Exploring Changes in Poverty in Zimbabwe between 1995 and 2001 using Parametric and Nonparametric Quantile Regression Decomposition Techniques

Katherine Amelia Eriksson

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Agricultural and Applied Economics

Jeffrey Alwang—Co-Chair
Christopher Parmeter—Co-Chair
Daniel Taylor

August 21, 2007
Blacksburg, Virginia

Keywords: nonparametric quantile regression, decomposition, Zimbabwe, poverty

Exploring Changes in Poverty in Zimbabwe between 1995 and 2001 using Parametric and Nonparametric Quantile Regression Decomposition Techniques

Katherine Amelia Eriksson

Abstract

This paper applies and extends Machado and Mata's parametric quantile decomposition method and a similar nonparametric technique to explore changes in welfare in Zimbabwe between 1995 and 2001. These methods allow us to construct a counterfactual distribution in order to decompose the shift into the part due to changes in endowments and that due to changes in returns. We examine two subsets of a nationally representative dataset and find that endowments had a positive effect but that returns account for more of the difference. In communal farming areas, the effect of returns was positive, while in urban Harare it was negative.

Acknowledgements

Thank you to Jeff Alwang, the best advisor I could have had, for guiding me towards this topic and for his infinite patience this year. Thanks also to Chris Parmeter for help with the methodology and for answering all my silly programming questions.

Thank you to Dan Taylor and the AAEC department for supporting me all year! I could not have survived the year without help from all the first year PhD students as well as my friends in other departments. Thanks finally to my parents for the support and encouragement throughout my entire academic career.

Table of Contents

Acknowledgements iii
Table of Contents iv
List of Figures v
List of Tables vi
1. Introduction 1
2. Background and Data 3
 2.1 Background 3
 2.2 Data 6
3. Changes in Endowments and Returns between 1995 and 2001 11
 3.1 Changes in endowments 11
 3.2 Changes in returns to endowments 14
4. Methodology 15
 4.1 Parametric Methodology 16
 4.2 Nonparametric Methodology 19
 4.3 Comparison to Alternative Methodologies 22
5. Results and Discussion 24
 5.1 Results of Parametric versus Nonparametric Model 24
 5.2 Coefficients 32
 5.3 Discussion 33
  5.3.1 Mashonaland 34
  5.3.2 Harare 35
6. Discussion and Conclusion 36
References 37
Appendix A. Coefficient Plots 39
Appendix B. Counterfactual CDF plots 43
Vita 49

List of Figures

Figure 2.1. Harare CDF 10
Figure 2.2. Mashonaland Communal CDF: 10
Figure 5.1 Results 24
Figure 5.2. Endowments, Returns and Residuals 25
Figure 5.5. 95% Confidence Intervals for Counterfactuals 28
Figure A.1 Mashonaland 39
Figure A.2 Harare 41
Figure B.1. Mashonaland Endowments 43
Figure B.2 Harare Endowments 45
Figure B.3. Mashonaland Coefficients 46
Figure B.4. Harare Coefficients 47

List of Tables

Table 2.1. Poverty Rates ... 9
Table 3.1. Highest Education Level Attained by any Household Member ... 11
Table 3.2. Employment Sector of the Household Head ... 12
Table 3.3. Percentage of Normal Rainfall in Mashonaland ... 13
Table 3.4. Proportion of Female Headed Households in Harare ... 13
Table 3.5. Proportion of Households without Children under 5 in Mashonaland ... 14
Table 5.1. Mashonaland Decomposition ... 26
Table 5.2. Harare Decomposition ... 26
Table 5.3. Mashonaland Endowments ... 29
Table 5.4. Harare Endowments ... 29
Table 5.4. Mashonaland Coefficients ... 33
Table 5.5 Harare Coefficients ... 33

1. Introduction

Since the late 1990's, abundant anecdotal evidence has implied that the Zimbabwean economy has been struggling, but no nationally representative data have been available to quantify changes in the welfare of households. This paper uses nationally representative data sets from 1995 and 2001 to explore changes in welfare between these two years and to examine the factors contributing to such changes. I examine the countervailing effects of better rainfall in 2001 compared with 1995 versus the declining economy which has experienced negative growth rates since 1999. My methodology enables us to identify the part of the shift in welfare due to variables I consider to represent both effects, namely rainfall and returns to communal farming to represent the improved rainfall and returns to education and formal sector employment to represent the declining economy.

The well known Oaxaca-Blinder mean decomposition decomposes the mean difference between two groups into the effects of endowments and the effects of returns. This counterfactual methodology allows one to identify these effects in the absence of panel data. We use quantile regression methods because I am interested in parts of the distribution in addition to the mean and want to allow the returns to the covariates to vary at different points of the distribution. In particular, changes in the lower part of the distribution tell us how those who were poor in 1995 compare to those who were poor in 2001. Similarly for the upper part of the distribution, I want to ask whether those well off in 2001 were better off than those well off in 1995. I apply Machado and Mata's (2005) quantile regression decomposition method and a nonparametric technique similar to that of Melly (2006). Machado and Mata's method has been widely applied in labor economics and other fields. In development, Nguyen et al (2007) apply it to describe rural/urban inequality in Vietnam before and after the *doi moi* reforms.

This paper makes two main econometric contributions. First, I extend Machado and Mata's method to account for the effects of specific coefficients. Second, I develop a nonparametric corollary of the parametric Machado and Mata method to account for specific endowments.

I consider two geographical regions of Zimbabwe—Matabeleland and Harare—which had vastly different changes over these years. Between the two survey years, overall education levels increased as a result of more Zimbabweans born after independence in 1980 completing their schooling in the post-independence school system. Unemployment levels in Harare rose and formal sector employment dropped following the closure of firms and significant job-cuts since 1999. Real wages in Harare in 2001 were also lower than they were in 1995. I seek to quantify the effects of these changes in the economy.

I find that poverty decreased in Mashonaland and that welfare unequivocally increased across the entire distribution. In Harare, poverty increased with lower welfare in 2001 in the bottom three quartiles of the distribution. I seek to explain these shifts in real per capita consumption, my measure of welfare, between the survey years. Both models indicate that returns explain a larger part of the shifts in welfare than endowments. This returns effect is positive in Mashonaland and the lower part of the Harare distribution but negative in the top quartile of the Harare distribution. By identifying the effects of returns and endowments, I can determine the effects of exogenous changes in the economy (changes in the exogenous covariates) and effects of changing prices of these exogenous variables, or returns to the covariates.

This paper is closely related to recent work on poverty in Africa utilizing panel data or repeated cross-sectional studies. Examples of panel studies in Africa are Dercon's studies in Ethiopia and those by Okidi in Uganda. Dercon (1998) measures the changes in poverty rates by village in Ethiopia between 1989 and 1994 and finds that rainfall can partly explain the shifts in welfare. McCulloch et al (2001) use a national cross-sectional survey in Zambia to analyze the effects of the Structural Adjustment Program in the early 1990s. This study is a unique contribution to this strand of literature in that it uses a larger data set than other studies and it applies methods allowing the consideration of the entire distribution along with a more detailed analysis of the reasons behind changes in the economy.

The structure of the paper is as follows. Section 2 provides background on Zimbabwe and describes the data and the variables used in the model. Section 3 presents changes endowments and returns between the two years. Section 4 explains the decomposition methodology in detail. Section 5 presents and discusses the decomposition results, and section 6 concludes.

2. Background and Data

2.1 Background

Zimbabwe is divided into ten provinces, two of which are the urban provinces of Bulawayo and Harare. The population is approximately 65 percent rural and 35 percent urban. Zimbabwe's agricultural land is broken into four main land-use sectors: communal farms, large scale commercial farms, small scale commercial farms, and resettlement areas. At independence in 1980, almost all large scale commercial farms were owned by white farmers, while Africans were largely concentrated in the less productive communal areas. After independence, the government purchased part of the large scale commercial farming land and instituted a resettlement program which is still in progress.

This paper focuses on the communal farming areas in Mashonaland East, Mashonaland West, and Mashonaland Central (the three provinces are hereafter referred to as just Mashonaland) and Harare urban areas. These two regions provide different insights into the Zimbabwean economy and the determinants of welfare between the two survey years. I am interested in the effects of the declining economy on farmers, who are somewhat insulated from the market, and on wage employment workers, who are fully connected to the economy. In communal areas, self employment in agriculture provides the main source of income, whereas in Harare most income is derived from wage employment and informal sources. The two regions also differ in their rates of extreme poverty, with 62% and 40% of households being extremely poor in Mashonaland in the two years but only 9% and 19% in Harare (see Table 2.1).

The majority of communal farms, whose agricultural production depends heavily on rainfall, are concentrated in low-rainfall zones.[1] Maize is the predominant crop grown in the main growing season between October/November and April/May. Other crops include cotton, groundnuts, sunflower, and paprika. In 2001, due to the late payment for maize in the previous season, many communal farmers diversified away from maize

[1] Zimbabwe has five agro-ecological zones defined by the amount of normal annual rainfall and which are suitable for different types of farming (e.g. crops or livestock grazing). Most commercial farms are found in the zones with more rainfall and also benefit from irrigation. See Vincent and Thomas (1960) for details.

into these secondary crops. Because of the rain-fed nature of crops in communal areas, I expect that rainfall plays a major role in determining welfare in these areas.

As noted, poverty increased in Harare between 1995 and 2001; this increase can be explained by a combination of urban-specific factors. During this period, urban areas experienced job-cuts, firm closures, declining real wages, and strikes. The government instituted a 2.5 percent increase in VAT and a five percent 'development levy' on wage earners in March 1998, resulting in a strike supported by 90 percent of formal sector workers who were already paying 42 percent of their income in taxes and had seen their real wages decline by 63 percent since the introduction of the structural adjustment program in 1991[2]. A 2001 survey by the Zimbabwe Chamber of Commerce found that in 2000, 400 firms closed and a further 350 down-sized, resulting in a loss of over 10,000 formal sector jobs. Inflation rates began climbing from 32 percent in 1998 to over 100 percent by the end of 2001. All of these factors contributed to political dissatisfaction in the cities. From the late 1980's when it merged with its main rival party ZAPU, until 1999, ZANU-PF did not have significant political opposition. However, the Movement for Democratic Change (MDC) was formed in 1999 by a coalition of trade unionists, commercial farmers, intellectuals, and the urban middle class. The 2000 parliamentary election was won by ZANU-PF with 62 seats to the MDC's 57 but was marked by widespread allegations of political violence. One would expect all of these factors to contribute to lower urban welfare in 2001.

---

[2] In common with most African countries, Zimbabwe instituted a structural adjustment program in 1991, called ESAP, which resulted in lower formal employment and lower wages. The expected post-liberalization increase in overall welfare had not happened as of 1995.

2.2 Data

The Zimbabwean Central Statistical Office (CSO) carries out a nationally representative survey of income and consumption expenditure (ICES) approximately every five years. I use the two most recent surveys, collected in July 1995 - June 1996 and January - December 2001. My final data-set has 17,527 households in 1995 and 19,221 households in 2001. This large sample size, rare in Africa, allows detailed analysis of subsets of interest. The survey over-sampled urban, small scale commercial farm, and resettlement areas in both survey years so survey weights are used in all estimation. In addition, we use official price data collected monthly by the Central Statistical office to construct our cost of living deflator.

For a number of methodological and theoretical reasons, I only analyze two subsets of the data, Mashonaland communal areas and Harare, during the months of July to October. This decision was the result of an extensive testing procedure in which I discovered heterogeneity between time periods and geographical regions which was not sufficiently accounted for by month and province dummies. Specifically, I consider only these areas and this time period in order to:

a) consider the part of the year in which consumption is fully determined by conditions in the previous growing season. In November and December, rainfall is used by households to start forming expectations about the next growing season. Similarly, in the first 4 months of the year, consumption is determined by a combination of prior year rainfall, which determines the stocks remaining for consumption, and current year rainfall, which determines expectations about the crop and therefore borrowing and smoothing behavior. I would expect the relationship between rainfall and consumption to be different in the May-October period than the other parts of the year. Finally, I do not use May and June because in the 1995/6 survey May and June were collected in 1996 but July-Oct were collected in 1995.[3]

b) reduce the sample size to a manageable size for the nonparametric analysis. Computing time for nonparametric analysis increases exponentially with the sample size[4], so a sample size of over 22,000 is not feasible.

c) utilize the survey weights while taking into account that they are inaccurate over the year due to uneven sampling from month to month.. For example, in 2001 urban areas were heavily sampled in the first half of the year but were under-sampled in the second half of the year.

The survey collects detailed information about household characteristics, consumption, assets, and agricultural activities. Expenditures on more than 200 food and 300 non-food items are collected. Households are visited by an enumerator on a weekly basis over a month and they keep a diary of daily expenditures. In addition to market purchases, the survey collects estimated values of own-account production, gifts, transfers, and in-kind consumption. Due to the one month recall period of the survey and the seasonality of agricultural incomes, I use real per capita expenditure [5], measured in Harare January 2001 dollars, as my measure of welfare.

Household consumption was constructed as the sum of expenditures on food, durable and nondurable goods, and expenditures on services and school fees[6]. For

---

[3] As noted above, the survey in 1995/6 was collected from July 1995-June 1996 but in 2001 the survey switched to the calendar year, January – December 2001.

[4] For example, doubling the sample size multiplies computation time by e^2 or approximately 7.3.

[5] See Deaton 1997

[6] While previous work on the earlier ICES data sets also included housing expenditure (Alwang et al 2001b), we made the decision not to include this because of the lack of explanatory power of the imputed rent values in the survey.

durable goods, I divide the average price of the good by the approximate life-span in months. The CSO collects monthly data on prices in urban and rural locations throughout Zimbabwe, which is used to construct provincial-level cost of living price deflators. Per capita consumption is then multiplied by the deflator so that all consumption is measured relative to Harare January 2001 prices. Unique deflators were constructed for each month and province using the price of a basket of food meeting minimum nutritional requirements. Finally, the Zimbabwean Meteorological Service collects detailed rainfall data in each area sampled which I use to construct my rainfall variable, the percentage of normal rainfall in the previous growing season.

I model the log of household real per capita consumption as a function of the log of the household's size, the log of the age of the household head, the highest education level attained by any member of the household, the sector in which the household head is employed, and rainfall (in communal areas only). I do not employ month dummies because I am looking at only four months in which any month to month variation in average consumption is likely caused by uneven sampling. Extensive model misspecification tests were carried out on a basic linear regression model with these variables.

My 1995 data was previously used together with the 1991-92 survey in Alwang et al (2001, 2003). They use the kernel density re-weighting method of Dinardo et. al (1996) to determine why welfare was lower in 1995 than in 1991. They find that, in rural areas, only a small part of the decline in welfare can be explained by droughts[7] of 1991/2 and 1994/5 and that most of the decline is explained by general equilibrium effects of a declining economy after the structural adjustment program put in place in the early 1990's. In urban areas, higher levels of human capital partly offset the general equilibrium effects.

Table 2.1 shows the percentage of households and individuals below the food poverty line and overall poverty line in July-October by year in the whole country as well as in each of the subsets I am analyzing[8]. Poverty decreased in the communal areas, large scale commercial farms, and resettlement areas. Poverty increased in Harare and Bulawayo. Poverty rates do not describe the whole distribution, however, so below I present the conditional density functions (CDFs) for Mashonaland communal areas and Harare. They show that while welfare increased across all quantiles in the Mashonaland communal areas, it decreased in the lower quartiles of the distribution in Harare and increased in the quartile.

Table 2.1. Poverty Rates[9]

	All	Mashonaland Communal	Matabeleland Communal	Harare Urban	LSC	Manicaland Communal	Resettlement Areas
1995							
Lower	.424	.626	.695	.092	.322	.640	.683
	(.022)	(.037)	(.036)	(.023)	(.034)	(.031)	(.038)
Upper	.626	.805	.845	.256	.568	.868	.889
	(.022)	(.030)	(.032)	(.034)	(.037)	(.023)	(.022)
2001							
Lower	.350	.400	.524	.191	.221	.432	.584
	(.019)	(.029)	(.031)	(.058)	(.038)	(.046)	(.061)
Upper	.576	.654	.755	.355	.454	.697	.781
	(.022)	(.033)	(.025)	(.080)	(.053)	(.038)	(.071)

[8] The lower food poverty line was calculated to be Z$630 (US$11.45 at the official rate) per capita per month in Harare January 2001 dollars. This represents the bare minimum a household must earn per person in order to provide the basic food requirements. The upper poverty line then is approximately 1.66 times this lower line, calculated from the average non-food expenditures of those spending exactly the food poverty line on food. That is, those spending Z$630 per capita on food spend on average Z$420 on non-food items as well, yielding an upper poverty line of Z$1050.

[9] The estimates and standard errors are calculated accounting for the survey weights and clustered sample design.

Figure 2.1. Harare CDF

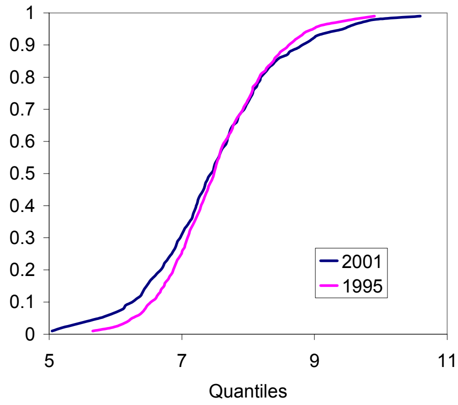
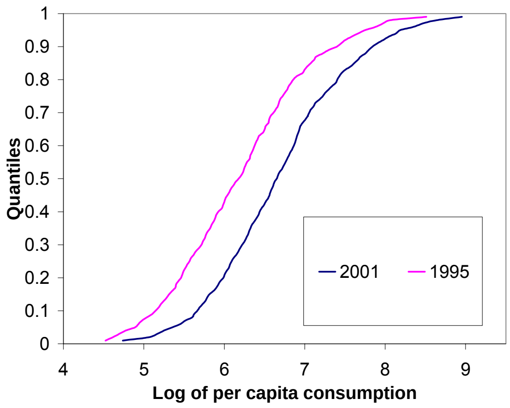


Figure 2.2. Mashonaland Communal CDF



I use a Kolmogorov-Smirnov test to test the statistical significance of the shifts in welfare. In Mashonaland, I find a p-value of .000137 so I reject the null hypothesis that the two CDFs come from the same distribution. In Harare, p=.7054 so I fail to reject the null hypothesis. Despite the insignificant change in the overall CDF, note that the change in poverty *is* significant (Table 2.1). In section 5, I find that the small overall shift is explained by opposite effects of endowments and returns.

3. Changes in Endowments and Returns between 1995 and 2001

3.1 Changes in endowments

The shifts in welfare I found are accounted for by changes in endowments and the returns to these endowments. Therefore, before considering the decomposition, I inspect these changes in order to determine which endowments and returns probably account for the increase in welfare. The endowments showing statistically significant shifts between 1995 and 2001 are education, employment, rainfall in Mashonaland, *infant* in Mashonaland, and *female* in Harare.

In both areas, overall levels of education increased (Table 3.1). After 1980, Zimbabwe instituted universal education and by the late 1990's had one of the highest literacy and secondary school enrollment rates in southern Africa. The increase in education levels can therefore be attributed to these new education policies which, by 2001, had yielded several cohorts educated under the new system. All else equal, I expect that the higher overall level of education raised consumption in 2001 over 1995, but, if the returns to education declined I would not necessarily see this in the data.

Table 3.1. Highest Education Level Attained by any Household Member

Education level	Mash. 1995	Mash. 2001	Harare 1995	Harare 2001
No School	.0313 (.0083)	.0318(.0068)	.0117 (.0077)	.0039 (.0025)
Some Primary	.2197 (.0224)	.1275 (.0127)	.0632 (.0144)	.0297 (.0139)
Primary	.1996 (.0181)	.1801 (.0190)	.1156 (.0101)	.0733 (.0165)
Some Secondary	.2651 (.0178)	.2634 (.0178)	.1750 (.0161)	.0863 (.0141)
10^{th} grade diploma	.2387 (.0230)	.2910 (.0197)	.3809 (.0348)	.4912 (.0571)
Some post 10^{th}	.0060 (.0049)	.0071 (.0030)	.0164 (.0057)	.0791 (.0232)
Secondary Diploma	.0374 (.0108)	.0750 (.0205)	.2067 (.0226)	.1761 (.0509)
Tertiary	.0021 (.0021)	.0240 (.0104)	.0305 (.0102)	.0605 (.0206)

*standard errors are given in parentheses

Between 1995 and 2001 there was a shift in Harare away from formal sector and government employment towards the informal sector and unemployment (Table 3.2). Unemployment doubled, government employment halved, and formal private sector employment decreased by seven percentage points. In Mashonaland, the only significant changes were an increase in government employment and slight decrease in informal and own account employment.

Table 3.2. Employment Sector of the Household Head

Employment State	Mash. 1995	Mash. 2001	Harare 1995	Harare 2001
Unemployed	.0741 (.0198)	.0688 (.0136)	.1029 (.0122)	.2223 (.0332)
Communal Farmer	.7939 (.0265)	.7648 (.0367)	0	0
Government	.0594 (.0169)	.1146 (.0307)	.1480 (.0187)	.0799 (.0093)
Parastatal & Coop	0	.0028 (.0022)	.0366 (.0076)	.0151 (.0052)
Formal	.0225 (.0078)	.0141 (.0063)	.4889 (.0442)	.4102 (.0372)
Informal & Own Account	.0501 (.0103)	.0347 (.0068)	.2166 (.0495)	.2724 (.0640)

In 1995, all parts of Mashonaland received less rain than normal in one of the worst drought years in recent Zimbabwean history (Table 3.3). In 2001, most areas received average or above average amounts of rain. I expect, therefore, that a large part of the increase in welfare can be explained by improved rain endowments. For the counterfactual analysis, I break rainfall into five classes.

Table 3.3. Percentage of Normal Rainfall in Mashonaland

Rainfall	1995	2001
<.70	.3703 (.0831)	.1369 (.0538)
Between .70 and .85	.4126 (.0914)	.0082 (.0083)
Between .85 and 1.15	.2170 (.0639)	.4766 (.0965)
Between 1.15 and 1.30	0	.1678 (.0734)
>1.30	0	.2103 (.0859)

Table 3.4 shows that, between my survey years, the percentage of female-headed households increased in Harare although this change is not statistically significant. I expect that overall welfare in Harare would have been higher in 2001 if this percentage had not changed.

Table 3.4. Proportion of Female Headed Households in Harare

	1995	2001
Female	.1737 (.0161)	.2170 (.0167)

While children over age five generally work in the household in communal areas and therefore contribute something to household income, households with children under age five have lower per capita income than those without children under age five. In my model, I use an indicator variable, *infant*, which takes the value 1 if there are no children under age five. In Mashonaland there was a slight increase in the number of households with no children, which I expect to have had a positive impact on consumption.

Table 3.5. Proportion of Households without Children under 5 in Mashonaland

	1995	2001
Infant	.4732 (.0131)	.4966 (.0134)

3.2 Changes in returns to endowments

I am also interested in how different returns in 2001 affected welfare. I therefore first consider the coefficients of the whole quantile regression process for the factors in which I are interested. My plots (Figures A.1, A.2) show the difference between the returns in 2001 and returns in 1995. In areas where the difference is significantly greater than zero, returns to that covariate were greater in 2001 than in 1995 and vice versa if the difference is less than zero.

In Mashonaland, the returns to female headed households decreased significantly. I see a similar pattern across the coefficients for different levels of education—returns increased at the lower end of the distribution but then decreased in the top three quartiles of most levels. Similarly for employment, households only experienced increased returns in the bottom 10-20% of the distribution. Returns to rainfall decreased except in the top quartile. I expect returns to rainfall to have fallen since returns were likely very high in 1995 given the extremely small amount of rainfall received.

In Harare, the difference in return to female household head was negative around the 75th quantile but otherwise remained approximately zero. I find opposite results for schooling to what I found in Mashonaland. In Harare, returns to all levels of school decreased at the lower end of the distribution but increased after that. The return to formal sector employment displays an erratic pattern, with negative changes at the bottom and top of the distribution but positive improvements in the middle. Those at the top end of the distribution in government employment received lower returns in 2001 than 1995. Finally, the returns to age decreased significantly in the upper part of the distribution. The age of the household head is sometimes considered a proxy for workforce experience, so these lower returns could signify lower compensation for experience at well paying jobs.

4. Methodology

My approach is a direct extension of the Oaxaca-Blinder mean decomposition. In the context of the standard linear regression model, the mean difference of Y between the two years can be rewritten as

$$\overline{Y_{01}} - \overline{Y_{95}} = (\overline{X_{01}}\hat{\beta}_{01} - \overline{X_{01}}\hat{\beta}_{95}) + (\overline{X_{01}}\hat{\beta}_{95} - \overline{X_{95}}\hat{\beta}_{95}) + \textit{residuals}$$

where the first bracket represents the part of the difference accounted for by the coefficients and the second represents the effect of the endowments or covariates. Because I am interested in the entire distribution, I extend the OB decomposition to consider each quantile rather than just the mean. Instead of the mean values of Y_{01} and Y_{95}, I use the θ^{th} quantile. Similarly, I estimate the predicted quantiles on the right hand side rather than the predicted mean of Y. Specifically, for θ in (0,1) I estimate

$$q_{01}(\theta) - q_{95}(\theta) = [\hat{q}_{01}(\theta) - \hat{q}_c(\theta)] + [\hat{q}_c(\theta) - \hat{q}_{95}(\theta)] + \textit{residuals}$$

by both parametric and nonparametric methods for $q_t(\theta)$, the θ^{th} quantile. I apply and extend Machado and Mata's (MM) parametric quantile decomposition method to construct the counterfactual. The nonparametric quantiles are estimated by inverting the unconditional CDF constructed using Racine and Li's conditional CDF method. The nonparametric decomposition is also extended to account for the effects of specific endowments.

I estimate both the parametric and nonparametric model because each has its advantages and drawbacks. The nonparametric model has two main advantages over the parametric model. First, it is robust to functional form misspecification, allowing for non-linearities not picked up in the parametric model. Second, Li and Racine's bandwidth selection method eliminates irrelevant covariates. The main drawback of the nonparametric model is that it does not allow identification of the effects of specific returns as the parametric model does. The models also give us slightly different results, with the parametric model leaving slightly more of the difference as unexplained residual, particularly in non-communal areas. In section 5, I discuss the fit of each model and possible reasons for the different results.

4.1 Parametric Methodology

For θ in (0,1) the conditional quantile is given by

$$Q_\theta(y \mid z) = z'\beta(\theta)$$

where y is the log of per capita consumption, z is the kx1 vector of covariates, and β(*θ*) is the kx1 vector of quantile coefficients. β (*θ*) solves

$$\min_\beta n^{-1} \sum\nolimits_{i=1}^{n} \rho_\theta (y_i - z_i'\beta)$$

where $\boldsymbol{\rho_\theta}$(u) is the check function

$$\rho_\theta(u) = \begin{cases} \theta u, & u > 0 \\ (\theta - 1)u, & u \le 0 \end{cases}$$

Given that the functional form is correctly specified, $Q_\theta(y \mid z)$ is a strongly consistent estimator of the population conditional quantiles (Bassett and Koenker 1982). As a function of θ, $Q_\theta(y \mid z)$ provides a full characterization of the conditional distribution of y.

As opposed to the conditional quantiles, my decomposition requires the unconditional quantiles, which MM obtain by integrating the conditional distribution to find the marginal distribution of consumption. Rather than using the sample quantiles of the original marginal distribution, integrating the conditional distribution gives us a marginal distribution which is consistent with the distribution of covariates, and which therefore allows us to perform counterfactual exercises.

To integrate the conditional distribution, MM apply the probability integral transformation theorem. For a uniform random variable U~U[0,1], $F^{-1}(U)$ has distribution F. Therefore for a sample $\theta_1, \theta_2, \ldots, \theta_n$ from the uniform distribution, the corresponding conditional quantiles $Q_{\theta_i}(y \mid z) = z'\beta(\theta_i)$ are a random sample from the conditional distribution. However, this sample is still conditional on the given z's. To obtain the marginal distribution I draw a random sample from the rows of the covariate matrix, Z(t). This is numerically equivalent to integrating the conditional distribution over the range of the covariates.

MM's specific procedure for simulating the marginal distribution of consumption at time t is as follows:

(1) Generate a random sample of size m from a uniform distribution, $u_1, \ldots, u_m$.

(2) For the covariates at time t and each u_i, estimate $Q_{u_i}(y \mid z,t)$ to obtain $\beta^t(u_i)$.

(3) Generate a random sample of size m with replacement from the rows of Z(t).

(4) Finally, $\{y^* = z_i^t{}'\beta^t(u_i)\}$ constitutes a random sample from the marginal distribution of y at time t.

To obtain the counterfactual marginal distribution that would prevail if the returns in 1995 were as in 2001 but the covariates were distributed as in 1995, MM estimate the coefficients at t=2001 but draw from the rows of Z(1995) to form $\{y_c^* = z_i^{95}{}'\beta^{01}(u_i)\}$. I then construct the quantile decomposition above by taking the quantiles of the marginal distributions where $q_c(\theta)$ comes from the counterfactual distribution.

In addition to the full decomposition, I also ask how much of the difference can be accounted for by both specific covariates, or endowments, and specific coefficients, or returns. For example, what would the 2001 distribution look like if the distribution of education was as it was in 1995? What would it look like if the returns to education were as in 1995?

MM extend their original simulation method to account for specific endowments as follows. Follow the steps above to create a sample of size m, $\{y_e^*\}$ from the 2001 data. For a categorical variable, repeat the following steps for each value of the variable, 1,…,J. For a continuous variable, break the range into any number of bins and repeat for each bin.

(i) Select the subset of $\{y_e^*\}$ where the covariate of interest takes value j;

(ii) Let $f_j(1995)$ be the relative frequency of class j in 1995. Resample with replacement from this subset to generate a sample of size $m^*f_j(1995)$.

This procedure yields a sample of size n which reweights the marginal distribution as if it had the relative frequencies of the other year.

I extend the MM methodology to account for one covariate receiving the return it would have received in the other year. To simulate this distribution, in step 2 I estimate the coefficients at each u_i for both years. Then I construct my coefficient vectors using the combination of returns in which I am interested. For example, if I am interested in the 2001 distribution with 1995 returns to school, I use the coefficients for the education variables from 1995 and the rest of the coefficients from the 2001 estimation. I sample from the rows of Z(2001) as before.

4.2 Nonparametric Methodology

My nonparametric technique differs from the parametric method in that I do not directly estimate the conditional quantile function. Instead, I estimate the conditional cdf, integrate the conditional cdf to obtain the unconditional cdf, and then invert the unconditional cdf to obtain the unconditional quantiles for the decomposition.

Li and Racine(2007) develop a method for estimating a conditional cdf with mixed data types that has not yet been widely applied. They propose the following kernel for an ordered discrete variable, z^{do}:

$$l(z_i^{do}, z_j^{do}, \lambda_{do}) = \begin{cases} 1, & z_i^{do} = z_j^{do} \\ \lambda_{do}^{|z_i^{do} - z_j^{do}|}, & z_i^{do} \neq z_j^{do} \end{cases}$$

and the following for an unordered discrete variable, x^{du}:

$$l(z_i^{du}, z_j^{du}, \lambda_{du}) = \begin{cases} 1, & z_i^{du} = z_j^{du} \\ \lambda_{du}, & z_i^{du} \neq z_j^{du} \end{cases}$$

For my continuous covariates I use a standard first-order Gaussian kernel.

Smoothing over the dependent variable as well as the independent variables, our estimator of F(y|z) is

$$F(y \mid z) = \frac{n^{-1} \sum G(\frac{y - Y_i}{h_0}) K_\gamma(Z_i, z)}{\mu(z)}$$

Where $\mu(z) = n^{-1} \sum K_\gamma(Z_i, z)$, $K_\gamma(Z_i, z)$ is the product of the individual kernels, and $G(v) = \int_{-\infty}^{v} w(u) du$ with density function *w(u).*

Most work on nonparametric conditional density estimation has used rule-of-thumb bandwidths, but I want to use the optimal bandwidths. An automatic data-driven bandwidth selection method does not exist for estimating the optimal bandwidths for a conditional cdf so Li and Racine advocate using Hall et al's (2004) conditional PDF method. They show that $h_s \sim n^{-1/(4+q)}$, $h_u \sim n^{-2/(4+q)}$, and $h_o \sim n^{-2/(4+q)}$ for continuous, unordered, and ordered variables, respectively The scale factors are then obtained using the conditional PDF method. I use the np package in R to compute the optimal bandwidths.

After obtaining the conditional cdf, I obtain the unconditional cdf by integrating over the range of covariates in that year:

$$F_t(y) = \sum_i F_t(y \mid z_{i,t})$$

Finally, I obtain the quantiles by inverting the unconditional cdf:

$$Q_t(\theta) = \{y_t : F_t(y) \geq \theta\}$$

To construct the counterfactual distribution, I use the dependent variable from 1995 but the covariates from 2001. This allows us to ask what the distribution would have been if the covariates in 2001 had the returns of 1995.

$$F_c = \sum_i F_{95}(y_{95} \mid z_{i,01}, t = 1995)$$

As before, I also want to simulate the counterfactual distribution if just one endowment was distributed as in the alternate year. My new method for the nonparametric model is conceptually identical to MM's method for the parametric model. Divide both continuous and discrete covariates into J classes as before. Let $f_j(1995)$ be the relative frequency of class j in 1995. Let n_{01} be the sample size in 2001. Estimate $F_{2001}(y|z_i)$ for each z_i in 2001. For each class j, take the subset of $F_{2001}(y|z_i)$ for which the covariate of interest takes value j. Resample from this subset with replacement to generate a sample of size n_{01}*$f_j(1995)$. This yields a new set of size n_{01} of conditional cdfs, reweighted as if the covariate was distributed as in 1995. To construct the counterfactual unconditional cdf, take the sum as before.

Finally, as Leibbrandt et al (2006) note, determining the effect of a specific return does not make conceptual sense in the nonparametric world. For example, suppose consumption was a function of urban/rural location, household size and the interaction between these two variables:

$$cons = \alpha_1 + \alpha_2 urban + \alpha_3 hhsize + \alpha_4 urban*hhsize$$

Then the return to urban location is $\alpha_2 + \alpha_4$*hhsize* and that to household size is $\alpha_3 + \alpha_4$*urban*. Changing the return to urban location alone then does not make sense since it also changes the return to household size. In a nonparametric model, I allow all variables to interact so I cannot identify the effect of one coefficient holding all else equal.

4.3 Comparison to Alternative Methodologies

This paper is most similar to Melly (2006) in that it compares parametric and nonparametric quantile regression decomposition results. However, I use a different estimator for the parametric model and construct the nonparametric estimator slightly differently. In addition, Melly does not employ the kernels and optimal bandwidth selection method I use. For the parametric model, rather than using a resampling method, Melly estimates the entire quantile regression process, constructs the conditional cdf from this, and then integrates it to get the unconditional cdf. He shows that his method is numerically equivalent to the MM procedure as $m \rightarrow \infty$. In finite samples the MM procedure has a higher MSE than Melly's estimator. I chose to use the MM procedure, however, because of the ease of constructing counterfactual distributions with just one coefficient or endowment distributed as in the other year, and because increasing m does not unrealistically increase computational time. Rather than using m equal to the sample size which most papers do, I use m=5000 for my simulations where my sample size is at most around 1000 in a given year. Finally, Melly's nonparametric method has one more step than mys. Rather than estimating the conditional cdf directly, he estimates the quantile regression function by local linear quantile regression, constructs the conditional cdf, and then follows the same steps as us. My construction of the conditional cdf is equivalent to local constant quantile regression so the results from my two estimators are likely to differ slightly.

Dinardo et. al.'s (DFL) semi-parametric kernel reweighting method is similar to my parametric and nonparametric methods. They reweight the marginal distribution as if the endowments had been distributed in another year by using a logit to predict the probability of being in one year given the set of endowments. The predicted probabilities provide the reweighting factors for the counterfactual kernel density. Like mys, this method determines the part of an observed difference due to differences in endowments. It can also be used to determine the part accounted for by a specific endowment. Leibbrandt, Levinsohn, and McCrary (2006) extend the DFL methodology to account for the part of the difference explained by the returns. The DFL method has been widely applied, both in labor and development economics. Rather than using a semi-parametric procedure like DFL which implicitly assumes a functional form when estimating the logit model, I prefer to either use just parametric or nonparametric techniques as in my methods.

Similar counterfactual methods are also applied by Donald et al (2000) and Fortin and Lemieux (2000). Donald et al use a proportional hazards model to estimate the conditional distributions. Fortin and Lemieux utilize an ordered probit to model the probability that the dependent variable exceeds some cut-off value.

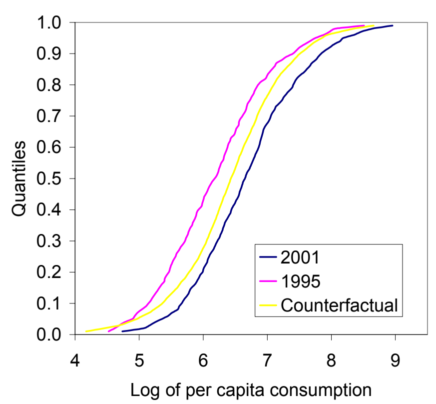
5. Results and Discussion

5.1 Results of Parametric versus Nonparametric Model

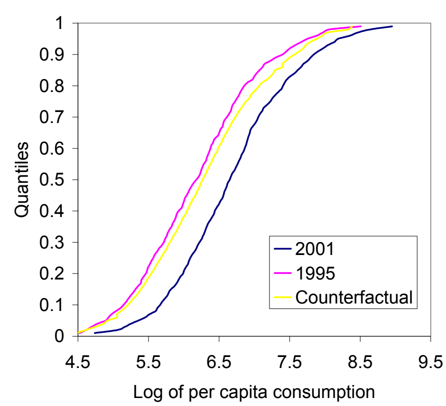
I follow the procedures outlined in section 4, decomposing the difference between 1995 and 2001 into the part due to changing returns and the part due to changing endowments. For each model, Figure 5.1 shows the empirical 1995 and 2001 CDFs and the counterfactual of the CDF that would have prevailed in each province had all 2001 returns been as in 1995. The horizontal distance between the 2001 and counterfactual curves represents the effect of the returns, while the distance between the counterfactual and 1995 curves represents the effect of the endowments. Figure 5.2 shows the contributions of endowments, returns, and residuals, the amount of the difference not accounted for by either, at each quantile.

Figure 5.1 Results

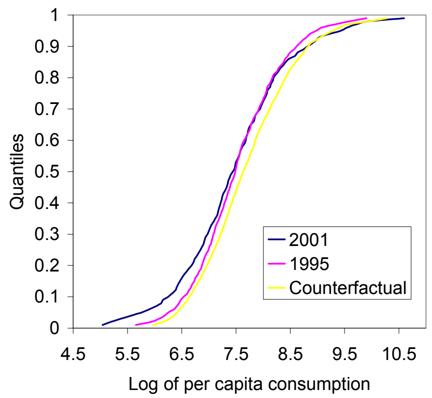
5.1.a Mashonaland Parametric



5.1.b Mashonaland Nonparametric



5.1.c Harare Parametric



5.1.d. Harare Nonparametric

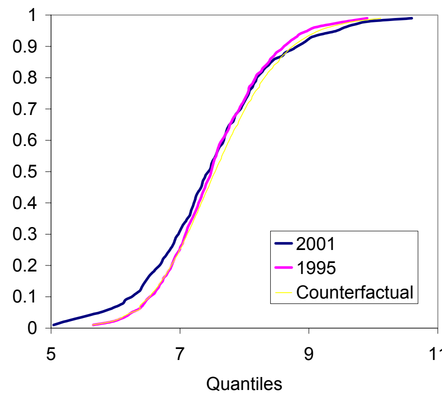
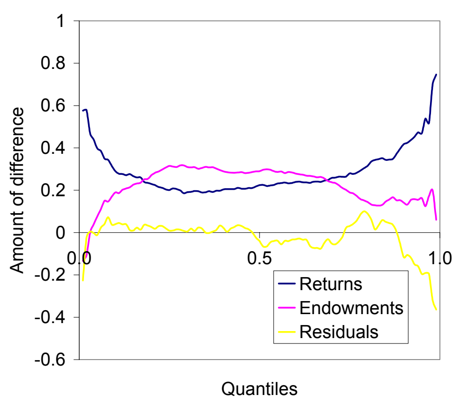
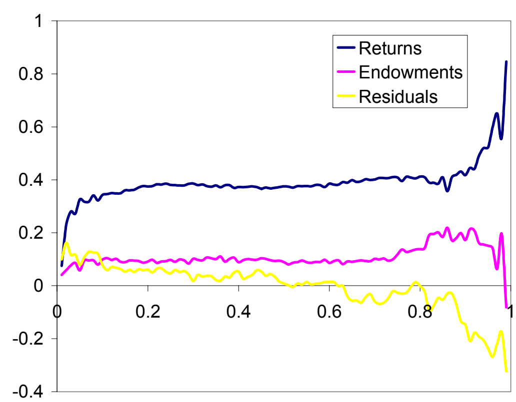


Figure 5.2. Endowments, Returns and Residuals

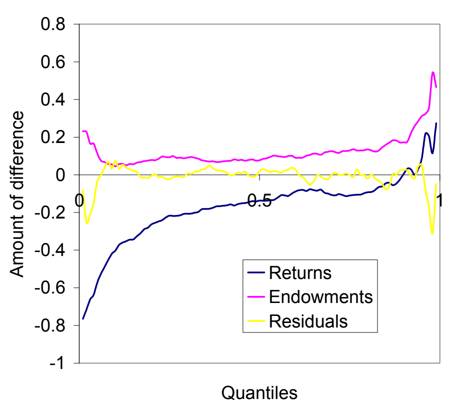
5.2.a Mashonaland Parametric



5.2.b Mashonaland Nonparametric



5.2.c Harare Parametric



5.2.d Harare Nonparametric

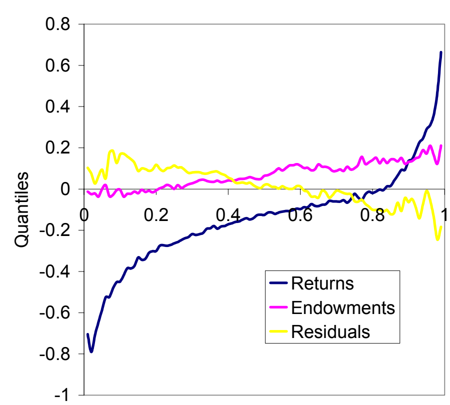


Table 5.1. Mashonaland Decomposition*

		Parametric			Nonparametric		
Quantile	Diff	Returns	Endow	Resid	Returns	Endow	Resid
10	.5172	0.2511* 0.4805 (0.0807)	0.2267* 0.4363 (0.0952)	0.0418 0.0804	.3718* .7156 (.0591)	.1005 .1933 (.0754)	.0473 .0911
25	.5248	0.2235* 0.4259 (0.0753)	0.2706* 0.5156 (0.0838)	0.0307 0.0585	.3975* .7574 (.0512)	.1082 .2062 (.0755)	.0191 .0364
50	.4731	0.1805* 0.3816 (0.0875)	0.2875* 0.6078 (0.0979)	0.0050 0.0106	.3807* .8048 (.0522)	.1193 .2521 (.0739)	-.0269 -.0569
75	.4998	0.2852* 0.5707 (0.1256)	0.1897 0.3965 (0.1278)	0.0164 0.0328	.4065* .8134 (.0874)	.1244 .2489 (.1045)	-.0311 -.0622
90	.4421	0.4441* 1.005 (0.1401)	0.1422 0.3217 (0.1455)	-0.1442 -0.3262	.3670* .8302 (.1031)	.1547 .3498 (.1186)	-.0796 -.1801

The first entry is the amount accounted for by the factor, the second entry in each cell of the decomposition columns is the proportion of the difference accounted for by that factor.
**estimates significantly different from 0 at the 5% level.*

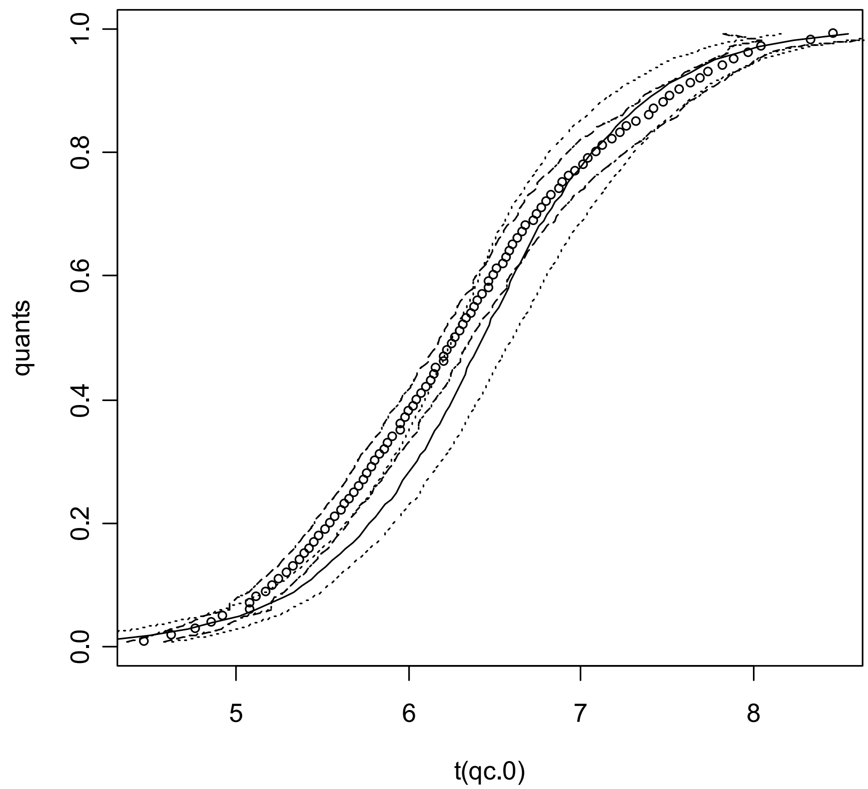
Table 5.2. Harare Decomposition

		Parametric			Nonparametric		
Quantile	Diff	Returns	Endow	Resid	Returns	Endow	Resid
10	-.2834	-0.4858* 1.7143 (0.1527)	0.0819 -0.2892 (.0879)	0.1204 -0.4251	-.411* 1.447 (.0818)	.071 -.25 (.0521)	.056 -.1971
25	-.1476	-0.3253* 2.2049 (.1051)	0.1159 -0.7857 (.0677)	0.0618 -0.4192	-.219* 1.479 (.0565)	.054 -.365 (.0464)	.017 -.1149
50	-.0416	-0.1640* 3.942 (.0853)	0.0785 -1.885 (.0606)	0.0439 -1.057	-.096* 2.341 (.0488)	.055 -1.341 (.0453)	0 0
75	.0193	-0.1275 -6.591 (.0893)	0.1175 2.109 (.0731)	0.0293 1.515	-.018 -.900 (.0612)	.043 2.15 (.0554)	-.005 -.250
90	.2045	-0.0712 -0.3485 (.1344)	0.1689 0.8260 (.0991)	0.1068 0.5225	-.136 .667 (.1039)	.073 .3784 (.0696)	-.005 -.0245

**estimates significantly different from 0 at the 5% level.*

Tables 5.1 and 5.2 present the exact estimates and standard errors for my estimates. The results of the two models differ slightly in Harare (Fig. 5.1.c, Fig. 5.1.d) but differ drastically in Mashonaland (Fig. 5.1.a, Fig. 5.1.b). Since no test currently exists to test for parametric or nonparametric quantiles, I perform two ad hoc tests of the difference between the two counterfactual CDFs in each province. First, I construct the 95% confidence band for the nonparametric counterfactual estimate. In Mashonaland, I find that the two CDFs are not significantly different because the confidence bands overlap everywhere (Figure 5.5). In Harare the CDFs are not significantly different. Second, I perform a Kolmogorov-Smirnov test for differences between the counterfactual distributions. With a null hypothesis that the two counterfactual CDFs are generated by the same distribution, I get a p value of .1609 in Mashonaland, so I cannot reject the null hypothesis. Similarly, in Harare there is no significant difference, with p = .9785.

Figure 5.5. 95% Confidence Intervals for Counterfactuals



Before exploring the reasons for the different results in Mashonaland, I also consider where the methods differ in the amount of the difference they attribute to specific endowments. Tables 5.3 and 5.4, and Appendix Figures B.1 and B.2, present the results for the endowments for which I found statistically significant shifts in section 3.1. The effect of the change in the specific endowment between 1995 and 2001 at a given quantile is captured by the difference between the 2001 quantile and the counterfactual quantile. The effect will be negative if welfare in 2001 would have been higher with the endowment distributed with 1995's distribution. In Mashonaland, the results differ between the parametric and nonparametric models for rainfall and *infant*. In Harare, the results are similar for all endowments regardless of method.

Table 5.3. Mashonaland Endowments

		Education		Employment		Rainfall		Infant	
Quantile	Diff	Par	NP	Par	NP	Par	NP	Par	NP
10	.5172	0.0194 0.0374 (0.0368)	.0109 (.0202)	0.0054 0.0105 (0.0281)	0 (.0117)	-0.0906 -0.1744 (0.0714)	0.0108 0.0209 (0.0569)	0.0642* 0.1236 (0.0228)	.1299* (.0188)
25	.5248	0.0032 0.0060 (0.0318)	.0232 (.0136)	0.0377* 0.0718 (0.0196)	-.0034 (.0097)	0.0233 0.0444 (0.0633)	0.0149 0.0285 (0.0445)	0.0503* 0.0958 (0.0194)	.0995* (.0135)
50	.4731	0.0507 0.1071 (0.0361)	.0394* (.0171)	0.0034 0.0071 (0.0194)	-.0024 (.0113)	0.0384 0.0812 (0.0622)	0.0250 0.0529 (0.0445)	0.0409* 0.0864 (0.0191)	.0996* (.0155)
75	.4998	0.0562 0.1125 (0.0469)	.0710* (.0255)	0.0451 0.0902 (0.0319)	-.0109 (.0164)	0.2494* 0.4991 (0.0868)	0.0433 0.0867 (0.0673)	0.0118 0.0237 (0.0250)	.1126* (.0248)
90	.4421	0.1118 0.2529 (0.0671)	.1120* (.0387)	0.1379* 0.3120 (0.0496)	-.0135 (.0245)	0.2239 0.5065 (0.1422)	0.0929 0.2101 (0.1008)	0.0245 0.0555 (0.0404)	.0244 (.0369)

**significantly different from 0 at the 5% level*

Table 5.4. Harare Endowments

Quantile	Diff	Education		Empl		Female	
		P	NP	P	NP	P	NP
10	-.2834	0.1497 -0.5284 (0.1718)	.0362 -.1276 (.0520)	-0.0242 0.0855 (0.1268)	-.0308 .1087 (.0396)	-0.0197 0.0693 (0.1417)	-.0109 .0384 (.0322)
25	-.1476	0.0834 -0.5655 (0.1127)	.0407 -.2759 (.0276)	0.0110 -0.0748 (0.1033)	-.0198 .1445 (.0244)	0.0055 -0.0372 (0.1034)	-.0149 -.0198 (.0251)
50	-.0416	0.0705 -1.696 (0.0830)	.0289 -.6937 (.0250)	0.0376 -0.9045 (0.0941)	-.0095 .2283 (.0253)	0.0253 -0.6080 (0.0879)	-.0181 -.0095 (.0251)
75	.0193	0.0795 4.114 (0.0984)	.0387 2.002 (.0319)	0.0089 0.4630 (0.1112)	-.1132 -.5857 (.0365)	-0.0422 -2.185 (0.1036)	-.0032 -.1132 (.0348)
90	.2045	0.1347 0.6588 (0.1696)	.0809 .3956 (.0585)	-0.0065 -0.0319 (0.1850)	0 0 (.0676)	-0.0250 -0.1226 (0.1553)	-.0096 0 (.1136)

There are three potential explanations for the differences due to methods. The first possibility is that the parametric model is misspecified; that is, in the true functional form there might be a non-linear relationship between the log of real per capita consumption and some of the covariates or some variables might interact. I explore this explanation for rainfall and *infant*. Second, the bandwidths for the nonparametric model may be suboptimal. I used the optimal bandwidth selection method but in the presence of outliers this method will tend to smooth more than without the outliers. To test this, I can recompute the bandwidths after identifying outliers and then recompute my estimates. However, the idea of an outlier does not make sense in the case of a categorical or dummy variable so in the context of *infant* it is unlikely the problem is with bandwidth selection. The third possible explanation is computational differences. The nonparametric model calculates very small numbers at some stages of the calculations so if the computer does not store enough digits the estimates could be different than those of the parametric model. This most likely does not account for the large differences but it could play a small part in explaining the differences between the two models.

I first consider whether the differences between the models can be accounted for by non-linearities in the parametric model. To see how this could affect the estimates, suppose that in the true functional form $y^*_{95}=\beta_{0,95}\ln(x_{95})$, but I have modeled it as $y_{95}=\beta_{1,95}x_{95}$. Suppose further that one data point, $x_{0,95}$ in 1995 satisfies $\beta_{0,95}\ln(x_{95})= \beta_{1,95}x_{95}= y_{95}$. The misspecified model will give an unbiased estimate of y_{95} at this particular $x_{0,95}$, but suppose that $x_{01}>x_{0,95}$ so that the counterfactual predicted value of the misspecified model is greater than the predicted counterfactual value of the true model. That is, x_{01} would have received a different return in 1995 than $x_{0,95}$ did due to the non-constant return to x. In this case, my parametric counterfactual value will be biased. However, the nonparametric model takes nonlinearities into account so it will provide a more believable prediction in this case.

There is good reason to believe that there are diminishing returns to rainfall past a certain point so this type of misspecification is reasonable for my rainfall variable. It is likely that the returns to *infant* vary according to education and employment sector so that there are interactions unaccounted for in the linear parametric model. Note that in order to perform the counterfactual analysis within the parametric model, I cannot include interactions which is a significant restriction to place on the model.

I test whether there are nonlinearities in rainfall two ways. First, I re-estimate the models with rainfall as an ordered discrete variable taking values (1) extremely low rainfall less than 70% of average; (2) less than average rainfall between 70% and 85% of average; and (3) average and above average rainfall. Second, I compute the models with the percentage of average rainfall entering in logarithmic form.

Using rainfall as an ordered discrete variable, I rerun the parametric and nonparametric models and use my Kolmogorov-Smirnov test to test the significance of the differences between the rainfall counterfactuals. With a p-value of .01893, at the 5% level I strongly reject the null hypothesis that the rainfall counterfactuals come from the same distribution. Because this functional form does not result in more similar predictions of the impact of rainfall than my original specification, it is unlikely that rainfall enters in this form.

Without even considering the counterfactual, I can reject the parametric model with rainfall entering in logarithmic form. I reject this specification because my first requirement of any model is that it predicts the data reasonably well which this model does not.

I reject the two most obvious specifications for rainfall so either there must be nonlinearities, in rainfall or other covariates, or the bandwidth selection procedure is suboptimal. I recompute the bandwidths after dropping the outliers with respect to rainfall and then reestimate the nonparametric model with all of the observations but the new bandwidths. The range of rainfall is from 9% of normal to 130%. I drop observations, all in 2001, with less than 50% of normal rainfall. Recomputing the 2001 bandwidths yields almost exactly the same bandwidths as before and statistically equivalent CDFs. Therefore, the difference is probably not due to suboptimal bandwidth selection.

The only reasonable explanation for the large differences in the results of the two models is significant non-linearities and interactions in the true functional form. Where the two models differ I are more confident in the nonparametric estimates.

5.2 Coefficients

While I can compare the results of the two models with respect to overall returns and endowments and with respect to specific coefficients, a large part of the difference is accounted for by coefficients. Only the parametric model allows identification of the effect of specific coefficients.

My choice of coefficients to consider is driven by the analysis in section 3.2. Tables 5.4 and 5.5 present the effects of specific coefficients. The constant term represents the general equilibrium effects not captured by the variables in my model. There is evidence of a nonlinear relationship with rainfall since 2001 households would have been unrealistically better off, particularly at low quantiles, with the high returns to rainfall from 1995. In both regions, female headed households received significantly lower returns and subsequently had lower welfare in 2001.

Table 5.4. Mashonaland Coefficients

Quantile	Diff	Constant	Female	Education	Employment	Rainfall
10	.5172	0.2952 0.5681	-0.1253 -0.2413	0.0292 0.0563	0.4089 0.7869	-0.6778 -1.304
25	.5248	0.3530 0.6726	-0.1273 -0.2426	-0.1686 -0.3214	0.1827 0.3482	-0.6681 -1.272
50	.4731	0.3348 0.7077	-0.0996 -0.2105	-0.4422 -0.9347	0.0772 0.1633	-0.5041 -1.066
75	.4998	-0.0107 -0.0214	-0.0923 -0.1848	-0.5207 -1.042	0.0521 0.1041	-0.3393 -0.6789
90	.4421	-0.5677 -1.284	-0.0099 -0.0223	-0.4261 -0.9639	0.0546 0.1236	-0.2294 -0.5189

Table 5.5 Harare Coefficients

	Constant	Female	Education	Employment	Lnage
10	1.108 -3.911 (1.148)	-0.0804 0.2839 (0.0499)	0.5632 -1.987 (0.7249)	-0.1582 0.5585 (0.1461)	0.2221 -0.7839 (1.140)
25	0.9535 -6.461 (1.199)	-0.0293 0.1987 (0.0396)	0.7627 -5.168 (0.6064)	-0.1288 0.8727 (0.1214)	0.2548 -1.727 (1.118)
50	0.3156 -7.583 (1.295)	0.0022 -0.0540 (0.0394)	0.7621* -18.31 (0.3505)	-0.0559 1.344 (0.1184)	0.1204 -2.893 (1.019)
75	-0.2791 -14.43 (1.464)	-0.0497 -2.569 (0.0396)	0.8136* 42.05 (0.3040)	-0.0813 -4.203 (0.1105)	-0.2084 -10.77 (0.9006)
90	-0.6471 -3.164 (1.540)	-0.0821 -0.4019 (0.0592)	0.8461* 4.138 (0.3831)	-0.1696 -0.8295 (0.1162)	-1.549 -7.575 (0.9655)

5.3 Discussion

The two regions are similar in that most of the difference is attributed to returns but they differ in the direction of the effect of returns. I need to explain why welfare increased at all quantiles in Mashonaland but only in the top part of the Harare distribution. This increase in inequality in Harare is a main concern of any analysis of the changes in the economy.

5.3.1 Mashonaland

Overall, the observed increase in welfare between the two survey years is best explained by a combination of improved endowments, general equilibrium effects unaccounted for by the variables in my model, and higher returns to education at the low end of the distribution. The positive effect of increased education levels at higher quantiles makes sense given that the increase in education was only at a level with a higher return, namely tenth grade education. As I expected, improved education endowments had a positive effect on welfare, rainfall had a positive effect over most of the distribution, and the larger number of households without children under five increased welfare as well. Increased returns to all employment sectors and their positive effect can be attributed to higher returns to communal farming in 2001 associated with a non-drought year.

I expected rainfall to be able to account for more of my observed shift. One potential reason rainfall was not more important at the low quantiles is that there was a higher correlation between consumption and rainfall amounts in 1995 than 2001. I would expect that in 1995 the low amounts of rainfall were extremely important in determining farm output and therefore consumption, but that in 2001, at least the minimum amount of rainfall was received and so other factors accounted for more of the variation in consumption.

5.3.2 Harare

Unlike in Mashonaland, where all parts of the distribution in 2001 were unequivocally better off than in 1995, in Harare welfare decreased in the bottom half of the distribution and increased in the top half. This explains the increase in poverty I find in Harare as well as the fact that mean income did not change significantly in Harare.

The main part of the difference in the bottom half is accounted for by negative returns while endowments acted opposite of the returns to raise welfare slightly. Negative returns to employment and female household head account for part of the negative returns effect. However, the effects of the returns to education and age and general equilibrium effects were positive in this part of the distribution. Overall, the negative returns are not fully accounted for by the variables in my model.

In the upper part of the distribution, the increased welfare was because the positive effect of the endowments was greater than the negative effect of the returns. The negative effect of the returns was driven by negative effects of the returns to employment, household head age, and female household head.

Throughout the whole distribution, education and female headed household endowments had the predicted effects. Increased levels of education had a positive effect across the whole distribution and the increase in the proportion of female headed households had negative effect. The positive effect of education is driven by more 10^{th} grade graduates and more people with some tertiary education. Surprisingly, the decrease in formal sector employment and rise in unemployment only had a slight negative effect, if any, on welfare.

6. Discussion and Conclusion

The most prevalent pattern across the two regions and two distributions is lower welfare for female headed households in 2001. There were more female headed households in Harare in 2001 than 1995. Decreasing returns to female headed households accompanied by more households acted to lower welfare in 2001 over what it would have been. Declines in returns to employment also acted to reduce welfare in Harare.

Improvements in both regions came from improved education endowments. Overall, most of the difference in both regions is due to returns, with a lot of the difference coming from the constant term. This returns effect is positive in Mashonaland and the lower quantiles of Harare and negative in the upper part of the Harare distribution. This paper has illustrated the usefulness of nonparametric quantile counterfactual techniques in assessing the adequacy of a parametric model. I applied both parametric and nonparametric methods to explain shifts in welfare in Zimbabwe. My findings that returns accounted for most of the shift are significant because they indicate that it was not rainfall, education, or employment endowments that account for the changes. In communal areas, the increase in welfare was due to better returns to communal farming coupled with increased returns to education in the lower part of the distribution.

References

Alwang, J., L. Ersado, et. al. (2001) “Changes in Poverty in Zimbabwe between 1990 and 1996: worsening outcomes under adverse conditions.” Development Southern Africa 18(5): 553-579.

Alwang, J., B. Mills, et al. (2003). "Changes in Well-being in Zimbabwe, 1990-96: Evidence using semi-parametric density estimates." Journal of African Economies **11**(3): 326-364.

Bassett, G and R. Koenker (1986) “Strong Consistency of Regression Quantiles and Related Empirical Processes”, Econometric Theory 104, 191-207.

Blinder, A. (1973). "Wage Discrimination: Reduced Form and Structural Variables." Journal of Human Resources **8**: 436-455.

Deaton, A. (1997) The Analysis of Household Surveys: A Microeconomic Approach to Development Policy. World Bank, Washington D.C.

Deininger, Klaus and Okidi, John (2003). “Growth and Poverty Reduction in Uganda, 1999–2000: Panel Data Evidence.” Development Policy Review **21** (4), 481-509.

Dercon, S. (2002). The Impact of Economic Reforms on Rural Households in Ethiopia. Poverty Dynamics in Africa, World Bank.

DiNardo, J., N. Fortin, et al. (1996). "Labor Market Institutions and the Distribution of wages, 1973-92: A Semiparametric Approach." Econometrica **64**(5): 1001-1044.

Donald, Stephen, et. al. (2000) “Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates” The Review of Economic Studies 67(4), 609-633.

Fortin, N and T. Lemieux (1998) “Rank Regressions, Wage Distributions, and the Gender Gap,” The Journal of Human Resources 33(3), 610-643.

Hall, P, Racine, J, & Li, Q. (2004) “Cross-validation and the estimation of conditional probability densities,” Journal of the American Statistical Association 99(2), 1015-1026.

Jolliffe, J. (2002) “Whose Education Matters in Determination of Household Income? Evidence from a Developing Country” Economic Development and Economic Change 50(2), 287-312.

Kinsey, B, J. W. Gunning, et. al. (1998) “Coping with drought in Zimbabwe: survey evidence on responses of rural households to risk”, World Development 26(1), 89-110.

Leibbrandt, M, Levinsohn, J & J McCrary. (2005) “Incomes in South Africa since the end of Apartheid” NBER Working Paper 11384.

Li, Q. and J. S. Racine (forthcoming). "Nonparametric Estimation of Conditional CDF and Quantile Functions with Mixed Categorical and Continuous Data." Journal of Business and Economic Statistics.

Liu, L. L. (2005). Pre-market Characteristics and Gender Wage Disparities among the Highly-Educated: Application and Comparison of Nonparametric Methodologies, Syracuse University.

Machado, J. a. J. M. (2005). "Counterfactual decomposition of changes in wage distribution using quantile regression." Journal of Applied Econometrics **20**(4): 445-465.

McCulloch, N., B. Baulch, and M. Cherel-Robson (2001). ‘Poverty, Inequality and Growth in Zambia during the 1990s’, WIDER Discussion Papers 2001/123, UNU/WIDER:Helsinki.

Melly, B. (2006). "Estimation of counterfactual distributions using quantile regression", Working Paper, Swiss Institute for International Economics and Applied Economic Research, University of St. Gallen.

Nguyen, B. et. al. (forthcoming). "A quantile regression decomposition of urban–rural inequality in Vietnam." Journal of Development Economics.

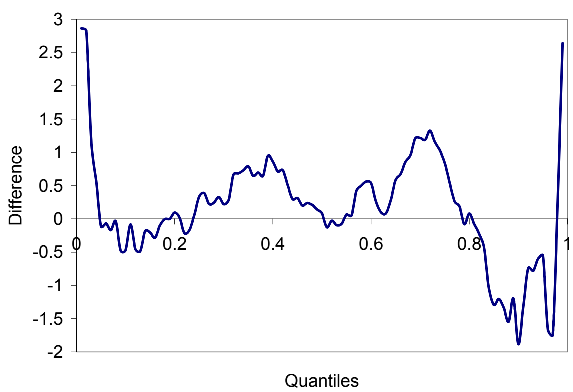
Oaxaca, R. (1973). "Male-Female Wage Differentials in Urban Labor Markets." International Economic Review **14**(3): 693-709.

Vincent, V. and R. G. Thomas. (1960) An agricultural survey of Southern Rhodesia: Part I: agro-ecological survey. Government Printer, Salisbury.

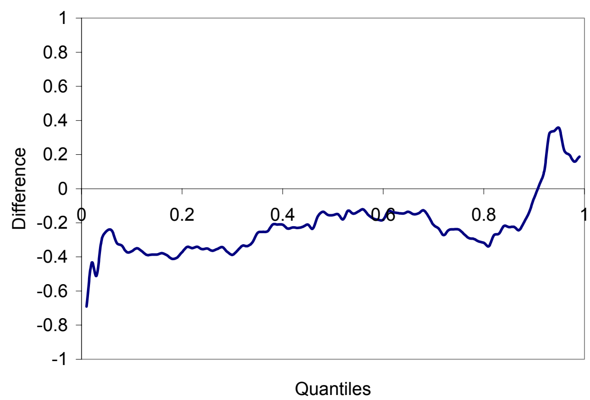
Appendix A. Coefficient Plots

Figure A.1 Mashonaland

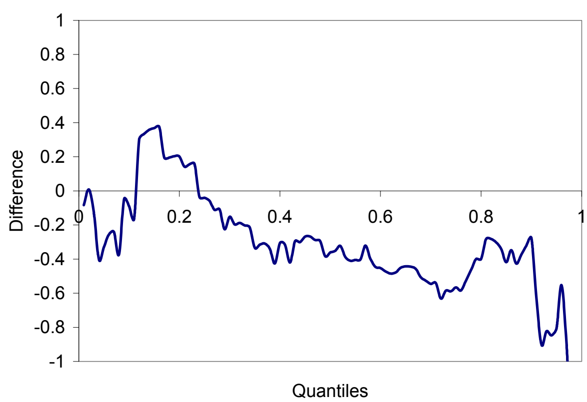
Constant



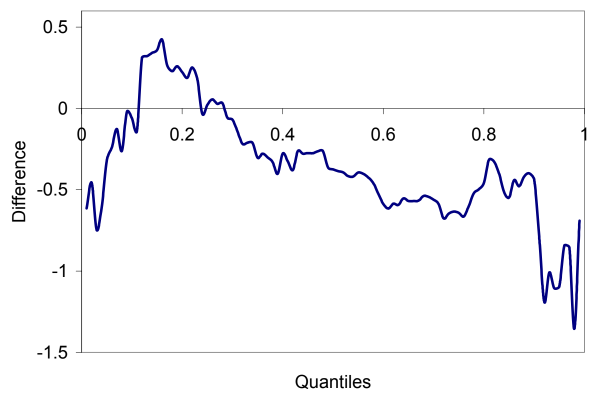
Female



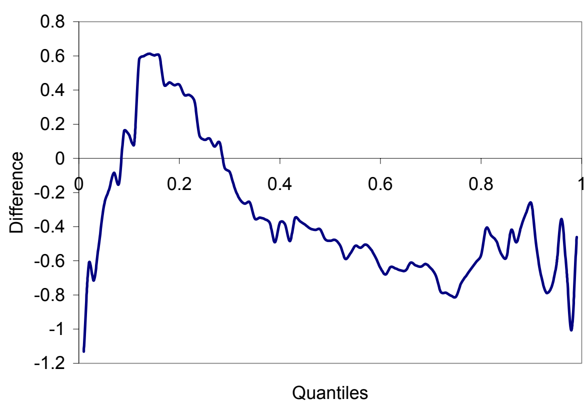
Some primary



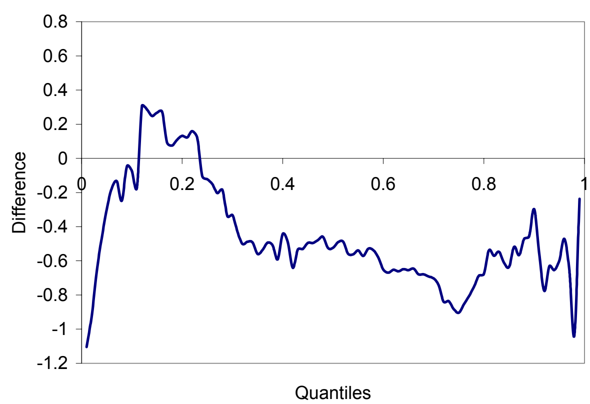
Primary



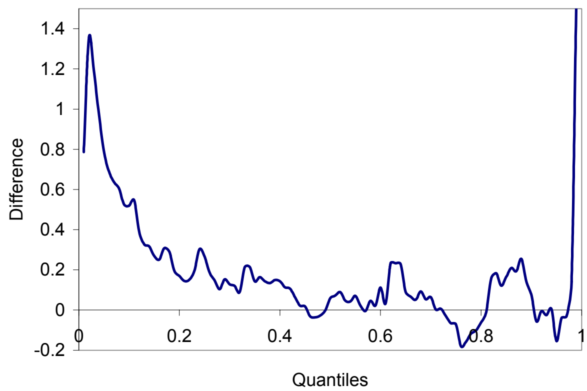
Some secondary



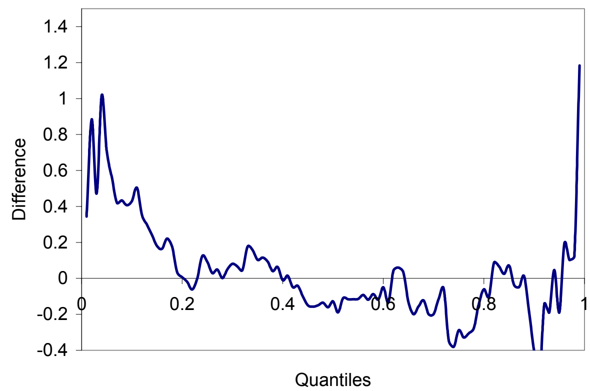
Tenth



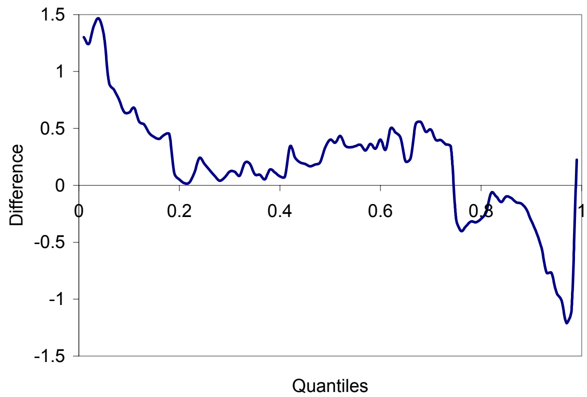
Communal Farming



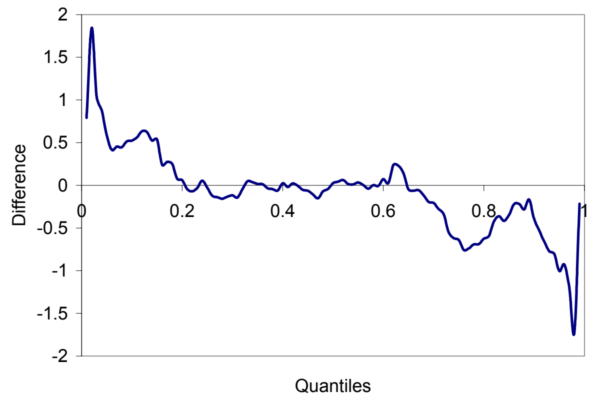
Government



Formal Sector



Informal and Own account



Rainfall

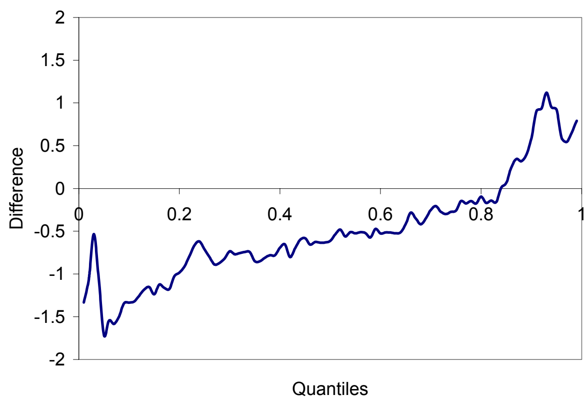
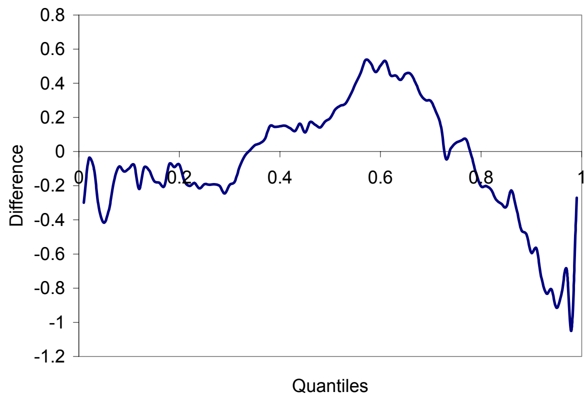
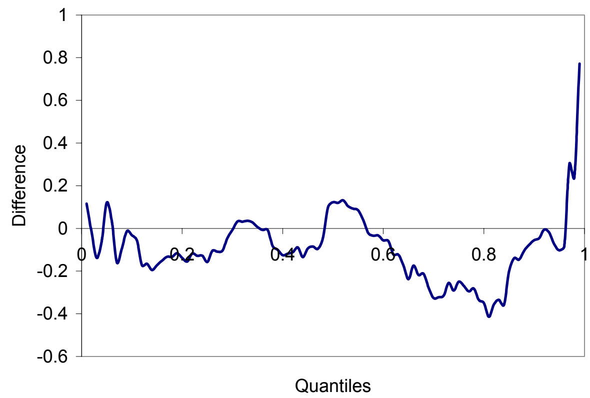


Figure A.2 Harare

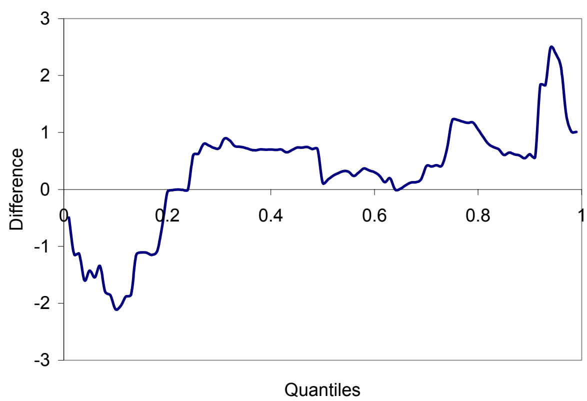
Constant



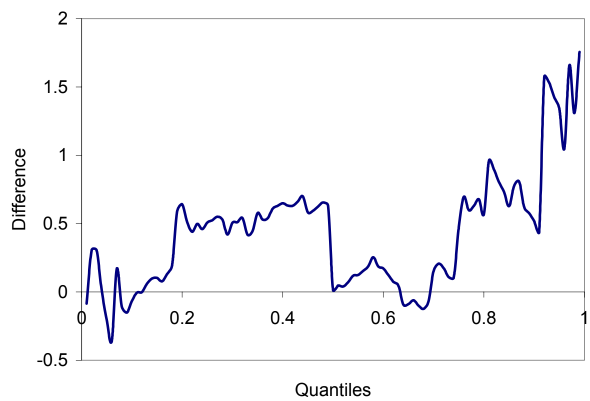
Female HH Head



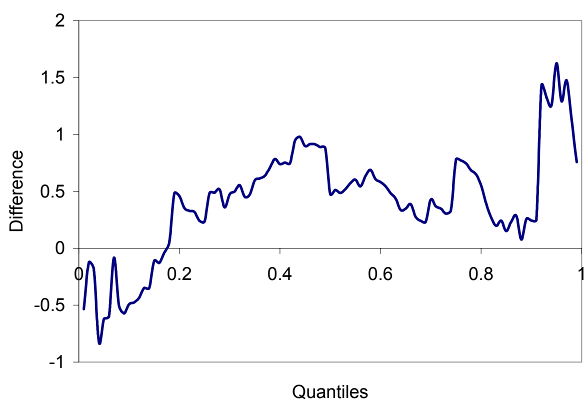
Some Primary



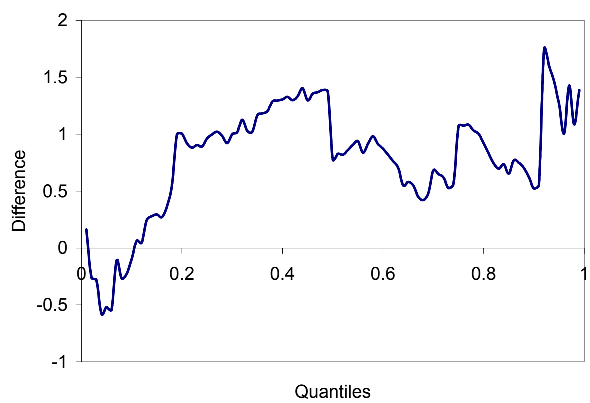
Primary



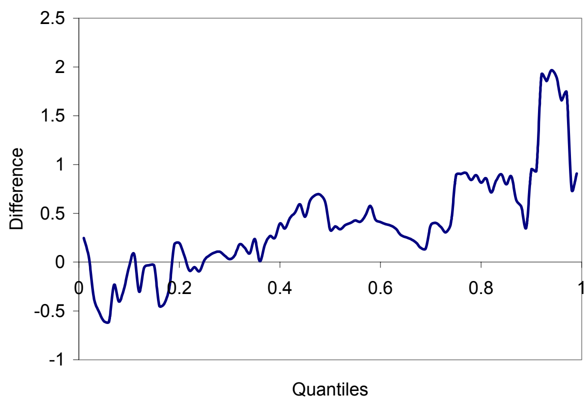
Some Secondary



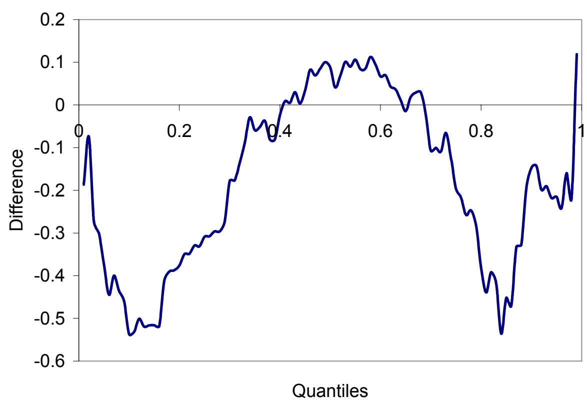
Tenth



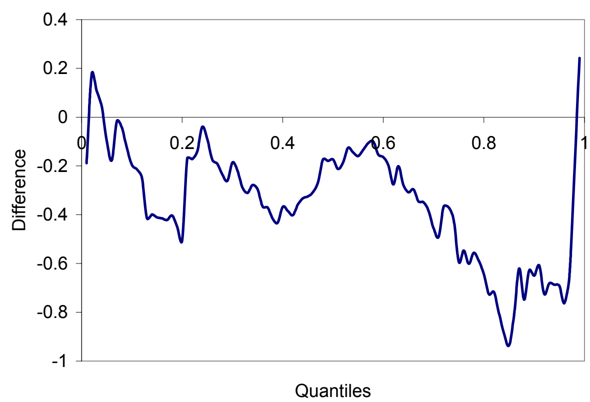
Tertiary



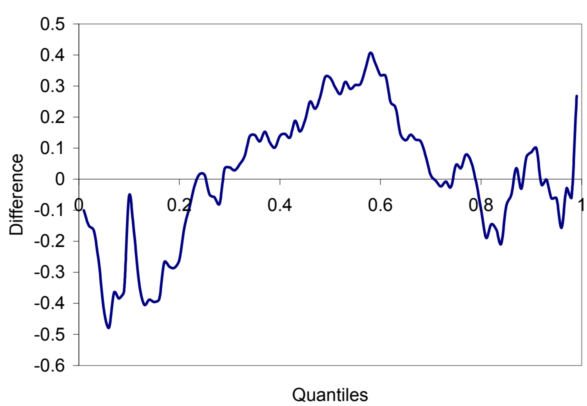
Formal



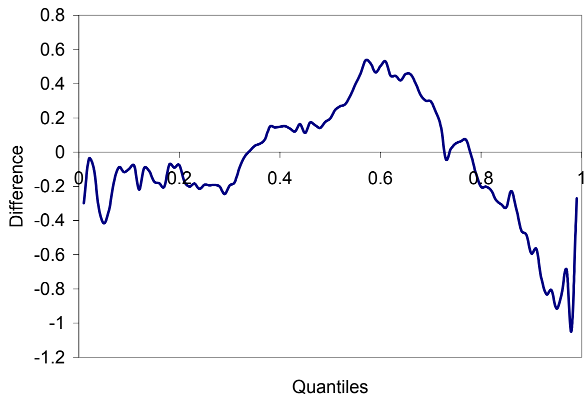
Government



Informal and Own account



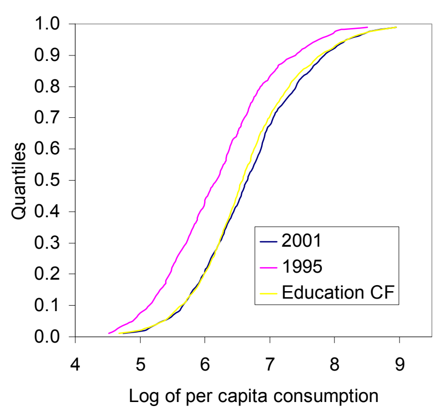
Log of HH head’s age



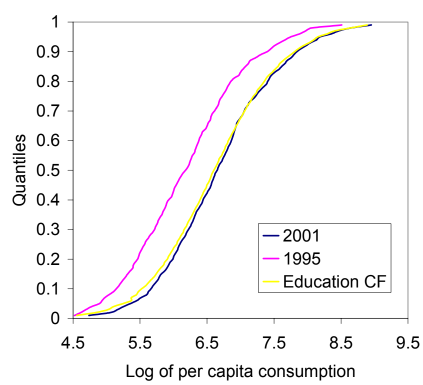
Appendix B. Counterfactual CDF plots

Figure B.1. Mashonaland Endowments

Education

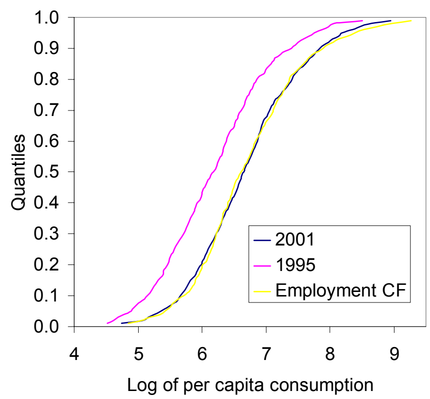


Parametric

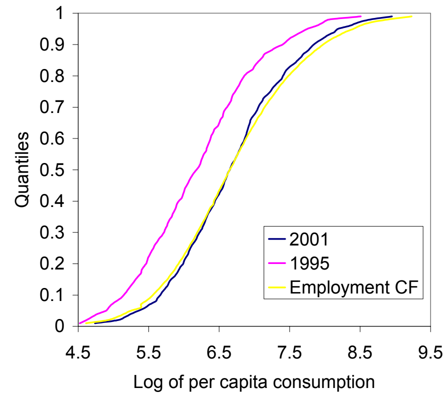


Nonparametric

Employment

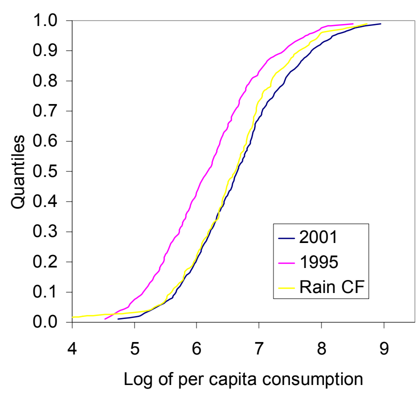


Parametric

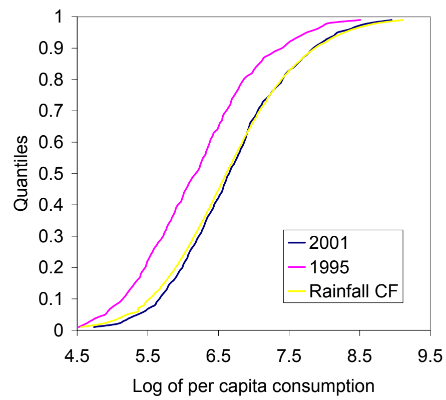


Nonparametric

Rainfall

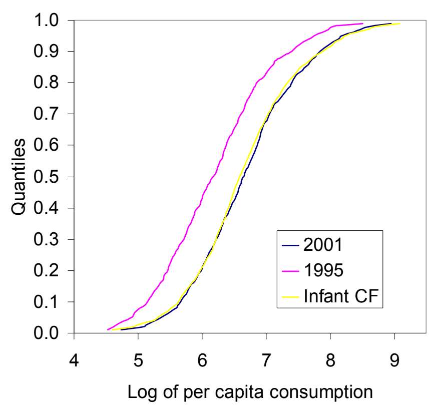


Parametric

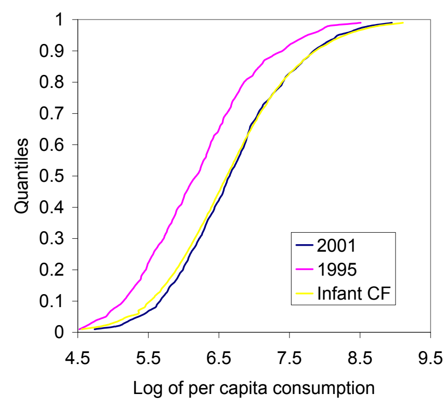


Nonparametric

Infant



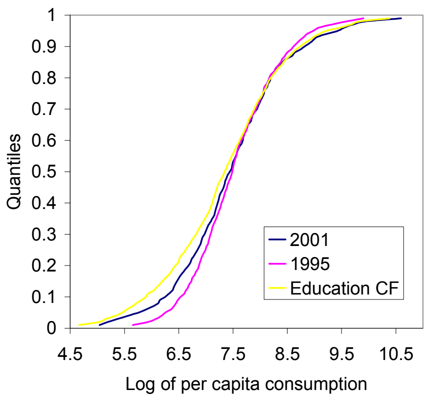
Parametric

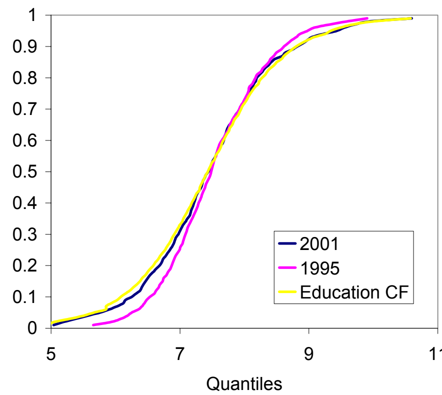


Nonparametric

<u>Figure B.2 Harare Endowments</u>

Education

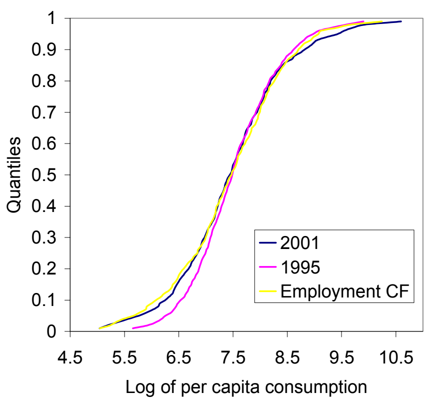


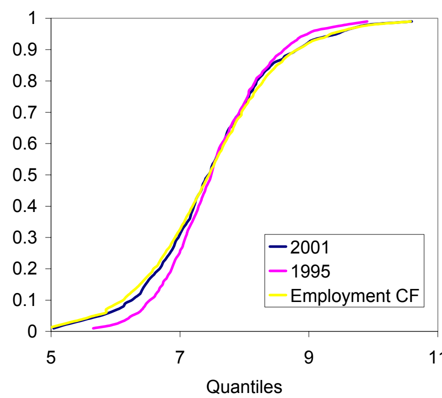


Parametric

Nonparametric

Employment

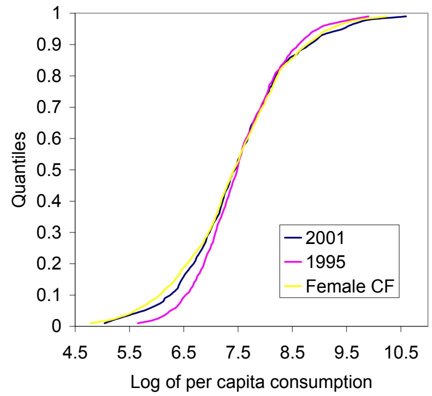




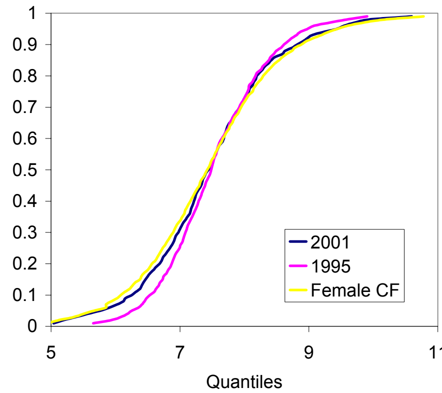
Parametric

Nonparametric

Female



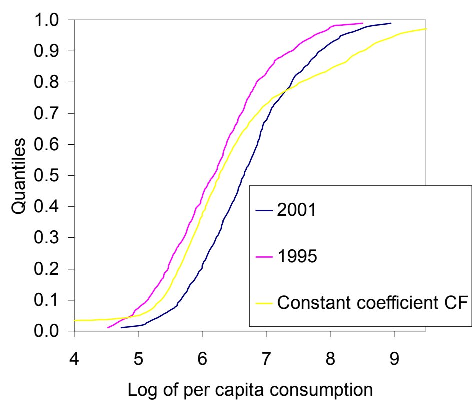
Parametric



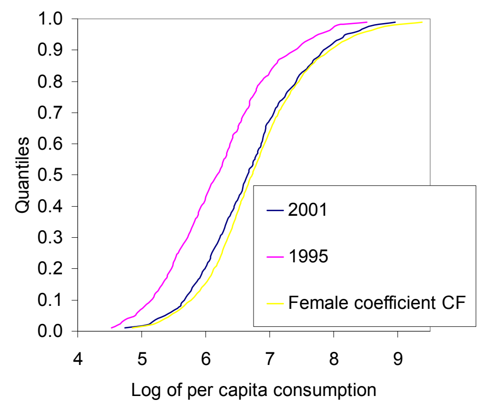
Nonparametric

<u>Figure B.3. Mashonaland Coefficients</u>

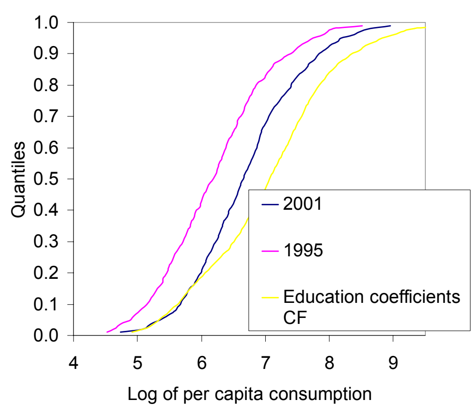
Constant



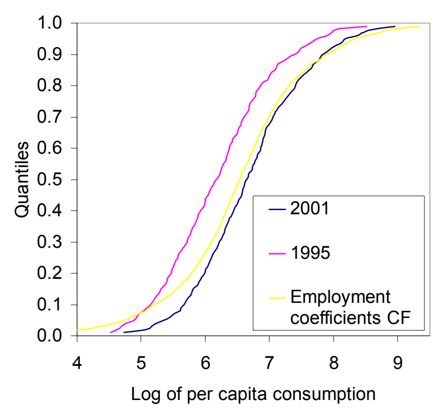
Female



Education



Employment



Rainfall

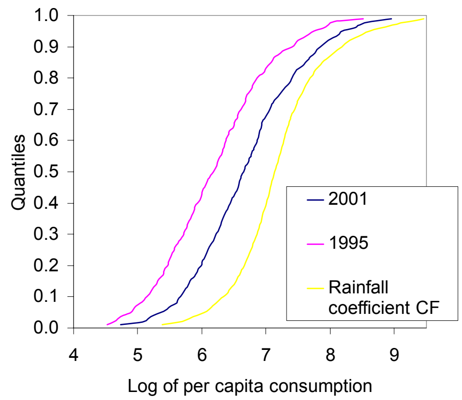
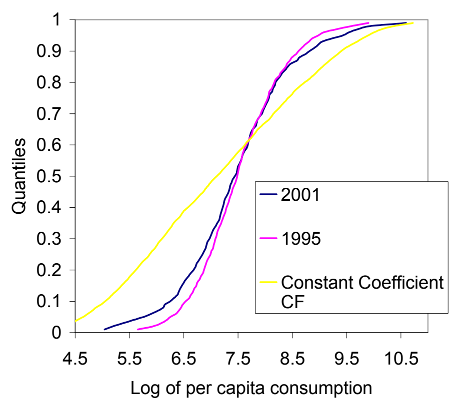
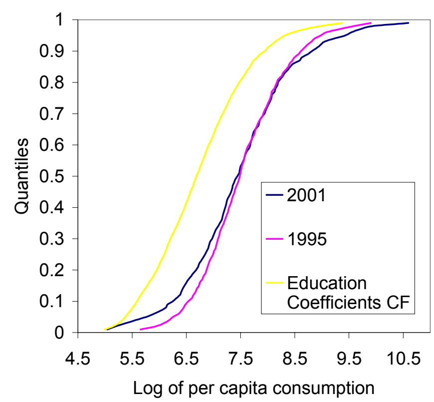


Figure B.4. Harare Coefficients

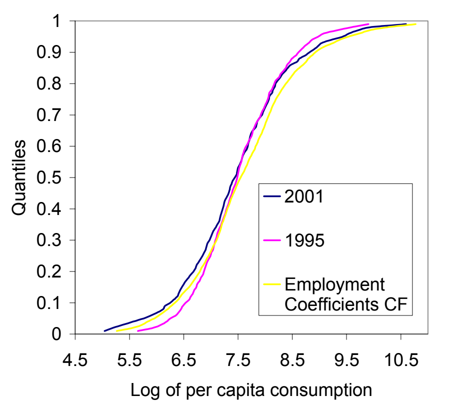
Constant



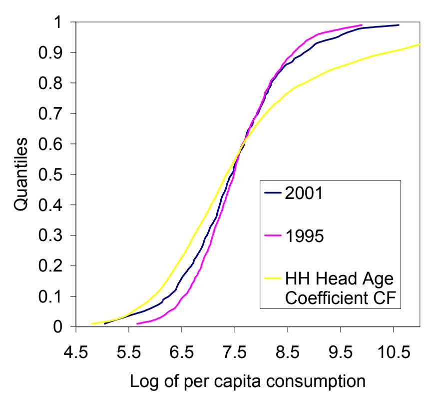
Education



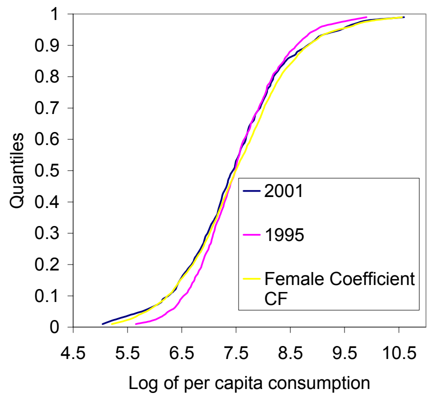
Employment



Household Head Age



Female



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Quantiles
4.5
5.5
6.5
7.5
8.5
9.5
10.5
Log of per capita consumption
2001
1995
Female Coefficient CF

Vita

Katherine Eriksson grew up in Blacksburg, Virginia. She received a BS with majors in Mathematics and Philosophy from Virginia Tech in 2004 and completed a BA in Philosophy, Politics, and Economics at Oxford University in 2006. She is currently a first year student in the PhD program in Economics at UCLA.