Essays on Applied Welfare Economics

Sutirtha Bandyopadhyay

Thesis submitted to the Indian Statistical Institute

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Essays on Applied Welfare Economics

Sutirtha Bandyopadhyay

June 2017

Thesis Supervisor: Prof. Bharat Ramaswami

Thesis submitted to the Indian Statistical Institute

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Acknowledgements

A doctoral thesis is a long journey. It is a path which cannot be traversed without the support of a number of people, whose encouragement plays an important role in reaching the final goalpost. First and foremost, I would like to thank Indian Statistical Institute for providing me an opportunity to conduct my research in an atmosphere which is intellectually stimulating and challenging. The research fellowship from the institute is gratefully acknowledged.

I would like to express heartfelt thanks to my supervisor Prof. Bharat Ramaswami for giving me the freedom to think and explore, and for being there whenever I faltered. This thesis would not have seen the light of day if it had not been for his constructive criticism, constant encouragement and guidance. I have learnt a lot from him, professionally and personally, and cannot thank him enough for his patience and endurance. I owe this research work to his foresight and academic rigor.

I would also like to thank all the faculty members of the Economic and Planning Unit at the Indian Statistical Institute, Delhi. It was an exceptionally enriching and joyful experience to have been taught by them. I would especially like to thank Prof. Tridip Ray, Prof. Prabal Roy Chowdhury, Prof.Arunava Sen, Prof. Abhiroop Mukhopadhyay, Prof. E Somanathan, Prof Farzana Afridi for their guidance and valuable comments at various stages of my research work.

I have also benefitted from discussion with several scholars. I would like to extend a special thanks to Prof. Ranjan Ray, Prof. Amita Majumdar, Prof. Parikshit Ghosh, Prof. Kam Yu, Prof. Debapriya Sen, Prof. Saikat Sinha Roy and Prof. Peter Robertson for their invaluable feedback.

During my graduate days I had the opportunity to attend several seminars and conferences. I would like to thank the participants of all these meetings for giving me feedback which has helped shape my research.

I am grateful to all my seniors and peers, who made this journey beautiful and joyful. This list is not exhaustive but I have to thank Soham Sahoo, Kanika Mahajan, Pawan Gopalakrishnan, Anuradha Saha, Priyanka Kothari, Subrato Banerjee, Dushyant Kumar, Mridu Prabal Goswami, Anup Pramanik, Digvijay Singh Negi, Dyotona Dasgupta, Anushree Saha, Prachi Singh, Kumarjit Saha, Bipul Saurabh, Gursharn Kaur, Pranabesh Das and Kaustav Nandy for having lend me an ear whenever I needed one. Sincere thanks to Mr. Athisii Kayina and to Mr. Ramesh Saini for their technical assistance. I would like to thank Mohammad Naved for his excellent research assistance during the field work of my thesis. I would also like to thank Mr. Chandan Kumar, Mr. Pramod Kumar and Ms. Shaheen Begum for helping me in surveying during the field work. I am grateful to Planning and Policy Research Unit (PPRU) for the research grant to conduct the field survey.

My deepest gratitude and thanks to my family who have stood by me in this arduous journey. I am thankful to my parents for giving me a life in which I had the luxury to choose what I wanted to do with it. Thank you for remaining patient enough and always encouraging me throughout this long journey. I cannot thank my elder brother, enough for all his words of wisdom whenever I felt overwhelmed by the task at hand. His love and faith in me, has encouraged me to carry on and do what I believe in. I must acknowledge that his presence in my life has made this journey easier. Lastly, I want to thank my sister in law whose constant inspiration and sisterly love has helped me to believe on myself.

To my Parents

Table of Contents

List of Figures

List of Tables

Chapter1

Introduction and Main Results

1.1 Background

The measurement of welfare forms the foundation of public policy analysis. It is an area where empirical investigations clearly benefits from theoretical insight and where theoretical concepts are brought alive and appropriately focused by the discipline of empirical relevance and policy design. While welfare measurement at the micro level is of independent interest, of greater concern is the well-being of groups of households/individuals.

A widely discussed issue in the literature of welfare economics is 'whose welfare to measure' when we are interested to measure the well-being of a group of individuals. The most common approach is to assume the existence of a representative agent. But sometimes this approach turns out to be unappealing due to distributional issues which are usually ignored under the representative agent framework. An alternative approach is to aggregate the welfare of individuals/households belong to a group. The aggregation as proposed by the literature can be done in various ways (Pollak, 1980).

The second chapter of my dissertation is a theoretical analysis on the issue of aggregation of individual welfare. This chapter investigates the impact of within group heterogeneity in budget share on aggregate/group cost of living index. A cost of living index is a measure of change in welfare when price changes. Although the theory of cost of living indices has been developed for individual welfare, policy interest and practical questions have invariably been concerned with group cost of living indices as a measure of changes in the welfare of that group. The important finding of this chapter is that heterogeneity in the distribution of budget share matters for group cost of living index. Even if two groups are identical in all other dimensions, the group cost of living index will be non-identical as the variability in budget share differs between two groups. Another principal finding of this chapter is that the impact of heterogeneity varies with the change in relative prices. Our results bear interesting relation to the 'representative consumer' literature. Statistical agencies that use average budget share to construct group cost of living index implicitly assumes no heterogeneity and the resulting bias is captured by our methods.

In applied welfare economics, we are typically interested in comparing the change in welfare across groups. A nice example is to consider the spatial variation in the welfare change due to any economic shock or public policy change. The third chapter of my thesis is in that direction. This chapter is an econometric evaluation of spatial impacts on prices and on wages, of India's trade liberalization in edible oils that began in the early 1990s. This chapter examines that how the impact varies across regions-especially in the regions close to ports relative to those in the hinterland and also in the high oilseeds producing regions relative to low oilseeds producing regions. While this chapter follows the literature in addressing the trade (or transport) costs that prevent the full pass-through of changes in border prices, the analysis breaks new ground in also identifying the spatial variation in competitive structures that also matter to the pass-through effects. First order welfare impact is computed for workers and consumers using the compensating variation measure.

Cost of living index/compensating variation are objective measures of welfare and fail to capture the perception of the individuals regarding the change in their well being. Therefore, in the fourth chapter of the thesis, we proceed to use subjective welfare measure instead of objective measures. This chapter is in the area of happiness and we have collected data from about thousands of households in low income neighborhoods of Delhi about happiness or life satisfaction. An important objective of this chapter is to investigate the response of the poor people on their self-reported life evaluation. This chapter also investigates the variation across gender of the relative importance of the factors that are correlated to the self reported measures of well being.

All these three chapters are in the area of applied welfare economics. In each of these chapters, we consider different issues related to applied welfare analysis in different contexts, using breadth of methodologies. An interesting resemblance across these three chapters is comparing welfare change between two populations when they are identical in all other aspects but different from each other only in terms of a single factor. In our second chapter, this distinguishing factor is the variability in budget share. The distinguishing criterion for the third chapter is distance from the port and competitiveness in the local market. In chapter 4, the distinction comes from gender. In the entire dissertation we have employed a diversity of approaches-economic theory, modern econometrics (dealing with the attribution of causality) and primary collection of data to obtain stylized facts.

The sections below contain an overview of each of the three essays in this thesis. I describe the motivation, methodology and the main findings for each essay and discuss them in detail in chapters that follow.

1.2 Does Heterogeneity Affect the Group Cost of Living Index?

Although the theory of cost of living indices has been developed for individual welfare, policy interest and practical questions have invariably been concerned with group cost of living indices as a measure of changes in the welfare of that group. A group cost of living index can be constructed in several ways. In literature, a natural and a widely used definition is to consider the aggregate/group cost of living index as an average of individual indices (Prais (1959), Muelbauer (1974), Nicholson (1975), Pollak (1980), Schultz et. al (2002), Fisher and Grilliches, (1995)).

We consider exact cost of living indices that are functions of budget shares and the change in prices. Even if we assume consumers face same prices, consumers may have different spending pattern leading to different budget shares. Can such heterogeneity in budget shares matter to the group cost of living index? Specifically, can the aggregate index be different for two populations that face the same prices (over two periods) but differ in the extent of heterogeneity in budget shares. This is the question that this chapter seeks to address.

The same question can be posed in a slightly different way. Assume the average budget shares to be same for both the populations. We also assume that variability exists in the distribution of budget share for one population but identical budget shares for everyone in the other population. In this situation the difference in the aggregate cost of living indices between two groups/populations boils down to the difference in the aggregate cost of living index and the cost of living index of an average/representative (with average budget share) individual for the same group/population. Often, there is ready access to group (or national) expenditure aggregates and it is easier to evaluate the cost of living index for an average individual. What would then be the bias – its sign and magnitude - if the group cost of living index were to be approximated by the cost of living index for a representative (average) individual? In this chapter we propose a methodology that answers that question.

Using the Rothschild-Stiglitz definition of mean preserving spread we show the impact of heterogeneity in budget shares on group cost of living index. This chapter finds that in most of the cases increase in the heterogeneity in budget shares increase the group cost of living index. Therefore, cost of living index is higher for a population with larger heterogeneity in budget shares. This result holds for a more general and important superlative index (that takes care of the substitution bias and is generated from non-homothetic preference). On further investigation, it turns out that the impact of heterogeneity is larger; greater is the change in relative prices.

As mentioned earlier, an important goal of this chapter is to characterize the bias that emerges from computing the cost of living index for a representative individual instead of the group cost of living index. Our theoretical framework shows that the bias depends on the variance in budget shares and the change in relative prices. Even if there is enough variability in budget share, the bias becomes negligible for small change in the relative prices. We also let the budget share be endogenous to relative prices and examine whether the earlier conclusion regarding the impact of heterogeneity continues to be valid.

Statistical agencies that use average budget shares to construct aggregate cost of living indices implicitly assume no heterogeneity and as we have already mentioned, the resulting bias is captured by our methods. Using Indian and US consumer expenditure data, we compute the bias. The bias computed from the real life data turns out to be small. It implies that even if the representative agent assumption is not justified theoretically, it may not be so much troubling from the empirical point of view.

1.3 Border Prices, Pass-Through and Welfare: Palm Oil in India

This chapter is regarding the edible oil trade liberalization in India. India is the world's largest importer of edible oils. This follows a sustained program, initiated in the 1990s, of eliminating quantitative restrictions, removing the monopoly of government agencies in oils imports. Among the oil imports, palm oil constitutes the largest share. Besides being the cheapest oil, the major palm oil exporting countries (Indonesia and Malaysia) are relatively closer than the major soya oil(which is the second largest imported edible oil in India) exporting countries.

Before 1994, all imports of edible oils were on government account as private trade was banned. The official policy was self reliance in oilseeds (which is an essential input to produce edible oil) production and there were government programs for promoting oilseeds production. India, however, continued to have one of the lowest oilseed yields in the world. The price support for oilseeds production was also less effective compared to the competing crops like rice and wheat and therefore oilseeds farmers could not make use of the best irrigated lands to improve productivity. Because of these concerns, Indian government liberalized importing edible oil in order to cut down its dependence on domestic oilseeds for producing edible oil. As a result of such import liberalization measures, India has gradually become the world's largest importer of edible oils and imports account for 70% of domestic consumption.

The objective of this chapter is to examine the impact of the border price (the cumulative outcome of world prices, tariffs and exchange rates) of palm oil on domestic edible oil price and domestic wage rates of agricultural labor. Change in the imported palm oil price affects the domestic edible oil price and hence the price and wage rate in oilseeds production. Since the agricultural markets for various commodities are integrated, the change in wage rate is likely to affect the aggregate agricultural wage rate and not just the wage rate in oilseeds production. In the early 1990s when trade liberalization began, oilseeds were grown on 13% of the cultivable land and were next in importance only to the cereals of rice and wheat. Therefore, at the time trade liberalization was initiated, the share of oilseeds in total agricultural production was presumably large enough for there to be appreciable effects on wages in all of agriculture.

This chapter is also a contribution to the literature on spatial impacts of trade liberalization. Constructing a theoretical model we show that the pass-through effect of the border price of palm oil on the domestic edible oil price and agricultural wage rate varies between port (coastal region) and hinterland (non-coastal region). The pass-through effect also varies between the high oilseeds producing regions and low oilseeds producing regions. The reason behind the spatial variation between port and hinterland is explained through the channel of transportation cost. On the other hand it is the spatial difference in the competitiveness of the local edible oil market that explains the varying pass-through impact between high and low oilseeds producing regions. The findings from our theoretical model regarding pass-through effects support the earlier literature (Nicita, 2009, Marchand, 2012, Atkin and Donaldson, 2015) that discuss the spatial impact of the change in border price. The key departure from the literature and the principal contribution of this chapter is that it exploits prior information about the domestic availability of substitutes to examine how that affects the spatial transmission of border prices to domestic prices and wages.

The model predicts that for limited substitutability between the locally produced edible oil and imported palm oil, the pass-through effect is higher in ports relative to inland. Similarly limited substitutability ensures the pass-through elasticity to be stronger in the high oilseeds producing regions relative to low oilseeds producing regions. We do not get such findings if local oil and imported oil are perfect substitutes.

Assembling a panel data set at the district level for five periods (1993-94, 1999-2000, 2004-05, 2007-08 and 2011-12), the price and wage pass-through effects are estimated and compared spatially. We run a district fixed effect regression controlling for time trend and many other factors.

From the basic specification of our regression, it turns out that an increase in the border price of palm oil (either from the increase in the world price or ad-valorem tariff rate or both) significantly (1% level) increases the domestic edible oil price. The pass-through elasticity in the low oilseed producing districts of non-coastal states is 0.63. If these (low oil producing) districts were located in coastal states, the pass-through elasticity would be significantly (1% level) higher by 0.12. Similarly, compared to the benchmark of low oil seed producing districts of non-coastal states, the pass-through is significantly higher by 0.15 in the high oil seed producing districts of non-coastal states. Therefore the empirical result is consistent with the imperfect substitute case in the theoretical model.

Turning to the wage regressions, it can be seen that the pass-through of palm oil border price on wages is also positive. The pass-through elasticity is 0.34 for the base category of a low oil producing district in a non-coastal state. But the differential wage impact across spatial categories is not as robust/prominent as compared to the price effect. The results for the price and wage regressions are robust to alternative specifications.

Finally first order welfare impacts for workers and consumers are computed using compensating variation measure. It turns out that spatially varying price and wage effects have important welfare implications. We find that the average compensating variation induced by the change in the palm oil world price or ad-valorem tariff rate vary spatially because of varying pass-through elasticity across regions.

1.4 Poverty, Gender and Well Being: A Study on the Slum Population in Delhi

This chapter is all about economics of happiness/subjective well being. Subjective well being/happiness as an empirical measure of welfare is gradually becoming more accepted by the economists and the policy makers. Although there is a reasonable amount of literature on happiness or life satisfaction, there are few research papers on life satisfaction among the poor. The most notable exceptions are the papers by Banerjee, Deaton and Duflo (2004) and Case and Deaton (2005). These authors find that poor tend to report high levels of happiness/life satisfaction. This is quite a surprising finding given their low standard of living, inconveniences in life and deprivation in terms of facilities they receive. On the other hand, their studies find that the poor report low levels of financial satisfaction. Banerjee, Deaton and Duflo (2004) and Case and Deaton (2005) conjecture that the poor people are adapted to their life they experience every day. Yet they are not adapted in the same way to their financial status.

These authors were confined to rural areas only. In rural areas, there is less number of rich/affluent households surrounding the poor people. Therefore the rural poor are unlikely to be aware of a good life and hence are presumably accustomed to the life they experience. But does the story of adaption hold universally? The poor in urban areas are geographically proximate to affluent neighbourhoods and the consumption of the wealthy. If, relative to rural poor, they are more aware and therefore, more aspiring of a more comfortable life, then would adaptation play a lesser role in reporting life satisfaction? This chapter reports on a recent survey on the low income population across the slums of Delhi intended to throw some lights on this issue.

Even if the poor over-report their well being, one may find enough variation in the reported life satisfaction score(in this chapter, we use the terms happiness, subjective well being, well being and life satisfaction interchangeably and consider these as equivalent concepts). If the reported happiness measure shows enough variability, then it is interesting to find out its correlates/determinants. But the more interesting thing is to see whether these correlations differ systematically between men and women. When societies offer different opportunities and liberty to men and women, they may experience life satisfaction differently and the factors that trigger it may also differ. The impact of any factor on subjective well being may also vary between male and female respondents because of divergent preferences. While recent work has drawn attention to the temporal and spatial variation in female well-being (relative to males); this chapter is the first study, to the best of my knowledge, to examine relative well-being of women among the poor.

Using a worldwide sample from Gallup World Poll, Graham and Chattopadhyay (2012) find that as one moves from lower income to higher income countries or from less educated to more educated cohorts, subjective well-being of women relative to men improves.

An interesting question is whether this relation reflects the impact of education alone or whether it is due to the country specific omitted factors especially relating to social norms and legal rights. Can we get a similar finding from our sample that has been drawn from a more homogeneous population? In our sample of urban slum dwellers, we can safely assume that there is no variation in the omitted variables relating to legal rights and social norms. What will happen to the well being of women relative to men in our data with improvement in income/education? This chapter seeks to address this question.

In order to conduct our survey, entire Delhi is stratified into zones (East, West, North and South) and slums are chosen randomly from each of the zones. From each of the slums listed in our survey, the households are chosen through the '*k*' th household approach. This is a systematic sampling with every '*k*'th element in the frame is selected. From each household, we attempt to interview a female and a male (20 years or above). However, often enough, there is either a female or a male available for interview and not both. The sample consists of 1278 respondents residing in 989 households across 29 slums in Delhi. 60% of the respondents are female and the rest are male. The survey was conducted during the entire month of March and first week of April, 2016.

In the questionnaire, the following question is asked to assess life satisfaction: In general terms would you say that you are satisfied with life? There are four choices to answer this question. The choices are 'not at all satisfied' (score 1), 'not very satisfied' (score 2), 'pretty satisfied'(score 3) and 'very satisfied' (score 4). Similar questions are asked on health satisfaction and financial satisfaction. Apart from asking questions on life, health and financial satisfaction, we ask a host of other questions. These include general information about the respondent and his/her family members, public facilities available in the slums and their qualities, mental and physical health of the individuals interviewed.

Analyzing our own survey data, we get quite similar results to Banerjee, Deaton and Duflo (2004) and Deaton and Case (2005). We find the reported life satisfaction of the urban poor to be on the higher side. Only 11.35% of the respondents report that they are 'not at all satisfied'. The most frequent response turns out to be 'pretty satisfied' (reported by 40.47% respondents). But there are more people reporting about low value of financial satisfaction. We find 46.71% respondents who are either 'not all satisfied' or 'not very satisfied' with their financial status. The percentage of people who say 'very satisfied' (i.e. the highest score) is only 6.26%. This is much less compared to the percentage of people who report the highest level of life and health satisfaction (16.9% and 23.9% respectively). Therefore our findings show that even in urban areas, low income people get accustomed to their poor living conditions and sufferings and hence don't complain about their life in general. The adaptation mechanism works even in an urban setting. But they express their concern more regarding their financial status.

We get enough variability in the reported well being data. The variables which are found to be correlated with the life satisfaction are financial satisfaction, health satisfaction, income, marriage, age, education, possession of assets (possession of refrigerator and two wheeler), public facilities (like functioning of drainage system) and mental health/psychological traits (like loneliness and stress). An ordered logistic regression is run to see whether these bivariate correlations persist after controlling for many other factors. Some of these variables show strong correlation with life satisfaction in the regression framework as well. These include financial satisfaction, health satisfaction, income, education, possession of fridge, functioning of drainage system, loneliness and stress.

Education shows the most interesting gender varying correlation in our data. The differential effect of education on life satisfaction across gender turns out to be statistically significant i.e. the reported life satisfaction score is significantly higher for an educated man compared to an educated woman (at 5% level of significance). This finding can be considered as a cross sectional counterpart of Stevenson and Wolfers (2009) regarding the declining female happiness in United States and the industrialized nations in Europe during 1970-2005 despite improvement in the objective measures for women in the same period of time(in our story, education is the objective measure). There can be several explanations for this finding. But the most probable may be the rising aspiration of women with the increase in the education. In that sense, our finding is supported by Lalive and Stutzer (2010) who find the life satisfaction of women to be higher in the traditional communities compared to liberal communities in Switzerland and explain their finding as a result of higher expectation of the women in liberal communities.

From our data, we also find an interesting gender varying bivariate correlation in terms of access to government subsidized foodgrain, sugar and kerosene (used for cooking and/or lighting) through ration shops. The access is determined by the possession of a 'ration' card. The average life satisfaction score of the women who possess a ration card is higher than those who don't possess it. But this does not turn out to be true for men. This finding can possibly be explained by the difference in preferences between men and women. The differential gender effect of possessing ration card does not turn out to be statistically significant in a regression framework when we control for other factors.

Chapter 2

Does Heterogeneity Affect the Group Cost of Living Index?

2.1 Introduction

Although the theory of cost of living indices has been developed for individual welfare, policy interest and practical questions have invariably been concerned with group cost of living indices as a measure of changes in the welfare of that group. Given a Bergson-Samuelson social welfare function, Pollak (1981) showed that a group cost of living index can be defined in a fashion analogous to the individual cost of living index. [1] However, as Pollak points out, the premise that society has preferences that can be summarized by a social welfare function does not have universal acceptance.

A natural and a more widely used definition is to consider the group cost of living index as an average of individual indices (Prais (1959), Muelbauer (1974), Nicholson (1975), Pollak (1980), Schultz et. al (2002), Fisher and Grilliches, (1995)). There has been some debate in the literature about whether the aggregate index should be an unweighted average (the so-called 'democratic' index) or the whether the individual indices be weighted according to that consumer's share of total expenditure (the so-called 'plutocratic' index). A democratic index, as it is argued, is more representative because it weights poor and rich consumers equally.

[1] Another approach that also is based on a social welfare function is to let the social cost of living index be that uniform scaling of every individual's expenditure that keeps social welfare constant across a price change (Crossley and Pendakur, 2010)

The practical challenge in constructing a group index is that it requires cost of living index for all individuals in the population. Often, there is ready access to group (or national) expenditure aggregates and it is easier to evaluate the cost of living index for an average individual. What would then be the bias – its sign and magnitude - if the group cost of living index were to be approximated by the cost of living index for a representative individual? This is the question that motivates this chapter.

The evaluation of a group cost of living is essential to welfare analysis in many contexts and our motivating question can be posed in those situations as well. Consider the welfare effects of trade liberalization. A natural metric to measure the change in welfare is to look at the compensating variation (due to the change in trade policy) as a proportion of initial expenditure (e.g., Porto, 2006). But this is the same as the cost of living index (between the pre-liberalization and post-liberalization prices) minus one. Here again, the correct measure for aggregate welfare change would be an average over individual welfare changes. But what if average individual characteristics are used to evaluate the welfare change? What would be the bias?

If all consumers faced the same prices and if they were identical, there is no aggregation problem. Any consumer is fully representative of the group. Aggregation is necessary only when consumers are heterogeneous. Even when they face the same prices, consumers may have different spending patterns leading to different budget shares. Can such heterogeneity matters to the group cost of living index? Specifically, can the aggregate index be different for two populations that face the same prices (over two periods) but differ in the extent of heterogeneity. This is the question that this chapter seeks to address. The answer to this question also addresses the motivation mentioned earlier: what would be the bias if the group cost of living index were approximated by the cost of living index for a representative individual.

We consider exact cost of living indices that are functions of budget shares (in base period or current periods or both) and the change in prices. All individuals are assumed to face the same prices but budget shares are allowed to vary between individuals. The chapter considers the implications of such heterogeneity for the group cost of living. The chapter finds that, for the most important cases, the aggregate index (whether democratic or plutocratic) increases when heterogeneity is higher. On further investigation, it turns out that the impact of heterogeneity is larger; greater is the change in relative prices.

The plan of the chapter is as follows. The next section defines the democratic and plutocratic cost of living indices. Section 2.3 constructs a numerical example with Cobb-Douglas preferences. It is seen that greater heterogeneity in the preference parameters (which is equivalent to budget share in the Cobb-Douglas case) leads to a higher group cost of living index. It is also seen that the magnitude of the effect varies with the change in relative prices. The extent of heterogeneity from real life data has been shown in section 2.4. Section 2.5 generalizes the example to all Cobb-Douglas preferences. The important case of superlative indices and in particular the Tornqvist index is considered in section 2.6. The following section shows why the impact of heterogeneity varies with the change in relative prices and why the magnitude of the impact is likely to be modest. Changing relative prices may, however, alter budget shares. Would that alter the relationship between the change in relative prices and the impact of heterogeneity? Section 2.8 shows that this is unlikely to be the case. Indeed, an implication of the analysis is that large changes in relative prices will certainly lead to greater heterogeneity in budget shares. The analysis is extended to *m* (>2) commodities in Section 2.9 and concluding remarks are gathered together in section 2.10.

2.2 Group Cost of Living Index

Following Fisher and Grilliches (1995), Pollak (1980), Schultz et. al (2002), a group cost of living index is defined as the following

$$S = \frac{\sum_{i=1}^{N} C^i(P^1, u^0)}{\sum_{i=1}^{N} C^i(P^0, u^0)}$$

where N is the population of the group and C^i denotes the expenditure function for the '*i*'th household. P^1 and P^0 are the current and base period price vector respectively. u^0 is the reference period utility level. The numerator is the sum of minimum total expenditure required for every household to attain its reference period utility level at current period prices. The denominator is the sum of minimum total expenditure required for each household to attain its reference period utility at base period prices. The group cost of living index can be alternatively written as

$$
\begin{aligned}
S &= \frac{\sum_{i=1}^{N} C^i(P^1, u^0)}{\sum_{i=1}^{N} C^i(P^0, u^0)} \\
&= \sum_{i=1}^{N} [\{\frac{C^i(P^0, u^0)}{\sum_{i-1}^{N} C^i(P^0, u^0)}\}\{C^i(P^{`1}, u^0)/C^i(P^0, u^0)\}] \\
&= \sum_{i=1}^{N} w^i I^i(P^1, P^0, u^0)
\end{aligned}
$$

where $w^i = \frac{C^i(P^0, u^0)}{\sum_{i=1}^{N} C^i(P^0, u^0)}$ and $I^i(P^1, P^0, u^0) = C^i(P^{`1}, u^0)/C^i(P^0, u^0)$

Therefore the group cost of living index can be written as the weighted average of the household cost of living indices where the weight for a household is its share in aggregate expenditure. In the literature, this is called the ‘plutocratic’ cost of living index. The plutocratic index gives greater weight to the cost of living indices of higher income households. As a result, the ‘plutocratic index’ could end up tracking the cost of living of a household with expenditure well above the mean (Muelbauer, 1974; Deaton, 1998). For this reason, many researchers prefer the ‘democratic’ cost of living index which is just an unweighted average of individual indices (Schultz et. al (2002), Muelbauer (1974), Deaton (1998), Prais (1959), Nicholson (1975)). The democratic index is defined as

$$D = (1/N)\sum_{i=1}^{N} I^i(P^1, P^0, u^0)$$

where *N* is the total number of households.

2.3 The Problem: An Example

Let there be two commodities, *x* and *y* and two periods (period '0' or the base period and period '1' or the current period) and five households. Each households is a utility maximizer with the common Cobb-Douglas utility function

$$u = x^{\alpha} y^{1-\alpha}$$

where *x* and *y* denote the quantity consumed of the two commodities by any household. As consumer optimization leads the Cobb-Douglas exponents to be equal to the budget shares of the respective commodities, the budget shares vary with the parameter *α*. Given the utility function, the cost of living index for any household takes the following form:

$$CLI = (\frac{P_x^1}{P_x^0})^{\alpha} (\frac{P_y^1}{P_y^0})^{1-\alpha}$$

In this example, the problem is to obtain the group cost of living index for a group of five households. In both periods, these households face the same prices for both the commodities. However, between period '0' and '1', prices change and they do so in the following manner:

$$\frac{P_x^1}{P_x^0} = 1.5 \text{ and } \frac{P_y^1}{P_y^0} = 1.2$$

where $\frac{P_x^1}{P_x^0}$ and $\frac{P_y^1}{P_y^0}$ denote the change in prices of commodity *x* and commodity *y* respectively between period `0' and`1'. In this example, the price of *x* relative to *y* rises by about 25%.

The first two rows of Table 2.1 display the *α* parameter and the corresponding cost of living indices for each of the households. The table is constructed by assuming that base period cost of living index is 100 for all the households. Table 2.1 clearly shows that the variation in budget share leads to a corresponding variation in cost of living index. The democratic cost of living index, which is nothing but the simple average of household indices, turns out to be 134.4. In this example the average budget share is 0.5. We can imagine an alternative situation when all households are identical with $\alpha = 0.5$. In this case, the democratic cost of living index is 134. Clearly then, in this example, the heterogeneity in budget share matters to the group cost of living index. Another interest in the comparison is if we suppose that democratic cost of living index is approximated by evaluating the cost of living index at the average budget share. The error in the approximation is 0.4.

To do a similar exercise for plutocratic indices, we also need information about the incomes of each of the household. This is shown in the fourth row of Table 2.1. It can be readily computed that the plutocratic budget share is 0.65 and the plutocratic cost of living index is 138.8. On the other hand, if all households are identical with an α equal to the plutocratic budget share, i.e., 0.65, then the plutocratic cost of living index is 138.6. Once again, the group cost of living index is higher when households are heterogeneous.

Suppose we continue the earlier example but with one difference. The change in the prices of the two commodities is now given by the following

$$\frac{P_x^1}{P_x^0} = 2 \text{ and } \frac{P_y^1}{P_y^0} = 1.2$$

The rise in the relative price of *x* is now greater by 66% as compared to 25%. The results are in Table 2.2. The democratic cost of living index is now 157. On the other hand, if all households have an α of 0.5, the group cost of living index is 155. The difference in the group cost of living index corresponding to the two different situations is 1.3%. If we were to compute the plutocratic cost of living index instead, then that would be 168.7 (with heterogeneity) and 167 (when all households have α =0.65). In the earlier example, when relative price did not change as much, the impact of heterogeneity was smaller. Of course, in the extreme if the relative price does not change at all, i.e. if all prices increased by the same percentage, the cost of living index would not depend on the budget share and heterogeneity would not matter (Mishra and Ray (2011) discuss the distributive impact of inflation and show that inflation vary across households only when relative price changes).

In this example, we see that heterogeneity matters and its impact seems to depend on the change in relative prices. Is this true more generally? We seek an answer to this question in the following sections.

2.4 Heterogeneity in Budget Shares in Survey Data

This section looks at the heterogeneity in budget shares in real world data. Table 2.3 reports the mean and standard deviation of the budget share of food in India as revealed by that country's national consumer expenditure survey (2004-05). For 94% rural and 95% urban households, budget share of food is within two standard deviations of the mean. However, the band is large enough to suggest considerable heterogeneity. 67.5% of the rural and 65% of the urban households lie within one standard deviation of the mean. The coefficient of variation is 20% in rural areas and 27% in urban areas. On the face of it, the heterogeneity in budget share of food seems to be large. For individual items (whether food or non-food), the heterogeneity might be substantially greater. Figure 2.1 and 2.2 shows the kernel density estimate of food share for rural and urban areas. The question is whether such heterogeneity is large enough to substantially affect the group cost of living index?

It might be thought that heterogeneity in budget shares would be lower for sub-groups of population that are homogeneous in some dimension. To check this, consider the occupation group 'Agricultural Laborers'. Households that belong to this category obtain most of their income from labor on farms. The expenditure survey data show that for this group, the mean and standard deviation of food share are 0.62 and 0.12 respectively. The two standard deviation band (that contains 96% of the agricultural labor households) is not very different from that of the entire rural population. Figure 2.3 plots the kernel density of food budget share for the occupation group of 'Agricultural Laborers'.

We don't find much difference in heterogeneity in budget share across households with different demographic compositions. The median household size from NSS (2004-05) rural sample turns out to be 5. The mean and standard deviation of the food budget share for the households with size below or equal to 5 and above 5 turns out to be identical (with mean food budget share 0.6 and standard deviation 0.12). These figures are also similar to what has been found for the entire rural sample. The two standard deviation band contains 94% of the households. When we focus on the urban areas, the median household size turns out to be 4. The mean and standard deviation of the food budget share of households with size below or equal to 4 comes out as 0.5 and 0.14 respectively. For the households with size above 4, the mean and standard deviation turns out to be 0.5 and 0.13. Again these figures resemble the figures for the entire urban sample and hence heterogeneity in budget share is not different when we consider households with different sizes (number of members). We also find the heterogeneity in budget share to be almost same across households with different number of children.

The heterogeneity in budget share for the United States consumer data is found to be quite similar with the Indian consumer expenditure data.[2] The data provides total expenditure and expenditure for various items for each consumer unit (i.e. household). We compute the budget share for food after dividing the food expenditure by the total expenditure. The mean and standard deviation of the food share for the U.S. consumers corresponding to the second quarter of 2014 is reported in Table 2.4. The mean budget share of food is much less in the United States compared to India but the variability (as measured by standard deviation) is quite similar to what we have found in Indian NSS data. As the mean is low, the coefficient of variation turns out to be higher. The two standard deviation band around the mean that contains 96% of the consumer units ranges from 0 to 0.39. The kernel density of food budget share for the U.S. consumers is shown in Figure 2.4.

2.5 The Cobb-Douglas Case

To generalize the example, consider a group of *N* households. The utility function for the 'i' th household is:

$$u = x^{\alpha^i} y^{1-\alpha^i} \quad \forall i = 1,2,\ldots,n$$

[2] In the United States, the consumer expenditure survey is conducted by the Bureau of Labor Statistics. Data is collected through two different surveys: the Quarterly Interview Survey and the Diary Survey. The Quarterly interview Survey provides quarterly expenditure data for each consumer unit and the Diary Survey is conducted over two consecutive one week periods with each respondent. The quarterly interview survey is a rotating panel where 20% of the existing consumer units have been replaced by 20% new consumer units in each and every quarter.

where α^i is the budget share of commodity x for the 'i'th household. The households in the group differ only in terms of α. Define $F(\alpha)$ to be the cumulative distribution of α. The cost of living Index for the 'i' th household is given by:

$$I(\alpha^{\mathrm{i}}) = \left(\frac{P_x^1}{P_x^0}\right)^{\alpha^i} \left(\frac{P_y^1}{P_y^0}\right)^{1-\alpha^i}$$

The distribution of α across households induces a distribution of the cost of living indices $I(\alpha)$ across them as well. Let $E[I(\alpha)]$ denote the average of $I(\alpha)$ in the population of N households. We consider $E[I(\alpha)]$ as the group cost of living index. It is assumed that in both the periods all households face the same prices for both the commodities.

Now following the definition of Rothschild and Stiglitz, consider a new distribution of budget share with the same mean but higher dispersion. Call the new distribution of budget share as $G(\alpha)$. $G(\alpha)$ is a mean preserving spread of $F(\alpha)$.

The second derivative of $I(\alpha)$ with respect to α gives us:

$$d^2 I(\alpha)/d\alpha^2 = \left(\frac{P_y^1}{P_y^0}\right)\left[\left(\frac{P_x^1}{P_x^0}\right)\left(\frac{P_y^0}{P_y^1}\right)\right]^{\alpha}\left[\ln\left(\left(\frac{P_x^1}{P_x^0}\right)\left(\frac{P_y^0}{P_y^1}\right)\right)\right]^2$$

Clearly the second derivative is positive and $I(\alpha)$ is convex in α. Therefore by Rothschild and Stiglitz:

$$E_G[I(\alpha)] \geq E_F[I(\alpha)]$$

where $E_G[I(\alpha)]$ is the expected value of $I(\alpha)$ corresponding to the distribution $G(\alpha)$ and $E_F[I(\alpha)]$ is the expected value of $I(\alpha)$ corresponding to the distribution $F(\alpha)$.

Thus the group cost of living index corresponding to a Cobb-Douglas utility function increases due to a mean preserving spread of α. This result explains the example in the earlier section.

2.6 Superlative Indices

For a cost of living index generated from Cobb-Douglas utility function, the budget share used as weight is the same for the base and current period. But in principle, we want the cost of living indices to use different budget shares for the base and the current period. The information on the current period budget share is important to evaluate the substitution effect of the change in prices. Following Diewert (1976), a cost of living index is called a ‘superlative price index’ if it is generated from a utility/expenditure function that is flexible. A flexible functional form is one which provides a local, second order approximation to the true functional form.[3] Superlative price indices are practically interesting because they consider the budget share for both the base and the current period as weights. For this reason, the superlative price indices are widely accepted and used by the statistical agencies as good approximation of the true cost of living.

Diewert (1976) showed that a family of superlative price indices can be generated from a class of utility function known as the ‘quadratic mean of order *r* utility function’. For each value of *r* ($r \neq 0$), we have a utility function and the corresponding cost of living index for that utility

[3] An arbitrary function $f(x)$ provides a local, second order approximation to the true function $f^*(x)$ about the point x^* if

$$f^*(x) = f(x^*)$$

$$\frac{\partial f^*(x^*)}{\partial x} = \frac{\partial f(x^*)}{dx}$$

and

$$\frac{\partial^2 f^*(x^*)}{\partial x \partial x'} = \frac{\partial^2 f(x^*)}{\partial x \partial x'}$$

function is a superlative price index. Diewert (1976) called these indices as 'quadratic mean of order r price indices'. For two commodities, the price index is

$$M = [s^0\{(\frac{P_x^1}{P_x^0})^{r/2} - (\frac{P_y^1}{P_y^0})^{r/2}\} + (\frac{P_y^1}{P_y^0})^{r/2}]^{1/r}[s^1\{(\frac{P_x^1}{P_x^0})^{-r/2} - (\frac{P_y^1}{P_y^0})^{-r/2}\} + (\frac{P_y^1}{P_y^0})^{-r/2}]^{-1/r}$$

where s^0 and s^1 are the budget share for commodity x in period '0' and period '1' respectively.

If we fix the budget share of commodity x in period '0' i.e. s^0, the 'quadratic mean of order r price indices' (i.e., M) becomes convex in period '1' budget share i.e. s^1. So considering a mean preserving spread of s^1 for a group of households assuming fixed s^0, increases the group cost of living index (i.e., the expected value of the household cost of living indices) corresponding to 'quadratic mean of order r price indices' for that group. On the other hand, if we assume that s^1 is fixed, M is convex in s^0 as $r < 1$. Therefore, a mean preserving spread of s^0 (given s^1 is fixed) increases the group cost of living index provided $r < 1$. The Fisher price index, popularly used by the statistical agencies throughout the world is a special case of 'quadratic mean of order r price indices' when $r = 2$.

The 'quadratic mean of order r utility function' is a class of utility functions which is homothetic. Hence, for given prices all the variation in budget share is driven by variation in utility function parameters. But in real world, preferences are non-homothetic. So, the most desired cost of living index is the one which is not only superlative but also generated from non-homothetic preferences. The variation in budget share would be because of variation in utility function parameters as well as variation in income.

An instance of such an index is the Tornqvist price index that is generated from a non-homothetic translog expenditure function (Diewert, 1976). The translog expenditure function provides a second order Taylor series approximation to an arbitrary expenditure function and hence has a flexible functional form. Therefore, the Tornqvist price index is a superlative price index that comes from non-homothetic preferences. From the point of view of the economic theory, it is a good approximation to the true cost of living index. Empirically, the Tornqvist price index and the Fisher price index are close approximations of each other which makes the Tornqvist price index equally desirable as the Fisher price index for the statistical agencies (Diewert, 1978). The functional form of the Tornqvist index for two commodities(x and y) is the following:

$$\frac{C(P^1,u^*)}{C(P^0,u^*)} = \left(\frac{P_x^1}{P_x^0}\right)^{(s^0+s^1)} \left(\frac{P_y^1}{P_y^0}\right)^{1-(s^0+s^1)/2} ; u^* = \sqrt{u^1 u^0}$$

where s^0 and s^1 are the budget share for commodity x in period '0' and period '1' respectively. The reference utility level u^* is the simple geometric mean of the base and current period utility level. It can be expressed as a function of one variable by letting

$$(s^0 + s^1)/2 = s$$

Then the Tornqvist index is

$$T(s) = \left(\frac{P_x^1}{P_x^0}\right)^s \left(\frac{P_y^1}{P_y^0}\right)^{1-s}$$

Now suppose the budget shares are heterogeneous in the population. In particular, suppose the population is distributed over *M* budget shares defined by the set $Z = \{Z_1\ Z_2, \ldots Z_M\}$. Define the indicator function for the '*i*'th household as

$J_i(Z_k) = 1$ if $s_i = Z_K\ \forall i = 1,2,\ldots,N$ and $K = 1,2,\ldots,M$ and $J_i(Z_k) = 0$, otherwise. Then define the density h(Z) as $h(Z_k) = (\frac{1}{N})\sum_{i=1}^{N} J_i(Z_k) \quad \forall\, k = 1,2,\ldots,M$. The democratic group cost of living index is *ET*(*Z*) where the expectations are defined over the density *h* (*Z*) i.e.

$$E[T(Z)] = \sum_{m=1}^{Z_m} T(Z_m)h(Z_m)$$

For the plutocratic group cost of living index, define a density *k* (*Z*) as

$$k(Z_K) = \sum_{i=1}^{N} (\frac{C_i}{\sum_{i=1}^{N} C_i}) J_i(Z_K)$$
$$\forall k = 1,2,\ldots,M$$

where C_i denotes the total expenditure of the '*i*'th household. Then the plutocratic group cost of living index is defined as *ET* (*Z*) where the expectations operator is defined with respect to the density *k* (*Z*) i.e.

$$E[T(Z)] = \sum_{m=1}^{Z_m} T(Z_m)k(Z_m)$$

Both the democratic and the plutocratic indices are expected values of the household cost of living indices. However, they differ with respect to the probability weights.

The second derivative of *T*(*s*) with respect to *s*:

$$d^2T(s)/ds^2 = (\frac{P_y^1}{P_y^0})[(\frac{P_x^1}{P_x^0})(\frac{P_y^0}{P_y^1})]^s[\ln((\frac{P_x^1}{P_x^0})(\frac{P_y^0}{P_y^1}))]^2$$

As the second derivative is positive, Tornqvist price index is convex in *s*. If we consider a mean preserving spread of *s*, the group cost of living index (whether democratic or plutocratic) corresponding to a Tornqvist index increases i.e. the expected value of the Tornqvist price indices for a group of households/individuals increases if all the households/individuals in that group have the translog expenditure function and a corresponding Tornqvist price index.

2.7 What Determines the Magnitude of Impact?

We now turn to the second issue highlighted by the example constructed in section 2.3 namely that the impact of heterogeneity is greater for larger change in relative prices. Is this generally true? We answer this for the special case when a population with heterogeneity is compared to a population without heterogeneity. This case is of interest because the outcome for a population without heterogeneity corresponds to the common practice when statistical agencies use the population average budget share to compute the cost of living index. Thus our answers are relevant to examining the bias that follows from this procedure.

Consider a population group where every household faces the same prices over periods '0' and '1' and is characterized by a translog expenditure function. The corresponding cost of living index for each household is the Tornqvist price index. Let $F(s)$ be the cumulative densities of budget share for the groups. The gap between the cost of living index for this group and the cost of living index for the average individual is given by

$$G = E[T(s)] - T[E(s)]]$$

Consider a second order Taylor series expansion of the Tornqvist index $T(s)$ around the average budget share $E(s)$:

$$T(s) = T[E(s) + s - E(s)] = T[E(s)]$$
$$+[s - E(s)]\frac{dT(s)}{ds} + (1/2!)[s - E(s)]^2 \frac{d^2T(s)}{ds^2} + R_2$$

where R_2 is the remainder term of the Lagrange form. Let $h = s - E(s)$. Using well known results, we can see that the remainder term is bounded and in particular,

$$|R_2| \leq (\frac{|h|^3}{3!})M$$

where M is a positive constant such that for all positive h, $|\frac{d^2T(t)}{ds^2}| \leq M \forall t \in [E(s), E(s)+h]$ and for all negative h, $|\frac{d^2T(t)}{ds^2}| \leq M \forall t \in [E(s)+h, E(s)]$. The second derivative of the Tornqvist index is

(1) $$d^2T(s)/ds^2 = (\frac{P_y^1}{P_y^0})[(\frac{P_x^1}{P_x^0})(\frac{P_y^0}{P_y^1})]^s[\ln(\frac{(\frac{P_x^1}{P_x^0})}{(\frac{P_y^1}{P_y^0})})]^2$$

It is clear that as long as the expected budget share and the change in relative price are finite, an upper bound M exists. Since $(\frac{|h|^3}{3!})M$ is $o(h^2)$ as $h \to 0$; $|R_2|$ is $o(h^2)$ as well. Therefore, for small h, the remainder term can be neglected.

Taking expectation on both sides of the Taylor series expansion and rearranging we get

$$E[T(s)] - T[E(s)] = (1/2!)E[s - E(s)]^2 \frac{d^2T[E(s)]}{ds^2}$$

The percentage gap g is obtained from dividing both sides of above equation by $T[E(s)]$

(2)
$$g = \frac{E[T(s)] - T[E(s)]}{T[E(s)]} = \frac{(1/2!)E[s - E(s)]^2 \dfrac{d^2 T[E(s)]}{ds^2}}{T[E(s)]}$$

Now, plugging (1) in equation (2), we obtain

(3)
$$g = \frac{E[T(s)] - T[E(s)]}{T[E(s)]} = (1/2!)E[s - E(s)]^2 [\ln(\frac{(\frac{P_x^1}{P_x^0})}{(\frac{P_y^1}{P_y^0})})]^2$$
$$= (1/2!)\,\mathrm{var}(s)[\ln(\frac{(\frac{P_x^1}{P_x^0})}{(\frac{P_y^1}{P_y^0})})]^2 = (1/2!)\,\mathrm{var}(s)[\ln(\frac{(\frac{P_x^1}{P_y^1})}{(\frac{P_x^0}{P_y^0})})]^2$$

The magnitude of the impact of heterogeneity depends on the variance of the distribution of budget share as well as on the change in relative price of commodity *x* in terms of commodity *y* between period '0' and '1'. If we denote $\frac{(\frac{P_x^1}{P_y^1})}{(\frac{P_x^0}{P_y^0})}$ as λ then expression (3) can be written as

(4)
$$g = (1/2!)\,\mathrm{var}(s)(\ln \lambda)^2$$

As $(\ln \lambda)^2 = (\ln \lambda - \ln 1)^2$, $(\ln\lambda)^2$ is zero when λ equals 1 i.e. when there is no change in the relative price. Larger is the departure of λ from 1, larger is the value of $(\ln\lambda)^2$. So the magnitude of the change in relative price is captured by the term $(\ln\lambda)^2$. Differentiating *g* with respect to $(\ln\lambda)^2$ we obtain,

(5)
$$\frac{dg}{d(\ln \lambda)^2} = (1/2!)\,\mathrm{var}(s)$$

$\frac{dg}{d(\ln \lambda)^2} > 0$ regardless the value of λ. So, given the variance of the budget share, larger is the change in the relative price, greater is the impact of heterogeneity.

Equation (4) helps us obtain the impact of heterogeneity. Both the base and current period budget share lies between 0 and 1 and hence the average of the two which is *s* also lies between 0 and 1. For a random variable *X* that lies between *a* and *b*, its maximum possible variance is $(b\text{-}a)^2/4$. As a result, the maximum variance of *s* is 1/4 which measures the highest level of heterogeneity in the population. For this level of heterogeneity, Table 2.5 computes the gap *g* for various values of the change in the relative prices. The benchmark case is when both prices increase in the same proportion so that the relative price is unchanged.[4]

Without loss of generality, we can normalize the base period price ratio between the two commodities to be 1. If the price ratio remains 1 in the current period i.e. period '1', heterogeneity has no impact. As the price of commodity *x* relative to commodity *y* rises from its initial level, heterogeneity begins to assert its impact. Similarly, the impact of heterogeneity is also observed as the price of commodity *x* relative to commodity *y* falls from its initial level.

In real life data, the variance of budget share(*s*) may well be smaller than 1/4. In that case, the impact of heterogeneity will be even smaller than what is displayed in Table 2.5. We report below illustrative calculations from Indian and U.S. expenditure data. One problem in applying the analysis to Indian data is that the Indian national expenditure survey is a cross-

[4] Notice, $P_x^1 / P_y^1 =$0.83=1/1.2 i.e. the reciprocal of 1.2. Similarly P_x^1 / P_y^1=0.71=1/1.4 i.e. the reciprocal 1.4. The same is true is for the other figures. Choosing the change in relative price this way ensures symmetry in the gap *g* in Table 2.5 regardless the direction of the change in relative price (i.e. whether relative price of commodity x in terms of commodity y increases or decreases).

section and not a panel. So we cannot report the exact value of the variance of *s* - however, we can obtain upper bounds of this value. The variance of the average budget share (between the base and current period) can be written in the following way:

$$
\begin{aligned}
V(s) &= V[(s^0 + s^1)/2] = (1/4)[V(s^0) + V(s^1) + 2\,\mathrm{cov}(s^0, s^1)] \\
&= (1/4)[V(s^0) + V(s^1)] + (1/2)\,\mathrm{cov}(s^0, s^1) \\
&= (1/4)[V(s^0) + V(s^1)] + (1/2)\left[\frac{\mathrm{cov}(s^0, s^1)}{\{V(s^0)V(s^1)\}^{(1/2)}}\right]\{V(s^0)V(s^1)\}^{(1/2)} \\
&= (1/4)[V(s^0) + V(s^1)] + (1/2)(R^2)^{(1/2)}\{V(s^0)V(s^1)\}^{(1/2)}
\end{aligned} \tag{6}
$$

where R^2 is the squared correlation between period 0 and period 1 budget share. In the 2004/05 expenditure survey, the variance of budget share for food is 0.018 while it is 0.02 in 2009/10. If we consider a limiting case of R^2=1, the variance of the average between share (between 2004/05 and 2009/10) $V(s)$ turns out to be 0.02. Using this value, Table 2.6 shows how the impact of heterogeneity varies as relative price changes from its initial level of 1. The interpretation is exactly similar to the earlier table. The magnitude of the impact of heterogeneity corresponding to each change in relative price is much less right now because of lower variance of average budget share. Figure 2.5 plots the numbers from Table 2.5 and Table 2.6.

The U.S. Consumer expenditure survey conducted by the Bureau of Labor Statistics is a quarterly panel data for the consumer units (a consumer unit is equivalent to a household). In section 2.4, we saw the distribution of food share for the second quarter in 2014. The distribution of food share is almost similar between the second and third quarter because these two quarters represent two adjacent time periods. The U.S. data is a rotating panel and hence 20% of the sample/consumer units in the second quarter is replaced by 20% new consumer units in the third quarter. For each of the common consumer units (for which we have data available for both the quarters), we calculate the average food budget share (average of two quarters). The variance of the distribution of this average food share comes out as 0.01 and therefore the estimates of the impact of heterogeneity would be comparable to those from the Indian data in Table 2.6.

2.8 Endogenising the Budget Share

In the last section, we showed that the impact of heterogeneity depends on the extent of change in relative price. The analysis did not, however, consider the fact that budget share itself can respond to the change in relative prices. In this section, we let the budget share be endogenous to relative prices and examine whether the conclusion of the previous section continues to be valid. In terms of equation (4), we have to allow the possibility that the change in relative prices affects the cost of living index not just directly (as analysed in the earlier section) but also through the variance of the budget share. From equation (4), we obtain

(7) $$\frac{dg}{d(\ln\lambda)^2} = (1/2!)\,\mathrm{var}(s) + (1/2)(\ln\lambda)^2\,\partial V(s)/\partial(\ln\lambda)^2$$

Once again we consider the Tornqvist price index. As noted earlier, the Tornqvist index is generated from the translog expenditure function. A translog expenditure function for two commodities(*x* and *y*) in period '*t*' is:

$$\begin{aligned}\ln C(u^t, P_x^t.P_y^t) &= a_0^t + a_x^t \ln P_x^t + a_Y^t \ln P_y^t + (1/2)a_{xx}^t(\ln P_x^t)^2 \\ &+ (1/2)a_{xy}^t(\ln P_x^t)(\ln P_y^t) + (1/2)a_{yx}^t(\ln P_y^t)(\ln P_x^t) + (1/2)a_{yy}^t(\ln P_y^t)^2 \\ &+ b_x^t(\ln P_x^t)(\ln u^t) + b_y^t(\ln P_y^t)(\ln u^t) + (1/2)b_{00}^t(\ln u^t)^2; t = 0,1\end{aligned}$$

where the parameters $a_x^t, a_{y,}^t a_{xx}^t, a_{xy}^t, a_{yx}^t, a_{yy}^t, b_x^t, b_{y,}^t$ satisfy the following restrictions:

$$a_{xy}^{t} = a_{yx}^{t}; a_{x}^{t} + a_{y}^{t} = 1; b_{x}^{t} + b_{y}^{t} = 0$$
$$a_{xx}^{t} + a_{xy}^{t} = 0; a_{yx}^{t} + a_{yy}^{t} = 0$$

These parameter restrictions ensure that *C*(*u*, *P*) is linearly homogeneous in *P*. From the above, the budget share equations for commodity *x* in period '0' and period '1' can be derived using the logarithmic version of Shephard's lemma and the symmetry restriction i.e. $a_{xy}^{0} = a_{yx}^{0}$

(8) $$s^{0} = a_{x}^{0} + a_{xx}^{0} \ln P_{x}^{0} + a_{xy}^{0} \ln P_{y}^{0} + b_{x}^{0} \ln u^{0}$$

and

(9) $$s^{1} = a_{x}^{1} + a_{xx}^{1} \ln P_{x}^{1} + a_{xy}^{1} \ln P_{y}^{1} + b_{x}^{1} \ln u^{1}$$

From equation (8), budget share for commodity *x* in period '0' can be rewritten as

(10) $$\begin{aligned} s^{0} &= a_{x}^{0} + a_{xx}^{0}(\ln P_{x}^{0} - \ln P_{y}^{0}) + a_{xx}^{0} \ln P_{y}^{0} + a_{xy}^{0} \ln P_{y}^{0} + b_{x}^{0} \ln u^{0} \\ &= a_{x}^{0} + a_{xx}^{0} \ln(\frac{P_{x}^{0}}{P_{y}^{0}}) + b_{x}^{0} \ln u^{0} \end{aligned}$$

where the last equality follows from the homogeneity restriction $a_{xx}^{0} + a_{xy}^{0}$. Similarly from equation (9), budget share for commodity *x* in period '1' can be written as

(11) $$s^{1} = a_{x}^{1} + a_{xx}^{1} \ln(\frac{P_{x}^{1}}{P_{y}^{1}}) + b_{x}^{1} \ln u^{1}$$

Adding equation (10) and equation (11) and then dividing it by two, we obtain the average budget share as

$$s = \frac{s^0 + s^1}{2} = \frac{a_x^0 + a_x^1}{2} + \frac{a_{xx}^0 \ln(\frac{P_x^0}{P_y^0}) + a_{xx}^1 \ln(\frac{P_x^1}{P_y^1})}{2} + \frac{b_x^0 \ln u^0 + b_x^1 \ln u^1}{2}$$

$$= \frac{a_x^0 + a_x^1}{2} + \frac{a_{xx}^1 [\ln(\frac{P_x^1}{P_y^1}) - \ln(\frac{P_x^0}{P_y^0})]}{2} + \frac{a_{xx}^0 \ln(\frac{P_x^0}{P_y^0}) + a_{xx}^1 \ln(\frac{P_x^0}{P_y^0})}{2} + \frac{b_x^0 \ln u^0 + b_x^1 \ln u^1}{2}$$

We denote

$$a_x = \frac{a_x^0 + a_x^1}{2}$$

and as before,

$$\lambda = \frac{(\frac{P_x^1}{P_y^1})}{(\frac{P_x^0}{P_y^0})}$$

Then the budget share equation becomes

$$s = a_x + (1/2) a_{xx}^1 \ln \lambda + (1/2)[a_{xx}^0 \ln(\frac{P_x^0}{P_y^0}) + a_{xx}^1 \ln(\frac{P_x^0}{P_y^0})]$$
$$+ (1/2)[b_x^0 \ln u^0 + b_x^1 \ln u^1]$$

Without loss of generality, we normalize the relative price to be unity in period ‘0’ i.e.

$$\frac{P_x^0}{P_y^0} = 1 \text{ and } \ln(\frac{P_x^0}{P_y^0}) = 0$$

This simplifies the budget share equation to

(12) $$s = a_x + (1/2)a^1_{xx} \ln \lambda + B; \text{ where } B = (1/2)[b^0_x \ln u^0 + b^1_x \ln u^1]$$

(12) shows that the variation in *s* will be because of variation in three factors. a_x is an intercept term. The second term is the variation in own price substitution effect. The third term measures the non-homotheticity effect because of which budget share is correlated with the level of utility (or income). The relative price matters only for the second term; however, the variation in other terms might matter if they are correlated with the variation in a^1_{xx}.

Taking expectation on both sides of equation (12) generates,

(13) $$E(s) = E(a_x) + (1/2)E(a^1_{xx}) \ln \lambda + E(B)$$

Subtracting equation (13) from (12), we get

$$s - E(s) = [a_x - E(a_x)] + (1/2)[a^1_{xx} - E(a^1_{xx})] \ln \lambda + [B - E(B)]$$

which leads to

$$\begin{aligned} E[s - E(s)]^2 &= E[a_x - E(a_x)]^2 + (1/4)E[a^1_{xx} - E(a^1_{xx})]^2 (\ln \lambda)^2 + E[B - E(B)]^2 \\ &+ E[\{a_x - E(a_x)\}\{a^1_{xx} - E(a^1_{xx})\}](\ln \lambda) + E[\{a_x - E(a_x)\}\{B - E(B)\}] \\ &+ E[\{B - E(B)\}\{a^1_{xx} - E(a^1_{xx})\}](\ln \lambda) \end{aligned}$$

Therefore, we get the expression for the variance of *s* as

(14) $$\begin{aligned} \operatorname{var}(s) &= \operatorname{var}(a_x) + (1/4)\operatorname{var}(a^1_{axx})(\ln \lambda)^2 + \operatorname{var}(B) + \operatorname{cov}(a_x, a^1_{axx})(\ln \lambda) \\ &+ \operatorname{cov}(a_x, B) + \operatorname{cov}(B, a^1_{xx})(\ln \lambda) \end{aligned}$$

As shown in the last section, the magnitude of the change in relative price is captured by $(\ln\lambda)^2$. The change in the variance of budget share of commodity *x* because of a change in relative price can be obtained by differentiating var(s) with respect to $(\ln\lambda)^2$ which turns out to be

$$(15)\qquad \frac{d\,\mathrm{var}(s)}{d(\ln\lambda)^2}=\frac{\partial\,\mathrm{var}(s)}{\partial\lambda}.\frac{\partial\lambda}{\partial(\ln\lambda)^2}=(1/4)\,\mathrm{var}(a^1_{axx})+\frac{\mathrm{cov}(a_x,a^1_{axx})}{2\ln\lambda}+\frac{\mathrm{cov}(B,a^1_{xx})}{2\ln\lambda}$$

The first term measures the impact of a higher relative price that happens through the substitution effects. This is unambiguously positive. The numerator of the second term measures the covariance between the intercept term and the substitution effect. There is no theoretical basis for signing the covariance. The covariance is negative if individuals with large intercept terms (leading to high budget shares) also have large (and negative) substitution effects. On the other hand, if large intercept terms are associated with small substitution effects, the covariance will be positive.[5] The numerator of the third term is about the covariance of the substitution effect and the non-homotheticity effect. It is similarly indeterminate.

Substituting (15) in (7), we obtain

$$(16)\qquad \frac{dg}{d(ln\lambda)^2}=\frac{Var(s)}{2}+\frac{(ln\lambda)^2}{8}V(a^1_{xx})+\frac{ln\lambda}{4}[Cov(a_x,a^1_{xx})+Cov(B,a^1_{xx})]$$

It is clear that for large enough λ, the sum of the last three terms will be positive and hence under those conditions, a higher relative price will increase the gap between the group cost of living index and the cost of living index for a representative consumer. Notice also that the impact of higher relative price is positive if either the variation in the substitution effect is zero or if the variation in both the intercept term and the non-homotheticity term is zero.

[5] It can be shown that $a^1_{xx}=s^1(1+\varepsilon^1_{xx})$ where ε^1_{xx} is the own price elasticity of demand in the first period. Then $cov(a^1_{xx},a_x)=cov(s^1,a_x)+cov(s^1\varepsilon^1_{xx},a_x)$. While the first term is positive, there is no natural assumption about the second term. Similarly $cov(a^1_{xx},B)=cov(s^1,B)+cov(s^1\varepsilon^1_{xx},B)$. The first term will be positive for commodities whose budget shares increase with utility. There is once again no theoretical supposition for the sign of the second term.

2.9 Extension to *m* (*m*>2) Commodities

In a two commodity world, an increase in the variability of budget share of commodity *x* also increases the variability of budget share of commodity *y*. Thus, heterogeneity in budget shares is characterized by the variability in budget share of either of the commodities. In an *m*-commodity world, however, populations are possibly heterogeneous with respect to (*m*-1) budget shares. For instance, consider a Tornqvist price index defined over three commodities: *x*, *y* and *z*.

$$T(s_x, s_y) = (\frac{P_x^1}{P_x^0})^{s_x} (\frac{P_y^1}{P_y^0})^{s_y} (\frac{P_z^1}{P_z^0})^{1-s_x-s_y}$$

$$= \{\frac{(P_x^1 / P_x^0)}{(P_z^1 / P_z^0)}\}^{s_x} \{\frac{(P_y^1 / P_y^0)}{(P_z^1 / P_z^0)}\}^{s_y} (\frac{P_z^1}{P_z^0})$$

$$s_x = \frac{s_x^0 + s_x^1}{2} \text{ and } s_y = \frac{s_y^0 + s_y^1}{2}$$

P_x^1 / P_x^0, P_y^1 / P_y^0 and P_z^1 / P_z^0 are the change in prices of commodity *x*, *y* and *z* respectively between period ‘0’ and period ‘1’. s_x is the average of period ‘0’ and period ‘1’ budget share for commodity *x*. s_y is the average of period ‘0’ and period ‘1’ budget share for commodity *y*. Since the budget share of all three commodities sum up to 1, we can write

$$s_z = 1 - s_x - s_y$$

Now consider a second order Taylor's series expansion of $T(s_x, s_y)$ around the expected/average budget share of commodity *x* and *y* i.e. $E(s_x)$ and $E(s_y)$.

$$
\begin{aligned}
T(s_x, s_y) &= T[E(s_x) + s_x - E(s_x), E(s_y) + s_y - E(s_y)] = T[E(s_x), E(s_y)] \\
&+ [s_x - E(s_x)](\frac{\partial T}{\partial s_x}) + [s_y - E(s_y)](\frac{\partial T}{\partial s_y}) + (1/2!)[s_x - E(s_x)]^2(\frac{\partial^2 T}{\partial s_x^2}) \\
&+ (1/2!)[s_y - E(s_y)]^2(\frac{\partial^2 T}{\partial s_y^2}) + [s_x - E(s_x)][s_y - E(s_y)](\frac{\partial^2 T}{\partial s_x \partial s_y})
\end{aligned} \tag{17}
$$

Now taking expectation on both sides of (17) and rearranging we get,

$$
\begin{aligned}
&E[T(s_x, s_y)] - T[E(s_x), E(s_y)] = (1/2!)E[s_x - E(s_x)]^2(\frac{\partial^2 T}{\partial s_x^2}) \\
&+ (1/2!)E[s_y - E(s_y)]^2(\frac{\partial^2 T}{\partial s_y^2}) + E[\{s_x - E(s_x)\}\{s_y - E(s_y)\}](\frac{\partial^2 T}{\partial s_x \partial s_y})
\end{aligned}
$$

where the first order terms in equation (17) become zero once we take expectation. Dividing both sides of equation (17) by $T[E(s_x), E(s_y)]$ we get the counterpart of equation (3) for the three commodity case

$$
\begin{aligned}
g &= \frac{E[T(s_x, s_y)] - T[E(s_x), E(s_y)]}{T[E(s_x), E(s_y)]} = (1/2!)\frac{\mathrm{var}(s_x)(\frac{\partial^2 T}{\partial s_x^2})}{T[E(s_x), E(s_y)]} \\
&+ (1/2!)\frac{\mathrm{var}(s_y)(\frac{\partial^2 T}{\partial s_y^2})}{T[E(s_x), E(s_y)]} + \frac{\mathrm{cov}(s_x, s_y)(\frac{\partial^2 T}{\partial s_x \partial s_y})}{T[E(s_x), E(s_y)]}
\end{aligned} \tag{18}
$$

The above expression measures the percentage gap between the average cost of living index and the cost of living index that corresponds to the average budget shares. As before, this gap measures the impact of heterogeneity in budget shares in the population. For the Tornqvist price index, we have the following expressions

$$\frac{(\frac{\partial^2 T}{\partial s_x^2})}{T[E(s_x), E(s_y)]} = [\ln\{\frac{P_x^1 / P_x^0}{P_z^1 / P_z^0}\}]^2$$

(19)
$$= \frac{(\frac{\partial^2 T}{\partial s_y^2})}{T[E(s_x), E(s_y)]} = [\ln\{\frac{P_y^1 / P_y^0}{P_z^1 / P_z^0}\}]^2$$

$$= \frac{(\frac{\partial^2 T}{\partial s_x \partial s_{yx}})}{T[E(s_x), E(s_y)]} = [\ln\{\frac{P_x^1 / P_x^0}{P_z^1 / P_z^0}\}][\ln\{\frac{P_y^1 / P_y^0}{P_z^1 / P_z^0}\}]$$

Substituting these expressions in equation (18), we get

$$\frac{E[T(s_x, s_y)] - T[E(s_x), E(s_y)]}{T[E(s_x), E(s_y)]} = (1/2!)\,\mathrm{var}(s_x)[\ln\{\frac{P_x^1 / P_x^0}{P_z^1 / P_z^0}\}]^2$$
$$+ (1/2!)\,\mathrm{var}(s_y)[\ln\{\frac{P_y^1 / P_y^0}{P_z^1 / P_z^0}\}]^2 + \mathrm{cov}(s_x, s_y)[\ln\{\frac{P_x^1 / P_x^0}{P_z^1 / P_z^0}\}][\ln\{\frac{P_y^1 / P_y^0}{P_z^1 / P_z^0}\}]$$
$$= (1/2!)[\mathrm{var}(s_x)\{\ln(\frac{P_x^1 / P_x^0}{P_z^1 / P_z^0})\}^2 + \mathrm{var}(s_y)\{\ln(\frac{P_y^1 / P_y^0}{P_z^1 / P_z^0})\}^2 + 2\,\mathrm{cov}(s_x, s_y)\{\ln(\frac{P_x^1 / P_x^0}{P_z^1 / P_z^0})\}\{\ln(\frac{P_y^1 / P_y^0}{P_z^1 / P_z^0})\}]$$
$$= (1/2!)\,\mathrm{var}[s_x\{\ln(\frac{P_x^1 / P_x^0}{P_z^1 / P_z^0})\} + s_y\{\ln(\frac{P_y^1 / P_y^0}{P_z^1 / P_z^0})\}]$$

The above expression is clearly positive if (i) the budget shares for commodity *x* and commodity *y* varies in the population and if (ii) the price for each of the commodities changes at different rates. The latter implies that relative prices change. In a two-commodity world, the impact of both these factors separate out in a multiplicative way. This is not possible in the three-commodity world. The generalization to the m-commodity world is straightforward and is the following.

$$\frac{E[T(s_1, s_2, ..., s_{m-1})] - T[E(s_1), E(s_2), ..., E(s_{m-1})]}{T[E(s_1), E(s_2), ..., E(s_{m-1})]} = (1/2!)\,\mathrm{var}\left(\sum_{i=1}^{m-1} s_i \ln \lambda_i\right) \tag{20}$$

where $\lambda_i = \ln\left[\frac{(P_i^1 / P_i^0)}{(P_m^1 / P_m^0)}\right]$

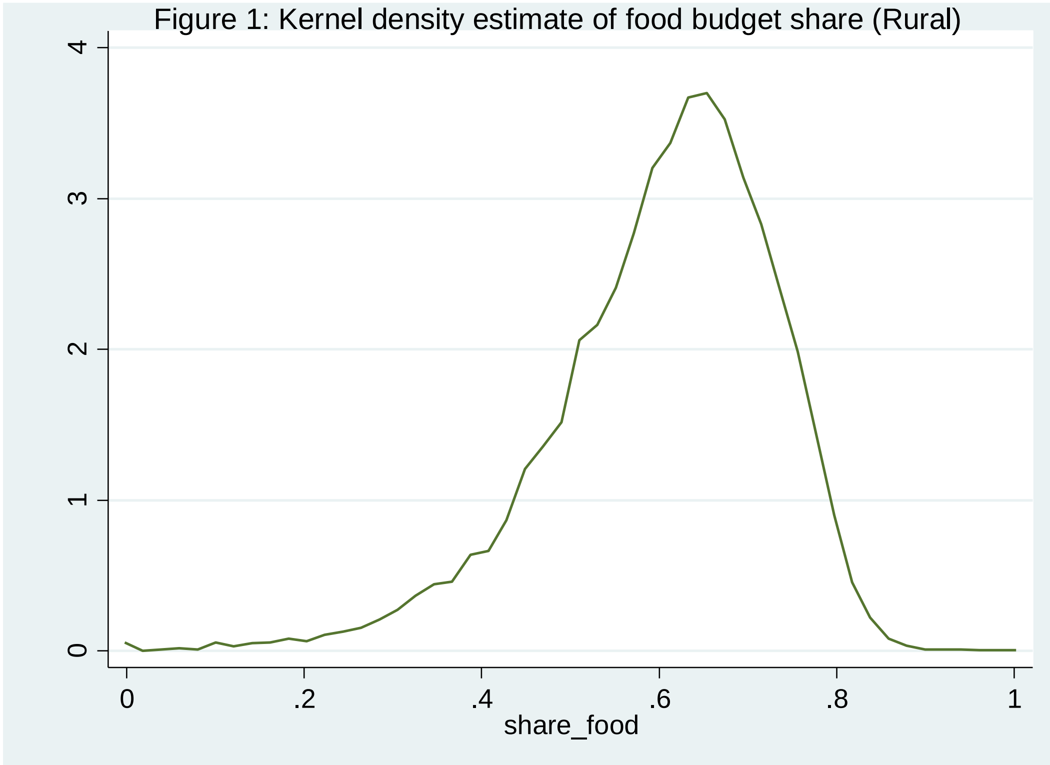
The difference is unambiguously positive. Again it is the variability of the term inside the square bracket in expression (20) that matters.

2.10 Conclusion

Many researchers have proposed the average of individual cost of living indices as a group cost of living index. Statistical agencies however rarely possess the distribution of budget shares needed to generate the distribution and hence the mean of individual cost of living indices. What is more easily known is the share of various commodities in aggregate expenditure that is nothing but the plutocratic budget share. The practical thing to do is to evaluate the cost of living index at the average budget share. The resulting bias has been the focus of this chapter.

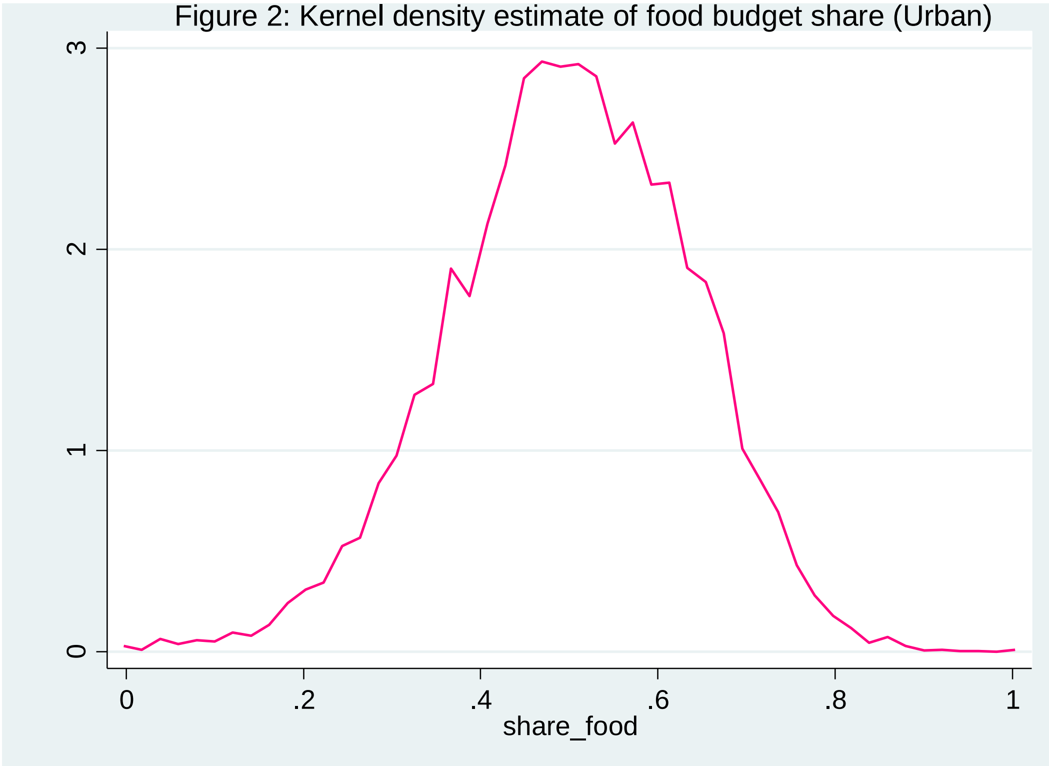
What this chapter has showed is that in most cases and for the important case of the Tornqvist index, the nature of the bias will be to underestimate the true group cost of living index. Furthermore, the magnitude of the bias depends on the extent to which the relative price structure changes between the base and current periods. For modest changes in relative prices, the bias is small and the group cost of living index is approximated well by a cost of living index evaluated at the average budget share.

Figure 2.1:- Kernel Density Estimate of Food Budget Share (Rural)



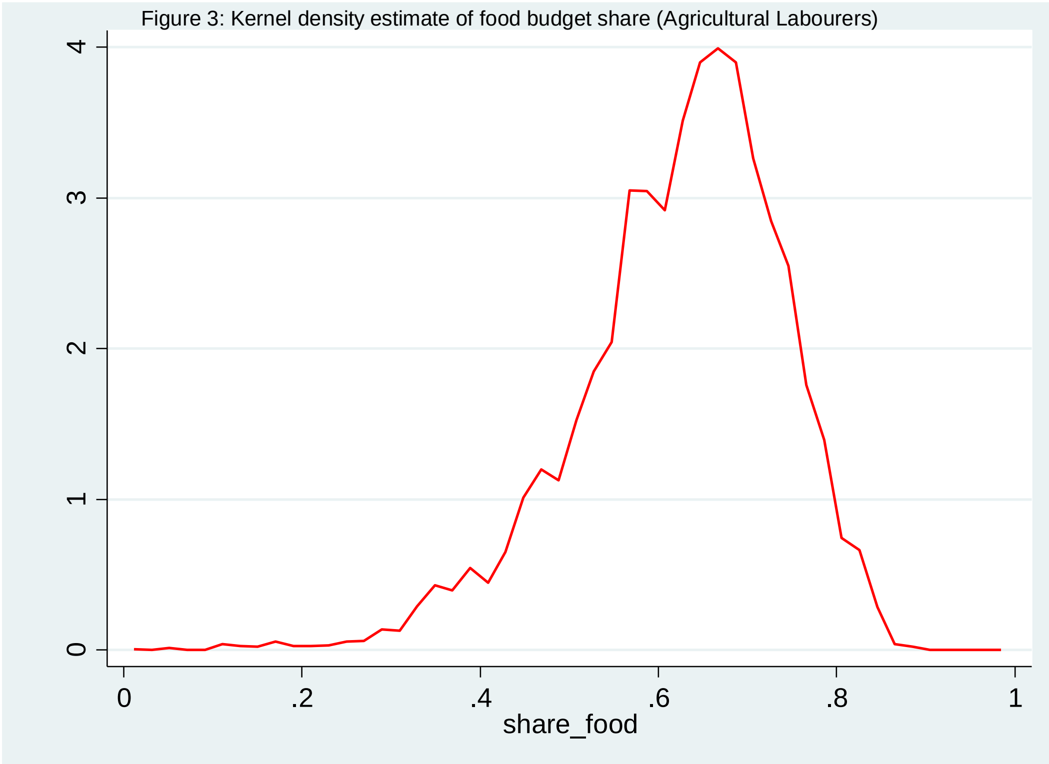
Note:-Drawn from National Sample Survey consumer expenditure survey data (2004-05)

Figure 2.2:- Kernel Density Estimate of Food Budget Share (Urban)



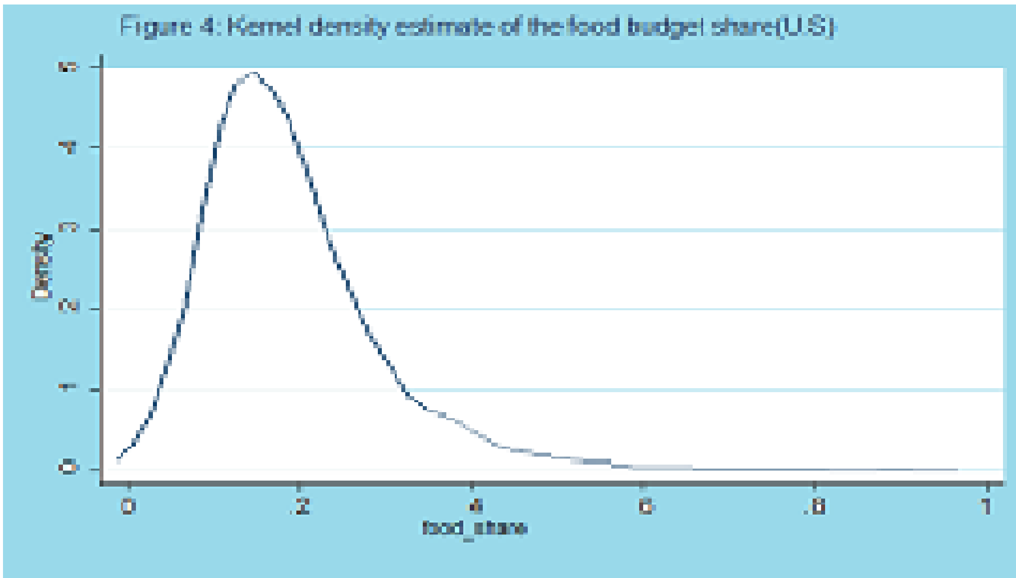
Note:- Drawn from National Sample Survey consumer expenditure survey data (2004-05)

Figure 2.3:- Kernel Density Estimate of Food Budget Share (Agricultural Labourers)



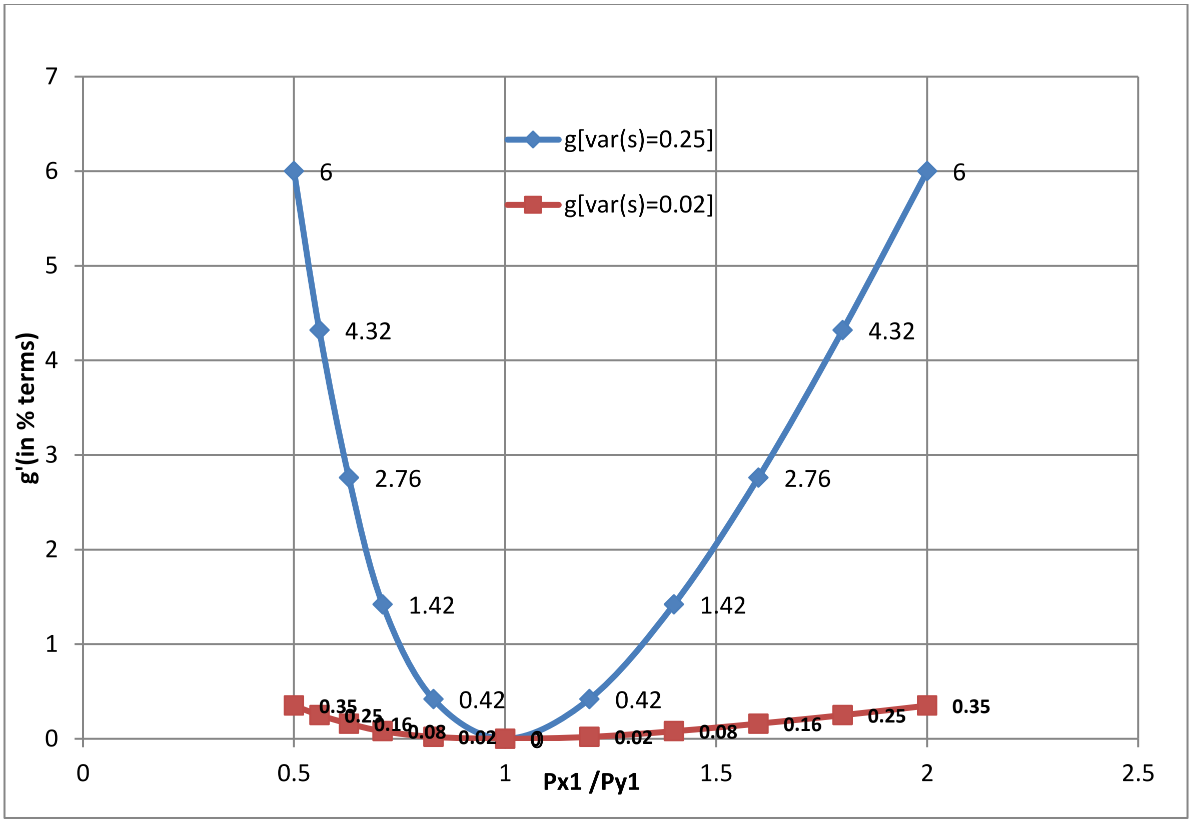
Note:- Drawn from National Sample Survey consumer Expenditure survey data (2004-05)

Figure 2.4:- Kernel Density Estimate of Food Budget Share (U.S)



Note:- Drawn from US consumer expenditure survey data(Second Quarter of 2014)

Figure 2.5:- Impact of Heterogeneity



Note:- Author's own calculation. The gap between the average of the cost of living indices and cost of living index at the average budget share is evaluated for different values of the relative price change. This graph is drawn for two different values of the variance of food budget share (the maximum variance which equals 0.25 and the variance of 0.02 which is calculated from the National Sample Survey consumer expenditure survey data (2004-05).

Table 2.1:- An Example

Household	1	2	3	4	5
α	0.1	0.2	0.5	0.8	0.9
Cost of Living Index	123	125	134	143	147
Democratic Cost of Living index	134.4	134.4	134.4	134.4	134.4
Income	100	200	300	400	500
Plutocratic Cost of Living Index	138.8	138.8	138.8	138.8	138.8

Note:- Author's own calculation

Table 2.2: Example with Greater Increase in Relative Price

Household	1	2	3	4	5
α	0.1	0.2	0.5	0.8	0.9
Cost of Living Index	126	133	155	181	190
Democratic Cost of Living Index	157	157	157	157	157
Income	100	200	300	400	500
Plutocratic Cost of Living Index	168.7	168.7	168.7	168.7	168.7

Note:- Author's own calculation

Table 2.3: Mean and Variability of Food Budget Share in India, 2004-05

Figures	Mean	S.D	Mean+2 S.D	Mean-2 S.D	CV
Rural	0.6	0.12	0.84	0.36	0.2
Urban	0.5	0.134	0.768	0.232	0.27

Note:- Author's calculation using the National Sample Survey consumer expenditure survey data (2004-05). S.D denotes the standard deviation and C.V stands for the coefficient of variation.

Table 2.4: Mean and Variability of Food Budget Share in United States, Second Quarter of 2014

Mean	S.D	Mean+2 S.D	Mean-2 S.D	CV
0.19	0.1	0.39	-0.01	0.53

Note:- Author's computations using the consumer expenditure survey data (second quarter of 2014) of the United States Bureau of Labor Statistics. S.D denotes the standard deviation and C.V stands for the coefficient of variation.

Table 2.5: The Impact of Maximum Heterogeneity

$\frac{P_x^1}{P_y^1}$	$\frac{P_x^0}{P_y^0}$	g(in %)
2	1	6
1.8	1	4.32
1.6	1	2.76
1.4	1	1.42
1.2	1	0.42
1	1	0
0.83	1	0.42
0.71	1	1.42
0.63	1	2.76
0.56	1	4.32
0.5	1	6

Note:- Author's own calculation. The gap between the average of the cost of living indices and cost of living index at the average budget share is evaluated for different values of the relative price change. This table is constructed at the maximum variance of food budget share which is equal to 0.25.

Table 2.6: The impact of Heterogeneity in Food Budget Shares in India

$\frac{P_x^1}{P_y^1}$	$\frac{P_x^0}{P_y^0}$	g(in %)
2	1	0.48
1.8	1	0.35
1.6	1	0.22
1.4	1	0.11
1.2	1	0.03
1	1	0
0.83	1	0.03
0.71	1	0.11
0.63	1	0.22
0.56	1	0.35
0.5	1	0.48

Note:- Author's own calculation. The gap between the average of the cost of living indices and cost of living index at the average budget share is evaluated for different values of the relative price change. This table is constructed at the variance of food budget share equals 0.02(calculated from the National Sample Survey consumer expenditure survey data (2004-05)).

Chapter 3

Border Prices, Pass-Through and Welfare: Palm Oil in India

3.1 Introduction

This chapter examines, the spatial impacts on prices and on wages, of India's trade liberalization in edible oils that began in the early 1990s. Starting from near-autarkic policies that prohibited import of either edible oils or oilseeds, restrictions were relaxed and tariffs reduced on edible oil imports. At the time of opening up the sector, oilseeds were grown on 13% of India's cultivable land and were next in importance only to the cereals of rice and wheat. India is now the world's largest importer of edible oils and imports account for 70% of domestic consumption.

While the impacts of liberalization in such a major sector are of interest in itself, the chapter is also a contribution to the literature on spatial impacts of trade liberalization. While this chapter follows the literature in addressing the trade (or transport) costs that prevent the full pass-through of changes in border prices, the analysis breaks new ground in also identifying the spatial variation in competitive structures that also matter to the pass-through effects.

In recent years, researchers have drawn attention to the possibility that the effects of trade liberalization and openness may vary within a country. For instance, transport costs may mean that changes in tariffs, world prices and exchange rates are not transmitted in the same manner to domestic prices in the hinterland relative to the ports (Atkin and Donaldson, 2015; Marchand, 2012; Nicita, 2009). Another strand of the literature has argued that trade exposure of a particular region depends on its employment composition and hence examines the differential impact of trade liberalization on local (regional) labour markets (Edmonds, Pavcnik and Topalova, 2010; Gaddis and Pieters, 2016; Goldberg and Pavcnik, 2007; Topalova, 2007; Topalova, 2010; Hasan, Mitra and Ural, 2007).

The literature has recognized that the pass-through of border prices depends on transport costs as well as the competitiveness of local markets. As Nicita (2009) pointed out "…consumer prices at the local level depend not only on border prices and tariffs (which are uniform across local markets), but also on local producer prices and transport costs (which are heterogeneous across local markets)." Similarly, Atkin and Donaldson (2015), show that price gaps between locations reflect not just transport costs but also mark-ups that vary across space. In principle, mark-ups are determined by the marginal costs of firms, demand conditions and the competitive environment (Atkin and Donaldson, 2015).

The focus in this chapter on just one sector (edible oils) enables the analysis here to offer a nuanced analysis on the transmission of border prices (the cumulative outcome of world prices, tariffs and exchange rates) on domestic prices and wages. The key departure from the literature and the principal contribution of this chapter is that it exploits prior information about the domestic availability of substitutes to examine how that affects the spatial transmission of border prices to domestic prices and wages.

The plan of the chapter is as follows. The next section discusses the episode of trade liberalization in edible oils in India. The section also discusses associated changes in the consumption of imported oils, its domestic substitutes and also the changes in land resources devoted to oilseeds – the primary input of domestic edible oils. The implications of these changes for spatial impacts and the principal questions for this chapter are posed in Section 3.3. This is followed in Section 3.4 by a review of how our analysis relates to the relevant literature. Section 3.5 shows that even a simple theoretical model can be ambiguous regarding its predictions about spatial impacts. The results depend on own elasticities of demand and the substitutability between imported and local oil. For sufficiently limited substitutability, it will be true that the pass-through effects (on the average consumer price for all edible oils) will be higher at the ports (where imported oil costs the least) relative to the hinterland and will also be higher at the regions that specialize in oilseeds (where the local oil costs the least). We explain the empirical strategy, variables and their data sources in section 3.6. Section 3.7 shows the results from econometric estimation. The implication of our findings on the welfare of workers and consumers is discussed in section 3.8. In Section 3.9, we write down the concluding remarks.

3.2 India's Liberalization of Trade in Edible Oils

In the early 1990s, India initiated a program of trade liberalization. Tariffs were substantially reduced and the proportion of manufacturing products subject to non-tariff barriers steadily declined. Agriculture, was, however, left out of the opening up process. The impetus for agricultural trade liberalization came with the completion of the Uruguay Round of multi-lateral trade negotiations in 1994. In the same year, India lifted quantitative restrictions on imports of edible oils.

Before 1994, all imports of edible oils were on government account as private trade was banned. Imports were contracted whenever domestic supplies fell short. The official policy at this time was self-reliance and government programs were launched to increase productivity and production of oilseeds. India, however, continued to have one of the lowest oilseed yields in the world. Oilseeds could not also compete for the best irrigated lands because of price support for the competing crops (rice, wheat).[1] In 1994, the government reversed policy and allowed free imports on private account subject to tariffs. Between 1990 (the pre-liberalization period) and 2001 (when quantitative restrictions were removed on all agricultural imports), the proportion of edible oils in agricultural imports increased from 18% to 42%. As a result, edible oils accounted for 60% of the increase in agricultural imports during this time.

The movement in tariffs on refined palm oil and soya oil (the two principal imported oils) is shown in Figure 3.1. Tariffs on crude oils are shown in Figure 3.2. Soya oil tariffs remain stable through the period while palm oil tariffs show considerable fluctuation. Gulati and Narayanan (2002) argue that tariffs spiked during the period 1999 to 2004 because of a crash in world prices. Tariffs on crude oils are usually lower in order to encourage domestic refining. Although the tariff rate of soya oil is typically lower than on palm oil, it is the latter that is mostly imported. Palm oil is cheaper than soya oil (see Table 3.1). In addition, the major palm oil exporting countries (Indonesia and Malaysia) are relatively closer than the major soya oil exporting countries (South America and the US). Figure 3.3 shows that the import of palm oil has increased at a much higher rate compared to the other imported oils.

The dramatic growth in the importance of palm oil can also been seen in the market share of different oils (Figure 3.4 and 3.5). In the 1970s, palm oil and soya oil were unimportant. The traditional oils of groundnut, mustard and cotton dominated the market. By the end of the

[1] See Dohlman, Persaud and Landes (2003) and Gulati and Narayanan (2002) for an account of oilseeds and edible oils policies prior to 1994 and for the period from 1994 to early 2000s.

century, palm oil was the leading oil followed by soya oil. The gap over the traditional oils increases by 2014/15.

The nationally representative consumer expenditure survey (from the National Sample Survey Organization or NSSO) does not collect palm oil or soya oil consumption figures. The survey reports the consumption of groundnut, mustard and coconut oils and all the other oils are lumped together in the ‘other edible oil’ category. The budget share of ‘other edible oil’ in total edible oil consumption has increased from 10% to 37% for all India between 1983-84 and 2009-10. The share of imported oils here is lower than indicated by aggregate consumption figures. Some part of the divergence is due to the fact that the NSSO survey only captures household consumption while aggregate consumption also includes the use of the oil in processed foods and restaurants.

Although India allows the imports of oilseeds (non-genetically modified), such imports have not been important for several reasons. Tariffs on oilseeds have generally been higher than on oils and their imports are also governed by phytosanitary regulations. The ban on genetically modified seeds also rules out the import of soybeans since it is predominantly genetically modified in the major exporting countries. Oilseeds producers therefore have not faced direct competition from foreign producers. However, the import of edible oils could have depressed the prices of domestically produced substitutes and thereby affected the demand and prices of domestic oilseeds.

In 1993/94, the three major oilseeds by area were groundnut, rapeseed-mustard and cotton and 26 million hectares (or 14% of total area) grew oilseeds. In 2010/11, 29 million hectares grew oilseeds. The increase, however, hides the fact that the area under the traditional oilseeds of groundnut and rapeseed-mustard declined by 30% and 12%. The aggregate area under oilseeds rose because of cotton and soybeans. In both cases, the returns to their cultivation are not derived solely from oil extraction. Cotton area expanded in the 2000s because the technology of Bt Cotton was profitable to growers (James, 2015). Despite competition from imported oils, soyabeans enjoyed robust demand because of domestic and overseas demand for soymeal feed (Landes, Persaud and Dyck, 2004).

The increased consumption of imported oils and palm oils, in particular, suggests that consumers benefitted from trade liberalization. Despite, the possibly downward pressure on prices, the effect on producers is not so clear because while traditional oilseeds stagnated or declined, producers may have had the opportunity to switch to the more dynamic sectors of cotton and soybean. The effect on wages is even less clear since labor could move to the faster growing sectors in agriculture and outside agriculture.

While edible oils imports were liberalized in the early 1990s (along with manufacturing imports), India maintained its quantitative restrictions on other agricultural imports citing balance of payment difficulties (Gulati and Narayanan, 2002). In a dispute, the WTO ruled against India's position and India lifted its quantitative restrictions on other agricultural imports in 2001. At the same time, India increased its applied tariff on several commodities because it had the benefit of high bound rates (Hoda and Gulati, 2013). Nonetheless, the tariffication of India's agricultural trade constituted a systematic trade reform and opened up the economy to agricultural imports.

3.3 The Spatial Impact of Palm Oil Border Price: The Question

Indonesia and Malaysia are the major suppliers of palm oil which is shipped to ports on the eastern and western coast of India. Table 3.2 lists the ports which account for most palm oil shipments. While in some instances, the ports are right next to cities (Chennai, Kolkata, Mumbai), much of India is populated inland to which the palm oil is transported by road or rail. The question is whether changes in border prices at the ports are felt equally inland. In this chapter, we address this issue in two different ways.

In the first specification, we compare the pass-through of border prices (on domestic prices and wages) between districts that belong to coastal states to districts that belong to non-coastal states (that do not have a coastal line). In a second specification, we use a continuous distance measure computed as the distance (in kilometers) between each district capital and nearest sea-port using the Haversine formula[2]. The Haversine formula measures the great-circle distance between any two points on a sphere by using the latitudes and longitudes of these two points. The great-circle distance is the shortest distance between any two points on the surface of a sphere (as opposed to the straight line through the sphere's interior). Atkin and Donaldson (2015) also use great circle distance in their paper to measure distance between the origin of import/production and destination. In a regression of domestic prices on border prices and other controls, the continuous distance measure is interacted with border prices.

However, this is not the only spatial impact that the chapter considers. While India's production of palm oil is negligible, palm oil imports face competition from domestically produced edible oils from groundnut, soybeans, and rapeseed-mustard among others.[3] The

[2] More detail about Haversine formula can be found in Robusto(1957), Blummelen (2013)
[3] Others include sesame, sunflower, safflower and castor.

competitive pressures may vary spatially depending on availability of local oils and on local preferences. India's traditional pattern of edible oil consumption is strongly linked to regional cuisines. Atkin (2013) has argued that regional differences in tastes are formed by variation in what foods are locally abundant. These arguments suggest that the availability of locally produced oils would (directly and indirectly through taste formation) increase the competitive pressure on the pricing of palm oils.

Spatially disaggregated data on locally produced oils is, however, not available. As a proxy we use spatially disaggregated data on the production of oilseeds. This is likely to be a good measure of variation in locally produced oils because most oil processing plants are located near the production regions. As a measure to protect employment, India restricted processing of traditional oilseeds (groundnut, rapeseed-mustard, sesame and safflower) to small production units. As a result, processing units comprising crushers and expellers are small-scale and located close to production areas (Dohlman, Persaud and Landes, 2003). Some oilseeds (e.g., soybean and cotton) are exempt from these restrictions but even here the solvent extraction plants and refineries tend to be located in the production regions (Madhya Pradesh and Maharashtra for soybeans and Andhra Pradesh, Maharashtra and Gujarat in the case of cotton). It is only the refineries that process imported crude palm oil that tend to locate themselves at the ports.

Figure 3.6 displays the district level map of India shaded according to the proportion of cultivated area under oilseeds. The regression analysis in this chapter considers whether the impact of border prices is different between the top oilseeds producing districts and the others. The top oilseed producing districts are defined to be those in the top quartile of the distribution of area shares.

It should be noted that our data does not allow us to identify the domestic palm oil price. The consumer expenditure survey data allows us to measure a domestic edible oil unit value (for districts which is the unit of analysis), which is the ratio of expenditures on all edible oils divided by the quantity of purchase. This is equivalent to the weighted average of prices of the imported and local oils, where the weights are the proportions in the edible oil consumption basket. The question of our analysis is to identify the impact of changes in palm oil border prices (whether due to tariffs, world prices or exchange rates) on the domestic edible oil unit value. The latter can change because of changes in the local prices of palm oil and other prices. It can also change because of changes in the proportion allocated to different oils. For evaluating consumer welfare, the weighted price of all edible oils is what is relevant.

3.4 Relation to Literature

In their survey of the voluminous pass-through literature, Goldberg and Knetter (1997) put down what they call the generic regression model of exchange rates and prices.

(1) $$p_t = \alpha + \delta X_t + \gamma E_t + \varphi Z_t + \varepsilon_t$$

where all variables are in logs, *p* is a price for a particular product, *X* is the primary variable of interest, *E* is the exchange rate, *Z* denotes other control variables and ε is an error term. The coefficient δ is the pass-through elasticity. In a study of the effect of trade liberalization in Mexico, Nicita (2009) adopts the framework of (1) and investigates the impact of world prices (in local currency), tariffs and trade costs (measured as distance from main port of entry) on consumer prices. Tariff is also interacted with distance to see how trade costs affect the pass-through elasticity. Nicita also includes as a control variable local producer prices to control for the competition from local substitutes.

Our regression specification is similar to that in Nicita. Besides using a distance measure, we also use an alternative specification where trade costs are proxied by a dummy variable for whether the district belongs to a coastal state. The more substantive difference is regarding local substitutes. Producer prices are endogenous. Our measure is the proportion of district area under oilseeds before the liberalization and that is plausibly more exogenous. Importantly, we also interact the border prices with the local substitute measure to investigate the effect of competition on pass-through.

Marchand (2012) also uses a specification similar to (1) to analyse the consequences of trade liberalization in India. In this paper, the pass-through from tariffs to domestic prices is estimated separately for urban and rural areas. The paper does not therefore explicitly model the effect of distance from ports or the effect of competition from local substitutes. (1) is estimated using a state-level panel data for two time points (1994 and 2000) and for 11 goods (including six food commodities but not edible oils, one industrial good, two beverages, energy and tobacco). This chapter is restricted to edible oils but uses a district-level panel data and includes five time points (1993/1994, 1999/2000, 2004/5, 2007/8 and 2011/12). Unlike this chapter, neither Nicita nor Marchand consider spatial effects in their wage regression.

The paper by Atkin and Donaldson (2015) is concerned with estimating trade costs from the data on spatial price gaps. The complication is that trade costs as well as mark-ups may vary with distance. The concern of this chapter is not with this inference problem. Instead, our goal is to see whether it is true that pass-through (and hence local mark-ups) varies with competitive pressure (which in turn may vary spatially) after controlling for distance.

Another strand of literature, mentioned in the introduction, has been concerned with the impact of trade liberalization on poverty, wages and outcomes such as schooling and child labour. The idea in this literature has been to identify the impact of tariffs by comparing industries or regions that are differentially exposed to trade (Goldberg and Pavcnik, 2007). The underlying premise is that labour is not mobile in the short run and so each region (or industry) can be regarded as a local labour market. In this literature, the spatial variation in the effects of trade is central to identification but is not of interest in itself.

Topalova (2007 and 2010) pioneered this approach and applied it to study the effects of India's trade liberalization. Topalova constructed a district specific measure of trade exposure as the employment weighted average of tariffs over all traded goods. Although tariffs are uniform for all districts, the employment composition is not and hence this measure captures a districts exposure to foreign trade. This is the key independent variable and the analysis seeks to uncover its effects on the dependent variable of interest. Such a measure has also been applied in other contexts.[4]

The employment weighted tariff measure assumes that a district's exposure to tariff change in any one sector is proportional to that sector's employment weight. However, if a tariff change also induces changes in derived demand for non-tradables, then the tariff measure may be misleading. Edible oils is a case in point. In her analysis, Topalova does include edible oils among the traded goods in the trade exposure measure. However, the employment in the edible oils sector is negligible relative to the workforce in the oilseeds sector (a non-traded sector). If a

[4] For the theoretical underpinnings of this approach, see Kovak (2013) who also lists the papers that use this approach.

change in edible oil tariff shifts the demand for local oilseeds, then its impact on wages and poverty could be greater than what would be implied by an employment weighted average tariff.

In this chapter, the wage regression is also a variant of (1) and our primary interest is the spatial variation in pass-through (on wages) across distance and across the availability of substitute oils. Since the tariffs of other commodities have also changed during this period, we construct and deploy an employment weighted average tariff measure for all other commodities as a control variable.

3.5 A Spatial Model of Pass-through

In this section, we consider a simple model of spatial pass-through in the presence of domestic substitutes. Following Greehnut, Hung and Norman (1987), the model is one of competition between two producers like in (of palm oil and the local oil respectively) but within a Hotelling framework.

In the model, the port and the hinterland are represented on a line. We assume that there is a single seller of palm oil who receives supplies at the port at a border price of c_1. The border price is the product of the world price, the ad-valorem tariff and the exchange rate. There is also a single seller of the locally produced edible oil. The production is located in the hinterland. The diagram below is a representation of the line model. In the line AB, the palm oil seller is located at point A (the port) and the local edible oil seller produces oil at point B. The distance of the line segment is d.

Diagram 3.1

A (Port) Total Distance=*d* *B*

Diagram 3.1. The Line Market AB

Consumers are uniformly distributed over the line. The local oil and palm oil are imperfect substitutes and the monopoly sellers of palm oil and the local edible oil compete in terms of setting their prices. In order to sell one unit of palm oil at a distance *r* from A, the seller incurs a transport cost of *tr* (where *t* is the transport cost of covering each unit of distance). Therefore cost of selling one unit of palm oil to the buyers located at distance *r* is $c_1 + tr$.

Assume the cost of producing per unit of local edible oil for the monopoly producer located at point *B* is c_2. This producer needs to bear a transport cost of $t\,(d\text{-}r)$ to sell to the buyers located at distance *r* from point *A.* Therefore the cost of producing one unit of edible oil and selling it to a buyer at location *r* is $c_2 + t\,(d\text{-}r)$.

In notation, the seller of palm oil is denoted as firm 1 and the producer of local edible oil as firm 2. Similarly, palm oil and the local oil are subscripted by 1 and 2 respectively. Demand functions for the two oils are

(2) $$q_1 = \alpha_1 - \beta_{11} P_1 + \beta_{12} P_2$$

(3) $$q_2 = \alpha_2 + \beta_{21} P_1 - \beta_{22} P_2$$

where P_1 and P_2 are the respective prices. These are linear demand functions with the following parameter restrictions:

$$\alpha_i > 0 \forall i = 1,2 \text{ and } \beta_{ij} > 0 \forall i, j = 1,2$$

Using demand functions (2) and (3), we can write down the profit functions for firm1 and firm 2 for sales at location *r* (i.e. distance *r* from port) as

$$\pi_1(r) = (P_1 - c_1 - tr)(\alpha_1 - \beta_{11}P_1 + \beta_{12}P_2)$$

$$\pi_2(r) = (P_2 - c_2 - t(d - r))(\alpha_2 + \beta_{21}P_1 - \beta_{22}P_2)$$

The first order conditions for the profit maximization are

$$\frac{\partial \pi_1(r)}{\partial P_1} = 0 \text{ and } \frac{\partial \pi_2(r)}{\partial P_2} = 0$$

Solving for these first order conditions, we can get the optimal prices charged by firm1 at location *r*. Optimal price for firm 1 is

(4) $$P_1(r) = \frac{2\alpha_1\beta_{22} + \alpha_2\beta_{12} + 2c_1\beta_{11}\beta_{22} + c_2\beta_{12}\beta_{22} + \beta_{12}\beta_{22}td + \beta_{22}(2\beta_{11} - \beta_{12})tr}{4\beta_{11}\beta_{22} - \beta_{12}\beta_{21}}$$

The denominator needs to be positive for the price to be positive. Now differentiating expression (4) with respect to distance (i.e. *r*) we get,

(5) $$\frac{\partial P_1(r)}{\partial r} = \frac{\beta_{22}(2\beta_{11} - \beta_{12})t}{4\beta_{11}\beta_{22} - \beta_{12}\beta_{21}}$$

Expression (5) is positive provided $2\beta_{11} - \beta_{12} > 0$ i.e. twice the own price effect of palm oil is higher than the cross price effect of palm oil with respect to local edible oil. As own price effects are usually greater, this condition is likely to be satisfied. Therefore, in the typical circumstance, the palm oil price increases in the distance from port.

The pass-through elasticity of palm oil price with respect to its border price is given by

(6) $$\frac{\partial \ln P_1(r)}{\partial \ln \mathrm{c}_1} = \left(\frac{\partial P_1(r)}{\partial \mathrm{c}_1}\right)\left(\frac{\mathrm{c}_1}{P_1(r)}\right) = \frac{2\beta_{11}\beta_{22}}{4\beta_{11}\beta_{22} - \beta_{12}\beta_{21}} \cdot \left(\frac{\mathrm{c}_1}{P_1(r)}\right)$$

As distance increases, the change in the pass-through elasticity for palm oil is given by

$$\frac{\partial}{\partial r}(\frac{\partial \ln P_1(r)}{\partial \ln \mathrm{c}_1}) = -\frac{2\beta_{11}\beta_{22}}{4\beta_{11}\beta_{22} - \beta_{12}\beta_{21}}.(\frac{\mathrm{c}_1}{(P_1(r))^2})(\frac{\partial P_1(r)}{\partial \mathrm{r}})$$

Therefore the pass-through elasticity of palm oil decreases with distance as the term $\frac{\partial P_1(r)}{\partial \mathrm{r}} > 0$. As seen in (5) earlier, to cover transport costs, the price charged by the palm oil seller is higher in the hinterland relative to port. But this means that the elasticity of demand facing the palm oil seller is increasing from the port to the inland. For this reason, the pass-through elasticity of palm-oil with respect to the border price falls with distance.

The optimal price charged by the local edible oil producer (at distance *r* from point *A*) is

(7) $$P_2(r) = \frac{2\alpha_2\beta_{11} + \alpha_1\beta_{21} + 2c_2\beta_{11}\beta_{22} + c_1\beta_{11}\beta_{21} + 2\beta_{11}\beta_{22}td + \beta_{11}(\beta_{21} - 2\beta_{22})tr}{4\beta_{11}\beta_{22} - \beta_{12}\beta_{21}}$$

Differentiating $P_2(r)$ with respect to *r*, we get

$$\frac{\partial P_2(r)}{\partial r} = \frac{\beta_{11}(\beta_{21} - 2\beta_{22})t}{4\beta_{11}\beta_{22} - \beta_{12}\beta_{21}}$$

The sign of the term $\frac{\partial P_2(r)}{\partial r}$ turns out to be negative because we expect $\beta_{21} - 2\beta_{22}$ to be negative i.e. the cross price effect of the local edible oil with respect to palm oil is likely to be less than two times own price effect of the local edible oil. Therefore, in the typical circumstance, the local oil price is highest at the port and declines as the distance from port increases.

The pass-through elasticity of the local edible oil with respect to palm oil border price is

$$\frac{\partial \ln P_2(r)}{\partial \ln \mathrm{c}_1} = (\frac{\partial P_2(r)}{\partial \mathrm{c}_1})(\frac{\mathrm{c}_1}{P_2(r)}) = \frac{\beta_{11}\beta_{21}}{4\beta_{11}\beta_{22} - \beta_{12}\beta_{21}}.(\frac{\mathrm{c}_1}{P_2(r)})$$

The pass-through elasticity of the local price is also positive because it is a substitute of the imported oil. The change in the pass-through rate for local edible oil with distance *r* is measured as

(8) $$\frac{\partial}{\partial r}\left(\frac{\partial \ln P_2(r)}{\partial \ln c_1}\right) = -\frac{\beta_{11}\beta_{21}}{4\beta_{11}\beta_{22}-\beta_{12}\beta_{21}}.\left(\frac{c_1}{(P_2(r))^2}\right)\left(\frac{\partial P_2(r)}{\partial r}\right)$$

The sign of expression (8) is positive because $\frac{\partial P_2(r)}{\partial r} < 0.$ This means the pass-through elasticity of the local oil with respect to the border price of palm oil increases as one moves toward hinterlands from port and the pass-through diminishes in the reverse direction. This is the exact reverse of the pass-through of palm oil price. But it happens for similar reasons. The price of the local oil is higher and therefore the demand for the local oil is more elastic as one moves towards the port from the hinterland.

As mentioned earlier, we consider, in the empirical analysis, the pass-through of the average edible oil price, i.e., a weighted average of the prices of all edible oils where the weights are the proportions in the edible oil consumption basket. Consumer welfare analysis requires us to measure the pass-through in the average edible oil price rather than on individual prices.

In our model, the counterpart is a weighted average of palm oil price and the price of the local oil. We can write it down as (in logarithmic form)

$$\ln P(r) = w_1(r)\ln P_1(r) + w_2(r)\ln P_2(r)$$

$w_1(r)$ and $w_2(r)$ are the budget share for the palm oil and local edible oil respectively,

$0 \le w_i(r) \le 1; i = 1,2$ and $w_1(r) + w_2(r) = 1.$

The pass-through elasticity of this average edible oil price with respect to palm oil world price or tariff rate of palm oil turns out to be

$$\frac{\partial \ln P(r)}{\partial \ln c_1} = \frac{\partial}{\partial \ln c_1}[w_1(r)\ln P_1(r) + w_2(r)\ln P_2(r)]$$
$$= \frac{\partial}{\partial \ln c_1}[w_1(r)\ln P_1(r) + (1-w_1(r))\ln P_2(r)]$$
$$= \frac{\partial}{\partial \ln c_1}[w_1(r)\ln P_1(r) + (1-w_1(r))\ln P_2(r)]$$
$$= \frac{\partial}{\partial \ln c_1}[w_1(r)(\ln P_1(r) - \ln P_2(r)) + \ln P_2(r)]$$
$$= w_1(r)(\frac{\partial \ln P_1(r)}{\partial \ln c_1} - \frac{\partial \ln P_2(r)}{\partial \ln c_1}) + (\ln P_1(r) - \ln P_2(r))\frac{\partial w_1(r)}{\partial \ln c_1} + \frac{\partial \ln P_2(r)}{\partial \ln c_1}$$

The above expression does not permit unambiguous theoretical predictions. Unlike the pass-through of either the imported or local oil, the pass-through elasticity of the average edible oil price need not be monotonic in distance or the cost of the local oil. To see this, consider the case where the oils are perfect substitutes. For the following restriction on the demand functions, we get the case of perfect substitutes.

$$\alpha_1 = \alpha_2 = \alpha$$
$$\beta_{12} = \beta_{21} = 1$$
$$\beta_{11} = \beta_{22} = 1$$

And the demand functions become

$$q_1 = \alpha - P_1 + P_2$$

$$q_2 = \alpha - P_2 + P_1$$

For positive consumption of both the commodities we need to assume that $(\alpha - P_i + P_j > 0; i, j = 1,2)$. If price of palm oil is less than the local edible oil i.e. $P_1 < P_2$, buyers will consume palm oil only. The reverse is true if the price of local edible oil is less than the palm oil (i.e. $P_1 > P_2$). Then consuming only local edible oil maximizes utility.

We have shown earlier that the price of palm oil increases monotonically as we move from port to hinterland. On the other hand, price of local edible oil decreases monotonically as we move from port to hinterland. For some configuration of parameters, we would get a case where two graphs cross each other as shown in Diagram 3.2 below.

Diagram 3.2

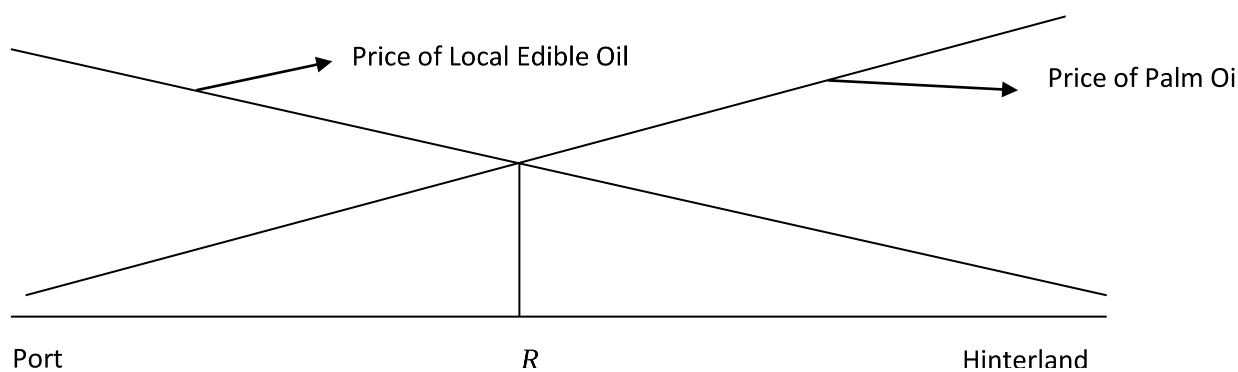


Diagram 3.2. Change in the Price of Palm Oil and Local Edible Oil as Distance from the Port Increases

Till point *R*, only palm oil is consumed. And after point *R*, only the local edible oil is consumed. Therefore, till *R* the pass-through elasticity of the average edible oil price coincides with the pass-through elasticity of the palm oil price, which is monotonically decreasing with the distance. At any distance greater than *R* the pass-through elasticity of the average edible oil price coincides with the pass-through elasticity of the local edible oil price which is monotonically increasing with the distance. At point *R* (i.e. exactly distant *R* from the port), both the prices are equal. We can assume, without loss of generality, that buyers at point *R* consume only palm oil. Therefore the pass-through curve of the average edible oil price will monotonically decline till it reaches point *R* and after that it monotonically increases. As a result we get a U-shaped pass-through curve as we move towards hinterland from port (see Diagram 3.3)

Diagram 3.3

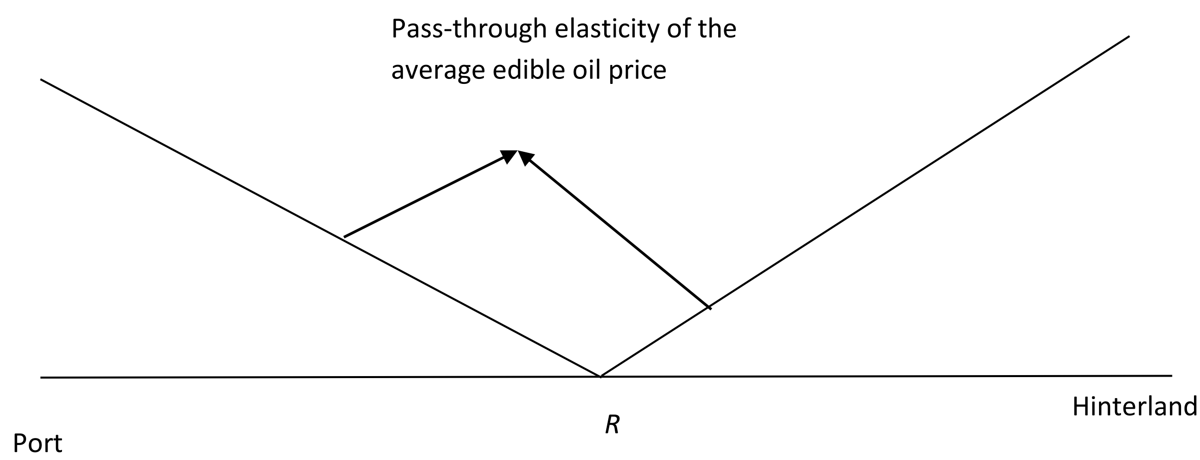


Diagram 3.3. Change in the Pass-through Elasticity of the Average Edible Oil Price as Distance from the Port Increases (Two commodities are Perfect substitutes of Each Other)

On the other hand, other special cases can be constructed where the pass-through elasticity of the average edible oil price is declining monotonically in distance as well as in the production cost of the local oil. If in the previous case, we assumed perfect substitutability, we will now assume zero substitutability.

Suppose the cross price effect of the local edible oil price with respect to palm oil price is zero i.e. $\beta_{21} = 0$

From the pass through elasticity equation, it can be seen that this implies

$$\frac{\partial \ln P_2}{\partial \ln c_1} = 0$$

Using this condition, the pass-through of the average edible oil price boils down to

$$\frac{\partial \ln P}{\partial \ln c_1} = w_1 \frac{\partial \ln P_1}{\partial \ln c_1} + (\ln P_1 - \ln P_2)\frac{\partial w_1}{\partial \ln c_1}$$

If the own price elasticity of demand for palm oil is unity, then it can be shown that the budget share of palm oil will be invariant to its own price and hence the second term vanishes.[5]

Therefore,

$$\frac{\partial \ln P}{\partial \ln c_1} = w_1 \frac{\partial \ln P_1}{\partial \ln c_1}$$

The change in the pass-through rate of the average edible oil price with distance (from the port) under these special conditions (mentioned above) is written as

(9) $$(\partial / \partial r)\frac{\partial \ln P}{\partial \ln c_1} = w_1 (\partial / \partial r)\frac{\partial \ln P_1}{\partial \ln c_1} + (\frac{\partial \ln P_1}{\partial \ln c_1})(\partial w_1 / \partial r)$$

The second term on the right hand side of equation (9) is zero if the palm oil budget share is invariant to distance. This follows from the fact that the palm oil price is unit elastic and if, in addition, we assume that the own price elasticity of local oil is unity.[6] Hence under these special conditions the pass-through elasticity of the average edible oil price with respect to the palm oil border price will be monotonically declining. This is illustrated in Diagram 3.4 below

[5] For a proof, see the appendix 3.A.

[6] For a proof, see the appendix 3.B.

Diagram 3.4

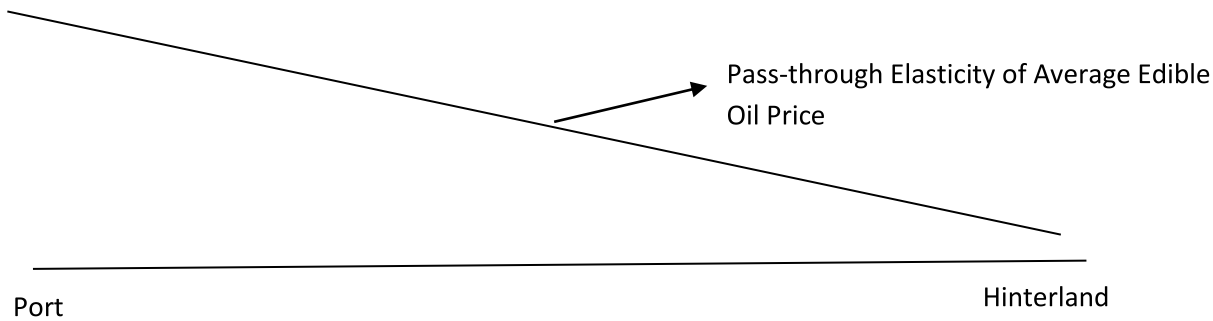


Diagram 3.4. Change in the Pass-through Elasticity of the Average Edible Oil Price as Distance from the Port Increases(Cross-Price Elasticity of Local Edible Oil with Respect to Palm Oil is Zero and the Budget Share of Palm Oil Does Not Respond to the Change in the World Price/Ad-Valorem Tariff Rate of Palm Oil))

We can also use the model to investigate how the pass-through elasticity varies with unit cost of production of the local edible oil (c_2). The interest in this comparative static comes from the fact that high oilseed producing regions are likely to offer lower production costs of local oil than the low oilseed producing regions because they offer easier access to raw materials (principally oilseeds). Intuition would suggest that regions with lower production cost would offer greater competition to imported supplies and hence the pass-through elasticity should be higher in those regions. In other words, we would expect that $\partial(\frac{dlnP}{dln\ c_1})/\partial c_2$ to be negative.

In the appendix, we demonstrate that the intuition is correct for the pass-through elasticity of both the palm oil and the local edible oil as both of these elasticities decreases with per unit cost of production for local edible oil (c_2). The reason is quite straightforward. Prices of both palm oil and local edible oil increases with c_2. Therefore more price elastic/price sensitive consumers are located where c_2 is higher and naturally that reduces the pass-through rate. However, these results do not imply that the pass-through elasticity of average edible oil price also declines with the per-unit production cost for local edible oil increases. Just like in the case of distance, the theoretical predictions are ambiguous but we can obtain insights under some special cases.

Suppose the cross-price effect of local edible oil with respect to palm oil price is zero, i.e., $\beta_{21} = 0$ and the own price effect of palm oil is unity. Then the pass-through elasticity of the average edible oil price boils down to

$$\frac{\partial \ln P}{\partial \ln c_1} = w_1 \frac{\partial \ln P_1}{\partial \ln c_1}$$

The change in the pass-through elasticity of the average edible oil price with respect to per-unit cost of local edible oil production under this special condition is written as

$$(\partial / \partial c_2)\frac{\partial \ln P}{\partial \ln c_1} = w_1 (\partial / \partial c_2)\frac{\partial \ln P_1}{\partial \ln c_1} + (\frac{\partial \ln P_1}{\partial \ln c_1})(\partial w_1 / \partial c_2) \qquad (10)$$

The first term in the right hand side of expression (10) is negative as the pass-through elasticity of the palm oil decreases with the per-unit cost of local edible oil production. The second term can be shown as zero under some additional conditions.[7] Therefore under these conditions, the direction of the change in the pass-through elasticity of the average edible oil price with respect to the per-unit cost of local edible oil production (c_2) coincides with the change in the pass-through elasticity of the palm oil price which is monotonically declining in c_2.

If, on the other hand, the cross price elasticity of palm oil with respect to local edible oil is not zero and is positive, then $\partial w_1 / \partial c_2$ is positive. If the cross price effect is so large such that $\partial w_1 / \partial c_2$ (i.e. the increase in the budget share of palm oil with the increase in c_2) dominates the

[7] This result is also demonstrated in the proof in appendix 3.C.

decline in the pass-through elasticity of palm oil with c_2, then the pass-through elasticity of the average edible oil price may even increase with the per unit cost of local edible oil production.

It can be shown that all analytical results of our theoretical model remain almost unchanged if we assume that the local edible oil producer is located near the port instead of being located in the hinterland. The only difference that would arise now is the fact that the local edible oil price also monotonically increases as we move from port to hinterland. Therefore the pass-through elasticity of local edible oil with respect to border price of palm oil declines monotonically towards hinterland. The conditions under which the average edible oil price declines from port to hinterland or declines from high to low oilseeds producing region remain same.

The effect of the change in palm oil border price may also have corresponding wage effects. Suppose, following the local labour market approach, we consider each district to be a segmented labor market. Then a change in the domestic edible oil price induced by the world price of palm oil or ad-valorem tariff rate affects the oilseeds price because oilseeds are inputs to edible oil production.

$$W_{oilseeds} = P_{oilseeds}{}^{*} MP_L$$

where $W_{oilseeds}$ is the wage rate of oilseeds production and $P_{oilseeds}$ is the price of oilseeds. MP_L stands for the marginal productivity of labor in oilseeds production.

Any change in the edible oils price shifts the marginal productivity curve in oilseed production and would also affect wages. As agricultural labor markets across commodities are integrated, the above chain of reasoning would suggest impacts on agricultural wages generally (and not just in oilseeds production alone). In the early 90s, oilseeds comprised 18% of the total agricultural production and 13% of the total agricultural areas. Therefore, at the time trade liberalization was initiated, the share of oilseeds in total agricultural production was presumably large enough for there to be appreciable effects on wages in all of agriculture. In the longer run, these differential wage effects may be muted by the movement of labour from oilseeds to more dynamic sectors within and outside agriculture.

3.6 Empirical Specification, Variable Definitions and Data Sources

We estimate the following basic specifications for the price and wage regressions:

(11)
$$\ln P_{dt} = \alpha_0 + \alpha_1 \ln c_{1t} + \alpha_2 \ln c_{1t} * coastal\_states$$
$$+ \alpha_3 \ln c_{1t} * high\_oilseeds\_producing\_districts + \beta Z_{dt} + f_d + \gamma_i T * \sum_{i=1}^{K} R_i + \varepsilon_{dt}$$

(12)
$$\ln W_{dt} = \delta_0 + \delta_1 \ln c_{1t} + \delta_2 \ln c_{1t} * coastal\_states$$
$$+ \delta_3 \ln c_{1t} * high\_oilseeds\_producing\_districts + \theta Z_{dt} + f_d + \mu_i T * \sum_{i=1}^{K} R_i + \varepsilon_{dt}$$

W_{dt} stands for real agricultural wage rate (nominal agricultural wage rate deflated by the consumer price index of agricultural laborers) in rural areas for district *d* at time period *t*. P_{dt} denotes domestic edible oil price (deflated by the consumer price index of agricultural laborers) for '*d*'th district at '*t*'th time period. The border price of palm oil at time point *t* (which is nothing but world price multiplied by one plus ad-valorem tariff rate and exchange rate) is denoted as c_{1t}. Both equation (11) and equation (12) are district fixed effect regressions where the district fixed effect is denoted as f_d. $T * R_i$ denotes the local time trend for the '*i*'th region where $R_i = 1$ and $R_j = 0 \forall j \neq i$ and *T* stands for the aggregate time trend. As we care about the local time trend for all regions (say there are K such regions), we take the sum over all these individual local/regional trends and write it as $T * \sum_{i=1}^{K} R_i$. In our empirical model, we use different definitions of regions in alternative specifications. In one specification, a district is considered as a region. But districts are part of larger sub-regions in India. According to the definition of the nationally representative sample survey (NSS) of India, these larger sub-regions are called state-regions/NSS-regions. In an alternative specification of our regression, we consider these state-regions/NSS-regions as regions.

A state-region is a spatial category within a state with similar topography and agro-climatic characteristics. There are 87 such state-regions defined by the nationally representative sample survey (NSS) of India. State-region time trends are example of local time trends. In alternative specification of our regression model, we have used district time trends (instead of state-region time trends) as local time trend. If both the aggregate and local time trends are included in a regression specification, then the coefficients of the local time trends capture the deviation from the aggregate trend. Finally ε_{dt} represents error term in the regression model.

The districts belong to non-coastal states and low oilseeds producing districts are considered as benchmark categories and we investigate whether the effect of the change in the palm oil price on domestic edible oil price and real agricultural wage rate is significantly different for the districts belong to coastal states and high oilseeds producing districts. The vector $\mathbf{Z_{dt}}$ includes a host of control variables at the district level.

We can divide these control variables into two parts: agricultural variables and non agricultural variables. Agricultural variables include total agricultural production and oilseeds production, percentage of villages irrigated in a district and average annual rainfall. The non-agricultural control variables are a composite tariff measure (to control for changes in tariffs for other commodities), a dummy variable for the period without quantitative restrictions (which takes the value one post-2001 and zero otherwise), the interaction of the composite tariff measure with the dummy of quantitative restrictions, the percentage of villages connected by bus and railways (as separate variables), the literacy rate and the share of scheduled caste and scheduled tribe population. In addition the regressions also control for the aggregate and local time trend (as we have already mentioned). To control for convergence, some regressions also include initial per-capita expenditure (i.e. per-capita expenditure corresponding to 1993-94).

Table 3.3 displays the descriptive statistics of the dependent and explanatory variables of regressions (11) and (12). The table also contains variable definitions and data sources. The details are elaborated in appendix 3.D which explains the data sources and the construction of variables.

Chaudhuri and Gupta (2009) questioned the representativeness of the district specific measures constructed from NSS data prior to 2004-05. This is because prior to 2004-05, there was no sub-stratification within a district in the NSS sample design. As a result, the entire district was treated as a strata. The absence of stratification generally increases the sampling variance and that is the reason why Chaudhuri and Gupta question the representativeness of district level estimates. However, this does not invalidate the use of district observations in a regression framework. As long as the sample is randomly chosen (whether without or with stratification), the estimates are unbiased. Unstratified sampling will increase the variance of estimates and therefore their significance. This is a handicap for district-level analysis but does not invalidate it. Indeed, district-level analysis using NSS data is frequently found in the literature such as in Topalova (2007, 2010) and in Duflo and Pande (2007).

We identify the parameters of the regression equations (11) and (12) under the assumption that conditional on district fixed effects, aggregate and local time trend and other control variables, the palm oil border price (the product of world price, exchange rate and the tariff rate) is exogenous to domestic edible oil price and wages. This identification strategy is questionable if India's tariff rates are endogenous to domestic production in which case the correlation between the border price and the domestic price may also reflect a reverse causality from domestic prices to border prices via tariffs. To address this concern, we examine the robustness of the results from the basic specification in regressions where we control for India's production of oilseeds.

The other concern for identification would stem from the possibility that aggregate demand shocks in India may affect world prices because India is a large economy in world consumption. These shocks are controlled by linear time trends in the regression – those which are common to all districts as well as time trends specific to a district.

3.7 Regression Results

The basic price and wage regression results are shown in Table 3.4a and Table 3.4b respectively. In each of these tables, we have three columns. In the first column of both tables, regressions are run using the aggregate time trend and state-region time trend.

In the second column of both tables, we replace state-region time trend by the district time trend. Apart from the district time trend, regressions in third column also have the initial district specific per-capita expenditure as a control variable (it is interacted with time trend to be used in the regression). The initial per-capita expenditure captures the convergence across districts (Topalova, 2010) in terms of per-capita income (if expenditure can be used as a proxy for income).

As shown by the price regression in Table 3.4a, an increase in the border price of palm oil (either from the increase in the world price or ad-valorem tariff rate or both) significantly (1% level) increases the domestic edible oil price in all the specifications. The pass-through elasticity in the low oilseed producing districts of non-coastal states is 0.63 in column 3 which is the preferred specification as it controls for district specific time trends and initial per capita expenditure. If these (low oil producing) districts were located in coastal states, the pass-through elasticity would be higher by 0.12 (significantly different from zero at the 1% level). Similarly, compared to the benchmark of low oil seed producing districts of non-coastal states, the pass-through is significantly higher by 0.15 in the high oil seed producing districts of non-coastal states.

These results show that the pass-through elasticity of domestic edible oil price with respect to palm oil price is significantly less in the hinterland as compared to port (coastal states represent port and non-coastal states represent hinterland). The pass-through elasticity also turns out to be larger for the high oilseeds producing districts relative to the low oilseeds producing districts. As mentioned earlier, the production cost of local oils is presumably lower in the high oilseeds producing districts.

As we saw earlier, the empirical pattern of results would be predicted in the theoretical model in the case when the imported and local oil are imperfect substitutes (apart from assumptions on own elasticity of demand). The empirical results are not consistent with the perfect substitutes case.

Turning to the wage regressions, it can be seen from Table 3.4b, that the pass-through of palm oil border price on wages is also positive (which is consistent with the positive pass-through elasticity of domestic oil price). The pass-through on wages is expectedly smaller. It is 0.34 (for the preferred specification in column 3) for the base category of a low oil producing district in a non-coastal state. But the differential wage impact across spatial categories is not as robust/prominent as compared to the price effect. There is no statistically significant difference in the wage pass-through between the high and low oilseeds producing districts. For the coastal states, wage effect is larger relative to non-coastal states at the 10% level. These results, imply that the wage effects of, say, a decline in the border price, have been more evenly distributed over spatial categories than the price effects. This could be because labour moved from oilseeds to other crops and other sectors.

Robustness

As mentioned earlier, if Indian tariffs adjust to domestic production, then the domestic oilseeds production may affect the palm oil border price via tariffs. To control for such feedback effects, we add the state level oilseeds production. It can be seen that the neither the price not wage regressions change appreciably (column 1 to column 3 in Table 3.5a and Table 3.5b). In another regression, we also add the national oilseeds production besides the state level production. Here too, the results remain robust in both the price and wage regressions (column 4 in Table 3.5a and Table 3.5b).

As a second robustness check, we use a continuous distance measure instead of the dummy variable for the coastal states. The price and wage regressions are run using that distance measure. As a continuous measure of distance, we compute the distance (in kilometers) between the head-quarter of each district and nearest sea-port using the Haversine formula. The

Haversine formula measures the distance between any two points on a sphere by using the latitudes and longitudes of these two points.

The results are shown in Table 3.6a and Table 3.6b. Notice that the price pass-through elasticity is comparable to those in the base specification that uses a coastal dummy. The interaction term between the palm oil price and the distance is significant at 10% level (when we control for the district time trend). The pass-through elasticity declines by 0.03, 0.05 and 0.08 from the port as we cross 300, 600 and 900 kilometers distance from the port respectively. All districts that are 600 km or more distant from the ports would be captured in the non-coastal states category (Haryana, Madhya Pradesh, Punjab, Rajasthan and Uttar Pradesh). Just like the earlier specifications, the pass-through on domestic price of oil turns out to be significantly (at 1%) larger for high oilseeds producing districts relative to low oilseeds producing districts (in Table 3.6a).

Looking at Table 3.6b, we find the results of the wage regressions. The interaction term between palm oil price and distance does not turn out to be significant for the wage regression. Similarly, no significant wage effect is found between the high and low oilseeds producing districts. These results are no different from what was found in the base specifications.

As a final robustness check, we run the price and wage regressions by generating three categories for area share devoted to oilseeds cultivation (instead of two categories that we have used so far). The results are shown in Table 3.7a and Table 3.7b. The three categories are districts belong to bottom 25% of the distribution of area shares devoted to oilseeds production, districts which are in between 25% and 75% of the distribution and the districts that belong to the top 25% in terms of area share devoted to oilseeds cultivation. Everything else is same as the base specification. The bottom 25% is considered as the benchmark category and the incremental price and wage impact is estimated for the other two categories. The results are found to be no different from what was found in two category classification of area shares devoted to oilseeds production. Compared to the benchmark category of districts belong to bottom 25% of area share, the price pass-through elasticity turns out to be significantly larger for the top two categories(in terms of area share in oilseeds cultivation). We do not see any significant differential wage pass-through impact across these three categories although the wage pass-through for the base category of districts belong to non-coastal states and bottom 25% of the distribution of area shares turns out to be statistically significant.

3.8 Spatial Differences in Welfare Change

From all the regression specifications, it is evident that the border price of palm oil has a positive and significant impact on domestic edible oil price and the effect differs across spatial categories. The differential wage effect across the spatial categories is not as significant or robust as compared to the price effect but the overall impact of the palm oil border price on the agricultural wage rate is positive and significant.

Suppose we consider a household as a consumer of edible oil or an agricultural laborer or both. Following Porto (2006), the change in welfare for '*j*'th household (measured by compensating variation as a proportion of initial expenditure) due to change in the relative price of edible oil (price of edible oil deflated by the consumer price index) between period '0' and '1' is written as

(13)

$$CV^j / e^j = [e^j(P_{eoil}{}^1, u^0) - e^j(P_{eoil}{}^0, u^0)] / e^j(P_{eoil}{}^0, u^0)$$
$$= [s^j - \sum_m \theta_m{}^j \varepsilon] d \ln P_{eoil}$$

s^j

e denotes the expenditure function and stands for the budget share of edible oil by '*j*'th household in the initial period (i.e. period '0'). P_{eoil} represents domestic edible oil price (deflated by consumer price index to adjust for inflation). Now

$$d \ln P_{eoil} = (\partial \ln P_{eoil} / \partial \ln c_1) d \ln c_1$$

$\partial \ln P_{eoil} / \partial \ln c_1$ is the pass-through elasticity on the domestic edible oil price where c_1 denotes the border price of palm oil. Elasticity of real agricultural wage rate with respect to domestic edible oil price is denoted as ε and θ_m^j represents the share of labor income (in total income) for '*m*'th member belongs to '*j*'th household. We can decompose the compensating variation into two parts. The term $s^j d \ln P_{eoil}$ is called the consumption effect and the term $\sum_m \theta_m{}^j \varepsilon d \ln P_{eoil}$ is denoted as the labor income/wage effect.

Now we aggregate these household effects to a region *R*. Suppose there are N^R households in region *R*. Using equation (13), we can write down the average compensating variation for region *R* as

$$\begin{aligned} CV^R &= (1/N^R)\sum_{j \in R}[s^j - \sum_m \theta_m{}^j \varepsilon^R] d \ln P_{eoil}{}^R \\ &= (1/N^R)\sum_{j \in R} s^j d \ln P_{eoil}{}^R - (1/N^R)\sum_{j \in R}\sum_m \theta_m{}^j \varepsilon^R d \ln P_{eoil}{}^R \qquad (14) \\ &= s^R d \ln P_{eoil}{}^R - \theta^R \varepsilon^R d \ln P_{eoil}{}^R \end{aligned}$$

where $s^R = (1/N^R)\sum_{j \in R} s^j ; (1/N^R)\sum_{j \in R}\sum_m \theta_m{}^j = \theta^R$

Therefore s^R stands for the average budget share of edible oil in region *R*. $d \ln P_{eoil}{}^R$ denotes the change in the domestic edible oil price in region *R* induced by the change in palm oil border price. The average share of labor income (in total income) in region *R* is represented as

θ^R. Elasticity of real agricultural wage rate with respect to edible oil price is ε^R. Just like the compensating variation for a household, average compensating variation for region *R* in equation (14) can be decomposed into two parts; $s^R d \ln P_{eoil}{}^R$ stands for the average consumption effect and $\theta^R \varepsilon^R d \ln P_{eoil}{}^R$ represents average labor income effect.

When there is a trade induced increase in the domestic edible oil price, the consumers lose but the agricultural laborers gain. Opposite happens when trade liberalization causes a decline in the domestic edible oil price. The negative sign in front of the labor income/wage effect indicates that consumption effect and labor income effect move in the opposite direction.

Using equation (14), we can compare the change in welfare between two regions induced by the change in palm oil border price. As a first comparison between two regions, we choose the coastal and the non-coastal regions. Suppose we focus on the consumption effect and the labor income effect separately.

Consider a hypothetical situation when the average budget share of edible oil is same across coastal and non-coastal region. Then the average consumption effect in the coastal states relative to non-coastal turns out to be

$$\frac{CV^{coastal}|_{consumption}}{CV^{non-coastal}|_{consumption}} = \frac{d \ln P_{eoil}{}^{coastal}}{d \ln P_{eoil}{}^{non-coastal}}$$
$$= [\frac{(\partial \ln P_{eoil} / \partial \ln c_1)^{coastal}}{(\partial \ln P_{eoil} / \partial \ln c_1)^{non-coastal}}.][\frac{d \ln c_1}{d \ln c_1}]$$
$$= \frac{(\partial \ln P_{eoil} / \partial \ln c_1)^{coastal}}{(\partial \ln P_{eoil} / \partial \ln c_1)^{non-coastal}}$$

Therefore under the assumption of equal budget shares, the ratio of the average consumption effect of the coastal region relative to non-coastal region is nothing but the ratio of their pass-through elasticities. From the estimated pass-through elasticities (estimates from the basic specification of the price regression), the ratio turns out to be 1.2 between coastal and non-coastal states.

Now we focus on the wage effect. If we assume that the average share of labor income is same across coastal and non-coastal states, then the average wage effect for the coastal states relative to non-coastal states turns out to be

$$\frac{CV^{coastal}|_{wage}}{CV^{non-coastal}|_{wage}} = \frac{\varepsilon^{coastal} d \ln P_{eoil}{}^{coastal}}{\varepsilon^{non-coastal} d \ln P_{eoil}{}^{non-coastal}}$$

$$= [\frac{(\partial \ln W / \partial \ln P_{eoil})^{coastal} (\partial \ln P_{eoil} / \partial \ln c_1)^{coastal}}{(\partial \ln W / \partial \ln P_{eoil})^{non-coastal} (\partial \ln P_{eoil} / \partial \ln c_1)^{non-coastal}}.][\frac{d \ln c_1}{d \ln c_1}]$$

$$= [\frac{(\partial \ln W / \partial \ln c_1)^{coastal} (\partial \ln c_1 / \partial \ln P_{eoil})^{coastal} (\partial \ln P_{eoil} / \partial \ln c_1)^{coastal}}{(\partial \ln W / \partial \ln c_1)^{non-coastal} (\partial \ln c_1 / \partial \ln P_{eoil})^{non-coastal} (\partial \ln P_{eoil} / \partial \ln c_1)^{non-coastal}}][\frac{d \ln c_1}{d \ln c_1}]$$

$$= \frac{(\partial \ln W / \partial \ln c_1)^{coastal}}{(\partial \ln W / \partial \ln c_1)^{non-coastal}}$$

Therefore the ratio of the average wage effect between coastal and non-coastal states turns out to be the ratio of the wage pass-through elasticity between coastal states and non-coastal states. From the estimated pass-through elasticities (estimates from the basic specification of the wage regression), the ratio turns out to be 1.52 between coastal and non-coastal states.

In a similar way, we also compute the ratio of the average consumption and wage effect between the high and low oilseeds producing districts using the estimated coefficients from the basic price and wage regression. The ratio of the average consumption and wage effect between high and low oilseeds producing districts come out as 1.24 and 1.2 respectively.

Therefore when we compare the coastal and non-coastal states, the average wage effect for the coastal states relative to non-coastal states turns out to be much larger than the average consumption effect for the coastal states relative to non-coastal states. But the relative average wage effect is not much different from the relative average consumption effect between high and low oilseeds producing districts. This can be explained from the fact that the differential wage pass-through rate is not significantly different between high and low oilseeds producing districts.

3.9 Concluding Remarks

This chapter discusses the impact of the trade liberalization of edible oil which is the most important imported agricultural commodity in India. When trade liberalization began in early 1990s, India started importing palm oil because of its cheap price and low transportation cost of importing from nearby locations of Indonesia and Malaysia. This chapter examines the effect of a change in the border price of palm oil on the domestic edible oil price and the agricultural wage rate.

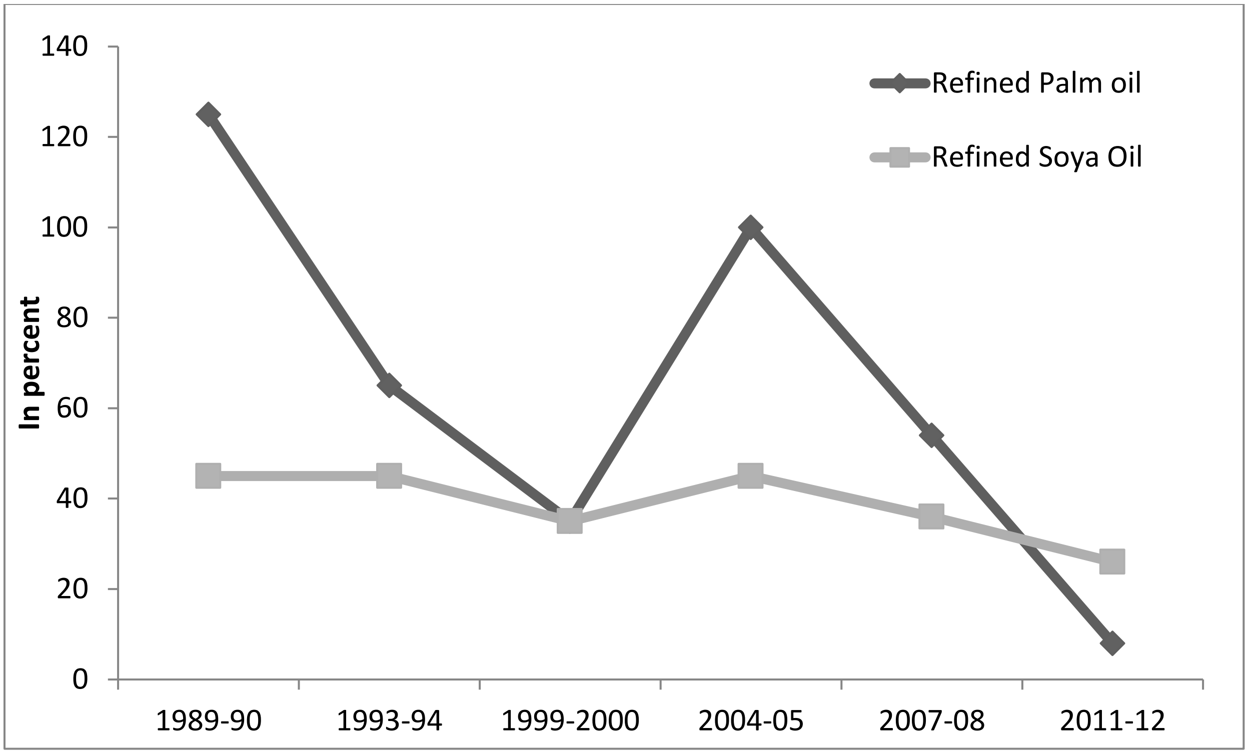
But these impacts vary spatially. We build a theoretical model to show that the pass-through effects of border prices on domestic price and wage rate differs between port and hinterland because of transportation cost. The model also identifies spatial variation in competitive structures that leads to varying pass-through elasticity between high and low oilseeds producing regions. The model predicts that for limited substitutability between the local oil and imported palm oil, the pass-through effect is higher in ports relative to inland. Similarly limited substitutability ensures that pass-through elasticity is stronger in the high oilseeds producing regions relative to low oilseeds producing regions. We do not get the above findings if local edible oil and imported oil are perfect substitutes.

From empirical investigation, we find a positive and statistically significant impact of the border price of palm oil (induced by the change in the world price, tariff rate of palm oil or exchange rate) on domestic edible oil price and real agricultural wage rate in the post trade liberalization period. The spatial difference in the price and wage effect is also found empirically. It turns out that the effect of the change in the border price is significantly larger for the coastal states relative to the non-coastal states and it is also significantly larger for the high oilseeds producing districts compared to low oilseeds producing districts. The spatial difference in the wage impact turns out to be weak compared to the spatial price effect. These findings imply that the wage effect is more evenly distributed over spatial categories than the price effects. As explained earlier, this could be because labour migrated from oilseeds to other crops and other sectors.

The spatially varying price and wage effects have important welfare implications. We find that the average compensating variation (both the consumption and labor income/wage effects of the compensating variation) induced by the change in the palm oil world price or ad-valorem tariff rate vary spatially because of varying pass-through elasticity across regions.

This chapter provides the motivation to investigate the spatial implication of the trade induced price change of other commodities which were important in the import basket. An interesting extension of this analysis is to include the producers of oilseeds in the picture and do a combined analysis with the consumers, producers and wage earners.

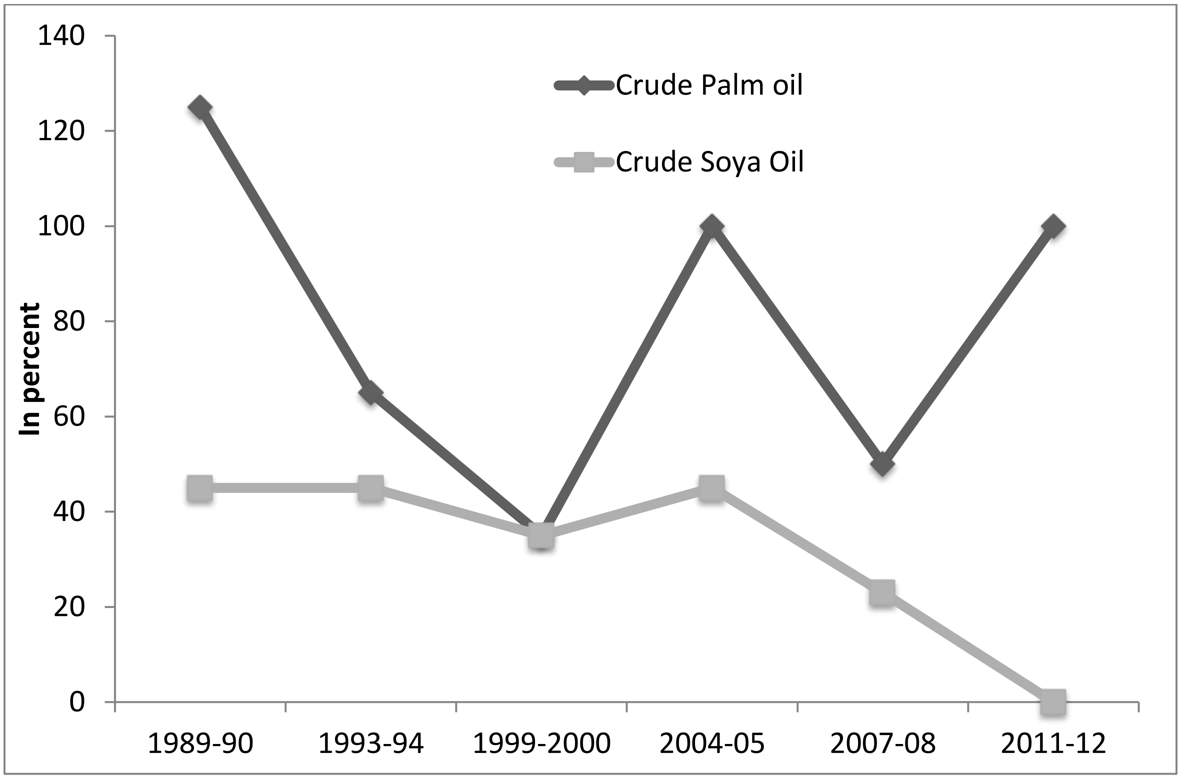
Figure 3.1:- Ad-Valorem Tariff Rate of Refined Palm Oil and Refined Soya Oil



Source:-World Integrated Trade Solution database.

Note: - The vertical axis represents the ad-valorem tariff rate in percentage and the horizontal axis denotes the years.

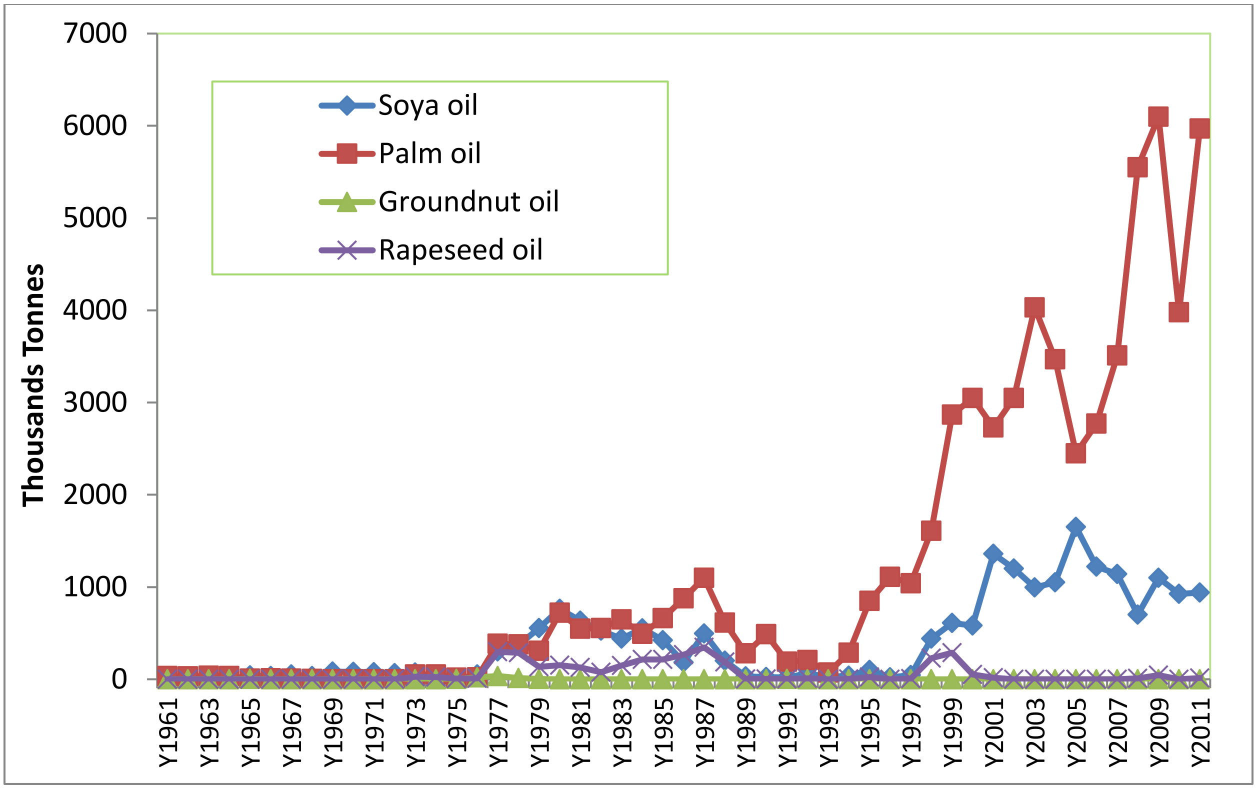
Figure 3.2:- Ad-Valorem Tariff Rate of Crude Palm Oil and Crude Soya Oil



Source:-World Integrated Trade Solution database.

Note: - The vertical axis represents the ad-valorem tariff rate in percentage and the horizontal axis denotes the years

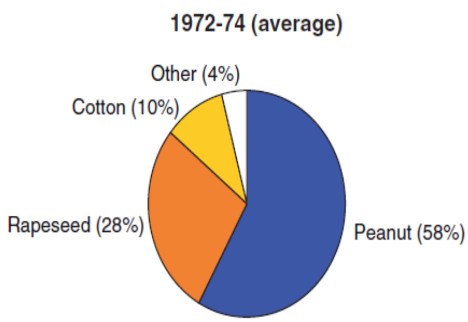
Figure 3.3:-Import of Different Types of Edible Oils in India

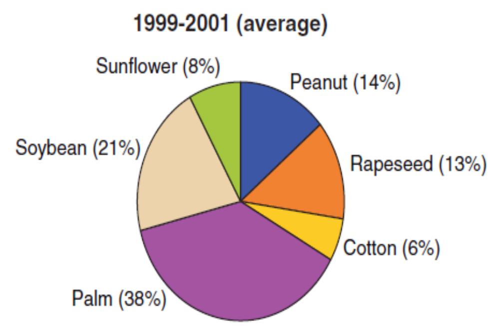


Source: - FAOSTAT

Note: - The vertical axis represents the import of different types of edible oils in thousand tonnes. The horizontal axis stands for the years.

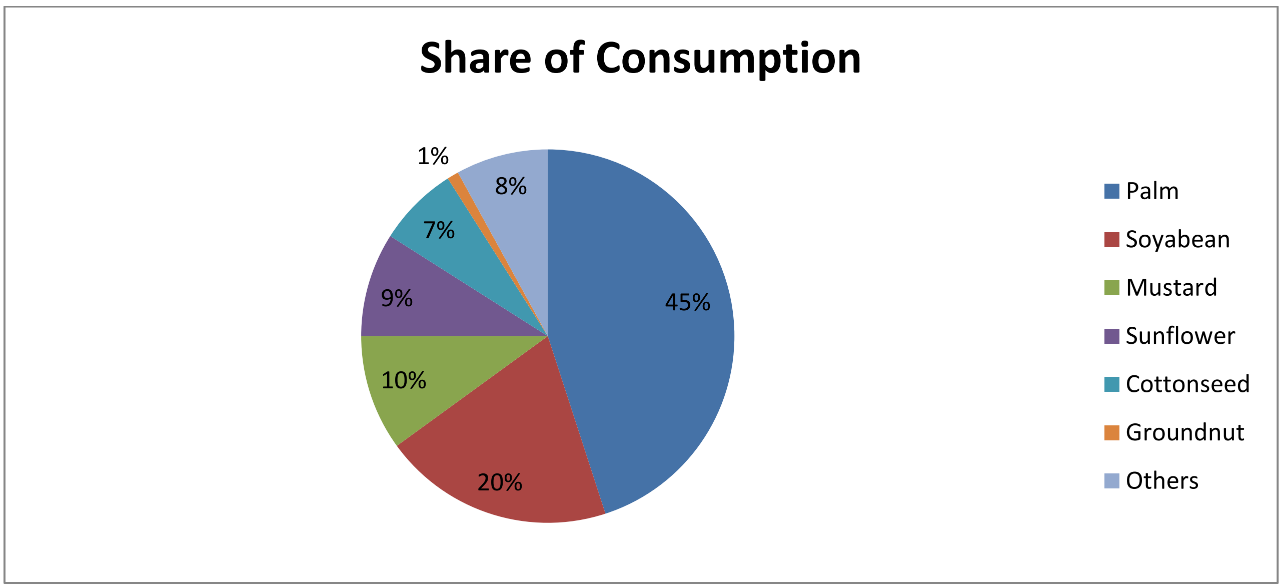
Figure 3.4:-Edible Oil Consumption in India





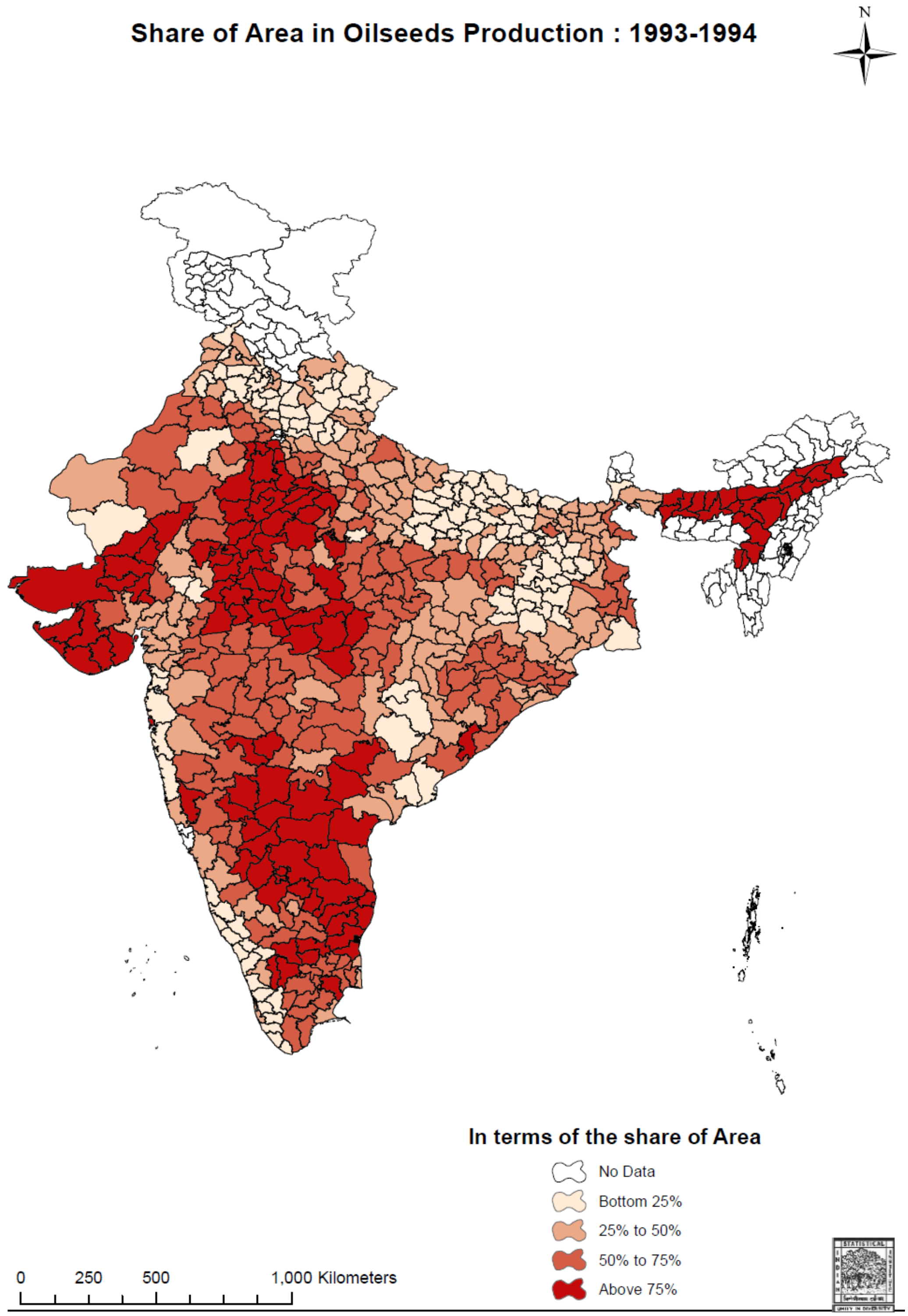
Source: - Production, Supply and Distribution database, USDA

Figure 3.5:-Edible Oil Consumption in India in 2014-15



Source: - GGN Research

Figure 3.6:-Share of Area in Oilseeds Production: 1993-94



Source: - IFPRI Database for Agricultural Area and Production 1993-94(Author's Calculation)

Note:- Ratio of the area devoted to oilseeds production to total agricultural production is plotted in the above figure

Table 3.1:-World Price of Different Types of Edible Oil in India

Year	Groundnut Oil	Palm Oil	Soybean Oil
1993-94	881	453	548
1999-2000	751	373	383
2004-05	1111	447	580
2007-08	1742	864	1070
2011-12	2212	1062	1263

Source: - World Bank Commodity Price Data

Table 3.2:-Ports Which Account for Most Palm Oil Shipments

Palm Oil Importing Ports	States
Kandla	Gujarat
Mumbai	Maharashtra
Jawaharlal Nehru Port	Maharashtra
Mangalore	Karnataka
Chennai	Tamil Nadu
Kakinada	Andhra Pradesh
Haldia	West Bengal
Krishnapatnam	Andhra Pradesh
Mundra	Gujarat
Kolkata	West Bengal

Source:- http://www.infodriveindia.com/india-import-data/edible-palm-oil-import-data.aspx

http://www.cybex.in/india-imports-data/Edible-Palm-Oil-Imports.aspx

Table 3.3:- Variable Definition and Summary Statistics

Variable	Definition	Mean	Standard Deviation	Source
Real wage rate	District specific real average daily wage rate for the agricultural laborers (in Rs.)	27.15169	15.04374	NSS
Real unit value	District specific average of the nominal unit values of edible oil deflated by the consumer price index	28.57182	4.580177	NSS
Rainfall	District specific average annual rainfall figures(in millimetres)	1199.648	714.6695	Gridded dataset of the Centre of Climatic Research at the University of Delaware
Oilseeds production	District specific average annual production of oilseeds(in tonnes)	25349.89	244885.7	IFPRI
Share of area in oilseeds production	District specific share of area in oilseeds production in total agricultural production in the initial/base period i.e. 1993-94	0.1310006	0.161887	IFPRI
Literate	District specific proportion of literate people (a person is defined as literate if he/she has at least secondary education)	0.1117669	0.0713828	NSS

St-Sc	District specific share of scheduled caste, scheduled tribe and other backward class population in total population	0.3195154	0.1708541	NSS
Bus	Proportion of villages in a district connected by bus	0.473835	0.3282191	Census
Rail	Proportion of villages in a district connected by rail	0.0215946	0.0232749	Census
Irrigation	Proportion of villages in a district under irrigation	0.445031	0.2961655	Census
Distance	Great circle distance(geodesic) between the head quarter of each district and nearest seaport	410.5883	291.595	Calculated using the latitudes and longitudes of the district head quarters and sea ports
CPI	State specific consumer price index for agricultural laborers	1.945291	0.7501382	EPW database
Initial expenditure	Monthly district specific average per-capita expenditure in the initial period i.e. 1993-94(in Rs.)	283.6814	74.01158	NSS
Composite tariff	District specific composite tariff measure(employment share weighted average of the ad-valorem tariff rates)	0.2386742	0.1570499	Average employment share at the 3 digit NIC level has been computed from NSS Employment-Unemployment

				round in the initial period (i.e. 1993-94). Tariff data is obtained from World Integrated Trade Solution (WITS) database
Palm oil tariff	Ad valorem tariff rate for palm oil	0.524	0.3067349	WITS
Palm oil world price	World price for palm oil ($/metric ton)	639.946	272.9615	World Bank Commodity Price Data

Table 3.4a:- Price Regression (Basic Specification)

VARIABLES	(1) ln edible oil price	(2) ln edible oil price	(3) ln edible oil price
ln Palm Oil Price	0.642***	0.629***	0.629***
	(0.0358)	(0.0482)	(0.0482)
Coastal_States*ln Palm Oil Price	0.129***	0.124***	0.124***
	(0.0310)	(0.0344)	(0.0344)
High Oilseeds Producing Districts*ln Palm Oil Price	0.0638***	0.148***	0.148***
	(0.0126)	(0.0337)	(0.0337)
Observations	1,476	1,476	1,476
R-squared	0.855	0.876	0.876
Number of districts	302	302	302
district fixed effects	yes	yes	yes
time trend	yes	yes	yes
state-region time trend	yes	no	no
district time trend	no	yes	yes
initial per capita expenditure	no	no	yes

Notes: - Log of the domestic edible oil price deflated by the consumer price index of agricultural laborers has been used as a dependent variable. The unit of analysis is a district. The dummy variable 'Coastal_States' takes the value one for districts belong to coastal states and the dummy variable 'High Oilseeds Producing Districts' takes the value one for those districts for which the share of area belongs to oilseeds production lie in the top quartile of the distribution of area shares in the initial period(i.e. 1993-94). Robust clustered standard errors are in parenthesis; ***,** and * indicates significance at the 1%, 5% and 10% levels respectively.

Table 3.4b:- Wage Regression (Basic Specification)

VARIABLES	(1) ln real wage rate	(2) ln real wage rate	(3) ln real wage rate
ln Palm Oil Price	0.265***	0.341***	0.341***
	(0.0988)	(0.130)	(0.130)
Coastal_States*ln Palm Oil Price	0.179**	0.176*	0.176*
	(0.0848)	(0.0962)	(0.0962)
High Oilseeds Producing Districts*ln Palm Oil Price	-0.0860	0.0686	0.0686
	(0.0613)	(0.110)	(0.110)
Observations	1,476	1,476	1,476
R-squared	0.596	0.724	0.724
Number of districts	302	302	302
district fixed effects	yes	yes	yes
time trend	yes	yes	yes
state-region time trend	yes	no	no
district time trend	no	yes	yes
initial per capita expenditure	no	no	yes

Notes:- Log of the daily nominal agricultural wage rate deflated by the consumer price index of agricultural laborers has been used as a dependent variable. The unit of analysis is a district. The dummy variable 'Coastal_States' takes the value one for districts belong to coastal states and the dummy variable 'High Oilseeds Producing Districts' takes the value one for those districts for which the share of area belongs to oilseeds production lie in the top quartile of the distribution of area shares in the initial period(i.e. 1993-94). Robust clustered standard errors are in parenthesis; ***, ** and * indicates significance at the 1%, 5% and 10% levels respectively.

Table 3.5a:-Robustness Check 1 for Price Regression (Controlling for the Aggregate Oilseeds Production)

VARIABLES	(1) ln edible oil price	(2) ln edible oil price	(3) ln edible oil price	(4) ln edible oil price
ln Palm Oil Price	0.645***	0.633***	0.633***	0.633***
	(0.0354)	(0.0490)	(0.0490)	(0.0490)
Coastal States*ln Palm Oil Price	0.128***	0.123***	0.123***	0.123***
	(0.0310)	(0.0345)	(0.0345)	(0.0345)
High Oilseeds Producing Districts* ln Palm Oil Price	0.0638***	0.144***	0.144***	0.144***
	(0.0125)	(0.0339)	(0.0339)	(0.0339)
Observations	1,476	1,476	1,476	1,476
R-squared	0.856	0.877	0.877	0.877
Number of districts	302	302	302	302
district fixed effects	yes	yes	yes	yes
time trend	yes	yes	yes	yes
state-region time trend	yes	no	no	no
district time trend	no	yes	yes	yes
initial per capita exp	no	no	yes	yes
state oilseeds production	yes	yes	yes	yes
national oilseeds production	no	no	no	yes

Notes: - Log of the domestic edible oil price deflated by the consumer price index of agricultural laborers has been used as a dependent variable. The unit of analysis is a district. The dummy variable 'Coastal_States' takes the value one for districts belong to coastal states and the dummy variable 'High Oilseeds Producing Districts' takes the value one for those districts for which the share of area belongs to oilseeds production lie in the top quartile of the distribution of area shares in the initial period(i.e. 1993-94). Robust clustered standard errors are in parenthesis; ***,** and * indicates significance at the 1%, 5% and 10% levels respectively.

Table 3.5b:-Robustness Check 1 for Wage Regression (Controlling for the Aggregate Oilseeds Production)

VARIABLES	(1) ln real wage rate	(2) ln real wage rate	(3) ln real wage rate	(4) ln real wage rate
ln Palm Oil Price	0.260***	0.330**	0.330**	0.330**
	(0.0988)	(0.128)	(0.128)	(0.128)
Coastal States*ln Palm Oil Price	0.181**	0.178*	0.178*	0.178*
	(0.0851)	(0.0965)	(0.0965)	(0.0965)
High Oilseeds Producing Districts* ln Palm Oil Price	-0.0859	0.0780	0.0780	0.0780
	(0.0613)	(0.112)	(0.112)	(0.112)
Observations	1,476	1,476	1,476	1,476
R-squared	0.597	0.725	0.725	0.725
Number of districts	302	302	302	302
district fixed effects	yes	yes	yes	yes
time trend	yes	yes	yes	yes
state-region time trend	yes	no	no	no
district time trend	no	yes	yes	yes
initial per capita exp	no	no	yes	yes
state oilseeds production	yes	yes	yes	yes
national oilseeds production	no	no	no	yes

Notes: - Log of the daily nominal agricultural wage rate deflated by the consumer price index of agricultural laborers has been used as a dependent variable. The unit of analysis is a district. The dummy variable 'Coastal_States' takes the value one for districts belong to coastal states and the dummy variable 'High Oilseeds Producing Districts' takes the value one for those districts for which the share of area belongs to oilseeds production lie in the top quartile of the distribution of area shares in the initial period(i.e. 1993-94). Robust clustered standard errors are in parenthesis; ***,** and * indicates significance at the 1%, 5% and 10% levels respectively.

Table 3.6a:-Robustness Check 2 for Price Regression (Using the Continuous Distance Measure Instead of Dummy Variable for Coastal States)

VARIABLES	(1) ln edible oil price	(2) ln edible oil price	(3) ln edible oil price
ln Palm Oil Price	0.671***	0.696***	0.696***
	(0.0389)	(0.0581)	(0.0581)
Distance*ln Palm Oil Price	-4.66e-05	-9.02e-05*	-9.02e-05*
	(3.54e-05)	(5.22e-05)	(5.22e-05)
High Oilseeds Producing Districts*ln Palm Oil Price	0.0676***	0.152***	0.152***
	(0.0126)	(0.0346)	(0.0346)
Observations	1,507	1,507	1,507
R-squared	0.844	0.864	0.864
Number of districts	302	302	302
district fixed effects	yes	yes	yes
time trend	yes	yes	yes
state-region time trend	yes	no	no
district time trend	no	yes	yes
initial per capita expenditure	no	no	yes

Notes: - Log of the domestic edible oil price deflated by the consumer price index of agricultural laborers has been used as a dependent variable. The unit of analysis is a district. The variable 'Distance' measures the distance in kilometers between each district head-quarter and its nearest sea-por. The dummy variable 'High Oilseeds Producing Districts' takes the value one for those districts for which the share of area belongs to oilseeds production lie in the top quartile of the distribution of area shares in the initial period(i.e. 1993-94). Robust clustered standard errors are in parenthesis; ***,** and * indicates significance at the 1%, 5% and 10% levels respectively.

Table 3.6b:-Robustness Check 2 for Wage Regression (Using the Continuous Distance Measure Instead of Dummy Variable for Coastal States)

VARIABLES	(1) ln real wage rate	(2) ln real wage rate	(3) ln real wage rate
ln Palm Oil Price	0.344***	0.480***	0.480***
	(0.107)	(0.142)	(0.142)
Distance*ln Palm Oil Price	-0.000115	-0.000305	-0.000305
	(0.000151)	(0.000194)	(0.000194)
High Oilseeds Producing Districts*ln Palm Oil Price	-0.0915	0.0285	0.0285
	(0.0611)	(0.111)	(0.111)
Observations	1,507	1,507	1,507
R-squared	0.596	0.722	0.722
Number of districts	302	302	302
district fixed effects	yes	yes	yes
time trend	yes	yes	yes
state-region time trend	yes	no	no
district time trend	no	yes	yes
initial per capita expenditure	no	no	yes

Notes: - Log of the daily nominal agricultural wage rate deflated by the consumer price index of agricultural laborers has been used as a dependent variable. The unit of analysis is a district. The variable 'Distance' measures the distance in kilometers between each district head-quarter and its nearest sea-por. The dummy variable 'High Oilseeds Producing Districts' takes the value one for those districts for which the share of area belongs to oilseeds production lie in the top quartile of the distribution of area shares in the initial period(i.e. 1993-94). Robust clustered standard errors are in parenthesis; ***,** and * indicates significance at the 1%, 5% and 10% levels respectively.

Table 3.7a:-Robustness Check 3 for Price Regression (Using Three Categories for the Share of Area Devoted to Oilseeds Cultivation)

VARIABLES	(1) ln edible oil price	(2) ln edible oil price	(3) ln edible oil price
ln Palm Oil Price	0.625***	0.576***	0.576***
	(0.0361)	(0.0493)	(0.0493)
Coastal States*ln Palm Oil Price	0.124***	0.110***	0.110***
	(0.0312)	(0.0360)	(0.0360)
top25_area_share*ln Palm Oil Price	0.0353***	0.0986**	0.0986**
	(0.0101)	(0.0389)	(0.0389)
between25_75_area_share*ln Palm Oil Price	0.0937***	0.219***	0.219***
	(0.0159)	(0.0434)	(0.0434)
Observations	1,476	1,476	1,476
R-squared	0.856	0.878	0.878
Number of districts	302	302	302
district fixed effects	yes	yes	yes
time trend	yes	yes	yes
state-region time trend	yes	no	no
district time trend	no	yes	yes
initial per capita exp	no	no	yes

Notes: - Log of the domestic edible oil price deflated by the consumer price index of agricultural laborers has been used as a dependent variable. The unit of analysis is a district. The dummy variable 'Coastal_States' takes the value one for districts belong to coastal states, the dummy variable 'top25_area_share' takes the value one for those districts for which the share of area belongs to oilseeds production lie in the top quartile of the distribution of area shares in the initial period (i.e. 1993-94) and the dummy variable 'between25_75_area_share' takes the value one for those districts for which this area share is between 25% and 75% of the distribution. Robust clustered standard errors are in parenthesis; ***,** and * indicates significance at the 1%, 5% and 10% levels respectively.

Table 3.7b:-Robustness Check 3 for Wage Regression (Using Three Categories for the Share of Area Devoted to Oilseeds Cultivation)

VARIABLES	(1) ln real wage rate	(2) ln real wage rate	(3) ln real wage rate
ln Palm Oil Price	0.272***	0.355**	0.355**
	(0.103)	(0.147)	(0.147)
Coastal States*ln Palm Oil Price	0.181**	0.180*	0.180*
	(0.0856)	(0.0987)	(0.0987)
top25_area_share*ln Palm Oil Price	-0.0144	-0.0261	-0.0261
	(0.0479)	(0.114)	(0.114)
between25_75_area_share*ln Palm Oil Price	-0.0981	0.0499	0.0499
	(0.0787)	(0.142)	(0.142)
Observations	1,476	1,476	1,476
R-squared	0.596	0.725	0.725
Number of districts	302	302	302
district fixed effects	yes	yes	yes
time trend	yes	yes	yes
state-region time trend	yes	no	no
district time trend	no	yes	yes
initial per capita exp	no	no	yes

Notes: - Log of the daily nominal agricultural wage rate deflated by the consumer price index of agricultural laborers has been used as a dependent variable. The unit of analysis is a district. The dummy variable 'Coastal_States' takes the value one for districts belong to coastal states, the dummy variable 'top25_area_share' takes the value one for those districts for which the share of area belongs oilseeds to production lie in the top quartile of the distribution of area shares in the initial period (i.e. 1993-94) and the dummy variable 'between25_75_area_share' takes the value one for those districts for which this area share is between 25% and 75% of the distribution. Robust clustered standard errors are in parenthesis; ***, ** and * indicates significance at the 1%, 5% and 10% levels respectively.

Appendices for Chapter 3

3. A Conditions Under Which Budget Share of Palm Oil is Invariant to Its Own Border Price

We proceed to investigate the conditions under which budget share of palm oil is invariant to its own price i.e. $\frac{\partial w_1}{\partial \ln c_1} = 0$. We can write down the following:

$$\frac{\partial w_1}{\partial \ln c_1} = c_1 \frac{\partial w_1}{\partial c_1}$$

where $$w_1 = \frac{P_1 q_1}{P_1 q_1 + P_2 q_2} = \frac{P_1 q_1}{I}$$

I stands for the total expenditure on palm oil and local edible oil i.e.

$$I = P_1 q_1 + P_2 q_2$$

q_1 and q_2 are the quantity consumed of palm oil and local edible oil respectively.

Therefore

(A1) $$c_1 \frac{\partial w_1}{\partial c_1} = c_1[(1/I)\{P_1(\partial q_1/\partial c_1) + q_1(\partial P_1/\partial c_1)\} - (P_1 q_1/I^2)(\partial I/\partial c_1)]$$

Now focus on the first term inside the square bracket in expression (A1).

(A2) $$\begin{aligned} &(1/I)\{P_1(\partial q_1/\partial c_1) + q_1(\partial P_1/\partial c_1)\} \\ &= (1/I)\{P_1(\partial q_1/\partial P_1)(\partial P_1/\partial c_1) + q_1(\partial P_1/\partial c_1)\} \\ &= (1/I)(\partial P_1/\partial c_1)\{P_1(\partial q_1/\partial P_1) + q_1\} \\ &= (1/I)(\partial P_1/\partial c_1)\{(P_1/q_1)(\partial q_1/\partial P_1) + 1\} \\ &= (1/I)(\partial P_1/\partial c_1)\{-\varepsilon_{11} + 1\} \end{aligned}$$

The entire expression (A2) will be zero if the own price elasticity of palm oil equals one (unit elastic) i.e. $\varepsilon_{11} = 1$.

The second term inside the square bracket in expression (A1) crucially depends on the term $\partial I / \partial c_1$. If $\partial I / \partial c_1 = 0$ then the entire second term becomes zero.

(A3)

$$\partial I / \partial c_1 = [P_1(\partial q_1 / \partial c_1) + q_1(\partial P_1 / \partial c_1)] + [P_2(\partial q_2 / \partial c_1) + q_2(\partial P_2 / \partial c_1)]$$

We can write down the first term in (A3) as

$$\begin{aligned}
&\{P_1(\partial q_1 / \partial c_1) + q_1(\partial P_1 / \partial c_1)\} \\
&= \{P_1(\partial q_1 / \partial P_1)(\partial P_1 / \partial c_1) + q_1(\partial P_1 / \partial c_1)\} \\
&= (\partial P_1 / \partial c_1)\{P_1(\partial q_1 / \partial P_1) + q_1\} \\
&= (\partial P_1 / \partial c_1)\{(P_1 / q_1)(\partial q_1 / \partial P_1) + 1\} \\
&= (\partial P_1 / \partial c_1)\{-\varepsilon_{11} + 1\}
\end{aligned}$$

Therefore the first term in (A3) becomes zero when palm oil price is unit elastic.

The second term in (A3) is written as

$$\begin{aligned}
&[P_2(\partial q_2 / \partial c_1) + q_2(\partial P_2 / \partial c_1)] \\
&= [P_2(\partial q_2 / \partial P_2)(\partial P_2 / \partial c_1) + q_2(\partial P_2 / \partial c_1)]
\end{aligned}$$

The above term will be zero if $\partial P_2 / \partial c_1 = 0$. It can be easily seen from our theoretical model that $\partial P_2 / \partial c_1 = 0$ when the cross price elasticity of local edible oil with respect to palm oil is zero i.e. $\beta_{21} = 0$

Therefore the zero cross price elasticity of local oil price with respect to palm oil along with unit elastic demand for palm oil ensures that $\frac{\partial w_1}{\partial \ln c_1} = 0$.

3. B Conditions Under Which Budget Share of Palm Oil is Invariant to Distance from Port

If $\frac{\partial w_1}{\partial \ln c_1} = 0$ i.e. the budget share of palm oil is invariant to its own price, the expression for the pass-through elasticity of average edible oil price boils down to

$$\frac{\partial \ln P}{\partial \ln c_1} = w_1 \frac{\partial \ln P_1}{\partial \ln c_1}$$

The change in the pass-through rate of the average edible oil price with distance (from the port) under this special situation is written as

$$(\partial / \partial r)\frac{\partial \ln P}{\partial \ln c_1} = w_1 (\partial / \partial r)\frac{\partial \ln P_1}{\partial \ln c_1} + (\frac{\partial \ln P_1}{\partial \ln c_1})(\partial w_1 / \partial r)$$

As shown earlier, the first term i.e. the pass-through elasticity of palm oil price is monotonically declining with distance. Now we show that under certain conditions, the second term becomes zero and hence the pass-through elasticity of average palm oil price declines with distance from the port. The second term becomes zero if $\partial w_1 / \partial r = 0$ (because $\frac{\partial \ln P_1}{\partial \ln c_1} > 0$)

Now

(B1) $$\frac{\partial w_1}{\partial r} = [(1/I)\{P_1(\partial q_1 / \partial r) + q_1(\partial P_1 / \partial r)\} - (P_1 q_1 / I^2)(\partial I / \partial r)]$$

First of all, we focus on the first term in equation (B1). Rewriting the first term we derive the following

(B2)
$$\begin{aligned}&(1/I)\{P_1(\partial q_1/\partial r)+q_1(\partial P_1/\partial r)\}\\&=(1/I)q_1(\partial P_1/\partial r)\{(P_1/q_1)(\partial q_1/\partial r)/(\partial P_1/\partial r)+1\}\\&=(1/I)q_1(\partial P_1/\partial r)\{(P_1/q_1)(\partial q_1/\partial P_1)+1\}\\&=(1/I)q_1(\partial P_1/\partial r)\{-\varepsilon_{11}+1\}\end{aligned}$$

Therefore the expression in (B2) becomes zero if the own price elasticity of palm oil is unity i.e. $\varepsilon_{11}=1$.

The second term of (B1) can be written more explicitly as

(B3)
$$\begin{aligned}&(P_1q_1/I^2)[\{P_1(\partial q_1/\partial r)+q_1(\partial P_1/\partial r)\}+\{P_2(\partial q_2/\partial r)+q_2(\partial P_2/\partial r)\}];(I=P_1q_1+P_2q_2)\\&=(P_1q_1/I^2)[\{P_1(\partial q_1/\partial r)+q_1(\partial P_1/\partial r)\}]\\&+(P_1q_1/I^2)[\{P_2(\partial q_2/\partial r)+q_2(\partial P_2/\partial r)\}]\\&=(P_1q_1/I^2)(\partial P_1/\partial r)q_1\{(P_1/q_1)(\partial q_1/\partial r)/(\partial P_1/\partial r)+1\}\\&+(P_1q_1/I^2)(\partial P_2/\partial r)q_2\{(P_2/q_2)(\partial q_2/\partial r)/(\partial P_2/\partial r)+1\}\\&=(P_1q_1/I^2)q_1(\partial P_1/\partial r)\{(P_1/q_1)(\partial q_1/\partial P_1)+1\}\\&+(P_1q_1/I^2)q_2(\partial P_2/\partial r)\{(P_2/q_2)(\partial q_2/\partial P_2)+1\}\\&=(P_1q_1/I^2)q_1(\partial P_1/\partial r)\{-\varepsilon_{11}+1\}+(P_1q_1/I^2)q_2(\partial P_2/\partial r)\{-\varepsilon_{22}+1\}\end{aligned}$$

The entire expression (B3) is unambiguously zero if the own price elasticity of both palm oil and local edible is unity i.e. $\varepsilon_{11}=1$ as well as $\varepsilon_{22}=1$.

Therefore the unit elasticity of palm oil price and local edible oil price ensures that the budget share of palm oil does not change with the distance from the port i.e. $\frac{\partial w_1}{\partial r}=0$. Hence the pass-through elasticity of average edible oil price monotonically declines with distance.

3. C The Change in the Pass-Through Rate With Respect to Per-Unit Cost of Local Edible Oil Production

The change in the pass-through elasticity of palm oil with respect to the per unit cost of local edible oil (c_2) is written as

(C1) $$\frac{\partial}{\partial c_2}\left(\frac{\partial \ln P_1}{\partial \ln c_1}\right) = -\frac{2\beta_{11}\beta_{22}}{4\beta_{11}\beta_{22}-\beta_{12}\beta_{21}}.\left(\frac{c_1}{P_1^2}\right)\left(\frac{\partial P_1}{\partial c_2}\right)$$

The expression (C1) is negative because $\frac{\partial P_1}{\partial c_2} > 0$. This implies that the pass-through elasticity of palm oil price with respect to its border price declines as the per unit cost of production of local edible oil increases. Therefore the pass-through elasticity is higher in high oilseeds producing regions where the per unit cost is less.

On the other hand, the change in the pass-through elasticity of local edible oil with respect to the per unit cost of local edible oil (c_2) is the following:

(C2) $$\frac{\partial}{\partial c_2}\left(\frac{\partial \ln P_2}{\partial \ln c_1}\right) = -\frac{\beta_{11}\beta_{21}}{4\beta_{11}\beta_{22}-\beta_{12}\beta_{21}}.\left(\frac{c_1}{P_2^2}\right)\left(\frac{\partial P_2}{\partial c_2}\right)$$

(C2) is negative as $\frac{\partial P_2}{\partial c_2} > 0$. Therefore the pass-through elasticity of local edible oil is also higher in high oilseeds producing region where the per unit cost is less compared to the low oilseeds producing regions.

The pass-through elasticity of the average edible oil price:

$$\frac{\partial \ln P}{\partial \ln c_1} = \frac{\partial}{\partial \ln c_1}[w_1 \ln P_1 + w_2 \ln P_2]$$

$$= w_1\left(\frac{\partial \ln P_1}{\partial \ln c_1} - \frac{\partial \ln P_2}{\partial \ln c_1}\right) + (\ln P_1 - \ln P_2)\frac{\partial w_1}{\partial \ln c_1} + \frac{\partial \ln P_2}{\partial \ln c_1}$$

$\frac{\partial \ln P_2}{\partial \ln \mathrm{c}_1} = 0$ if $\beta_{21} = 0$ i.e. the cross price elasticity of local edible oil with respect to palm oil price is zero. It has been already shown that $\frac{\partial w_1}{\partial \ln c_1} = 0$ when the palm oil price is unit elastic and $\beta_{21} = 0$

.

Therefore, under these special conditions

$$\frac{\partial \ln P}{\partial \ln c_1} = w_1 \frac{\partial \ln P_1}{\partial \ln c_1}$$

The change in the average edible oil price with respect to c_2 turns out to be

(C3) $$(\partial / \partial c_2) \frac{\partial \ln P}{\partial \ln c_1} = w_1 (\partial / \partial c_2) \frac{\partial \ln P_1}{\partial \ln c_1} + (\frac{\partial \ln P_1}{\partial \ln c_1})(\partial w_1 / \partial c_2)$$

The first term of (C3) is negative as we have already shown. The second term becomes zero if $(\partial w_1 / \partial c_2) = 0$.

Now

(C4) $$\frac{\partial w_1}{\partial c_2} = [(1/I)\{P_1(\partial q_1 / \partial c_2) + q_1(\partial P_1 / \partial c_2)\} - (P_1 q_1 / I^2)(\partial I / \partial c_2)]$$

The first term in expression (C4) is explicitly written as

(C5) $$\begin{aligned} &(1/I)\{P_1(\partial q_1 / \partial c_2) + q_1(\partial P_1 / \partial c_2)\} \\ &= (1/I)\{P_1(\partial q_1 / \partial P_1)(\partial P_1 / \partial c_2) + q_1(\partial P_1 / \partial c_2)\} \end{aligned}$$

Therefore the term written above becomes zero if $\partial P_1 / \partial c_2 = 0$. From our theoretical model, it can be easily checked that $\partial P_1 / \partial c_2 = 0$ if the cross price elasticity of palm oil price with respect to local edible oil price is zero i.e. $\beta_{12} = 0$.

From the second term in expression (C4), we can write down

(C6)

$$
\begin{aligned}
&(P_1 q_1 / I^2)\{P_1(\partial q_1 / \partial c_2) + q_1(\partial P_1 / \partial c_2)\} \\
&+ (P_1 q_1 / I^2)\{P_2(\partial q_2 / \partial c_2) + q_2(\partial P_2 / \partial c_2)\} \\
&= (P_1 q_1 / I^2)\{P_1(\partial q_1 / \partial P_1)(\partial P_1 / \partial c_2) + q_1(\partial P_1 / \partial c_2)\} \\
&+ (P_1 q_1 / I^2)\{P_2(\partial q_2 / \partial P_2)(\partial P_2 / \partial c_2) + q_2(\partial P_2 / \partial c_2)\} \\
&= (P_1 q_1 / I^2)\{P_1(\partial q_1 / \partial P_1)(\partial P_1 / \partial c_2) + q_1(\partial P_1 / \partial c_2)\} \\
&+ (P_1 q_1 / I^2)(\partial P_2 / \partial c_2)\{(P_2 / q_2)(\partial q_2 / \partial P_2) + 1\} \\
&= (P_1 q_1 / I^2)\{P_1(\partial q_1 / \partial P_1)(\partial P_1 / \partial c_2) + q_1(\partial P_1 / \partial c_2)\} \\
&+ (P_1 q_1 / I^2)(\partial P_2 / \partial c_2)\{-\varepsilon_{22} + 1\}
\end{aligned}
$$

The term $(P_1 q_1 / I^2)\{P_1(\partial q_1 / \partial P_1)(\partial P_1 / \partial c_2) + q_1(\partial P_1 / \partial c_2)\}$ in (C6) turns out to be zero as $\partial P_1 / \partial c_2 = 0$. The term $\partial P_1 / \partial c_2 = 0$ if $\beta_{12} = 0$. The term $(P_1 q_1 / I^2)(\partial P_2 / \partial c_2)\{-\varepsilon_{22} + 1\}$ equals zero for unit elasticity of local edible oil price. Therefore unit own price elasticity of both palm oil and local edible oil and the zero cross price elasticities assure that $\partial w_1 / \partial c_2 = 0$

If $\partial w_1 / \partial c_2 = 0$, then we get

$$(\partial / \partial c_2) \frac{\partial \ln P}{\partial \ln c_1} = w_1 (\partial / \partial c_2) \frac{\partial \ln P_1}{\partial \ln c_1}$$

Now

$(\partial / \partial c_2) \dfrac{\partial \ln P}{\partial \ln c_1} < 0$ as $\dfrac{\partial \ln P_1}{\partial \ln c_1} < 0$ (pass-through elasticity of average edible oil price declines with per unit cost of production for local edible oil).

3. D Data Sources and Construction of Variables:-

The primary dataset used in this chapter are the Employment and Unemployment and Consumer Expenditure rounds (1993-94, 1999-00, 2004-05, 2007-08 and 2011-12) of NSS in India. NSS is a cross-sectional dataset which is representative of India's population. The entire NSS sample is divided into two parts; rural sample and urban sample. In this chapter, we consider the rural sample only. In rural areas, the first stratum is a district. Villages are primary sampling units (PSU) and are picked randomly in a district over an entire agricultural year (July to June) over quarters to ensure equal spacing of observations across the year. The households are randomly chosen in the selected PSU's. The district level analysis spans 304 districts that includes14 major states in the sample: Punjab, Haryana, Uttar Pradesh (includes Uttarakhand), Madhya Pradesh (includes Chhattisgarh), Bihar (includes Jharkhand), Gujarat, Rajasthan, West Bengal, Maharashtra, Andhra Pradesh, Karnataka, Orissa, Tamil Nadu and Kerala.

The geographic boundaries refer to the boundaries of the parent districts as Indian districts have been split into two or more districts over time. Districts across NSS rounds (1993-94, 1999-00, 2004-05, 2007-08 and 2011-12) have hence been merged into their parent districts according to district boundaries in the 1991 census.

The dependent variable for the price regression is the domestic edible oil price. We deflate it by the consumer price index of agricultural laborers (at the state level) to adjust for inflation. This inflation adjusted domestic edible oil price is used as dependent variable in the price regression. The district specific unit value of edible oil, computed from the NSS Consumer Expenditure Survey is considered as the domestic edible oil price at the district level. NSS Consumer Expenditure Survey provides information on the total expenditure and quantity consumed of edible oil at the household level. The total expenditure and quantity consumed can be summed up at the district level. Dividing the district level total expenditure by the district specific total quantity consumed, we obtain the unit value of edible oil (for a district).

The real agricultural wage rate is used as a dependent variable in the wage regression. The district specific daily real agricultural wage rate is obtained through deflating the district specific daily nominal wage rate by the state specific consumer price index of agricultural laborers (CPI-AL). District specific daily nominal wage rate is calculated from the NSS Employment and Unemployment Survey data for all the time periods mentioned earlier.

We get the data on agricultural production from International Food Policy Research Institute (IFPRI) database. Deflating the area belongs to oilseeds production by the total agricultural production, we obtain the share of area belongs to oilseeds production for each district. A district is said to be a high oilseeds producing district if its share of area in oilseeds cultivation in the initial period of our study (i.e. 1993-94) lies in the top quartile of the distribution of area shares.

A coastal district is the one that belongs to a coastal state. Andhra Pradesh, Gujarat, Karnataka, Kerala, Maharashtra, Orissa, Tamil Nadu and West Bengal are coastal states in our sample.

Ad-valorem tariff rate of palm oil is taken from WITS (World Integrated Trade Solution) data. We get the world price of palm oil from the World Bank commodity price data. We have used the exchange rate adjusted palm oil price in the regressions i.e. palm oil price multiplied by the exchange rate (rupee to dollar exchange rate; Source RBI database on Indian Economy).

Just like Topalova (2010) we also create a composite tariff measure at the district level to use it as a control variable in our regressions. This composite tariff measure is the employment weighted average of tariff rates. It controls for the change in tariff rates of all other commodities (apart from the imported edible oil i.e. palm oil). In order to construct that composite tariff measure, we compute the average employment share for each district at the three digits (NIC code) level from NSS employment-unemployment survey data. The employment share corresponds to the initial period i.e.1993-94. Data on ad-valorem tariff rate at the 6 digit commodity level is available from WITS (World Integrated Trade Solution) database. We match 3 digit NIC codes with the 6 digit trade codes to calculate the tariff rate at the 3 digit NIC code level(for almost 200 commodities). Then we multiply the tariff rate (computed at the 3 digit level) with the employment share (also at the 3 digit level) and sum it up for all commodities to compute the composite tariff rate.

India eliminated quantitative restrictions for most of the commodities by 2001(Goldar, 2005; Mehta, 2000). Therefore in the period after 2001, we have an economy free of quantitative restriction and that may have some impact in the determination of prices and wages. We capture the effect of quantitative restrictions by generating a dummy variable that takes the value one for the period after 2001and zero for the period before 2001. We also interact this dummy with the composite tariff measure and use it in the regressions to control for their joint impact on the dependent variables.

The rainfall figures are taken from the gridded dataset of the Centre of Climatic Research at the University of Delaware, which includes monthly precipitation values on 0.5 degree intervals in longitude and latitude centered on 0.25 degree. This grid value is achieved by spatial interpolation using data from nearby weather stations and other sources of rainfall data. District level monthly rainfall estimates were arrived at by averaging the monthly precipitation value of all the grid points lying within the geographic boundaries of a district in a year. The geographic boundaries refer to the 1991 Indian census boundaries as Indian districts have been over time split into two or more districts.

We obtain the data on share of villages connected by bus, rail and share of villages irrigated in a district (each as separate variable) from the census data of India. These are infrastructural variables at the district level. For the period 1993-94, we use the census data of 1991. For the period 1999-2000, 2004-05 and 2007-08, the census data of 2001 is used. For 2011-12, we use the 2011 census figures.

Chapter 4

Poverty, Gender and Well-Being: A Study on the Slum Population in Delhi

4.1 Introduction

Subjective well-being/happiness as an empirical measure of welfare is gradually becoming more accepted by the economists and the policy makers. It seems appropriate, therefore, to examine some of its implications. Although there is now a growing literature on happiness or life satisfaction, there are few papers on life satisfaction among the poor. The most notable exceptions are the papers by Banerjee, Deaton and Duflo (2004) and Case and Deaton (2005).

In their paper, Banerjee, Deaton and Duflo (2004) have dealt specifically with the life satisfaction of the poor people in rural Udaipur in Rajasthan. In a similar study, Case and Deaton (2005) evaluate the level and determinants of health satisfaction for the poor in India and South Africa. These authors find that poor tend to report high levels of life satisfaction and health satisfaction. This is quite a surprising finding and contradictory to the usual belief that self-reported life and health satisfaction can be high only for the wealthy people. On the other hand, their studies also find that the poor report low levels of financial satisfaction. The poor people are presumably adapted to the life and sickness they experience, in that they don't see themselves as particularly unhappy or unhealthy. Yet they are not adapted in the same way to their financial status.

The first objective of this chapter is to revisit this finding in an urban context. The poor in urban areas are geographically proximate to affluent neighbourhoods and the consumption of the wealthy. If, relative to rural poor, they are more aware and therefore, more aspiring of a more comfortable life, then would adaptation play a lesser role in reporting life satisfaction? To answer this question, I conducted a survey on the low income population across the slums of Delhi.

In this chapter, I report results that show that even in an urban setting, reported life and health satisfaction scores are on the high side. On the other hand, like the rural poor, the proportion of the slum population reporting low levels of financial satisfaction turns out to be larger compared to the proportion of people reporting low levels of life and health satisfaction. Despite the higher than expected life satisfaction scores, there is enough variability for us to figure out the economic and non-economic correlates of well-being.

The second objective of the chapter is to see whether the self-reported well-being measures and their determinants differ systematically between men and women. When societies offer different opportunities and liberty to men and women, they may experience life satisfaction differently and the factors that trigger it may also differ. The impact of any factor on subjective well-being may also vary between male and female respondents because of divergent preferences. While recent work has drawn attention to the temporal and spatial variation in female well-being (relative to males); this chapter is the first study, to the best of my knowledge, to examine relative well-being of women among the poor.

Analyzing the data on life satisfaction during the period 1970 to 2005, Stevenson and Wolfers (2009) find that women's happiness has declined both absolutely and relative to men in United States. On the other hand, lives of women in the United States have improved over the past 35 years by many objective measures. They call it as the paradox of women's declining relative well-being.

A contrary finding from a cross-sectional analysis comes from Graham and Chattopadhyay (2012). Using a worldwide sample from Gallup World Poll, they find that self-reported well-being for women relative to men is higher in developed countries and for more educated cohorts. This finding can be interpreted as the increase in the subjective well-being of the women relative to men as the objective measures (like income and education) improve. An open question is whether this relation reflects the impact of income and education alone or whether it also due to country specific omitted factors especially relating to legal rights and social norms.

In the sample of urban slum-dwellers that I collected, we can plausibly assume that there is no variation in the omitted variables relating to legal rights and social norms. If so, the analysis in this chapter captures the impact of income and education without contamination by variation in social norms. Our analysis finds that women's relative well-being falls as income and education increase. The relative decline turns out to be more prominent for education as compared to income. This result is a cross sectional counterpart of what has been shown by Stevenson and Wolfers (2009) from the long time series data on life satisfaction for United States and industrialized nations in Europe. Like Stevenson and Wolfers (2009), we do not offer any definite answer to what is driving our results. Perhaps the most probable reason behind our finding is the rising aspiration/expectation of women with increase in education and income. In that sense our finding resembles Lalive and Stutzer (2010) who find the life satisfaction of women to be higher in the traditional communities compared to liberal communities in Switzerland and explain their finding as the result of higher expectation of the women in liberal communities.

Our analysis also reports on the gendered impacts of other factors on well-being. For instance, eligibility for a food subsidy scheme matters more to women's well-being than to men. This can be seen as evidence of divergent preferences between women and men.

The plan of this chapter is as follows. In the next section we relate this chapter to the relevant literature. Section 4.3 outlines the survey and section 4.4 provides a brief description of the questionnaire. In section 4.5, we explain the living conditions of the slums in Delhi. Section 4.6 reports about our general finding regarding the well-being measures. A few interesting bivariate correlation between life satisfaction and some of the economic and non-economic factors have been shown in section 4.7. Section 4.8 discusses whether these correlations vary between men and women. The ordered logistic model for our regression is explained in section 4.9 and we describe the variables for our regression in section 4.10. Basic regression results are shown in section next to that and the concluding remarks are gathered together in Section 4.12.

4.2 Relation to Literature

The measurement and analysis of subjective well-being has a long history in the social sciences. In the past, contributions by the economists have been relatively slim. Recent years, however, have seen a flowering of work.

There is a vast literature that focuses on the correlation between life satisfaction and income. The relationship between happiness and income is puzzling. In most of the cross-sectional survey, a positive correlation between happiness and income has been found (Easterlin (2001), Argyle (2003), Diener (1984)). When one turns to the change in happiness over time, however, a seeming contradiction arises to the positive income-happiness relationship. On average, even if income and economic circumstances improve substantially, no corresponding advancement in subjective well-being is noticed (Easterlin (1974, 2001)).

Easterlin argues that there are two main forces behind this paradoxical finding. The first one is the fact that the individuals consider relative income rather than absolute income as a determinant of their well-being. If income of everyone increases proportionately, relative income of a person does not change and as a result happiness remains unaltered as well. The second factor as explained by Easterlin is the material aspiration. Subjective well-being does not change if aspiration increases proportionately with income. This surprising result is known as the ‘Easterlin Paradox’ in the happiness literature. Di Tella and MacCulloch (2008) show that the Easterlin Paradox survives even after controlling for variables such as life expectancy, hours worked, inflation and unemployment. The Easterlin Paradox is a highly debated issue and there are many proponents (Oswald (1997)) and opponents (Hagerty and Veenhoven (2003), Stevenson and Wolfers (2008)) of it.

Another interesting application of subjective well-being/life satisfaction approach in economics is the monetary valuation of non-market goods. Typically in this approach, self-reported life satisfaction is regressed on the non-market goods of interest, income and other covariates. Using the coefficients of the non-market commodity and income it is possible to calculate utility constant trade off ratios between the non market good and income. Luechinger

(2009) uses the life satisfaction approach to value air quality for Germany, combining individual level panel and high resolution sulphur dioxide data. Luechinger and Raschky (2009) monetize utility losses caused by floods in 17 OECD countries between 1973 and 2004. Van Praag and Baarsma (2005) studied the external effects due to aircraft noise nuisance at the Amsterdam Airport Shiphol and found a trade-off ratio between income and exposure to noise. Air pollution has been found to reduce happiness (Welsch, 2006). Carbonell and Gowdy (2007) show the relationship between the subjective measures of well-being and individual environmental attitude using the British Household Panel Survey. As a measure of individual environmental attitude, they consider the concern about ozone pollution and concern about species extinction.

Fritjers and Van Praag (1998) estimated the effects of climate on both welfare and well-being in Russia. They also explored the costs and benefits of climate change measured as an increase of one, two or three degrees in temperature combined with an overall increase in precipitation of 5% or 10%. They find climate to be one important determinant of household's standard of living in Russia.

The use of subjective well-being measure has been advocated by many economists (like Easterlin (1974) and Helliwell (2002)). On the other hand, Deaton and Stone (2013) question the clarity of the subjective well-being measures. They argue that response to the well-being questions is sometimes deeply affected by the wording of the questions or by the context in which they are put.

A serious concern regarding the subjective well-being measure is the interpersonal comparisons of welfare. Widely used measures of subjective welfare ask survey respondents to rate their life satisfaction/happiness on an ordinal scale. However, different people may well have different ideas about what it means to be satisfied with one's life, leading them to interpret survey questions on subjective well-being differently. If the scale of the life satisfaction question is not understood the same way by different respondents it is unclear what meaning can be attached to such measures.

Although the problem of interpersonal comparability seems to be a threat to the accuracy of the subjective well-being measures, the evidences show that the application of these well-being measures may not be that erroneous. In the early 1960s, social psychologist Hadley Cantril carried out an intensive survey in fourteen countries worldwide, rich and poor, capitalist and communist, asking open-ended questions about what people want out of life – what they would need for their lives to be completely happy. Despite enormous socio-economic and cultural disparities among countries, what people said was strikingly similar. In every country, material circumstances, especially level of living, is considered the most important factor for a happy life. This is followed by happy family life, good health and job satisfaction. Therefore the comparability issue of the ordinal rankings on the life satisfaction question across respondents may not be as serious as it is thought. Ravallion, Himelein and Beegle (2016) propose a methodology to correct the comparability issue in the ordinal rankings on the well-being measures. But little difference is found in the coefficients on the covariates of life satisfaction or in their statistical significance even after correcting for the heterogeneity in the ordinal rankings.

Another major problem in the use of subjective well-being as a welfare measure is the finding that poor report high levels of subjective well-being and health status even though they report low financial satisfaction (Banerjee, Deaton and Duflo(2004) and Case and

Deaton(2005)). This is a surprising finding and supports Deaton and Stone's (2013) concern regarding the accuracy and validity of the subjective well-being measures.

An important feature of the paper by Banerjee, Deaton and Duflo (2004) and Case and Deaton (2005) is that they have focused on the rural areas. The rural poor may be geographically distant from affluent households and therefore may not be aware of a comfortable life. This could be a reason why they report high levels of subjective well-being. But does the story of adaption hold universally? If we focus on the urban slums, then the poor are in close proximity to affluent households and urban aspirations are likely to be different from that of the rural poor.

In this chapter, our story starts from here. Do the urban poor report high life and health satisfaction score just like the rural poor as shown by Banerjee, Deaton and Duflo (2004) and Case and Deaton (2005)? What about the reported value for financial satisfaction? This chapter seeks to answer these questions.

Despite the adaptation phenomenon, Case and Deaton (2005) find variability in the self-reported health status and show that it is adversely affected by all health conditions. They also find that self-reported health status is positively correlated with the literacy of women and holding of assets. This finding motivates us to check the variability of the self-reported well-being measures in our survey data and figure out the determinants of these self-reported measures if any variability exists. Therefore, even among the poor, subjective well-being can be correlated to objective economic performances.

The association between the subjective well-being and its correlates may differ systematically between men and women. The systematic difference may arise when societies offer different opportunities to men and women. It may also arise because of the divergent preference across gender. Unfortunately the literature on systematic difference between genders is small.

Stevenson and Wolfers (2009) find that women's happiness has declined both absolutely and relative to men in United States between 1970 and 2005. This happens despite the improvement of the women lives in United States in terms of many objective measures. Those measures include increasing female labor force participation, rise in the women's real wage rate relative to men, increasing freedom in the family life and many other things. Stevenson and Wolfers (2009) call it as the paradox of women's declining relative well-being. They conjecture several explanations for their finding. The decline in female happiness may be because of doubling of the total workload i.e. women's movement in the paid labor force without any shift away from household production. It may be also due to the fact that with the increase in gender equality over time, the reference group for the women expands i.e. it includes men also and as a result the women may find their relative position lower than when their reference group includes only women. Women's rising aspiration can also play a key role. There can be many other factors that may lead to the finding of declining female happiness.

The finding by Stevenson and Wolfers (2009) is restricted to United States and some industrialized countries in Europe. Using datasets for a wide selection of countries spanning the period, 1981 to 2009, Lima (2013) finds that for 71% of the countries, female happiness has increased relative to men. Therefore, Stevenson and Wolfer's (2009) finding is not validated globally.

Using a worldwide sample from Gallup World Poll, Graham and Chattopadhyay (2012) find that as one moves from lower income to higher income countries or from less educated to more educated cohorts, life satisfaction of women increases more than men. We can also interpret this finding as the increase in the life satisfaction of the women relative to men when objective measures (like income and education) improves.

Graham and Chattopadhyay (2012) is a cross-country analysis and its results may be driven by omitted country specific factors such as gender rights and social norms. Will we get the same result if we concentrate on a more homogeneous population in a particular region and at a particular point of time with some fixed/given gender norms? Our sample of slum dwellers in Delhi who are homogeneous in terms of wealth and living conditions provides an excellent opportunity to answer this question. In our setting, where we get rid of the country/region specific unobserved characteristics by considering a homogeneous group of people, change in the well-being of women relative to men with better objective outcomes (like increase in income and education) is likely to be driven by the rising aspiration level of women. The increased aspiration with the improvement in the objective measures can also be interpreted in terms of the changing internalized norms that allow women to feel greater control over their lives.

Lalive and Stutzer (2010) who find the life satisfaction of women to be higher in the traditional communities compared to liberal communities in Switzerland, explain their finding through the channel of changing internalized norms that result in higher expectation of the women in liberal communities. Mettauci and Lima (2014) also argue in favor of internalized norms as a determinant of women's well-being. They suggest that political rights do not translate into subjective well-being for women unless accompanied by conducive social norms that allow women to feel greater control over their lives.

Before concluding this section, it should be mentioned that like the literature, we have used the terms subjective well-being, well-being, life satisfaction, satisfaction and happiness interchangeably. The literature does argue for a distinction between experienced (emotions, happiness) and remembered well-being (life satisfaction). [1] Our analysis in this chapter is based entirely on remembered evaluation as captured by life satisfaction evaluation.

4.3 The Survey

In 2009-10, the Centre for Global Development Research in New Delhi collaborated with the Planning Commission, Government of India to conduct a slum survey in Delhi. The objective of that survey was to investigate the conditions of life and facilities in low income neighborhoods of Delhi. A report was also published based on their findings in 2011. The report listed 477 slums out of almost four thousand slums located in Delhi. The report also reported the results of survey in 65 slums randomly sampled after stratifying Delhi into 5 zones. These zones are South, East, West, North and Central.

This chapter's sample strategy is based on these 65 slums. We follow the same stratification. However, our survey sampled from four zones (South, East, West, and North) and dropped the central zone, which has very few slums.

For each of the four zones, we have randomly chosen half of the slums surveyed by the Centre for Global Development Research (CGDR). For example, there are 20 slums surveyed by them in the East zone. We randomly choose 10 out of it. Following the same strategy for all other zones, we end up listing total 29 slums for our survey.

[1] Deaton and Stone (2013), Helliwell, Layard and Sachs (2013)

From each of the slums listed in our survey, the households are chosen through the '*k*' th household approach. According to this approach, we start from the northernmost point for each of the slums and then select every '*k*'th household by moving clockwise and right hand side. This is a systematic sampling with every '*k*'th element in the frame is selected, where *k*, the sampling interval is calculated as:

$$k=N/n$$

where *n* is the sample size and *N* is the population size (i.e. sample and population of households in a slum). For each slum, we set a target to interview 35 households (i.e. n=35). We have an idea about the total number of households in each slum. From the above informations, we calculate the sampling interval i.e. *k*. The value of *k* in our study turns out to be 4, 5 or 6 depending on the population of households in each slum. Whenever we have found a household (which is in our sample) locked (i.e. no one is present in the house during the interview), the next one is selected. This can be made clearer with the following example. Consider a particular slum. Suppose the '*p*'th household from the starting point (northernmost point of a slum in our case) is the '*q*'th sample. If the '*p*'th household is found locked during the time of interview, 'p+1'th household is considered as the '*q*'th sample. If 'p+1'th household is also found locked, then 'p+2'th household is chosen and we continue this way. Following this sampling strategy, we have covered the entire area for most of the slums.

From each household, we attempt to interview a female and a male (20 years or above). However, often enough, there is either a female or a male available for interview and not both. We surveyed 989 households across 29 slums during the entire month of March and first week of April, 2016. The total number of respondents surveyed is 1278. Number of households where we interview both a man and a woman are 289. The number of female and male respondents are 771 (60% of the total sample) and 507 (40% of the total sample) respectively. The map in Figure 4.1 shows the coverage of the area for our survey. Table 4.1 and Table 4.2 list all the slums surveyed.

4.4 The Questionnaire

The survey asked the following question to assess life satisfaction: In general terms would you say that you are satisfied with life? There are four choices to answer this question. The choices are 'not at all satisfied' (score 1), 'not very satisfied' (score 2), 'pretty satisfied' (score 3) and 'very satisfied' (score 4). The wording of the question and the choices are exactly the same as the Eurobarometer and Latinobarometer survey. Similar questions are also asked about health and financial satisfaction.

Eurobarometer is a series of public opinion surveys conducted regularly on behalf of the European Commission since 1973. The data from the Eurobarometer survey has been used in many scholarly articles on subjective well-being (for example, Welsch, 2006). Latinobarometer is an annual public opinion survey that involves some 20000 interviews in 18 Latin American countries, representing more than 600 million inhabitants. The South African Quality of Life Trends Study also asks a similar question on life satisfaction. [2]

[2] We also piloted a life satisfaction question based on the Cantril ladder. Respondents were shown a ladder ranging from step 1 to step 10 where step 1 is supposed to represent the worst possible life and step 10 represents the best possible life. Each respondent is asked to locate his/her position on the ladder by choosing any one of the ten steps. Such Cantril ladder measures are used in the World Value Survey as well as in the Gallup World Poll. In our pilot, we found that households did not understand the Cantril ladder well and it required repeated intervention by the survey investigator to elicit a meaningful response. As such investigator intervention might contaminate the response, we did not pursue the Cantril ladder measures any further.

There are five sections in the questionnaire. The first section asks a series of questions regarding the general information. These include several socio-economic and socio-demographic characteristics of the respondent and his/her family members like marital status, age, relation to head, religion, social group, education level, vocational training and working status. This section also asks about the usual place of residence of the respondent and reason for leaving the usual place of residence if he/she is not from Delhi by birth.

The second section consists of questions on income and occupation of the respondent and his/her family members. There are also questions on family expenditure, job satisfaction and the financial satisfaction of the respondent. The third section contains only the life satisfaction question. The fourth section investigates on various public facilities available in a slum.

The fifth and the final section is on health. It has two sub-parts. The first sub-part is on various psychological traits of the respondent. The second one is on physical health. Apart from the question on health satisfaction, there are also questions regarding the disease and the sudden health shocks that the respondent and his/her family members have faced recently (last one year).

4.5 Living Conditions of the Slums in Delhi

In this section, I describe the living conditions in the slums of Delhi. I draw upon the discussion in the literature as well from the survey reported in this chapter.

The key findings from the CGDR report mentioned earlier are the following. Most of the slum dwellers are long term migrants. Migration is valued for access to greater economic opportunities. According to the CGDR report, the increase in average per-capita income due to migration is 317% which is quite large.

Although there is a large financial gain, living in slums is accompanied with poor delivery of public services and poor policing of law and order. Only 43% slums report regular visits by government sweepers. The West zone gets the minimum attention where only 17% slums report visits by the government sweepers. In 54% slums, there is no common garbage facility inside the slum. Only 44% slums have the facility of street lights. About 89% slums are without any government dispensary inside the slums. Only 39.41% slums have been covered by NGO or any other welfare/charitable organization. As regards public order, frequent quarrels and harassment of women is reported by 65% and 80.45% slum dwellers respectively.

The article by Banerjee, Pande and Walton (2012) on Delhi's slums also sheds some light on the deprivation of the slum dwellers. In their survey, 42% of the households do not have access to safe drinking water at home. Only 14% of the households report having a toilet inside their homes. More than half the households (60%) are not connected to sewage systems. 30% of the respondents are dissatisfied with the cleanliness of public toilets. 75% report law and order as a major concern. Theft, gambling and alcoholism are the most frequently cited problems.

Education and health are the two most important components of social sector services. Banerjee, Pande and Walton (2012) find that 48% of the adults living in the slums are illiterate. The health status of the poor slum dwellers is poor as well. 93% of the respondents report a visit to a health clinic or other health facility in the last six months. Although the government hospitals are not very far for most of the slums in Delhi, 60% of the respondents report poor quality of services.

Turning to the economic lives of the poor slum dwellers as reported by Banerjee, Pande and Walton (2012), surprisingly high levels of unemployment is observed (22% of who are 18 and above). Reported per-capita household income indicates significant variability and inequality within the slum population with a Gini coefficient of 0.41and a poverty incidence of over 50% according to the official poverty line.

The average family monthly income and average per-capita monthly income in our sample turns out to be Rs. 12920 and Rs. 2665 respectively. The average per-capita monthly income for the entire Delhi in the year 2014-15 is estimated to be Rs. 21000 (source: Statistics India) which is almost eight times higher than what has been earned by an average earner living in the slums of Delhi. This figure is a clear indicator of the relative deprivation suffered by a poor slum dweller. We also find enough variability in the income distribution. The standard deviation for the monthly family income is Rs. 9500. The coefficient of variation comes out to be 74% and it shows significant inequality in the income distribution.

As has already been mentioned, section 4 of the questionnaire asks a series of questions to a respondent regarding the availability and quality of public facilities in the slums. There are public facilities that are common to everyone who lives in the same slum. Examples of such common public facilities are the ‘availability of a common public toilet inside a slum’, ‘availability of a separate public toilet for women’, ‘distance of that slum to the nearest primary school’, ‘distance of that slum to the nearest government hospital’ and ‘existence of an NGO/charitable organization working in that slum’.

Only 12 out of 29 slums we surveyed are found to be covered by NGO or any other charitable organization providing community services. There are 4 slums without any government dispensary/primary health centres. Health camps and child immunization camps also do not take place very often. Only 17% of the respondents are found to be aware about any health camp in the last six months. Only 23% of the respondents acknowledge the occurrence of any child immunization camp in the past six months.

Some of the public facilities also show variability in terms of availability/quality within a slum. The quality rating is that of the respondents and reflects their experience and perception. Examples include 'functioning of the drainage system in front of the home of the respondent', 'existence of street light near respondent's home and if it exists then whether it works or not', 'cleanliness of the public toilet', 'existence of a dustbin nearby the home of the respondent', 'regularity of the visit by the government sweepers in the slums' and 'availability and quality of water supply'. Table 4.3 reports the percentage of individuals who are dissatisfied due to the unavailability or poor quality of some of these public facilities.

42.1% of the respondents complain about the unavailability of a common dustbin (garbage facility) and 65.4% of the respondents report irregular visits by the government sweepers. These findings are consistent with the Centre for Global Development Research report (2011). Malfunctioning of the drainage system and the absence of the water supply at home has been reported by 56.6% and 35.5% of the respondents respectively. A large fraction of people also express their dissatisfaction regarding the poor quality of drinking water (18%) and unclean public toilets (22.4%). These findings mirror the findings by Banerjee, Pande and Walton (2012).

Safety is another serious concern for the poor slum dwellers. 41% respondents say that they feel unsafe in the slums. The most frequent problems are quarrel, theft and harassment of women. 32% of the respondents feel that the women can't move freely/alone even within the slum. Although the intensity of the safety issue varies across slums, it still remains a general matter to be worried.

Although most of the children are enrolled in school, adult literacy still remains a great problem (38% of those who are interviewed are completely illiterate i.e. can't even sign). The picture on health is not that bright as well. 24% of the respondents report to suffer from some disease (like heart disease, diabetes and high blood pressure).

Therefore from all the discussions above, it is clear that the poor slum dwellers suffer enormous deprivation in terms of standard of living and public facilities they receive. Does it adversely affect their subjective well-being? We move into the next subsection to answer this question.

4.6 Well-Being Measures

When we consider the life satisfaction question, only 11.35% of the respondents report that they are 'not at all satisfied' and 42.49% of the respondents report that they are 'not at all satisfied' or 'not very satisfied'. The modal response is 'pretty satisfied' (reported by 40.47% respondents). The results are displayed in Figure 4.2. Therefore, the reported life satisfaction score is on the high side and there is a clear resemblance between our finding and what has been found by Banerjee, Deaton and Duflo (2004) in rural Udaipur. As remarked by those authors, such life satisfaction levels are in line with those reported in the rich countries. Nonetheless, as Table 4.4 reports, the variability in life satisfaction scores is also high.

The finding on health satisfaction score is quite similar to the life satisfaction score. Only 8.37% report that they are 'not at all satisfied' with their health status. Only 32.47% are found to be 'not at all satisfied' or 'not very satisfied' with their health status. Just like life satisfaction, the modal response turns out to be 'pretty satisfied' (43.27% report that). These findings are displayed in Figure 4.3. Once again, our findings are similar to that in Banerjee, Deaton and Duflo (2004) and Case and Deaton (2005). Like in the case of life satisfaction, health satisfaction scores also exhibit considerable variability (Table 4.4).

Turning to the self-reported financial satisfaction level, we find more people reporting low value of financial satisfaction compared to life and health satisfaction. We find 46.71% respondents who are either 'not all satisfied' or 'not very satisfied' with their financial status. The percentage of people who say 'very satisfied' (i.e. the highest score) is only 6.26%. This is much less compared to the percentage of people who report the highest level of life and health satisfaction (16.9% and 23.9% respectively). Again this finding is consistent with Banerjee, Deaton and Duflo (2004) and Case and Deaton (2005).

But the financial satisfaction as reported by the slum dwellers of Delhi is much more than what is found by Banerjee, Deaton and Duflo (2004) and Case and Deaton (2005). Banerjee, Deaton and Duflo (2004) find that the modal response for the self-reported financial status is at the bottom rung of a ten rung ladder and more than 70 percent of people live in households that self-reported themselves in the bottom three rungs of a ten rung ladder. The variability in the self-reported financial satisfaction status is shown in Figure 4.4 and Table 4.4.

From the findings above, it is evident that, much like the rural sample in Banerjee, Deaton and Duflo (2004) and Case and Deaton (2005), urban slum-dwellers also report life and health satisfaction scores that are on average higher than financial satisfaction scores. These findings are consistent with the earlier literature which conjectures that poor people may be adapted to the life and sickness they experience. The same adaptation mechanism appears to be at work even in an urban setting.

On the other hand, just like in the earlier literature, self-reported values for financial satisfaction tend to be markedly lower. It implies that the poor people are more concerned about their financial status compared to health or life in general. Regardless of the pattern of reporting (over reporting or under reporting), well-being measures are found to have sufficient variability. What explains this variation? To answer this question, we proceed to the next subsection.

4.7 Bivariate Correlations

In this section, I report on the bivariate correlations of life satisfaction with other variables. Life satisfaction score is positively correlated with the other two self-reported measures i.e. health satisfaction and financial satisfaction. The positive association between financial satisfaction and life satisfaction is illustrated in Figure 4.5. As financial satisfaction increases from level 1 to 4, the corresponding increase in average life satisfaction score is almost 84%. The Spearman rank correlation coefficient between life satisfaction and financial satisfaction turns out to be 0.45 and it is statistically significant. The fact that financial satisfaction and life satisfaction is positively correlated supports Banerjee, Deaton and Duflo (2004).

Self-reported health satisfaction turns out to be another important correlate of life satisfaction. Figure 4.6 shows the positive association between these two. As health satisfaction increases from level 1 to 4, the corresponding increase in average life satisfaction score is almost 56%. The rank correlation coefficient between self-reported life and health satisfaction score turns out to be 0.35 and it is statistically significant.

Thus, I find the association between financial satisfaction and life satisfaction to be stronger than the association between health satisfaction and life satisfaction. In addition, I find that for each level of increase in financial satisfaction, increase in life satisfaction is much larger compared to each level of increase in health satisfaction. For example when health satisfaction score increases from 1 to 2, the corresponding increase in the average life satisfaction is 19.6%. On the other hand if financial satisfaction score increases from 1 to 2, the corresponding increase in life satisfaction is 24.6%. Therefore the impact of financial satisfaction is larger. I get similar findings as we move from level 2 to 3 or level 3 to 4. Table 4.5 and 4.6 show the findings.

As shown in Figure 4.7, financial satisfaction and monthly family income are positively correlated. Therefore, we also expect a positive correlation between monthly family income and life satisfaction. Figure 4.8 shows that this does obtain. Our finding supports all the earlier literature that discusses the positive cross-sectional relationship between income and life satisfaction (like Easterlin (2001), Argyle (1999), Diener (1984)).

The Pearson correlation coefficient between life satisfaction and financial satisfaction (0.45) turns out to be much larger compared to the correlation coefficient between monthly family income and life satisfaction (0.1) (both the correlations are statistically significant). Therefore self-reported financial status turns out to be a stronger correlate than income and this supports what has been found by Banerjee, Deaton and Duflo (2004). Thus, financial satisfaction has a larger impact on life satisfaction than either health satisfaction or income.

Education level seems to be an important determinant of life satisfaction. I define a person ‘educated’ if he/she has at least the secondary education. The average life satisfaction score for the educated cohort is 2.76 which is almost 7% higher than the average life satisfaction score of the uneducated cohort (which is 2.58). Figure 4.9 shows the same finding in a bar diagram. Age is found to be negatively correlated with the self-reported life satisfaction measure (see Figure 4.10). Figure 4.10 fits a linear line. Even if we fit a quadratic regression line, we don't get the U shaped happiness-age relationship that is widely found in the literature (like Blanchflower and Oswald (2008)). Most of the respondents in our sample are found to be married. The average life satisfaction of the married respondents (average life satisfaction score 2.68) turns out to be slightly higher than the average life satisfaction of the unmarried, widowed or separated (average life satisfaction score 2.6).

The survey questionnaire asks each respondent about the possession of a refrigerator and the possession of a two wheeler (scooter or motor bike). From the CGDR Survey Report (2011), we find sufficient variation in the possession of these two assets among the slum dwellers in Delhi (other assets like television, radio, mobile phone are owned by almost everyone and assets such as computer, car, or washing machine are owned by too few people and hence lack variation in the possession) . Therefore I have specifically chosen these two assets to see whether their possession have any impact on the subjective life evaluation of the poor slum dwellers. Our survey data also finds reasonable variation in the possession of these two assets. Around 46% of the respondents report ownership of a refrigerator while 19% of the respondents own a two wheeler. Possession of these assets is found to increase the subjective well-being (see Table 4.7).

Mental health/psychological traits have been considered important determinants of happiness in the literature (Carbonell and Gowdy (2007)). Our survey asks a host of questions regarding the psychological traits of the respondents. The average life satisfaction of those who feel lonely is 2.26 in our data. This is much less than the average life satisfaction of those who don't feel lonely which is 2.78. The average life satisfaction score for the respondents who 'don't feel stressed at all', 'often feel stressed but not always' and 'always feel stressed' are 2.98, 2.65 and 2.09 respectively. Therefore stress is found to be negatively associated with life satisfaction and this is consistent with the findings by Banerjee, Deaton and Duflo (2004) in rural Udaipur.

Finally, I focus on the effects of the public facilities and services available to the poor slum residents on their self-reported well-being. As discussed earlier, the availability and quality of some of the facilities vary within a slum. A nice example is the functioning of drainage system. It is found from our survey data that those who have complained about the drainage function (i.e. it does not work properly in front of their home) are also the ones who report lower life satisfaction scores (see Table 4.8). The higher average life satisfaction score for those who have reported proper functioning of drainage system is indicative of the influence of public facilities on the reported life satisfaction score. This finding is at variance with Banerjee, Deaton and Duflo (2004) who find the availability and quality of the public facilities to be uncorrelated to happiness in their data.

4.8 Do Men and Women Respond Differently?

This section considers the question whether the correlates of life satisfaction (reported in the earlier section) differ between men and women.

Self-reported financial status turned out to be an important correlate of life satisfaction in the last section. The positive association between the financial satisfaction and life satisfaction is found to be almost same for men and women (see Figure 4.11). The Spearman rank correlation coefficient between life satisfaction and financial satisfaction is also identical (0.43 for men and 0.46 for women).

The correlation between life satisfaction and health satisfaction turns out to be slightly higher for men compared to women (Spearman rank correlation coefficient is 0.41 for men and 0.32 for women). Figure 4.12 illustrates the positive association between life satisfaction and health satisfaction for men and women. For low level of health satisfaction, life satisfaction turns out to be higher for women. But life satisfaction for men and women converges for higher level of health satisfaction.

Health satisfaction and financial satisfaction are subjective measures. Do the life satisfaction response between men and women vary with respect to objective measures? I consider income first of all. The gradient of the income-life satisfaction line turns out to be steeper for men relative to women (see Figure 4.13). It implies that the impact of family income on life satisfaction is larger for men relative to women.

For education, we have similar findings. It turns out that as we move from uneducated to educated cohort; self-reported well-being for women increases less than men i.e. the ratio of women's average life satisfaction to men's average life satisfaction is lower for the educated cohort compared to uneducated cohort. This is displayed in Table 4.9.

The finding that relative (to men) well-being of women fall with income and education is a cross sectional counterpart to the finding of declining female happiness in United States (Stevenson and Wolfers (2009)). Possible explanations can be many. Women earn only 18% of the total income in our sample and their command of resources may not increase commensurately with household income (Motiram and Osberg (2010) discuss the gender inequalities in tasks and the constraints faced by urban women in labor market participation). Higher incomes and higher education may also increase aspirations. The finding is different from that of Graham and Chattopadhyay (2012). As argued earlier, our sample is unlikely to contain variation in social and cultural factors and hence the correlations reported here are free of their effect.

Our finding resembles Lalive and Stutzer (2010) who find the life satisfaction of women to be higher in the traditional communities compared to liberal communities in Switzerland and explain their finding as a result of higher expectation of the women in liberal communities.

Gender differences in correlates could also possibly arise from differences in preferences. An instance of this is that men and women seem to value the access to government subsidized foodgrain, sugar and kerosene (used for cooking and/or lighting) differently. The access is determined by the possession of a 'ration' card. 15% of our sample do not possess a ration card. The average life satisfaction score of the women who possess a ration card is higher than those who don't possess it. But this does not turn out to be true for men and for the overall sample (see Table 4.10).

Other correlates discussed in the last section are similar in their impacts between men and women. These include age, the presence of a functioning drainage system, possession of a refrigerator or a two wheeler, and the states of 'not feeling lonely' or 'not feeling stressed'.

Marriage brings more happiness in our data for women but not for men. This finding is also consistent with Graham and Chattopadhyay (2012). A probable reason behind the positive happiness-marriage relationship for women is some kind of social respect and security that a woman usually gets after her marriage.

4.9 Ordered Logit Model

This section investigates the correlations in a regression framework controlling for all variables. The outcome variable of interest is the self-reported four scale life satisfaction score by the respondents. This is a categorical variable with monotonic ordering. Hence the ideal regression specification is the ordered logistic (or the ordered probit regression model). The ordinary least square regression can be used as an alternative specification.

When there are two categories of the dependent variable (say success and failure) and the probability of success is denoted as *P*, the regression model is defined in the following way:

(1) $$P_i = E(Success = 1 \mid X) = \beta_1 + \beta_2 X_i$$

where the subscript '*i*' in equation (1) denotes the '*i*'th observation. $P_i = E(Success = 1 \mid X)$ means that the '*i*'th observation report success with a probability *P*. X stands for independent variable or the predictor of success.

Under logistic distribution, equation (1) is written as

(2) $$P_i = \frac{1}{1 + e^{-Z_i}} = \frac{e^{Z_i}}{1 + e^{Z_i}}$$

where $Z_i = \beta_1 + \beta_2 X_i$.

If P_i denotes the probability of success, then the probability of failure is written as

(3) $$1 - P_i = \frac{1}{1 + e^{Z_i}}$$

Therefore we can write

(4) $$\frac{P_i}{1-P_i} = e^{Z_i}$$

Now $\frac{P_i}{1-P_i}$ is simply the odds ratio in favor of success i.e. the ratio of probability of success to the probability of failure. Thus if P_i =0.8, it means that odds are 4 to 1 in favor of success.

Taking the log of (4), we obtain

(5) $$L_i = \ln(\frac{P_i}{1-P_i}) = Z_i = \beta_1 + \beta_2 X_i$$

L is the log of the odds ratio. It is also called the logit. It is not only linear in *X* but also linear in parameters. Therefore the coefficient β_2 can be interpreted as the change in the log of the odds in favor of success as *X* increases.

In our case, the categorical variable life satisfaction has more than two categories. In the four scale life satisfaction score, there are four categories. These four categories are 'not at all satisfied' (score 1), 'not very satisfied'(score 2), 'pretty satisfied'(score 3) and 'very satisfied' (score 4). Any respondent can answer one of these four choices. I denote the probabilities of reporting score1, score2, score3 and score 4 as P_1, P_2, P_3 and P_4. These probabilities sum up to 1. We can compare these four categories with the two category case as described earlier by assuming the probability of reporting 'very satisfied' (score 4) as *P* i.e. $P_4=P$. Therefore the sum of the probabilities of reporting the other three choices becomes P_1+P_2+ P_3 =1-*P*. The logit equation is now

$$L_i = \ln(\frac{P_i}{1-P_i}) = \ln(\frac{P_{4i}}{P_{1i}+P_{2i}+P_{3i}}) = \beta_1 + \beta_2 X_i \tag{6}$$

Equation (6) is an ordered logistic equation where the coefficient β_2 is interpreted as the increase in the log of odds of being in highest level of satisfaction i.e. 'very satisfied' when the independent variable *X* increases by one unit. In equation (6), we have considered only one independent variable/predictor. Replacing the single independent variable by a vector of independent variables **X** and adding a stochastic disturbance term ε_i in equation (6) we get,

$$\ln(\frac{P_{4i}}{P_{1i}+P_{2i}+P_{3i}}) = \beta_1 + \beta_2 \mathrm{X}_i + \varepsilon_i \tag{7}$$

Instead of considering the probability *P* as the probability of reporting 'very satisfied', we can also consider it as the sum of the probabilities of reporting 'very satisfied' and 'pretty satisfied' (i.e. P_3 +P_4). Then the probability 1-*P* represents the sum of the probabilities of reporting 'not at all satisfied' and 'not very satisfied' (P_1 +P_2). The corresponding ordered logistic regression equation can be written as

$$\ln(\frac{P_{4i}+P_{3i}}{P_{1i}+P_{2i}}) = \beta_1^* + \beta_2^* \mathrm{X}_i + \varepsilon_i \tag{8}$$

In equation (8), the '*j*'th component of the coefficient vector β_2^* i.e. β_{2j}^* is interpreted as the increase in the log of odds in favor of being 'very satisfied' or 'pretty satisfied' when the independent variable X_j increases by one unit. On the other hand in equation (7), the '*j*'th component of the coefficient vector β_2 i.e. β_{2j} is interpreted as the increase in the log of odds in favor of being 'very satisfied' when the independent variable X_j increases by one unit. The important assumption of the ordered logistic regression is that $\beta_{2j} = \beta_{2j}^{*} \; \forall j = 1,2,...,n$ (where n is the total number of independent variables in the regression model). This assumption is called the proportional odds assumption. The proportional odds assumption says that equation (7) and equation (8) are equivalent ways to represent an ordered logistic regression specification and the magnitudes and statistical significance of the coefficients from both the specifications are identical.

Similarly, the ordered logistic regression can also be written as

$$\ln\left(\frac{P_{4i} + P_{3i} + P_{2i}}{P_{1i}}\right) = \beta_1^{**} + \beta_2^{**}\mathrm{X}_i + \varepsilon_i \quad (9)$$

β_{2j}^{**} captures the increase in the log of odds in favor of being 'very satisfied', 'pretty satisfied' or 'not very satisfied' when the independent variable X_j increases by one unit. Again the proportional odds assumption implies that $\beta_{2j} = \beta_{2j}^{*} = \beta_{2j}^{**} \; \forall j = 1,2,...,n$ (n is the number of independent variables used in the regression). As an alternative specification, one can also use the OLS regression and compare the results that we find from the ordered logistic regression.

4.10 Variables in the Regression

As has already been mentioned, equation (7), equation (8) and equation (9) are equivalent ways to estimate the ordered logistic model because of the proportional odds assumption. The dependent variable in our regression equation is the self-reported life satisfaction score. We are particularly interested in those independent variables which are shown to have some correlation with the self-reported life satisfaction in the last section. These are financial satisfaction, health satisfaction, monthly family income, education, age, marital status, possession of assets like refrigerator and two wheeler (scooter or motorbike), loneliness, stress, functioning of the drainage system and possession of a ration card.

Apart from these variables, I have also controlled for a host of other factors. These include individual/respondent specific control variables, household/family specific control variables, mental health/psychological traits of the respondent and availability and quality of public facilities. Table 4.11, Table 4.12, Table 4.13 and Table 4.14 describe the independent variables of interest and other control variables along with the summary statistics (mean and standard deviation).

The independent variables are of two types; continuous variables and binary variables. The mean of a binary variable provides the proportion of people reporting a particular category (say yes/no).

4.11 Regression Results

Firstly, regressions are run for the entire sample with an intercept dummy for gender (women). Table 4.15 shows the results of the basic ordered logistic regression for the entire sample. Three columns are shown in the regression Table 4.15. In the first column, the regression is run using the slum fixed effects. In the second column, slum fixed effects are replaced by the categorical variables on slum specific common public facilities. Although the slum specific common facilities are same for everyone in a slum, the reported response may vary across respondents within a slum depending on whether he/she is aware of a facility. The third column controls for both i.e. slum fixed effects as well as slum specific common public facilities. Therefore the third column provides the most robust specification. As already mentioned, I report the coefficients of only those independent variables that are found to be correlated with the reported life satisfaction score (as shown in section 4.7). Rests of the independent variables are used as controls and I do not report their coefficients (these other control variables have been defined in Table 4.11, Table 4.12, Table 4.13 and Table 4.14).

Among the reported coefficients, health satisfaction, logarithm of family income, marriage, malfunctioning of drainage system, loneliness and stress turns out to be significant correlates of life satisfaction at 1% level. Possession of refrigerator turns out to be significant at 5% level and education is significant at 10% level. Possession of ration card, possession of two wheeler, gender and age turns out to be insignificant. The results are robust across all three specifications.

The exponential of the regression coefficients in Table 15 estimates the change in the odds ratio in favor of high level of reported well-being versus low level of reported well-being. For example, if family income increases by 1%, the odds of high level of life satisfaction versus low level of life satisfaction is greater by 48(computed from column 3 of Table 15). As one becomes educated, the odds of reporting high level of life satisfaction is greater by 14.59(again computed from column 3 of Table 15). These large odd ratios suggest that the increase in income or education substantially improve the reported life satisfaction. Similarly, we can compute the change in the odds-ratio for the other predictors of reported well-being.

In the regressions of Table 4.15, I did not include financial satisfaction as an independent variable. The reason is the high collinearity between monthly family income and financial satisfaction. In order to see the impact of financial satisfaction on life satisfaction, I replace monthly family income by financial satisfaction and run the same regressions. The results are reported in Table 4.16. Just like Table 4.15, Table 4.16 has three different columns with three different specifications. These three specifications are same as Table 4.15.

In Table 4.16, I only report the coefficients of health satisfaction and financial satisfaction. The coefficients of the rest of the variables are not reported as these are similar to in magnitude to the coefficients in Table 4.15. The main reason for including the coefficients of health satisfaction along with the coefficient of financial satisfaction is to compare their impact on subjective life evaluation. All three specifications find positive and statistically significant coefficients for both health satisfaction and financial satisfaction (significance is at 1% level for both the variables). But the impact of financial satisfaction turns out to be stronger and this is consistent with what was seen in the earlier section.

Table 4.15 and 4.16 showed the results of an ordered logistic regression. An alternative specification should be the ordered probit regression. I have run the ordered probit regressions as well and the sign and the statistical significance for the coefficients almost turn out to be the same. Therefore, the results of the ordered probit regression are not reported here. Instead, I report the results of an OLS regression where the self-reported life satisfaction is considered as a continuous variable. The results are shown in Table 4.17.

The sign and the statistical significance of the coefficients in the OLS regressions are almost identical with the ordered logistic model (only some minor changes in the level of significance is noticed for a few variables). The specification in three columns of Table 4.17 is equivalent to Table 4.15 and Table 4.16.

Gender Effects

Does the impact of these independent variables vary between men and women? In order to see that, I interact all the independent (slope) variables in the regression equation with the female dummy. This female dummy takes a value of 1 if the respondent is a female and 0 otherwise. This interaction term captures the differential impact across gender for all the independent variables. The result is shown in Table 4.18. I only report the coefficients of the interaction terms. The first column shows the results of an ordered logistic model and the second column shows the result of an OLS model. The slum fixed effects are controlled in both the specifications.

The interaction term between the dummy 'educated' and dummy 'female' turns out to be negative and significant at 5% level. It implies that compared to an educated man, log of odds in favor of reporting a higher level of life satisfaction is significantly less for an educated woman. This finding is robust to both the specifications (both ordered logistic and least square specification) and supports what we have found in terms of bivariate correlation.

A probable reason as mentioned earlier for the above finding can be the rising aspiration level of women with education and is consistent with Lalive and Stutzer (2010). Expectations grow with education. If the societal norms are not good enough for women, then they can't fulfill those aspirations. This may have some depressing effects on the subjective life evaluation of the educated women (same kind of argument has been made by Mettauci and Lima, 2014).

Apart from education, the interaction term between age and female dummy turns out to be negative and significant at 10% level for the OLS specification. The interaction effect between squared age and female dummy is negative and significant at 10% for both ordered logit and OLS specification. The only other independent variable which shows statistically significant gender differential impact is loneliness. The coefficient of the interaction term turns out to be negative and significant at 1% level. This shows that the positive effect of not feeling lonely on life satisfaction is significantly less for women compared to men. Although income and possession of ration card have shown interesting gender varying bivariate correlation, the differential gender effect does not turn out to be statistically significant when we control for other factors. None of the other interaction terms turn out to be statistically significant.

4.12 Concluding Remarks

There is not a substantial amount of literature that looks into the subjective life evaluation of the poor people. This chapter is a step towards that relatively unexplored field of research. The papers by Banerjee, Deaton and Duflo (2004) and Deaton and Case (2005) are among the very few papers that have investigated the life satisfaction of the poor. They find that poor tend to report high levels of happiness inspite of living a distressful life and facing a whole lot of inconveniences every day. They consider the adaptation to the life circumstances as the principal factor behind their finding. Their research was confined to rural areas only.

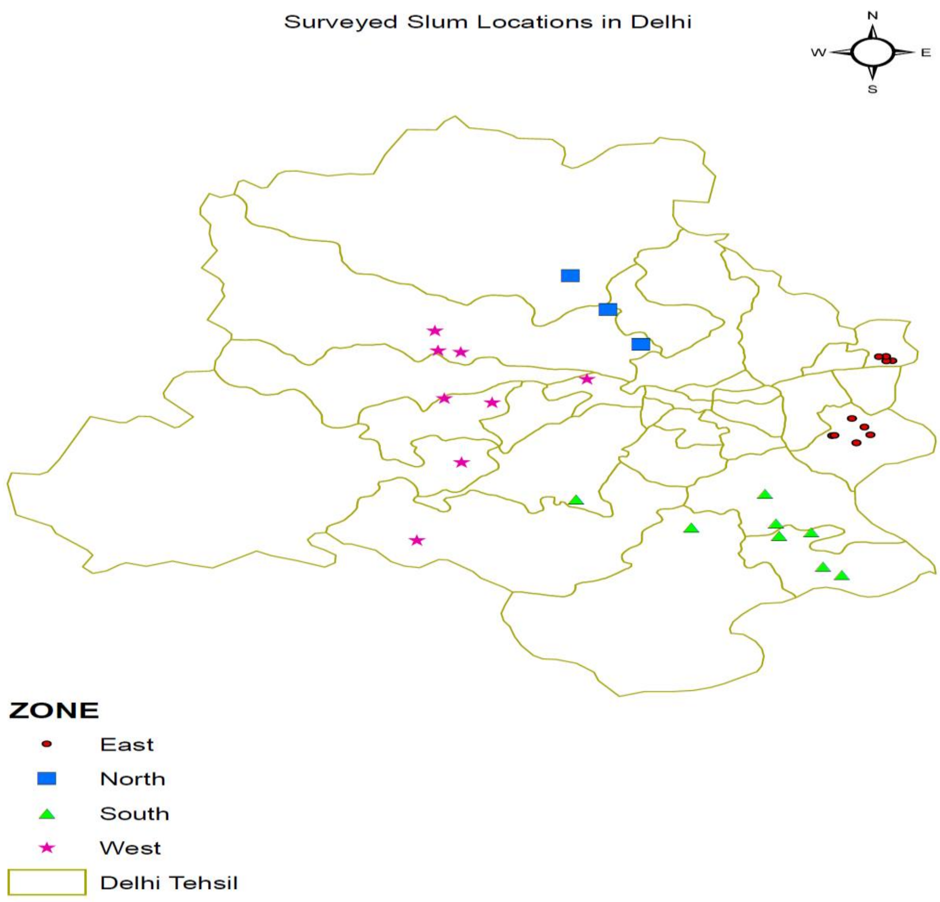
This chapter has extended their research to an urban setting where adaptation mechanism is, *a priori*, less likely. Nonetheless, our sample of poor urban households seems to evaluate life and health satisfaction in ways similar to the poor rural households studied in the literature. We also find the reported life satisfaction of the urban poor to be on the higher side. But there are more people who complain about their financial status instead of complaining about the overall life. Therefore financial concern is very important for the poor. All these findings are consistent with Banerjee, Deaton and Duflo (2004) and Deaton and Case (2005) and hence indicates that an urban environment does not introduce too much change. The adaptation mechanism prevails in urban areas as well.

Regardless of the pattern of reporting (over reporting or under reporting), well-being measures are found to have sufficient variability and it is important to explain this variability. The analysis finds a host of factors which are correlated to the self- reported well-being measures. These include financial satisfaction, health satisfaction, income, marriage, age, education, possession of assets (possession of refrigerator and two wheeler), public facilities (like functioning of drainage system) and mental health/psychological traits (like loneliness and stress). Some of these correlations persist in a regression framework even after controlling for other factors.

Our point of departure from the literature on life satisfaction among the poor is that this chapter also looks at how life satisfaction differs between men and women. If society offers different opportunities to men and women and if the preferences vary across gender, we may expect variation across gender of the relative importance of the factors that are correlated with the self-reported well- being measures and that's what we have found from our data. Among all the gender varying correlates, education turns out to be most interesting. As we move from uneducated to educated cohort, increase in the life satisfaction of men turns out to be higher relative to women. It implies that with the increase in education, reported well-being of women relative to men falls. The regression analysis confirms the differential effect of education on life satisfaction across gender to be statistically significant i.e. the reported life satisfaction score is significantly higher for an educated man compared to an educated woman.

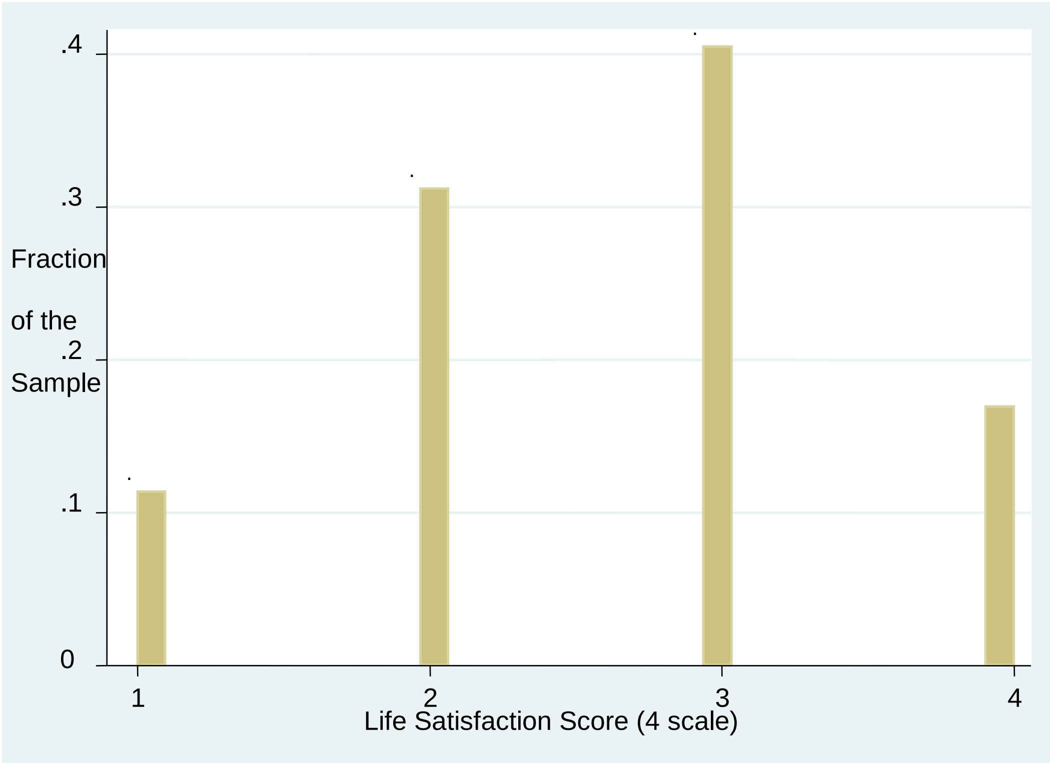
This finding can be considered as a cross sectional counterpart of what has been found by Stevenson and Wolfers (2009) regarding the declining female happiness in United States and the industrialized nations in Europe during 1970-2005 despite improvement in the objective measures in the same period of time(in our case, education is an objective measure of well-being). The chapter's finding also resembles Lalive and Stutzer (2010) who find the life satisfaction of women to be higher in the traditional communities compared to liberal communities in Switzerland and explain their finding as a result of higher expectation of the women in liberal communities. A similar explanation in our case would suggest while education increases the aspirations of women and men, societal norms constrain women more than men. The challenge for future research is to investigate the foundations of such explanations and to locate life satisfaction within social norms.

Figure 4.1:- Surveyed Slum Locations in Delhi



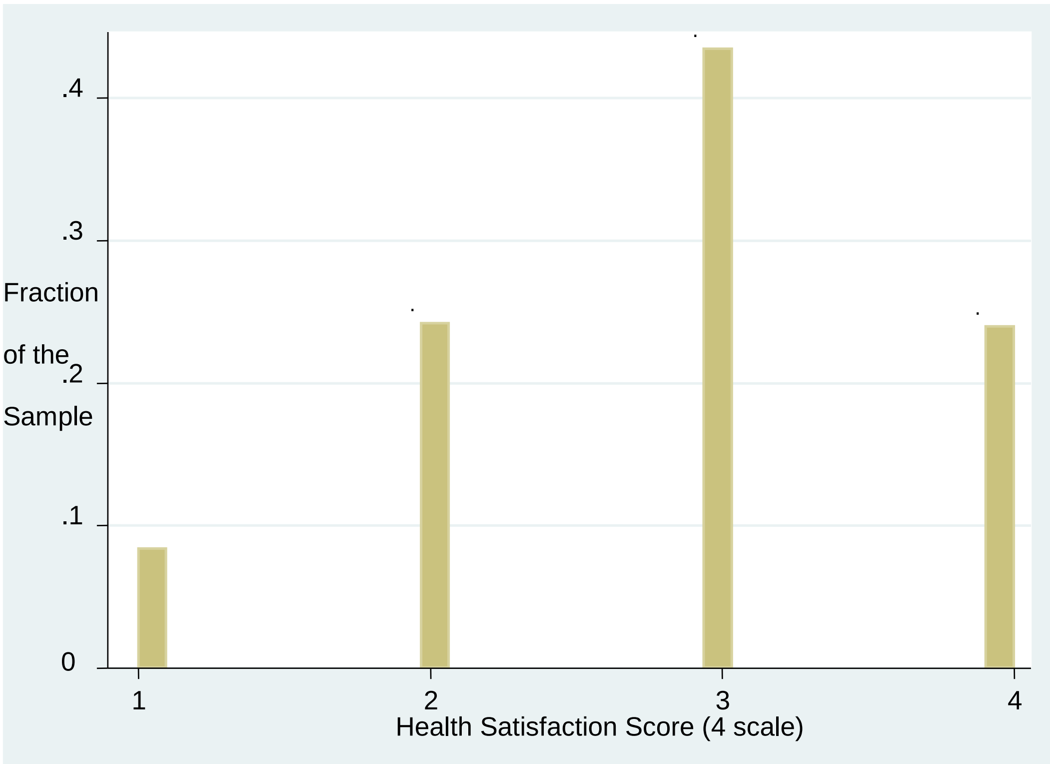
Note: - Drawn through the Geographical Information System using the latitudes and longitudes of the slums

Figure 4.2:- Fraction of People Reporting Different Level of the Life Satisfaction Score



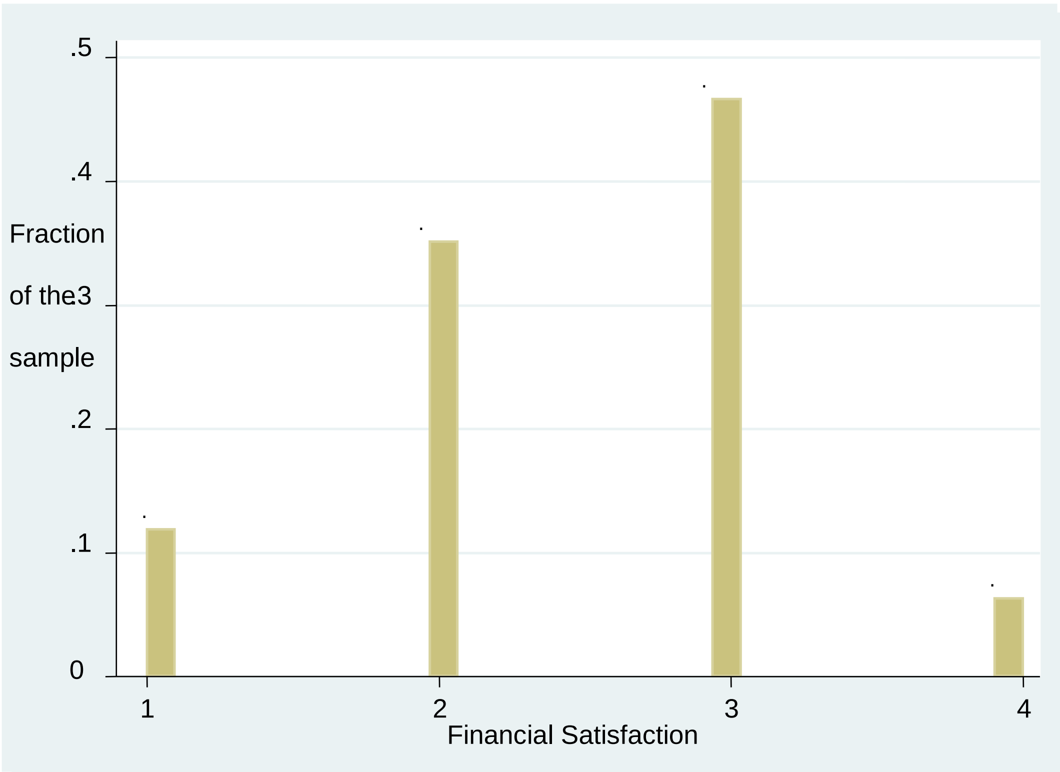
Note: - Drawn from our own survey data. The horizontal axis represents different levels of life satisfaction. The vertical shows the fraction of the respondents reporting those values.

Figure 4.3:- Fraction of People Reporting Different Level of the Health Satisfaction Score



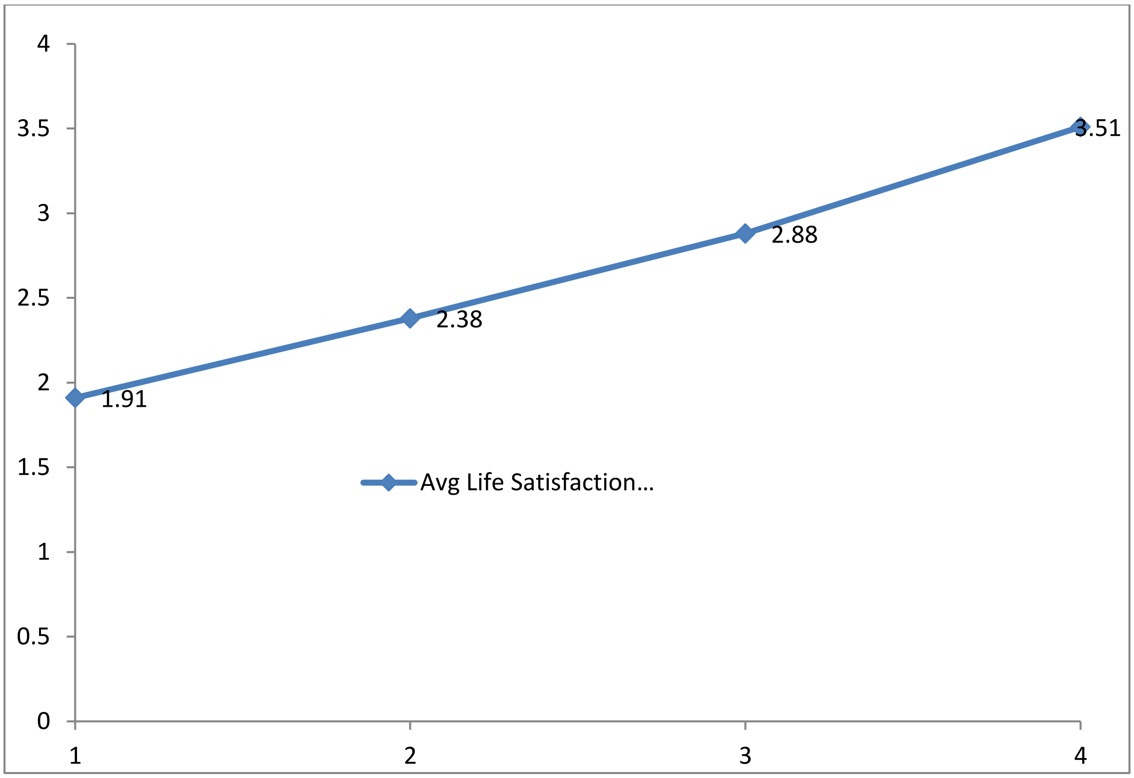
Note: - Drawn from our own survey data. The horizontal axis represents different levels of health satisfaction. The vertical shows the fraction of the respondents reporting those values.

Figure 4.4:- Fraction of People Reporting Different Level of the Financial Satisfaction Score



Note: - Drawn from our own survey data. The horizontal axis represents different levels of financial satisfaction. The vertical shows the fraction of the respondents reporting those values.

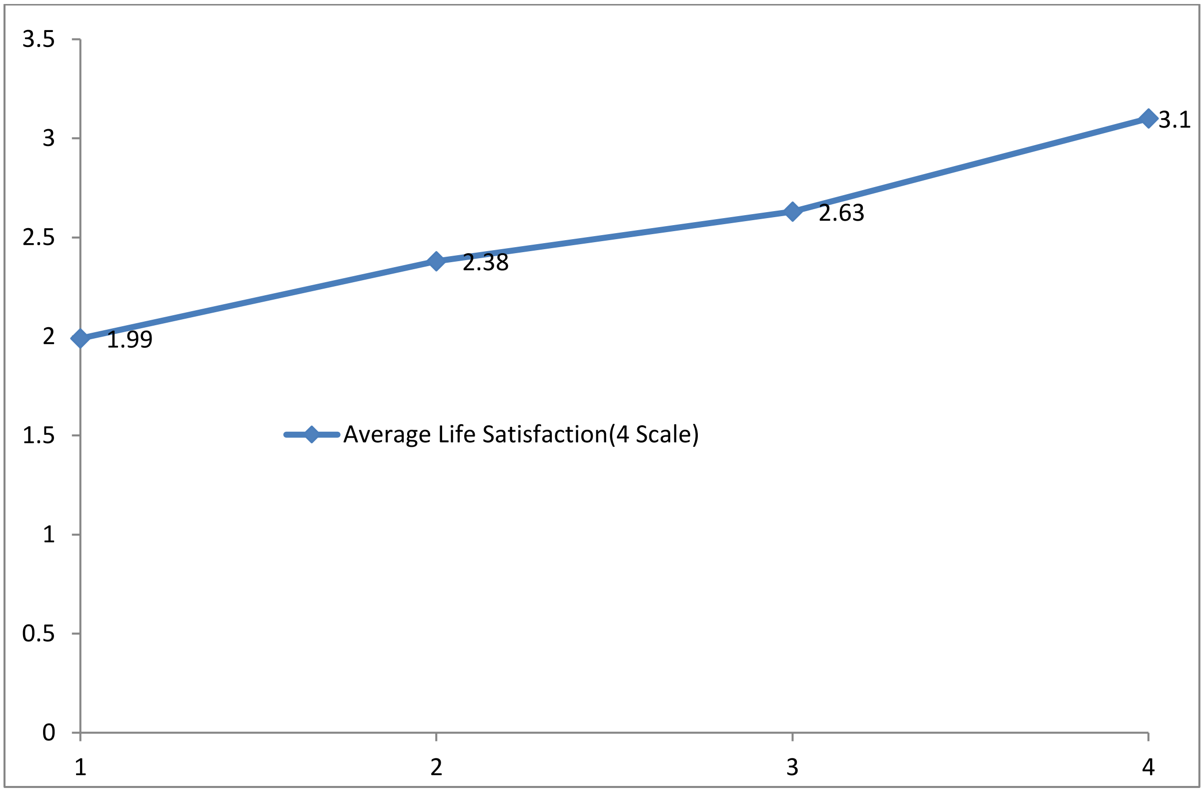
Figure 4.5:- Average Life Satisfaction Score for Different Levels of Financial Satisfaction



Financial Satisfaction (horizontal axis)

Note: - Drawn from our own survey data. The horizontal axis represents different levels of financial satisfaction. The vertical shows the average life satisfaction score for each level of financial satisfaction.

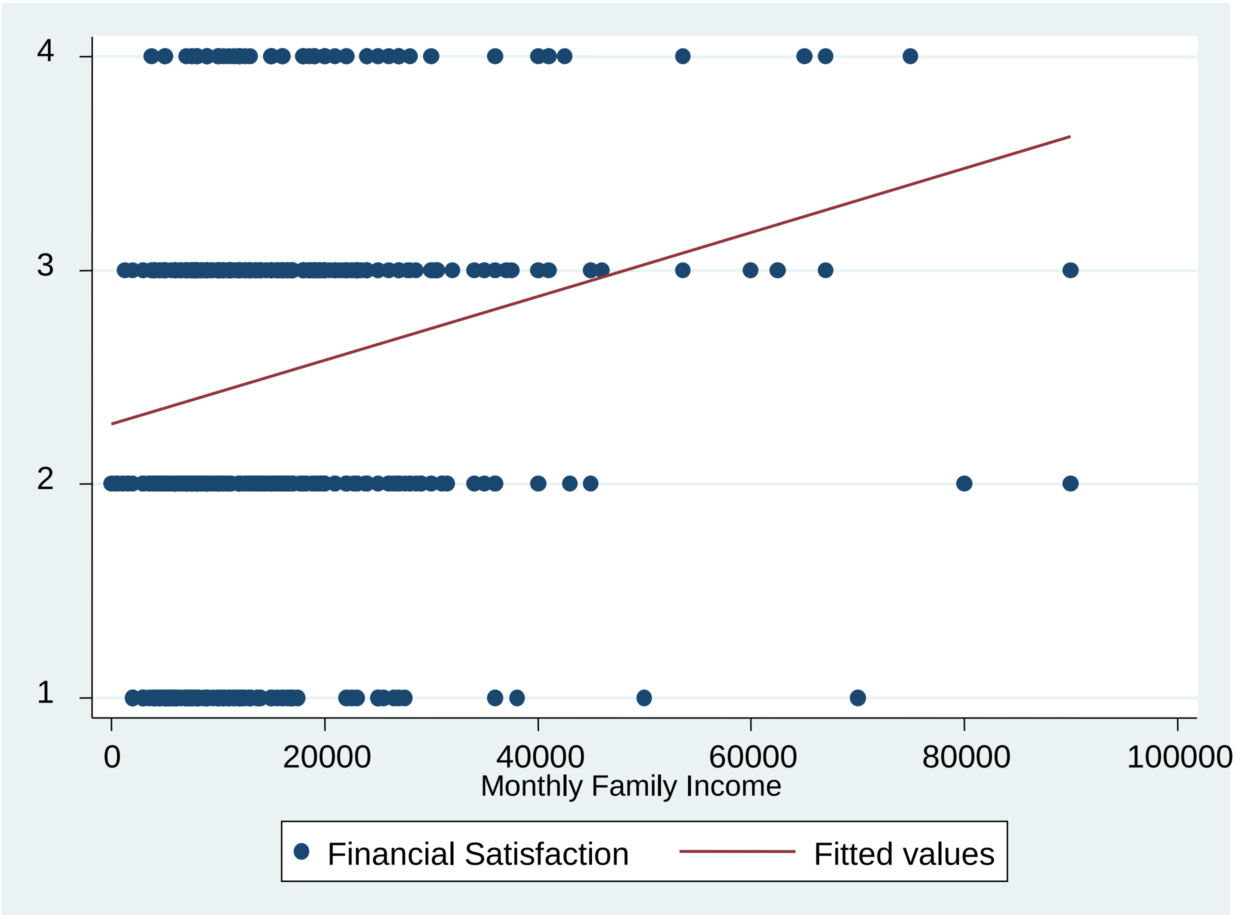
Figure 4.6:- Average Life Satisfaction Score for Different Levels of Health Satisfaction



Health Satisfaction (horizontal axis)

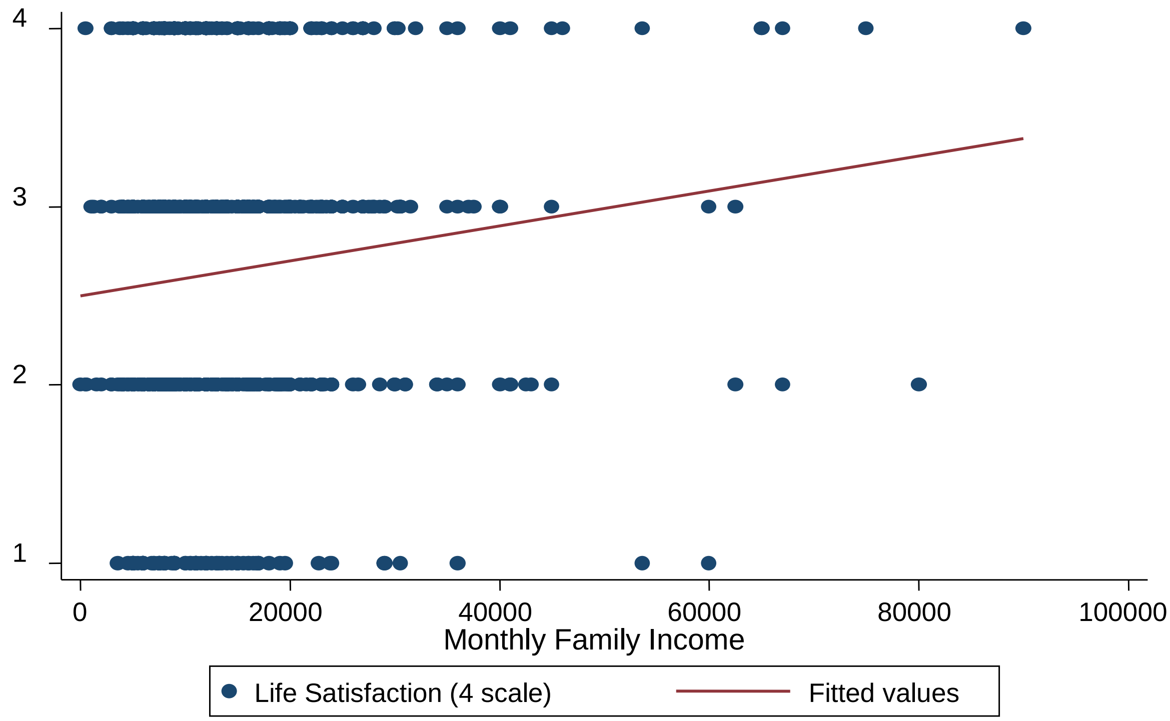
Note: - Drawn from our own survey data. The horizontal axis represents different levels of health satisfaction. The vertical shows the average life satisfaction score for each level of health satisfaction.

Figure 4.7:- Financial Satisfaction and Monthly Family Income are Positively Correlated



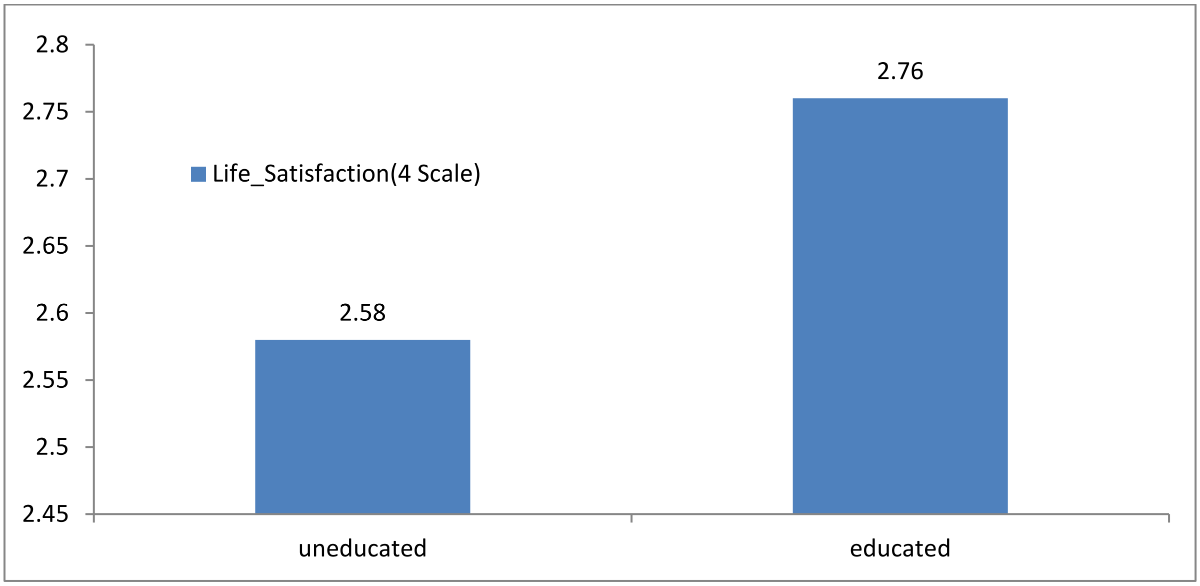
Note: - Plotted using our own survey data. Bivariate linear regression lines are fitted to the above plot.

Figure 4.8:- Life Satisfaction and Monthly Family Income are Positively Correlated



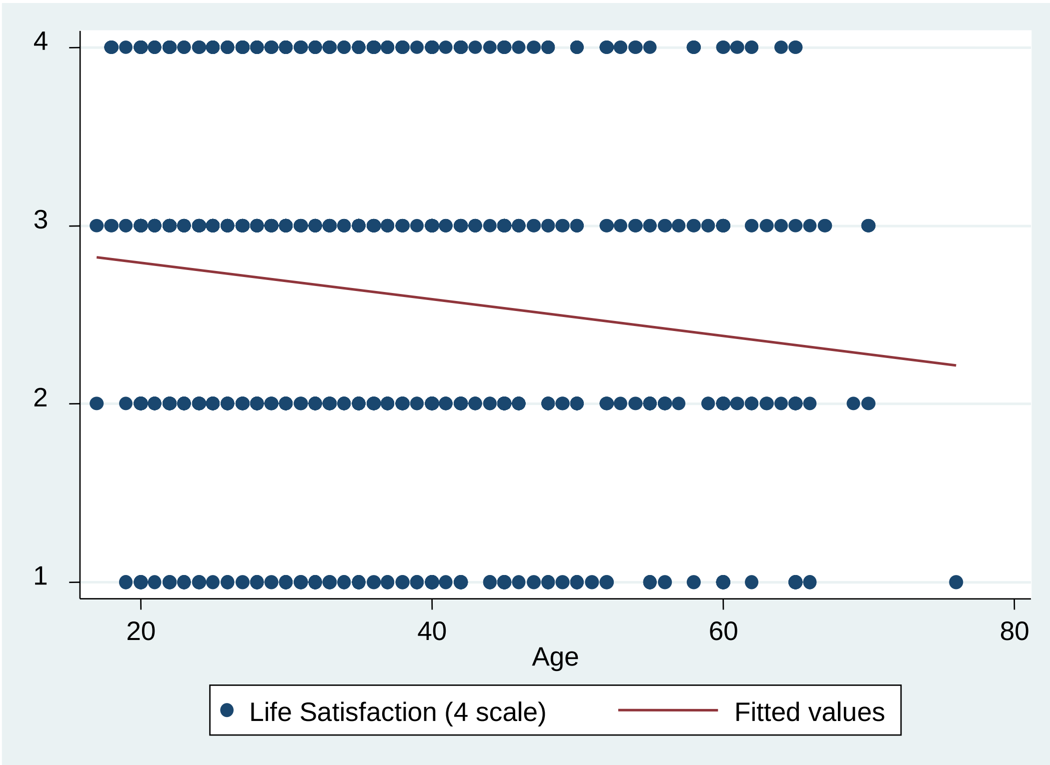
Note: - Plotted using our own survey data. Bivariate linear regression lines are fitted to the above plot.

Figure 4.9:- Average Life Satisfaction Score for 'Educated' and 'Uneducated' Cohort



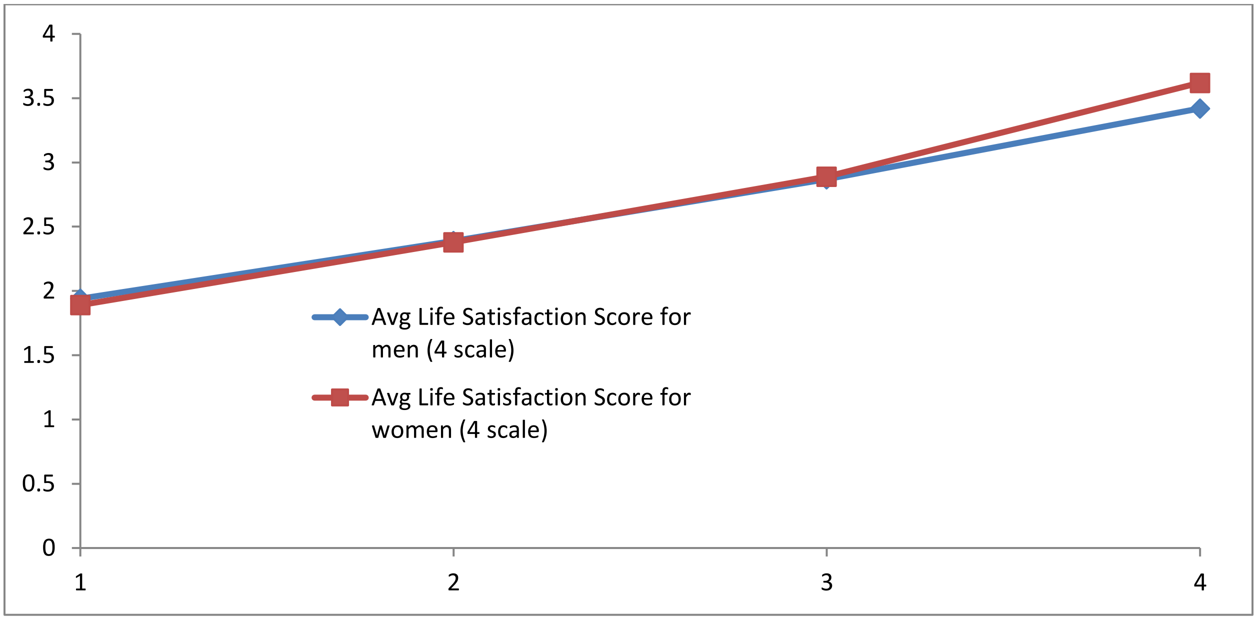
Note: - Plotted using our own survey data. The bar 'uneducated' and 'educated' represents the 'uneducated' and 'educated' cohort in our sample respectively.

Figure 4.10:- Life Satisfaction and Age are Negatively Correlated



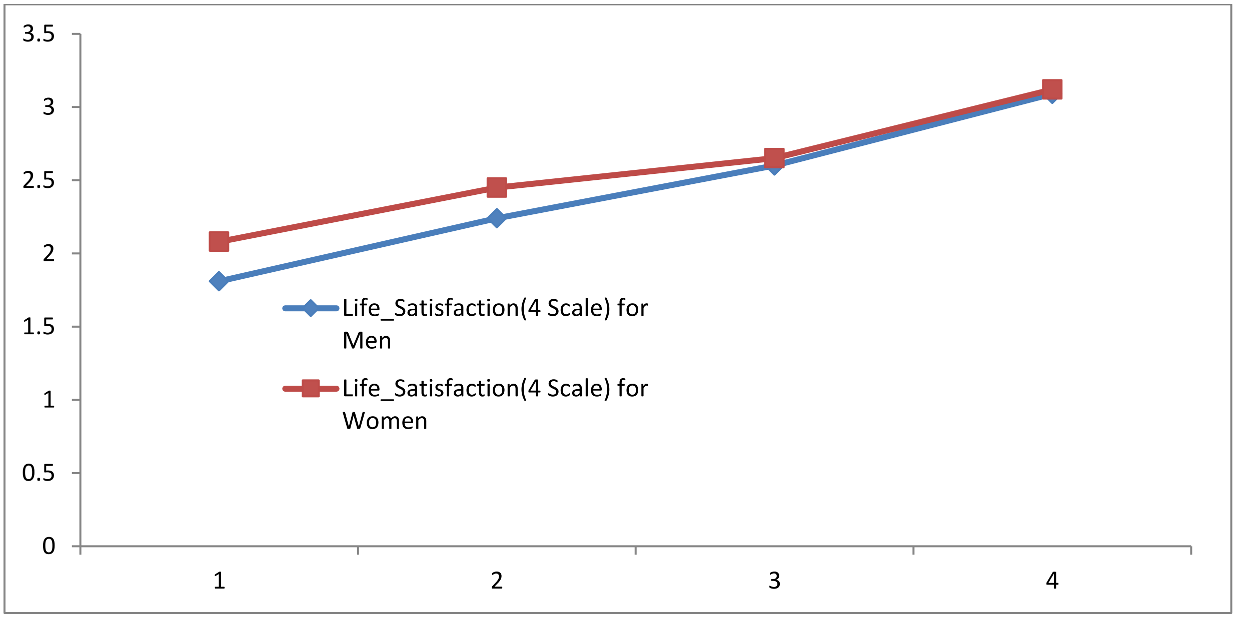
Note: - Plotted using our own survey data. Bivariate linear regression lines are fitted to the above plot.

Figure 4.11:- Average Life Satisfaction Score for Different Levels of Financial Satisfaction (Men and Women)



Note: - Drawn from our own survey data. The horizontal axis represents different levels of financial satisfaction. The vertical axis shows the average life satisfaction score for each level of financial satisfaction.

Figure 4.12:- Average Life Satisfaction Score for Different Levels of Health Satisfaction (Men and Women)



Note: - Drawn from our own survey data. The horizontal axis represents different levels of health satisfaction. The vertical axis shows the average life satisfaction score for each level of health satisfaction.

Figure 4.13:- Correlation between Life Satisfaction and Monthly Family Income: Men and Women

Note: - Plotted using our own survey data. Bivariate linear regression lines are fitted to the above plot.

Table 4.1:-List of Slums Surveyed-1

ZONE	SLUM NAME	LOCATION
East	Aradhak Nagar Camp	Behind Shahadra Border
East	Sonia Camp	Dilshad Garden
East	Rajiv Camp Mini Market	Trilok Puri
East	Shashtri Mohalla	Shashi Garden
East	Deepak Colony, Block E-103	Near Ahauchalaya
East	Mazdoor Nagar Camp	I P Extension
East	J. J. Bharti Camp	East Vinod Nagar
East	Ram Prasad Vishmil camp	Shashi Garden
East	Shahid Bhagat Singh camp	Kalyan puri
East	Dr. Rajender Prasad Camp	G.T.B. Hospital Delhi
North	J.J Camp Bhagwan Pur	Libas Pur
North	Kabir Nagar and Kishore Nagar jj Cluster	Rana Pratap Nagar
North	JJ Colony Sari Peepasl Thala	Adarsh Nagar

Table 4.2:-List of Slums Surveyed-2

ZONE	SLUM NAME	LOCATION
South	Malviya Nagar Corner Camp	Malviya Nagar
South	Sarvodaya Camp	Kalka Ji
South	JJ. Indira Camp	Sriniwaspuri
South	Nehru camp	Govind Puri
South	New Sanjay Camp E-33	Okhla Ph-II
South	Bhanwar Singh Camp	Vasant Vihar
South	Sonia camp part ii	Prahlad Pur
South	V.P. Singh camp	Tugalkabad
West	Udyog Nagar Camp	Preera Garhi
West	Indira Camp Part 2	Vikash Puri
West	Tilak Nagar Industrial area	Subhash Nagar
West	JJ Camp Block D-4	Sultan Puri
West	Bhim Nagar Jwalapur Camp	Pira Gadhi
West	Prem Nagar Camp	Patel Nagar
West	Rajeev Gandhi Camp Saad Nagar Ph-2	Saad Nagar
West	Nehru Camp	Brijwasan Village.

Table 4.3:- Percentage of Individuals Who Report Unavailability or Poor Quality of The Following Public Facilities.

Public Facilities	**Percentage of Respondents Reporting**
No dustbin around	42.1%
Malfunctioning of the drainage system	56.6%
Irregular visits by the Govt. sweepers	65.41%
Street light does not exist/not functioning	57.3%
No toilet in home	70.6%
No water supply at home	35.5%
Poor quality drinking water	18%
Unclean public toilet	22.4%

Note: - This Table is constructed from our own survey data.

Table 4.4:-Variability in the Satisfaction Score

Satisfaction Measure	Mean	SD	Mean+2SD	Mean-2SD
Life Satisfaction	2.63	0.89	4.41	0.85
Health Satisfaction	2.83	0.89	4.61	1.05
Financial Satisfaction	2.47	0.78	4.03	0.91

Note: - Author's calculation from the survey data. SD is the standard deviation.

Table 4.5:-Percentage Change in Self Reported Life Satisfaction as Self Reported Health Satisfaction Changes

Change in the level of health satisfaction	From 1to 2	From 2to 3	From 3to 4
Percentage increase in average Life Satisfaction(4 scale)	19.6	10.5	17.9

Note: - Author's computation from the survey data.

Table 4.6:-Percentage Change in Self Reported Life satisfaction as Self Reported Financial Satisfaction Changes

Change in the level of financial satisfaction	From 1to 2	From 2to 3	From 3to 4
Percentage increase in average Life Satisfaction(4 scale)	24.6	21	21.9

Note: - Author's computation from the survey data.

Table 4.7:-Average Life Satisfaction Score for Those Who Possess and Those Who Don't Possess the Following Assets (Refrigerator and Two Wheeler)

Possession of an asset	Refrigerator	Two Wheeler
Possess	2.72	2.86
Don't Possess	2.55	2.58

Note: - Author's computation from the survey data.

Table 4.8:-Average Life Satisfaction Score and Functioning of the Drainage System

Average Life Satisfaction Score	Overall Sample
Drainage system functioning well	2.78
Drainage system not functioning well	2.52

Note: - Author's computation from the survey data. The respondents have reported whether the drainage system in front of their home function properly.

Table 4.9:-Average Life Satisfaction Score (4 scale) for Educated and Uneducated Cohorts: Comparison between Men and Women

Zone	Uneducated men	Uneducated women	Educated men	Educated women	Ratio between uneducated women and men	Ratio between educated women and men
East	2.41	2.66	2.84	2.78	1.1	0.98
West	2.49	2.56	2.96	2.48	1.03	0.84
South	2.7	2.7	2.84	2.82	1	0.99
North	2.08	2.53	2.38	2.47	1.22	1.04
Overall Sample	2.46	2.63	2.84	2.67	1.07	0.94

Note: - Author's computation from the survey data. Educated group/cohort consists of those people who have at least secondary education.

Table 4.10:-Average Life Satisfaction Score and Possession of Ration Card

Average Life Satisfaction Score(4 Scale)	Men	Women	Overall
Possess a ration card	2.6	2.66	2.64
Don't possess a ration card	2.68	2.53	2.59

Note: - Author's computation from the survey data.

Table 4.11:-Definition of Variables and Their Summary Statistics: Respondent Specific Variables

Variable Name	Description of the Variable	Mean	Standard Deviation
Health Satisfaction	A categorical variable with four scale/four choices. A bottom scale response gets the score of 1and the top scale response gets the score 4.	2.83	0.89
Financial Satisfaction	A categorical variable with four scale/four choices. A bottom scale response gets the score of 1and the top scale response gets the score 4.	2.47	0.78
Age	Age of the respondent.	35.87	11.8
Married	It represents the marital status of the respondent. Married is a binary variable. It takes the value 1 if the person is married and takes the value 0 otherwise.	0.82	0.39
Educated	A binary variable that stands for the education level of the respondent. It takes the value of 1 for a person who has at least secondary education. Otherwise it is assigned a value 0.	0.28	0.45
Employed	A binary variable that takes the value 1 for an earner and 0 for a non-earner.	0.47	0.5
Household Head	A binary variable that takes the value 1 if the respondent is the household head and it takes the value 0 otherwise.	0.31	0.46
Smoke	A binary variable that takes the value 1 if an individual smokes and takes the value 0 otherwise.	0.15	0.36
Drink Alcohol	A binary variable that takes the value 1 if an individual drinks alcohol and takes the value 0 otherwise.	0.11	0.31
Chew Tobacco	A binary variable that takes the value 1 if an individual chews tobacco and takes the value 0 otherwise.	0.14	0.35

Female	A binary variable that takes the value 1 if the respondent is a female and takes the value 0 otherwise.	0.6	0.49

Note:-Author's computation from the survey data. Apart from age, all other variables are binary variables. The mean of these binary variables represent the proportion of people reporting the value/category 1. For health satisfaction and financial satisfaction, there are more than two categories. But each category represents a score. Therefore, the reported mean and standard deviation for health satisfaction and financial satisfaction are the mean and standard deviation of the reported scores.

Table 4.12:-Definition of Variables and Their Summary Statistics: Household Specific Variables

Variable Name	Description of the Variable	Mean	Standard Deviation
Monthly Family Income	Sum of the income earned by all the members in a household.	12920	9500
Household Size	Number of members in a household.	5.35	1.94
Male Children	Total number of male children in a household.	0.86	0.93
Female Children	Total number of female children in a household.	0.84	1.02
General	A binary variable that takes the value 1 if the respondent belongs to general category, otherwise it takes the value 0.	0.24	0.42
Non-Hindu	A binary variable with the value 1 for the non-Hindu respondents and 0 otherwise.	0.15	0.36
Not Possessing Refrigerator	A binary variable that takes the value 1 if the household do not own a refrigerator and 0 otherwise.	0.54	0.5
Not Possessing Two Wheeler(Scooter/Motorbike)	A binary variable that takes the value 1 if the household do not own a two wheeler and 0 otherwise.	0.81	0.39
Not Possessing Ration Card	A binary variable that takes the value 1 if the household do not possess a ration card and 0 otherwise.	0.15	0.36

Note: - Author's computation from the survey data. 'Not Possessing Refrigerator', 'Not Possessing Two Wheeler' and 'Not Possessing Ration Card' are all binary variables. For these binary variables, the mean represents the proportion of people reporting the value/category 1.

Table 4.13:-Definition of Variables and Their Summary Statistics: Mental Health/Psychological Traits of the Respondents

Variable Name	Description of the Variable	Mean	Standard Deviation
Not Feeling Lonely	A binary variable that takes the value 1 if the respondent does not feel lonely and takes the value 0 otherwise.	0.71	0.45
Can't Concentrate	A binary variable that takes the value 1 if the respondent can't concentrate and takes the value 0 otherwise.	0.16	0.37
Unable to Take Decision	A binary variable that takes the value 1 if the respondent is unable to take any decision and takes the value 0 otherwise.	0.21	0.41
Don't Feel Confident in Work	A binary variable that takes the value 1 if an individual does not feel confident and takes the value 0 otherwise.	0.05	0.23
Can't Overcome Difficulties	A binary variable that takes the value 1 for an individual who is unable to overcome difficulties and 0 otherwise.	0.67	0.47
Someone in the Family Died in the Past Two Years	A binary variable that takes the value 1 if any household member has died within the past two years and 0 otherwise.	0.19	0.39
Don't Feel Secured	Takes the value I if the respondent doesn't feel secured in the slum and 0 otherwise.	0.41	0.49
Stress	Takes the value 1 if the respondent feels stressed and 0 otherwise.	0.72	0.45
Hours Working	Number of hours a respondent work in a day.	6.8	3.5
Hours Sleeping	Number of hours a respondent sleep in a day.	6.8	1.5

Note: - Author's computation from the survey data. All variables except 'Hours Working' and 'Hours Sleeping' are binary variables. For these binary variables, the mean represents the proportion of people reporting the value/category 1. 'Hours Working' and 'Hours Sleeping' are continuous variables.

Table 4.14:-Definition of Variables and Their Summary Statistics: Public Facilities

Variable Name	**Description of the Variable**	**Mean**	**Standard Deviation**
Health Camp	Takes the value 1 if the respondent reports that at least one health camp has been held in his/her slum in the last six months, otherwise value of 0 is assigned.	0.17	0.38
Child Immunization Camp	Takes the value 1if the respondent reports that at least one child immunization camp has been held in his/her slum in the last six months, otherwise it takes the value of 0.	0.23	0.42
NGO Working	It takes the value 1 if the respondent reports to be aware about any NGO facility in his/her slum; otherwise a value of 0 is assigned.	0.2	0.4
Drinking Water Quality	Value is 1 when the respondent is satisfied with the quality of the drinking water, otherwise the value is 0.	0.78	0.41
Clean Toilet	A binary variable with a value 1 if the respondent reports the public toilet to be clean and with a value 0 if he/she reports the public toilet to be unclean.	0.6	0.49
Anganwadi (Child Care Centre) Exist	Takes the value 1if such a centre exists in the slum, otherwise it takes the value of 0.	0.81	0.39
Anganwadi (Child Care Centre) Mid Day Deal	Takes the value 1 if the respondent says that Anganwadi provides mid-day meal, otherwise a value of 0 is assigned.	0.72	0.45
Public Toilet Woman	When a public toilet exists in the slum, this variable takes the value 1 and it is 0	0.83	0.37

	otherwise.		
Government Dispensary	When a government dispensary exists in the slum, this variable takes the value 1 and it is 0 otherwise.	0.24	0.43
Drainage Not Functioning	It takes the value 1 if the respondent complains about the malfunctioning of drainage in front of his/her home; otherwise it takes the value 0.	0.57	0.5
No Dustbin	It takes the value 1 if the respondent complains that there is no dustbin near his/her home, otherwise it takes the value 0.	0.43	0.5
Sweeper Not Cleaning	Takes the value 1 if a slum dweller complains about the irregular visits by the government sweeper, otherwise it is 0.	0.66	0.47
Street Light Not Working	Takes the value 1 if the street light does not exist or does not function properly and 0 otherwise.	0.57	0.49

Note: - Author's computation from the survey data. All these variables are binary variables. The mean represents the proportion of people reporting the value/category 1.

Table 4.15:- Ordered Logistic Regression: Determinants of Life Satisfaction

VARIABLES	(1) life satisfaction	(2) life satisfaction	(3) life satisfaction
2.health satisfaction	0.831***	0.868***	0.871***
	(0.273)	(0.265)	(0.275)
3.health satisfaction	1.077***	1.093***	1.110***
	(0.276)	(0.270)	(0.279)
4.health satisfaction	2.098***	2.147***	2.158***
	(0.312)	(0.308)	(0.316)
ln monthly family income	0.376***	0.387***	0.387***
	(0.103)	(0.104)	(0.105)
not possessing ration card	-0.212	-0.161	-0.205
	(0.161)	(0.165)	(0.169)
not possessing two wheeler	-0.212	-0.178	-0.200
	(0.168)	(0.167)	(0.176)
not possessing refrigerator	-0.251**	-0.312**	-0.258**
	(0.126)	(0.128)	(0.131)
female	-0.202	-0.268	-0.246
	(0.284)	(0.279)	(0.290)
married	0.816***	0.652**	0.762***
	(0.280)	(0.275)	(0.295)
age	0.00412	0.00402	0.00438
	(0.0374)	(0.0379)	(0.0393)
squared age	-7.79e-05	-4.15e-05	-5.84e-05
	(0.000457)	(0.000465)	(0.000483)
educated	0.271*	0.277*	0.268*
	(0.158)	(0.154)	(0.161)
drainage not functioning	-0.512***	-0.484***	-0.547***
	(0.141)	(0.132)	(0.145)
not feeling lonely	0.567***	0.589***	0.529***
	(0.155)	(0.159)	(0.162)
sometimes feeling stressed	-0.460***	-0.448***	-0.439***
	(0.157)	(0.153)	(0.160)
always feeling stressed	-1.321***	-1.334***	-1.371***
	(0.243)	(0.241)	(0.248)
Observations	1,214	1,201	1,201
slum fixed effects	yes	no	yes
common public facilities	no	yes	yes

Note: The dependent variable is the life satisfaction score based on four scale. The unit of analysis is an individual. In the first column, regressions are run after controlling for the slum fixed effects. In the second column slum specific common public facilities are controlled. Both slum fixed effects and slum specific facilities are controlled in the third column. Robust standard errors are in parenthesis, ***, ** and * indicate significance at 1%, 5% and 10% level respectively.

Table 4.16:- Ordered Logistic Regression: Impact of Financial Satisfaction on Life Satisfaction

VARIABLES	life satisfaction	life satisfaction	life satisfaction
2.health satisfaction	0.701**	0.758***	0.749***
	(0.275)	(0.266)	(0.277)
3.health satisfaction	0.824***	0.837***	0.883***
	(0.276)	(0.269)	(0.280)
4.health satisfaction	1.819***	1.835***	1.886***
	(0.311)	(0.307)	(0.316)
2.financial satisfaction	0.949***	0.916***	0.941***
	(0.263)	(0.264)	(0.267)
3.financial satisfaction	1.937***	1.876***	1.884***
	(0.275)	(0.278)	(0.278)
4.financial satisfaction	3.508***	3.560***	3.612***
	(0.458)	(0.447)	(0.458)
Observations	1,212	1,199	1,199
slum fixed effects	yes	no	yes
common public facilities	no	yes	yes

Note: The dependent variable is the life satisfaction score based on four scale. The unit of analysis is an individual. In the first column, regressions are run after controlling for the slum fixed effects. In the second column slum specific common public facilities are controlled. Both slum fixed effects and slum specific facilities are controlled in the third column. Robust standard errors are in parenthesis, ***, ** and * indicate significance at 1%, 5% and 10% level respectively.

Table 4.17:- OLS Regression: Determinants of Life Satisfaction

VARIABLES	(1) life satisfaction	(2) life satisfaction	(3) life satisfaction
2.health satisfaction	0.330***	0.342***	0.332***
	(0.0994)	(0.0966)	(0.0989)
3.health satisfaction	0.431***	0.435***	0.431***
	(0.101)	(0.0984)	(0.100)
4.health satisfaction	0.795***	0.804***	0.795***
	(0.109)	(0.108)	(0.109)
ln family monthly income	0.139***	0.143***	0.138***
	(0.0418)	(0.0415)	(0.0414)
not possessing ration card	-0.0761	-0.0649	-0.0789
	(0.0626)	(0.0630)	(0.0643)
not possessing two wheeler	-0.0757	-0.0663	-0.0701
	(0.0639)	(0.0638)	(0.0659)
not possessing refrigerator	-0.0868*	-0.105**	-0.0840*
	(0.0486)	(0.0485)	(0.0493)
female	-0.0641	-0.0954	-0.0795
	(0.108)	(0.106)	(0.108)
married	0.313***	0.245**	0.281**
	(0.108)	(0.106)	(0.111)
age	-0.00133	-0.00229	-0.00224
	(0.0142)	(0.0143)	(0.0146)
squared age	8.41e-06	2.47e-05	2.50e-05
	(0.000171)	(0.000172)	(0.000176)
educated	0.0860	0.0855	0.0797
	(0.0591)	(0.0584)	(0.0598)
drainage not functioning	-0.181***	-0.180***	-0.194***
	(0.0537)	(0.0494)	(0.0538)
not feeling lonely	0.232***	0.236***	0.210***
	(0.0613)	(0.0613)	(0.0625)
sometimes feeling stressed	-0.160***	-0.147***	-0.140**
	(0.0589)	(0.0567)	(0.0585)
always feeling stressed	-0.478***	-0.477***	-0.480***
	(0.0885)	(0.0870)	(0.0883)
Observations	1,214	1,201	1,201
R-squared	0.330	0.338	0.355
slum fixed effects	yes	no	yes
common public facilities	no	yes	yes

Note: The dependent variable is the life satisfaction score based on four scale. The unit of analysis is an individual. In the first column, regressions are run after controlling for the slum fixed effects. In the second column slum specific common public facilities are controlled. Both slum fixed effects and slum specific facilities are controlled in the third column. Robust standard errors are in parenthesis, ***, ** and * indicate significance at 1%, 5% and 10% level respectively.

Table 4.18:- Determinants of Life Satisfaction: Differential Effect between Men and Women

VARIABLES	(1) life satisfaction	(2) life satisfaction
2.health satisfaction*female	-0.484	-0.160
	(0.581)	(0.203)
3.health satisfaction*female	-0.701	-0.270
	(0.586)	(0.205)
4.health satisfaction*female	-1.005	-0.349
	(0.635)	(0.218)
ln monthly family income*female	-0.234	-0.0842
	(0.231)	(0.0834)
not possessing ration card*female	-0.417	-0.168
	(0.321)	(0.120)
not possessing two wheeler*female	0.337	0.109
	(0.330)	(0.125)
not possessing refrigerator*female	-0.143	-0.0667
	(0.260)	(0.0962)
married*female	0.766	0.273
	(0.547)	(0.205)
age*female	-0.120	-0.0497*
	(0.0750)	(0.0270)
squared age*female	0.00156*	0.000629*
	(0.000918)	(0.000327)
educated*female	-0.622**	-0.253**
	(0.309)	(0.115)
drainage not function*female	-0.0115	0.00372
	(0.250)	(0.0938)
not feeling lonely*female	-0.819***	-0.313***
	(0.309)	(0.117)
sometimes feeling stressed*female	-0.251	-0.104
	(0.334)	(0.121)
always feeling stressed*female	-0.447	-0.172
	(0.475)	(0.173)
Observations	1,214	1,214
R-squared		0.348
regression specification	Ordered Logit	OLS
slum fixed effects	yes	yes

Note: The dependent variable is the life satisfaction score. The unit of analysis is an individual. The regressions are run using the entire sample. The first column shows the results of an ordered logistic regression and the second column shows the results of an OLS regression. For both the regressions, slum fixed effects are used as control variables. Robust standard errors are in parenthesis, ***, ** and * indicate significance at 1%, 5% and 10% level respectively.

Chapter 5

Summary and Conclusion

In this thesis, we explain and analyze three important and basic issues in applied welfare economics in three different chapters. Although all these three chapters pose new questions in new contexts and use different methodologies, the common theme of all these three chapters is welfare evaluation and comparing change in welfare across groups.

In the first chapter; we consider the impact of individual heterogeneity on group cost of living index. Heterogeneity in consumption pattern leads to heterogeneity in budget shares and we have considered its impact on the group cost of living index. The different spending pattern may arise from varying preferences or varying income across people. As shown in chapter 2, higher heterogeneity matters if change in relative prices is large enough. The central statistical agencies for almost all the countries ignore this issue of heterogeneity and compute the cost of living index for a representative individual. The resulting bias is captured by our methodology.

It turns out that the bias will be negligible for small change in relative prices. But it is theoretically plausible for the bias to be large and if the statistical agencies neglect that large bias, then the implemented policies can be wrong. The bias happens to be negligible for the two commodity example that has been provided in chapter2. But for two commodities (food and non-food in our story), commodities are defined at the broad level and hence the budget share (for either food or non-food) usually lack enough variation. As a result, the bias due to considering the index of a representative agent (defined at the average budget share) as the group cost of living index turns out to be small if we concentrate on two commodity classification. But whenever we classify commodities at a more disaggregated level, the variation in the budget share is expected to be much higher and it may generate a larger bias provided the change in relative price is large enough. Therefore an extension of our example to n commodities may turn out to be quite interesting and useful. Another interesting work that can be done from this second chapter is to empirically determine the factors that explain the heterogeneity in budget share i.e. what are the factors that cause the budget share to vary.

The third chapter of our thesis deals with the welfare implication of the edible oil trade liberalization in India. We have shown in chapter 3 that the domestic edible oil price and the agricultural wage rate increases as the border price of palm oil (which is the cumulative outcome of the world price of palm oil, ad-valorem tariff rate and exchange rate) increases. But the effects are not uniform across the country and we notice spatial variation in the impact. The effect of border price of palm oil on domestic edible oil price turns out to be larger in the coastal states of India and also in the top oilseeds producing districts in India. The spatial effect of the border price on agricultural wage rate does not turn out to be as strong as compared to the impact on domestic edible oil price. We provide a theoretical foundation of such spatial variation and show the corresponding welfare implication if such spatial effect exists.

Chapter 3 is immensely important in terms of policy implication. If different regions get affected differently due to trade induced price change, then government needs to plan its compensation scheme for different regions separately. Therefore we can relate our research to political economy of trade policy. In this chapter, we have focused on edible oil. Similar analysis can be done for other commodities which are important in the import basket of the

Indian consumers. Spatial heterogeneity may be even more prominent for some of these other commodities and again that may have a huge welfare and policy implication.

Right now, we are doing a host of extensions of chapter 3 as a part of post-doctoral research. One of them is disaggregating the district level average edible oil price (which we have used as the dependent variable in our price regression) into two categories. As we have mentioned earlier, nationally representative consumer expenditure survey data (NSS) of India does not allow us to obtain price of palm oil at the district level. The NSS data disaggregates all the edible oil into two parts: traditionally consumed edible oils in India (groundnut, rapeseed-mustard and coconut) and other edible oils (palm oil is included in other edible oil). Therefore we can obtain unit value (which we have used as price) of traditionally consumed edible oil and other edible oil at the district level and instead of using the average edible oil price (unit value of edible oil constructed at the district level by considering all edible oils together) as a dependent variable, two separate price regressions can be run using the price of traditional oil and other oil respectively. The price of other edible oil closely approximates the price of palm oil.

Another extension of the empirical methodology where we are working right now is to check the robustness of the results after controlling for the lagged values of the border price of palm oil. So far only the contemporaneous values of the border price of palm oil has been considered in the regression analysis.

Are we truly capturing the pass-through elasticity of the border price of palm oil or something else? The way to answer this question is to use the border price of other imported commodities in the regression instead of palm oil and compare the results. We have already started working in that direction.

Finally, an interesting extension of this chapter is to incorporate the producers of oilseeds in the picture and analyze the impact of the border price of palm oil on the domestically produced price of oilseeds. Then it will be possible to do a combined welfare analysis with the consumers, producers/farmers and wage earners all together. But for that, we need to collect the data of domestically produced oilseeds at the district level and we have started searching for that.

The fourth chapter of my thesis is on economics of happiness. One of the key findings of the chapter is that the poor tend to report high levels of life satisfaction despite an inconvenient life they face regularly. The reported high level of happiness score resembles what we usually see for the developed countries (like United States). Adaptation to the daily life circumstances can be considered as a principal factor behind the above finding.

Although reported life and health satisfaction score is high, same is not true for financial satisfaction. Number of people who report low level of financial satisfaction is much higher than those who report low level of life satisfaction and health satisfaction. Regardless of the number of people who report very low or very high level of satisfaction, sufficient variation in the satisfaction score is found and that's why it is important to figure out the determinants of the variation in the satisfaction score. Family income, health satisfaction, financial satisfaction, education, possession of assets, psychological traits and quality and availability of public facilities turn out to be correlates of self reported well being.

A more interesting thing is to investigate the variation across gender of the relative importance of the factors that are correlated with the self-reported well- being measures. This is worthy to investigate when societies offer different opportunities and liberty to men and women and preferences vary across gender.

In our analysis, education turns out to be the most interesting gender varying correlate. We find that with increase in education, reported well being of women relative to men falls. This finding is consistent with the findings of Stevenson and Wolfers (2009) regarding the declining female happiness (relative to male) in United States and the industrialized nations in Europe during 1970-2005 despite the fact that the women has done much better than men in terms of objective measures in the same period of time (in our story, education is an objective measure of well-being). Stevenson and Wolfers (2009) did a time series analysis while our data is cross-sectional. Therefore our finding is a cross-sectional counterpart of what has been found by Stevenson and Wolfers. The chapter's finding also resembles Lalive and Stutzer (2010) who find the life satisfaction of women to be higher in the traditional communities compared to liberal communities in Switzerland and explain their finding as a result of higher expectation of the women in liberal communities. A similar explanation in our case would suggest while education increases the aspirations of women and men, societal norms create barrier for women more than men. One of our future research goals is to investigate the foundations of such explanations. It is quite interesting to think of a measure that captures the increase in aspiration for women as education level increases.

Recently we have another interesting finding from our survey data. It turns out that average life satisfaction score is much higher for an employed man compared to an unemployed man irrespective of the education level(i.e. it is true for both educated and uneducated cohort). The opposite is true for women i.e. the average of the self reported well being is less for the employed women compared to unemployed women and again this is true regardless the level of education. In a regression framework (using only the sample of women respondents) it turns out that the subjective well being score is significantly higher for the housewives compared to the earning women.

One plausible reason for the above finding as mentioned by Stevenson and Wolfers (2009) can be the double burden of workload for the working women. They need to manage both the workplace and home when they are employed. Just like education, the story of increasing expectation can also play a big role here.

As shown in chapter 4, the impact of some factor on life satisfaction can also vary across gender because of divergent preferences. An instance provided in the chapter is the access to government subsidized foodgrain, sugar and kerosene (used for cooking and/or lighting) through public distribution system. In future we will try to find out more instances of such preference heterogeneity across gender and will research on its implication on self reported well being. Although there are literature which touch upon the issue of gender and well-being, to our knowledge, this work is the first one that discusses the well being of women relative to men among the poor.

In this thesis, we pose three important questions in the area of applied welfare economics and try to solve these using suitable methodologies. A few interesting extensions and related future research is in the pipeline and works in those issues are in progress.

Bibliography

Ackerman, F., Wise, T. A., Gallagher, K. P., Ney, L., and Flores, R. (2003). *Free trade, corn, and the environment: environmental impacts of US-Mexico corn trade under NAFTA*. Tufts University.

Argyle, M. (2003). 18 Causes and Correlates of Happiness. *Well-being: The foundations of hedonic psychology*, *353*.

Atkin, D. (2013). Trade, tastes and nutrition in India. *The American Economic Review*, *103*(5), 1629-1663.

Atkin, D., and Donaldson, D. (2015). *Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs* (No. w21439). National Bureau of Economic Research.

Banerjee, A., Deaton, A., and Duflo, E. (2004). Health, health care, and economic development: Wealth, health, and health services in rural Rajasthan. *The American economic review*, *94*(2), 326.

Banerjee, Abhijit, Rohini Pande and Michael Walton (2012); "Delhi's Slum Dwellers: Deprivation, Preferences and Political Engagement among the Urban Poor"; IGC Working Paper (2012).

Blanchflower, D. G., and Oswald, A. J. (2008). Is well-being U-shaped over the life cycle?. *Social science and medicine*, *66*(8), 1733-1749.

Boskin, M. J., and Hurd, M. D. (1985). Indexing Social Security benefits a Separate Price-Index for the Elderly. *Public Finance Quarterly*, *13*(4), 436-449.

Case, A., and Deaton, A. (2005). Health and wealth among the poor: India and South Africa compared. *The American economic review*, *95*(2), 229-233

Chaudhuri, S., and Gupta, N. (2009). Levels of living and poverty patterns: a district-wise analysis for India. *Economic and Political Weekly*, 94-110.

Crawford Ian and Zoe Smith (2002),"Distributional Aspects of Inflation", The Institute for Fiscal Studies, June 2002

Crossley, T. F., and Pendakur, K. (2010). The common-scaling social cost-of-living index. *Journal of Business & Economic Statistics*, *28*(4), 523-538

Deaton, A. (1998). Getting prices right: what should be done?. *the Journal of economic Perspectives*, *12*(1), 37-46.

Deaton, A. (1989). Rice prices and income distribution in Thailand: a non-parametric analysis. *The Economic Journal*, *99*(395), 1-37.

Deaton, A. (2008). Income, health, and well-being around the world: Evidence from the Gallup World Poll. *The journal of economic perspectives*, *22*(2), 53-72.

Deaton, A., and Stone, A. A. (2013). Two happiness puzzles. *The American economic review*, *103*(3), 591-597.

Di Tella, R., and MacCulloch, R. (2008). Gross national happiness as an answer to the Easterlin Paradox?. *Journal of Development Economics*, *86*(1), 22-42.

Diener, E(1984); "Subjective Well-Being", *Psychological Bulletin*, Vol-95, No3, 542-575

Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of econometrics*, *4*(2), 115-145.

Diewert, W. E. (1978). Superlative index numbers and consistency in aggregation. *Econometrica: Journal of the Econometric Society*, 883-900

Diewert, W.Erwin(1981),"The Economic Theory of Index Numbers:A Survey", Essays in the theory and Measurment of Consumer Behaviour in Honour of Sir Richard Stone, edited by A. Deaton, London: Cambridge University Press,1981,pp.163-208

Dohlman, E. N., Persaud, S., and Landes, R. (2003). *India's edible oil sector: Imports fill rising demand*. US Department of Agriculture, Economic Research Service.

Duflo, Esther, and Rohini Pande. "Dams." *The Quarterly Journal of Economics* 122, no. 2 (2007): 601-646.

Easterlin, R. (1974). 'Does economic growth improve the human lot ? Some empirical evidence.' In Nations and Households in Economic Growth : Essays in Honour of Moses AbramoWit (ed. P. A. David and M. W. Reder). New York and London: Academic Press.

Easterlin, R. A. (2001). Income and happiness: Towards a unified theory. *The economic journal*, *111*(473), 465-484.

Edmonds, E. V., Pavcnik, N., and Topalova, P. (2010). Trade adjustment and human capital investments: Evidence from Indian tariff reform. *American Economic Journal: Applied Economics*, *2*(4), 42-75.

Ferrer-i-Carbonell, A., and Gowdy, J. M. (2007). Environmental degradation and happiness. *Ecological Economics*, *60*(3), 509-516.

Fisher, F. M., and Griliches, Z. (1995). Aggregate price indices, new goods, and generics. *The Quarterly Journal of Economics*, 229-244.

Friedman, J., and Levinsohn, J. (2002). The distributional impacts of Indonesia's financial crisis on household welfare: A "rapid response" methodology. *The World Bank Economic Review*, *16*(3), 397-423.

Frijters, P., and Van Praag, B. M. (1998). The effects of climate on welfare and well-being in Russia. *Climatic Change*, *39*(1), 61-81.

Gaddis, I., and Pieters, J. (2016). The gendered labor market impacts of trade liberalization: evidence from Brazil. *Journal of Human Resources*.

Ghosh, N. (2009). Effects of tariff liberalization on oilseed and edible oil sector in India: Who wins and who loses. *Online). www. taerindia. com/pdf/working_paper_no*, *2*.

Goldar, B. (2005). *Impact on India of Tariff and Quantitative Restrictions under WTO* (No. 172). Indian Council for Research on International Economic Relations, New Delhi, India.

Goldberg, P. K., and Knetter, M. M. (1997). Goods Prices and Exchange Rates: What Have We Learned?. *Journal of Economic Literature*, *35*(3), 1243-1272.

Goldberg, P. K., and Pavcnik, N. (2007). Distributional effects of globalization in developing countries. *Journal of economic Literature*, *45*(1), 39-82.

Graham, Carol (2009); Happiness Around the World: “The paradox of happy peasants and miserable millionaires”; Oxford University Press (2009)

Graham, Carol (2011);”The Pursuit of Happiness: An Economy of Well Being”; Brookings Institution Press (2011)

Graham, C., and Chattopadhyay, S. (2012). *Gender and well-being around the world: Some insights from the economics of happiness* (No. 2012-010).

Greenhut, M. L., Norman,G, ., and Hung, C,S. (1987). *The economics of imperfect competition: A spatial approach*. Cambridge University Press.

Gulati, A., and Narayan, S. (2002, June). Managing Import Competition when LDCs Liberalize Trade: The Indian Experience. In *IATRC Symposium on The Developing Countries, Agricultural Trade and the WTO, Whistler, British Colombia*.

Gulati, A., Sharma, A., and Kohli, D. S. (1996). Self-sufficiency and allocative efficiency: case of edible oils. *Economic and Political Weekly*, A15-A24.

Hagemann, R. P. (1982). The variability of inflation rates across household types. *journal of Money, Credit and Banking*, *14*(4), 494-510.

Hasan, R., Mitra, D., and Ural, B. P. (2007, August). Trade liberalization, labor-market institutions and poverty reduction: Evidence from Indian states. In *India Policy Forum* (Vol. 3, No. 1, pp. 71-122). National Council of Applied Economic Research

Helliwell, J. F. (2002). *Globalization and well-being*. UBC press.

Helliwell, J. F., Layard, R., and Sachs, J. (2013). World happiness report [2012].

Hoda, A., and Gulati, A. (2013). India’s agricultural trade policy and sustainable development. *ICTSD Issue Paper*, *49*.

Jacoby, H. G. (2016). Food prices, wages, and welfare in rural India. *Economic Inquiry*, *54*(1), 159-176

James, C., Teng, P., Arujanan, M., Aldemita, R. R., Flavell, R. B., Brookes, G., and Qaim, M. Invitational Essays to Celebrate the 20th Anniversary of the Commercialization of Biotech Crops (1996 to 2015): Progress and Promise.

Kovak, B. K. (2013). Regional effects of trade reform: What is the correct measure of liberalization?. *The American Economic Review*, *103*(5), 1960-1976.

Lalive, R. and Stutzer, A. (2010). Approval of Equal Rights and Gender Differences in Well-Being. *Journal of Population Economics*, 23, 933-962.

Landes, M., Persaud, S. C., and Dyck, J. H. (2004). India's poultry sector.

Lima, S. V. (2013). Essays on Economics and Happiness.

Luechinger, S. (2009). Valuing air quality using the life satisfaction approach. *The Economic Journal*, *119*(536), 482-515.

Luechinger, S., and Raschky, P. A. (2009). Valuing flood disasters using the life satisfaction approach. *Journal of Public Economics*, *93*(3), 620-633.

Mackie, Christopher and Charles L. Schultz (Editors),"At What Price?: Conceptualizing and Measuring Cost Of Living and Price Indexes", Panel on conceptual, measurement and other statistical issues in developing cost of living indexes, National Academy Press(2002)

Marchand, B. U. (2012). Tariff pass-through and the distributional effects of trade liberalization. *Journal of Development Economics*, *99*(2), 265-281.

Matteucci, N., and Vieira Lima, S. (2014). Women and happiness.

Mehta, R. (2000). Removal of QRs and Impact on India's Import. *Economic and Political Weekly*, 1667-1671.

Michael, R. T. (1979). Variation across households in the rate of inflation. *Journal of Money, Credit and banking*, *11*(1), 32-46.

Miller, C. K., and Idson, T. (1994). The Effect of Household Characteristics on Household-Specific Inflation Rates: An Application to Trends in Child Poverty and Educational Real Wage Differentials.

Mishra, A., and Ray, R. (2011). Prices, Inequality, and poverty: Methodology and Indian evidence. *Review of Income and Wealth*, *57*(3), 428-448.

Motiram, S., and Osberg, L. (2010). Gender inequalities in tasks and instruction opportunities within Indian families. *Feminist Economics*, *16*(3), 141-167.

Muellbauer, J. (1974). *The political economy of price indices*. University of London Birkbeck College.

Nicholson, J. L. (1975). Whose cost of living? *Journal of the Royal Statistical Society*

Series A (General), *138*(4), 540-542.

Nicita, A. (2009). The price effect of tariff liberalization: Measuring the impact on household welfare. *Journal of Development Economics*, *89*(1), 19-27.

Ogle, J. (2013). Corn, Free Trade and the Mexican Quest for Food Security. *PSU McNair Scholars Online Journal*, *3*(1), 19

Oswald, A. J. (1997). Happiness and economic performance. *The Economic Journal*, *107*(445), 1815-1831.

Pollak, R. A. (1980). Group cost-of-living indexes. *The American Economic Review*, *70*(2), 273-278.

Pollak, R. A. (1981). The social cost of living index. *Journal of Public Economics*, *15*(3), 311-336.

Porto, G. G. (2006). Using survey data to assess the distributional effects of trade policy. *Journal of International Economics*, *70*(1), 140-160.

Prais, S. J. (1959). Whose cost of living?. *the review of economic Studies*, *26*(2), 126-134.

Ravallion, M. (1990). Rural welfare effects of food price changes under induced wage responses: theory and evidence for Bangladesh. *Oxford Economic Papers*, *42*(3), 574-585.

Ravallion, M., Himelein, K., and Beegle, K. (2016). Can Subjective Questions on Economic Welfare Be Trusted?. *Economic Development and Cultural Change*, *64*(4), 000-000.

Report on the slums of Delhi by CGDR and Planning Commission(2011)

Robles, M., Torero, M., and Cuesta, J. (2010). Understanding the Impact of High Food Prices in Latin America [with Comment]. *Economia*, *10*(2), 117-164.

Robusto, C. C. (1957). The cosine-haversine formula. *The American Mathematical Monthly*, *64*(1), 38-40.

Rothschild, M., and Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic theory*, *2*(3), 225-243.

Slesnick, D. T. (1998). Empirical approaches to the measurement of welfare. *Journal of Economic Literature*, *36*(4), 2108-2165.

Son, H. H., and Kakwani, N. (2006). Measuring the impact of prices on inequality: With applications to Thailand and Korea. *The Journal of Economic Inequality*, *4*(2), 181-207.

Stevenson, B., and Wolfers, J. (2009). The Paradox of Declining Female Happines. *American Economic Journal: Economic Policy*, *1*(2), 190-225.

Stevenson, B., and Wolfers, J. (2008). *Economic growth and subjective well-being:*

Reassessing the Easterlin paradox (No. w14282). National Bureau of Economic Research.

Stutzer, Alois and Bruno S Frey (2002); “Happiness and Economics”; Princeton University Press (2002)

Topalova, P. (2007). Trade liberalization, poverty and inequality: Evidence from Indian districts. In *Globalization and Poverty* (pp. 291-336). University of Chicago Press.

Topalova, P. (2010). Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India. *American Economic Journal: Applied Economics*, *2*(4), 1-41.

Van Brummelen, G. (2013). *Heavenly mathematics: The forgotten art of spherical trigonometry*. Princeton University Press.

Van Praag, B., and Baarsma, B. E. (2005). Using happiness surveys to value intangibles: The case of airport noise. *The Economic Journal*, *115*(500), 224-246.

Veenhoven, R., d Hagerty, M. (2003). Rising happiness in nations 1946–2004: A reply to Easterlin. *Social Indicators Research, 79*, 421–436.

Welsch, H. (2006). Environment and happiness: Valuation of air pollution using life satisfaction data. *Ecological Economics*, *58*(4), 801-813.

Zahniser, S., and Coyle, W. T. (2004). *US-Mexico corn trade during the NAFTA era: new twists to an old story*. US Department of Agriculture, Economic Research Service.