio $e^{-0.63} = 0.53$ while has an impact on those families who reached at least most elementary level or high school. If we focus on those families with heads who reached elementary the highest ($E_3 = 1, E_2 = 0$), the odds of them improving if they received the program is $e^{0.21} = 1.23$ times that of those who did not receive the program. On the other hand, among those families who have heads who reached high school ($E_3 = 0, E_2 = 1$), their odds of improvement if they received the program is $e^{0.45} = 1.57$ times that if they did not benefit from the program.

After finding out that GHFP is more helpful among those families with heads with lower educational attainment (no grade to secondary) than those families with heads who reached at least tertiary, we now continue on interpreting the effect of CARP.

- *Receiving CARP 1998 only compared to receiving neither 1998 nor 1999 CARP*

  The odds of improvement among families who benefited from CARP in 1998 but not in 1999 compared to those who were not able to access them both is

  $$\hat{\theta}\left(\{C_1 = 1, C_2 = 0\}, \{C_1 = 0, C_2 = 0\}\right) = \exp\{1.25 - 0.04A\}$$

given all other independent variables at a fixed level. This means that the odds ratio depends on the age of the family head and that as age of the head increases, the ratio decreases. In other words, the older the head of the family who received CARP in 1998, its edge in improvement over the family who did did not receive any CARP decreases. Furthermore, when the head is a middle-aged person who has reached 31 and a quarter years old, the odds starts to favour those who did not receive the program since the value of the odds ratio at that point is $e^{1.25-0.04\times31.25} = 1$. This can be seen in Figure 3.1.

If we examine the distribution of family head's age in figure 3.2, we can see that they are dominated young persons with 16 years old as the youngest (Table 3.19). This shows that lower values of age have a great influence on the model. Moreover, if we compare families with heads aged 16 years old, the odds of moving out of poverty if a family received CARP in 1998 is $e^{1.25-0.02\times16} = 1.84$ times that of those who did not receive. On the other hand, if we compare those families headed by 94-year olds, the odds of improvement of those who received the program is $e^{1.25-0.04\times94} = 0.08$ compared to those who did not receive the program.

The finding makes sense since most likely, family heads who are mostly young people will till the agricultural land being distributed through CARP. Those heads who are relatively older and was granted with the program will likely underutilise the land and the positive impact of the program in improvement will not be higher than those who did not receive the program.

Min	1st Q	Median	Mean	3rd Q	Max	Std. Dev.
16	37	44	46.25	54	97	13.06

Table 3.19: Descriptive statistics, Age of family head

Figure 3.1: Plot of age versus odds ratio of those who received CARP in 1998 compared to those who did not receive CARP in 1998 and 1999

- *Receiving CARP 1999 only compared to receiving neither 1998 nor 1999 CARP*

  If we look at a family who received CARP in 1999 only compared to a family who did not benefit from both, we have the following ratio of the odds

  $$\hat{\theta}\left(\{C_1 = 0, C_2 = 1\}, \{C_1 = 0, C_2 = 0\}\right) = \exp\{0.24\}$$

  This means means that those who received CARP in 1999 has odds of improving 1.27 times that of a family who did not receive both programs.

Figure 3.2: Histogram of Age of Family Heads who received 1998 CARP

- *Receiving CARP in 1998 and 1999 compared to receiving neither*

  Looking at those families who received the program in both years 1998 and 1999, their chance of improvement compared to those who did not receive both programs can be expressed by

$$
\begin{aligned}
\hat{\theta}\left(\{C_1 = 1, C_2 = 1\}, \{C_1 = 0, C_2 = 0\}\right) &= \exp\{0.94 - 0.04A + 0.24 + 1.25\} \\
&= \exp\{2.43 - 0.04A\}
\end{aligned}
$$

  Similar to comparing a family who received only 1998 CARP to a family who did not receive both, the odds ratio here depends also on age of the family head and that the odds ratio decreases for increasing age. However, the difference is the odds ratio is higher due

to the amount augmented by the program follow up in 1999. Figure 3.3 demonstrates that although the odds ratio is decreasing, the odds favours those who received the program in both years until their heads become a senior citizen [3] or about 60.75 years old. In that case, the odds of improvement of those who received both programs against those who did not receive the program in any of 1998 and 1999 is 1.

In comparison with the finding in the recipient of 1998 CARP case, if a family with median-aged (44 years old) head is a recipient of CARP 1998 but stopped being a beneficiary in 1999, its chance of improvement is not as high as that when they did not receive 1998 CARP with odds ratio 0.6. However, if the family received a 1999 follow up, then its odds of improvement becomes $e^{2.43-0.04\times44} = 1.95$ times those who did not receive both programs.

We can recover other comparisons for different beneficiary scenarios such as receiving both 1998 and 1999 program compared to receiving only 1999 or receiving only 1998 using the aforementioned odds ratios. If a family received only 1999 CARP, its odds of improvement is $e^{0.24}/e^{1.24-0.04A} = e^{0.04A-1}$ times that of a family who received only the 1998 one. The odds ratio increases on age (refer to Figure 3.4) and favours those who have received only 1999 CARP over the 1998 only when the head reached 25 years old. If a family received both, then its odds of moving out of poverty is $e^{2.43-0.04A}/e^{1.24-0.04A} = 3.3$ times that of a family which received only 1998. Finally, if the same family which received both programs is compared to a family which received only the 1999 one, then the odds ratio is $e^{2.43-0.04A}/e^{0.24} = e^{2.19-0.04A}$. The ratio shows that the chance of improvement favours only those who received the 1999 over those who received both when the family head is about 55 years old.

Now, let us see how scholarship programs affect the dynamics of the

---

[3]Defined as those persons aged 60 years old and above.
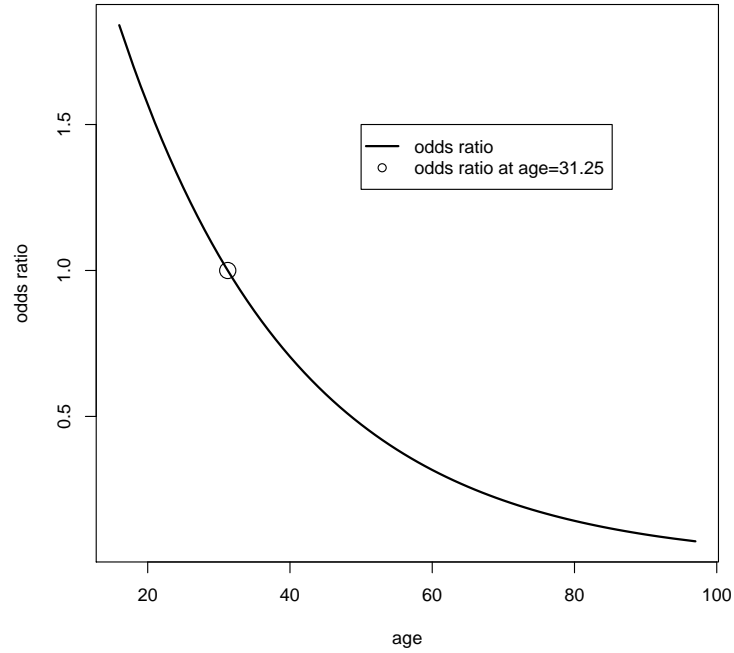
Figure 3.3: Plot of age versus odds ratio of those who received CARP in 1999 compared to those who did not receive CARP in 1998 and 1999

families in the panel.

- *Receiving 1998 scholarship for tertiary education*

  If we examine those families who received 1998 scholarship program, their odds of improvement compared to those who did not benefit from the program can be computed as

  $$\hat{\theta}\left(\{S_1 = 1\}, \{S_1 = 0\}\right) = \exp\{0.19 + 1.15G\}$$

  which depends on the gender of the family head. The estimate of the odds ratio here shows that among those families headed by males, the chance that a family who received scholarship in 1998 is $e^{0.19} =$

Figure 3.4: Plot of age versus odds ratio of those who received CARP in 1999 only compared to those who received CARP in 1998 only

1.21 times that of a family who did not. In the case of female headed families, those who have received the program have odds of improving $e^{1.34} = 3.82$ times that of those who were not able to receive it. The edge among the females is considerably higher than that among males.

- *Receiving 1999 scholarship for tertiary education*

  The odds ratio of improvement of those families who received 1999 scholarship compared to those who did not receive scholarship is expressed by

  $$\hat{\theta}\left(\{S_2 = 1\}, \{S_2 = 0\}\right) \quad = \quad \exp\{-0.32 - 0.83E_2 + 1.9E_3 + 1.02U\}$$

Similar to the previous case, receiving 1999 scholarship will not have more impact among those whose heads reached at least college level residing in rural area due to the odds ratio $e^{-0.32} = 0.73$. Should these families reside in urban area, then receiving the 1999 scholarship will have odds of improvement $e^{0.7} = 2.01$ compared to not benefiting from it.

On the other hand, among those families with heads who reached high school residing in rural area, receiving the 1999 scholarship program does not help them move out of poverty more than when they did not receive it since the odds ratio is $e^{-1.15} = 0.32$. Even if they reside in urban area, the chance of moving out of poverty if they received the program is not more than those who did not receive it due to the odds ratio $e^{-0.13} = 0.88$.

Finally, the impact of the program is visible among those families headed by persons who reached at most elementary level. If these families reside in rural area, the odds of improving among these families who received 1999 scholarship is $e^{1.58} = 4.84$ times that of these families who did not receive it. If these families reside in the urban area, the odds of improving of those who received the program is $e^{2.6} = 13.46$ times those who did not receive the program.

Improvement of those who received 1998 scholarship depends on gender of the family head wherein odds ratio of improvement among females is higher than that of males. In terms of benefiting from 1999 scholarship, the odds ratio depends on educational attainment of the head and urbanity. Impact is visible among families whose heads reached at most elementary particularly in urban. Apparently, scholarship program impacts those who need it the most and also where it can be utilised optimally, i.e. urban area since universities and colleges are most likely located.

We have drawn how programs affect improvement of families and in doing so, we found some relationship with family characteristics. To com-

plete the picture, focusing on each of the family characteristics and holding the rest of the covariates fixed at one level gives us the following findings:

- Families headed by females have higher chance of moving out of poverty compared to those headed by males as evidenced by the following expression for the odds ratio

$$\hat{\theta}\left(\{G=1\},\{G=0\}\right) \;=\; \exp\{0.32+1.15S_1\}$$

  In particular, among those families who did not receive 1998 scholarship, the odds of improving among female-headed families is 1.38 times that of male-headed ones. Furthermore, odds favours those families with female heads even more among those who received 1998 scholarship with odds ratio 4.35 ($= e^{0.32+1.15}$).

- Urban families generally have higher chance improving compared to those in rural areas. We can verify it through

$$\hat{\theta}\left(\{U=1\},\{U=0\}\right) \;=\; \exp\{0.46+1.02S_2-0.19E_2-0.35E_3\}$$

  Odds ratio is highest among those families whose heads reached college level and have availed scholarship, 4.39 ($= e^{0.46+1.02}$) which means their odds of improvement is more than four times those in rural areas. Lowest ratio is among those with heads who reached at most elementary level and did not receive scholarship, 1.12 ($= e^{0.46-0.35}$) but still they have more chance of improving than those in rural areas.

- In terms of age, families with older heads have higher chance of moving out of poverty compared to younger ones among those non-beneficiaries of 1998 CARP. However, among those beneficiaries of 1998 CARP, older ones have less chance of moving out of poverty compared to younger ones. To illustrate, for every 10-year difference

in age of the family head, the odds ratio is

$$\hat{\theta}\left(\{A = a + 10\}, \{A = a\}\right) \;\; = \;\; \exp\{0.2 - 0.4C_1\}$$

This means that among non-recipient families of 1998 CARP, the improvement odds of the family with older head is 1.22 ($= e^{0.2}$) times that of the younger one. However, among recipients, the odds favours the families with the younger head with odds of improvement 1.22 ($= e^{-(0.2-0.4)}$) times those with older ones.

- Those families with heads who reached higher levels of education have higher chance of moving out of poverty compared to those with lower ones. To wit, let us just compare families with heads who reached college with those with heads who reached at most elementary

$$\hat{\theta}\left(\{E_3 = 0, E_2 = 0\}, \{E_3 = 1, E_2 = 0\}\right) \;\; = \;\; \exp\{0.40 - 1.9S_2 - 0.84F_1 + 0.35U\}$$

Based on above, among rural families who were neither recipients of 1999 scholarship nor 1998 GHFP, those headed by persons who reached college have odds of improvement 1.49 ($= e^{0.40}$) times those with heads who reached at most elementary level. In the same subgroup but in the urban area, the odds favours those headed by college level attendees even more with more than twice the odds of those headed by persons who reached at most elementary. The odds only favours those with heads who reached at most elementary among those who received either scholarship, GHFP or both.

- Household condition is the only family characteristic variable which has no interaction with any of the other covariates. Hence, the effect of ($H_3 = 1, H_2 = 0$) insignificant based on the Wald test with p-value 0.76. On the other hand, those living in poor condition housing

has odds of moving out of poverty $e^{0.05} = 1.65$ times those living in moderate condition housing.

After culling all of these information, we now have a richer picture of how do programs affect well-being as well as characteristics of families and their households. In the next section, let us see how these programs contribute to improving income of the families and how do they relate to other variables.

3.3 Multiple linear regression among 1997 poor families

After finding out factors that affect moving out of poverty, here we aim to analyse status of families in terms of change in their income. This will be useful in determining what variables are related in increase or decrease in income. The raw measure of income is family income, but we will use the per capita income (PCI) in 1999 'pci99' as the response variable for scalability and comparability which is computed by dividing the total family income by family size.

On the other side, the explanatory variables will include the PCI in 1997 and 1998 as predictors. Although it is expected that they must have strong correlation with the response variable, it can be useful in describing how much change in the response variable will be brought by values of 1997 and 1998 income. Since we are looking at a longitudinal data, there will be an apparent change in magnitude of income due to inflation. Hence, the PCI variables will be transformed into distance from poverty line to reflect the deficit / surplus of families compared to the standard in each year. Now we have our Y as the distance of 1999 PCI from standard while M_1 and M_2 are distances of 1997 and 1998 PCI from respective standards[4].

---

[4]Poverty threshold in each year.

These together with other explanatory variables in Table 3.20 will make up our model.

j	x_j	Variable(level)	Description	Value
1	G	gender	family head's gender	1 - female; 0 - male
2	U	urb	Urbanity	1 - urban; 0 - rural
3	A	age99	family head's age	
4	C_1	carp98	acquired agricultural land thru CARP in 1998	1 - yes; 0 - no
5	C_2	carp99	acquired agricultural land thru CARP in 1999	1 - yes; 0 - no
6	F_1	hsfg98	acuired house / lot through GHFP in 1998	1 - yes; 0 - no
7	F_2	hsfg99	acuired house / lot through GHFP in 1999	1 - yes; 0 - no
8	S_1	scholp98	availed scholarship for tertiary education in 1998	1 - yes; 0 - no
9	S_2	scholp99	availed scholarship for tertiary education in 1999	1 - yes; 0 - no
10	H_2	hhcond(2)	fair household condition	1 - if hhcond = 2; 0 - otherwise
11	H_3	hhcond(3)	good household condition	1 - if hhcond = 3; 0 - otherwise
12	E_2	heduc(HS)	family head's highest educational attainment is high school	1 - if heduc = 'HS'; 0 - otherwise
13	E_3	heduc(NG-Elem)	family head's highest educational attainment is at most elementary	1 - if heduc = 'NG-Elem'; 0 - otherwise
14	M_1	dpci97	distance of PCI in 1997 from standard	
15	M_2	dpci98	distance of PCI in 1998 from standard	

Table 3.20: Explanatory Variables for the Multiple Linear Regression

Given our previous analyses, we adopt the revised categories for educational attainment of the family head to have sufficient number of observations in each subgroup should we analyse factor interactions. We again focus our analysis on a subpopulation which is the poor families in 1997. This may help in determining how do family incomes change among those who started poor in the panel. We first examine the distribution of

distance of PCI from standard in each year through Figures 3.5, 3.6, and 3.7.

Figure 3.5: Histogram of M_1, 1997 poor families

PCI distance in 1997, M_1, is expected to be skewed to the left since we are looking at families who are poor in that year. However, it is noticeable through Figures 3.6 and 3.7 that the histograms in 1998 and 1999 are skewed to the right. It can be verified that this can bring problems on the residuals of the model most especially due to the fact that we are using the 1999 income as the response. Recall that in linear regression case of the GLM, we assume that our response variable Y follows the normal distribution[5]. Although we are most concerned in stabilising the response

---

[5]In classical linear regression, only the residuals are assumed to be normally distributed, centered at zero and with constant variance. However, in this study, stabilisa-

Figure 3.6: Histogram of M_2, 1997 poor families

variable and achieving its normality, we need some state of comparability among the three income variables. Hence, this signals that we need to transform the three variables.

We attempt to stabilise the data using logarithmic [6] transformation of PCI distances in 1999, 1998, and 1997 and we use the variables $\log(Y)$, $\log(M_2)$, and $\log(M_1)$ instead. However, this will not be directly applicable due to negative values (deficits) when below the standard. This can be remedied by adding an arbitrary constant c on the three variables. The chosen figure to be added to the three variables is $c = 16,000$ defining new

---

tion of the response variable as well as the other income variables are needed in attaining these assumptions on the residuals. See Figures B.1 and B.2.

[6]We are referring to the natural logarithm here and usage of $\log(\cdot)$ is of base e

Figure 3.7: Histogram of Y, 1997 poor families

variables $Y^* = Y + c$, $M_1^* = M_1 + c$, and $M_2^* = M_2 + c$ and new logarithms $\log(Y^*)$, $\log(M_1^*)$, and $\log(M_2^*)$. After applying this, we have Figures 3.8, 3.9 and 3.10.

It is now evident that after stabilising the income variables, the response variable $\log(Y^*)$ now resemble a bell-shaped curve (Figure 3.10) which satisfies our estimation assumption in the linear regression. From this point forward, we will refer to $\log(Y^*)$ as income in 1999, $\log(M_2^*)$ income in 1998 and $\log(M_1^*)$ as income in 1997 for simplicity.

Figure 3.8: Histogram of $\log(M_1^*)$, 1997 poor families

3.3.1 Model estimation

After satisfying data assumptions, we will repeat the procedure done in logistic regression. That is, given our explanatory variables in Table 3.20, we start with the simple additive model and perform forward recursion afterwards. Table 3.21 shows the initial estimates from this procedure.

It can be seen that some programs are not significant such as GHFP in 1998 and 1999 and scholarship in 1998 (F_1, F_2, S_1). However, we will attempt to find a fitting interaction model using forward recursion on the initial model. Since we have identified several interactions with limited information in our logistic regression analysis, we will omit them also in the upper scope of our iteration. As a result, we have Table 3.22 showing

Figure 3.9: Histogram of $\log(M_2^*)$, 1997 poor families

the history of effects being added while there is improvement in the AIC. Based on this, we derived a three-factor interaction namely income in 1998, urbanity and 1998 CARP ($\log(M_2^*) \times U \times C_1$).

After identifying possible interactions for the model, it can be verified that the interaction between 1998 income and 1998 scholarship ($M_1 \times S_1$) and 1999 GHFP (F_2) are not significant (refer to Table B.3). Hence, we drop them from the model after forward selection, giving us coefficient estimates in Table 3.23. It appears that we have a considerably complicated model which includes all the programs but F_2 plus a three-factor interaction $\log(M_2^*) \times U \times C_1$.

In validation of the state of the model residuals, the plot of the error estimates versus fitted values $\widehat{\log(Y^*)}$ in Figure 3.11 shows a cloud pattern

Figure 3.10: Histogram of $\log(Y^*)$, 1997 poor families

which tells us that there is no alarming problem with the error terms. Furthermore, Figure 3.12 shows that the residuals follow a bell shaped curve. Hence, we can write our final model as in Equation (3.3)

x_j	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	Wald Z	$P(\mathcal{Z} > Z)$	
(Intercept)	0.47	0.261	1.796	0.073	.
$\log(M_1^*)$	0.54	0.03	18.099	0.000	***
$\log(M_2^*)$	0.40	0.014	27.732	0.000	***
G	0.04	0.019	2.076	0.038	*
U	0.02	0.01	1.579	0.114	
A	0.003	0.004	6.960	0.000	***
C_1	-0.09	0.042	-2.051	0.04	*
C_2	0.10	0.039	2.691	0.007	**
F_1	0.01	0.033	0.342	0.732	
F_2	0.01	0.033	0.224	0.822	
S_1	0.04	0.037	0.999	0.318	
S_2	0.09	0.05	1.748	0.081	.
H_2	-0.03	0.031	-1.105	0.269	
H_3	0.00	0.031	0.099	0.921	
E_2	-0.02	0.012	-1.365	0.172	
E_3	-0.02	0.013	-1.829	0.067	.

Signif. codes: 0.000 *** 0.001 ** 0.01 * 0.05 . 0.1   1
AIC: 4106.2   Deviance: 669.12   Df: 5285

Table 3.21: Estimates for the initial Linear Regression, 1997 poor families

$$
\begin{aligned}
\widehat{\log(Y^*)} \;=\; & 0.68 + 0.46\log(M_1^*) + 0.45\log(M_2^*) + 0.04G - 0.65U + 0.002A + 1.29C_1 \\
& 0.07C_2 + 1.23F_1 - 0.04H_2 + 0.003H_3 + 0.82E_3 - 0.42E_2 - 3.53S_1 + 0.09S_2 \\
& + 0.05[\log(M_2^*) \times E_2] - 0.09[\log(M_2^*) \times E_3] - 0.14[\log(M_2^*) \times C_1] \\
& - 0.07(U \times E_2) - 0.05(U \times E_3) + 0.38[\log(M_1^*) \times S_1] \\
& + 3.75(U \times C_1) + 0.20(U \times C_2) + 0.001(U \times A) + 0.16(F_1 \times E_2) \\
& + 0.06(F_1 \times E_3) - 0.14[\log(M_2^*) \times F_1] + 0.12[\log(M_1^*) \times U] \\
& - 0.05[\log(M_2^*) \times U] - 0.43[\log(M_2^*) \times U \times C_1]
\end{aligned}
\tag{3.3}
$$

Using this, we will proceed on analysing the effects and draw meaning from the model. We will utilise the formula we established in Equation 2.8

k	Step	Δdf	$G^2(\mathcal{M}_{k-1}|\mathcal{M}_k)$	df	$G^2(\mathcal{M}_k)$	AIC
1				5285	669.118	4106.155
2	$+ \log(M_2^*) \times (E_2 + E_3)$	2	2.210	5283	666.907	4092.615
3	$+ \log(M_2^*) \times C_1$	1	1.015	5282	665.892	4086.538
4	$+ U \times (E_2 + E_3)$	2	0.817	5280	665.075	4084.03
5	$+ \log(M_1^*) \times S_1$	1	0.386	5279	664.689	4082.953
6	$+ U \times C_1$	1	0.398	5278	664.291	4081.778
7	$+ U \times C_2$	1	0.562	5277	663.729	4079.291
8	$+ A \times U$	1	0.363	5276	663.366	4078.391
9	$+ F_1 \times (E_2 + E_3)$	2	0.574	5274	662.792	4077.799
10	$+ \log(M_2^*) \times S_1$	1	0.275	5273	662.517	4077.6
11	$+ \log(M_2^*) \times F_1$	1	0.250	5272	662.266	4077.596
12	$+ \log(M_1^*) \times U$	1	0.253	5271	662.014	4077.574
13	$+ \log(M_2^*) \times U$	1	0.502	5270	661.512	4075.555
14	$+ \log(M_2^*) \times C_1$	1	0.681	5269	660.831	4072.094

Table 3.22: AIC history in adding effects for the Linear Regression Model, 1997 poor families

to interpret findings from the model. Change in $\log(Y^*)$ will be examined given sets of $\mathbf{x}$ values from two different scenarios, thus we have $\hat{\Delta}(a, b)$ as in Equation 3.4

$$\hat{\Delta}(a, b) = \widehat{\log(Y_a^*)} - \widehat{\log(Y_b^*)} = \log(\widehat{Y^* \mid \mathbf{x}} = \mathbf{x}_a) - \log(\widehat{Y^* \mid \mathbf{x}} = \mathbf{x}_b) \quad (3.4)$$

3.3.2 Model interpretation

Before we interpret what the model tells us given our estimates, it is worth noting that there can be some findings that will not be consistent with our existing notion since we are focusing on a subgroup of our data, that is the poor in 1997. For instance, 1999 income will not necessarily be increasing on 1998 income when we focus on some programs being received. Furthermore, we first note that increase / decrease in the response variable does not directly translate to change in income, but the logarithm of

x_j	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	Wald Z	$P(\mathcal{Z} > Z)$	
(Intercept)	0.68	0.407	1.669	0.095	.
$\log(M_1^*)$	0.46	0.04	11.559	0.000	***
$\log(M_2^*)$	0.45	0.029	15.524	0.000	***
G	0.04	0.019	2.109	0.035	*
U	-0.65	0.521	-1.255	0.209	
A	0.002	0.001	4.253	0.000	***
C_1	1.29	0.895	1.446	0.148	
C_2	0.07	0.043	1.530	0.126	
F_1	1.23	0.855	1.443	0.149	
H_2	-0.04	0.031	-1.362	0.173	
H_3	0.003	0.031	-0.090	0.929	
E_2	-0.42	0.308	-1.376	0.169	
E_3	0.82	0.302	2.703	0.007	**
S_1	-3.53	1.921	-1.839	0.066	.
S_2	0.09	0.05	1.812	0.070	.
$\log(M_2^*) \times E_2$	0.05	0.033	1.419	0.156	
$\log(M_2^*) \times E_3$	-0.09	0.032	-2.718	0.007	**
$\log(M_2^*) \times C_1$	-0.14	0.094	-1.501	0.133	
$U \times E_2$	-0.07	0.025	-2.672	0.008	**
$U \times E_3$	-0.05	0.026	-1.772	0.076	.
$\log(M_1^*) \times S_1$	0.38	0.203	1.860	0.063	.
$U \times C_1$	3.75	1.716	2.188	0.029	*
$U \times C_2$	0.20	0.096	2.097	0.036	*
$U \times A$	0.001	0.001	1.786	0.074	.
$F_1 \times E_2$	0.16	0.077	2.052	0.040	*
$F_1 \times E_3$	0.06	0.083	0.713	0.476	
$\log(M_2^*) \times F_1$	-0.14	0.091	-1.519	0.129	
$\log(M_1^*) \times U$	0.12	0.059	2.019	0.043	*
$\log(M_2^*) \times U$	-0.05	0.029	-1.784	0.075	.
$\log(M_2^*) \times U \times C_1$	-0.43	0.184	-2.314	0.021	*

Signif. codes: 0.000 *** 0.001 ** 0.01 * 0.05 . 0.1 1

AIC: 4069.9 　　　　　　　　Deviance: 661.05　Df: 5271

Table 3.23: Estimates for the final Linear Regression Model, 1997 poor families

Figure 3.11: Residuals versus Fitted, Regression (1997 poor families)

it. Hence, we talk about change here and not magnitude since we cannot directly recover the amount in terms of 1999 income. In addition, our $\hat{\Delta}(a, b)$ apparently appears to be difficult to interpret or to draw meaning from. Thus, it is better to interpret them by exponentiation. That is, since $\widehat{\log(Y_a^*)} - \widehat{\log(Y_b^*)} \approx \widehat{\log(Y_a^*/Y_b^*)}$ (Equation 3.4), and we ensured that all Y_i^*s are positive, then $\exp\{\hat{\Delta}(a, b)\} \approx \widehat{Y_a^*/Y_b^*}$ will enable us to talk about the ratio of 1999 incomes and we can draw results similar to what we have done in logistic regression. Using this, we have some of the aspects being implied by the model in terms of change in 1999 income at different scenarios of intervention programs given all other explanatory variables at a fixed level. We begin by analysing the impact of F_1.

- *Receiving 1998 GHFP*

Figure 3.12: Histogram of residuals, Regression (1997 poor families)

Difference in 1999 income among those who have received GHFP in 1998 compared to those who did not is given by

$$\hat{\Delta}(\{F_1 = 1\}, \{F_1 = 0\}) \quad = \quad 1.23 + 0.16 E_2 + 0.06 E_3 - 0.14 \log(M_2^*)$$

which depends on the 1998 income and educational attainment of the head. The difference in 1999 income decreases as 1998 income increases. This means that the program will have an impact on the 1999 income among families with relatively low income in 1998. In addition, highest difference can be seen among those with heads who have finished high school while lowest among those who have reaced college. This can be seen in Figure 3.13. To illustrate an ex-

ample, let us look at the plot and go back to the original measurement of our variables. Those who received GHFP with family heads who reached college are likely to have higher 1999 income if the *actual* distance of their 1998 PCI from standard is below $e^{8.79} - c = 6,568.23 - 16,000 = -9,431.77$. Since the sought value of 1998 PCI distance is negative, it means that those who have managed to increase their 1999 PCI when they receive the program belong to the poor group in 1998.

Figure 3.13: Plot of $\exp\left\{\hat{\Delta}(\{F_1 = 1\}, \{F_1 = 0\})\right\}$ versus $\log(M_2^*)$
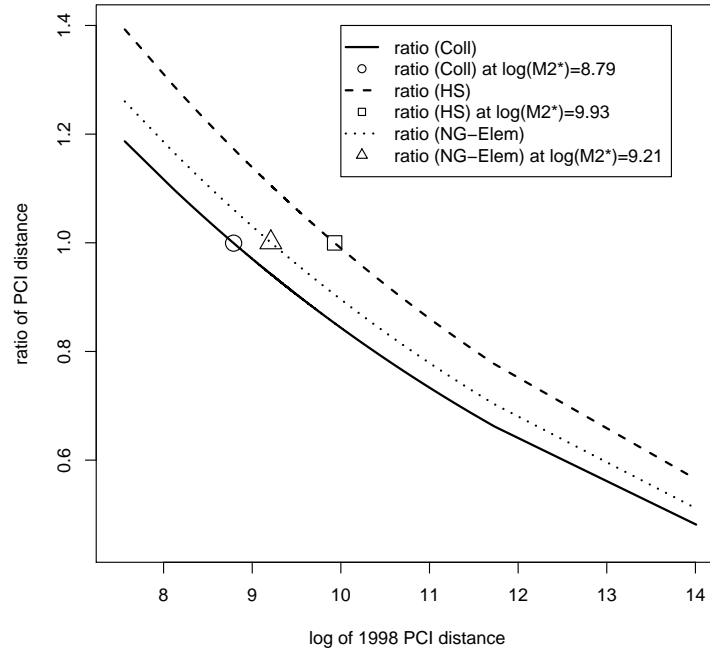
Hence, based on the model, GHFP has impact among those with 1998 income below poverty line with deficit amounting to 9,431.77 or higher. If the program was received by a family with deficit lower than that or by

a family which does not belong to the impoverished, then their 1999 PCI distance from standard will not be more improved than those who did not receive the program. We examine the phenomenon further by analysing 1998 and 1999 CARP.

- *Receiving 1998 CARP*

  The impact on 1999 income depends on 1998 income and urbanity as expressed by the following equation

  $$\hat{\Delta}(\{C_1 = 1\}, \{C_1 = 0\}) = 1.29 - 0.14 \log(M_2^*) + U[3.75 - 0.43 \log(M_2^*)]$$

  Similar to GHFP, $\hat{\Delta}(\cdot)$ here decreases on $\log(M_2^*)$ in both urban and rural areas. In particular, the decline in the ratio is steeper in urban area than in rural area as shown by Figure 3.14. Furthermore, for poorest families in urban areas, the 1999 income among those who received CARP 1998 is twice that of those who did not receive the program. In the same way, the highlighted points in Figure 3.14 show until what amount of 1998 income the edge stays with families who received 1998 CARP.

  Although less steep than urban, there is a higher threshold for 1998 income in order to maintain recipients' edge over those who did not receive the program.

- *Receiving 1999 CARP*

  The expression for the change in $\widehat{\log(Y^*)}$ is simpler if we vary 1999 CARP compared to varying 1998 CARP as shown by the following equation

  $$\hat{\Delta}(\{C_2 = 1\}, \{C_2 = 0\}) = 0.07 + 0.2U$$

  The change depends on urbanity and it can be seen that positive impact on income can be realised given that a family received 1999

Figure 3.14: Plot of $\exp\left\{\hat{\Delta}(\{C_1 = 1\}, \{C_1 = 0\})\right\}$ versus $\log(M_2^*)$

CARP and even more if it resides in urban area. That is, those in rural areas who received 1999 CARP has 1999 income approximately 1.072 times those who did not receive CARP. Among families in urban area, the 1999 CARP makes recipients' 1999 income 1.31 times those who were not beneficiaries.

We have seen that 1998 CARP effects positively on recipients on a given neighbourhood of 1998 income wherein threshold to maintain edge of those recipients of the program is higher for rural areas higher than in urban areas. Families in urban areas have steeper ratios of 1999 income with respect to 1998 income compared to rural areas. CARP 1999, on the other hand, helps increase 1999 income with increase more visible in urban

areas.

Now, let us proceed on analysing how does scholarship program augment income of families.

- *Receiving 1998 scholarship for tertiary education*

  Almost similar to 1998 CARP and 1998 GHFP, the change in 1999 income depends on previous income, but here its the 1997 one

  $$\hat{\Delta}(\{S_1 = 1\}, \{S_1 = 0\}) \quad = \quad -3.53 + 0.38 \log(M_1^*)$$

  The expression implies that the change in 1999 income increases on 1997 income. This can be seen clearly in Figure 3.15. Moreover, the highlighted point on the trend line shows at what amount of 1997 income recipients of scholarship program in 1998 start to have an edge over those who did not receive the program. Since we focused on poor families in 1997, we expect these families who managed to increase their income to be poor. Among the most well-off among the poor in 1997, those who have received the scholarship program in 1998 can have 1999 PCI distance standard $e^{-3.53+0.38 \max\{\log(M_1^*)\}} = 1.16$ times those who did not receive the program

- *Receiving 1999 scholarship for tertiary education*

  Finally, scholarship 1999 affects the 1999 income without dependence on other covariates, i.e.

  $$\hat{\Delta}(\{S_2 = 1\}, \{S_2 = 0\}) \quad = \quad 0.09$$

  In other words, holding other explanatory variables constant, the 1999 scholarship enables recipients to have 1999 income 1.09 times that of those who were not recipients.

After having interpreted the effect of intervention programs, we have found some differences in their effects in different levels of other variables

Figure 3.15: Plot of $\exp\left\{\hat{\Delta}(\{S_1 = 1\}, \{S_1 = 0\})\right\}$ versus $\log(M_1^*)$

particularly 1997 and 1998 income. Let us conclude this by analysing other remaining family characteristics, again holding others constant.

- Gender has no interaction with any other explanatory variable and conforming to the finding in Section 3.2, female headed families have higher incomes than those families with male heads. Families with female heads have 1999 income $e^{0.04} = 1.04$ times those with male heads.

- For every 10-year old difference the difference in 1999 income is

$$\hat{\Delta}(\{A = a + 10\}, \{A = a\}) \quad = \quad 0.02 + 0.01U$$

which means in rural areas, income in 1999 of families is 1.02 times those families with heads with heads 10 years younger. In addition, in urban areas, the 1999 income of families is 1.03 times those with heads 10 years younger.

- Similar to Section 3.2, household condition did not interact with other explanatory variable, and in addition, the effect of $(H_3 = 1, H_2 = 0)$ is not significant based on the Wald test. To compare those with poor household condition with moderate household condition, we have

$$\hat{\Delta}(\{H_2 = 0, H_3 = 0\}, \{H_2 = 1, H_3 = 0\}) \quad = \quad 0.04$$

which means in the end of the period, those with poor household condition have 1999 income 1.04 times those families in moderate household condition.

Although we have extracted a relatively more complicated model in this section compared to the model in the logistic regression section, the model proves to be useful in comparing groups of families and inferring about their incomes. We have found out that programs indeed helps in increasing income of the families with the help of other explanatory variables such as income in 1997 and 1998. Furthermore, we have found some unconventional finding such as families in poor condition having higher income in 1999 compared to families living in moderate household condition. Let us summarize these findings in the next section.

3.4   Summary of findings

Improvement of families, in general, is positively affected by 1998 GHFP, 1998 and 1999 CARP and 1998 and 1999 scholarship. 1998 GHFP helps improve households whose heads reached at most high school while the effect of 1998 CARP differs depending on 1999 CARP. The effects of 1998

and 1999 CARP generally depends on age of the family head except when analysing the effect of 1999 CARP compared to not receiving CARP in both years. 1998 CARP must be targeted on optimal family head age since for older family heads, the effect decays. On the other hand 1998 scholarship effect varies depending on gender of the family head where female-headed families have the edge while 1999 scholarship effect varies depending on location of the family and educational attainment of the family head. Those in urban areas have the edge particularly those with heads who reached at most elementary.

In terms of family income, intervention programs help increase 1999 income. GHFP, CARP and scholarship in 1998 affect 1999 income depending on previous incomes. 1998 GHFP and CARP help those with lower income families in 1998. Moreover, the effects of 1998 and 1999 CARP depends on urbanity with the 1999 one's higher when a family resides in urban area. The effect on income in 1999 of scholarship in 1998 increases on 1997 income while scholarship in 1999 affects 1999 income constantly.

These are just some of the aspects being implied the model and we just focused on those of essence. Now that we have gathered what the models tell us, let us conclude these in the next chapter.

Chapter 4

Conclusions

There is a paradox in poverty in the Philippines such that the incidences go lower while the magnitudes get higher. While its measurement is one challenge, its alleviation has been one of the primary targets of different administrations in the country at the national and local level. In aid of this endeavor, this study focused on assessing which programs had been effective in alleviating poverty given income based poverty, specific intervention programs and other family characteristics.

Aside from income, we have defined and used the term improvement in the preceding chapters in order to describe how intervention programs and aspects in the household affect their welfare. We have employed descriptive methods as well as logistic and regression modeling to characterise effects particularly of intervention programs. It must be noted, however, that the procedures done were based on a limited set of available predictors and response variable (income and poverty status) measured in different kinds of surveys.

Our first model, logistic regression, led us to believe that, generally, improvement of families is influenced positively by intervention programs 1998 GHFP, 1998 and 1999 CARP and 1998 and 1999 scholarship. Those families headed by persons who reached at most high school that acquired their house / lot through 1998 GHFP were able to improve their welfare.

1998 CARP, on the other hand affects improvement depending whether 1999 CARP was also availed.

Age of the family head differs how 1998 and 1999 CARP improves family welfare save when analysing the effect of 1999 CARP compared to not receiving CARP in both years. Timing is of essence in availing CARP especially in 1998 since if it is not optimally targeted in terms of age, the positive effect decays for older family heads. This explains that, since CARP is about availing agricultural land, those who avail it must have someone who can maximise the land.

On the other hand, gender of the family head varies the 1998 scholarship effect wherein female-headed families have the edge. The effect of 1999 scholarship is different between urban and rural families and varying educational attainment of the family head. Those in urban areas have the edge particularly those with heads who reached at most elementary. This is given that most universities and colleges are located in cities or town centres.

We also arrived at other relevant findings on improvement. One of them is that female-headed families have higher chance of moving out of poverty compared to male-headed ones. In addition, urban families have higher chance of improving compared to those residing in rural areas. In terms of age, families with older heads have higher chance of moving out of poverty compared to those with younger ones. Moreover, the higher the educational attainment, the higher the chance that a family can move out of poverty. Finally, those in poor household condition can improve more than those in moderate household condition.

We have answered questions relating to how do intervention programs as well as family characteristics affect improvement in welfare of families. In terms of increasing family income, we used linear regression to determine how intervention programs and aspects in the family help increase 1999 income.

GHFP, CARP and scholarship in 1998 affects 1999 income depending

on previous incomes. Lower income families can be assisted by 1998 GHFP and CARP. Furthermore, the change in 1999 income brought by 1998 and 1999 CARP is different in rural and urban. That is, 1999 income increases more if a family resides in urban area. Also, for increasing 1997 income, 1998 scholarship increases 1999 income while 1999 scholarship affects 1999 positively but constantly.

Similar to logistic regression, another relevant result in linear regression is that female headed families have higher incomes than those families with male heads. Also, older heads have families with higher income and even higher when residing in urban area. Finally, those families living in poor household condition have higher incomes in 1999 compared to those living in moderate condition.

Our findings proved that utilising statistical model building, specifically those with interactions, is useful in monitoring the impacts of intervention programs and identifying characteristics of families that benefit the most. From the models acquired, it appears that the government programs really help improve families and augment their incomes particularly if properly targeted. Although these programs are still existing in the country and information on access to these programs are available, there is insufficient longitudinal data to assess the latest impact of these programs. Hence, it is recommended that enhanced data for monitoring be made available.

It is mentioned that one crucial factor, that must have affected the models, is the difference in income measurement between the two specific surveys: FIES (1997) and APIS (1998 and 1999). Although it is a gargantuan task, mitigating the disparities between these two can improve the logistic and regression models. This is one of the reasons why we acknowledged the multidimensional nature of poverty as well. This means further work can be done in terms of improving poverty measurement particularly when appropriate data is available. In this way, assessment of impact of programs, particularly its different effects on varying aspects in

the family, can be made richer.

Bibliography

[1] AGRESTI, ALAN. *An Introduction to Categorical Data Analysis*, 2nd ed. John Wiley & Sons, Heboken, N.J., 2007.

[2] ASIAN DEVELOPMENT BANK. *Key Indicators for Asia and the Pacific 2010*, 41st ed. Asian Development Bank, Manila, Philippines, 2010.

[3] ASSELIN, LOUIS-MARIE. *Analysis of Multidimensional Poverty: Theory and Case Studies*. Springer and International Development Research Center, Ottawa, Canada, 2009.

[4] BANCOLITA, JOEL, AND ALVARADO, MARIA NORIAN. Developing Composite Indices Using CBMS. *Proceedings of the 2006 CBMS Network Conference* (November 2006).

[5] BARRIOS, ERNIEL B. AND GINALYN B. SURATOS. Regional Clustering of Poverty in the Philippines. *UPSS Working Papers*, 2008-06 (November 2008).

[6] CAPONES, ERLINDA M. Institutionalizing the Community-Based Monitoring System (CBMS) in the Philippines. *6th PEP Research Network General Meeting Proceedings* (June 2007).

[7] DUCANES, GEOFFREY M., AND BALISACAN, ARSENIO M. Multidimensional Poverty in the Philippines: Trend, Patterns, and Determinants. *International Conference on The Many Dimensions of Poverty* (August 2005).

[8] ERICTA, CARMELITA, AND FABIAN, EMMA. A Documentation of the Philippines' Family Income and Expenditure Survey. *PIDS Discussion Paper Series*, 2009-18 (July 2009).

[9] ERICTA, CARMELITA, AND LUIS, JEREMIAS. Impact of Agrarian Reform on Poverty. *PIDS Discussion Paper Series*, 2009-20 (July 2009).

[10] HOSMER, DAVID W., AND LEMESHOW, S. *Applied Logistic Regression*, 2nd ed. John Wiley & Sons, Canada, 2000.

[11] JOHNSON, RICHARD A., AND DEAN W. WICHERN. *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, N.J., 2002.

[12] MINA, C. D. Unearthing Poverty with MARS: Application of Multivariate Adaptive Regression Splines (MARS) in Identifying Household Poverty Correlates of Poverty in the Philippines. Master's thesis, School of Statistics, University of the Philippines, Philippines, 2008.

[13] NATIONAL STATISTICAL COORDINATION BOARD. Local Estimation of Poverty in the Philippines, November 2005.

[14] NATIONAL STATISTICAL COORDINATION BOARD. Official Poverty Statistics, April 30, 2010. http://www.nscb.gov.ph.

[15] R DEVELOPMENT CORE TEAM. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. ISBN 3-900051-07-0.

[16] REYES, CELIA AND EVAN DUE. *Fighting Poverty with Facts: Community Based Monitoring Systems*. International Development Research Center, Ottawa, Canada, 2009.

[17] REYES, CELIA M. Impact of Agrarian Reform on Poverty. *PIDS Discussion Paper Series*, 2002-02 (January 2002).

[18] REYES, CELIA M. A Study on Chronic and Transient Poverty. *MIMAP Research Paper*, No. 58 (2003).

[19] REYES, MELISSA LOPEZ. Construction of Unidimensional and Multidimensional Poverty Indices Using the FIES and APIS Data. *9th National Convention on Statistics (NCS) Proceedings* (2004).

[20] SCHELZIG, KARIN. *Poverty in the Philippines: Income, Assets, and Access.* Asian Development Bank, Manila, Philippines, 2010.

[21] SEN, AMARTYA. *Social Exclusion: Concept, Application, and Scrutiny.* Asian Development Bank, Manila, Philippines, 2000.

[22] SHEATHER, SIMON J. *A Modern Approach to Regression with R.* Springer Science+Business Media, LLC, New York, USA, 2009.

Appendix A

Acronyms and Definitions

Term	Description
ADB	Asian Development
AIC	Akaike Information Criterion
APIS	Annual Poverty Indicators Survey
ARMM	Autonomous Region in Muslim Mindanao
ASE	Asymptotic Standard Error
CAR	Cordillera Administrative Region
CARP	Comprehensive Agrarian Reform Program
CBMS	Community-based Monitoring System
CPI	Consumer Price Index or Composite Poverty Index
DILG	Department of Interior and Local Government
FGT	Foster-Greer-Thorbecke
FIES	Family Income and Expenditure Survey
GHFP	Government housing or other housing financing program
GLM	Generalised Linear Model
KALAHI	Kapit-bisig Laban sa Kahirapan
MCA	Multiple Correspondence Analysis
MDG	Millennium Development Goal

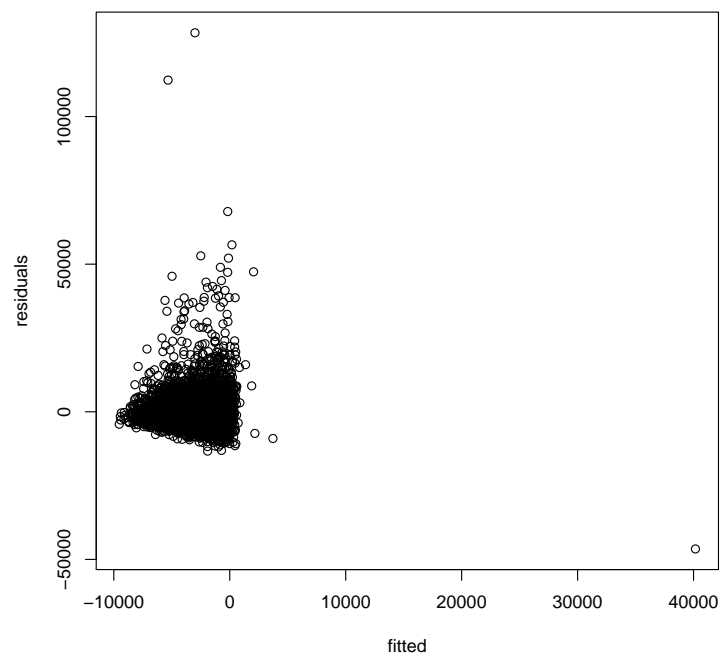| Term	Description
MLE	Maximum Likelihood Estimator / Estimation
MTPDP	Medium-term Philippine Development Plan
NAPC	National Anti-poverty Commission
NSCB	National Statistical Coordination Board
NSO	National Statistics Office
PCA	Principal Components Analysis
PCI	Per Capita Income
Per capita	Per head
SRA	Social Reform Agenda
Threshold, poverty	amount needed by an individual to meet food and non-food needs
Threshold, food	amount needed by an individual to meet food needs

Appendix B

Tables and Figures

Figure B.1: Residuals versus Fitted, $Y^* = \beta_0 + \beta_1 M_1^* + \beta_2 M_2^* + \epsilon$ (1997 poor families)

X_j	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	Wald Z	$P(\mathcal{Z} > Z)$	
(Intercept)	-2.05	0.236	-8.679	0.000	***
G	0.35	0.116	3.042	0.002	**
U	0.30	0.069	4.401	0.000	***
A	0.02	0.003	8.288	0.000	***
C_1	-0.16	0.295	-0.554	0.579	
C_2	0.37	0.246	1.510	0.131	
F_1	-0.02	0.224	-0.073	0.942	
F_2	0.11	0.224	0.513	0.608	
S_1	0.34	0.23	1.492	0.136	
S_2	0.41	0.297	1.371	0.170	
H_2	-0.48	0.211	-2.289	0.022	*
H_3	-0.06	0.205	-0.287	0.774	
E_2	-0.30	0.082	-3.631	0.000	***
E_3	-0.53	0.091	-5.848	0.000	***

Signif. codes: 0.000 *** 0.001 ** 0.01 * 0.05 . 0.1 1

AIC: 5410.1   Deviance: 5382.1   Df: 5287

Table B.1: Estimates for the initial Logistic Regression Model using revised categories

X_j	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	Wald Z	$P(\mathcal{Z} > Z)$	
(Intercept)	-2.14	0.244	-8.756	0	***
G	0.32	0.118	2.749	0.006	**
U	0.46	0.117	3.964	0	***
A	0.02	0.003	8.383	0	***
C_1	1.27	1.16	1.091	0.275	
C_2	0.24	0.283	0.845	0.398	
F_1	-0.65	0.428	-1.519	0.129	
F_2	0.09	0.226	0.390	0.696	
H_2	-0.50	0.211	-2.353	0.019	*
H_3	-0.06	0.206	-0.313	0.755	
E_2	-0.21	0.118	-1.760	0.078	.
E_3	-0.40	0.121	-3.292	0.001	***
S_1	0.19	0.257	0.748	0.455	
S_2	-0.32	0.642	-0.497	0.619	
$E_2 \times S_2$	-0.83	0.79	-1.056	0.291	
$E_3 \times S_2$	1.90	0.868	2.188	0.029	*
$U \times S_2$	1.02	0.696	1.465	0.143	
$G \times S_1$	1.16	0.768	1.505	0.132	
$E_2 \times F_1$	1.07	0.535	1.992	0.046	*
$E_3 \times F_1$	0.84	0.595	1.419	0.156	
$A \times C_1$	-0.04	0.024	-1.548	0.122	
$E_2 \times U$	-0.19	0.165	-1.183	0.237	
$E_3 \times U$	-0.35	0.174	-2.007	0.045	*
$C_1 \times C_2$	0.93	0.643	1.450	0.147	

Signif. codes: 0.000 *** 0.001 ** 0.01 * 0.05 . 0.1   1
AIC: 5404.6   Deviance: 5356.6   Df: 5277

Table B.2: Estimates for the Logistic Regression Model derived after forward Selection with revised categories
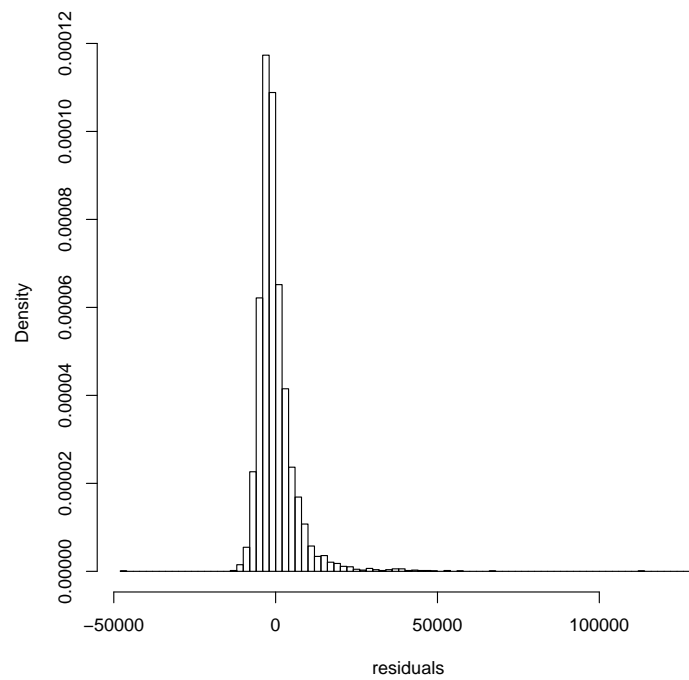
Figure B.2: Histogram of residuals, $Y^* = \beta_0 + \beta_1 M_1^* + \beta_2 M_2^* + \epsilon$ (1997 poor families)

X_j	$\hat{\beta}_j$	$SE(\hat{\beta}_j)$	Wald Z	$P(\mathcal{Z} > Z)$	
(Intercept)	0.67	0.407	1.639	0.101	
$\log(M_1^*)$	0.46	0.04	11.518	0.000	***
$\log(M_2^*)$	0.45	0.029	15.579	0.000	***
G	0.04	0.019	2.109	0.035	*
U	-0.66	0.521	-1.270	0.204	
A	0.00	0.001	4.258	0.000	***
C_1	1.30	0.895	1.449	0.147	
C_2	0.07	0.043	1.530	0.126	
F_1	1.23	0.855	1.441	0.150	
F_2	0.00	0.033	0.124	0.901	
H_2	-0.04	0.031	-1.370	0.171	
H_3	0.00	0.031	-0.095	0.924	
E_2	-0.41	0.308	-1.325	0.185	
E_3	0.84	0.303	2.771	0.006	**
S_1	-3.40	1.923	-1.770	0.077	.
S_2	0.09	0.05	1.825	0.068	.
$\log(M_2^*) \times E_2$	0.04	0.033	1.368	0.171	
$\log(M_2^*) \times E_3$	-0.09	0.032	-2.786	0.005	**
$\log(M_2^*) \times C_1$	-0.14	0.094	-1.504	0.133	
$U \times E_2$	-0.07	0.025	-2.689	0.007	**
$U \times E_3$	-0.05	0.026	-1.804	0.071	.
$\log(M_1^*) \times S_1$	0.48	0.218	2.214	0.027	*
$U \times C_1$	3.78	1.716	2.204	0.028	*
$U \times C_2$	0.20	0.096	2.108	0.035	*
$U \times A$	0.00	0.001	1.788	0.074	.
$F_1 \times E_2$	0.16	0.077	2.041	0.041	*
$F_1 \times E_3$	0.06	0.083	0.723	0.470	
$\log(M_2^*) \times S_1$	-0.12	0.088	-1.327	0.185	
$\log(M_2^*) \times F_1$	-0.14	0.091	-1.519	0.129	
$\log(M_1^*) \times U$	0.12	0.059	1.974	0.048	*
$\log(M_2^*) \times U$	-0.05	0.029	-1.656	0.098	.
$\log(M_2^*) \times U \times C_1$	-0.43	0.184	-2.331	0.020	*

Signif. codes: 0.000 *** 0.001 ** 0.01 * 0.05 . 0.1 1

AIC: 4072.1   Deviance: 660.83   Df: 5269

Table B.3: Estimates for the Linear Regression Model after forward selection, 1997 poor families