ECONOMIC IMPLICATIONS OF CHANGING HOUSEHOLD MANAGEMENT OF BASIC NEEDS FOR POVERTY ALLEVIATION AND SUSTAINABILITY POLICY, AN ILLUSTRATION OF FOOD AND ENERGY

Dissertation

to attain the doctoral degree (Dr. rer. pol.)

of the Faculty of Forest Sciences and Forest Ecology/Agricultural Sciences Georg-August-Universität Göttingen

Submitted by

Lucie Louise Maruejols

Born on 18.07.1985 in Rennes

UNIVERSITÄT GÖTTINGEN

Göttingen, November 2022

1st Reviewer: Prof. Xiaohua Yu, Ph. D.

2nd Reviewer: Prof. Dr. Meike Wollni

3rd Reviewer: Prof. Dr. Thomas Herzfeld

Additional Examiner: Prof. Dr. Stephan von Cramon-Taubadel

Date of Oral Examination: 19.07.2022

SUMMARY (ENGLISH)

The sharp reduction of extreme poverty worldwide and improvement to living standards in the last 20 years are remarkable and imply that a large number of people in developing countries are able to afford more than their minimum subsistence. However, without adequate protection mechanisms, this progress remains fragile and can easily be reverted, as seen by the devastating socio-economic effects of the COVID-19 pandemic restrictions and disturbances to the economies. Households situated just above the poverty line remain vulnerable, which points to the evident need for continuous public support to secure their position out of poverty. However, as income increases sufficiently to exit extreme poverty, households' preferences shift and their strategies to fulfill their basic needs as well. These modifications have implications for the design of appropriate support programs, to help individuals stabilize and improve their quality of life, but also encourage sustainable choices.

This dissertation expands knowledge on mechanisms of household consumption and fulfillment of basic necessities, at various levels of income. Specifically, food and energy consumption are examined, as they are evidently required for basic human development and well-being, but their consumption patterns change noticeably with income, to better quality diets and to modern energy sources respectively. Accounting for these changes is a condition for the success of interventions to secure decent living standards and sustainability. Specifically, the four essays of this work provide illustrations for rural populations of selected Asian countries and considers the choice of income versus non-income forms of support under different settings.

In particular, essay 1 examines the changing preferences of households for home-produced food over market purchases at various income levels. A theoretical framework places this arbitrage under the constraints of minimum food and minimum cash requirements to be fulfilled for survival. A nonlinear relationship appears between the share of home food in total food expenditure and income. Therefore, contrary to the view that home food is associated with subsistence, the share of home-produced food does not always shrink with better incomes. Using data from rural India, a semiparametric analysis shows the pattern is stronger for households whose main activity is not farming. Analysis of market changes or policy intervention on food consumption patterns should acknowledge that higher home-food consumption may not always represent a sign of impoverishment.

Essay 2 addresses the persisting perception of poverty among Chinese households despite rising incomes. In order to alleviate this feeling, support programs require first an identification of which households are concerned with subjective poverty. Using machine learning techniques, it is possible to make good predictions of these households and identify the factors associated with subjective poverty. The analysis offers specific attention to modest-income households, who may feel poor but not be identified as such by objective poverty lines. The findings highlight that a combination of low income, low endowment (land, consumption assets) and unusual large expenditure (medical, gift) constitutes the key predictor of feeling poor for the middle-income households. Therefore, poverty alleviation policy should continue to focus on increasing incomes, but improvements in non-income domains such health system can also relieve the feeling of income inadequacy.

Next, rising incomes are thought to offer households the opportunity to transition from a situation of energy poverty, marked by consumption of harmful and inefficient energy sources (charcoal, kerosene, biomass) to healthier and more sustainable choices, such as electricity. However, many rural households of developing countries continue to rely on biomass despite rising incomes and having access to grid electricity. The next two essays address obstacles to rural households' energy transition.

The third essay investigates the reasons ethnic minorities continue to use lower amounts of electricity than majority group. The case of highly ethnically-diverse Vietnam offers an opportunity to examine the barriers to energy transition despite availability of grid electricity and rising income thanks to economic growth. A machine learning technique operates selection of most relevant covariates, which are used in a mediation analysis to show the direct and indirect effects of ethnicity on energy consumption. The results reveal different effects of rising incomes between the dominant and the minority groups, which highlight the existence of non-income barriers for ethnic minorities to increase electricity use and operate a transition toward clean and safe energy use.

The fourth essay examines the policy instruments of direct income transfer versus a continuation of subsidized progressive prices, to support consumption of electricity by rural users in India. The essay proposes an ad hoc analytical framework to include the increasing block price structure in a modified LA-AIDS model and compute demand elasticities with respect to marginal price and income. First, the resulting price elasticities are higher with respect to marginal prices than average price, and the gap is quantified. Second, rural consumers are

more sensitive to price change but less sensitive to income change than urban consumers, which suggests that tariffs should be differentiated between rural and urban sectors. Third, targeting urban, large and medium users with price increases would be effective in revenue collection, because their demand is more inelastic, especially in the context of relatively low progressivity of rates within these categories. Price increases applied uniformly and/or compensating poor users by direct cash transfer tend to benefit urban and large users disproportionally.

Overall, these four essays make contributions to the research fields of nutrition and energy use of vulnerable households, in the context of fluctuating income levels. First, they offer thematic and policy contributions by explaining preferences for home-produced food consumption, understanding the factors associated with the feeling of poverty after households have exited extreme poverty, lifting the veil on the ethnic factors that maintain traditional energy consumption patterns, and understanding rural users' demand response to changes in progressive electricity prices. Second, the essays offer methodological contributions, by building a theoretical model of household preferences that integrates arbitrage between various basic needs and various consumption options, by developing new empirical frameworks to capture the effect of ethnicity and deal with increasing block price structure in elasticities studies, and by applying new empirical approaches that rely on machine learning techniques to improve accuracy and robustness of estimations.

ZUSAMMENFASSUNG (DEUTSCH)

Der starke Rückgang der extremen Armut weltweit und die Verbesserung des Lebensstandards in den letzten 20 Jahren sind bemerkenswert und implizieren, dass sich eine große Anzahl von Menschen in Entwicklungsländern mehr als das Existenzminimum leisten können. Ohne angemessene Schutzmechanismen bleibt dieser Fortschritt jedoch fragil und kann leicht rückgängig gemacht werden, wie die verheerenden sozioökonomischen Auswirkungen der COVID-19-Pandemiebeschränkungen und -störungen auf die Volkswirtschaften zeigen. Haushalte, die knapp über der Armutsgrenze liegen, bleiben gefährdet, was auf die offensichtliche Notwendigkeit einer kontinuierlichen öffentlichen Unterstützung hinweist, um ihre Position aus der Armut zu sichern. Wenn das Einkommen jedoch ausreichend steigt, um der extremen Armut zu entkommen, ändern sich die Präferenzen der Haushalte und auch ihre Strategien zur Erfüllung ihrer Grundbedürfnisse. Diese Änderungen wirken sich auf die Gestaltung geeigneter Unterstützungsprogramme aus, um Haushalten dabei zu helfen, ihre Lebensqualität zu stabilisieren und zu verbessern, aber auch nachhaltige Entscheidungen zu fördern.

Diese Dissertation erweitert das Wissen über Mechanismen des Haushaltskonsums und der Befriedigung von Grundbedürfnissen für verschiedene Einkommensniveaus. Insbesondere wird der Lebensmittel- und Energieverbrauch untersucht, da sie offensichtlich für die grundlegende menschliche Entwicklung und das Wohlbefinden erforderlich sind, sich ihre Konsummuster jedoch mit dem Einkommen deutlich ändern, hin zu jeweils einer qualitativ hochwertigeren Ernährung und modernen Energiequellen. Die Berücksichtigung dieser Veränderungen bedingt den Erfolg von Interventionen zur Sicherung eines angemessenen Lebensstandards und Nachhaltigkeit. Insbesondere liefern die vier Essays dieser Arbeit Illustrationen für die ländliche Bevölkerung ausgewählter asiatischer Länder und betrachten die Wahl von Einkommens- versus Nicht-Einkommensinterventionen unter verschiedenen Bedingungen.

Essay 1 untersucht gezielt die sich ändernden Haushaltspräferenzen für selbst produzierte Lebensmittel gegenüber Markteinkäufen auf verschiedenen Einkommensniveaus. Ein theoretischer Rahmen stellt diese Arbitrage unter die Beschränkungen von Mindestnahrungs- und Mindestbargeldanforderungen, die zum Überleben erfüllt werden müssen. Es zeigt sich ein nichtlinearer Zusammenhang zwischen dem Anteil der selbst produzierten Lebensmittel an den

gesamten Nahrungsmittelausgaben und -einnahmen. Entgegen der Ansicht, dass die selbst produzierten Lebensmittel mit dem Lebensunterhalt verbunden ist, sinkt der Anteil der selbst produzierten Lebensmittel nicht immer mit einem besseren Einkommen. Unter Verwendung von Daten aus dem ländlichen Indien zeigt eine semiparametrische Analyse, dass das Muster bei Haushalten, deren Haupttätigkeit nicht die Landwirtschaft ist, stärker ausgeprägt ist. Die Analyse von Marktveränderungen oder politischen Interventionen in Ernährungsmuster sollte anerkennen, dass ein höherer Lebensmittelverbrauch zu Hause nicht immer ein Zeichen von Verarmung sein muss.

Essay 2 befasst sich mit der anhaltenden Armutswahrnehmung chinesischer Haushalte trotz steigender Einkommen. Um dieses Gefühl zu lindern, erfordern Förderprogramme zunächst eine Identifizierung, welche Haushalte von subjektiver Armut betroffen sind. Mithilfe von Techniken des maschinellen Lernens ist es möglich, gute Vorhersagen über diese Haushalte zu treffen und die Faktoren zu identifizieren, die mit subjektiver Armut verbunden sind. Die Analyse widmet Haushalten mit bescheidenem Einkommen besondere Aufmerksamkeit, die sich möglicherweise arm fühlen, aber durch objektive Armutsgrenzen nicht als solche identifiziert werden. Die Ergebnisse verdeutlichen, dass eine Kombination aus niedrigem Einkommen, geringer Ausstattung (Land, Konsumgüter) und ungewöhnlich hohen Ausgaben (medizinische Versorgung, Geschenke) den wichtigsten Prädiktor dafür darstellt, dass sich Haushalte mit mittlerem Einkommen arm fühlen. Daher sollte sich die Armutsbekämpfungspolitik weiterhin auf die Erhöhung der Einkommen konzentrieren, aber auch Verbesserungen in Nicht-Einkommensbereichen wie Gesundheitsausgaben können das Gefühl der Einkommensunzulänglichkeit lindern.

Außerdem wird angenommen, dass steigende Einkommen den Haushalten die Möglichkeit bieten, von einer Situation der Energiearmut, die durch den Verbrauch schädlicher und ineffizienter Energiequellen (Holzkohle, Kerosin, Biomasse) gekennzeichnet ist, zu gesünderen und nachhaltigeren Entscheidungen wie Elektrizität überzugehen. Allerdings sind viele ländliche Haushalte in Entwicklungsländern trotz steigender Einkommen und des Zugangs zum Stromnetz weiterhin auf Biomasse angewiesen. Die nächsten beiden Aufsätze befassen sich mit Hindernissen für die Energiewende ländlicher Haushalte.

Der dritte Essay untersucht die Gründe, warum ethnische Minderheiten weiterhin weniger Strom verbrauchen als die Mehrheitsgruppe. Der Fall des ethnisch sehr heterogenen Vietnams bietet die Gelegenheit, die Hindernisse für die Energiewende trotz der Verfügbarkeit von

Netzstrom und steigenden Einkommen dank des Wirtschaftswachstums zu untersuchen. Eine maschinelle Lerntechnik führt eine Auswahl der relevantesten Kovariaten durch, die in einer Mediationsanalyse verwendet werden, um die direkten und indirekten Auswirkungen der ethnischen Zugehörigkeit auf den Energieverbrauch aufzuzeigen. Die Ergebnisse offenbaren unterschiedliche Auswirkungen steigender Einkommen zwischen Haushalten der Mehrheits- und der ethnischen Minderheitengruppe was die Existenz von Nicht-Einkommensbarrieren für ethnische Minderheiten hervorhebt, um den Stromverbrauch zu erhöhen und einen Übergang zu einer sauberen und sicheren Energienutzung zu ermöglichen.

Der vierte Essay untersucht die politischen Instrumente des direkten Einkommenstransfers im Vergleich zu einer Fortsetzung von subventionierten, progressiven Preisen, zur Unterstützung des Stromverbrauchs durch den ländlichen Kundenstamm in Indien. Der Essay schlägt einen analytischen Ad-hoc-Rahmen vor, um die steigende Blockpreisstruktur in ein modifiziertes LA-AIDS-Modell einzubeziehen und Nachfrageelastizitäten in Bezug auf Grenzpreis und Einkommen zu berechnen. Erstens sind die resultierenden Preiselastizitäten höher, wenn Grenzpreise betrachtet werden, und der Abstand zu Elastizitäten basierend auf Durchschnittspreisen wird quantifiziert. Zweitens reagieren ländliche Verbraucher empfindlicher auf Preisänderungen, aber weniger empfindlich auf Einkommensänderungen als städtische Verbraucher, was darauf hindeutet, dass die Tarife zwischen ländlichen und städtischen Sektoren differenziert werden sollten. Drittens wäre eine Ausrichtung auf städtische, mittlere und große Nutzer mit Preiserhöhungen bei der Einnahmenerhebung effektiv, da ihre Nachfrage unelastischer ist, insbesondere im Zusammenhang mit einer relativ geringen Tarifprogression innerhalb dieser Kategorien. Einheitlich angewandte Preiserhöhungen und/oder Ausgleichszahlungen für arme Nutzer durch direkte Bargeldtransfers kommen städtischen und großen Nutzern tendenziell überproportional zugute.

Insgesamt leisten diese vier Aufsätze Beiträge zu den Forschungsfeldern Ernährung und Energienutzung von vulnerablen Haushalten im Kontext schwankender Einkommensniveaus. Erstens bieten die Essays thematische und politische Beiträge, indem sie Präferenzen für den Konsum von selbst produzierten Lebensmitteln erklären, die Faktoren verstehen, die mit dem Armutsgefühl verbunden sind, nachdem Haushalte aus extremer Armut herausgekommen sind, den Schleier über die ethnischen Faktoren lüften, die traditionelle Energieverbrauchsmuster aufrechterhalten, und Nachfragereaktion der ländlichen Nutzer auf Änderungen der progressiven Strompreise verstehen. Zweitens bieten die Essays methodische Beiträge, indem sie ein theoretisches Modell der Haushaltspräferenzen aufbauen, das die Arbitrage zwischen

verschiedenen Grundbedürfnissen und verschiedenen Konsumoptionen integriert, neue empirische Rahmen entwickelt, um den Effekt der Ethnizität zu erfassen und mit zunehmender Blockpreisstruktur in Elastizitätsstudien umzugehen, und durch die Anwendung neuer empirischer Ansätze, die auf Techniken des maschinellen Lernens beruhen, um die Genauigkeit und Robustheit von Schätzungen zu verbessern.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

First and foremost, I am extremely grateful to my supervisor Prof. Xiaohua Yu for giving me a chance to pursue doctoral studies with him. I benefited greatly from his exceptional knowledge, his innovative mind, and his amazing approachability. Thank you for the great guidance into the world of research, constant support and professional advice. I am also extremely thankful to Ms. Jana Nowakowsky for her efficient administrative support, for her care and attention. Her steady 'gute Laune' lit up many of my days.

In the last few years, I have been fortunate to collaborate with Prof. Hanjie Wang, whose comments helped me become a better researcher, and with Lisa Höschle, whose abilities and cheerfulness make an excellent research partner.

I am lucky to be part of a strong and friendly working group in Göttingen, which carries out inspiring work. I am especially indebted to Jasmin Wehner and Lisa Höschle for their help during the final stages of this work. I also enjoyed crossing paths with students from other research groups. Big or small, frequent or not, all these interactions make work and life better.

My family and my friends gracefully provided support and took interest in my work. I am especially grateful to my mother for her attention and frequent visits. Pursuing Ph. D. studies would not have been possible without the influence and constant encouragement of my spouse Dr. Jitendra Bajpai. I am also grateful to my two wonderful children.

Finally, one thought did not leave me during my entire doctoral studies. It is that of many women who do not have - or take - the chance to feel the gratification of professional achievement, especially after having children. It is sometimes difficult to balance expectations from all sides, but it is well worth it. Male partners, when you provide encouragement and share domestic tasks, you play an important role in this.

In memory of my grandmothers.

CHAPTER 1

GENERAL INTRODUCTION

1.1. Background

1.1.1. Poverty alleviation and policy

Between 1990 and 2018, extreme poverty was reduced from 27.7 percent of the global population to just 8.6 percent, representing a drop of 1,036 million people.[1] This was achieved largely thanks to economic growth and coordinated global efforts around the United Nations' Millennium Development Goals and successively the Sustainable Development Goals (SDGs).

However, the population groups who exited extreme poverty did not become rich. Their existence may not be as precarious as before, but the vast majority still needs support from governments, non-profit organizations and other providers to secure and improve their quality of life. Their vulnerabilities, and the need for social protection, became especially blatant with the Covid-19 pandemic, whose effect reversed some of the poverty reduction success achieved in the last years. The income loss due to lockdowns and health-related restrictions on the economies exacerbated the challenges of climate change and political stability, and led between 119 million and 124 million additional people to fall into poverty in 2020, of whom 60 percent are in Southern Asia (United Nations, 2021). Without adequate social protection for low-income people, gains in income and human development remain fragile and subject to shocks.

Therefore, policy focus on eradicating extreme poverty is closely connected with ensuring that people remain out of poverty. To help fortify the improvement in living conditions and support building sustainable and resilient lives, a key component for policy frameworks and support programs is to ensure that the households can secure provision of their basic needs, for example access to food, water, health, education, energy, and others which form part of the 17 Sustainable Development Goals adopted by the United Nations in 2015 (United Nations, 2015).

[1] Extreme poverty is defined by the World Bank as living on less than 1.90 international dollars per day. Source: https://databank.worldbank.org/reports.aspx?source=poverty-and-equity-database, visited on 7.05.2022.

1.1.2. Income transfers and other forms of support

To help households fulfill the basic needs that contribute to human development, a number of policy instruments have been used over the years, and the discussion about the best form of intervention, overall categorized between in-kind versus cash transfers, remains open. Provision of in-kind support is widespread and encourages consumption of particular goods, by people who need it the most (Cunha et al., 2019). Conceptually, preference for this option is mostly driven by paternalistic attempts to induce certain behavior among recipients, and by attempts to manage externalities that may not enter a single household's decision, such as long-term health outcome of child nutrition or environmental-friendly choices (Currie and Gahvari, 2008; Gentilini, 2016). On the other hand, economists have a preference for the supposed efficiency of cash transfers. Cash transfers give the recipients the freedom to spend the cash where it creates the highest utility for them, thus creating fewer distortions to welfare. Income support is also typically easier to administer and does not carry the stigma attached to receiving in-kind aid for the recipients (Blackorby and Donaldson, 1988; Currie and Gahvari, 2008; Hidrobo et al., 2014). In term of repercussion on local prices and non-recipients, Cunha, De Giorgi, and Jayachandran (2019) predict that the effect of either instrument depends on the local price and market condition context.

Empirical evidence shows positive effects of both in-kind transfers (Slesnick, 1996; Aaberge et al., 2019) and cash transfers to reduce poverty, increase standards, and reduce inequalities. However, the only few studies able to make robust empirical comparisons between the two forms of intervention conclude that the optimal choice depends on the context, for example whether the recipient's local economy is well-integrated and well supplied, or rather remote and less competitive. As summarized by Cunha, De Giorgi, and Jayachandran (2019), many studies concluded that cash and kind transfers can have the same overall effect on total expenditure, but different types of goods are consumed. Therefore, superiority is also subject to the aim of a policy instrument, for example in-kind transfers of food are found to increase consumption of calories while vouchers or cash lead to improvement in dietary diversity (Hoddinott et al., 2013; Hidrobo et al., 2014; Gentilini, 2016).

In practice, cash transfers are being increasingly adopted in the Global South (Tiwari et al., 2016), and many studies showed the positive effect of both conditional (Fiszbein and Schady, 2009; García and Saavedra, 2017) and unconditional transfers for consumption, food security, strengthening rural livelihoods and education (Baird, Ferreira, Özler, and Woolcock, 2013;

Burchi, Scarlato, and D'Agostino, 2018; Handa et al., 2018). However, the literature that compares directly conditional and unconditional transfers is limited or inconclusive, for example Pega et al. (2017). Benefits of income support also depend on whether transfers consist of regular and predictable amounts, or irregular, lump-sum payments (Tiwari et al., 2016; Haushofer and Shapiro, 2016) and whether short-term or long-term effects are considered (Baird et al., 2019).

All in all, both income and in-kind transfers seem to be advantageous, and the superiority of one or the other depends on the context. Part of this context lies in the heterogeneity of recipients' resources as they enter the programs. For example, Cooper et al. (2020) points to heterogenous effect of health-related income transfers on different sub-groups of recipients. However, an explicit discussion about how to best support vulnerable individuals in relation to their income or deprivation level is generally absent from the above-mentioned studies, although needs vary with income level.

1.1.3. Consumer preferences and income

In terms of poverty reduction, economic development and human development, the widely referred-to capability approach developed by Amartya Sen (Sen, 1993, 1988) implies that enhancement of living conditions is best achieved when people have the freedom to make choices that matter to them (Alkire, 2005). According to this view, a policy can best support vulnerable people not by imposing consumption choices uniformly or blindly providing aid, but rather by enabling recipients to make and realize their choices. For support programs, this can take the form of lifting obstacles to make an option available, and letting individuals the freedom to choose among available options. This necessitates to understand households' needs and preferences, especially at different stages of income and poverty.

Generally, Engel's law (1857) characterizes the relation between a household expenditure on a specific good and total household expenditure, and specifies that, as income increases, spending on certain goods decreases while it stays the same or increase on other goods. Much literature has been dedicated to finding the appropriate functional form of the Engel curves for each category of goods, with main contributions of the Working-Leser form (Leser, 1963) and quadratic form Banks et al. (1997). Certain goods, in particular food and domestic fuel, have clear linear decreasing relationship between the log expenditure and the respective budget shares, indicating that household relative preferences shift away from theses goods as income grows.

In addition, the internal composition of each consumption group also changes with the resource level, as exampled here with consumption of food and energy.

Food consumption is marked by structural change along the income range, with a transition from a calorie-dominated consumption at low-income level toward more diversified diets and nutritional improvement attainable at higher income levels. Indeed, survival relies on consumption of sufficient calories, therefore, households under limited budget choose the most affordable sources of calories in priority, such as staple foods, and continue to increase its consumption with income (Deaton and Subramanian, 1996). As recently exampled in Nsabimana, Bali Swain, Surry, and Ngabitsinze (2020) and multiple times before, diets of low-income households are dominated by higher amounts of carbohydrate and starch, and are poor in micronutrients. At some point however, consumption of calories approaches saturation, and households shift preferences toward non-caloric attributes such as taste and micronutrients components (Jensen and Miller, 2010). This was shown to result in calorie-income elasticities that diverge at different income stages (Zhou and Yu, 2014), as well as in a non-linear relationship between nutrition intake and income in the form of economic growth (Tian and Yu, 2015), althought the connection between calories and income is not always well-defined (Deaton and Drèze, 2009). Overall, as income reaches a sufficiently high level, new types of food enter consumption, which result in more diversified diets (Gupta et al., 2020; Sibhatu et al., 2015; Tian and Yu, 2015).

Energy consumption patterns also change drastically as income grows. Described by Leach (1992) and Dowd (1989), this 'energy transition'[2] is a shift in energy consumption of developing countries households from patterns dominated by solid fuels (traditional biomass and coal), to increasing shares of modern sources such as kerosene, electricity and natural gas. Solid fuels still represented 85 percent of rural energy use in years 2000's in China and India, while electricity represented about 10 percent of expenditures. Urban and richer households are typically ahead of rural households in their transition, with more even shares of traditional and modern fuels (Pachauri and Jiang, 2008). Leach (1992) indicated that the main obstacles to the transition reside in access to modern energy sources and to the equipment necessary to use that energy, rather than in relative prices of the various fuels. A collection of evidence added that income is clearly a major driver of the fuel switching behavior (Ekholm et al., 2010; Han et al.,

[2] More recently, the term 'energy transition' has been used to describe shifts of national level energy systems from the use of high carbon emitting, non-renewable sources to use of low carbon emitting renewable sources. Although related, this transition is slightly different than that operating at the household level, which is the main focus here.

2018; Muller and Yan, 2018; Pachauri and Jiang, 2008; Yawale et al., 2021). In details, the transition was thought to take the form of an 'energy ladder' that better-off households climb when they gradually replace traditional fuel sources by kerosene, LPG and later electricity and piped gas (Hanna and Oliva, 2015; Hosier and Dowd, 1987; Van der Kroon et al., 2013). However, empirical evidence shows that households in fact adopt an 'energy stacking' strategy, whereby they integrate additional use of modern sources when their income allows it, but without completely discarding traditional sources (Choumert-Nkolo et al., 2019; Guta, 2012).

All in all, these transitions demonstrate that, although basic needs are universal and independent from income level, fulfilling them is not satisfied in consistent ways by all households, which results in different consumption patterns at different income levels. Awareness of these patterns, transitions and varying preferences is key to design effective support programs.

1.2. Research question

The previous section reminds us how extreme poverty has been drastically reduced, but gains remain fragile. It is critical to support vulnerable people in securing satisfaction of the needs that contribute to better living standards. However, the best forms of support depend on context, and part of this context is the household's own resources level, which determine how households satisfy their basic needs.

The four essays of this dissertation examine how to best support vulnerable households achieve better living conditions, while accounting for households' preferences and consumption patterns. Illustrations with food and energy are proposed; although basic needs, living conditions and human development depend on achievements in many additional areas.[3]

Access to food is evidently required for basic subsistence, and is so critical that it forms the second SDG after ending poverty. Particularly, SDG 2.1 aims to "end hunger and secure access by all people […] to safe nutritious and sufficient food all year round." Access to sufficient calories is necessary for survival and functioning, but transition to more diversified diets brings higher consumption of micronutrients. Better diets and nutrition are linked to reduced malnutrition prevalence and better health outcomes (Arimond and Ruel, 2004; Headey and

[3] In this view, the second essay adopts a general definition of deprivation which is not restricted specifically to food or energy.

Ecker, 2013; Lim et al., 2012), but also to long-term drivers of economic growth (Hoddinott et al., 2008). In addition, households often orient their productive activities toward food access, either via home production of food or via income generating activities. For example, 75 percent of the world agricultural land is estimated to be operated by family farms.[4]

Energy is also necessary, to accomplish daily life tasks such as cooking, lighting, communicating or powering medical devices, but also to develop small businesses. SDG 7 recognizes its importance and aims to "ensure access to affordable, reliable, sustainable and modern energy for all'. In particular, SDG 7.1 emphasizes a primary reliance on *clean* fuels and technology, such as electricity and natural gas which have lower carbon emissions than traditional biomass and solid fuels. Transition to clean fuels has major consequences for the health of individuals: household indoor air pollution from solid fuels was one of the three leading risk factors for global disease burden in 2010, according to Lim et al. (2012). It also has consequences on deforestation of certain areas and global warming (Masera et al., 2015). In this domain, literature on energy poverty identifies that, given the recent progress in rural electrification (almost 100 percent households are now connected to an electricity source in India, China, Vietnam…), the main obstacles lie in the hours of electricity supply available per day and its affordability for poor households (Agarwal et al., 2020; H. Wang et al., 2021). In addition, the field of energy justice identifies the population groups that bear unequitable burden of energy use and are not able to operate energy transition (Jenkins et al., 2016; Reames, 2016).

Specific to both food and energy, individuals can procure these goods either by purchases from the market, or by self-production and self-collection. Food can be home-grown by households with access to land, and firewood, agricultural residues or dung cakes can be freely collected to provide cooking energy, heat and light. This access to own in-kind resources is often not explicitly considered in design or evaluation of support intervention, although it represents non-negligible shares of household consumption and therefore modifies consumption patterns of purchased goods (Bhattacharyya and Timilsina, 2010). Therefore, analyses accounting for households' resources and deprivation status should include in-kind resources.

In summary, a clear improvement to individual living standards can be achieved by sufficient consumption of nutritious food and modern energy. Specifically, the essays of this dissertation highlight the mechanisms that allow or prevent households' satisfaction of basic needs, relative

[4] https://www.fao.org/land-water/overview/covid19/smallholders/fr/, visited on 11.05.2022.

to their resources and income level. In particular, this work aims to make policy recommendations on the choice of income transfers as opposed to other forms of support for modest-income households.

1.3. Content of dissertation

1.3.1. Conceptual framework

Figure 1.1 proposes a conceptual framework of the different steps and mechanisms through which basic needs are identified and fulfilled by the households. This forms the context of the dissertation.

First, individuals or households seek to identify their basic needs and the resulting goods and services they will consume in priority. Although this is an individual decision, general measures are available. For example, between 1700 to 2400 calories per day is often considered a requirement for an adult male doing strenuous work (Kaicker and Gaiha, 2013), and a consumption of 30 to 50 kWh per month and per household is deemed necessary to cover basic domestic energy needs (Goldemberg, 1990). More generally, the World Bank places a universal threshold at 1.90 USD per day per person to cover all minimum needs.

Second, households aim to fulfill these needs with their monetary and non-monetary resources. They can arbitrate between consumption of own products, for example home production of food and self-collection of firewood, or consumption of market-purchased goods. Finally, consumption of market goods naturally necessitates adequate supply of the good and affordable price. Otherwise, the household may choose to revisit their balance between market and non-market good consumption. Overall, it is clear that household choices made at each step depend on income and resources available to the household.

The chart also shows examples of possible interventions and support program (in-kind, direct cash transfer, subsidized price and public good such as infrastructure) and how they interact with the household decision-making process. As household decisions change according to their resources and needs, so will the effect of the various interventions.

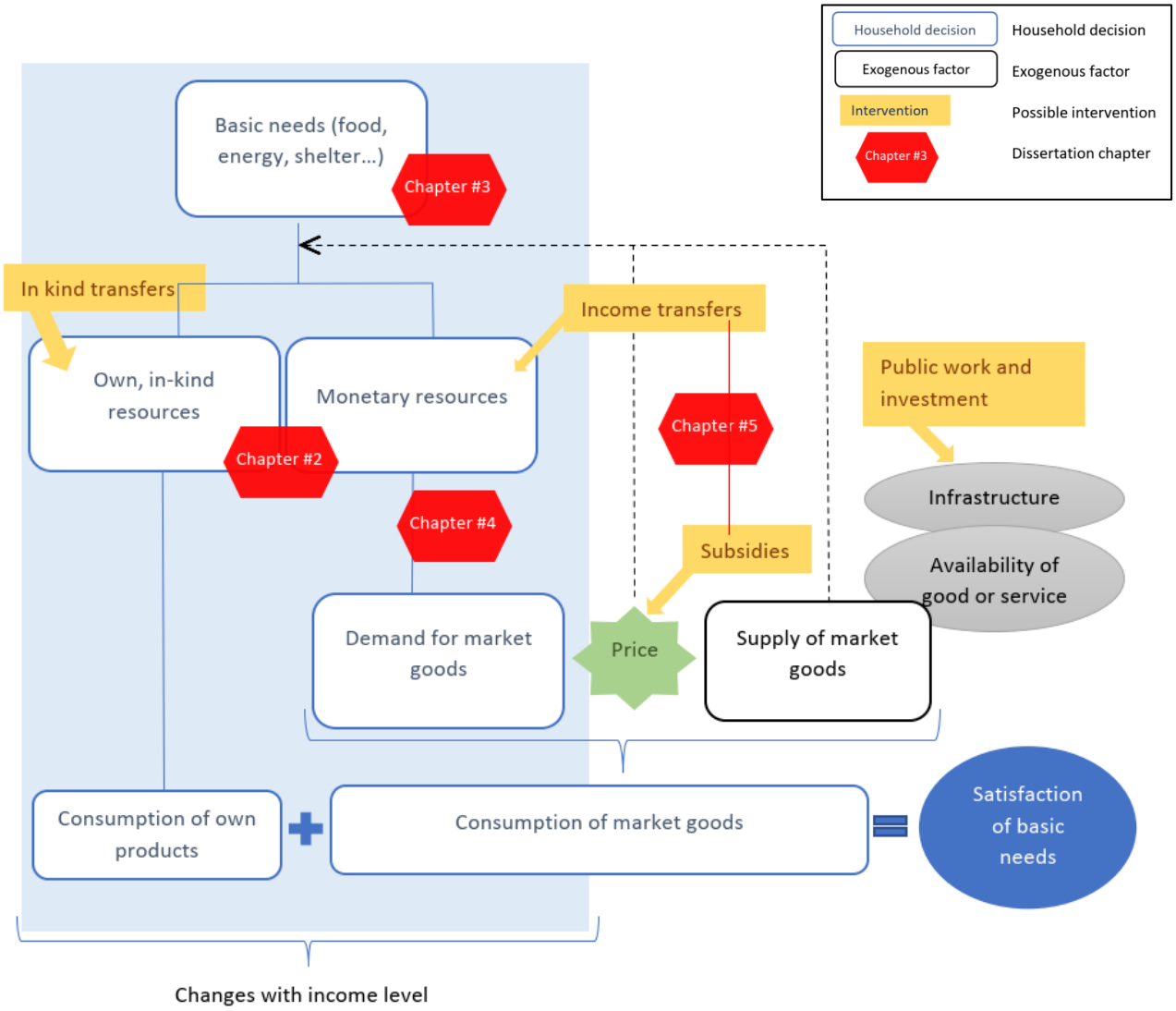

Figure 1. 1: Conceptual framework of the dissertation

1.3.2. Essays

Figure 1 allows visualizing where the chapters of this dissertation fit within the context of poverty, basic needs and support intervention.

The first essay (chapter 2) builds a theoretical framework that explains households' arbitrage between market-purchased and self-produced foods in context of basic needs. Specifically, the chapter shows that, contrary to expectations, consumption of home-produced food does not necessarily decrease as income grows. In particular, for some household categories, consumption of home-produced food is higher at moderate income than at very low income level. Therefore, access to non-monetary resources (land), together with income level, determine preferences for market-purchased food over in-kind foods. This has implication for

the interpretation of own-food consumption, for example in the design or evaluation of support programs, policies, or market changes.

The second essay (chapter 3) is concerned with the identification of poor households in order to best respond to the feeling of deprivation in a population. In the Chinese context where extreme poverty has been mostly eradicated, households' own definition of poverty as well as their definition of basic needs deviates from objective poverty standards. Therefore, the chapter explores techniques to predict which households feel poor and the factors associated with the feeling of poverty. The analysis detects both income and non-income factors to be associated with subjective poverty, and thus identifies potential non-income factors that could be targeted for alleviating the feeling of income inadequacy.

The third essay (chapter 4) examines obstacles to rural households' transition from biomass and coal consumption to electricity, once access to the electricity grid is granted and income is sufficient to operate the change. In fact, the paper finds that ethnic minorities are disproportionally affected by the presence of non-income barriers that prevent them from increasing electricity consumption as much as the dominant group with similar income level.

The last essay (chapter 5) explores the properties of two policy tools, namely subsidized electricity rates and income transfer, to maintain affordability of electricity to newly connected, rural and small users, in order to promote more sustainable energy usage and replacement of harmful traditional fuels usage, but also sufficient collection of revenue for the utilities. The results propose adjustments to the progressivity of electricity rates to account for the specificities of the vulnerable and rural users.

1.3.3. Contributions

The essays make a number of theoretical, methodological and policy contributions.

The first essay builds for the first time a theoretical framework explaining consumption of own-produced food, not as a side effect of production decisions, but as a conscious consumption arbitrage of households aiming to satisfy basic needs. Satisfaction of food needs is not seen in isolation from other basic needs, as the model accounts for the arbitrage of households between various conflicting needs. In addition, the model integrates previous theoretical explanations of own food consumption as a result of market uncertainties.

Contributions to methodological approaches are found in the third essay, which emphasizes the potential indirect effect ethnicity can have on electricity consumption through other determinants. Specifically, the methodological framework captures the interactions between income and ethnicity on the outcome and uncovers the nonlinear effect of income on electricity consumption. In the fourth essay, a new solution is proposed to account for the specificities of prices in step structure, where customers of different volumes pay different marginal prices, for the evaluation of elasticities. This completes previous techniques which used instrument variable approaches to account for endogeneity, or calculated elasticities at pre-determined volumes of consumption.

Contribution to estimation techniques are found in the second essay which compares the attributes and performance of several machine learning techniques in the prediction of poverty when it is subjectively defined, and show that this perception can be predicted with good levels of accuracy. A machine learning technique is also employed in essay 3 for the selection of the most relevant explanatory factors among a high number of potential covariates. This allows selecting the variables not only based on researcher's assumptions and findings from other contexts, but rather on the variables that matter most within the specific data available.

Policy contributions are found in each essay. The first essay brings a new light on the behavior of consumption own-produced food and shows that the interest of cash versus in-kind transfer depends on the recipients' consumption patterns and own resources. In particular, farming households have little interest in additional in-kind calorie-rich foods when their resources increase and prefer to consume foods they can purchase. On the other hand, other workers with vulnerable income prefer increasing in-kind calorie-rich foods when more resources become available. Results from essay two show that the feeling of income inadequacy is clearly linked to income level for very low-income households, for who measures of direct income support are clearly beneficial. However, for modest-income households, the feeling of income inadequacy is also associated with large expenditure (e.g. medical, gift). Therefore, non-income support and improving social protection such as health insurances could help improve wellbeing of these households. The third essay shows that raise in income translate into smaller raise in electricity consumption for ethnic groups than for members of the dominant groups. This indicates that households from ethnic minorities face non-income barriers that prevent them from transitioning to modern fuel sources as fast as their counterparts of the dominants group. Clearly, focusing on increasing incomes is not sufficient and support programs should emphasize lifting these non-income barriers. Finally, essay four shows that rural and small users

of electricity have different income and price elasticities than urban and large users, which has implications for the current government's policy to make uniform changes to prices while compensating poor users by direct income transfers. Such changes would harm small and rural users disproportionally, while not bringing as much revenue collection as a more discriminated rise of electricity price would.

CHAPTER 2

HOUSEHOLD PREFERENCES FOR HOME FOOD CONSUMPTION SUBJECT TO FOOD AND CASH CONSTRAINTS, UNDER MARKET IMPERFECTIONS: AN INVERSE-U RELATIONSHIP WITH INCOME[5]

Abstract

Production of food for own consumption makes up an important part of rural households' diets in emerging and less privileged economies. However, the role of home products in nutrition and the drivers of households' arbitrage between home and market-purchased food are little understood. This creates an obstacle to analyze the effects of market changes or prospective interventions on nutrition, poverty and welfare outcomes. This essay proposes a framework that shows how minimum food and minimum cash requirements modify the preferences for or against home-produced food over market-purchased food, under market imperfections. These food and cash constraints generate an inverse U-shape relationship between the share of home food in total food expenditure and income. Using data from rural India, a semiparametric analysis shows the inverse U-shape pattern is stronger for households whose main activity is not farming, who first increase home food consumption when their income grows, before reducing it at higher income levels. This confirms home-produced food can be preferred and contributes to satisfaction of basic food needs. The possibility to cultivate provides households some flexibility to manage basic food needs.

[5] The authors of this essay are Lucie Maruejols and Xiaohua Yu. All authors contributed to conceptualize the research idea. Lucie Maruejols conducted the data analysis, results discussion and writing the paper. Xiaohua Yu provided critical feedback and revisions. We are grateful to Prof. Sebastian Vollmer for comments on an earlier version of this essay.

2.1. Introduction

Misconceptions about poor households' food choices and their subsistence strategies is a barrier to making appropriate policy interventions. Generally, to ensure that households are able to meet their basic nutritional needs, countries have developed programs with strong focus on affordable calories (Behrman et al., 1988; Deaton and Drèze, 2009; Jensen and Miller, 2010; Headey and Ecker, 2013), which sometimes limits effectiveness on improving nutrition (Bhagowalia and Chandna, 2011; Krishnamurthy et al., 2017; Pingali et al., 2017; Chhotray et al., 2020). These policies assume that households turn to exterior sources to fulfill their food needs, but ignore non-market mechanisms. On the contrary, the widespread inclusion of home-produced food in diets indicates that consuming these products is part of the households' strategies to fulfill their basic requirements.

Home food consumption is typically interpreted as the result of subsistence-oriented farming from households who have limited access to market, while agricultural households who follow market-oriented farming can afford market-purchased foods, and consume better diets (Kirimi et al., 2013; Frelat et al., 2016; Jones, 2017; Sibhatu and Qaim, 2017, Ogutu et al., 2020). However, home food consumption persists among the rural population of many countries, and surprisingly also when commercialization and income grow (Ogutu et al., 2020). In parallel, purchased food plays an important role not only for market-oriented farmers but also for households with low income and marketization level (Von Braun and Kennedy, 1994).

The underlying reasons for these patterns are generally poorly understood (de Janvry and Sadoulet, 2011). Existing household production models have integrated the possibility for farming households to consume part of their production (Nakajima, 1986; Von Braun and Kennedy, 1994; Janvry et al., 1991**;** Sadoulet and de Janvry, 1995) but the decision to consume own products is mostly viewed as a result of production choices (i.e. labour and time decisions, finding the optimal level of marketization, choice of crop). The existing models do not fully explain the mechanisms underlying home food consumption. In particular, it remains unclear how home consumption behavior changes when income increases and basic requirements for food are gradually met.

This is not trivial because consumption of home-produced foods potentially interacts with the nutrition policies in place, or at least their evaluation, in several ways. First, Janvry et al. (1991) explain how the selling and purchasing prices, relative to the shadow value of food, determine the amount of production allocated to own-consumption. For example, India's system of

subsidized grains and procurement price (Aditya et al., 2017; Chhotray et al., 2020; George and McKay, 2019) is already known to alter choices in crops production (Morales et al., 2021) and consumption (Pingali et al., 2017) by altering relative prices. However, it remains unclear how these interactions change when income grows and basic food requirements are fulfilled.

Second, the response of producer-consumer households to food prices variations or policies is difficult to anticipate due to the ambiguity of their dual role (Janvry et al., 1991; Sadoulet and de Janvry, 1995; Ravallion, 2000; Taylor and Adelman, 2003; Gillespie et al., 2012). The existing studies mostly document the production decisions of farmers and the resulting income effects, but the overall effects often remain unclear, partly because the underlying role and importance of home-produced food in food consumption is unknown.

Third, delivery of nutritional support to vulnerable households can take the form of either cash or in-kind transfers (Farrington and Slater, 2006; Khera, 2014; Narayanan, 2011). Cash is increasingly used as replacement or together with in-kind transfers, for example in the Indian context since the 2013 National Food Security Act (Sinha and Patnaik, 2016). So far, evaluations of the household responses to cash and in-kind transfers do not fully account for access to home consumption. It appears that the superiority of either in-kind or cash transfers depends firstly on context (Narayanan, 2011) and on socio-economic and subsistence status of the households (Khera, 2014, Ghatak et al., 2016). Therefore, anticipating own-consumption behavior to fulfill food needs is necessary in the analysis of basic food needs satisfaction and the design and evaluation of nutritional support programs.

This essay proposes a model of consumption arbitrage between home-produced and market-purchased food, where home food consumption is used to mitigate costs in providing basic food requirements. A consumer utility maximization model is used and is consistent with household production theory and with consumers valuing calories but also taste and other food attributes. Basic needs are modelled with a subsistence constraint, which reflects the minimum amount of energy (calories) below which surviving is possible but difficult and uncomfortable. The model tolerates living with a sub-optimal level of calories but then fulfilling this need becomes a priority, obviating other pursuits such as adding taste and diversity in the diet (Jensen and Miller, 2010). The model shows that households do not always substitute toward purchased food when more income becomes available, due to the subsistence concern.

This model is further integrated into a larger framework where households must also satisfy basic non-food needs. Indeed, food needs are not managed in isolation of other needs. Expenses

such as housing costs, energy, debt repayments, medical expenses and others, play a role on the food security of households (Headey and Ecker, 2013). Fulfilling non-food needs is represented by a need for cash, which can at times be stronger than the need for additional calories. Indeed, Gelan (2006) shows that some in-kind food transfers were sold for cash by their recipients, revealing either that the type of food given is inappropriate or that the households have a preference for cash. Balasubramanian (2015) further shows that families treat the amount saved from subsidized food as cash and spend it on goods other than cereals, in particular non-food goods. Poor households, despite small scale and limited resources, make rational decisions and allocate their resources efficiently (Schultz, 1980) so as to maximize satisfaction of both food and non-food needs. In any case, the model examines how the cash constraint affects food consumption in general and the choice between own products or market products in particular, as it becomes more stringent.

In addition to food and nonfood constraints, households decisions are modelled under conditions of market imperfections, as missing markets are already associated with persistence of subsistence farming (Janvry et al., 1991). All in all, the model explains how preference for affordable calories first increases the home-produced food consumption, before the satisfaction of basic needs lowers it. An empirical illustration with nationally representative food expenditure data from India confirms the nonlinear relationship between the share of home food in diet and household income.

2.2. Background: home food consumption vs. market purchases

2.2.1. Different nutritional roles

The nutritional benefits of subsistence farming (production for own-consumption) have been recognized to be limited but valuable when markets are difficult to access (Luckett et al., 2015; Sibhatu et al., 2015) or unfavorable (high prices) (Baiphethi and Jacobs, 2009). Due to low incomes, households who rely on subsistence farming and home food consumption achieve lower nutritional outcomes than their commercialized counterparts (Frelat et al., 2016; Jones, 2017; Sibhatu and Qaim, 2017, Ogutu et al., 2020). Nevertheless, empirical evidence shows that commercialized households do not abandon their home food consumption altogether as their income increases, while subsistence farmers do not fill all their needs with self-produced foods only.

Instead, both own consumption and purchased food hold important roles in diets across all levels of income and marketisation. Purchased food often constitutes more than half of food consumption of rural households in developing countries, even for small holders, at low income levels, or in high subsistence context (Jones, 2017; Luckett et al., 2015; Rais et al., 2009; Sibhatu et al., 2015; Sibhatu and Qaim, 2018). Specifically, Jones (2017) reports that, in a mostly subsistence context of Malawi, the proportion of purchased food in diet is independent from the levels of farm commercialization and earnings. Purchased food also continues to play an important role when plenty of own production is available, for example in harvest or post-harvest season of rural Ethiopia, where always at least one third of total calories are purchased (Sibhatu and Qaim, 2017).

Despite this, consumption of own products also contribute importantly to household diets, for example up to 58 percent of yearly calorie consumption in rural Ethiopia (Sibhatu and Qaim, 2017). Contrary to popular belief, subsistence or own consumption is not associated with the most severe nutrition deficiencies. Ogutu et al. (2020) find that more commercialized households do not reduce their own consumption. Rather, the income generated from market activities is used to buy additional food, not to replace own consumption, implying limited subsitution of own consumption with purchased food. From a diet quality perspective, own consumption is particularly valuable for certain types of micronutrients (e.g. Vitamin A). As cited in Sadoulet (1991), Braun et al. (1989) observed that "peasant households continue to grow their own food, and more so as income rises, in spite of increasing involvement in the production of cash crops".

Consumption of home food and of market-purchased food seem to coexist for most farming households, suggesting they fill different roles in househods' diets. Consumption of calorie-rich staples from own production typically provides the basic calorie needs of the households (Von Braun et al., 1991; Sibhatu and Qaim, 2017). In a large study in Sub-Saharan Africa, Frelat et al. (2016) show that own consumption acts as the main source of calorie for food-insufficient households, but also for food-adequate households. On the other hand, purchased food provides more diverse and nutritious products than staple-dominated own consumption (Sibhatu et al., 2015; Sibhatu and Qaim, 2017; Gupta et al., 2020). Specifically, income enables households to purchase micronutrient-rich foods such as fruits vegetables and livestock products and add more diversity to the diet than own consumption (Ogutu et al., 2020).

Evidence for India is narrow and contrasted. Sharma (2006) shows that the diets of agricultural laborers and smallholders are similar, suggesting that own consumption is not relevant, while Gillespie et al. (2012) indicates that households without access to subsistence consumption (landless farmers and non-agricultural households) are more food insecure and spend more on food than agricultural households. Further, Gupta et al. (2020) finds that on-farm production diversity on home gardens and livestock is associated with better diets. Ownership of livestock and crop diversification are linked to better consumption of animal products, dairy products and diet diversity (Bhagowalia et al.; 2012, Headey et al., 2012; Kadiyala et al., 2014).

These observations indicate that households value own consumption, even when increasing incomes offer a better accessibility to market products. Despite this, the fact that own consumption and market-purchased food fill different roles in househods' diets has not been formalized.

2.2.2. Theoretical background

In existing models of farm production, home consumption is explained by three groups of factors relating to production decision and marketization: time allocation of production inputs between various uses (cash crop, home production, off-farm work, leisure); risk in parallel to basic food requirements; and gap between shadow price and market price.

First, in frameworks laid out by Evenson (1978) and Nakajima (1986) own and purchased consumption are considered to be close substitutes, and production decisions depend on time allocation. The household chooses home consumption as long as marginal valuation of home food is higher than the good's market price, where marginal valuation of home food arises from the allocation of family labour time between on and off-farm activities and the associated off-farm wage.

In a second set of literature, it is the presence of risk in both income generation and food prices that explains how risk-adverse families increase subsistence production beyond an optimal point, because retaining production for own consumption appears less risky. Von Braun, de Haen, Blanken (1991) and Von Braun and Kennedy (1994) explain household's allocation of other production resources between subsistence farming and cash crops by the price risks inherent to greater market integration. The expected marginal revenue of specializing further in cash crops is balanced with the extra risk involved with markets, but decreasing marginal productivities prevent households from specializing fully into either subsistence or market

production. A step toward food preferences and minimum subsistence requirement is made by Finkelshtain and Chalfant, (1991), Fafchamps (1992), and Von Braun (1995) who add that risk can also arise from the market-purchased food and that greater marketization result in greater exposure to this consumption risk. The risk becomes especially critical in the context of households' need to procure minimum amounts of food to ensure their survival. Subsistence is thus considered an insurance against hunger in the presence of risks such as higher prices at a different season, fluctuation or poor quality of supply on the market, and taste.

Third, the persistence of own-consumption can also be explained by the price band formed between the sale and the purchase price of food due to market imperfections such as transaction costs, mark-ups but also risk and uncertainties. Janvry et al. (1991) explain that when the shadow price for food falls within the price band, the household does not use the market and instead own-consume the food. When the shadow price is below the sale price, the household sells a share of its product and can consume the rest. The arbitrage between sale and own consumption depends on the match of the sale price and shadow price. As the household increases sale and reduces own consumption, the shadow price of own food (its marginal utility) rises, thus the household will continue to sell the commodity up to the level where the shadow price equates the sale price. The shadow price being different for each household, the decision to own-consume or to use the market is heterogeneous across households. Arslan and Taylor (2009) and Embaye et al. (2018) also point that the shadow price of home-produced food can be higher than its market value because of transaction costs but also cultural attributes attached to home production. This rationally explains why some farmers continue to engage in subsistence farming at levels higher than deemed efficient. This also implies that perfect substitutes for home foods may not exist on the market. Imperfect information about food safety is another reason why farmers may place a higher value on home-produced foods than on market-purchased food (Hoffmann and Gatobu, 2014).

All in all, these approaches imply that satisfaction of basic food need is a priority, which modifies household production and consumption choices, especially in presence of risky environment or markets failures. Households are ready to give up potential income benefits from selling crops in exchange for security in staple food supply. However, this minimum food requirement is not modelled explicitly in the aforementioned approaches and the interaction between these food choice and income are not well-known.

2.2.3. Relationship with income

In general, households modify their calories and staple consumption as income grows and basic needs are gradually satisfied. In particular, there are some evidence of nonlinearity between staple demand, calorie or other nutrition indicators with income (Behrman et al., 1988; A. Ramezani, 1995; Deaton and Subramanian, 1996; Gibson and Rozelle, 2002; Tian and Yu, 2015). The subsistence requirement has been specifically shown to affect nutrient intake by Behrman et al. (1988), who set that preferences for food variety are constrained at low income levels, in favor of more uniform low-cost calories sources. It follows that households typically opt for more expensive foods with lower calorie content when income grows, so that calories increase more slowly than income. Consequently, better economic status is associated with declining calorie share and staple food shares in diets, as exampled in the recent Indian context (Bhagowalia et al., 2012; Kaicker and Gaiha, 2013). However, the strength of the pattern may not be uniform across regions and across income quintiles and the kink might occur only at higher strata of income in highly nutrition-deficient context, with lower income households actually increasing the share of staple in their diet when extra income becomes available (Von Braun and Kennedy, 1994). More precisely, Jensen and Miller (2008) show that an inverse U-shape curve can exist for the consumption of staple food in nutrition-deficient contexts. In response to a price decrease for the staple good, the better-off households decrease their demand, but lower-income households increase it to satisfy their caloric needs, thereby exhibiting a Giffen behavior.

It remains open how households' arbitrage between home and market foods changes with income. As home food consumption is often made prominently of staple food, similar nonlinear relationship between home-produced food consumption and income can be expected to occur. In the high-subsistence context of Malawi (Von Braun and Kennedy, 1994), households with the highest degree of subsistence are in the middle income range, they are neither the poorest, nor the richest. Lower income households, especially those with small lands, are more dependent on grain purchases. The authors put forward that the inability to produce sufficiently to satisfy all nutrition requirements leaves households with a more pressing need to generate income, and thus to sell crop, meaning that less product is available for own consumption (Von Braun and Kennedy, 1994; Rais et al., 2009). On the other hand, subsistence consumption is preferred over keeping cash for future purchases of food by better-off households, as an insurance mechanism against risks, thereby increasing the share of subsistence consumption in diet as income rises.

In the next sections, we propose to formalize the arbitrage between home food and market-purchased food consumption along the income range. We explicitly integrate both a subsistence constraint for staples (Jensen and Miller, 2010) and a shadow price approach (Janvry et al., 1991) to show how market imperfections and subsistence constraint interact and affect the choice of home vs. market food. Then we extend the model to include a minimum requirement of nonfood commodities (cash constraint) and show how households modulate their food consumption choices according to these multiform survival requirements.

2.3. A model of food consumption arbitrage

The model is built in four sections that integrate the constraints step-wise. The first section sets the basic model and examines the effect of market imperfections on the arbitrage between home food consumption and market-purchased food consumption. The second part adds a food subsistence constraint to the model and shows how market imperfections interact with the food constraint, thereby influencing the food choice arbitrage. In the third part, the model temporarily zooms out to a more general form which includes consumption of nonfood commodities in order to include the cash constraint. The effect of both food and nonfood constraints can then be observed on the arbitrage between food and nonfood consumption, and in the fourth part, on the arbitrage between home food and market-purchased food consumption. The effect of rising income and relaxing the constraints are examined. A fifth section offers a mathematical illustration.

2.3.1. Equilibrium conditions under imperfect market conditions

The starting point is a farm household model inspired from Nakajima (1986) and Sadoulet and de Janvry (1995) that combines the activities of producing, working and consuming. The rural household maximizes a utility function subject to an endogenous budget constraint whereby the total farm production provides both monetary income to purchase commodities from market and in-kind income (food) to be consumed directly by the household.

The utility function is as follow:

$$u = u(x_1, x_2, Y)$$

Subject to:	$q_1 = q(L_F)$	The production function
	$P_{x_2}x_2 + P_yY = P_{q_1}(q_1 - x_1) + P_w(L_O)$	The cash constraint
	$T = L_F + L_O$	The time constraint

The latter two constraints can be merged into:

$$P_{x_2}x_2 + P_yY = P_{q_1}(q_1 - x_1) + P_w(T - L_F) \quad \textbf{(1)}$$

where:

x_1: home consumption of food, and x_2 the market-purchased consumption of food, purchased at price P_{x_2}.

Y: nonfood market purchased commodities, bought at price P_y.

T: the total time, divided between work on own farm L_F , off-farm work L_o, valued at the opportunity cost P_w.

q_1: farm production of food products, sold to the market at farm gate price P_{q_1}.

The utility function is assumed to be homothetic which implies that $\frac{U_{x_1}}{U_{x_2}}$ is constant along rays from the origin (Jensen and Miller, 2010) and declines as $\frac{x_1}{x_2}$ grows. The marginal rate of substitution between two goods depends on their ratio $\frac{x_1}{x_2}$. The first order condition (FOC) produces the marginal rate of substitution $\frac{U_{x_1}{}^*}{U_{x_2}{}^*} = \frac{p_{q_1}}{p_{x_2}}$ [6], where the marginal utility attached to the consumption of home foods represents its shadow price, the internal price that the household would be ready to pay to consume this commodity (Janvry et al., 1991). On the other hand of the equation, p_{q_1} represents the opportunity cost of home food consumption to the household, in terms of foregone revenue from crop sale. By virtue of decreasing marginal utilities, a household will continue to increase consumption of good 1 relative to good 2 until its shadow price equates its opportunity cost, relative to good 2.

[6] The sign * represents market imperfections.

In well-functioning markets, $p_{q_1} = p_{x_1}$, where p_{x_1} is the cost of purchasing the commodity x_1 from the market and $\frac{U_{x_1}}{U_{x_2}} = \frac{p_{x_1}}{p_{x_2}}$ at equilibrium. However, market imperfections such as transaction costs, mark-ups by intermediaries, risk, searching costs, and others, result in a price band such that $p_{q_1} = \delta p_{x_1} < p_{x_1}$ with $0 < \delta < 1$. Then equilibrium conditions are fulfilled when:

$$\frac{U_{x_1}{}^*}{U_{x_2}{}^*} = \frac{p_{q_1}}{p_{x_2}} \; or \; \frac{U_{x_1}{}^*}{\delta[U_{x_2}{}^*]} = \frac{p_{x_1}}{p_{x_2}} = \frac{U_{x_1}}{U_{x_2}} \qquad \textbf{(2)}$$

This implies that $\frac{U_{x_1}{}^*}{U_{x_2}{}^*} < \frac{U_{x_1}}{U_{x_2}}$, the optimal marginal rate of substitution of x_1 and x_2 differs under imperfect market conditions such that $\frac{x_1{}^*}{x_2{}^*} > \frac{x_1}{x_2}$. As a result, the presence of market frictions and a price band increase the consumption of home-produced food relative to market-purchased food compare to a perfect market conditions, as illustrated on Figure 2.3 with the change from point A to B. The difference reflects the lower opportunity cost of home products than if the household sourced the same product from the market.

2.3.2. Adding a food constraint

Let's consider that food consumption is subject to a subsistence requirement $\bar{X}$, a minimum amount of consumption below which life is possible but marked by hunger. The subsistence level depends primarily on consuming a sufficient amount of calories. Other attributes of food such as taste, diversity or nutrient content contribute to the utility derived from the consumption of food, but they do not contribute to satisfying the minimum subsistence requirement. As food products can be sourced from either home or market, the subsistence level can be satisfied either way and is expressed by $\bar{X} = c_1 . x_1 + c_2 . x_2$ where c_1 and c_2 are the average calorie content of home-produced food and market-purchased food respectively. The subsistence constraint is binding when the level of food consumption a household can attain given his or her income is not sufficient to meet the subsistence level $\bar{X}$. Similar to Jensen and Miller (2010), the model allows households to consume below their subsistence level, but the gap in food consumption to the subsistence level represents a penalty to utility.

To integrate this constraint, the household sets consumption levels $\widetilde{x_1}$ and $\widetilde{x_2}$ that maximize its utility while also minimizing a penalty for the gap between the subsistence level $\bar{X}$ and the household caloric intake (Jensen and Miller, 2010)[7]. The household is now concerned with:

$$Max\ \ V\ (x_1, x_2, Y) = U(x_1, x_2, Y) - f(\bar{X} - c_1.\ x_1 - c_2.\ x_2) \tag{3}$$

s.t. the same constraints as above.

With $f(.)$ defined as in Figure 2.1, with $f'(.)\ > 0\ and\ \frac{\partial f(.)}{\partial\ x_i} = -\ c_i f'(.) < 0$.

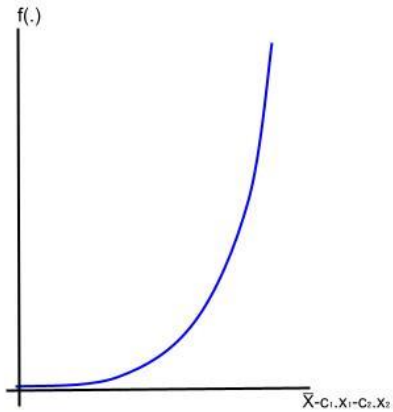

Figure 2. 1: Subsistence penalty

The penalty increases with the gap in calories in the household's consumption, but an increase in consumption of x_1 or x_2 reduces the penalty. The properties $f(0) = 0$ and $f'(.) = 0$ for $\bar{X} - c_1.\ x_1 +\ c_2.\ x_2 \leq 0$ guarantee that the penalty disappears when the calories requirement is fulfilled, in which case the household is only concerned with maximizing its utility. On the other hand, if the calorie gap increases, the household's problem is dominated by reducing the penalty (Jensen and Miller, 2010).

At equilibrium, the FOC of the penalized utility function gives:

$$V_i = \frac{\partial V\ (x_i)}{\partial\ x_i} = \frac{\partial U\ (x_i)}{\partial\ x_i} + \frac{\partial f(.)}{\partial\ x_i} = \ \theta\ p_i \tag{4}$$

where θ is the Lagrangian operator.

[7] The sign $\widetilde{\ }$ represents situation under constraints.

2.3.2.1. Perfect markets

Under the assumption of well-functioning markets where $p_{q_1} = p_{x_1}$, we have

$$\frac{V_1}{V_2} = \frac{\widetilde{U_{x_1}} + c_1 f'(.)}{\widetilde{U_{x_2}} + c_2 f'(.)} = \frac{p_{x_1}}{p_{x_2}} \quad \textbf{(5)}$$

or

$$\frac{\widetilde{U_{x_1}}}{p_{x_1}} = \frac{\widetilde{U_{x_2}}}{p_{x_2}} - f'(.)\left[\frac{c_1}{p_{x_1}} - \frac{c_2}{p_{x_2}}\right] \quad \textbf{(6)}$$

Taking $\left[\frac{c_1}{p_{x_1}} - \frac{c_2}{p_{x_2}}\right] = a$, the difference in calorie content per unit of price between own and purchased food, and rearranging, we obtain at equilibrium:

$$\frac{V_1}{V_2} = \frac{\widetilde{U_{x_1}}}{\widetilde{U_{x_2}} - p_{x_2} a f'(.)} = \frac{p_{x_1}}{p_{x_2}} \quad \textbf{(7)}$$

Thus, the term $-p_{x_2} a f'(.)$ represents the subsistence effect. Given that $\frac{U_{x_1}}{U_{x_2}} = \frac{p_{x_1}}{p_{x_2}}$ represents the equilibrium condition without a subsistence constraint, we can compare $\frac{\widetilde{U_{x_1}}}{\widetilde{U_{x_2}}}$ with $\frac{U_{x_1}}{U_{x_2}}$ to see the effect of subsistence on the home vs. market-purchased food ratio.

In case $a = 0$, the household produces a mix of crops with a similar calorie content to what can be found on the market, and the subsistence requirement has no effect on the equilibrium. More realistically, a farming household that produces mostly staples food would face $\frac{c_1}{p_{x_1}} > \frac{c_2}{p_{x_2}}$, that is the home products contain more calories per unit than market-purchased food. In this case $a > 0$ and $\frac{\widetilde{U_{x_1}}}{\widetilde{U_{x_2}}} < \frac{U_{x_1}}{U_{x_2}}$.

In short, under well-functioning markets, the subsistence constraint increases the share of home consumption in the diet if the household produces staple crop (represented by point A' on Figure 2.2), but would reduce the demand of home products if the household produced cash crop (represented by point A'' on Figure 2.2). The precise effect depends on a, the calorie-content difference between home products and market-purchased food, such that:

$$\frac{\widetilde{x_1}}{\widetilde{x_2}}[cash\ crop] < \frac{x_1}{x_2} < \frac{\widetilde{x_1}}{\widetilde{x_2}}[staple\ crop].$$

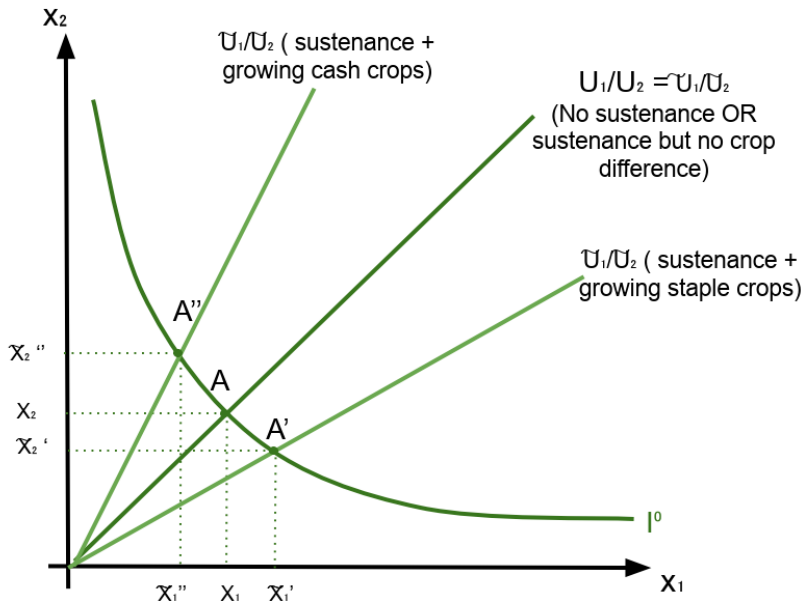

Figure 2. 2: Subsistence effect under perfect market

2.3.2.2. Imperfect markets

Market imperfections can either reinforce or mitigate the subsistence and crop choice effects described above. The calorie-content difference now depends on the price band, as well as on crop choice. Remember that $p_{q_1} = \delta p_{x_1}$ where $0 < \delta < 1$; this implies $\frac{c_1}{p_{q_1}} > \frac{c_1}{p_{x_1}}$. The difference in calorie content arising simply from the lower opportunity cost for home-produced food can be expressed by $\frac{c_1}{p_{q_1}} - \frac{c_1}{p_{x_1}} = \frac{c_1}{p_{x_1}} . \frac{1-\delta}{\delta}$.

Taking equilibrium conditions of a standard case, we have:

$$\frac{V_1}{V_2} = \frac{\widetilde{U_{x_1}}^*}{\widetilde{U_{x_2}}^* - p_{x_2} b f'(.)} = \frac{p_{q_1}}{p_{x_2}} \qquad \textbf{(8)}$$

where $-p_{x_2} b f'(.)$ is the effect of the subsistence constraint on food arbitrage with the calorie content difference $b = \left[\frac{c_1}{p_{q_1}} - \frac{c_2}{p_{x_2}}\right]$. The term b can be decomposed into the extent of the price

band $\frac{c_1}{p_{x_1}}.\frac{1-\delta}{\delta}$, plus the crop difference a. Comparing this to the results obtained from $\frac{U_{x_1}{}^*}{U_{x_2}{}^*} = \frac{p_{q_1}}{p_{x_2}}$ will show the effect of the subsistence penalty under market imperfections.

We can also compare this to the equilibrium conditions in a perfect market, by:

$$\frac{V_1}{V_2} = \frac{\widetilde{U_{x_1}}^*}{\delta\left[\widetilde{U_{x_2}}^* - p_{x_2} b f'(.)\right]} = \frac{p_{x_1}}{p_{x_2}} \qquad \textbf{(9)}$$

This makes explicit that market imperfections affect the ratio of products both through the utility maximizing process, *via* the δ before the denominator, and through the subsistence penalty, via the term $\frac{1-\delta}{\delta}$ present in b.

Here again, the crop choice affects b and thus the ratio of home over market-purchased foods. If the farmer grew a crop mix mirroring exactly that of local markets such that $a = 0$, then $b = \frac{c_1}{p_{x_1}}.\frac{1-\delta}{\delta} + a > 0$. This gives $\frac{\widetilde{U_{x_1}}^*}{\widetilde{U_{x_2}}^*} < \frac{U_{x_1}{}^*}{U_{x_2}{}^*}$ and thus $\frac{\widetilde{x_1}^*}{\widetilde{x_2}^*} > \frac{x_1{}^*}{x_2{}^*}$. Under market imperfections, the subsistence constraint increases the preference for home-produced food, even if the household produces a crop mix similar to what is found on the market. Furthermore, this subsistence concern reinforces the market imperfection effect identified above in increasing the relative share of home-produced food such that $\frac{\widetilde{x_1}^*}{\widetilde{x_2}^*} > \frac{x_1{}^*}{x_2{}^*} > \frac{x_1}{x_2}$.

In the typical case where the household produces staple crops, such that $a > 0$ and $b > a > 0$. The same conclusion as above apply, but producing staple crops reinforces again the preference for home-produced food, on top of the effect of market imperfection. On Figure 2.3 this is illustrated by the move from B to C, which is valid only for farmers growing food with a calorie content at least as high as that found on the market (i.e. dominated by staple food).

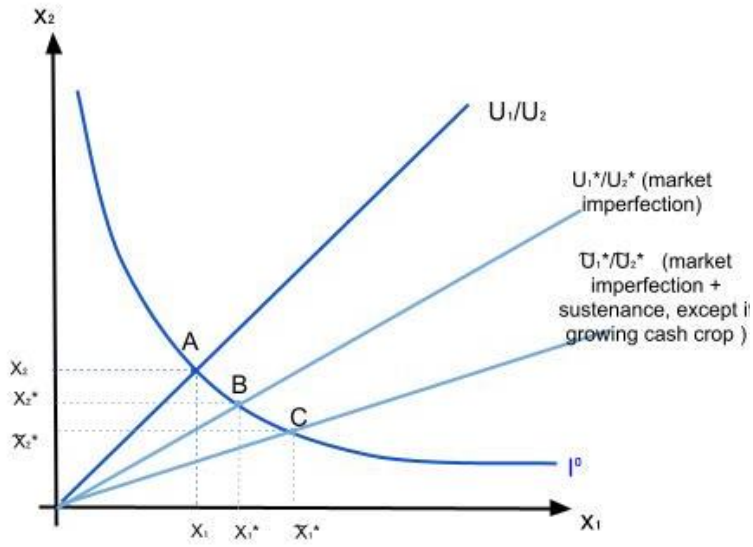

Figure 2. 3: Market imperfections and subsistence constraint

In summary, two factors increase the preference for home-produced food are observed so far: market imperfections and the subsistence constraint, where the subsistence constraint depends itself on the strength of the market imperfections as well as on the calorie-content difference between home-produced crops and market-available foods. In a case where markets are imperfect and the subsistence constraint is binding, and households produce staple food, the preference for home-products is stronger, as each of these elements reinforces the preference for home products. In other words, home products are clearly preferred under market imperfections; but if the household produces staples, this effect is further amplified via the subsistence constraint. If the household produces non-staple food, the subsistence constraint would counteract the imperfect market effect and the household may consume either more or less home-produced food, depending on the relative strength of the opposite effects.

For the purpose of unambiguity, the subsequent analysis focuses on staple-producing households and retains the realistic assumption of market imperfections.

2.3.3. Adding a cash constraint

Household do not solely face a food subsistence constraint for their survival. They must also afford basic nonfood commodities, such as housing rentals, energy, school fees, etc. This requirement can be modelled by a cash penalty, analogue to the food penalty, of the form $g(\bar{Y} - Y)$ where $\bar{Y}$ is the minimum amount of nonfood expenses a household must afford. The following steps show how the satisfaction of these basic needs affect the ratio x_1/x_2.

In order to model this cash constraint, we adopt a 2-stage budgeting approach (Gorman, 1959; Strotz, 1957) whereby the household first determines consumption between food and nonfood and later on allocate its food budget between home food and market purchases. For the purpose of modelling the 1st stage, food expenditure is aggregated into $X = x_1 + x_2$, and characterized by a price $p_X = h(P_{q_1}, P_{x_2})$.

The utility maximization problem now includes both a subsistence penalty and a cash penalty, which reflect the need to ensure minimum amounts of food and nonfood consumption.

$$Max\ V\ (X,Y) = u(X,Y) - f(\bar{X} - c_x X) - g(\bar{Y} - Y) \qquad \textbf{(10)}$$

s.t. the same constraints as above.

The Lagrangian and first order conditions give that:

$$\frac{V_X}{V_Y} = \frac{\widetilde{U_X} + c_X f'(.)}{\widetilde{U_Y} + g'(.)} = \frac{p_X}{p_Y} \qquad \textbf{(11)}$$

The relative consumption of food and nonfood products thus depends on the relative strength of each constraint. If the subsistence constraint is stronger, $\frac{\widetilde{U_X}}{\widetilde{U_Y}} > \frac{U_X}{U_Y}$ and $\frac{\tilde{X}}{\tilde{Y}} < \frac{X}{Y}$ so the consumption on nonfood commodities increases relative to food products, compared to when no constraint exists. The inverse is observed if the subsistence constraint is binding but not the nonfood constraint. So the cash constraint and food constraint have opposite effects on the ratio of food over nonfood consumption, as illustrated on Figure 2.4, with a move from A, the point where both constraints are either null or have equal weights, to the points A' (if the food constraint is stronger) or A'' (if the cash constraint is stronger). In other words, as the cash constraint becomes stronger, it crowds out food expenditure and replaces it with nonfood expenditure.

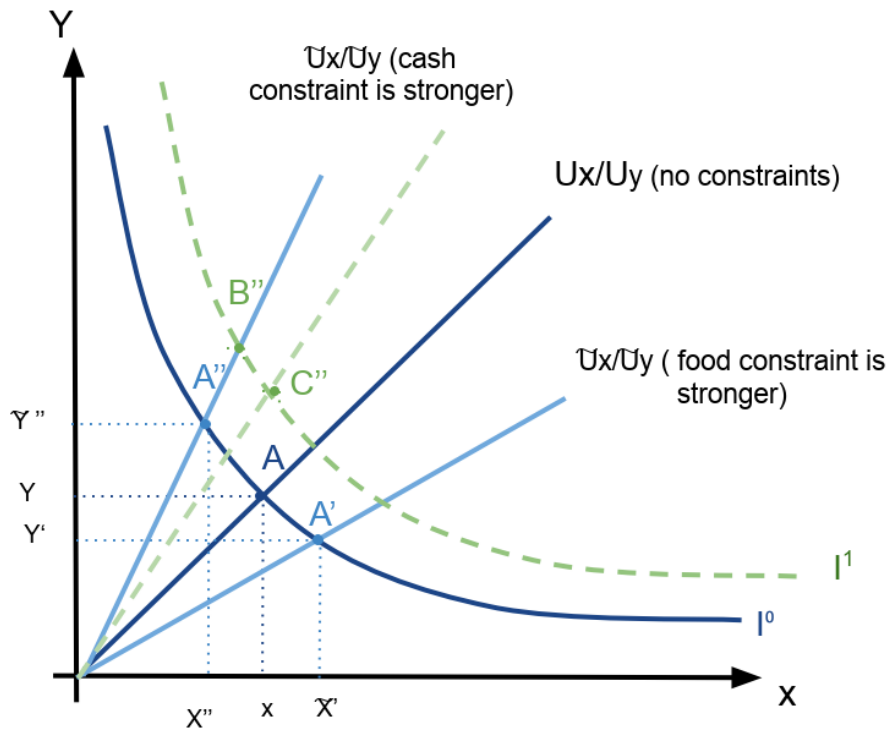

Figure 2. 4: Stage I - MRS between food and nonfood commodities

This has implication for the food sourcing choice of stage II (Figure 2.5). The presence of the cash penalty increases preference for Y over X. At a given income level, strong cash penalty would lower food consumption in order to satisfy the requirement for Y and bring it to the level I^{-1} (blue dashed lines). The reduction in X affects both x_1 and x_2 (point A' to B' on Figure 2.5), thereby increasing the gap in calories to the subsistence level. As the food penalty becomes stronger, the preference for x_1 over x_2 is reinforced (point B' to C'). In short, a cash constraint has an income effect, which reduces x_1 and x_2, and a penalty effect, which increases x_1/x_2. Therefore, the cash constraint amplifies the effect of the subsistence constraint on x_1/x_2, by reducing the budget share allocated to food.

- *Proposition 1: both subsistence and cash penalty increase preference for x_1 over x_2.*

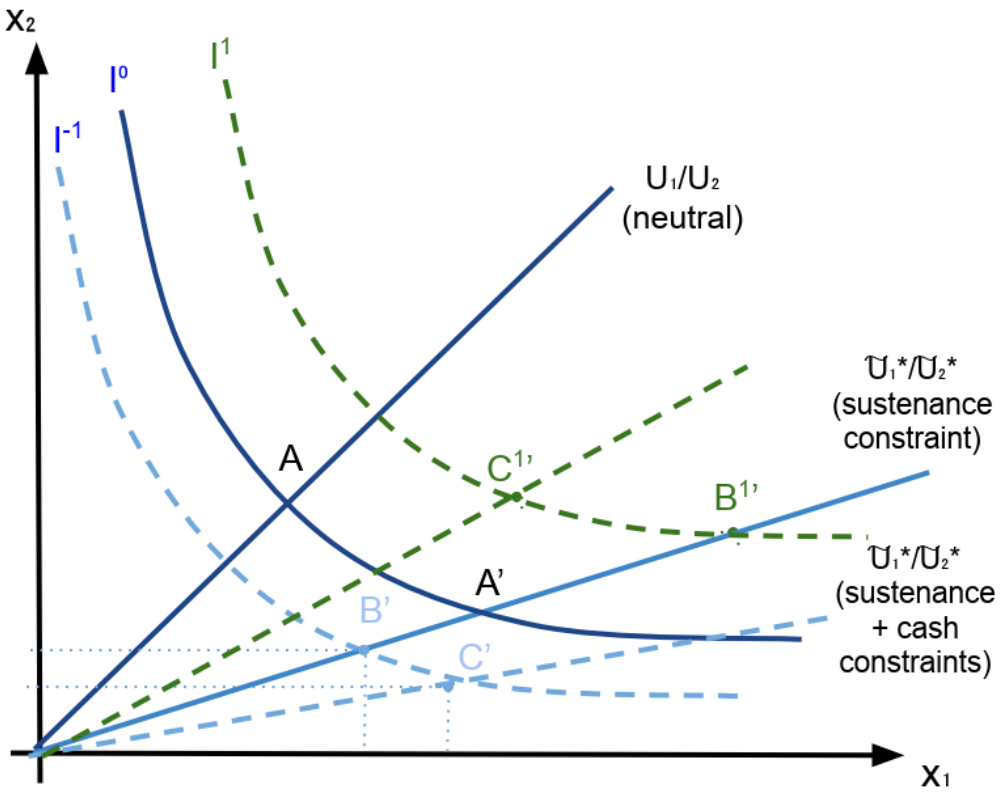

Figure 2. 5: Stage II – MRS between home and. market food, under cash penalty

2.3.4. Effect of increasing income

We now turn to observe how these arbitrages change as income increase. At the first stage (green dashed lines of Figure 2.4), a higher income would allow the household to afford bundles of consumption satisfying the indifference curve I^1. If the cash constraint is stronger, the household will move from point A'' to point B''. However, an increase in consumption of Y will lower the cash penalty, and bring the MRS closer to the equilibrium level Ux/Uy, as shown by the move to C''. The consumption of both food and nonfood goods increase, but the overall effect depends on income effect (which keeps X/Y stable), and the penalty effect, which increases X/Y compare to A''. In short, preferences change as income increases *via* the reduction in penalties. Thus, the marginal effect of rising income includes both an income effect and a penalty effect.

At the second stage, households allocate an extra unit of food budget between home food and market food. Assuming the household remains constrained by the subsistence penalty, it would increase consumption from bundle A' to bundle B^{1}' (green lines in Figure 2.5), so that both x_1

and x_2 rise. However, the reduction in the food penalty brings the MRS closer to its initial level, for example C[1]', thereby creating equilibrium conditions at a lower ratio x_1/x_2 than at A'.

- *Proposition 2: as income increases x_1/x_2 lowers, i.e. the consumption of home food shrinks relative to market purchased foods.*

It can be noted that if the income gain comes from non-farm work and the household is already experiencing $q_1 = x_1$(it consumes all it produces), it is not possible to increase x_1 in value. Then x_2 will increase even more than described above, and the ratio x_1/x_2falls even faster. On the other hand, if the income gain comes from greater q, then it is possible to increase x_1 as desired. The extra q will be split between x_1 and x_2 (and Y) depending in the prevalent MRS.

The above observation has implication for the share of home product consumption in the total food consumption, or put mathematically $p_{q_1}x_1/(p_{q_1}x_1 + p_{x_2}x_2)$. We know that the food and cash constraints bring x_1/x_2 higher than its equilibrium level, suggesting a preference for x_1. Thus, at low levels of income, when the penalty is high, the high x_1/x_2 will cause an extra expenditure on x_1 to increase in value faster than expenditure on x_2. This causes the share of home food consumption in the total food expenditure to increase (purely under the income effect). But as income and food consumption increase, the reduction in penalties gradually brings x_1/x_2 back its equilibrium level. Home food expenditure grows at a decreasing rate while food expenditure on market-purchased food will start to grow faster in value (reduction of the penalty effect). When this effect becomes strong enough, home food consumption's share in total food expenditure decreases. The resulting curve of $p_{q_1}x_1/(p_{q_1}x_1 + p_{x_2}x_2)$ will then be increasing with income at first, and then decreasing.

- *Proposition 3: the share of home food consumption among total food consumption follows an inverse-U shape as income rises.*

All in all, the choice between home food and market-purchased is affected by income levels and mostly by the interplay of the cash and food subsistence constraints, and possibly by the input availability and agricultural production capacity. In the following model, we show that all these variables enter the demand equations and result in a nonlinear relationship between the share of home food consumption and income.

2.3.5. Full model illustration

We assume a Stone-Geary Utility function to illustrate further how the food and cash constraints modify the consumer maximization problem.

$$\text{Max } V(x_1, x_2, Y) = (x_1 - c_1)^{\alpha_1}(x_2 - c_2)^{\alpha_2} \quad (Y - \overline{Y})^{\beta} \tag{12}$$

$$\text{s.t. } P_{x_1}x_1 + P_{x_2}x_2 + P_y Y = M \tag{13}$$

With the FOC, we can solve this function as:

$$P_{x_1}x_1 = \alpha_1(M - P_{x_1}c_1 - P_{x_2}c_2 - P_Y\overline{Y}) + P_{x_1}c_1 \tag{14a}$$

$$P_{x_2}x_2 = \alpha_2(M - P_{x_1}c_1 - P_{x_2}c_2 - P_Y\overline{Y}) + P_{x_2}c_2 \tag{14b}$$

$$P_Y Y = \beta(M - P_{x_1}c_1 - P_{x_2}c_2 - P_Y\overline{Y}) + P_Y\overline{Y} \tag{14c}$$

With (14a) and (14b), we then can give the following equation for own consumption share:

$$R = \frac{P_{x_1}x_1}{P_{x_1}x_1 + P_{x_2}x_2} = \frac{\alpha_1(M - P_{x_1}c_1 - P_{x_2}c_2 - P_Y\overline{Y}) + P_{x_1}c_1}{(\alpha_1 + \alpha_2)(M - P_{x_1}c_1 - P_{x_2}c_2 - P_Y\overline{Y}) + P_{x_1}c_1 + P_{x_2}c_2} \tag{15}$$

Differentiating (15) with M, we have,

$$\frac{\partial R}{\partial M} = \frac{(\alpha_1 P_{x_2}c_2 - \alpha_2 P_{x_1}c_1)}{[(\alpha_1 + \alpha_2)(M - P_{x_1}c_1 - P_{x_2}c_2 - P_Y\overline{Y}) + P_{x_1}c_1 + P_{x_2}c_2]^2} \tag{16}$$

Equation (16) shows that the effect of income M on own consumption share R depends on the sign of $(\alpha_1 P_{x_2}c_2 - \alpha_2 P_{x_1}c_1)$

In the first case, if $\alpha_1 P_{x_2}c_2 - \alpha_2 P_{x_1}c_1 > 0$, income growth could lead to higher own consumption share. We link this to the behavior of very poor households, with limited agricultural production capacities. Indeed, in such situation food is mostly drawn from purchased food because of households' limited resources to grow their own foods. In addition, some food cannot be purchased and are not typically produced by households (e.g. seasonings, meat products, dairy products, or other processed food). The cash burden of having to pay for the incompressible purchased food is such that the subsistent level of purchased food (first term of the equation) is relatively high compare to the

households' ability to produce food for own consumption (second term of the equation). Furthermore, crops grown are sold to the market to generate the cash needed to purchase the food.

For these households, an extra amount of income (either monetary of in-kind resources) translates into a less pressing need to generate cash in order to purchase food, which therefore allows to increase the amount of own consumption.

The increase of own consumption (second term of equation) reaches a level at which it surpasses subsistent purchased food (first term of the equation). This brings the household to the second stage where $\alpha_1 P_{x_2} c_2 - \alpha_2 P_{x_1} c_1 < 0$, where income growth could lead to lower own consumption share. This can be interpreted as households having reached a point where subsistent own consumption having reached a desired maximum and households prefer to turn to more purchased food when additional income is available. Here there is no cash pressure to acquire purchasable food, because sufficient income is available to these farmers.

We can have the following simulation (Figure 2.6), where subsistent own food consumption is set to 100, subsistent purchased food consumption is set to 500+(income*0.005)^2, where the squaring reflects that incompressible cash expenditure tend to increase with income level, and subsistent other consumption is set at 300. The parameters are set as $\alpha_1 = 0.1$, $\alpha_2 = 0.3$ and $\beta = 0.6$.

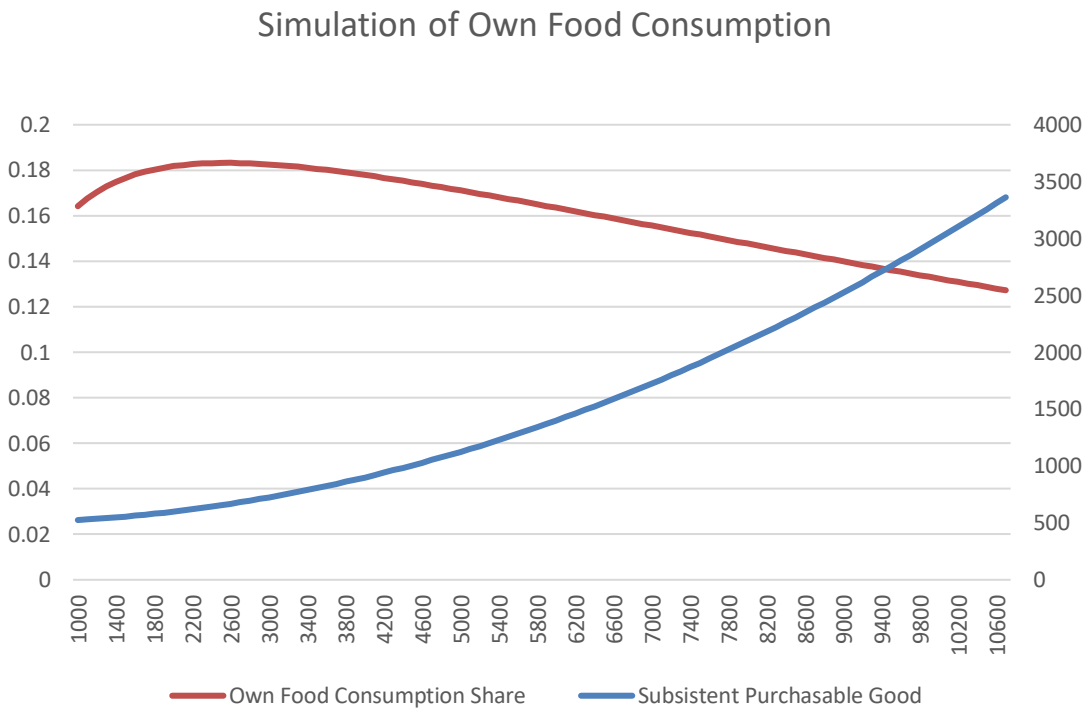

Figure 2. 6: Simulation of own food and purchased food consumption

The simulation confirms the proposition of a nonlinear relationship between the share of own food consumption in food expenditures and income. In the next section, an illustration with data from India is used to test the hypothesis as well.

2.4. Empirical evidence

2.4.1. Income and own consumption patterns: summary statistics

Data for rural India come from the 68th round of the National Sample Surveys (NSS)[8] on Household Consumer Expenditure, carried out in 2011-2012. Close to 60 percent of rural households cultivate some land and include some of their own products in their diet. About 46 percent of this sample is made of households who rely on agriculture as a main source of income (subsequently "farmers"), while the rest (self-employed in other areas, regular salary earners, or casual workers) cultivate land as a secondary activity. The farmers possess larger cultivated areas, and more of them use irrigation (Table 2.1). Monthly income per capita is clearly lowest for casual workers and highest on average for households receiving a regular salary as a main

[8] Household Consumer Expenditure: NSS 68th Round, NSSO, Ministry of Statistics & Programme Implementation, Government of India, http://microdata.gov.in/nada43/index.php/catalog/1

source of income. Nutrition variables[9] show that farmers and regular salary earners consume higher amounts of calorie per day on average. Casual workers on the other hand have the smallest average calorie consumption and have the largest proportion of households that consume fewer than 2700 calories per day (considered by some as the minimum requirement, see Reddy, 2010). In terms of diet quality (shares of protein and fat in calories), no large difference is observed across groups.

With respect to food consumption, all households spent on average around 40 percent of their total expenses on food. Finally, the share of food coming from own production is highest for farmers, averaging around 40 percent. For this group, observations are almost normally distributed around the mean of (orange area of Figure 2.7). For other groups, most households have small shares of own consumption, under 20 percent, and a few households have large shares. Therefore, own consumption is a non-negligible part of food consumption, mostly for farmers, but also for other types of households.

[9] Calculated based on food recall for the last 7 or 30 days, adjusted for meals taken outside and meals served to guests; average per consumer unit, accounting for sex and age of the household members, and days away from the household.

Table 2. 1: Summary statistics for various household types

	Main source of Income				
	Self-employed in agriculture (farmers)	**Self-employed in non-agriculture**	**Regular salary earner**	**Casual labour in agriculture**	**Casual labour in non-agriculture**
Number of household observations	*15961*	*7123*	*5369*	*1662*	*3622*
Income and endowment					
Monthly income per capita	1842	1793	2274	1251	1341
Presence of regular salary earner in the household	4.8%	7.1%	99.6%	1.4%	3.0%
Size of cultivated land, ha	2,027	0,698	0,818	0,446	0,38
Irrigate land, % households	67%	49.9%	47.4%	50.8%	45.9%
Food consumption					
Total food expenditures	779	744	889	536	561
Food in total expenditures	42.3%	41.5%	39.1%	42.9%	41.9%
Own consumption, % of food	39%	23%	25%	17%	22%
Nutrition Status					
Calories / day / individual	3046	2906	3023	2800	2850
Households below 2700 cal/day	39.2%	44.5%	38.7%	51.3%	46.3%
Protein in calories	2.7%	2.7%	2.7%	2.6%	2.7%
Fat in calories	2.1%	1.9%	2.1%	1.8%	1.9%

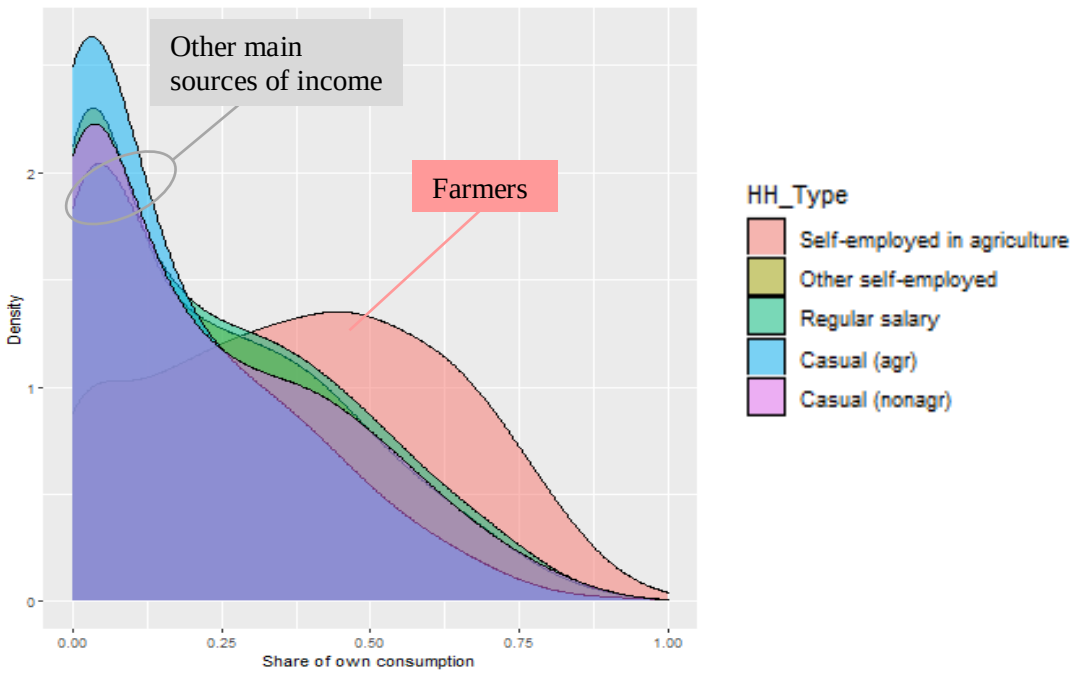

Figure 2. 7: Density of own consumption in food expenditures, per household type

2.4.2. Semiparametric analysis

Tian and Yu (2015) explain that parametric models cannot capture the non-linear relationship between nutrition and income, while fully nonparametric model can be difficult to interpret and their high dimensionality leads to large variance of the estimates. Alternatively, semiparametric methods can be used to model relationships when the function between the output and one independent variable is unknown. We adopt a spline smoothing method (Wand et al., 2005), which is based on piecewise polynomial regression whereby a polynomial is fit at each piece delimited by knots. The model is specified to assign a free form function (non-parametric) for the relationship with income and a linear (parametric) form to all other determinants. In the following, the sample focuses on rural households that consume positive amounts of home-produced food (e.g. pure cash crop growers who would consume none of their own production are excluded). Income and food consumption are measured per capita, and food consumption per capita is adjusted for demographic of household composition: age, sex and days away for each member, as well as for meals taken outside the home and meals served to guests.

To begin with, three univariate regressions are used to examine the relationship between income and a) the value of home food consumption, b) the share of home food consumption in total food consumption (in value) and c) the ratio of home food consumption over the market-purchased food consumption (in value). The resulting nonlinear relationships, shown in Figure 2.8.a to 2.8.c, present wide confidence intervals at both ends of the income distribution. The large intervals for high-income households likely reflect the scarcity of data at this income level, while the large intervals at very low income levels can be interpreted as revealing the heterogeneity of behavior for those households. With tight confidence intervals at the middle income level, the results support the general predictions that:

a) home food consumption increases in absolute value as income increases (up to some point, see discussion),

b) the share of home food in the total food expenses rises and then decreases according to the prevalence, and then release, of the food constraint,

c) the ratio of home food over market-purchase is higher (around 1) at low income levels but gradually decreases as income grows.

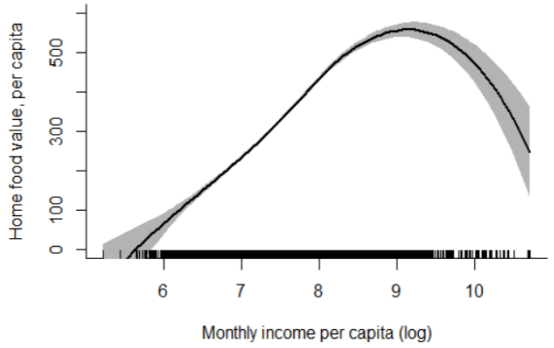

Figure 2.8.a: Value of home food consumption

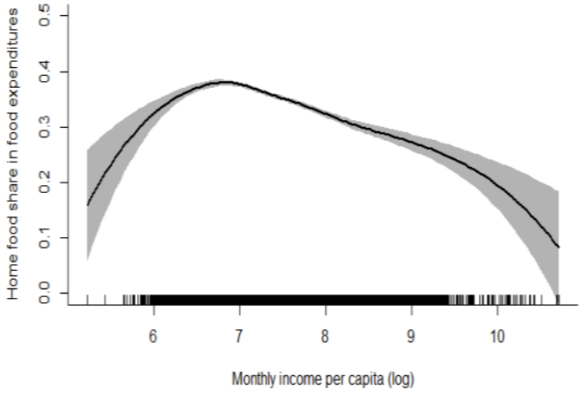

Figure 2.8.b: Share of home food in diet

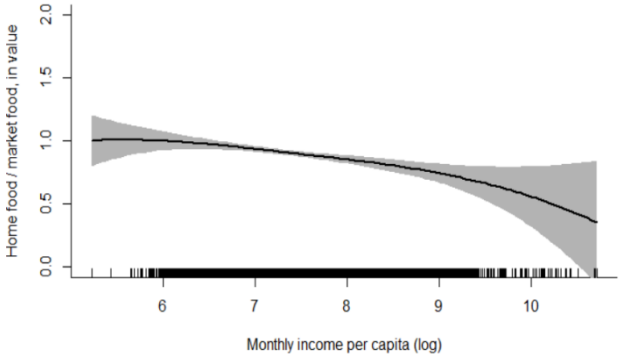

Figure 2.8.c: Ratio of home food over market purchased food

Figure 2. 8: Univariate nonlinear regression, whole sample

Next, results are presented separately for households whose main source of revenue is farming (14633 observations) and other households, who draw income mostly from other, precarious sources of income (casual workers or self-employed in non-agricultural sector).[10] This second group is subsequently referred to as non-farming households and represent 10295 observations. The distinction allows highlighting potential heterogeneity that could arise from access to farming production capacity. We also turn to a multivariate analysis to control for explanatory variables and to include a control function for the endogeneity of income. The resulting curves, shown in Figure 2.9 for farming households and Figure 2.10 for non-farming households, do indeed show some difference for outcomes b) and c). While the inverse U pattern for the share of home food is observable for non-farming households, starting at around 0.2 (Figure 2.10.b), the group of farmers exhibit mostly a decreasing trend, but which starts at higher value (around 0.45) of home consumption in total food expenses (Figure 2.9.b). This is reflected in the ratio of home over market purchased food, which clearly decreases for farmers, while a light inverse U-shape can be seen for non-farming households. The absolute value of home food consumption (figures a) has a similar pattern for both groups of households, but at clearly higher levels for farming households.

[10] Households whose main income comes from a regular salary earner are excluded, as they are typically better-off and less concerned by subsistence concern.

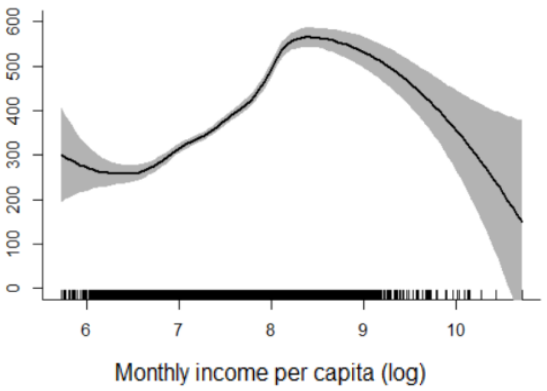

Figure 2.9.a: Value of home food consumption

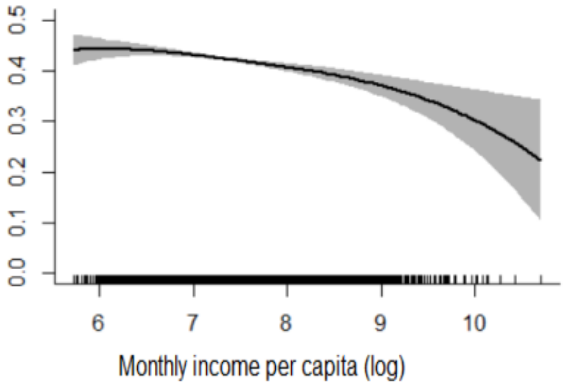

Figure 2.9.b: Share of home food in diet

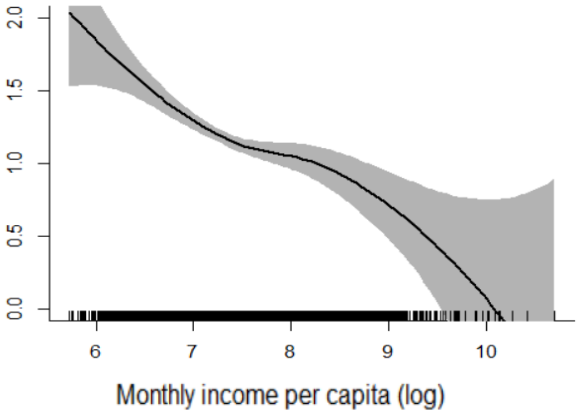

Figure 2.9.c: Ratio of home food over market purchased food

Figure 2. 9: Multivariate nonlinear regression, farming households

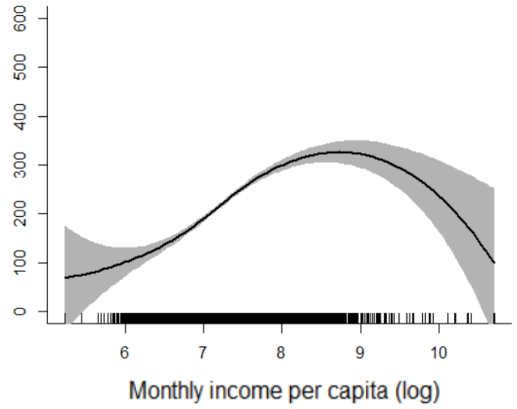

Figure 2.10.a: Value of home food consumption

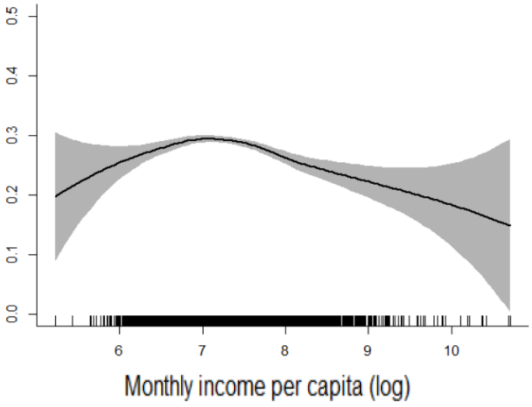

Figure 2.10.b: Share of home food in diet

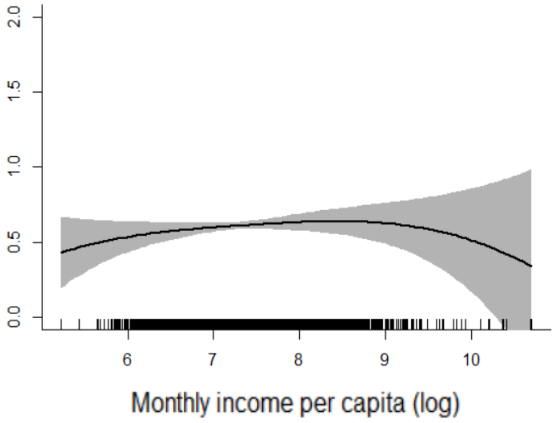

Figure 2.10.c: Ratio of home food over market purchased food

Figure 2. 10: Multivariate nonlinear regression, non-farming households

In other words, farmers do not need to increase home food consumption, likely because their access to home food products is sufficient to satisfy their caloric requirement already at low income levels. On the other hand, non-farming households increase home food consumption faster than market food at the first place when income rises, suggesting that very low-income households are restrained in their capacity to source food from their own production, and prefer to increase home consumption first when more resources become available.

The coefficients and significance for the other variables entering the model, presented in Table 2.2, help to further understand the role of inputs and prices.

Table 2. 2: Semiparametric multivariate analysis, by household type

PANEL A: Farming households

	Value of home food consumption		**Share of home food in total food expenses**		**Home food expenses/ market food expenses**	
n=14545	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*
Intercept	-1476 *	860	-0.179 ***	0.0654	-0.269	1.072
Land size, per capita (log)	32.53 ***	1.96	0.033 ***	0.0015	0.201 ***	0.0156
Regular salary earner	6.76	9.96	-0.004	0.0078	0.024	0.0792
Nb adult workers	29.42 ***	1.43	0.016 ***	0.0011	0.123 ***	0.0114
Nb semi-retired workers	-16.12 *	8.964	-0.039 ***	0.0070	-0.305 ***	0.0713
Price of cereals (log)	-51.38 ***	13.24	-0.102 ***	0.0104	-0.176 *	0.1053
Price of animal (log)	101.5 ***	4.06	0.071 ***	0.0031	0.467 ***	0.0323
Price of vegetables (log)	-34.78 ***	4.46	-0.051 ***	0.0035	-0.326 ***	0.0354
Price of fresh fruits (log)	-44.08 ***	3.62	-0.039 ***	0.0028	-0.235 ***	0.0288
Control function residuals	215.1 ***	11.41	0.012	0.0089	0.517 ***	0.0907

PANEL B: Non-farming households

	Value of home food consumption		**Share of home food in total food expenses**		**Home food expenses/ market food expenses**	
n=10242	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*
Intercept	-1107 ***	147.7	-0.129	0.1634	-1.613 ***	0.4221
Land size, per capita (log)	27.4 ***	1.411	0.033 ***	0.0013	0.103 ***	0.0082
Regular salary earner	26.22 ***	7.989	0.009	0.0075	0.024	0.0466
Nb adult workers	20.81 ***	1.396	0.014 ***	0.0013	0.055 ***	0.0081
Nb semi-retired workers	6.21	8.891	-0.017**	0.0083	-0.097 *	0.0518
Price of cereals (log)	-25.47**	11.71	-0.064 ***	0.0110	-0.257 ***	0.0682
Price of animal (log)	59.57 ***	3.769	0.059 ***	0.0035	0.187 ***	0.022
Price of vegetables (log)	-32.41 ***	3.99	-0.042 ***	0.0037	-0.122 ***	0.0232
Price of fresh fruits (log)	-37.77 ***	3.328	-0.035 ***	0.0031	-0.128 ***	0.0194
Control function residuals	122.7 ***	9.677	0.016 *	0.0091	0.120 **	

The availability of production inputs, specifically the size of cultivated land (reported per capita) and the number of workers (number of household members aged 18 to 65), increase the value, share and ratio of home consumption. The model also accounts for the presence of 'semi-retired' workers (age 55 to 70) relative to the number of workers in the household, to reflect the effect of labor with possibly lower productivity and lower off-farm employment possibilities. It is associated with lower home food consumption, although at a lower significance level for non-farming households. Having a regular salary earner is insignificant on the split sample models, likely due to the low number of households having a regular salary in the selected samples, but it was clearly associated with lower home food consumption on a specification with the full sample (not reported here).

Regarding prices[11], when products which are typically grown on farms (cereals, vegetables and fruits) become more expensive, own consumption decreases (in value, share and ratio), which suggests that farmers take advantage of higher selling prices to sell more and consume less of their own products, or that the extra income from higher price enables them to shift toward more purchases. On the other hand, an increase in the price of milk, milk products and animal protein (eggs, fish and meat), which are less often produced by rural families, leads to an increase in home consumption in all three models and for both sub-samples.

Finally, a control function is used to account for the endogeneity attached with income. Indeed, the decision on home food consumption affects the household's capacity to sell its produce and generate income, and vice versa. This simultaneity between income generation and home consumption creates correlation between the error term and the regressor which leads to biased results. Blundell and Powell (2011) show that for nonparametric and semiparametric estimations, a control function is preferred over the use of two stage least squares to handle endogeneity. Control function is successfully applied to nutrition in Ogutu et al. (2020) or to the study of inverted U-shapes in Aghion et al. (2005). Three variables are selected to instrument income: expenditure on non-durables, except foods (e.g. fuel and light, entertainment, medical, transport, rent …); expenditure on durable goods (clothing, bedding, footwear, education …), and the number of durable goods (furniture, small and large appliances, vehicles, jewelry…) owned by the household. The diagnostics results (Sargan test, Wu-Haussmann test and weak instrument test) confirm the presence of endogeneity and suggest that

[11] Prices are calculated with survey information on quantity consumed and expenditure for each good and each household, then aggregated by food group (using budget shares) at the district level.

the selected instruments are appropriate for the model. The coefficients on the residuals of the control function are mostly significant and positive. Alternative specifications show that the overall shapes of the nonlinear curves are not affected by the inclusion or removal of the control function.

The results confirm, both theoretically and empirically, that it is not the poorest of the rural households who rely the most on home food consumption, but rather households of the middle-income range. This confirms the prediction by von Braun (1995) that higher income would likely result in increased volume of subsistence production.

2.5. Discussion of results

2.5.1. Access to production inputs

The model predicts that households have a disproportionate preference for home consumption as long as their basic food requirements are not satisfied, however they can consume home food only up to the limit of their own production. In the results, larger agricultural production possibilities (size of land, number of workers) are indeed associated with larger own consumption. According to the model, households favor market purchases either when their basic food needs are satisfied and they can afford market food, or when they have reached the maximum amount of food they could source from their cultivating capacities. Thus, consumption of home food of households below the food requirement level can be limited by access to production capacities before it reaches saturation.

Indeed, farmers, who have larger production capacities (larger land and more frequent irrigation), display much higher shares of home food consumption in their diets than their counterparts. For them, home food consumption decreases as income increases (Figures 2.9.b and 2.9.c), as opposed to following the inverse U shape proposed theoretically and observed for non-farming households. This can be interpreted as the fulfillment of the desirable level of home food consumption already at low income levels. For farmers all over the income range, increases in income translate into faster increase in purchased food than home produced food. For them, it is likely that the preference for food diversity and taste obtained from the market is stronger than the need to satisfy basic needs via additional low-cost staples. Indeed, the farmers have

higher calories consumption per person on average and are less often (39 percent) calorie-deprived (under 2700 cal/day) than their non-farmers counterparts (44 percent to 50 percent).[12]

For non-farming households, increasing shares of home food consumption at low income levels indicates their home food consumption is not sufficient to meet their food requirement. It can be argued that access to land is one tool to improve their nutritional situation, by allowing the households to consume more of their own production. Access to land is often advocated as important for food security (de Janvry and Sadoulet, 2011; Rammohan and Pritchard, 2014). However, land fragmentation and landlessness are issues difficult to solve and may not present the most efficient way to suport nutritionally-deprived households.

2.5.2. Implication of the cash constraint for cash vs. in-kind transfers

The cash consumption, as modelled here, suggests that households might chose to sell crops before they meet their basic food need, when need for cash is strong enough to reduce food consumption. Thus, for households with poor nutrition status and strong cash constraint, unconditional cash transfers could alleviate undernutrition by giving households the opportunity to satisfy some of their non-food consumption requirements. Households could free up income or crops and allocate them to food, either from home or market according to their preferences. Thus, decision makers and programs administrators do not need to assess subsistence level of households, which is difficult to judge (Headey and Ecker, 2013; Kaicker and Gaiha, 2013; Tian and Yu, 2015). Cash transfer spent on non-food can thus also indirectly help lift some nutrition deficiencies.

Whether food transfers, in the form of in-kind staples or subsidized grains, will be consumed or sold depends on the strength of the cash and food constraints for each household. Assuming the food constraint is binding (which motivated the policy support), households might still choose to sell the food if their cash constraints is stronger, as additional income is more crucial than added calories. This is especially true if they already have access to a source of affordable staples via home production. On the contrary, the preference for in-kind calorie-rich food support is likely stronger for households with limited access to own consumption.

[12] Still excluding households with regular salary as main income source.

2.5.3. Omitted factors

A few aspects relevant to the topic have been so far excluded from this analysis, among them the prominent role of gender. Indeed, the income gain associated with the shift from subsistence farming to commercialization does not always fully translate in gain for nutrition, because of a gender bias in the intra-household redistribution of income (von Braun and Kennedy, 1994; Carletto et al., 2017; Ogutu et al., 2020 and others). With increased commercialization, income is shown to be more controlled by men, which has adverse effect on nutrition because women orient their nutrition choices toward better quality than men.

Furthermore, some studies have shown that availability of non-farm income plays a special role on the arbitrage between own and purchased consumption, especially to smooth out the seasonal availability of own products (Sibhatu and Qaim, 2017). Own consumption supplies higher shares of calories during harvest and post-harvest times, but is substituted by purchased calories during lean season when own production is less available. The seasonal differences is also observed in Luckett et al. (2015). During the post-harvest (dry) season, purchased food contributes most to diet diversity and own consumption the least.

Next, the model assumes that cultivating households prioritize own consumption of staples to satisfy calories requirements, and turn to market purchases for diversified and micronutrient-rich foods once their income allows it. However, the model also allows for the possibility that households shift to micronutrient-rich own consumption as income increases. For example, the data for this essay shows a fast increase of home-sourced milk products consumption as income increases (not reported here). Consequently, if households adapt to the release of their food constraint by producing diversified micronutrient-rich foods, the predicted shift from home consumption to market purchases is postponed to higher income levels, as home products also provide some of the attributes (diversity, taste, and nutrients) that are desired once the caloric requirement is fulfilled.

Last, reliance on home food consumption is often explained by difficult access to market, both to generate income from selling crop and to purchase a diversified diet. Better access to markets is generally recommended in order to improve nutrition, rather than improving diversity of subsistence farming (Gupta et al., 2020; Koppmair et al., 2017; Sibhatu et al., 2015; Sibhatu and Qaim, 2018). The proposed framework is consistent with prior findings, because the difficulty to access markets materializes in larger market imperfections, increasing reliance on home consumption. In addition, access to market seems to be less of an issue for the Asian case,

which enjoys relatively stronger market linkage, than in the Sub-Saharan context, where market access is lower and subsistence higher (Sibhatu and Qaim, 2018, Gupta et al., 2020). In the present data, the role of markets, as measured by the average estimated travelling time to cities above 50,000 inhabitants, was tested in alternative OLS specifications but it was found to be insignificant to explain shares of own consumption in diet.

2.6. Conclusion

Consumption of home-produced food is generally interpreted as an indication of subsistence farming, which is bound to disappear if farming households can shift to commercialized agriculture and are able to generate sufficient income to purchase quality diets. On the contrary, evidence shows that for some households, home food consumption first increases when income increases, before shrinking. At the same time, purchased food plays a vital role at all income levels, including at the poorest.

The present study analyzes own consumption in the context of rural households' strategies to fulfill their basic food needs, rather than as a result as production choice between subsistence and market-oriented farming. The theoretical framework shows how a binding food constraint orients households' preference toward calorie-rich home-produced foods, and then to purchased foods when basic needs are fulfilled. Additionally, the model considers household's choice in the wider context of all basic needs, by including a cash constraint to cover uncompressible non-food expenses. For poorer households, a rigid cash constraint may crowd out other expenses, including food, which in turn reinforces the gap in calories and therefore the preference for home foods. The model finally integrates that market imperfections amplify the preference for home food when staples are produced, and also form a cause of home consumption.

Empirical analysis with data from rural Indian households for 2012 confirms the presence of an inverse U-shape relationship between the share of home food in diet and income, with a stronger effect for non-farming households. This has implication for the analysis of nutrition and wellbeing outcomes. Rising home food consumption of very nutrition-deficient households can be sign of improvement, while it indicates deterioration for households who were already above the subsistence level. In such case, increased reliance on home food means that households had to forgo the more nutritious and diversified market-purchased foods. This stands parallel to the

ambiguous interpretation of rising calories consumption (Deaton and Drèze, 2009) where an increase in caloric intake may indicate greater food security, if households move away from survival constraint; or deteriorating diet quality, if households switch to higher shares of calorie-rich staple foods.

Nutrition policy design can benefit from accounting for the heterogeneity in people's need and strategies of subsistence. Rural households for whom cultivation is not the main activity (casual workers and other self-employed households) and who are not able to meet their basic needs would seem to benefit from increased own consumption, or other forms of support to increase basic food intake. On the other hand, farming households, who seem to have sufficient access to own production, display a preference for purchased foods.

CHAPTER 3

COMPARISON OF MACHINE LEARNING PREDICTIONS OF SUBJECTIVE POVERTY IN RURAL CHINA[13]

Abstract

Despite rising incomes and reduction of extreme poverty, the feeling of being poor remains widespread in China. Support programs can improve wellbeing, but they first require identifying who are the households that judge their income is insufficient to meet their basic needs, and what factors are associated with subjective poverty. For this purpose, this study aims to establish whether subjective poverty, which is more of an inner feeling than a visible characteristic, can be well predicted. Using microdata from China in which households report the income level they judge is sufficient to make ends meet, three machine learning algorithms, the random forest, support vector machines and LASSO regression, are applied to a set of socio-economic variables to predict subjective poverty status. For the first time, we show that prediction techniques are reliable to identify subjective poverty, with a performance of 85.29 percent of correct predictions with the random forest. The analysis offers specific attention to the modest-income households, who may feel poor but not be identified as such by objective poverty lines. A combination of low income, low endowment (land, consumption assets) and unusual large expenditure (medical, gift) constitutes the key predictors of feeling poor. To reduce the feeling of poverty, policy intervention should therefore continue to focus on increasing incomes. In addition, improvement in non-income domains such as health expenditure can also relieve the feeling of income inadequacy.

[13] The authors of this essay are Lucie Louise Maruejols, Hanjie Wang (corresponding), Qiran Zhao, Yunli Bai, Linxiu Zhang. Lucie Maruejols conceptualized the research idea, conducted the data analysis, results discussion and writing the paper. Hanjie Wang contributed to the data analysis, results interpretation, writing, and provided critical revision of the draft. Qiran Zhao, Yunli Bai, Linxiu Zhang performed data collection and preparation. We are thankful for two anonymous reviewers who provided comments on an earlier draft of this essay. This essay is under review at *China Agricultural Review of Economics.*

3.1. Introduction

The characterization of people suffering from poverty is the first step to poverty-alleviation programs (Dubois, 2012). Typically, such identification relies on data obtained from detailed field surveys, which are expensive and difficult to conduct, and may thus lack frequency or comprehensive geographical coverage. To fill this data gap, poverty prevalence can be predicted using limited subsets of consumption variables (Christiaensen et al., 2012; Mathiassen, 2013; and Nichols, 2018), or satellite data (Jean et al., 2016; Perez et al., 2017; Piaggesi et al., 2019; Ni et al., 2020; Ayush et al., 2020; Yeh et al., 2020; Tang et al., 2018). These predictions perform well and, because they are affordable and easier to conduct, open the possibility for more targeted and more timely identification of the need for anti-poverty measures.

Although essential, measures of poverty lines set by experts or governments are quite limiting, especially in context of rising incomes and improving living standards, as the feeling of poverty may extend beyond the satisfaction of objectively-set basic needs (Goedhart et al., 1977; Flik and van Praag, 1991; Van Praag and Ferrer-i-Carbonell, 2008). Instead, some proposed to establish poverty lines and national poverty ratios directly from households' own appreciation of what income levels correspond to living in poverty (Bishop et al., 2006; Wang et al., 2020). The resulting subjective poverty line better reflects people's own experience of deprivation and typically identifies different sets of poor households, who can be targeted with anti-poverty support (Mahmood et al., 2019; Wang et al., 2020). However, prediction of subjective poverty, although critical and different than objective income poverty, has never been carried out.

Therefore, this essay proposes to predict subjective poverty for the first time, for the case of rural China where, given the stark reduction in absolute poverty, subjective poverty is becoming more relevant. As the country enters a new era of poverty, the historical focus on providing basic subsistence (food and clothing), improving agricultural productivity and supporting industrial development and infrastructure, needs to make ways to a new strategy, with improvements in others areas that correspond to the needs of the modern population. The focus of this work is on rural households in particular, where income poverty is more prevalent than among urban households (Gustafsson et al., 2004) and people survive on much less income, and where perception of poverty differs from the poverty identified with the official objective poverty measures (Wang et al., 2020). Plus, the sample extends beyond the regions that were targeted by the government for anti-poverty measures recently, which are not the only ones concerned by subjective poverty (Liu et al., 2017). Machine learning techniques are particularly

well suited to predictions and are generally recognized to perform well. Among them, random forest is popular for its good predictions and because it can be replicated by government agencies, NGOs and other stakeholders (Browne et al., 2021; Htet et al., 2021). In addition, the present essay explores another classification tool, the support vector machines (SVM), and a regularized regression approach, the LASSO regression, to compare their ability to predict subjective poverty correctly. Therefore, this work joins a growing body of literature that examines performance of various machine learning technique on different contexts of poverty around the world (Alsharkawi et al., 2021; Li et al., 2019; Sani et al., 2018; Wijaya et al., 2020; Yin et al., 2021). Furthermore, an analysis of the factors that make the largest contribution to the predictions can inform policy efforts on the context that surrounds poverty. There is only limited literature on predictors of absolute poverty, and nothing on the predictors of subjective poverty, which likely differ. Thus, identifying different targets and different predictors can bring valuable information for the design of specific poverty-alleviation strategies that address people's feeling of poverty.

Section 2 provides contextual information about subjective poverty and prediction exercises. Section 3 presents the machine learning techniques and the data. Section 4 offers the results and section 5 concludes.

3.2. Framework

3.2.1. Objective and subjective poverty measures

Subjective poverty relies on judgment from the households themselves. In this approach, households are considered to be good judges of their situation and the conditions that constitute 'living in poverty'. Generally, in these self-assessed, income-based approaches, households provide a qualitative assessment of their situation ('good', 'bad') and poverty lines can be established by calculating the income levels that correspond to the feeling of poverty. Specifically, the Leyden Poverty Line (LPL) relies on the Income Evaluation Question (IEQ), where households are asked to indicate which levels of income correspond to a range of qualitative statement (very bad, bad, medium…). Alternatively, the Subjective Poverty Line (SPL) is obtained from the Minimum Income Question (MIQ or MINQ) where survey respondents are asked to indicate the income level that is necessary to meet basic living requirements, or to 'make ends meet' (Goedhart et al., 1977; Flik and van Praag, 1991). Some methodological obstacles to these approaches include households' underestimation of their

actual after-tax income and sample selectivity because the poor are typically under-represented in socio-economic data collection (Kapteyn et al., 1988). However, these subjective poverty lines are useful as they reflect the diverse needs of households with diverse conditions (one person or lone parent, number of earners,…) and provide differentiated poverty thresholds (de Vos and Garner, 1991).

Notably, self-assessed poverty provides a good alternative to the limits of more traditional objective poverty lines (Van Praag and Ferrer-i-Carbonell, 2008; Wang et al., 2020). First, measures of absolute extreme poverty, such as the World Bank's US $ 1.90 expenditure per day per person using 2011 purchasing power parity price, present valuable but partial picture of poverty. Among others, the underlying definition of 'basic needs' which are supposedly covered by $ 1.90 per day is debatable and how much income is necessary to achieve it depends on context. For example, the prices faced by the poor differ from the aggregated prices used to calculate the purchasing power parity (Deaton, 2010; Deaton and Dupriez, 2011). When more context-sensitive approaches are adopted, such as national poverty lines, measurements often rely on fulfillment of basic calorie needs, which are difficult to establish (Deaton and Drèze, 2009; Zhou and Yu, 2014; Yu and Abler, 2016). Furthermore, monetary income of rural households in developing countries context are complex to measure and incomplete (Van Praag and Ferrer-i-Carbonell, 2008). Also, absolute poverty lines do not inform on the depth of poverty (Brady, 2003; Kakwani, 1993; Sen, 1976).

Second, when quality of life and income tend to rise, extreme poverty lowers, but the experience of poverty can certainly persist. Indeed, the feeling of poverty also comes from expectations households have set for themselves, independently from objectively-set universal or national poverty lines. Indeed, the concept of minimally necessary income varies with income (Goedhart et al., 1977) and country (de Vos and Garner, 1991). As a result, people do not necessarily feel more satisfaction when income increases. These observations relate to the "Easterlin Paradox", which points that income is positively correlated with subjective well-being on the short-term, but is not associated with better wellbeing on the long-term (Easterlin, 1974), in particularly when economic inequalities increase together with income growth (Easterlin and Connor, 2020; Oishi and Kesebir, 2015). Relative measures of poverty (for example as a percent of the median) are useful but still rely on an objective judgment of what constitutes income poverty and provide an incomplete picture of the experience of income sufficiency by the households themselves.

Therefore, measuring poverty using subjective judgment better reflects the actual level of deprivation felt by the concerned households and can be a good complement to the information obtained otherwise (Mahmood et al., 2019). In fact, measuring poverty subjectively provides different, generally higher, levels of poverty than objective poverty lines (de Vos and Garner, 1991, Herrera et al., 2006a). Although objective and subjective poverty are closely related, the different lines do not identify the exact same set of households (Mahmood et al., 2019). For example in rural China, just 82 percent of the objective poor feel subjectively poor, while 29 percent of the rural households who are not objectively poor feel subjectively poor (Wang et al., 2020). Policies that aim at improving people's feeling of income sufficiency may thus consider different target recipients than policies focused on objective income poverty.

3.2.2. Determinants of poverty

Depending on whether policy aims at lifting incomes of the lowest-income households, for example to increase consumption, or at the satisfaction of a particular group, for example urban voters (Gustafsson et al., 2004), different poverty-alleviation strategies may be adopted. Information on the different sets of drivers that contribute to each type of poverty can help design specific strategies.

First, monetary income is clearly linked to subjective poverty, but its role is ambiguous as aspirations grow with the income level. Although wealthier households are objectively less poor, they report higher levels of necessary income to satisfy basic needs than poorer households (Herrera et al., 2006a; Wang et al., 2020). Beyond income, socio-economic factors also have strong influence on objective and in particular subjective poverty. De Vos and Garner (1991) find for the context of developed country (US and Netherland) that household income but also recent changes to income, household composition, age, education, sex, region, and fixed expenditures affect the minimum income level considered sufficient. In the context of Madagascar and Peru, Herrera et al. (2006b) find that health, education, job, family structure, social structure and trajectories, as well as incomes of neighbor and past income influence wellbeing. Past history and relative economic situation, as well as type of employment, are also found to matter for feeling poor in urban Ethiopia, where subjective poverty remains despite economic improvement and a reduction of objective poverty (Alem et al., 2014). Similar factors (education level, life cycle, and labour market status) are found to affect subjective income poverty for urban households of China (Gustafsson et al., 2004).

Comparing the determinants of both objective and subjective income-poverty, Mahmood et al. (2019) find that many are common to both, with similar direction but differences in effect strenght. However, some determinants exhibit opposite influence. For example, larger households are associated with higher objective poverty, and in some case with higher appreciation of the minimum necessary income (Goedhart et al., 1977), but it can also be associated with increased subjective wellbeing in some cases (Mahmood et al., 2019; Winkelmann, 2005; Wang et al., 2020). Similarly, in urban Ethiopia, receiving money from remittance improves objective poverty but not subjective poverty (Alem et al., 2014). Therefore, socio-economic and demographic determinants that drive objective and subjective poverty may differ, and thus constitute different levers to reduce poverty.

However, the ambiguity or inconsistency of certain determinants across studies suggest strong regional and context-dependent effects. For example, higher levels of education increase the subjective poverty standard set by households in rural China, although the subjective poor have lower education on average (Wang et al., 2020). In Pakistan however, higher education is linked to higher wellbeing (Mahmood et al., 2019). Household head age increases subjective poverty in Pakistan (Mahmood et al., 2019) but reduces it in China (Wang et al., 2020), alhtough it is possible its effect is nonlinear (Gustafsson et al., 2004). Clearly the identification of determinants and the direction of their effect is context dependent, which calls for context-specific studies.

3.2.3. Poverty in China: new policy strategy

In 2021, China's government announced the end of poverty, achieved notably with an impressive economic growth in the last 20 years and its program "The Targeted Poverty Alleviation" (TPA). Over the decades, a succession of poverty-reduction policies adapted to the different forms of poverty while targeting the recipients more narrowly overtime. Anti-poverty measures first focused on providing subsistence (food and clothing) to the rural population, where large shares of population faced extreme poverty at the end of World War II until the 1970's. Then, institutional reforms of the period 1978-1985 shifted from collective management to more autonomy for the farmers and has led to higher productivity. Next, anti-poverty efforts took a more local turn with the state-designation of target counties, which were mostly the remote deep rocky mountainous areas, border areas and minority areas of central and western China, which are the most affected by deprivation in basic food and clothing. Since 2013, the TPA focuses on removing all extreme poverty at household and village level, and ensuring that

anti-poverty measures reaches the designated poor (Yao et al., 2004; Liu et al., 2017; Guo et al., 2019).

Absolute extreme poverty is thus almost eradicated in 2021. However, despite economic growth and improvement of basic living standards, the feeling of poverty likely persists among certain households. Wang et al. (2020) showed that the households who feel poor are not always the same as households considered poor by objective poverty lines. Therefore, in order to avoid a policy failure of people continuing to feel poor despite the economic growth and the government's efforts, the government's credibility to end poverty now rests on its capacity to target subjective poverty, wellbeing and happiness (Liu et al., 2017).

As extreme poverty disappears, the new policy focus moves away from fulfilling minimum materialistic conditions toward improvement in non-materialistic factors, for example social status, relative economic position or other factors that affect wellbeing and satisfaction of life. For this purpose, it is critical to identify the households who might consider themselves poor, because traditional anti-poverty measures may not be effective anymore to reduce subjective poverty. A better understanding of what exogenous factors are associated with feeling of income inadequacy can inform new policy strategy.

3.2.4. Poverty prediction: targeting beneficiaries, identifying levers

Delivering poverty relief requires first and foremost an identification of who and where the poor are, which typically results from surveying the population. However, the infrequent and uneven geographical coverage of household surveys has been recognized as an obstacle to the relevant, complete and timely identification of poverty prevalence. To remedy the lack of data, some authors (Christiaensen et al., 2012; Mathiassen, 2013; McBride and Nichols, 2018) proposed to track poverty predictors, mainly consumption variables, to obtain poverty estimates from a subset of variables that would be more easily collected. Indeed, small scope surveys carried frequently between larger-scale and more comprehensive surveys can collect information on a subset of key variables that are sufficient to predict poverty in the absence of large and precise data. Despite usefulness and good performance, this approach still depends on field collection of survey data, which is likely constrained in scale and frequency.

As a response, a range of literature emanating from earth science and computer science examined the predictive power of satellite imagery using deep learning and convolutional neural network (CNN) approaches in the prediction of poverty rates, for example obtained from the

Demographic and Health Surveys. Daytime imagery combined with nighttime lights allows the models to identify features such as pools, road, or houses, which are treated as close proxy to economic development and in turn provide accurate and fine-scale prediction of poverty rates (Jean et al., 2016; Perez et al., 2017; Piaggesi et al., 2019; Ni et al., 2020; Ayush et al., 2020). However, these models are difficult to replicate because they rely on expensive data and require advanced technical knowledge. Perez et al. (2017), Tang et al. (2018), Yeh et al. (2020) and Ayush et al. (2020) show instead acceptable performance with lower resolution images, or even vegetation index, that are more freely accessible. Nevertheless, these approaches remain difficult to implement for non-specialists. Instead, Browne et al. (2021) focus on reproducible examples that can be implemented by relief agencies, NGOs or governments, thereby using only publicly available data and parsimonious models to make prediction of poverty and malnutrition. For this purpose, the random forest technique is shown to produce comparable or superior results to the deep learning approaches, but is simpler in use. Htet et al. (2021) confirm the good prediction performance of random forest to predict poverty in Myanmar: it performs comparably well to gradient boosting and xgboost, and much better than linear regression and ridge. Yin et al. (2021) also find that random forest outperforms other techniques, namely support vector machines and deep neural network. Other authors chose alternative machine learning algorithm to perform similar task of poverty classification, for example Kshirsagar et al. (2017) use regularized regression with elastic net estimation.

The advantages of predicting poverty using a limited number of predictors (either frequent subsets of socio-economic indicators or satellite data) are clear and far-reaching. Policy makers can rely on these variables to assess poverty prevalence without difficult, extensive and costly field data collection efforts. Such predictions can participate to the identification of program beneficiaries on a more frequent basis and potentially in wider areas (Browne et al., 2021).

Next to targeting beneficiaries, policy makers need to devise the tools to reduce poverty. For this, an understanding of the drivers of poverty is necessary. While random forest cannot formally establish causal links between an outcome and an explanatory variable, an examination of the variables that contribute to predictions provides useful contextual information about poverty (H. Wang et al., 2021). However, only few of the afore-mentioned studies go beyond the prediction results and report on factors that contribute to making correct predictions, as most include only day- and night-time imagery. In this regard, only Browne et al. (2021) report its largest predictors of malnutrition and poverty for 11 low and lower-middle income countries, which are physical geography features, followed by vegetation and climate. Food price and

conflict have low importance in prediction, but data construct (low in-country variation) might be responsible for this.

3.2.5. Literature gap and contribution

Existing prediction work about poverty in China is limited to how nighttime lights from a surveyed area can predict county level poverty in a non-surveyed area (Xu et al., 2021), how geo-spatial information can make poverty predictions for the Guizhou Province (Yin et al., 2021), urban poverty (Li et al., 2021), or county-level poverty (Li et al., 2019). Importantly, human wellbeing is not limited to income poverty, but is rather related to a sense of deprivation that can be better captured by measures of subjective poverty, multidimensional poverty or inequalities. A few studies examine the ability to predict multidimensional poverty so far (Pokhriyal and Jacques, 2017; Yeh et al., 2020; Alsharkawi et al., 2021, Tingzon et al., 2019). However, no study has tried to predict subjective poverty and identify what are its most important predictors.

3.3. Machine learning algorithms

The current study aims to establish whether personal judgement of poverty experienced by the households can be well predicted, and does so by comparing the ability of several machine learning techniques to make correct predictions. Given the availability of both input and output vectors, supervised machine learning algorithms are appropriate to attempt predicting subjective poverty in rural China.

Accuracy of predictions depend on a model's ability to balance the tradeoff between bias and variance, and to avoid overfitting (Belkin et al., 2019). Ideally, an algorithm makes predictions that are close to actual values (low bias) and the level of accuracy is consistent when predictions are carried out with new input datasets (low variance). To judge the performance of a machine learning technique, data is typically split between a train set, used to build the algorithm, and a test set, used to verify the accuracy of predictions when the algorithm is faced with new input data.

An algorithm can make highly accurate prediction and guarantee low bias when it fits precisely the features of the training data. But when a model matches the training data too closely, it loses flexibility to adapt to new input data, and overfitting occurs. In this case, predictions made with

the test set, which contains slightly different features than the training set, produce more misclassification (high variance) and thus a lower performance.

Therefore, a balance in the level of model complexity must be found to make reliable predictions. The model should be complex enough to capture the important trends that determine the outcome, but without focusing too closely on secondary factors, so that it maintains an ability to make good predictions with new sets of data. Various machine learning techniques have been developed to achieve low variance while maintaining low bias, and thus avoid overfitting. Several of them are explored in the following section.

Finally, machine learning approaches are sometimes criticized for their failure to make the underlying mechanism of the prediction process interpretable and to generate 'black box' predictions (Storm et al., 2020). In order to verify the plausibility of the predictions, and understand the relationship between the features and the outcome, we select models that can explicitly identify the features responsible for the predicted outcome. This also presents the advantage to identify potential target factors for policy intervention.

3.3.1. Random forest algorithm

Considerable literature, in recent years, has shown an increased interest in the decision tree model for its interpretability and adaptability to non-linear relationship, which could provide explanations for economic phenomena. At each step, this classification technique choses the variable that best separates the data according to the group they belong. Specifically, the algorithm finds the variable that minimizes the Gini impurity, a measure of the probability of making incorrect classification. Nevertheless, one major drawback of this approach is the risk of overfitting problem as the decision tree model tends to precisely fit the training dataset.

To overcome the drawbacks of single decision tree, the random forest algorithm (Breiman, 2001) employs an ensemble methodology to train the model, by combining many decision trees. Using bootstrapping technique, the random forest algorithm trains multiple decision trees on randomly selected samples from original dataset, using a random subset of predictors with each tree. Results are then aggregated with a majority rule. The optimal number of trees to include in the forest is chosen so as to minimize the prediction inaccuracy. For this, the out-of-bag (OOB) error measure is used, which indicates the mean prediction error on each bootstrap sample. The out-of-bag sample consists in the entries that were not included in the bootstrapped

sample (typically 1/3 of original data). For each tree, the algorithm trained on the bootstrapped sample makes prediction on the out-of-bag sample and the share of wrong predictions is reported as the OOB error rate. Mean OOB are calculated for forests of various size and the forest with the number of trees that minimizes the mean OOB is selected.

This technique preserves the low bias advantage of the single tree technique. However, the aggregation process also limits the risk of overfitting as the predictions are not based on a single set of data but rather on bootstrapped samples of the data. Using bootstrapped datasets provides some variation in the input data and therefore avoids that the resulting algorithm matches too closely one single dataset. Compared with individual decision tree, training the algorithm using several trees therefore reduces the variance of the results to new input data. The aggregated result preserves the low bias as it still relies the features of the original data, but also delivers low variance.

Explicitly, the random forest algorithm can be defined as in (H. Wang et al., 2021):

Let a set $(\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_n})$ denote the features of the training dataset, and $(\mathbf{y_1}, \mathbf{y_2}, \cdots, \mathbf{y_n})$ denote the outputs. Then we use bootstrap aggregating or bagging, where a random sample is selected B times from the training dataset.

First, for $b = 1$ to B:

(a) Select a bootstrap sample S_b from the training data.

(b) Grow a classification tree T_b to the bootstrapped data S_b.

Second, after B times iterations, the algorithm outputs the ensemble of trees $\{T_b\}_1^B$.

Finally, we let $\hat{P}_b(x)$ presents the class prediction of the bth decision tree. With this, we can use the majority voting to make a final prediction:

$$\hat{P}_{rf}(x) = majority\ vote\{\hat{P}_b(x)\}_1^B$$

Random forest is among the most popular techniques to conduct prediction: the tree structure is able to handle non-linear relationship between the outcome and the feature, the Gini impurity allows ranking the predictors according to their contribution and facilitates interpretation, finally using forest instead of tree increases performance of model on new data. However, we also explore the potential of alternative machine learning techniques that rely on different mechanisms.

3.3.2. Support vector machine

The support vector machine (SVM), introduced in Cortes and Vapnik (1995), is a classification technique that assigns new data to either side of a hyperplane built with input data (Lesmeister, 2015). First, support vector *classifiers* find the hyperplane that best separates the training data by maximizing the distance (the margin) between the hyperplane and the points on the edge of each category, called the supporting vectors. By allowing some misclassification, the choice of hyperplane is made less sensitive to possible outliers in the input data, and thus produces less variance. The exact choice of hyperplane is determined by cross-validation, which establishes how much misclassification should be allowed to obtain the combination of best classification and lowest variance. However, the simplicity of a hyperplane can be limiting when the relationship between variables is non-linear and the choice of a flat hyperplane produces large misclassification. Instead, in support vector *machines*, nonlinear transformations are applied to the data, for example polynomial transformation (Hofmann, 2006). This allows separating the groups more effectively and finding nonlinear decision boundaries, which form support vector classifiers of higher dimensions. More precisely, with the kernel trick, the technique takes the data from a space where it is not possible to find a separating hyperplane and translates it into another space where such a hyperplane exists.

The interest of this technique is its ability to handle outliers and overlapping classification because misclassification is allowed. It is known to perform well with small samples of data. Fauziah and Ruliana (2022) and Boto Ferreira et al. (2021) find that it performs better than the other classification technique in poverty-related applications.

3.3.3. Least absolute shrinkage and selection operator

Last, the regularization technique of least absolute shrinkage and selection operator (LASSO) (Tibshirani, 2016) aims to improve prediction by reducing variance of mean squared error in a regression approach (Lesmeister, 2015). Specifically, the operator finds the coefficients that minimize the residual sum of squares (RSS) plus a penalty, where the penalty consists of a tuning parameter lambda λ and the sum of absolute values of the coefficients β_i.

$$\widehat{\beta}_\iota = \arg\min(RSS + \lambda \sum |\beta_i|)$$

This minimization process thus introduces a bias that tends to shrink the coefficients, therefore reducing the variance of out-of-sample predictions to new input data. In the fact, the LASSO

effectively brings some coefficients to zero, which makes it an interesting technique for feature selection, where only the variables that explain most of the variance in the outcome are retained by the model. Chernozhukov, Hansen, and Spindler (2016) propose an improvement of the LASSO technique where the penalty level λ is not chosen by cross-validation, but is a function of the predictor vectors. This 'rigorous LASSO' is therefore data-driven and grounded in theory:

$$\widehat{\beta}_i = \arg\min(RSS + \frac{\lambda}{n}\sum|\theta\beta_i|)$$

where θ represents the weights or penalty loadings, which are chosen so that they insure basic equivariance of coefficient estimates, i.e. the weights account for the fact that predictors may have different variances and prevent lambda from being affected by this. Implementation of this technique for a logistic regression is shown by Belloni et al. (2013) and Belloni et al. (2016).

The strength of this technique lies in building parsimonious models that can handle multicollinearity, reduce overfitting and improve accuracy. The regressors retained by the model also help interpretability.

All in all, the three methods aim to improve prediction accuracy by reducing variance due to overfitting. The classification tree finds, at each step, the variable that separates the data with the least amount of impurity, and classifies new data according to the identified predictors. Low variance is achieved via aggregation of many trees. The support vector machines instead classify new data points based on a separation drawn from the input data. Variance is reduced by tolerating some misclassification at the input data stage, which avoids that new data from being too strongly affected by outliers. On the other hand, the LASSO regression approach shrinks the model to only the few variables that best capture the features of the data, to create a lower variance of the squared error on new data.

3.4. Data

To examine the subjective poverty in rural China, we use a nationally representative survey data of 2025 rural households in 2016. The data is collected by the Center for Chinese Agricultural Policy, Chinese Academy of Science, covering five representative provinces: Jilin, Shannxi, Jiangsu, Sichuan, and Hebei. Specifically, considering the five major agro-ecological zones in China, the survey randomly selected a sample province from each zone to represent the zone.

Within the sample province, the survey divided all counties into five groups ranking by the per capita gross value of industrial output (GVIO) and then randomly selected one representative county from each group. Likewise, the survey randomly further selected two towns from each county, two villages from each town, and 20 sample households from each village. Accordingly, the survey finally collected data for 2025 representative rural households.

The output vector in this study is the subjective poverty status of rural households. Among the various measures for identifying the individuals' subjective poverty in the existing literature, we use the MIQ (Minimum Income Question) method as this approach is more understandable for the respondents. Particularly, the survey question goes as: "*Please offer an income amount below which you will feel poor for a household as yours*". The information provided by the respondents can be regarded as the subjective poverty standard of the rural households as it reflects the individual living condition, subjective well-being, and regional development level (Wang et al., 2020). Notably, we further divide the self-reported minimum income by the household family size in response to the heterogeneity of family size. Henceforth, the subjective poverty status of rural households could be easily identified by comparing per capita income and subjective poverty standard. That is, if the per capita subjective poverty standard of a rural household surpasses per capita income, we attach a subjective poverty status to the household. As a result, we construct a binary variable of subjective poverty as the output vector, where 44 percent of the households are considered to experience subjective poverty.

Regarding the input vectors, we consider a wide range of socioeconomic characteristics of rural households, including economic characteristics, family characteristics, human capital, social capital, and major expenditure. Explicitly, the economic characteristics refer to per capita income[14], land size, house value, productive asset value, and consumption asset; the family characteristics refer to family size, number of elders, number of children, number of labor forces, whether the head is a village leader, and whether the head is a party member; the human capital refers to average education level and average health condition of family members; the social capital refers to the number of friends or relatives working in the government as well as working as managers in a state enterprise; the major expenditure refers to education

[14] Income is calculated following the official standard of the China National Bureau of Statistics, and consists of wage income, operating income (both agricultural and non-agricultural activities), property income and transfer income.

expenditure, medical expenditure, gift expenditure, and wedding expenditure. Table 3.1 presents the definitions of variables.

Table 3. 1: Definition of variables

Variables		Definition	Unit
Output	Subjective poverty status	Whether per capita income surpasses subjective poverty standard? Yes=0, No=1	dummy
Family characteristics	Family size	The number of family members	person
	Number of elders	The number of family members whose age is above 65	person
	Number of children	The number of family members whose age is below 15	person
	Number of labor forces	The number of family labor forces	person
	Village leader	Whether a village leader? Yes=1, No=0	dummy
	Party member	Whether a party member? Yes=1, No=0	dummy
Human capital	Health condition	Average health condition of family members (1=very good, 2= good, 3=general, 4=bad, 5= very bad)	category
	Education level	Average school years of family members	year
Economic characteristics	Per capita income	Log of per capita income of household	yuan
	Land Size	Log of the area of cultivated land	mu
	House value	Log of the value of the house	yuan
	Productive asset value	Log of the value of the productive asset	yuan
	Consumption asset value	Log of the value of the consumption asset	yuan
Social capital	Government organization	How many friends or relatives working on the government organization?	person
	Enterprise's manager	How many friends or relatives working as a manager on the enterprise?	person
Major expenditure	Education expenditure	Log of the total education expenditure	yuan
	Medical expenditure	Log of the total medical expenditure	yuan
	Gift expenditure	Log of the total gift expenditure	yuan
	Wedding expenditure	Log of the total wedding expenditure	yuan

3.5. Results and discussion

3.5.1. Whole sample predictions

The whole sample is split to obtain 1215 training data (60 percent) and 810 test data (40 percent). To start, we need to choose the optimal number of trees to use in the random forest

algorithm in order to reduce misclassification errors. The minimum out-of-bag (OOB) error is found with a number of 168 trees (see Figure 3.1), which indicates that aggregation over this number of trees provides the most accurate predictions.

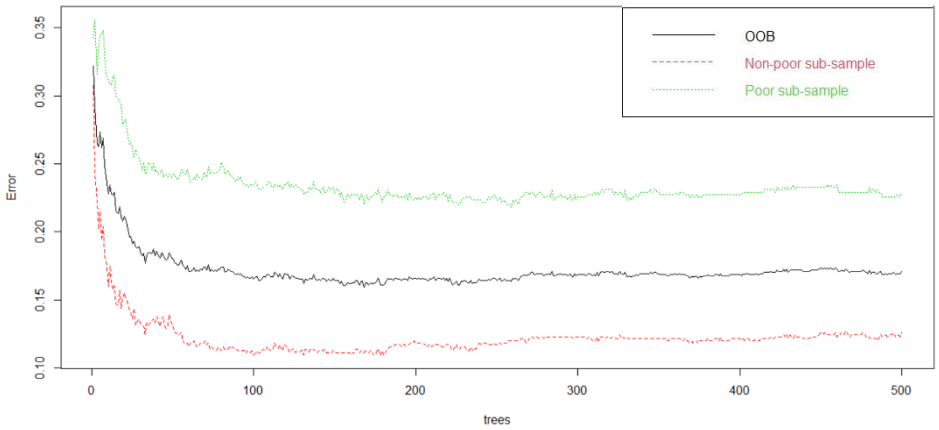

Figure 3. 1: Optimal tree number

Table 3.2 presents the prediction results for subjective poverty in rural China, built with 168 trees. For the whole sample, the random forest algorithm provides an overall out-of-bag (OOB) error of 17.52 percent on the train set, which indicates the accuracy (ACC)[15] of the out-of-bag sample is 82.48 percent. When we use the test dataset to conduct the best-trained model, the predictive accuracy (ACC) slightly increases to 85.29 percent. In other words, the trained random forest algorithm can correctly predict the subjective poverty in rural China with a high accuracy of 85.29 percent, providing a powerful tool to target the population who present a subjective poverty status.

Table 3. 2: Prediction for subjective poverty

Performance measures	Random Forest	Support Vector Machines (Linear)	Support Vector Machines (Polynomial)	LASSO Logit
Misclassification error (train set) *	17.52%	17.93%	19.82%	16.95%
Accuracy (ACC, test set)	85.29%	85.29%	84.30%	84.44%
Sensitivity (True positive rate, test set)	83.30%	83.89%	82.78%	82.97%

* Out of Bag (OOB) Error

[15] Accuracy is defined as the percentage of correct predictions.

To continue with another classification tool, the linear support vector machines data achieves very similar results. Using 10-fold cross validation, the support vector machines using a linear approach establishes that best performance is achieved by allowing 17.92 percent of misclassification, giving an accuracy on the test data of 85.29 percent as well. Alternative to the linear SVM, predictions done with the polynomial SVM also achieve high performance, but not as high as the linear model, with accuracy of 84.3 percent.

The regularized regression approach, using the rigorous LASSO, shrinks all coefficients, effectively bringing some to zero and leaving only the most important predictors in the regression. In this model, 4 variables out of the 19 predictors are selected. Predictions made with the 4 predictors reach a similarly high level of accuracy, with 84.44 percent correct predictions on the test set.

According to the above prediction results, we have substantiated the potential of machine learning techniques to predict subjective poverty, as the accuracy is relatively high and similar for the three methods. However, the socioeconomic characteristics entered in the models include a wide range of variables, providing limited interest for policy making. The next section investigates which of the variables are important for subjective poverty prediction.

3.5.2. Variables importance and economic explanations

To reveal the underlying mechanism of subjective prediction results, the random forest model and the LASSO regression provide useful information. First, the random forest model calculates the mean decrease in the Gini index to mirror each variable's contribution to making correct predictions. With this, we can easily capture the key predictors for the subjective poverty prediction. Table 3.3 indicates the ranked importance of the main variables to the random forest classification. Explicitly, the income of rural households accounts for the largest contribution to the classification, which indicates that income level is strongly correlated with subjective poverty status. Second, the LASSO regression retains the variables with sufficient explanatory power to explain variance in the outcome, without over fitting the model. Here the model selected income per capita, size of the household, gift expenditure and house value.

Table 3. 3: Variables of importance and main predictors

	Random forest	**LASSO regression**
Rank	*Variables of Importance*	*Predictors*
1	Income	Income
2	Consumption asset	Size of the household
3	Land	Gift expenditure
4	Labor	House Value
5	Medical expenditure	
6	Education level	

Descriptive statistics for these important variables are reported (see Table 3.4). It is clear that the mean values of key predictors are significantly different between the subjective poor and the subjective non-poor. Overall, it is striking that the two techniques both achieve high and similar prediction accuracy but that the variables responsible for the prediction mechanisms differ. Only income is common to both approaches, which emphasizes its primary role in the feeling of poverty.

Table 3. 4: Descriptive statistics of important variables

Important Variables	**Non-poor** ***(n=1124)*** *Mean*	**Poor** ***(n=901)*** *Mean*	*Mean diff*
Per capita income [a,b]	9.50	7.55	1.95***
Medical expenditure [a]	7.24	7.60	-0.36***
Consumption asset [a]	0.35	-0.23	0.58***
Land Size [a]	1.35	0.60	0.76***
Education level [a]	7.02	6.37	0.65***
Number of labor force [a]	3.03	2.20	0.83***
Size of the household [b]	4.49	3.72	0.77***
Gift expenditure [b]	7.29	6.94	0.35***
House value [b]	2.29	2.10	0.20***

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

a: main predictor of random forest, b: selected by LASSO logit

As expected, per capita income of the subjective non-poor is significantly higher than the subjective poor. The economic explanation possibly lies in the fact that with the growth of income, rural households are more likely to meet the subjective poverty standard, resulting in a lower incidence rate of subjective poverty. This has been substantiated by the findings of existing literature (e.g. Stevenson; Mahmood et al., 2018; Wang et al., 2020). In addition to income, consumption asset, land and education level also make a great contribution to the random forest classification. Such findings match expectations that rural households with more material capital and higher education level tend to have the ability to improve their economic situation. Apart from this, medical expenditure plays a vital role in the subjective poverty

prediction in rural China, especially for middle- and high-income groups. Table 3.4 clearly reveals that, on average, the subjective poor have a heavier financial burden of medical expenditure than the subjective non-poor. It is generally known that rural households in China have a high burden of medical expenses due to the imperfection of medical security system (Fang et al., 2010), and that it contributes to poverty (Guo et al., 2019). Consequently, the rural households with high medical expenditure would need more money to compensate for it, increasing their subjective poverty standard. Once the income level of those households cannot meet the subjective poverty standard, they would fall into subjective poverty. As such, it is understandable that medical expenditure could perform well in the prediction of subjective poverty within the random forest algorithm.

Generally, our result show that the prediction of subjective poverty in rural China mainly relies on the income of rural households, and therefore corresponds to some extent to objective income poverty. Compare to existing research, this conclusion nevertheless provides a new perspective to understand poverty in rural China. Although the extreme poverty was eradicated in 2020, the subjective poverty perception is still highly related to the income of rural household, which means that the Chinese government should pay more attention to the sustainable growth of income but not only the eradication of extreme poverty.

3.5.3. Split income groups

Moreover, considering the heterogeneity of different income groups, we split the whole sample into three sub-groups: low-income, middle-income and high-income groups. Specifically, the low-income group is identified by the 20 percent of households with the lowest income, while the high-income group is on the top of 20 percent. Table 3.5 reports that most households of the lowest income range, with income going up to 2467 yuan per capita per year (or about 1.02 USD/day),[16] declare living in poverty. Expectedly, almost all households of the richest group feel sufficiently wealthy. The opinions of the middle-income group are more mixed.

[16] 6.644 Yuan/US$ in 2016, Source: PPPs and exchange rates, OECD data

Table 3. 5: Summary statistics for different income groups

Income group	Number of observations	Subjective Poverty	Income (in Yuan/capita/year)			Income (in USD/ capita /day)	
		Mean	*Mean*	*Min*	*Max*	*Min*	*Max*
Whole	2025	0.44	10891	1	121952	0.00	50.29
Low	405	0.98	1014	1	2467	0.00	1.02
Middle	1216	0.40	8978	2467	17499	1.02	7.22
High	405	0.05	26492	17500	121952	7.22	50.29

Table 3.6 shows prediction results for each group performed by random forest, the technique which had performed best on the whole sample. The out-of-bag (OOB) errors of the three groups are 2.46 percent, 24.25 percent and 5.33 percent respectively, which indicate that subjective poverty is straightforward to predict among members of the lowest and highest income levels, as seen by the very low error rates, but that predictions for the middle-income group are subject to more uncertainty. In particular, the forest makes a wrong prediction for about 1 in 4 households of this group regarding their subjective poverty status. This is confirmed with predictions made on the test set, where the accuracy rates for low income and high-income groups are up to 98.14 percent and 95.03 percent respectively, while the accuracy for the middle-income group is 73.25 percent.

Table 3. 6: Random forest prediction for subjective poverty by income group

	Low income	Middle income	High income
Out of Bag (OOB) Error	2.46%	24.25%	5.334%
Accuracy (ACC, test set)	98.14%	73.25%	95.03%

An observation of the main predictors, in Table 3.7, reinforces the primary role of income in predicting subjective poverty. In addition, it is clear that for low-income households demographic variables participate to the feeling of poverty, as number of children and family size are among the top predictors. On contrary, these variables do not appear for the middle-income group, where it is markers of large expenditures (medical, gift, consumption) that contribute to make predictions of subjective poverty.

Table 3. 7: Variables of importance from random forest by income group

Rank	Low income	Middle income	High income
1	Income	Income	Medical expenditure
2	Children number	Consumption asset	Income
3	Enterprise	Medical expenditure	Productive asset
4	Family size	Education level	Land
5	Education level	Gift expenditure	Health
6	Land	Land	Consumption asset

3.5.4. Focus on middle-income group

This group deserves specific attention because it contains most of the households that feel poor but are not identified as such by their income level. Indeed, almost all households of the low-income group are classified as being subjectively poor, and the main predictor for this group is, unsurprisingly, income. It appears evident that lifting incomes should be the focus of policy intervention for this group. On the contrary, the case of households just above a certain income level, for example here above the 20 percent poorest households, is less straightforward. First, if income-based poverty measures do not identify them as poor, they will likely not benefit from support program that could help improve their welfare. Several studies have shown that the people who feel poor are not necessarily recognized as poor by objective measures. Second, some households with similar levels of income may on the contrary feel that their income is adequate, as seen mostly with the left half of the red-shaded area on Figure 3.2, indicating that other factors may drive the feeling of poverty. Therefore, policy programs aimed solely at increasing income to relieve poverty may either be wasteful (targeting people who do not feel poor) or incomplete, as other drivers of poverty are not addressed. Therefore, this section aims to understand what, beyond income, drives the feeling of poverty for households with a modest, but not the lowest, income.

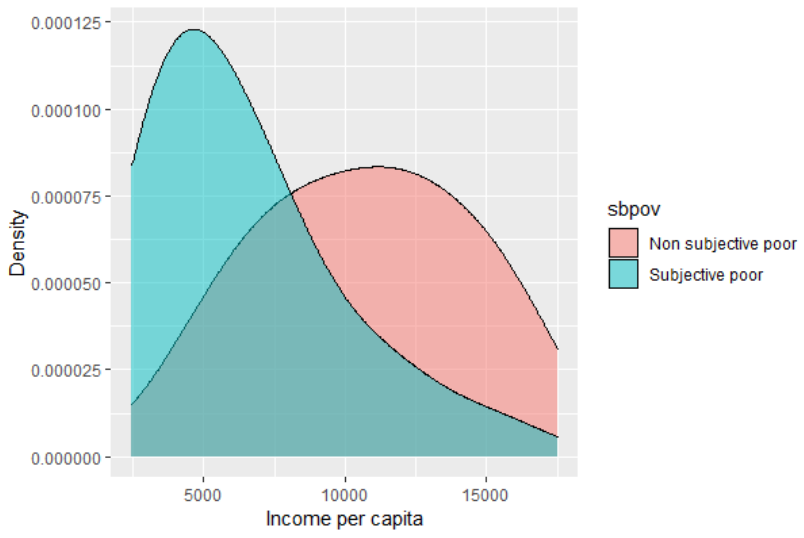

Figure 3. 2: Distribution of observations by income and subjective poverty status

Expectedly, subjective poverty predictions using random forest algorithm are not as reliable for the middle-income group as for other groups, where most of the households belong to the same poverty status. We explore if the SVM approach or the LASSO logit regression can improve prediction of subjective poverty for this income group. As seen in Table 3.8, the performance of the LASSO logit regression, with an accuracy on the test set of 75.7 percent, is slightly better than the prediction ability of random forest, 73.25 percent.

Table 3. 8: Prediction for subjective poverty for middle-income group

Performance measures	Random Forest	Support Vector Machines (Linear)	Support Vector Machines (Polynomial)	LASSO Logit
Accuracy (ACC, test set)	73.25%	72.4%	73%	75.7%
Sensitivity (True positive rate, test set)	59.3%	57.2%	57.7%	62.5%

Furthermore, results from the LASSO logit are valuable because of their interpretability. Out of the 19 predictors of the model, the LASSO algorithm selected 5 variables to explain most of the variance in the data: income, education level and gift expenditure (which are common with the top predictors selected by random forest), as well as whether the household is a village leader and size of the household. Random forest selected consumption asset, medical expenditure and land as main predictors on top of income, education level and gift expenditure. Summary statistics for these variables, provided in Table 3.9, show that middle-income households that feel poor show a combination of lower income but higher gift and medical expenditure than households who do not feel poor. However, they possess smaller lands and

fewer consumption assets (television, computer, camera, car, air conditioner…). Finally, households who feel poor are more likely to be leaders of their village, and to have smaller families than households whose income is judged adequate.

Table 3. 9: Descriptive statistics of key predictors for middle-income group

	Subjective non-poor (*n=731*)				**Subjective poor (*n=485*)**			
	Mean	*St. dev.*	*Min*	*Max*	*Mean*	*St. dev.*	*Min*	*Max*
Per capita income [a,b]	10658	3741	2500	17499	6446	3367	2467	17494
Gift Expenditure [a,b]	3568	4402	1	35000	4437	5659	1	60000
Education level [a,b]	6.55	2.35	0	13.8	6.63	2.57	0	15
Medical Expenditures [a]	9441	29453	1	510600	10513	23041	1	300000
Land Size [a]	11.18	13.5	0.01	109	9.78	10.45	0.01	67
Consumption asset [a]	2.86	4.61	0	44.39	2.79	5.4	0	48.39
Leader [b]	0.07	0.26	0	1	0.1	0.3	0	1
Size [b]	4.54	1.82	1	13	3.8	1.7	1	10

a: main predictor of random forest, b: selected by LASSO logit

3.6. Conclusion and policy implications

Improving the subjective well-being is a key issue in rural China, after the eradication of extreme poverty at the end of 2020. Most of the existing literature in the field of subjective poverty focuses on the definitions and measurement of subjective poverty. To the best of our knowledge, no previous study has paid attention to the prediction of subjective poverty, although predictive work has a potential to help in the design of policies by targeting recipients and factors. Against this background, we employ three machine learning algorithms, the random forest model, support vector machines and LASSO regression, to predict subjective poverty in rural China.

The random forest algorithm incorporating socioeconomic characteristics could accurately predict 85.29 percent of subjective poverty. The two other techniques achieve similarly high performance. In particular, the essay insists on middle-income households where the study of subjective poverty is more relevant than objective poverty, because people with similar income level have differing opinion on whether they live in poverty. In particular, this group contains the households that may feel poor but are ignored from support program because they are not identified as poor by objective poverty lines. When applied to split income groups, performance of the random forest falls to 73.25 percent accuracy for this group, but the LASSO logit

regression outperforms the random forest on this group, with an accuracy of 75.7 percent. Interestingly, the analysis reveals different key poverty predictors for this group than for the lowest-income group.

Importantly, both random forest and LASSO regression provide useful information about the features involved in the prediction mechanism. Generally, we find that the income of rural households has the greatest contribution to the subjective poverty prediction for all levels of income, which indicates that the government should continue to promote the growth of income in rural area even though extreme poverty has been eradicated. On top of income, demographic characteristics (size of household and number of children) are also part of feeling poor for low-income households in rural China. Regarding middle-income households, it is rather socioeconomic characteristics that are strongly associated with subjective poverty. In particular, these households report income insufficiency when they are faced with a combination of low income, low in-kind endowment but high occasional expenditure, primarily for gifts and medical treatments. Given that large expenditure is more likely transitory than permanent, further research is needed to know whether the subjective poverty among middle-income group is enduring and should be the concern of policy makers. The analysis also suggests that if increasing incomes became difficult and uncertain to realize through policy support, the feeling of poverty could be reduced by focusing efforts on other domains, such as improving affordability of medical care or financial support for exceptional expenditures.

The contributions of this essay are twofold. On the one hand, we evidence the potential of several machine learning techniques in the field of subjective poverty prediction. Predictions conducted with random forest, support vector machines and LASSO logit show good accuracy performance. Generally, the two classification approaches could deal with non-linearity present in the relationship between subjective poverty and socioeconomic characteristics. However here, the linear support vector machine outperformed its polynomial approach, suggesting that nonlinearity is not a major trait of the present data. This could support why the LASSO regression approach also performs well. The techniques also differ in how they handle the bias-variance tradeoff. While random forest tries to minimize misclassification, and compensate for overfitting risk by aggregation over a high number of trees, the SVM and LASSO approaches intentionally allow some degree of bias, determined in relation to the variance by cross-validation. Applied to the present data, this contrast did not create major differences in results.

On the second hand, the accurate prediction of subjective poverty could help policymakers efficiently identify the subjectively poor population, providing valuable information for ex-ante intervention. Results from both random forest and LASSO logit provide complementary information on key predictors, which helps better understand the underlying mechanism of subjective poverty prediction process. Monitoring determinants of subjective poverty can help indicate its prevalence in certain regions or among certain social groups without using frequent, extensive and time-intensive surveys. Furthermore, the predictors provide contextual information to use as leads for policy action. This facilitates targeting of support programs or additional data collection efforts.

CHAPTER 4

VIETNAM BETWEEN ECONOMIC GROWTH AND ETHNIC DIVERGENCE: A LASSO EXAMINATION OF INCOME-MEDIATED ENERGY CONSUMPTION[17]

Abstract

Ethnic divergence in energy consumption is a phenomenon that threatens the potential gain in wellbeing that developing countries can achieve with strong economic growth and rise in income. However, the underlying mechanism preventing ethnic groups from accomplishing a successful transition to modern fuels is not yet well-understood, requiring an in-depth analysis of interaction effects between ethnicity and rise in income. The case of highly ethnically-diverse Vietnam offers an opportunity to examine the role of ethnicity on the energy transition, at the stage when grid coverage is almost universal and income begins to rise thanks to economic growth. A methodological framework exposes the direct effect of ethnicity as well as the indirect, income-related effect of ethnicity on the ability of rural households to increase their electricity consumption. Using data from Vietnam's 2010 Household Living Standards Survey, feature selection is conducted with a machine learning technique, the *least absolute shrinkage and selection operator* (LASSO). Furthermore, a mediation analysis shows that income acts as a full mediator of ethnicity with respect to electricity consumption and as partial mediator for electricity expenditures. The results thus reveal a positive interaction effect between income and ethnicity, where Kinh are more able than non-Kinh to translate extra income into higher electricity usage. Our results highlight the urgent need to identify and address the non-income barriers that create ethnic disparities in the ability of poor households to increase electricity use.

[17] The authors of this essay are Lucie Maruejols, Lisa Höschle and Xiaohua Yu. Lucie Maruejols conceptualized the empirical framework, performed the data analysis and results discussion, and wrote the manuscript. Xiaohua Yu helped conceptualize the idea, access the data and provided critical feedback. Lisa Höschle contributed to the data analysis, results interpretation, writing, and provided critical revision of the draft. This chapter is under review at *Energy Economics*. We are grateful to Dr. Linda Steinhuebel and two anonymous reviewers for comments on an earlier draft.

4.1 Introduction

Since the early 1990s, Vietnam experiences a dynamic economic growth, with annual GDP growth rates around 6 percent (World Bank, 2021), accompanied by a fast-increasing electricity demand. Production capacities and investment in infrastructure were developed to meet the growing electricity demand. As a result, around 95 percent of all rural households are technically in condition to access the electric grid (Nguyen et al., 2019) and electricity access in rural areas is generally better than for other utilities (Dang, 2012). However, high electrification rates do not guarantee electricity access for all Vietnamese households. Particularly members of rural and ethnic minorities face barriers which decelerates the much-needed transition from traditional energy consumption - dominated by biomass - to a modern, safe and clean energy consumption characterized by the use of electricity.

Despite great macroeconomic growth in the last 20 years, energy costs have risen more than income in Vietnam, especially in low-income deciles (Nguyen et al., 2019). Consequently, poor households are exposed to higher energy expenses, resulting in increased access barriers to modern energy sources. Furthermore, additional mechanisms seem to be at work, which condition energy poverty. In particular, Nguyen et al. (2019) show that it is predominantly ethnic minorites that are subject to energy poverty. Over the period between 2004 to2016, most Vietnamese households successfully transitioned to modern fuels owing to their improved economic status. However, the observed trend did not apply to the poor and ethnic minorities, where benefits from improved access to electricity were cancelled out by energy expenditure (driven by coal) increasing faster than income over the same period. Inequalities with regards to energy consequently increased. It is therefore of particular importance to support mentioned sub-populations in their transition process towards modern energy (electricity) usage and reduce reliance on coal and biomass. However, the underlying reasons for a slower transition process compared to the dominant ethnic group remain uncertain. Clearly, the persistence of energy poverty among ethnic groups calls for a closer look at the mechanisms and determinants of electricity usage across social groups.

So far, research has been primarily concerned with the nexus between health, environmental implications or rural-urban disparities and energy sources. Income has been mapped out as one of the main drivers for successful transition, indicating that rising income accelerates the transition process towards modern energy sources (Barnes and Qian, 1992; Dowd, 1989; Fitzgerald et al., 1990). However, additional factors have been shown to influence fuel choices, such as low awareness and lack of familiarity with electricity (Bernard, 2012; Ranganathan,

1993), bureaucratic obstacles or unstable grid connection and reliability (Lee et al., 2016) as well as the impact of neighbors' responsiveness and adaptation to the electric grid (Bernard and Torero, 2015). As a result, access barriers that prevent households from realizing their transition emerge, even if income is rising (Gertler et al., 2016). Furthermore, a persistent ethnic divide in many countries represents a major source of marginalization and discrimination with respect to access to economic opportunities (Heath and Cheung, 2007). Yet, aspects of ethnicity and race impacting upon a comprehensive household transition process have been so far neglected in the literature. As a result, income remains one of the main determinants of energy transition and increasing electricity usage in the developing world (Mcneil and Letschert, 2010; Gertler et al., 2016). In the context of Vietnam, where a persistent racial energy usage divide can be observed despite substantial economic growth and good infrastructure coverage, further examination of the income channel, affecting energy usage through ethnicity, is consequently required. Looking beyond mere electricity connection, the present study uses data from the 2010 Vietnam's Household Living Standards Survey to understand the interaction between rising income and ethnic determinants in order to identify the barriers that may prevent ethnic minorities from transitioning to higher electricity consumption in Vietnam.

Consequently, this essay contributes to an improved understanding of electricity consumption of ethnic minorities in three ways. Firstly, electricity consumption and more generally, energy poverty, have been framed as multidimensional issues that depend on a great number of socio-economic determinants. In order to reduce the dependence of the results on many predictors, we aim to build a parsimonious model where only the most important determinants are selected by using the data-driven process *least absolute shrinkage and selection operator* (LASSO), a machine learning technique. This process reduces variance and allows the model to more closely match the Vietnamese case. Secondly, we confirm that income is a valid channel of ethnic distinction using, inter alia, the mediation approach by Baron and Kenny (1986). Finally, the inclusion of an interaction term between ethnicity and income in the empirical approach reveals the different effect of increasing income on electricity consumption for majority and minorities.

4.2 Background

4.2.1. Ethnicity-related issues in Vietnam

According to the Minority Rights Group International and based on 2009 census data, ethnic composition in Vietnam is primarily driven by the ethnic Viet or Kinh, which represent 85.7 percent of the total population. Besides, main minorities in the country are Tay, Thai, Muong, Khmer Krom, Hmong, Nung and Hoa, all at least representing one percent of the total population (Minority Rights Group International, n.d.). Up to 54 diverse ethnic groups are recognized, but ethnic minorities are strongly associated with poverty and considered to be traditionally backward by the authorities (Taylor, 2008). On the contrary, individuals belonging to the majority group hold both economic and political power over the country and count with a considerably advantageous position to access infrastructure or institutions (Imai et al., 2011). This also reflects in poverty rates, affecting minority groups over-proportionally compared to members of the majority group (Dang, 2012), but also shows in gaps related to life expectancy, health and nutrition outcomes as well as living standards (Pham Thai et al., 2010; Mbuya et al., 2019), or credit and consumption (Nguyen et al., 2020). Although considerable advances have been made in terms of poverty reduction, ethnic minorities still represent more than half of the total poor in Vietnam (Pham Thai et al., 2010). What is more, it is predominantly members of the majority ethnic group that benefited from economic growth while the share of ethnic minorities in the poor population has been constantly growing since the early 1990s.

The government has not been inactive in the matter and programs that support ethnic minorities in many domains were implemented (Dang, 2012), such as the Ethnic Minority Planning Framework (Power Design Center, 2005) and others. However, policy implementations that aimed to target minority groups did, in most cases, not achieve their aim. On that account, Pham Thai et al. (2010) stresses that uniform approaches and policies were not adequate in addressing poverty outcomes in Vietnam. Minority groups did not benefit from past governmental interventions to the same degree as individuals of the majority group did.

Consequently, Van de Walle and Gunewardena (2001) aim to investigate underlying mechanisms that determine ethic disparities in Vietnam. Minority groups are primarily located in remote, mountainous areas and suffer from insufficient public service provision and infrastructure, but it is their returns to productive factors that shape ethnic disparities. Minorities are often to be found in areas that are less productive, harder to cultivate, far away from major markets as well as non-agricultural labor markets and scarce in public service provision. Furthermore, members of minority groups face adverse effects of lower returns to education

and land, not only due to their remote location but also because of their socio-economic status as minority. In other words, the authors find evidence for ethnic divergences in Vietnam both because of a geographic factor of influence as well as socio-economic differences (Walle and Gunewardena, 2000). Singhal and Beck (2015) further focus attention on divergent origins of household income, with non-Kinh members being less likely to occupy well-paid, non-agricultural positions. Further, Mbuya et al. (2019) identify low coverage of government programs among ethnic groups as a reason for persisting differences in malnutrition between dominant and minority groups. Poor accessibility (topology of terrain dominated by hills and mountains) as well as language and culture barriers, which lower trust towards governmental institutions and thus reduce participation in programs, can all explain lower program coverage. Besides, remoteness contributes to reduced awareness and only limited knowledge diffusion about existing support programs (Dang, 2012). In addition, Thanh Tung Nguyen et al. (2020) point out difficulties in accessing formal credit, suggesting reduced financial integration of minority groups into the formal economy. In relation to energy, Feeny et al. (2021) identify that the over-reliance on agricultural income of the central and northern regions and the vulnerability of crop yields to extreme climate events, such as heat waves, reinforce the propensity towards energy poverty of mentioned regions, which are mainly populated by ethnic groups and dominated by agriculture livelihood.

Overall, the effects of poverty and low income on diverse human development outcomes have been vastly investigated (Anand and Sen, 1997; Arimah, 2004), and the link between ethnicity and poverty has attracted major scholarly attention (Awaworyi Churchill and Smyth, 2017; Gang et al., 2002). In line, Rafi et al. (2021) demonstrated how being energy poor impacts adversely upon multiple human capital outcomes, such as for example increased likelihood of severe diseases, malnutrition, lower school enrollment rates or human capital formation. Yet, only limited research has been conducted on the relationship between ethnicity and energy poverty. The next section shows that ethnic differences in electricity usage are not limited to Vietnam or to developing countries, but are rather widespread, which calls for consistent and research-based policy action.

4.2.2. Ethnicity and energy in other countries

In Australia, greater ethnic diversity at the neighborhood level, measured inter alia by the Herfindahl's index, is associated with increased energy poverty. In particular, the share of

household income spent on energy, as well as on the number of households spending more than 10 percent of their income on energy is greatly driven by ethnicity (Churchill and Smyth, 2020).

Similarly, in the US, great heterogeneity - racial, social and geographical - is observed for energy consumption and energy poverty patterns. African-American households are more prone to suffer from energy poverty (about one-third of them) than White or Asian households, with little observed variability over the past 25 years, except from an increasing share of energy poor White households between 1990 and 2015 (Q. Wang et al., 2021).

Energy poverty is also strongly prevalent among marginalized groups of the highly hierarchical society of India (Acharya and Sadath, 2017). Sedai, Nepal, and Jamasb (2021) capture specificities of energy poverty in a developing country context, where electricity usage depends on infrastructure such as access to a grid connection, but also on sufficient hours of power supplied each day. With data for 2005 and 2012, they find that marginalized groups (Hindu Schedule Caste/Schedule Tribe and Muslims) were less likely than dominant groups (Hindu of other casts) to have had an electricity connection in the past. However, these groups enjoyed a higher likelihood of electricity access in 2012 as well as a better reliability of electricity supply in terms of hours of electricity in a day. This renders evidence for a positive effect of public policies to reduce the social gap, with an accelerated access rate for disadvantaged groups during the period of an intense rural electrification program. However, the authors found that the welfare effect of better electricity, although positive, was not as strong for socially marginalized groups as it was for dominant groups. For example, better electricity is associated with increased income, larger assets, and moving out of poverty, but the observed effect was smaller than for the dominant group. This indicates that other barriers come into play concerning the reduction of welfare disparities across communities. This work is a step forward, pointing out that policies may not have the same effect on various socio-economic groups, but that existing social discrimination, here via the factors of religion and social status (caste), are not immutable.

In contrast, Pelz, Chindarkar, and Urpelainen (2021) find that for 6 North Indian states, inequalities of grid connection between Schedule Caste/Schedule Tribe and dominant groups remained constant from 2015 to 2018, while improvements in the quality of power (daily hours of power and monthly power outages) were lower among minorities than for the dominant group. This indicates that changes may not have been robust over time and region. For the case of Sri Lanka, higher multidimensional energy poverty rates are found among Tamils, who represent about 22 percent of the sample, than among Sinhalese, who form about 77 percent of

the sample population (Jayasinghe et al., 2021). In Nepal as well, ethnic fractionalization is associated with more severe energy poverty, which particularly affects low-caste individuals (Paudel, 2021). Significant disparities based on ethnicity are also found for South Africa where people of African descent face higher energy poverty levels than people of White or Asian/Indian descent (Ismail and Khembo, 2015).

Clearly, members of ethnic minorities routinely experience increased difficulties concerning the access and availability of electricity compared to majority groups. In order for policy makers to address these discriminations, the mechanisms through which ethnicity affects energy usage must be identified.

4.2.3. Channels of effects

Ethnicity can drive energy poverty via several channels. In the context of developed countries, where energy poverty is mostly characterized by high prices and low income, Churchill and Smyth (2020) provide a conceptual framework that identifies trust as a possible link between energy poverty and ethnic diversity. First, different environmental values may lead to diverging practices with respect to, for example, insulating homes. Second, ethnic minorities experience lower income due to rent-seeking behavior, which is more prevalent in ethnically diverse societies, or to the lack of opportunities and discrimination of on the job market. Third, lack of trust toward ethnic communities may result in weaker infrastructure or hinder the delivery of public goods, and thus result in higher energy prices. Finally, the authors show that trust at the neighborhood level is more important than income to explain higher energy poverty of more ethnically diverse neighborhoods. Similarly, long-lasting mistrust between policy makers and marginalized communities of Bedouin (Israel) and Roms (Romania) is also pointed out as a decisive factor in persisting energy poverty experienced by these communities, together with a lack of interest by policy makers and a general level of informality adopted by these communities (Teschner et al., 2020).

Likewise, Aklin et al. (2021) point that the social bias of the local implementation officers, mainly belonging to dominant group, against discriminated communities can explain the poor local implementation of the program. Indeed, they find that the proportion of the socially-discounted scheduled castes at village level in India decreases its probability of receiving support from the government-led electrification program RGGVY (Rajiv Gandhi Grameen Vidyutikaran Yojana).

In contrast to these human-based factors, some racial disparities seem to be rooted in more technical aspects of energy usage. Brockway et al. (2021) highlight how technical capacities of the grid to accommodate distributed solar energy generation can be a vector of racial discrimination. The authors show that the ability of local residents to adopt solar photovoltaic systems is more highly limited for racially and socially disadvantaged groups because of racial patterns in the grid construction. Therefore, the concerned households miss out on the opportunity offered by solar photovoltaic systems to lower electricity expenditures.

Further evidence from the United States shows that the higher energy burdens experienced by African-American households are mainly due to poor energy efficiency performance of their dwellings (Drehobl and Ross, 2016). In particular, African-American and Hispanic households are more vulnerable to fuel poverty because urban housing patterns reflect a persisting racial segregation in access to housing. The housing stock these households generally occupy is among the least energy efficient (Reames, 2016). Low energy efficiency coupled with lower income are thus causes of higher energy poverty for socially-disadvantaged groups in the US.

Therefore, racial differences in access to modern energy can clearly result from institutional frameworks and patterns (trust, representation, housing discrimination). However, no study has so far examined how income, the main driver of energy transition in developing countries, also contributes to ethnic divergences in relation to energy, on top of the already existing racial discriminations related to income.

4.2.4. Role of income in energy transition

Barnes et al. (2004, p. 22) summarizes the energy transition process stating three subsequent components, where income is decisive. Labeled as energy ladder theory (Leach, 1992; Smith, 1987), numerous studies investigate the decisive components of energy transition (Hanna and Oliva, 2015; Hosier and Dowd, 1987; Van der Kroon et al., 2013). Initially, energy demand is assumed to be met by firewood whereas in a next step, the supply constantly decreases due to deforestation activities, resulting in the substitution of wood by charcoal and kerosene. Lastly, the transition process profits from evolving energy markets, which provide newly industrialized and urban areas with electricity. Importantly, it is firstly urban zones that undergo mentioned transition process while rural areas often lack necessary income requirements to keep pace with the development process (Barnes et al., 2004, p.22). In a similar manner, Leach (1992) links the transition towards modern energy sources to key components of economic development and growth. With an accelerating development process and increasing urbanization, the average

income level is expected to increase and hence facilitate the transition towards modern energy sources, predominantly in urban areas. In practice, households tend to adopt fuel stacking, i.e. they continue using multiple fuel sources simultaneously as they climb the energy ladder, before they can fully transition (Choumert-Nkolo et al., 2019).

All in all, income remains an important factor that determines households' transition to electricity in developing countries. With rising available income, a surge in energy consumption follows (Chang, 2014; Sari and Soytas, 2007). McNeil and Letchert (2010) describe households energy transition process as a function of income following a S-shaped curve. Based on a theoretical framework developed by Gertler et al. (2016), it is income that drives energy consumption in developing countries. In their framework, the authors describe the impact of economic growth in the developing world, resulting in rising income, on shifting demand patterns for energy. Once households pass a critical point in available income, a process of energy-using acquisitions of durable consumer goods is initiated. These so-called "first-time owners of energy-using assets" (Gertler et al., 2016; p. 1398), are those that drive the energy transition process in the developing world.

Empirically, Pachauri and Jiang (2008) investigate the case of India and China, where they particularly highlight the persistence of major gaps in electricity access between urban and rural households for both country cases (see also Dong and Hao (2018) for China). On that account, it is still primarily rural areas that rely heavily on traditional energy sources and hence rarely consume commercial energy. Main reasons for a decelerated rural household energy transition are considerably lower income levels resulting in divergent spending preferences, higher energy prices compared to urban areas and inferior energy supply, which is often unreliable and unstable, as well as higher opportunity costs with regards to traditional fuel sources (Pachauri and Jiang, 2008).

Additionally Han et al., (2018) claim, investigating the case of China, that it is not only income that determines the pace of rural energy transition, but factors such as the size of the household, the educational level of the head as well as the number of available electronic appliances significantly determine the speed and mode of transition. This is in line with the assumption that the ability of transitioning from energy poverty to an electricity-dominated consumption is a multidimensional approach, with determinants that include household characteristics, habits and fuel choices as well as results of public policies for access to market and affordability (Ashagidigbi et al., 2020; Jayasinghe et al., 2021; Mendoza et al., 2019; Nussbaumer et al., 2013, 2012; Sadath and Acharya, 2017 and others). In line, Nguyen et al. (2019) therefore

conclude that economic development and the speed of transition are linked with each other, conditioning the nature of the energy transition process.

In the context of developing countries, income has been singled out as a key determinant for electricity consumption (Barnes et al., 2004; Leach, 1992, Gertler et al., 2016). Thus, to summarize the role of income for energy transition, a multidimensional perspective is adopted, which controls for household characteristics. In particular, income is identified as a channel through which other factors may influence energy poverty and electricity usage. Frictions based on ethnic association or household characteristics result in mitigated energy consumption and are channeled through income. Consequently, rising household income acts as a mediator for energy consumption, contingent on ethnicity and socioeconomic background. In this context, Cheng et al. (2021) confirm that income is a channel through which children that experienced famine are less likely to suffer from energy poverty later in life. Similarly, H. Wang et al. (2021) shows that vulnerability of agriculture-oriented communities to rain is the most important predictor of energy poverty for North India, because rain is associated with rural incomes. In relation to ethnicity, income appears to be a mediating factor of severe energy poverty experienced by low-caste individuals in Nepal (Paudel, 2021).

Therefore, income plausibly plays an important direct, but also indirect, role in energy usage. The next sections develop a methodological approach to capture the role of income as a mediator in the relation between ethnicity and energy usage, with empirical evidence from Vietnam.

4.3 Methodological framework

4.3.1. Base Model

Energy consumption and expenditure depend on a large number of determinants, among them income, energy prices and socio-economic factors. Given the evident ethnic differences in electricity usage in many countries, as well as the established divergence in living standards between Kinh and non-Kinh households in Vietnam (Baulch et al., 2012, 2008), a relationship between ethnicity and electricity appears to be consistent with the overall environment. However, additional socio-economic factors that are important concerning the energy behavior of households are likewise associated with ethnicity; primarily income, cultural habits and economic opportunities. In order to identify how ethnicity may be a barrier towards increased electricity usage and to energy transition, the analysis aims to include both direct and indirect

effects of ethnicity on electricity usage. Further, the study focuses solely on the indirect effect of income, as the main determinant of energy transition.

Energy consumption q can be expressed as a function of the ethnic group the household belongs to k, income x , electricity prices p, as well as a vector D of additional covariates that will be selected by a data-driven process.

$$q = f(k, x\ , p, D)$$

Therefore, belonging to an ethnic minority is expected to have a direct effect on energy consumption and expenditure. However, as ethnic minorities face disadvantages in other domains as well, it is clear that income is a function of ethnic belonging due to possible racial discriminations, such that:

$$x = x(k)$$

This in turn has implication for energy consumption. Thus, we pose that ethnic belonging also affects energy consumption and expenditure indirectly through the channel of income. Therefore, to examine the marginal effect of ethnicity on energy consumption and expenditure, both direct and indirect effects must be accounted for. Assuming that prices and the other factors D are exogenous to ethnicity, the total differentiation gives:

$$\frac{dq}{dk} = \frac{\partial q}{\partial k}|_{Fixed\ x} + \frac{\partial q}{\partial x}|_{Fixed\ k} \cdot \frac{\partial x}{\partial k}$$

Where $\frac{\partial q}{\partial k}|_{Fixed\ x}$ represents the direct effect of ethnicity on electricity while $\frac{\partial q}{\partial x}|_{Fixed\ k} \cdot \frac{\partial x}{\partial k}$ captures the indirect effect of ethnicities on electricity via the channel of income.

Applying a log transformation to income, the effect of ethnicity on electricity can be estimated empirically with the following model, where an interaction term between ethnicity and income serves to capture the income effect of ethnicity on electricity usage:

$$q = \beta_0 + \beta_1 k + \beta_2 lnx + \beta_3 k.lnx + \sum_{i=4}^{j} \beta_i D_i \qquad (1)$$

In order to grant robustness to the empirical analysis, two additional steps are performed prior the empirical estimation. First, the most relevant features to be used as controls in the estimation are identified and selected via the machine learning data-driven process LASSO. Second, to verify the assumption that income is a valid channel for the effect of ethnicity on electricity

consumption, a formal test of the mediation effect is employed. Additional details on these methodologies are provided below.

4.3.2. Feature selection with LASSO

Electricity usage results from a large set of determinants including socio-economic status of the households, electrical equipment, habits, or energy prices. However, not all determinants carry the same importance, and selecting the right set of controls is best achieved when context-specific information is accounted for. In such context of high dimensionality, the machine learning technique of *least absolute shrinkage and selection operator* (LASSO) provides a formal and efficient method for feature selection. The technique relies on the assumption of approximate sparsity of most models, which indicates that among the high number of regressors of a model, only some of them are relevant to capture the main features of the regression. This approach has already been used for the purpose of feature selection in predicting energy consumption in buildings (Jain et al., 2014), solar power generation (N. Tang et al., 2018) or human decision making towards energy usage (Das et al., 2019), among others. This form of feature selection was found to generate improved accuracy compared to conventional methods, henceforth, we employ it for the present research question.

Specifically, the LASSO approach is a regularization technique that aims to reduce the variance arising from the standard ordinary least squares (OLS) approach by introducing some bias in the coefficients (Lesmeister, 2015). To do so, the model not only minimizes the residual sum of squares (RSS), but also a penalty that consists of a tuning parameter lambda and the sum of absolute values of the coefficients β_i.

$$\min(RSS + \lambda \sum {\beta_i}^2)$$

In the minimization process, lambda increases and the coefficients lower, so that the prediction becomes less and less sensitive to inputs values, thus diminishing the problem of over fitting and high variance. The process of lowering coefficients does not affect all variables equally and the LASSO approach allows some coefficients to drop to zero, an advantage which enables to drop variables with little explanatory power and limit the model to its most meaningful features.

The right dose of penalty is determined by the value lambda, comprised between 0 and 1, and is typically chosen via cross-validation. However, machine learning techniques are often criticized for lacking theoretical background. To remedy this, Chernozhukov, Hansen, and Spindler (2016) propose an improvement to the LASSO technique that relies on theoretically

grounded data-driven choice of penalty level lambda, called the rigorous LASSO and implemented in the *rlasso* package. Furthermore, in order to remove the potential bias created by shrinking the coefficients to zero, a post-LASSO estimator is used, that applies ordinary least squares to the data after removing the regressors that were not selected by the LASSO (Belloni and Chernozhukov, 2013).

4.3.3. Mediation Effect

In this step, the validity of income as a mediation factor for ethnicity on electricity usage is determined. In the literature, the theoretical foundations of the mediation effect have been first summarized by Baron and Kenny (1986). Impacting upon the dependent variable, the mediator is assumed to alter the outcome through its influence on the independent variable. Central to the framework is the assumption of a mediation effect, which either reinforces or mitigates the strength of the effect of the independent variable on the dependent variable (Baron and Kenny, 1986). Accordingly, the authors define the role of the mediator as additional, explanatory component which investigates the underlying reason and mechanism of an observed effect (Baron and Kenny, 1986; Preacher and Hayes, 2004). In order to statistically test for and confirm the occurrence of a mediation effect, Barron and Kenny (1986) assume the absence of measurement errors concerning the mediator variable as well as the independence between independent variable and mediator. Further, the following three conditions need to hold:

- Firstly, when running a regression of the mediator on the independent variable, a significant effect of the independent variable on the mediator needs to be established. Formally, the first condition can be expressed as:

$$\ln \hat{x} = \alpha_1 + a\,k + \varepsilon_1$$

- Secondly, when running a regression of the dependent on the independent variable, a significant effect of the independent variable on the dependent variable needs to be established. The second condition is formally expressed as follows:

$$\hat{q} = \alpha_2 + b\,k + \varepsilon_2$$

- Lastly, when running a regression of the dependent variable on both the independent and the moderator variable, a significant effect of the meditator variable on the dependent variable needs to be established. Further, the effect of the independent on the dependent variable should be smaller in the last regression compared to the second model.

  Formally, the equation estimates:

$$\hat{q} = \alpha_3 + c\,k + d\,lnx + \varepsilon_3,$$

where α_i represents the intercept, a, b, c and d are parameters that measure the strength of effect and ε_i represents the error term.

In this context, Preacher and Hayes (2004) advocates for formal statistical testing in order to confirm the presence of a mediation effect. Although both the methodologies by Barron and Kenny (1986) as well as the Sobel test (1982) are frequently used in the literature, they highlight potential limitations, which is in line with MacKinnon et al. (2002). The authors stress the reduced statistical power of Barron and Kenny´s methodology compared to the Sobel test. The Sobel test (1982) constitutes a statistical approach, testing the null hypothesis of an indirect effect of the mediator of zero magnitude. However, it assumes normal distribution of the sample as well as a large sample size, which is not always applicable to the underlying data sample (Preacher and Hayes, 2004). Formally, the Sobel test can be expressed as:

$$s_{cd} = \sqrt{c^2 s_d^2 + d^2 s_c^2 + s_d^2 s_c^2}$$

where s_{cd} represents the standard error of the indirect effect.

Testing for normal distribution of the dependent variables' electricity consumption and electricity expenditures in the last 12 months, we observe a non-parametric distribution of the variables of interest. To account for mentioned limitations, a nonparametric bootstrap approach is additionally employed, which computes confidence intervals for the indirect effect and is not contingent on the distributional aspects of the underlying data structure (Preacher and Hayes, 2004; Tingley et al., 2014).

4.4 Empirical results

4.4.1. Data

The data employed in the analysis comes from Vietnam's Household Living Standards Survey 2010. The ethnic divide in living standards is has not been subject to substantial changes over the years, but rather consists of a long-term phenomenon in Vietnam. Furthermore, energy use of minorities continues to be dominated by traditional energy sources, such as coal and biomass (Nguyen et al., 2019). Therefore, making use of the latest publicly available cross-sectional data from 2010 is appropriate. This biennially conducted survey covers 69.360 households distributed over 3.133 communes in Vietnam and is representative for the country's entire population, at national, regional, provincial, urban and rural stage (ILOSTAT, 2017). The

design of the survey follows a three-stage stratified cluster based on enumeration areas defined for the 1999 Population and Housing Census (IOLSTAT, 2017). With the aim to evaluate differences in living standards and to improve the effectiveness of national programs to combat poverty and inequalities, interviews were conducted with the head of households as well as influential official at the communal level.

In respect to electricity use, the nationally representative results of the survey show that 97.2 percent of households use electricity for lighting and that this share remains high in rural areas (96.2 percent) as well as among households of the lowest income quintile (91.6 percent), representing a constant improvement since 2002 (Vietnam General Statistics Officce, 2010). Among households of the Northern Midlands and Mountain areas, characterized by remote communities, the share of households not using electricity lowers to 91.1 percent, and down to 83.8 percent in the specific Tây Bắc/ North West sub-region. Households which do not use electricity instead rely on electric generators or batteries as well as kerosene lamps for their lighting needs. The general large share of population relying on electricity indicates the good infrastructure, since more than 98.9 percent of communes are connected to the grid through the national electricity network (Vietnam General Statistics Officce, 2010).

The sample available consists of 82 percent Kinh households, Vietnam's largest and dominant ethnic group, while the remaining population is split between more than 50 ethnicities, each representing a small share of the population.[18] Preliminary summary statistics confirm that Kinh benefit from higher rates of connection to the national electric grid, as well as about twice the amounts of consumption and expenditure on electricity compared to Non-Kinh households.

Given the clear gap between urban and rural areas usage patterns of electricity, and that about 88 percent of non-Kinh live in rural areas, the analysis focuses on rural households only. The sample retained consists of 6373 rural households with a non-zero electric consumption, 18.8 percent of which belong to groups other than Kinh.

[18] Kinh, 82.2%; Tay, 3.5%; Thai, 2.5%; Chinese, 0.8%; Khmer, 1.6%; Muong, 1.4%; Nung, 1.6%; Hmong, 1.4%; Dao, 1.2%

Table 4. 1: Electricity usage and socio-economic indicators for rural households

	Non-Kinh	Kinh
Electricity		
Consumption, last month, per capita, in kWh	12.7	24.8
Expenditure, last month, per capita, in thousand Dong	9.9	22.0
Expenditure, last 12 months, per capita, in thousand Dong	110.7	254.6
Household characteristics		
Yearly real consumption, per capita, in thousand Dong	8266	14948
Household size	4.5	3.8
Illiteracy rate	0.20	0.32

n=6373

The gap in electricity usage between Kinh and non-Kinh appears is large, for example expenditure on electricity is more than twice higher for Kinh than for non-Kinh (Table 4.1). With respect to electricity consumption, non-Kinh clearly consume smaller amounts than Kinh (Figure 4.1).

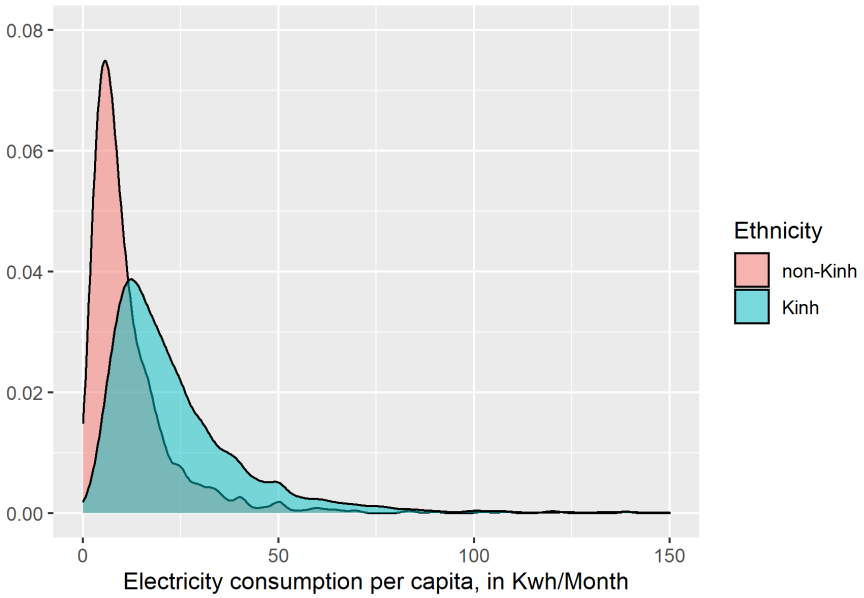

Figure 4. 1: Density of electricity consumption by ethnic groups, rural sample

4.4.2. Model selection

In order to select appropriate controls among the vast amount of information collected by the survey, LASSO regression is used (employing both rigorous LASSO and post-LASSO estimator, see section 3.1) for both outcomes 'electricity consumption in kWh per capita in the

last month', and 'electricity expenditure per capita over the last 12 months. Appendix A provides summary statistics for the 36 variables that are initially included in the LASSO regression.

Table 4.2 shows the variables identified by LASSO as contributing the most to the variance of each outcome. The models selected respectively 9 and 15 out of the 36 explanatory variables. Second, both models are found to have joint significance of the coefficients, as the null hypothesis, tested with a sup-score statistic (Chernozhukov et al., 2016), is rejected in both cases. The R-squared adjusted values reach 0.3 and 0.5 for each outcome respectively.

Table 4. 2: Features selected by LASSO technique

rLASSO coefficients	**Electricity consumption per capita (last month)** *Estimates*	**Electricity expenditure (last 12 months)** *Estimates*
Intercept	-77.814	-2853.75
Expenditure (log)	11.893	310.772
Household size	-2.208	101.176
Income: regular wage (1 member)	-2.6	x
Income: household production (1 member)	x	-100.674
Income: household production (2 members)	-2.82	-59.325
Income: own business (1 member)	x	87.365
Number of appliances	0.231	16.654
Fridge	9.437	452.127
Air Conditioner	20.308	1132.902
Washing machine	4.507	358.172
Computer	x	101.272
Electricity Price	-3.538	111.846
Poverty status	x	-15.988
Expenditure on kerosene	x	0.026
Size of the dwelling (squared meters)	x	1.079
Location in Midlands and Northern Mountainous Areas	x	-108.791
Residual standard error	*19.57*	*547.5*
Adjusted R-squared:	*0.3042*	*0.524*
Joint significance test	***	***

As expected, income is found to be a good predictor for electricity consumption and expenditure, as well as related sources to income (regular wage, business activity, own production). Further, the total number of appliances owned by the households, as well as ownership of large appliances (fridge, air conditioning, and washing machine), are found to be important predictors for electricity consumption and expenditure. Ownership of small appliances however is not found to have sufficient association with electricity consumption and expenditure. Among socio-economic variables, it is not surprising to find that household size is

selected by the LASSO approach as it has often been associated with energy consumption (Ozughalu and Ogwumike, 2019; Sharma et al., 2019; H. Wang et al., 2021). However, education, gender, age, marital status and other measures of family structure such as presence of young or older dependents do not appear to have sufficient predictive power to be retained in the selected models. Electricity prices have a sufficiently strong association with both consumption and expenditure to be included in the reduced models. Regarding areas with higher shares of ethnic minorities, being located in the central highlands is not selected by the model, while being located in the Midlands and Northern Mountainous areas is found to have sufficient explanatory power for electricity expenditure, but not for consumption.

The coefficients also adopt the expected signs, with higher electricity consumption and expenditure being associated with higher income, ownership of appliances as well as larger dwelling. At the same time, income generation from household production, poverty status and dwelling in the Midlands and Northern Mountainous areas are associated with lower energy consumption and/or expenditure. Electricity prices tend to reduce consumption while at the same time increase expenditure. Surprisingly, drawing income from a regular wage is associated with lower consumption. The effect of family size lowers consumption per capita but increases expenditure (measured at the household level).

4.4.3. Mediation through income

Next, a mediation analysis, as used in previous studies (Alesina and Zhuravskaya, 2011; Cheng et al., 2021; Churchill and Smyth, 2020), helps verify the validity of income as a channel of racial differences on both energy consumption and energy expenditures.

Consolidating statistical results of the mediation analysis, the effect of income as a significant channel on energy consumption can be confirmed. Assessing results that follow the approach by Baron and Kenny, (1986), we observe a positive impact of being a member of the majority ethnicity on income by a factor of 0.6 (Table 4.3, panel A). Besides, the effect of majority ethnicity on energy consumption is statistically significant, indicating an increased energy consumption by 12.05 kWh per month of the Kinh minority. When additionally including income into the third equation, the ethnicity term becomes insignificant, yet, a substantial increase in energy consumption with raising income can be observed. On that account, the effect of ethnicity on the dependent variable energy consumption is larger in the third equation than in the second, which is in line with Baron and Kenny (1986). When including control variables into our equations, the observed effects still hold, albeit decrease in magnitude. Following this,

we can observe a mediation effect of income through ethnicity in Figure 4.2. The slope for energy consumption for members of the majority ethnicity (Ethnicity = False) is steeper compared to their non-Kinh counterparts, indicating a faster uptake in energy consumption with raising income. This confirms the assumption of divergent energy consumption patterns among Vietnamese ethnicities.

Running a Sobel test yields insignificant results with a p-value of 2.98. However, the dependent variable energy consumption is not normally distributed, which is why we additionally employ a non-parametric bootstrap approach. The results confirm the mediating effect of income as a channel on energy consumption. With a total effect of 240.89 kWh, ethnicity significantly affects electricity consumption. Since the direct effect (Average Direct Effect) is insignificant, we assume full mediation. In other words, the effect of ethnicity on electricity consumption is fully explained by income, with a significant average causal mediation effect, or indirect effect, of 204.7 kWh. Further, if electricity consumption was only determined by income – conditional on ethnicity - household would experience a remarkable rise in electricity demand. However, it seems reasonable to assume that additional covariates impact upon electricity consumption. When including control variables into the model, the full mediation effect still holds but greatly decreases in magnitude. The indirect effect of income consequently results in an increased energy consumption of 39.27 kWh.

Table 4. 3: Mediation test for income and electricity consumption

PANEL A: Consumption

	Condition 1	**Condition 2**	**Condition 3**	**Condition 1 inc. control**	**Condition 2 inc. control**	**Condition 3 inc. control**
Coefficient	*Estimates*	*Estimates*	*Estimates*	*Estimates*	*Estimates*	*Estimates*
Intercept	8.87 *** (0.02)	12.74 *** (0.66)	-164.35 *** (4.33)	9.14 *** (0.02)	31.45 *** (1.04)	-80.63 *** (5.39)
Ethnicity Dummy	0.60 *** (0.02)	12.05 *** (0.73)	0.13 (0.71)	0.30 *** (0.01)	3.20 *** (0.68)	-0.46 (0.68)
Income			19.97 *** (0.48)			12.26 *** (0.58)
Control	No	No	No	Yes	Yes	Yes
Observations	6373	6373	6373	6373	6373	6373
R^2 adjusted	0.165	0.041	0.243	0.486	0.286	0.332

PANEL B: Expenditure

	Condition 1	**Condition 2**	**Condition 3**	**Condition 1 inc. control**	**Condition 2 inc. control**	**Condition 3 inc. control**
Coeffcient	*Estimates*	*Estimates*	*Estimates*	*Estimates*	*Estimates*	*Estimates*
Intercept	8.87 *** (0.02)	441.34 *** (22.32)	-5432.21 *** (147.33)	9.26 *** (0.02)	-32.82 (34.87)	-2809.35 *** (176.54)
Ethnicity Dummy	0.60 *** (0.02)	466.88 *** (24.77)	71.42 ** (24.21)	0.19 *** (0.01)	127.44 *** (22.32)	70.49 ** (22.17)
Income			662.46 *** (16.46)			299.81 *** (18.70)
Controls	No	No	No	Yes	Yes	Yes
Observations	6373	6373	6373	6373	6373	6373
R^2 adjusted	0.165	0.053	0.245	0.600	0.516	0.535

Notes: 1, Statistical significance: * $p<0.05$; ** $p<0.01$; *** $p<0.001$.
2, Standard errors in the brackets.

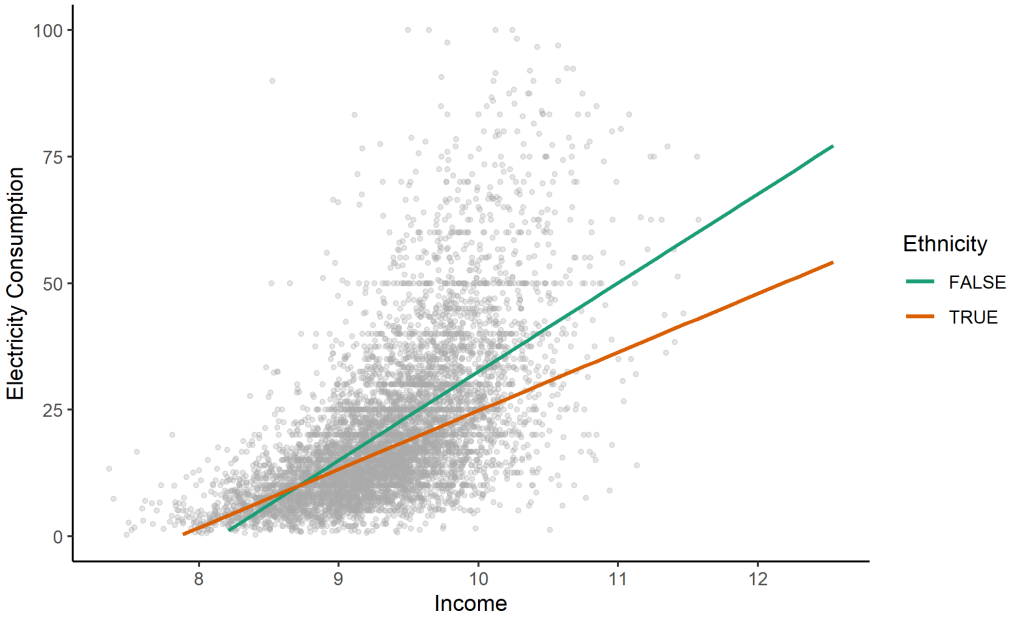

Figure 4. 2: Income and electricity consumption, by ethnicity

Similarly, income is a channel that impacts upon energy expenditures. Evaluating established conditions by Baron and Kenny (1986), we observe a small but positive effect on income if individuals belong to the majority group (Table 4.3, panel B). Likewise, the impact of ethnicity on energy expenditures is positive for members of the majority group by a factor of 466.88 (measured in thousand Dong). Expenditures are also positively associated with both income and ethnicity. Controlling for additional covariates, the effects hold but decrease in magnitude. Figure 4.3 shows different expenditure patterns comparing majority and minority ethnic groups in Vietnam. With increasing income, households of the majority group tend to raise their energy expenditures more extensively than comparable households of the minority group. Consequently, energy expenditures follow as similar trend as energy consumption with substantial divergences between ethnic groups, channeled through income.

Based on the non-normal distribution of the dependent variable, the Sobel test once again yields insignificant results with a p-value of 7.59. Nevertheless, non-parametric bootstrapping establishes a partial mediation effect of income on energy expenditures. A significant total effect confirms a relationship between ethnicity and energy expenditures. Accordingly, a significant indirect effect further underlines the mediating role of income. A significant direct effect indicates that although income acts as a mediator, a direct relationship between ethnicity

and electricity expenditures can still be established. As a result, in the case of energy expenditures, income acts as a partial mediator.

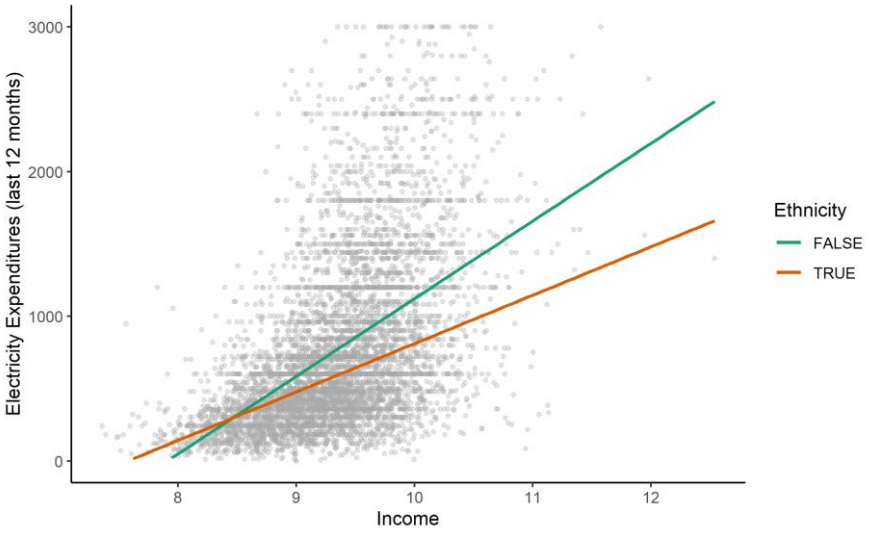

Figure 4. 3: Income and electricity expenditure, by ethnicity

4.4.4. Regression results

This section reports the results of estimating the model *(1)*, which includes an interaction term between ethnicity and income to reflect the indirect effect of ethnicity on electricity through income, and confirms findings from section 4.3. This section proposes three sets of regressions with increasing numbers of control variables. In the first set of regressions, which takes ethnicity as the only regressor of energy consumption and expenditure, the coefficients on Kinh (value of 1 for belonging to the majority group, 0 to one of the ethnic minorities) are positive, indicating that belonging to the majority group increases electricity use (Table 4.4).

Next, the dual role of income (direct and indirect) is added to regression sets 2 and 3, which exclude and include other controls respectively. The interaction term is significant, confirming that the effect of income on electricity differs whether the household belongs to the majority or to a minority ethnic group. The effect of income on electricity for non-Kinh is captured with the positive coefficient on expenditure (14.4 without controls and 8.3 with controls), while the effect of income for Kinh is obtained by taking the coefficient on expenditure plus the interaction term. Given that the coefficient on the interaction term is positive (7 without controls and 4.9 with controls), the effect of income is larger for Kinh than for non-Kinh. An extra unit of income leads to a larger increase of electricity consumption for members of the majority

group than for members of the ethnic groups. The findings give substance to Baulch et al. (2012), who indicate that socio-economic endowments as well as returns to endowments did improve faster for Kinh households over the period between 1993 until 2004. Similar observations are made for electricity expenditures, which confirm that ethnicity plays a role on electricity usage both directly and indirectly via a differentiated income effect. The coefficients on all other control variables adopt the expected signs.[19] Our results are in line with previous findings, where members of the majority group show improved outcomes with respect to the overall standard of living, life expectancy, health and nutrition (Pham Thai et al., 2010; Mbuya et al., 2019) as well as to access to credits or purchasing power (Nguyen et al., 2020) compared to non-Kinh households. As a result, Kinh members benefit to a larger degree from available resources, which also reflects in electricity usage. Summarized by Singhal and Beck (2015), indications for welfare convergence in Vietnam are thus not present between different ethnic groups.

[19] Under this configuration, the negative coefficient on Kinh represents the effect of being Kinh when income is 0, which is not very interpretable as no household has zero income.

Table 4. 4: Indirect income effects of ethnicity on electricity consumption

	Electricity consumption per capita (last month)			Electricity expenditure (last 12 months)		
	Set 1 (Ethnicity only)	Set 2 (Ethnicity and income)	Set 3 (All controls)	Set 1 (Ethnicity only)	Set 2 (Ethnicity and income)	Set 3 (All controls)
Coefficient	*Estimates*	*Estimates*	*Estimates*	*Estimates*	*Estimates*	*Estimates*
Intercept	12.74 ***	-114.74 ***	-45.80 ***	441.34 ***	-2829.75 ***	-2184.53 ***
	(0.66)	(9.59)	(9.77)	(22.32)	(325.12)	(297.46)
Kinh	12.05 ***	-62.70 ***	-44.88 ***	466.88 ***	-3224.83 ***	-707.92 *
	(0.73)	(10.87)	(10.42)	(24.77)	(368.45)	(299.15)
Expenditure (log, per capita)		14.38 ***	8.35 ***		368.94 ***	229.97 ***
		(1.08)	(1.08)		(36.6)	(32.65)
Kinh*Expenditure (log, per capita)		**6.99 *****	**4.94 *****		**366.84 *****	**86.46 ** **
		(1.21)	**(1.16)**		**(40.92)**	**(33.14)**
Household size			-2.26 ***			94.82 ***
			(0.18)			(5.46)
Regular wage (1 pers.)			-3.06 ***			na
			(0.5)			na
Household production (1 pers.)			na			-89.67 ***
			na			(22.28)
Household production (2 pers.)			-2.35 ***			-33.87
			(0.54)			(19)
Business (1 pers.)			na			82.20 ***
			na			(15.53)
Number of appliances			0.2			17.10 ***
			(0.13)			(3.7)
Fridge			9.40 ***			454.65 ***
			(0.67)			(19.15)
Air conditioning			22.30 ***			1106.30 ***
			(2.1)			(60.04)
Washing machine			5.25 ***			376.36 ***
			(1.12)			(32.24)
Computer			na			113.65 ***
			na			(28.47)
Electricity prices			-3.53 ***			109.90 ***

			(0.3)			(8.6)
Poverty status			na			-17.33
			na			(22.64)
Kerosene expenditures			na			0.02 ***
			na			(0)
Size of dwelling, in sq. meters			na			1.06 ***
			na			(0.21)
Northern			na			-90.97 ***
			na			(20.98)
Observations	6373	6373	6373	6373	6373	6373
R^2 adjusted	0.041	0.247	0.334	0.053	0.254	0.535

Note: 1, Statistical significance: * p<0.05; ** p<0.01; *** p<0.001.
2, Standard errors in the brackets.

4.5 Discussion of results

4.5.1. Lifestyle and basic needs for electricity

This section aims to clarify the lower increase of electricity consumption by non-Kinh groups than by Kinh when income rises, and identify potential limits to upward consumption.

It is sometimes argued that members of ethnic groups adopt lifestyles that are less energy-intensive compared to lifestyles of other groups dominated by urban culture, suggesting that smaller amounts of modern energy are sufficient to ensure satisfaction of ethnic groups' basic needs. This could constitute a reason ethnic members do not increase their electricity consumption as much as their counterparts from majority groups when income rises. To verify this, we can exploit the information collected in the survey as to whether households subjectively feel they face shortage of electricity to satisfy their basic needs. 'Basic needs' are not defined in the questionnaire, and neither are the related quantities of electricity necessary to meet those needs. Households rather report their own appreciation, in relation to their actual electricity consumption, reflecting their vision of lifestyle and basic needs. About the same proportion of Kinh (24.9 percent) and non-Kinh (25.9 percent) report that their electricity consumption is insufficient. Figure 4.4 further shows that among households who report sufficient electricity consumption (electricity shortage = 0), consumption levels are lower for non-Kinh than for Kinh households. This suggests that most non-Kinh are able to fill their basic needs with lower amounts of electricity than Kinh, indicating different basic needs between ethnic groups. A similar pattern is observed among households who report experiencing electricity shortage (electricity shortage =1), where Kinh consider electricity insufficient at higher levels of electricity consumption than non-Kinh. The perception of shortage is thus

linked to ethnic affiliation, which indicates diverging basic energy needs. This, in turn, is a possible reason for a decelerated uptake in electricity usage rates for non-Kinh households, for given income levels.

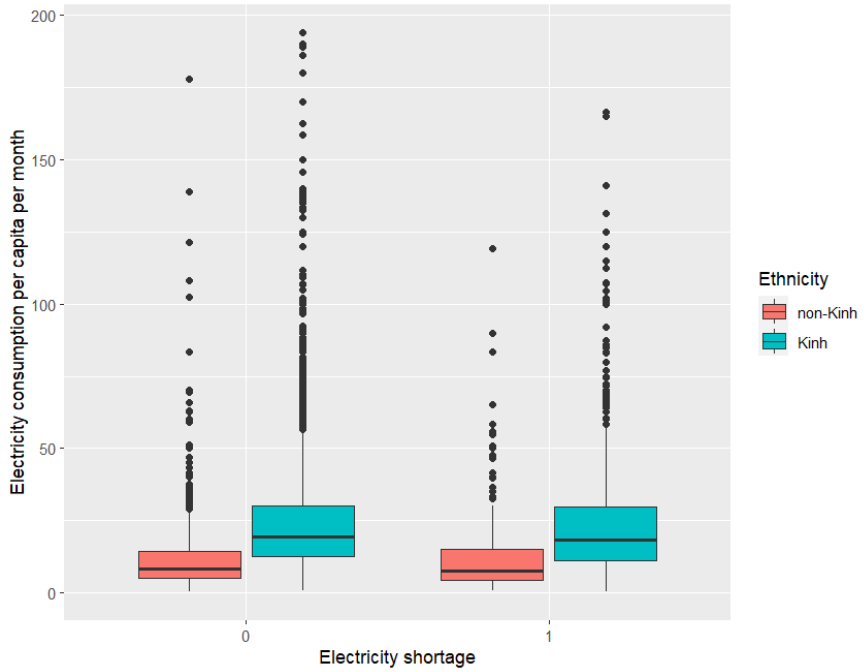

Figure 4. 4: Electricity consumption and electricity sufficiency/shortage, by ethnicity

4.5.2. Appliances

Electricity cannot be used without appliances to transform it into useful service. Therefore, electrical equipment of a household, such as light bulbs, electric fans, televisions, fridge or washing machine, is a key factor that determines the ability of households to increase their electricity consumption and define electricity consumption patterns. Access to markets to purchase appliances is thus critical for households who want to access the services offered by modern electricity (watching television, storing food in a fridge…) and is thus more challenging for households living in rural and remote communities. This section shortly examines the appliance stock of Kinh and non-Kinh. Non-Kin households possess on average smaller number of appliances (3.1) than Kinh households (4.9). As seen in Figure 4.5, this pattern is visible at all levels of electricity consumption, expect very high levels. Therefore, even when Kinh and non-Kinh households have the ability to consume similar levels of electricity, the non-Kinh houses are less well-equipped than houses of the Kinh majority. At this stage the interpretation requires caution. A more difficult access to markets to purchase appliances, a lower desire to

purchase electric appliances and pursue modern energy-intensive lifestyle, or other priorities could all be reasons for members of ethnic communities to own fewer appliances than their Kinh counterparts.

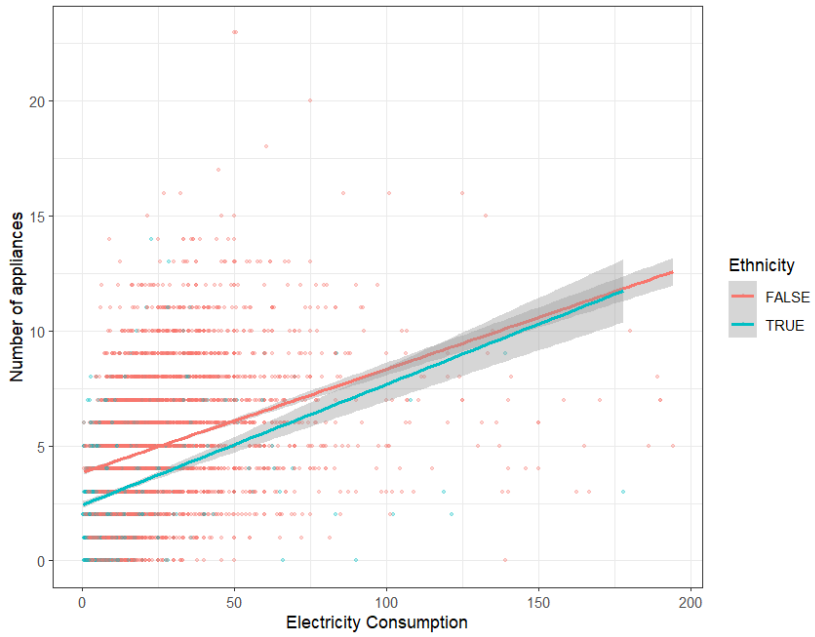

Figure 4. 5: Appliance ownership and electricity consumption, by ethnicity

4.5.3. Robustness checks

4.5.3.1 Urban households

Among the non-Kinh households of the survey, 12 percent lived in urban area. The same analysis is performed for the 2620 urban households of the initial survey to see if similar patterns of income effect are observed for them. First, the LASSO analysis reveals that the variables identified as main features of the regression are close to those for the rural sample, shown in Table 4.5.

Table 4. 5: Features selected by LASSO technique for urban households

rLASSO coefficients	**Electricity consumption per capita (last month)** *Estimates*	**Electricity expenditure (last 12 months)** *Estimates*
(Intercept)	-151.173	-10545.1
Expenditure (log)	21.028	1026.72
Household size	-4.771	350.73
Income source: household production (1 member)	-1.917	-162.32
Income source: household production (2 members)	0.025	x
Number of appliances	0.869	64.72
Fridge	4.702	-17.36
Air Conditioner	26.427	1528.79
Washing machine	2.81	166.8
Computer	x	284.6
Electricity Price	-6.028	369.94
Location in Midlands and Northern Mountainous Areas	x	-356.03
Location in Central Region	-5.715	-484.55
Residual standard error	30.17	1511
Adjusted R-squared:	0.4186	0.5101
Joint significance test	***	***

The models are significant and retain sufficiently explanatory power (adjusted R-squared at 0.42 and 0.51 for consumption and expenditure respectively, as well as joint significance of the coefficients). For urban households, being located in the central region is sufficiently, but negatively, associated with energy usage to be retained by the model. Poverty status, expenditures on kerosene, and size of the dwelling are not predictors strong enough to be kept. Among income sources, only those that relate to household own production are selected, while obtaining a regular salary from wage or running its own business is not deemed sufficiently associated with electricity usage. The same appliance variables are selected as for the rural households, at the exception of having a computer.

Second, the mediation analysis generates different results as for rural households, both in terms of magnitude and significance. With increasing income, electricity consumption increases by 37.55 kWh for urban households when ethnicity only is included as control. Taking into consideration selected control variables, the effect holds but decreases in size (23.75 kWh). Majority ethnicity appears to have a significant, positive impact on electricity consumption. Yet, the effect disappears when the control variables are included. Thus, in an urban context, ethnic belonging does not seem to impact upon electricity consumption. Consequently,

proposed conditions by Baron and Kenny (1986) only hold when we consider electricity consumption without additional control variables.

As previously, the Sobel test renders insignificant results (p-value of 1.40), which can be attributed to the non-normal distribution of our sample. Although the bootstrapping approach indicates full mediation of income on electricity consumption, the results become insignificant when selected control variables are included.

With regards to electricity expenditures, a similar trend is observed where ethnicity is not a significant determinant for electricity expenditures in an urban context. The conditions stated by Baron and Kenny (1986) do not hold, which is further confirmed by an insignificant Sobel test (p-value of 8.8) and by the bootstrapping approach, confirming the insignificant indirect effect of income neither on electricity expenditures nor consumption based on ethnicity.

Finally, the regression models with interaction between ethnicity and income confirm the absence of significant direct and indirect role of ethnicity on electricity consumption and expenditure.

Table 4. 6: Indirect income effect on electricity consumption and expenditure, urban households

	Electricity consumption per capita (last month)			Electricity expenditure (last 12 months)		
	Set 1 (Ethnicity only)	**Set 2 (Ethnicity and income)**	**Set 3 (All controls)**	**Set 1 (Ethnicity only)**	**Set 2 (Ethnicity and income)**	**Set 3 (All controls)**
Coefficient	*Estimates*	*Estimates*	*Estimates*	*Estimates*	*Estimates*	*Estimates*
Intercept	34.12 ***	-286.37 ***	-170.39 ***	1473.08***	-12331.52 ***	-12986.9 ***
	(2.95)	(34.03)	(33.46)	(211.92)	(2776.43)	(2633.5)
Kinh	12.92 ***	-43.26	-9.34	748.02 ***	-3742.8	1545
	(3.06)	(35.62)	(34)	(220.1)	(2905.56)	(2666.41)
Expenditure (log, per capita)		33.75 ***	23.40 ***		1453.67 ***	1287.28 ***
		(3.58)	(3.49)		(291.65)	(273.85)
Kinh*Expenditure (log, per capita)		**4.13**	**0.38**		**386.19**	**-196.31**
		(3.73)	**(3.55)**		**(304.05)**	**(278.49)**
Household size			-4.04 ***			423.49 ***
			(0.47)			(37.21)
Household production (1 prs.)			-1.52			-226.27
			(2.01)			(121.62)
Household production (2 prs.)			-1.11			na
			(2.46)			na
Number of appliances			0.60 *			67.78 ***
			(0.25)			(20.03)
Fridge			7.04 ***			80.89
			(1.53)			(120.93)
Air conditioning			26.59 ***			1719.43 ***
			(2)			(157.8)
Washing machine			2.05			109.66
			(1.6)			(126.65)
Electricity prices			-6.84 ***			340.54 ***
			(0.88)			(69.7)
Central			-5.39 *			-457.78 *
			(2.35)			(186.4)
Northern			na			-454.52 **
			na			(156.14)
Computer			na			217.64
			na			(124.11)
Observations	2620	2620	2620	2620	2620	2620
R^2 adjusted	0.006	0.349	0.432	0.004	0.159	0.316

Note: 1, Statistical significance: * p<0.05; ** p<0.01; *** p<0.001. 2, Standard errors in the brackets.

The resulting coefficients, shown in Table 4.6, are mostly of the expected signs, at the exception of a negative association between having a computer and electricity consumption. The estimates on ethnic minority and the interaction term become insignificant for the urban sample when income, alone or with the other control variables, is included (sets 2 and 3). The income effect of ethnicity on electricity usage thus disappears for the urban dwellers. For them, it appears that an income increase has the same effect on electricity usage, whether the household belongs to the majority group or to an ethnic group. Therefore, we can deduct that for members of ethnic group who move to urban centers, the opportunities to increase electricity usage offered by a higher income are the same as the members of the majority group.

4.5.3.2 Racial distinction

Given the great plurality of ethnic groups in Vietnam, other definitions of majority and minority ethnic groups can be adopted to distinguish how racial profile affects electricity use. Ethnic groups can be grouped under larger categories. The Kinh together with the Muong, Tho, Chut form the major Vietic ethnic group, which represents 83.7 percent of the total sample. Thus, while all Kinh are Vietic, most but not all Vietic are Kinh. The above analysis was conducted using the Vietic/Other distinction as racial divide, instead of Kinh/non-Kinh, but produces similar results that are not reported here.

4.6 Conclusion

It is central to human development and environmental sustainability that poor and rural households increase their electricity usage, as replacement of polluting, unsafe and inefficient kerosene or biomass fuel sources. The modern Vietnam economy experienced relative dynamic growth, accompanied by a reduction in poverty rates and considerable improvement in living standards. Particular effort has been made to expand infrastructure, which resulted in almost universal access to the electric grid for rural households. Nevertheless, ethnical minorities as well as rural and remote communities still suffer from lower access rates to electricity compare to their majority counterparts. What is more, even connected to grid electricity, members of ethnic minorities continue to consume prominently lower amounts of electricity compared to members of the Kinh majority. Ethnicity has known effects of discrimination and marginalization on access to economic opportunities, human development and living standards. The present study, in line with previous research, helps clarify the link between ethnicity and energy use.

In particular, the research questioned the role of ethnic belonging on the most important factor that drives household transition to electricity: rising income. It is well established that members of socially disadvantaged groups often experience lower income levels, but the present analysis further investigated how rising incomes affect both members of the majority and the minority groups differently in relation to their electricity use.

This study proposes a robust methodological framework where feature selection is operated with the least absolute shrinkage and selection operator (LASSO) and where a formal analysis of mediation is conducted to establish whether income also acts as a mediation factor for ethnicity on electricity.

Based on data for rural households from the Vietnam's 2010 Household Living Standards Survey, the results highlight the positive impact of being a member of the majority group on electricity consumption. According to the non-parametric bootstrap approach, complementing the methodology suggested by Baron and Kenny (1986), we find significant evidence that income fully explains the effect of ethnicity on electricity consumption, while it only acts as partial mediation effect in the context of electricity expenditures. The empirical approach then confirms a different relationship between income and electricity for Kinh and non-Kinh households. An extra unit of income results in a larger increase of electricity consumption for members of the majority group than for non-Kinh households. Although income remains the most important factor for households to increase electricity usage, higher income levels do not translate into the same opportunities for Kinh and non-Kinh to increase electricity consumption.

The analytical application of our approach to urban households in Vietnam (where 12 percent of the non-Kinh live) yields contrary results, where the indirect effect of ethnicity on income is no longer statistically significant, indicating that members of ethnic groups located in urban areas are no longer exposed to ethnic adversity with regards to electricity consumption.

Consequently, the present study suggests that policies attempting to resolve racial differences in electricity consumption should look beyond infrastructure and grid connection, but rather identify additional barriers that ethnic minorities face with respect to energy use. As the case of Vietnam has shown, basic needs of minority and majority households appear to be characterized by disparities, resulting in different electricity usage. Income has been identified as a main channel, through which ethnicity affects the use of electricity. However, income is additionally related to socioeconomic characteristics, which differ between urban and rural households. Lifestyle patterns of non-Kinh households appear to diverge from Kinh households, and this

requires further investigation of the link between ethnic divide, living standards and economic growth. Besides, Kinh and non-Kinh households might not only diverge in electricity consumption but also in the consumption of other services. Non-Kinh households seem to diversify increasing income into other goods and services than electricity. The effect is strongest for rural and remote non-Kinh households. Consequently, inquiries emerge about why and how those households choose to spend or save extra income, demanding for supplemental research.

Of particular importance we consider the implementation of tailor-made policy programs, which so far have not met minorities energy needs to a satisfying degree. We suggest a holistic approach of governmental programs particularly for remote, non-Kinh households. It is immanent that those programs take into consideration the remote character of these households, which has been previously identified as a distinct component in electricity use. Furthermore, considerations raised by other scholars, such as the lack of familiarity with electricity (Ranganathan, 1993; Bernard, 2012), the potential impact of neighbor's behavior (Bernard and Torero, 2015) as well as a possible lack of knowledge about the benefits of modern electricity use should all be taken into consideration when designing future governmental programs. New insight can also be gained through the economic assessment of opportunity costs between traditional energy sources in remote areas and modern electricity. Therefore, specific policy recommendations, based on obtained research results, entail the supplemental collection of qualitative data in remote households of the country.

CHAPTER 5

PROGRESSIVITY OF INCREASING BLOCK PRICES FOR ELECTRICITY CONSUMPTION IN INDIA: AN ANALYTICAL FRAMEWORK OF TWO-STAGE BUDGETING WITH INCOME EFFECT[20]

Abstract

Under the existing increasing block price (IBP) system that governs the Indian residential electricity sector, most of the new, small and rural users who recently joined the utilities' customer base due to progress in rural electrification face smaller marginal rates than the urban customers. Current policy changes encourage price increase and uniformizing to improve the weak financial situation of the electricity sector. This could lift the lowest rates and reduce the progressivity of marginal rates. Using NSSO microdata from 2012 and published electricity tariffs, we propose an ad hoc analytical framework to include the IBP structure in a modified LA-AIDS model and compute Marshallian, Hicksian and unconditional demand elasticities with respect to marginal price and income. Accounting for the IBP structure, the resulting income and own price response are 0.57 and -0.39 respectively, which can be compared to 0.62 and -0.33 when flat average prices are used. Rural consumers are more sensitive to price change but less sensitive to income change, which support implementation of differentiated tariffs between rural and urban sectors. An analysis of the current IBP structure endorses increasing progressivity of the rates between 50 and 200 monthly units, with the design of narrower and/or higher blocks. This would be effective in revenue collection, reduce subsidy leakage to rich households and maintain protection of vulnerable households. Price increases applied uniformly and/or compensating poor users by direct cash transfer tend to benefit urban and large users disproportionally.

[20] The authors of this essay are Lucie Maruejols and Xiaohua Yu. All authors contributed to conceptualize the research idea. Lucie Maruejols conducted the data analysis, results discussion and writing the paper. Xiaohua Yu provided critical feedback and revisions. We are grateful to Prof. Bernhard Brümmer and anonymous reviewers for comments on an earlier draft.

5.1. Introduction

Despite recent improvement, the Indian electricity sector is characterized by underdevelopment, inadequate supply and weak financial situation of the utilities distribution, in part because residential electricity prices have been kept artificially low and revenue collection have fallen far from cost recovery levels (IEA, 2021). Policy changes in the last 20 years (Ministry of Power Government of India, 2020, 2018, 2003) press for an increase of residential rates in order to recover the financial losses accumulated in the 1990's and provide right economic incentives to the sector (Acharya and Sadath, 2017; Bhattacharyya and Ganguly, 2017). The most recent measure (2020) proposes to compensate the price increases by direct cash transfer to vulnerable users. This could potentially reduce the price discrimination at user level inherent to the increasing block prices (IBP) system widely adopted in India and other developing countries, which reserves low unit prices for low levels of consumption and charges higher rates for larger consumption amounts. The resulting flatter rates could correct the flaws of the IBP, which is known to fail at redistributing income towards the most vulnerable customers, but to benefit more the wealthiest users (Mayer et al., 2015).

On the contrary, recent studies suggest maintaining progressive tariffs, because of heterogeneities in household consumption patterns across income groups and rural-urban sectors (Gundimeda and Köhlin, 2008; Chindarkar and Goyal, 2019). Sedai et al. (2021b) observe different benefits of increased hours of electricity supply between the rural and urban households and support adjusting progressive tariffs according to data-driven recommendations rather to uniform nationwide principles.

Therefore, it remains unclear whether the electricity rates' progressivity should be flattened, maintained or increased. In particular, the specificities of the block tariffs in relation to the growing customer base of rural customers, who face substantially lower marginal rates due to their low consumption levels and differentiated tariff schedules (i.e. the set of marginal prices at various consumption quantities for a type of customer), has been ignored from previous estimations. Using expenditure data from the 2012 National Sample Surveys (NSS)[21], this essay aims to close this gap by addressing the role of progressive rates in the price and income

[21] Household Consumer Expenditure: NSS 68th Round, NSSO, Ministry of Statistics & Programme Implementation, Government of India, http://microdata.gov.in/nada43/index.php/catalog/1

responses estimations, and find opportunities for modulating the existing IBP to increase revenues while protecting vulnerable users.

Elasticities are derived using the linearized version (Buse, 1994) of the almost ideal demand system (Deaton and Muellbauer, 1980), or LA-AIDS model (Bhuvandas and Gundimeda, 2020; Hasiner and Yu, 2016; Lin and Liu, 2013; Ngui et al., 2011). To account for the specificity of block prices, we propose a simple adjustment to the LA-AIDS whereby the *ex-post* average price is replaced by the marginal rate paid by each consumer, collected from published rate schedules in each jurisdiction, and used together with the rate structure premium approach of Taylor (1975) and Nordin (1976). For comparison and robustness, elasticities are also calculated according to a standard approach where price is the survey information *ex-post* average price. Response's heterogeneity is highlighted across the urban-rural sectors and across various volumes of consumption, which is more relevant to the tariff design than groups based on income.

Non-linearity of the price has been found to affect results of elasticities, but to be too small to matter (Henson, 1984), albeit for developed countries. In the context of a low-income country, energy poverty and affordability are major policy concerns (Jain et al., 2018; Wang, Maruejols, and Yu, 2021), and families spend larger shares of their budget on energy expenditures. This study is among the firsts which attempt to quantify the income effect bias resulting from smaller rates (Sun and Lin, 2013; Khanna et al., 2016).

In the following, Section 2 offers background information. The model is detailed in Section 3, followed by the data used in Section 4, and the results of the analysis in Section 5.

5.2. Background

5.2.1. Major trends in residential electric sector and customer heterogeneity

Huge investment and better recovery of costs are necessary to meet the bulging electricity demand of the developing Indian economy (Buckley, 2015; IEA, 2021). Consumption of electricity by the domestic sector more than doubled between 2007-2008 and 2016-17[22], while it is projected to increase by almost 5 percent in the coming decades, in part driven by newly

[22] Report: Energy statistics, 2018 (twenty fifth issue) Central Statistics Office Ministry Of Statistics And Programme Implementation Government Of India New Delhi, accessed 14.10.2019.

electrified households of the rural areas (IEA, 2021; Parikh and Parikh, 2011). Indeed, there has been undeniable progress in recent years to provide electricity access to rural dwellers, with programs such as DDUGJY (village-level electrification) or Saubhagya (household electrification), such that in 2020 it was estimated that only 2.4 percent of Indian households remained unelectrified (Agarwal et al., 2020). This brings many new customers to the utilities: close to 30,000 new villages have been connected between 2014 and 2019.[23] Between October 2017 and March 2019, more than 26 million new households received electricity connection under the Saubhagya scheme, most of them in the rural and underdeveloped states of Uttar Pradesh, Bihar, and Odisha (Statista and Ministry of Power (India), 2021). Therefore, to meet the growing demand with sufficient provision and an efficient distribution, the utilities require a price structure which achieves both effectiveness (collecting additional revenue) and equity (protecting vulnerable customers, especially new rural users). However, the rate increase intended by the policy reforms since 2003 was politically difficult to implement in the face of poverty and poor provision. Average residential prices have even fallen from 4.36 INR/kWh in 2005 to 3.7 INR/kWh in 2012 (Chindarkar and Goyal, 2019; Mayer et al., 2015) and the gap between costs and revenue has not been reduced between 2003 and 2011 (Khurana and Banerjee, 2015).

At the same time, revenue loss also results from extensive electricity theft, about a fifth of the total electricity production in India. Electricity theft is mostly driven by socio-economic determinants (Razavi and Fleury, 2019), attitude of electricity customers and utility employees (Saini, 2017) and their collusion (Sharma et al., 2016), as well as social norms that do not condemn electricity theft. Increasing prices is sometimes not implemented for fear that it would lead to higher theft, as in some countries (Jamil and Ahmad, 2014), although the effect is found insignificant in India (Gaur and Gupta, 2016). However Gaur and Gupta (2016) note that good management and governance of utilities, as well as power tariffs that are designed according to the regional socio-economic conditions of the users, seem more important than socio-economic conditions themselves to prevent theft.

An efficient rate structure must thus account for the particularities of the new rural customer base. First, rural consumption is much lower than urban consumption at same income levels (Bhattacharyya, 2006). Electricity is used mostly for lighting among poorer households, whereas high levels of electricity use indicate high appliance use, especially color television,

[23] https://www.statista.com/statistics/1232956/india-number-of-electrified-villages/ , accessed 07.08.2021

refrigerator, and washing machine (Chunekar and Sreenivas, 2019; IEA, 2021), which are mostly adopted by urban households (IEA, 2021; Yawale et al., 2021). In particular, the fast adoption of air conditioning units (AC) by urban households (Osunmuyiwa et al., 2020), which and are a major driver of electricity usage (Singh et al., 2018) but remain mostly absent from rural areas (Khosla et al., 2021; Kumar et al., 2021), could reinforce the gap between urban and rural consumption patterns. Differences in the energy efficiency performance of the appliances also contribute to the urban-rural gap (Murthy et al., 2001).

Second, important differences rest in the choice and availability of energy sources. Similar shares of urban and rural households rely on kerosene for lighting, but the share of rural households using electricity for lighting is smaller. Rural consumption patterns are also marked by high reliance on biomass and on non-commercial energy such as firewood, dung cake, crop residue and agricultural waste (Yawale et al., 2021). This can be explained by lower provision, lower quality and lack of reliability of rural electric supply which is endemic in most Indian states (Kennedy et al., 2019, Sedai et al., 2021). The physical access to equipment to transform end-use energy (delivered to the household) into useful energy (which can be directly used by the household) is also critical to the household's choice of an energy source (Pachauri et al., 2004; Pachauri and Spreng, 2004). Therefore, market access to purchase electrical equipment and appliances can encourage or restrict the household's preference for electricity.

5.2.2. Block Tariffs

Most Indian states (33 out of 35 in 2012) have long adopted increasing-block prices for their residential customers, where lower rates are offered for small consumption levels. In theory, this tool enables low-income users to afford basic energy needs while the electricity provider recovers its costs from residential consumers with larger consumption. The higher unit rates also encourage energy conservation practices among large consumers. In the Indian context, electricity rates are determined by each state electricity regulatory commission, based on three main guiding principles. First, they aim to bring rates close to the average cost of supply, in accordance with the Electricity Act of 2003, (e.g. Delhi Electricity Regulatory Commission). Second, many states (e.g. Bihar, Gujarat, Madhya Pradesh, Jharkhand…) offer a special tariff for customers under a certain income level (the below poverty line or 'BPL' tariff, or the 'Kutir Jyoti' rate), which is usually valid for the first 30 to 50 kWh of consumption per month and is

aimed to cover basic needs.[24] Third, an effort is made to reduce the number of unmetered connections, in the states where they still exist. However, only little information is publicly available about the rationale behind the choice of the various block rates and boundaries. It is unclear if these parameters result from data-driven analysis of the customer base or are the result of historical structures.

There is an incredible variety of price structure among the Indian states, reflecting the variety of local conditions. Some states offer 2 blocks only (Jharkhand), but most states have 3 or 4 blocks, sometimes up to 8, with the first boundary often at 30, 50 or 100 kWh per month. Some states differentiate customers between urban and rural areas (e.g. Uttar Pradesh, Gujarat…) while some states (Puducherry) offer a single flat rate. In other cases, customer types are differentiated by their total consumption and assigned different tariffs from the first unit of consumption, meaning that larger customers pay higher rates from all their consumption. Example of the diversity of state specific IBP structures are available in Appendix A. Therefore, rural customers face tariffs that can be quite different, and lower, from the tariffs of their urban counterparts.

However, block prices have generally been found to be problematic. Theoretically, they fail to provide users the incentives that reflect marginal costs (Boland and Whittington, 1998). Empirically, it was shown that better-off users benefit more from the price support than poor households, that insufficient revenues are collected and that an overall deadweight loss of customer welfare is created (Borenstein, 2009; Cardenas and Whittington, 2019; Singh et al., 2005, 2019; You and Lim, 2017). Specifically, Mayer et al. (2015) calculates that 93 percent households receive subsidized electricity in India (price below cost recovery) and rich households receive equal or larger subsidy in value. Only 13 percent of the subsidy reaches the poors, because of their low access rates to electricity, low consumption levels, and higher unit price than wealthier households due to fixed charges. For the same reasons, it also becomes clear that the IBP system also disporportionally benefits urban users compare to rural users. To counter this, the recent policy amendments that aim to bring tariffs close to cost of recovery could remove some of the heterogeneity (Chindarkar and Goyal, 2019) and possibly reduce the mistargeting of the subsidy toward richer households. However, the effect of implementing higher and more uniform prices in the rural sector has not yet been examined in the literature.

[24] 30 kWh / month procure the equivalent of 6–8 hours of lighting for 2-3 units, 8–10 hours of fan use, 2–3 hours of TV use, and mobile/radio charging (CEEW, 2018).

5.2.3. Elasticities in the literature

Price and income responses of residential electricity demand is generally found to be relatively inelastic with estimates inferior to 1 (in absolute value), with short run demand being less elastic than long-term demand. Some examples, including estimates for India, are provided in Table 5.1.

Table 5. 1: Previous residential elasticity estimates (non-exhaustive)

Authors	Date	Location	Price elasticity	Income elasticity	Methodology
Espey and Espey	2004	mostly OECD	-0.35 (ST) -0.85 (LT)	0.28	Meta-analysis
Miller and Alberini	2016	US States	-0.2 to -0.8		
Labandeira et al.	2017	OECD and non OECD	-0.203 (ST) -0.52 (LT)		Meta-analysis / all sectors
Sun and Ouyang	2016	China	-0.38	0.63	LA-AIDS model
Hung and Huang	2015	Taiwan	-0.454 to -0.857 (ST) -1.149 to -1.1318 (LT)	0.291 and 0.205 (ST) 0.737 and 0.315 (LT)	Fixed effect / summer and non-summer
Ngui et al.	2011	Kenya	-0.631	0.85	LA-AIDS model
Bose and Shukla	1999	India	-0.65 (ST)	0.88	Double-log demand functions with lagged model
Filippini and Pachauri	2004	India	-0.29 to -0.51 (urban only)	0.60 to 0.64	Household production theory / seasonal differences
Gundimeda and Köhlin	2008	India	-0.58 to -0.44 (urban) -0.42 to -0.44 (rural)	0.66 to 0.89 (urban) and 0.53 to 0.72 (rural)	LA-AIDS
Harish et al.	2014	India	-0.3 to -0.4		OLS regression and fixed effects
Saha and Bhattacharya	2018	India	-0.29 (urban) -0.49 (rural)		Fixed effect model
Singh et al.	2018	India	-0.72 (urban)		Focus on appliances
Chindarkar and Goyal	2019	India	-0.25 (urban) -0.47 (rural)		Fixed effect model /quadratic specification

ST: short term; LT: long term

Regarding the particularities of rural users, Bhattacharyya and Timilsina (2010) question the relevance of the standard models to address the specific features of developing countries. Specifically, they advise to include traditional substitutes and non-monetary transactions, which are widely used in developing countries, as they otherwise create an incorrect allocation decision. Also, estimation should allow for heterogeneous consumer characteristics, as demonstrated by Sun and Ouyang (2016) or Reiss and White (2005) for income levels.

Gundimeda and Köhlin (2008), Saha and Bhattacharya (2018) and Chindarkar and Goyal (2019) evaluate elasticities separately for urban and rural groups in India. Rural own-price elasticities for electricity are reported higher than for the urban sector in the two most recent studies, but slightly lower than urban estimations in the first study. Access to non-commercial energy and use of biomass was included in the earlier study but not in the last two. These studies also explicitly assume that the difference between marginal and average prices is small enough that it can be omitted from the estimations, and so none of them account for the lower marginal rates faced by rural customers or the endogeneity issue due to IBP.

5.3. Model

5.3.1. Demand estimation with non-linear prices in literature

Many researchers (Houthakker, 1951; Houthakker et al., 1974) have early on identified the econometric problems linked to block prices. First, this price structure creates a non-linear budget constraint where the average price depends on every intramarginal price of the price schedule. Second, the users, aware that they will pay higher unit price if they consume above a certain threshold, might voluntarily limit their consumption under that threshold. Therefore, household decision on quantities and price are made simultaneously, causing a problem of endogeneity and leading to biased estimates if OLS estimation techniques are used. Taylor (1975) underlined the importance of using the marginal prices in the estimation of demand, and insisted that all price rates shall be taken from actual tariff schedules and not be *ex post* prices. Nordin (1976) added that the use of a rate structure premium (RSP), the difference between the observed expenditure and a hypothetical expenditure under a flat rate, is important to capture the income effects resulting from changes in intramarginal prices. Other researchers (Hausmann et al., 1979; Henson, 1984; McFadden et al., 1977, Billings, 1982, and more recently Hung and Huang, 2015) have thus frequently adopted this Taylor-Nordin approach. They use marginal rate and RSP from actual rate schedules as instrument variables (IV) to predict demand and avoid the problem of endogeneity.[25]

[25] Opinions remain divided as to whether marginal, average or other measures of price should enter demand estimation models. Wilder and Willenborg, 1975, consider that customers rarely know the marginal price and that elasticities are the same under marginal and average price. In comparative studies, households were found to

By comparing IV procedures to OLS models, these studies found that the presence of endogeneity affects results, and creates bias, although the income effect of the RSP is considered too small to matter (Henson, 1984). Miller and Alberini (2016) found that OLS overestimated elasticities compared to an IV specification using average price of electricity. But the IV approach using actual rates continues to create biased estimates when marginal rate and RSP are calculated at predetermined consumption levels, especially when the block price structure is steep. In response, Reiss and White (2005) built a system of equations where demand is estimated at each marginal price from actual rate schedules.

5.3.2. Marginal price and rate structure premium

In order to capture the impact of the nonlinear pricing structure on the consumer demand, we follow the Taylor-Nordin approach but we aim to apply, for each household, the actual marginal rate they face instead of relying on predetermined consumption levels. Each household's consumption level, state, sector (urban/rural) and poverty status are matched against actual rate schedules to find its relevant marginal price. To resolve the potential endogeneity of marginal price, each household's marginal price is instrumented by the average marginal price of all other households in same district who consume similar amounts of electricity per month, as inspired from Ito (2014). For this, electricity consumption is ranked in categories ranging from 0-25, 25-50, 50-75, 75-100, 100-125, 125-150, 150-200, 200-250, and above 250 kWh/Month. Significant diagnostics test confirm that this approach forms a good instrument for the marginal price (Weak instrument test) and that the instrument variable approach should be used (Hausmann test).

Second, a rate structure premium (RSP) is included to account for the effect of the shape of the price structure in blocks. It is calculated for each household as the difference between actual expenditure under the block prices and a hypothetical expenditure (plain orange area of Figure 5.1) where all units are paid at a flat rate, using the lowest block of the schedule as a benchmark price (p^B)[26]. Under increasing tariffs, the RSP (blue-stripped area of Figure 5.1) represents a

---

respond to a perceived average price which reflects the rate structure premium (Shin, 1985), or to an expected marginal price (Borenstein, 2009). Thus, households are not completely ignorant of the marginal rate structure they face. Lately, it has been found that average price is a better predictor of demand than the marginal price (Ito, 2014). However, these studies relate to the developed context.

[26] To check for robustness, calculations using the hypothetical expenditure if all units had been charged at each customer's current marginal rate is also performed, following Henson, 1984, Billings, 1982; Hung and Huang,

form of penalty, or tax levied by the government on the users for consuming above a certain threshold. In the following, this amount is referred to as the tax.

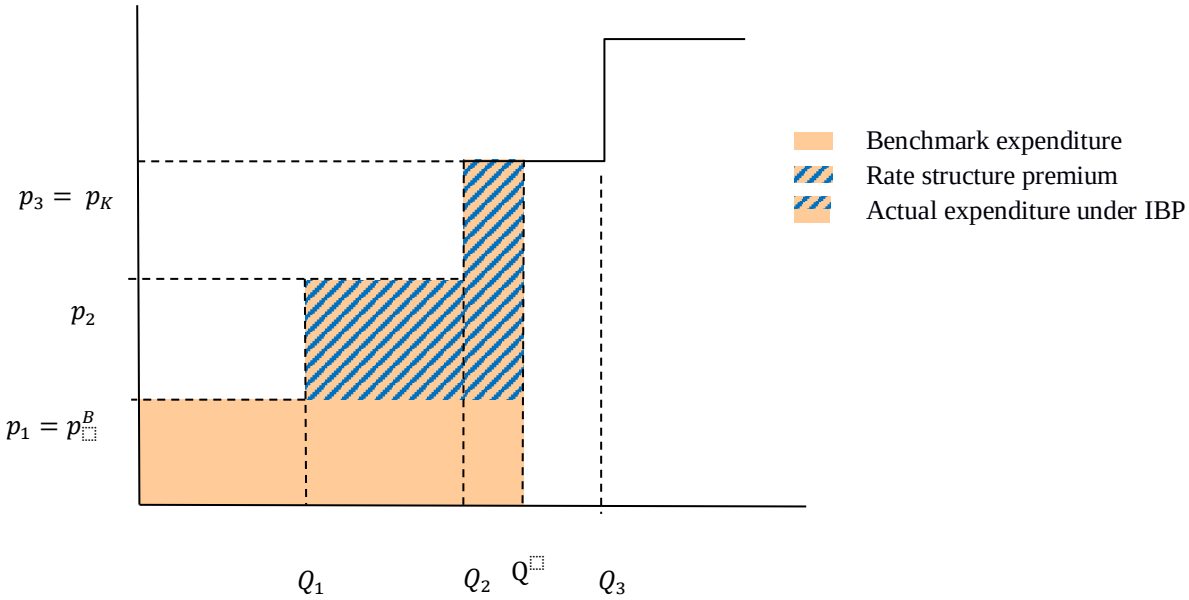

Figure 5. 1: Actual and benchmark expenditure under IBP

For a consumption level Q in block K, such that $Q_{K-1} < Q < Q_K$ at marginal price p_K, the tax can be expressed as:

$$TAX = \left[\sum_{k=1}^{K} p_k \left(Q_k - Q_{k-1}\right) - p_K (Q_K - Q) \right] - p^B Q \qquad (1)$$

Where p_k are the intramarginal prices. The RSP captures the income effect of the price structure in blocks, as the tax represents a reduction in the available income resulting from higher intramarginal prices. Thus, we have $Q = f(Y^*, p_K)$, the demand for electricity Q is a function of the marginal electricity price p_K and available income of the household Y^*, where:

$$Y^* = Y - TAX. \qquad (2)$$

2015). Consistent with Henson (1984), we assume that using p^B as marginal rate for the consumer's consumption does not create a difference large enough to induce a block-switching behavior.

The available income is the total expenditure Y, less the amount levied by the government for high electricity consumption (TAX). In the case of a developing country with disadvantaged users, the income effect of the RSP can be more substantial than it was previously found for developed countries (Henson, 1984).[27]

For comparison and robustness, we also build a 'base model' that computes elasticities by ignoring endogeneity and non-linearity altogether. In this standard approach, demand response is calculated with respect to average prices, calculated *ex post* from survey information, and a standard measure of total expenditure as income.

5.3.3. Two-stage budgeting model

Following these preliminary steps, elasticities are obtained using a two-stage budgeting model similar to Hasiner and Yu (2016). Under the assumption of weakly separable consumer utility functions, households first decide the share of their budget to spend on each category of expenses (food, housing, energy, …) and in a second stage, they allocate their budget share between specific items within each group.

5.3.3.1 Stage I elasticities

Stage I elasticities, the elasticities of the group energy (j) with respect to the overall household budget, are calculated first. A functional form of the budget share, or Working-Leser model (Leser, 1963; Working, 1943), is adopted:

$$W_j^* = \eta_j \ln P_j^* + \beta_j \, lnY^* + \sum_{n=1}^{N} \epsilon_{jn} d_n \qquad (3)$$

where:

W_j^* is the share of energy expenditures in the household budget Y*;

d_n is a set of N socio-economic and demographic variables, for which ϵ_{jn} are parameters;

[27] Electricity bills often comprises of at least two parts, a fixed fee and a variable fee. Under an efficient rate system, the fixed fee is aimed at recovering the fixed costs of the utility and the variable charge to recover the variable cost of the utility provider (Coase, 1946). In the present analysis, we are concerned with the effect of the increasing marginal rate, which corresponds to the variable part of the bill. Fixed charge plays no role in the calculation of the tax.

And P_j^* is the Stone Price index for group *j*, built from the prices p_i^* of all its composing items *i*, weighted by their respective mean budget shares $\overline{w}_i$:

$$\log P_j^* = \sum_i \overline{w}_i \ln p_i{}^* \qquad (4)$$

The non-linearity (indicated by *) is accounted for at this stage by using the marginal price of electricity instead of the average price in the Stone Price index. Its income effect is included by using Y^*, where the total expenditure is deflated from the tax. That is, the stage I captures the income penalty from the tax. The expenditure on electricity and on the overall energy category are taken from the survey and are not affected by these changes. However, the share of energy in total expenditure W_j^* is calculated over $Y^*$and thus differs in the base and IBP specifications.

The first-order derivatives provide the own price and income elasticities of energy as:

$$E_j = \frac{\eta_j}{W_j{}^*} - 1 \text{ (price elasticity, stage I)} \qquad (5)$$

$$E_{jY} = \frac{\beta_j}{W_j{}^*} + 1 \text{ (expenditure elasticity, stage I)} \qquad (6)$$

5.3.3.2 Stage II elasticities

The LA-AIDS (Deaton and Muellbauer, 1980; Buse, 1994) is appropriate to model consumer demand (Ryan and Plourde, 2009) and offers the possibility to account for alternative sources of energy (here kerosene, gas, oils, coal, candles, firewood and biomass). Similar to Ngui et al. (2011) or Hasiner and Yu (2016), a Shephard's lemma is applied to a cost function and the expenditure share for the ith fuel (electricity) in group *j* (energy) is obtained:[28]

$$w_i = \omega_{io} + \sum_{n=1}^{N} \omega_{in} d_n + \sum \gamma_{il} \ln p_i^* + \beta_i \ln\left(\frac{y}{P_j^*}\right) \qquad (7)$$

where:

w_i is the share of expenditure of good *i* in energy budget;

γ_{il} is a parameter for the ith good and the lth good;

[28] The usual restrictions of adding-up, homogeneity and symmetry are imposed on the model, and demographic information is included using the translating method, modifying the intercept.

y is the total expenditure on energy.

As in the first stage, the non-linearity of prices is reflected by using the marginal price of electricity instead of the average price in the Stone Price index. The expenditures on electricity and energy are not affected by this; thereby the share of electricity in energy expenditure is the same for both base and IBP models. The income effect of the penalty TAX being already captured at stage I, it is only controlled for in the estimation of this stage, as follow:

$$w_i{}^* = \omega_{io} + \sum_{n=1}^{N} \omega_{in} d_n + \sum \gamma_{il} \ln p_i^* + \beta_i \ln\left(\frac{y - TAX}{P_j^*}\right) \qquad \textbf{(8)}$$

Following Buse (1994), we can use the as-if expenditure $y - TAX$, rather than y itself in estimation, and obtain the Marshallian own price and income elasticities from the coefficients estimated:

$$e_i = \frac{\gamma_i}{w_i} - \beta_i - 1 \text{ (price elasticity, stage II)} \qquad \textbf{(9)}$$

$$e_{iy} = \frac{\beta_i}{w_i} + 1 \text{ (expenditure elasticity, stage II)} \qquad \textbf{(10)}$$

These Marshallian uncompensated elasticities are transformed to Hicksian compensated elasticity using the Slutsky equation (11):

$$e_i^c = e_i + e_{iy} \,.\, w_i \text{ (compensated price elasticity, stage II)} \qquad \textbf{(11)}$$

5.3.3.3 Unconditional elasticities

After calculating the stage I and stage II elasticities, we are able to compute the unconditional elasticities, with the methodology employed in Hasiner and Yu (2016):

$$n_i = e_i + e_{iy} \,.\, w_{iy}(1 + E_j) \text{ (unconditional price elasticity)} \qquad \textbf{(12)}$$

$$n_{iY} = e_{iy} . E_{jY} \text{ (unconditional income elasticity)} \qquad \textbf{(13)}$$

Finally, uncompensated price elasticity is transformed to compensated unconditional elasticities with the formula:

$$n_i \ = n_i + n_{iY}.w_{iY} \text{ (compensated unconditional price elasticity)} \qquad (14)$$

5.3.4. Bias calibration

Section 3.3 proposed a two-stage model to estimate price and income elasticities with the marginal prices and as-if expenditure $Y - TAX$. However, changes in TAX as a result of changes in the marginal price were not considered. We can calibrate the possible bias by assuming a standard model:

$$\text{Ln}\, Q_i = \alpha_i + \beta_i \ln p_i + \gamma_i \ln(Y^*) \qquad (15)$$

Thus β_i and γ_i are the marginal price and income elasticities of demand for good i, established using equations (12) to (14).

To account for changes in TAX, we replace the expenditure with $Y^* = Y \ - TAX$ where $\text{TAX} = \text{TAX}\ (p_i, Q_i)$, and we obtain:

$$\ln Q_i = \alpha_i + \beta_i \ln p_i + \gamma_i \ln(Y - TAX) \qquad (16)$$

Now taking the derivative of (16) with respect to $\ln p_i$ and $\partial \ln(Y - TAX) = \frac{\partial(Y-TAX)}{(Y-TAX)}$ gives:

$$\frac{\partial \ln Q_i}{\partial \ln p_i} = \frac{\partial Q_i / Q_i}{\partial p_i / p_i} = \beta_i - \frac{\gamma_i}{Y^*}\left[\frac{\partial TAX}{\partial \ln p_i}\right] \qquad (17)$$

Replacing TAX by its definition in equation (1) and taking its derivative with respect to the marginal price p_i, we have:

$$\frac{\partial TAX}{\partial p_i} = (Q_i - Q_{K-1}) \qquad (18)$$

which is the quantity consumed in the last block. Thus, $\frac{\partial TAX}{\partial \ln p_i} = \frac{\partial TAX}{\partial\, p_i / p_i} = \ p_i\ (Q_i - Q_{K-1})$ which is the expenditure in the last block of consumption

Thus, writing $\varepsilon_p = \frac{\partial Q_i / Q_i}{\partial p_i / p_i}$ the 'IBP-consistent' own-price elasticity, we have:

$$\varepsilon_p = \hat{\beta} - \hat{\gamma}_{iY}\left[\frac{p_i.Q_i - p_i.Q_{K-1}}{Y^*}\right] \quad \textbf{(19)}$$

where $\hat{\gamma}$ and $\hat{\beta}$ are the estimated unconditional income and price elasticities estimated in section 3.3.3. Clearly, we find that own-price elasticity ε_p could bias up by accounting for the effect of TAX as results of change in marginal price. in the bias. In other words, the actual IBP-consistent elasticity could be smaller in absolute value than found without the bias in (14).

5.4. Data

5.4.1. Survey information

Data on expenditures and consumption quantities was collected through the 68th round of the National Sample Surveys (NSS)[29] on Household Consumer Expenditure, carried out in 2011-2012 and covering over 100 000 households in whole India. The survey reports, for the period over the last 30 days, both market and non-market expenditure and quantities, such as of home-grown food, and self-collected firewood or dung cake, which can make a substantial part of the total consumption in developing countries. For this study, the term 'energy' groups consumption of electricity, wood, natural gas (LPG), kerosene (both subsidized and market quantities), and other fuels (coal, charcoal, petrol, diesel), as well as candle, dung cake and biogas. All in all, the sample is restricted to those having positive expenditure of electricity, or 88,281 observations. Electricity is mainly used for lighting (chosen by 87 percent of households), powering appliances, and marginally for cooking (0.3 percent of the households).

As seen in Table 5.2, rural electricity consumption is lower than urban consumption and more than 1 in 5 rural electricity users is not able to satisfy its basic electricity needs, or 30 kWh/month. Furthermore, electricity dominates urban households' energy budget (43 percent), while it represents a much smaller share of rural households' energy expenditure, which is instead dominated by woodchips and firewood. Additional summary statistics per state are given in Appendix B and show diverse patterns across states.

[29] Household Consumer Expenditure: NSS 68th Round, NSSO, Ministry of Statistics & Programme Implementation, Government of India, http://microdata.gov.in/nada43/index.php/catalog/1

Average *ex post* prices are calculated based on the reported consumption quantities and expenditures, and amount to 3.00 INR/kWh (or about 0.037euro/kWh) for electricity.[30]

Table 5. 2: Summary statistics, variables of interest, 68th NSS.

	Whole sample		**Urban sample**		**Rural sample**	
	(n=88281)		***(n=40025)***		***(n=48256)***	
	Mean	*St. Deviation*	*Mean*	*St. Deviation*	*Mean*	*St. Deviation*
Consumption of electricity, kWh/Mo.	85	83	106	102	67	57
Share of obs. below 30 u./Mo.	0.18	0.38	0.13	0.33	0.22	0.41
Price of electricity (average), INR/kWh	3.00	1.34	3.20	1.38	2.83	1.28
Stone Price index	12.49	8.05	14.91	7.34	10.49	8.07
Expenditure on electricity, INR/Mo.	247	302	332	386	178	181
Total expenditure on energy	662	420	716	478	617	359
Total expenditure of household	9549	8088	10739	9132	8561	6955
Share of electricity in energy expenditure	0.35	0.19	0.43	0.21	0.28	0.15
Share of energy in total expenditure	0.08	0.04	0.08	0.04	0.09	0.04
Share of electricity in total expenditure	0.03	0.02	0.03	0.02	0.02	0.02
Lighting with electricity, share of obs.	0.99	0.11	1.00	0.07	0.98	0.13

5.4.2. IBP variables

To build the IBP model, all elements of the increasing-block tariffs (block boundaries and marginal price at each level[31]) are collected from approved tariff schemes, found in the 'Aggregate Revenue Requirement' publications of the websites of each state regulatory commission. As changes to the approved tariffs often occur during the survey period (01.07.2011 to 30.06.2012), each household is matched with the tariff scheme in place during the quarter of its interview. When the information back to 2011-2012 is not publicly available (Chattisgarh, Sikkim), information available for the period as close as possible afterwards (2013-2014) is used instead. In total, there are 109 different tariff schedules for the period 2011-2012. The marginal price assigned to each household reaches an average of 2.38 INR/kWh. The highest marginal price of each scheme reaches on average 4.71 INR/kWh across all states. Most

[30] The local price of exchanged firewood and woodchip is used as proxy for the price of dung cake and gobar gas, for which price cannot be derived.

[31] Fixed charges are also generally published along with the variables rates, but are sometimes determined based on elements not available in the survey (e.g. load). Plus, the inconsistency across states makes it difficult to assess with certainty the real amount of fixed charge applied to the customers. It is therefore excluded.

households consume in more than 1 block, thereby facing tiered-pricing, but the rates above 200 kWh/month are relevant to only a minority of users.

To obtain the benchmark expenditure, each customer is assigned to a benchmark price from the tariff that applies to its geographical location (state and, if applicable, district) and sector (urban or rural). To compute the actual expenditure under IBP, marginal rates from the same schedule are assigned to each household, but information on the household consumption level and whether they possess a ration card help presume which households are entitled to the reduced BPL rates (or 'Kutir Jyoti').[32] The difference between the actual expenditure IBP and the benchmark expenditure provides the amount of premium, or tax, the households pay for their consumption above a certain threshold. For 744 observations (less than one percent of the sample), the premium turns out to be negative which is inconsistent with the assumptions made by the approach. We assume this situation results from either measurement error of the survey information or from a failure to correctly match a household with its correct tariff. Therefore, these observations are removed from the analysis.

A further validity check can be found as follows. Calculating the actual expenditure with the tariffs (all intramarginal rates) allows identifying the variable part of the bill only, while the expenditure reported in the survey reflects the variable plus the fixed charges. The difference between the two provides a value for the fixed part of the bill, whose distribution can be checked for reliability. As it turns out, the fixed cost is positive for 73 percent of the observations. We remove the other observations, whose distribution across quantity and expenditure values is similar to the overall sample, thereby indicating that their removal doesn't create a strong sample bias. Additionally, 15 high outliers are removed, leaving a sample of 63563 observations. Therefore, although the matching process can't be perfect due to lack of information, these steps provide some confidence that measurement error on survey information or a mismatch of actual tariff schedule are kept at minimum within the remaining sample.[33] Summary statistics of the above-mentioned variables for the retained sample are provided in Appendix C.

[32] Some households, especially small users in rural areas, are connected through unmetered connection and thus pay only a fixed amount. We do not know which connections are metered and which are unmetered, and thus we consider that all are metered and face the variable charges of the schedule.

[33] The fixed part of the bill does not cause further estimation problem, because in both IBP and base models, total expenditure of electricity entering estimation is the same (as reported in the survey).

5.4.3. Controls

Socio-economic and demographic variables obtained from the survey information are included as controls in both stage I and stage II of the analysis, and presented in Table 5.3. In addition, dummy variables are included for each state, and for the quarter during which the data was collected to account for seasonality effects on elasticities (Filippini and Pachauri, 2004).

Table 5. 3: Control variables used in OLS for stage I and LAIDS for stage II

	Whole sample		**Urban sample**		**Rural sample**	
	(n=63563)		***(n=28358)***		***(n=35205)***	
	Mean	*St. Deviation*	*Mean*	*St. Deviation*	*Mean*	*St. Deviation*
Rural	0.55	0.50				
Household size	1.30	0.51	1.25	0.47	1.34	0.53
Age of household head	46	13	45	13	47	13
Education: primary or less	0.44	0.50	0.33	0.47	0.53	0.50
Possess ration card	0.79	0.41	0.71	0.46	0.85	0.35
Regulary salary earner	0.32	0.47	0.44	0.50	0.23	0.42
Owned dwelling	0.83	0.38	0.68	0.47	0.95	0.22
Number of different types of equipment owned	2.76	1.81	3.32	1.98	2.31	1.52
Own many appliances (>10)	0.16	0.37	0.25	0.44	0.09	0.28
Cooking with LPG	0.46	0.50	0.70	0.46	0.26	0.44
Cooking with traditional fuel	0.46	0.50	0.18	0.38	0.68	0.47
Lighting with electricity	0.99	0.12	0.99	0.08	0.98	0.15

5.5. Results

5.5.1. IBP elasticities

Elasticities for the whole, urban and rural samples are calculated for an IBP model, and compared to a base model, which ignores the specificities of IBP and uses *ex post* average prices.

Parameters for the stage I (presented Table 5.4 for the IBP regression), obtained with an OLS estimation, help control for the main determinants of the household total energy expenditures decision. Results are consistent with the view that a better economic status allows households to spend relatively less on basic necessities and more on other goods and services (significant negative coefficients on income and regular salary, but positive coefficient on ration card). Being a rural household reduces the share of domestic energy expenditures in the overall

budget, while household size, and having a low education (primary or less) increase it. Most state dummies are significant but are not reported here. The price level affects positively and significantly the energy expenditure share. Households that own their dwellings spend relatively more on energy than households in rented dwellings, possibly because electricity costs can be included in the rent as a lump sum, thereby increasing the energy expense burden of landlords while decreasing that of tenants. Overall significance and signs of coefficients are similar across the whole, urban and rural samples.

Table 5. 4: OLS regression results for stage I, IBP model

	Whole sample		**Urban sample**		**Rural sample**	
	Estimate	*Std. Error*	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*
Intercept	3.65E-01***	0.0026	3.08E-01***	0.0041	4.16E-01***	0.0036
Stone Price Index	7.17E-03***	0.0002	1.28E-02***	0.0004	3.97E-03***	0.0003
Log Income	-3.80E-02***	0.0003	-3.43E-02***	0.0004	-4.20E-02***	0.0004
Rural	-6.61E-03***	0.0003	na	na	na	na
Household size	2.60E-03***	0.0003	2.56E-03***	0.0004	3.80E-03***	0.0004
Age of household head	2.10E-04***	0	3.10E-04***	0	1.13E-04***	0
Education: low	2.26E-03***	0.0003	2.55E-03***	0.0004	1.50E-03***	0.0004
Possess ration card	4.18E-03***	0.0003	5.07E-03***	0.0005	1.90E-03***	0.0005
Regulary salary earner	-8.19E-04**	0.0003	-1.08E-03**	0.0004	-1.14E-03**	0.0004
Owned dwelling	1.33E-02***	0.0004	1.23E-02***	0.0005	1.27E-02***	0.0008
Number of owned equipment category	1.08E-03***	0.0001	2.56E-03***	0.0002	-2.61E-04	0.0002
Own many appliances	6.97E-04	0.0006	-3.38E-03***	0.0007	2.67E-03**	0.0008
Cooking with LPG	1.39E-02***	0.0005	5.32E-03***	0.0007	1.72E-02***	0.0008
Cooking with traditional fuel	2.41E-02***	0.0005	2.72E-02***	0.0007	1.99E-02***	0.0008
Lighting with electricity	1.48E-06	0.001	-2.60E-04	0.0023	-3.15E-04	0.0012
Sub Round 1	-3.77E-03***	0.0003	-1.79E-03***	0.0005	-5.52E-03***	0.0005
Sub Round 2	-8.33E-04*	0.0003	7.44E-05	0.0005	-1.66E-03***	0.0005
Sub Round 3	2.77E-04	0.0003	-9.14E-04.	0.0005	1.27E-03**	0.0005
R-squared Adjusetd	*0.42*		*0.46*		*0.40*	

Significant levels : 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

The top part of Table 5.5 presents the resulting elasticities for stage I. In stage II, most coefficients, calculated with the LA-AIDS model[34], are significant and the elasticities are of the expected sign, with negative own-price elasticities for all types of energy, and positive expenditure elasticities. Lastly, the last panel of Table 5.5 shows that IBP unconditional own-price elasticities are larger than the base estimates, while the IBP income elasticities are smaller (in absolute values), and that the confidence intervals do not overlap. Thus, ignoring the specificities of the IBP structure leads to under-estimating price elasticities. Overall, the difference between IBP and based model is more marked at stage II than stage I.

The resulting income elasticities of the base model are consistent with earlier studies (price elasticities between -0.2 and -0.5 and income elasticities between 0.5 and 0.8), in particular the most recent price response estimation of Chindarkar and Goyal (2019) of -0.47 for rural dwellers.

[34] Arne Henningsen, Demand Analysis with the "Almost Ideal Demand System" in R: Package micEconAids, https://CRAN.R-project.org/package=micEconAids

Table 5. 5: Electricity demand elasticities

		Whole sample	Urban sample	Rural sample
		Est. *(Conf. interval)*	*Est.* *(Conf. interval)*	*Est.* *(Conf. interval)*
Stage I				
Income	IBP Model	0.55	0.59	0.50
		(0.54 : 0.55)	(0.58 : 0.60)	(0.49 : 0.51)
	Base model	0.55	0.60	0.50
		(0.55 : 0.56)	(0.59 : 0.61)	(0.49 : 0.51)
Own Price *	IBP Model	-0.91	-0.85	-0.95
		(-0.92 : -0.91)	(-0.86 : -0.84)	(-0.96 : -0.94)
	Base model	-0.94	-0.87	-0.97
		(-0.94 : -0.93)	(-0.88 : -0.86)	(-0.98 : -0.96)
Stage II				
Income	IBP Model	1.05	1.15	0.92
		(1.04 : 1.05)	(1.14 : 1.16)	(0.91 : 0.93)
	Base model	1.12	1.2	0.97
		(1.11 : 1.12)	(1.19 : 1.2)	(0.96 : 0.98)
Own Price *	IBP Model	-0.44	-0.4	-0.49
		(-0.45 : -0.42)	(-0.42 : -0.39)	(-0.51 : -0.47)
	Base model	-0.37	-0.31	-0.4
		(-0.38 : -0.35)	(-0.33 : -0.29)	(-0.42 : -0.38)
Unconditional				
Income	IBP Model	0.57	0.68	0.46
		(0.57 : 0.58)	(0.67 : 0.68)	(0.46 : 0.46)
	Base model	0.62	0.72	0.49
		(0.61 : 0.62)	(0.71 : 0.72)	(0.49 : 0.49)
Own Price *	IBP Model	-0.39	-0.31	-0.47
		(-0.4 : -0.38)	(-0.32 : -0.29)	(-0.48 : -0.45)
	Base model	-0.33	-0.22	-0.38
		(-0.34 : -0.31)	(-0.23 : -0.2)	(-0.4 : -0.36)

*Compensated Own Price Elasticities

Lastly, the calibration step (equation 19) indicates that when expenditure in the last block w.r.t. income increases, demand will become more elastic. The elasticity thus depends on the location of the household within the block, i.e. if consumption Q_i is close to the next block, elasticity will be larger in absolute terms than the estimates presented, whereas if it consumes close to the lower limits of its block, the effect will be limited. This effect also depends on income level and is expected to be larger for low-income households whose electricity spending represents a higher share of their income. Therefore, ignoring the calibration of the IBP elasticities would have limited consequences for households with high income but would predict a less elastic reply for some low-income households. However, with increasing block prices, wealthier households also experience higher unit rate for their last block consumption than low-income

households do, which mitigates the previous interpretation. Therefore, the effect of the calibration is small and differs for each household. It is therefore not estimated.

5.5.2. Rural specificities

At each stage of the calculation, prices responses are larger (by up to 0.15 for the unconditional elasticities) for the rural than the urban sample. Gain in revenue collection would thus be best achieved from urban customers, as rural demand would decrease faster in case of a uniform rate increase. The household overall welfare effect of rising prices also depends on the ability to switch fuels and find substitutes. The electricity price response of the demand for kerosene, the most prevalent substitute to electricity, is also obtained from the LA-AIDS model and shows that urban households substitute more easily toward kerosene (0.26) than rural households (0.14) when electricity becomes more expensive (Table 5.6), which is consistent with Singh et al. (2018). Therefore, rural households tend to reduce their electricity demand more than their urban counterparts and do not operate fuel substitution to the extent of their urban counterparts. Given that rural households already have lower consumption of electricity, a reduction of their end-use electrical service could bring them closer to satisfying just the basic needs or could likely increase the share of households that are not able to meet basic needs using electricity.

Table 5. 6: Stage II compensated Hicksian electricity price response

Budget shares	Urban sample	Rural sample
Woodchip	0.27	0.19
Electricity	-0.40	-0.49
LPG	0.33	0.30
Candle	0.26	0.18
Fuels	0.95	0.24
Coalpdct	0.43	-0.34
Biomass	-0.35	0.12
Kerosene	0.26	0.14

As income changes, changes in rural demand are observed to be more limited (0.46) than urban demand (0.68). Indeed, rural households' consumption can be bounded both downward and upward. First, for a number of rural households, electric consumption is so low that it covers only basic needs, or barely above. These households thus have a limited scope to reduce their consumption, even when income drops (Khandker et al., 2012). On the higher end, even as income grows, a rural household's consumption growth is limited by the lack of sufficient and

reliable supply. This is supported by evidence form Sedai et al. (2021) who show that increasing the hours of electricity supply to the rural sector would benefit more the higher-income households than lower income groups, pointing to the unmet demand of this group. Furthermore, the lack of end-use devises is another limiting factor, as appliances penetration is still lower in rural areas. Without as many entertaining devices or household appliances to power, increases in income are not necessarily translated into the same increases in demand as in the urban sector.

In summary, the recent policy changes of compensating increased prices by a direct cash transfer to needy households would have differentiated effects on both urban and rural sectors. The demand reduction following higher prices is larger by rural households, but their increase in demand from higher income is smaller than for their urban counterparts. As long as this rural-urban pattern persists, such uniform policy change leads to a deterioration of rural consumption greater than that of urban dwellers. Observing this heterogeneity, the differentiation of tariffs schedules between the urban and the rural sectors, which exists in some but not all states, appears to be fair to rural users.

5.5.3. Discussion: Progressivity of tariffs

Instead of uniform change in prices, modifying the degree of progressivity of the IBP (i.e. the gaps between price levels at various consumption levels) via discriminated price changes can be another tool to improve revenue collection. Several authors have recently reinforced the necessity to keep heterogeneous and progressive tariffs in place (Chindarkar and Goyal, 2019; Sedai et al., 2021) to reflect the diversity of users and usage patterns, although the current IBP schemes have poor revenue collection performance and are inefficient at targeting of poor households with the subsidized service (Mayer et al., 2015).

To maintain a protection of vulnerable customers but reduce inefficiencies, Mayer et al. (2015) proposes to increase tariffs for above-poverty households, so as to allow subsidizing below-poverty households. Here we examine how and where, in the current block rates, it would be possible to maintain a large magnitude of the subsidized rates (rates far below cost recovery) but reduce their prevalence (fewer households benefit), in other words how to keep low prices for the poorest households, but only them, while raising tariffs for large users.

This requires first an examination of the current state of the price discrimination for various user groups, and across various types of tariff designs. First, for the states that offer the same

tariff structure to all their customers, independently of their characteristics, Figure 5.2 shows that, although unit prices do increase with consumption in most states, the increases are in fact rather small. Corresponding numbers in panel A of Appendix D reveal that many states offer the same rate at 100 kWh per month (approximately the urban consumption) than at 50 kWh per month (about the baseline amount). Among states that apply progressivity at this level, the unit price difference is larger than 1 rupee/kWh for only two states. There is more progressivity (on average 31 percent increase) in the rates between 100 and 200 monthly units of consumption.

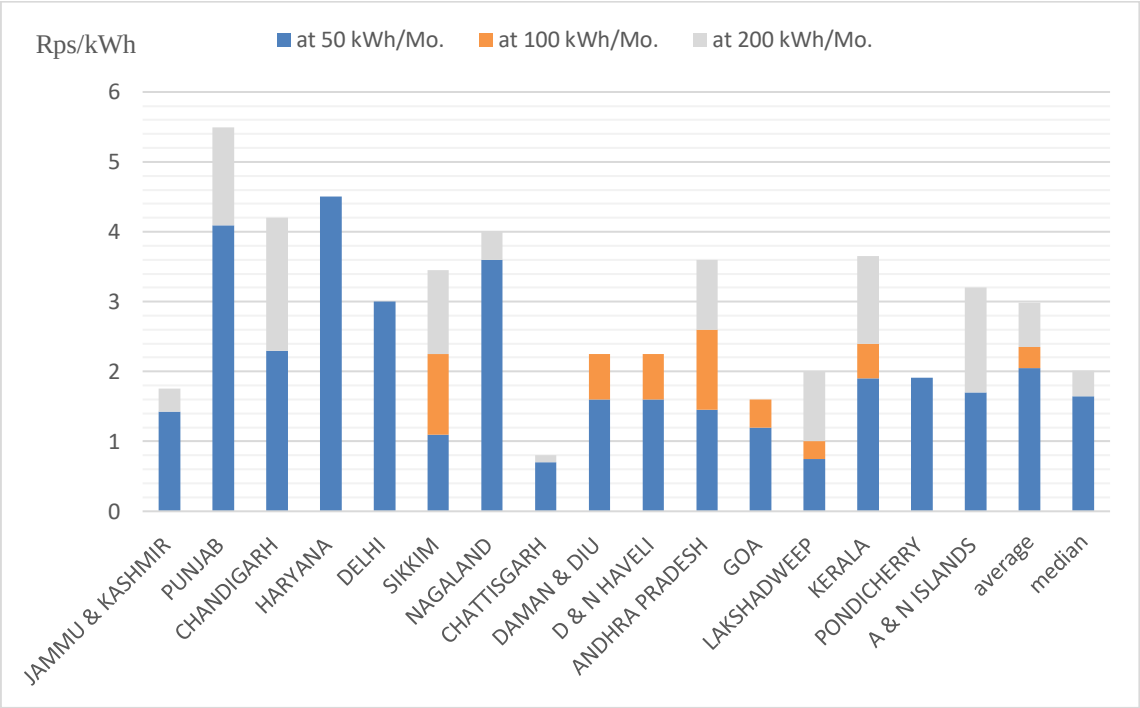

Figure 5. 2: Marginal price at various monthly consumption levels, own calculation from 2012 tariff schedules (For states that offer the same tariff structure to all their customers, independently of their characteristics).

Second, for states that discriminate their customers by poverty status and offer a below-poverty line (BPL) rate for the first 30 or 50 units (Figure 5.3), there is often a marked difference between the BPL rate and rate offered to regaular customers for 50 unit of monhtly consumption (68 percent on average). However, the steps in regular tariff beyond 50 units are rather small, with only 11 percent on average between 50 and 100 units per month, and 25 percent on average between 100 and 200 units per month (see panel B of Appendix D for details per state). These states therefore make a strong discrimination between the BPL and non-BPL rates, but apply

very little progressivity to the non-BPL rate for higher levels of consumption. Finally, some states without a BPL tariff choose to offer differentiated tariffs for poor households and are able to set the regular tariffs slightly higher (not shown here).[35,36]

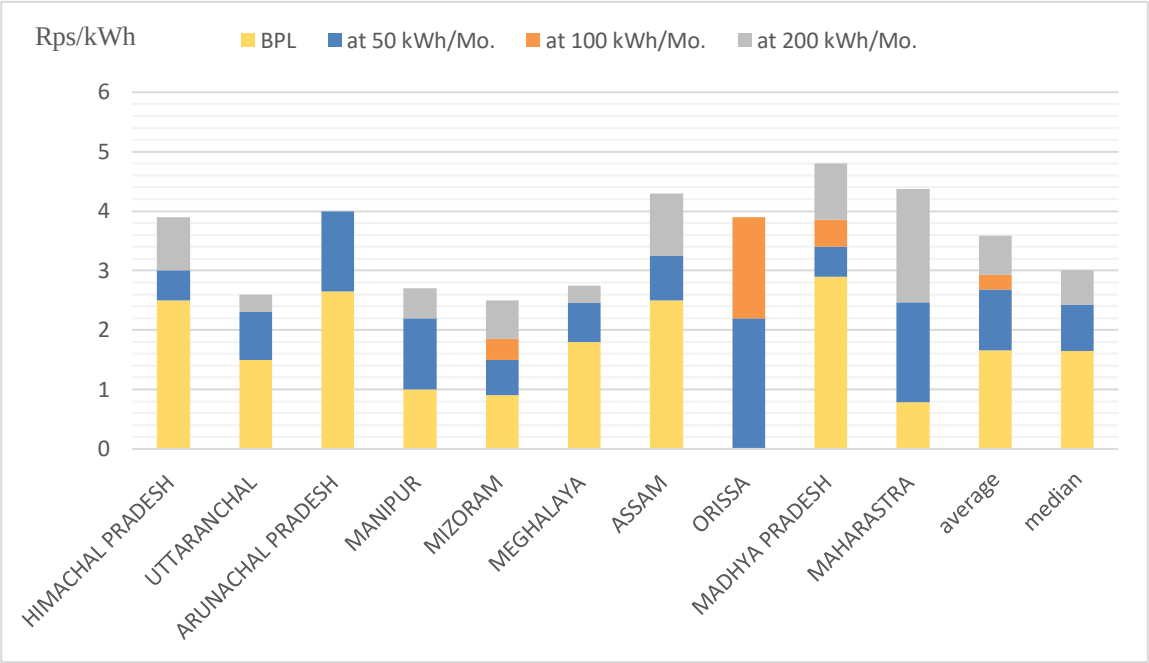

Figure 5. 3: Average unit price at various monthly consumption levels, with BPL tariffs, own data of 2012.

The overall limited progressivity of the tariffs leads to question if more - and not less – price discrimination is needed in the tariffs, in the form of higher prices for larger consumption, esepcially within the blocks relevant to most of the customer base (below 200 kWh/Month). By implementing higher steps at each block, better outcomes could theoretically be obtained, where the low rates remain in place for small users, but only make up a small portion of the medium to large users' consumption. To establish how customer heterogeneity would accentuate or mitigate such changes, we propse an examination of consumer price response by user size. IBP-consistent elasticities are calculated for the rural and urban sector, for three classes of customers

[35] For example, 6 out of 10 of them charge more than 3 rupees/unit at 200 units of consumption, againts 3 out of 16 for the first group of states. The average unit prices are higher, but the differences for the first group of state remain nevertheless small.

[36] Another smaller group of states offered tariffs that combined special BPL tariff along a regular tariff discriminated based on customer size and/or rural-urban sector. These schemes are complex and unique to each state, so that trends are difficult to establish, apart from the fact that rural tariffs are set lower than urban tariffs at each consumption level and that when customer size discrimination is present, larger users pay higher unit prices from the first unit on.

based on their monhtly consumption: small (0-50 units), medium (50-100 units) and large (100-200 units). Small customers are those that consume up to just the basic necessities. They form a third of urban and two third of rural customers. Presenting elasticities by user size, as opposed to income groups, shows directly what revenue changes could the distribution companies expect from rising prices, as they know precisely the usage volume of their cutomer base, but not their income. Table 5.7 shows that price response is noticeably higher for small users than the medium users of both the urban and rural sectors. However, large users also have a more responsive demand than medium users.

Table 5. 7: IBP electricity demand elasticities by user size

	Urban sample			**Rural sample**		
	Small (0-50) *n=10189*	Medium (50-100) *n=10286*	Large (100-200) *n=6238*	Small (0-50) *n=21395*	Medium (50-100) *n=10647*	Large (100-200) *n=2855*
Stage I - Income Elasticites	0.54	0.34	0.27	0.50	0.38	0.35
Stage I - Own Price Elasticites	-0.74	-0.80	-0.88	-0.94	-0.92	-0.95
Stage II - Income Elasticites	1.00	1.15	1.16	0.86	0.83	0.92
Stage II - Own Price Elasticites*	-0.49	-0.33	-0.39	-0.48	-0.42	-0.54
Unconditional - Income Elasticites	0.54	0.39	0.32	0.43	0.32	0.32
Unconditional - Own Price Elasticites*	-0.39	-0.22	-0.30	-0.46	-0.39	-0.51

*Compensated Own Price Elasticities

Due to the relatively inelastic demand of medium users in both sectors, increasing tariffs for these users would likely result in higher revenue collection, relative to other groups. However, a clear rural-urban distinction remains, which supports a higher price increase for urban customers, whose demand is less responsive, than rural customers. In particular, increasing tariffs for large urban users remains a possibility, for its low price response compare to rural users. Finally, providing cash support to low users in exchange for removing the low electricity rates they face, is sound because their demand is the most income elastic of all groups. Nevertheless, it remains clear that rural customers do not translate income bonuses into increased electricity consumption to the same extent as their urban counterparts.

All in all, combining observations of limited progressivity of the rates, in particular those relevant under 200 units/month, and that urban medium to large consumers levels are least price

responsive, confirm there is room for bringing price hikes to these consumers. In practice, this results in tariff design with narrower and/or higher blocks. Reducing the block width (frequency at which price increases along the usage ladder) allows the subsidized rates to be redistributed to fewer households. However, further price hikes beyond the level of 200 monthly units are of little importance, as this concerns only a minority of users.

5.6. Conclusion

In most Indian states, the electricity pricing system is designed in increasing block prices (IBP), which are known to be highly inefficient at targeting poor households as large amounts of subsidy leak toward rich and urban households. In the last two decades, policy changes have called for lifting the low residential prices, which is to be compensated by direct cash transfers to poor households.

The recent addition of many small rural users due to rural electrification progress, who pay smaller marginal rates under the IBP, calls for an examination of the potential effects of these changes on the new customer base. Accounting for progressivity in the block rates, this study estimates demand response to both average and marginal price, as well as income response, along the urban-rural divide and for various customer sizes. The results show that the proposed changes would affect small and rural customers worse than their richer and urban counterparts, due to higher price response, lower substitution toward other fuels and higher income elasticity.

An analysis of the existing progressiveness reveals it is rather limited for users under 200 units per month, except when 'below poverty line' rates are in place. This observation, combined with the heterogeneity in price response highlighted in this study, suggests increasing rates for urban and for medium to large users (50 to 200 monthls unit) in priority. Incidentally, the results support keeping or implementing separate urban-rural tariff schedules. Maintaining or expanding the practice of separate tariffs schemes would not solve the problem of subsidy leakage, which would disappear only with flat tariffs, but could help reserve the low unit rates to users who need it the most, and target price increase to groups that would generate the highest revenue collection.

This study is the first to show the extent of the bias arising from using average prices instead of marginal price in the estimations of elasticities for India. Overall, responses to marginal price are higher, in absolute value, than response to average price, by 0.06 to 0.09. Looking forward,

the response to marginal price is likely to become more relevant with the advent of smart meters and the development of smart grids, which produce real-time information and enable users to monitor and adjust their consumption more precisely to the price structure they face.

CHAPTER 6

GENERAL CONCLUSION

6.1. Summary

Despite great improvement to reduce worldwide extreme poverty in the last 20 years, the increasing difficulties in accessing basic commodities due to the recent Covid-19 pandemic and the war in Ukraine, on top of the growing effects of climate change, remind us that gains in living standards are fragile. Food and energy in particular are strongly affected by major escalation of prices and difficult provision, which threatens food security and transition to modern energy use for vulnerable population groups of developing countries.

Given these conditions, how can policies maintain households' access to sufficient basic commodities and secure decent living standards? There is no general answer to whether in-kind transfers, direct income transfers or other forms of support are most appropriate to lift quality of life, but it is clear that their effectiveness depends on the recipients' response to these interventions. Naturally, the forms of support that fit the recipients' own needs and resource level are more likely to reach the expected goals, but this requires an understanding of households' preferences and their strategies to fulfill basic needs. To begin answering this, four essays of this thesis examine how modest-income households adapt their perception of basic needs and their consumption of food and energy to small increases in income.

Specifically, the first essay (chapter 2) shows how rising income modifies the households' resource utilization and arbitrage between market-purchased and home-produced food. The second essay (chapter 3) deals with how people define basic needs and poverty at different stages of income, and aims to improve the targeting of subjectively poor households. The last two essays highlight some obstacles to households' use of electricity to meet their basic energy needs, in place of traditional and harmful solid fuels. Specifically, the third essay (chapter 4) highlights income and non-income barriers faced by rural and ethnic minority households, while the fourth essay (chapter 5) analyzes the suitability of income support compared to subsidies.

The results from these analyses provide some theoretical and empirical contributions that support policy making as well as research on the choice and design of income and non-income forms of public support toward households.

6.2. Contributions

6.2.1. Main findings and policy contributions

The preference of households for own-products over market-purchased foods has not been studied thoroughly before, and consumption of own-grown products was mostly regarded as a side-result of subsistence farming orientation. Therefore, the share of own products in a diet was expected to decrease when farmers adopt market-oriented farming and their income grows. However, the first essay proposes a theoretical framework analyzing households' arbitrage between the two sources of food under constraints of satisfying basic needs and market imperfection. The framework predicts a nonlinear relationship between the share of home-grown foods in the diet and total expenditure, in form of an inverse U-shape. The empirical results show that households whose main activity is farming do turn to market-purchased food when their resources grow, which indicates that nutritional support in form of in-kind food would likely be misplaced. On the contrary, there is a clear difference in basic needs strategy for households whose main activity is not farming, who tend to increase food from own procurement when their resources move from a very low stage to a better-off stage. Policy design for food support should be aware of this.

The second essay is concerned with the shift from widespread extreme poverty to other forms of more subtle poverty in China and how the government can best support households in the challenges of their newfound economic situation. This relies first on an ability to recognize occurrence of subjective poverty, if possible widely and frequently. In this perspective, the essay aims at predicting which households perceive themselves as poor, and what are the factors associated with subjective poverty. It results that subjective poverty can be fairly well predicted, by focusing on a set of both income and non-income indicators. Therefore, certain non-income factors that contribute to the feeling of poverty, essentially large health expenditure, could be the target of support measures by the government.

The third essay contributes to small but growing literature on race and energy justice. The ethnic difference in electricity consumption are investigated for the case of Vietnam, where, at same

income level, ethnic minorities continue to consume lower amounts of electricity despite general access to the grid. In particular, ethnicity is found to affect electricity consumption indirectly, via income. In particular, increases in income do not allow members of ethnic minorities to increase their electricity consumption as much as members of the ethnic dominant group. This points to the existence of non-income barriers that are yet to be uncovered. In this case, supporting the energy transition of ethnic minorities toward modern fuels with income support would not be sufficient. Support policies should also target the non-income barriers that prevent ethnic minorities from increasing their electricity consumption as much as other groups do at same levels of income.

The fourth essay aims to reconcile increasing residential electricity prices in India while protecting small and rural customers, in the context of increasing block prices. In particular, it demonstrates that the proposed policy changes to replace subsidized electricity by cash transfers would harm rural users more than urban users, while being less effective at increasing revenues among rural users because of different elasticities. Although differences in usage pattern and price response between the two sectors were known, they were so far not analyzed in the context of block prices. The essay analyzes the diverse forms that increasing block prices take in India and their resulting progressivity. In particular, the essay points to a rather limited progressivity of rates for the range of consumption between 50 and 200 units per month, which would represent an appropriate target to increase rates first. This essay supports the discrimination of rate schedules between the urban and rural sectors.

6.2.2. Methodological contributions

The first essay develops a theoretical framework that expands existing models of household preferences in the context of fulfilling basic needs. The framework shows how market imperfections, basic survival requirements and different food characteristics (calorie content) affect the choice between consumption of own production and market purchases. Furthermore, the essay places the choice of the households in the context of satisfaction of other non-food basic needs (e.g. energy, clothing, education…). As a result, the trade-off that the households must operate between food and non-food needs also affects their preferences for own versus market foods. This results in an inverse U-shape relationship that was not accounted for before in research and analysis of household food consumption behavior. This is followed by a semiparametric analysis of data for India, that confirms the non-linear relationship between consumption of own consumption and income, albeit for certain types of households only. To

account for endogeneity, the methodology adopts a control function, which has been shown to be more appropriate for nonparametric estimations than a two stage least square approach.

The second essay proposes a review of several machine learning techniques to establish whether the subjective poverty, which relates to individual perception of poverty rather than to an objectively-set standard, can be well predicted. The essay compares techniques of support vector machines, random forests, and LASSO regression and finds that the random forest performs the best, with 85.29 percent of correct predictions. Specifically, the analysis offers attention to the modest-income households, who may feel poor but not identified as such by objective poverty lines.

The third essay first employs a machine learning technique of feature selection, the LASSO approach, in preparation of the empirical estimation. Although many socio-demographic, economic, usage and equipment variables are known to affect electricity consumption, it is not clear which factors have the largest influence in each specific context. The LASSO technique allows selecting the variables that contribute the most to the variance of the outcome in the applicable data. The second step of the empirical framework consists in a model where income interacts with ethnicity, in order to capture how ethnicity affects consumption both directly and indirectly. This is implemented together with a mediation analysis, which confirms the indirect effect of ethnicity on consumption via income.

The fourth essay develops an empirical framework that accounts for the structure of block prices in estimating price responses. Using a two-stage budgeting approach, price responses are calculated both with respect to average price, to form a base comparison case, and to marginal prices. The bias that results from the specificities of the block price structure can thus be clearly measured for the first time in a context of a developing country. Accounting for the price structure in block is particularly relevant when estimating the potential effect of a price increase on various types of customers, who face different marginal prices.

6.3. Future considerations

The first essay is concerned with the changes in the amount of own-produced food in the diet at different income levels, but the data indicates that the composition of the own-produced food also changes with income, although it is not examined in details in the present work. Further research on the composition of the own food consumption is necessary to understand the

patterns of these changes with income, and their effects on nutritional outcomes and implications for support programs. This could also complete the theoretical model developed in this work. Additionally, the results show that very low-income non-farming households increase their consumption of own food when they have the capacity to do so, which suggests a preference for in-kind foods. However, more research is needed to establish the best form of support for these households specifically, whether in-kind or income transfers, or access to cultivating possibilities (home garden, shared garden). Finally, the model addresses the arbitrage between conflicting food and non-food basic needs, but the extent to which non-food needs (i.e. the cash constraint) suppresses fulfillment of food needs is not explored in details. Little is known empirically about the strength of the cash constraint and the arbitrage that households operate between various basic commodities.

The second essay showed that subjective poverty for not-so-poor households is associated with major occasional expenditures such as health and gift. Financial pressures arising from social norms, i.e. making substantial gifts during special occasions, are difficult to regulate. However, it is likely that providing the population with health insurance or improving affordability of health care would improve wellbeing among the middle-income class of China. Further research should go in this direction, for example on which recipients should be targeted in priority. A further comment relates to the relevance of the concept of subjective poverty as a standard measurement for policy making. It is likely that some wealthy people continue to feel poor even although they have reached comfortable living standards and met their basic needs from an objective perspective. The concept of subjective poverty nevertheless brings valuable insights for policy making, so poverty standards combining both subjective and objective elements should be further researched. Furthermore, to govern public policies and orient support intervention, a definition of poverty that includes subjective perception with the complex time dimension of poverty, e.g. intergenerational poverty and poverty cycles (Yu and Huang, 2021), should be investigated.

In the context of household energy transition (last two essays), utilities and their regulators play an important role in how electricity is priced and made accessible to vulnerable and/or discriminated households. However, the current cost-of-service business model dominating the electric sector has limitations. Alternative solutions that exist on limited scales have shown some advantages. Privatization and deregulation of the service part of electricity provision (contract, billing and customer service) could lead to better rationalization of tariffs, while not meddling with the supply of electricity. Private players could be more efficient at finding

solutions to balance the ability to pay for service by small users with sufficient revenue collection. Another solution exists with pre-paid electricity cards, present in England and Australia for example. These cards have been linked to poor social outcomes and energy insecurity (Longden et al., 2021). They could nevertheless be used as a transition tool for very low-income households who may not be able to afford a permanent electricity connection that entails monthly fixed costs. The flexibility offered by a pre-paid solution could give households more control over their expenses and exempt them from having to make long commitments. Additional options exist in the form of alternative business models for utilities, better adapted to remote or discriminated communities, and which foster transition to renewable energy (He et al., 2022). All in all, these alternative options present risks and advantages that need to be well investigated before large scale implementation.

Finally, the domains of food and energy have many interactions, one of them discussed in the first essay where food and energy compete for households' budget. A further interaction at the household level lies in the energy used for cooking. Little is known on the changes to household diet and nutritional status resulting from energy becoming expensive or unavailable (Ruel et al., 2010). Furthermore, indoor pollution generated by burning solid fuels have implication for nutrition outcome such as stunting (Islam et al., 2021). Further research is needed on nutritional programs that could improve both sufficiency and quality of food, and also address how it is prepared at the same time.

On the aggregate level, the interface of food and energy domains lies in their reliance on land and water, all under major threat of climate change (D'Odorico et al., 2018). Adoption of agrovoltaic systems is one of the promising solutions for a sustainable management of resources, which requires further research (Barron-Gafford et al., 2019).

BIBLIOGRAPHY

Aaberge, R., Eika, L., Langørgen, A., Mogstad, M., 2019. Local governments, in-kind transfers, and economic inequality. J. Public Econ. 180, 103966. https://doi.org/https://doi.org/10.1016/j.jpubeco.2018.09.015

Acharya, R.H., Sadath, A.C., 2017. Implications of energy subsidy reform in India. Energy Policy 102, 453–462. https://doi.org/10.1016/j.enpol.2016.12.036

Aditya, K.S., Subash, S.P., Praveen, K. V., Nithyashree, M.L., Bhuvana, N., Sharma, A., 2017. Awareness about Minimum Support Price and Its Impact on Diversification Decision of Farmers in India. Asia Pacific Policy Stud. 4, 514–526. https://doi.org/10.1002/app5.197

Agarwal, S., Mani, S., Jain, A., Ganesan, K., 2020. State of Electricity Access in India. Counc. Energy, Environ. Water.

Aghion, P., Bloom, N., Blundell, R., Griffith, R., Howitt, P., 2005. Competition and Innovation: an Inverted-U Relationship. Q. J. Econ. 120, 701–728. https://doi.org/10.1093/qje/120.2.701

Aklin, M., Cheng, C.Y., Urpelainen, J., 2021. Inequality in policy implementation: Caste and electrification in rural India. J. Public Policy 41, 331–359. https://doi.org/10.1017/S0143814X20000045

Alem, Y., Köhlin, G., Stage, J., 2014. The persistence of subjective poverty in urban ethiopia. World Dev. 56, 51–61. https://doi.org/10.1016/j.worlddev.2013.10.017

Alesina, A., Zhuravskaya, E., 2011. Segregation and the Quality of Government in a Cross Section of Countries. Am. Econ. Rev. 101, 1872–1911. https://doi.org/10.1257/aer.101.5.1872

Alkire, S., 2005. Why the Capability Approach? J. Hum. Dev. 6, 115–135. https://doi.org/10.1080/146498805200034275

Alsharkawi, A., Al-fetyani, M., Dawas, M., Saadeh, H., Alyaman, M., 2021. Poverty Classification Using Machine Learning : The Case of Jordan. Sustainability 13, 1–16.

Anand, S., Sen, A., 1997. Concepts of Human Development and Poverty: a Multidimensional Perspective, in: Poverty and Human Development: Human Development Papers. New York: United Nations Development Programme ; Pp. 1-20. pp. 173–190.

Arimah, B., 2004. Poverty Reduction and Human Development in Africa. J. Hum. Dev. 5, 399–415. https://doi.org/10.1080/1464988042000277260

Arimond, M., Ruel, M.T., 2004. Dietary Diversity Is Associated with Child Nutritional Status: Evidence from 11 Demographic and Health Surveys. J. Nutr. 134, 2579–2585. https://doi.org/10.1093/jn/134.10.2579

Arslan, A., Taylor, J.E., 2009. Farmers' subjective valuation of subsistence crops: The case of traditional Maize in Mexico. Am. J. Agric. Econ. 91, 956–972. https://doi.org/10.1111/j.1467-8276.2009.01323.x

Ashagidigbi, W.M., Babatunde, B.A., Ogunniyi, A.I., Olagunju, K.O., Omotayo, A.O., 2020.

Estimation and determinants of multidimensional energy poverty among households in Nigeria. Sustain. 12, 1–13. https://doi.org/10.3390/SU12187332

Awaworyi Churchill, S., Smyth, R., 2017. Ethnic Diversity and Poverty. World Dev. 95, 285–302. https://doi.org/https://doi.org/10.1016/j.worlddev.2017.02.032

Baiphethi, M.N., Jacobs, P.T., 2009. The contribution of subsistence farming to food security in South Africa. Agrekon 48, 459–482. https://doi.org/10.1080/03031853.2009.9523836

Baird, S., Ferreira, F.H.G., Özler, B., Woolcock, M., 2013. Relative Effectiveness of Conditional and Unconditional Cash Transfers for Schooling Outcomes in Developing Countries: A Systematic Review, Campbell Systematic Reviews. https://doi.org/10.4073/csr.2013.8

Baird, S., McIntosh, C., Özler, B., Hampshire, N., Nw, A., 2019. When the money runs out: Do cash transfers have sustained effects on human capital accumulation? J. Dev. Econ. 140, 169–185. https://doi.org/https://doi.org/10.1016/j.jdeveco.2019.04.004

Balasubramanian, S., 2015. Is the PDS Already a Cash Transfer ? Rethinking India ' s Food Subsidy Policies. J. Dev. Stud. 51, 642–659. https://doi.org/10.1080/00220388.2014.997221

Banks, J., Blundell, R., Lewbel, A., 1997. Quadratic Engel curves and Consumer Demand. Rev. Econ. Stat. 79, 527–539. https://doi.org/https://doi.org/10.1162/003465397557015

Barnes, D., Qian, L., 1992. Urban interfuel substitution, energy use, and equity in developing countries: some preliminary results. Int. Issues Energy Policy, Dev. Econ.

Barnes, D.F., Krutilla, K., Hyde, W.F., 2004. The Urban Household Energy Transition. Urban Househ. Energy Transit. https://doi.org/10.4324/9781936331000

Baron, R.M., Kenny, D.A., 1986. The Moderator-Mediator Variable Distinction in Social Psychological Research. Conceptual, Strategic, and Statistical Considerations. J. Pers. Soc. Psychol. 51, 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Barron-Gafford, G.A., Pavao-Zuckerman, M.A., Minor, R.L., Sutter, L.F., Barnett-Moreno, I., Blackett, D.T., Thompson, M., Dimond, K., Gerlak, A.K., Nabhan, G.P., Macknick, J.E., 2019. Agrivoltaics provide mutual benefits across the food–energy–water nexus in drylands. Nat. Sustain. 2, 848–855. https://doi.org/10.1038/s41893-019-0364-5

Baulch, B., Chuyen, T.T.K., Haughton, D., Haughton, J., 2008. Ethnic minority development in Vietnam. J. Dev. Stud. 43, 1151–1176. https://doi.org/10.1080/02673030701526278

Baulch, B., Pham, H.T., Reilly, B., 2012. Decomposing the Ethnic Gap in Rural Vietnam, 1993-2004. Oxford Dev. Stud. 40, 87–117. https://doi.org/10.1080/13600818.2011.646441

Behrman, J.R., Deolalikar, A.B., Wolfe, B.L., 1988. Nutrients: Impacts and determinants. World Bank Econ. Rev. 2, 299–320. https://doi.org/10.1093/wber/2.3.299

Belkin, M., Hsu, D., Ma, S., Mandal, S., 2019. Reconciling modern machine-learning practice and the classical bias–variance trade-off. Proc. Natl. Acad. Sci. U. S. A. 116, 15849–15854. https://doi.org/10.1073/pnas.1903070116

Belloni, Alexandre; Chernozhukov, Victor; Wei, Y., 2013. Honest confidence regions for a regression parameter in logistic regression with a large number of controls (No.

CWP67/13), cemmap working paper. London. https://doi.org/http://dx.doi.org/10.1920/wp.cem.2013.6713

Belloni, A., Chernozhukov, V., 2013. Least squares after model selection in high-dimensional sparse models. Bernoulli 19, 521–547. https://doi.org/10.3150/11-BEJ410

Belloni, A., Chernozhukov, V., Wei, Y., 2016. Post-Selection Inference for Generalized Linear Models With Many Controls. J. Bus. Econ. Stat. 34, 606–619. https://doi.org/10.1080/07350015.2016.1166116

Bernard, T., 2012. Impact Analysis of Rural Electrification Projects in Sub-Saharan Africa. World Bank Res. Obs. 27, 33–51. https://doi.org/10.1093/wbro/lkq008

Bernard, T., Torero, M., 2015. Social Interaction Effects and Connection to Electricity: Experimental Evidence from Rural Ethiopia. Econ. Dev. Cult. Change 63, 459–484. https://doi.org/10.1086/679746

Bhagowalia, P., Chandna, A., 2011. Food Subsidies and Nutritional Status: Evidence from ICRISAT Data. Sel. Pap. Prep. Present. 2016 Agric. Appl. Econ. Assoc. Boston, MA, July 31-August 2.

Bhagowalia, P., Headey, D., Kadiyala, S., 2012. Agriculture, Income, and Nutrition Linkages in India. IFPRI Rep.

Bhattacharyya, R., Ganguly, A., 2017. Cross subsidy removal in electricity pricing in India. Energy Policy 100, 181–190. https://doi.org/10.1016/j.enpol.2016.10.024

Bhattacharyya, S.C., 2006. Energy access problem of the poor in India : Is rural electrification a remedy ? Energy Policy 34, 3387–3397. https://doi.org/10.1016/j.enpol.2005.08.026

Bhattacharyya, S.C., Timilsina, G.R., 2010. Modelling energy demand of developing countries: Are the specific features adequately captured? Energy Policy 38, 1979–1990. https://doi.org/10.1016/j.enpol.2009.11.079

Bhuvandas, D., Gundimeda, H., 2020. Welfare impacts of transport fuel price changes on Indian households: An application of LA-AIDS model. Energy Policy 144, 111583. https://doi.org/https://doi.org/10.1016/j.enpol.2020.111583

Billings, R.B., 1982. Specification of Block Rate Price Variables in Demand Models. Land Econ. 58, 386. https://doi.org/10.2307/3145946

Bishop, J.A., Luo, F., Pan, X., 2006. Economic transition and subjective poverty in urban China. Rev. Income Wealth 52, 625–641. https://doi.org/10.1111/j.1475-4991.2006.00209.x

Blackorby, B.C., Donaldson, D., 1988. Cash versus Kind , Self-Selection , and Efficient Transfers. Am. Econ. Rev. 78, 691–700.

Blundell, R., Powell, J.L., 2011. Endogeneity in Nonparametric and Semiparametric Regression Models (Chapt 8). Dev. Econom. 193–214. https://doi.org/10.1002/9781119954231.ch5

Boland, J.J., Whittington, D., 1998. The Political Economy of Increasing Block Tariffs in Developing Countries. EEPSEA Spec. Tech. Pap. 15.

Borenstein, S., 2009. To what electricity price do consumers respond? Residential demand elasticity under increasing-block pricing. Prelim. Draft April 1–37.

Boto Ferreira, M., Costa Pinto, D., Maurer Herter, M., Soro, J., Vanneschi, L., Castelli, M., Peres, F., 2021. Using artificial intelligence to overcome over-indebtedness and fight poverty. J. Bus. Res. 131, 411–425. https://doi.org/https://doi.org/10.1016/j.jbusres.2020.10.035

Brady, D., 2003. Rethinking the sociological measurement of poverty. Soc. Forces 81, 715–752. https://doi.org/10.1353/sof.2003.0025

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brockway, A.M., Conde, J., Callaway, D., 2021. Inequitable access to distributed energy resources due to grid infrastructure limits in California. Nat. Energy 6, 892–903. https://doi.org/10.1038/s41560-021-00887-6

Browne, C., Matteson, D.S., McBride, L., Hu, L., Liu, Y., Sun, Y., Wen, J., Barrett, C.B., 2021. Multivariate random forest prediction of poverty and malnutrition prevalence. PLoS One 16, 1–23. https://doi.org/10.1371/journal.pone.0255519

Buckley, T., 2015. India's Electricity-Sector Transformation, Institute for Energy Economics and Financial Analysis.

Burchi, F., Scarlato, M., D'Agostino, G., 2018. Addressing Food Insecurity in Sub-Saharan Africa The Role of Cash Transfers. Poverty and Public Policy 10. https://doi.org/https://doi.org/10.1002/pop4.233

Buse, A., 1994. Evaluating the Linearized Almost Ideal Demand System. Am. J. Agric. Econ. 76, 781–793. https://doi.org/10.2307/1243739

Cardenas, H., Whittington, D., 2019. The consequences of increasing block tariffs on the distribution of residential electricity subsidies in Addis Ababa, Ethiopia. Energy Policy 128, 783–795. https://doi.org/https://doi.org/10.1016/j.enpol.2019.01.033

Carletto, C., Corral, P., Guelfi, A., 2017. Agricultural commercialization and nutrition revisited: Empirical evidence from three African countries. Food Policy 67, 106–118. https://doi.org/10.1016/j.foodpol.2016.09.020

Chang, S.C., 2014. Effects of financial developments and income on energy consumption. Int. Rev. Econ. Financ. 35, 28–44. https://doi.org/10.1016/j.iref.2014.08.011

Cheng, Z., Guo, L., Smyth, R., Tani, M., 2021. Childhood Adversity and Energy Poverty. https://doi.org/Available at SSRN: https://ssrn.com/abstract=3951000 or http://dx.doi.org/10.2139/ssrn.3951000

Chernozhukov, V., Hansen, C., Spindler, M., 2016. High-dimensional metrics in R. arXiv:1603.01700 1–35.

Chhotray, V., Adhikari, A., Bahuguna, V., 2020. The political prioritization of welfare in India: Comparing the public distribution system in Chhattisgarh and Jharkhand. World Dev. 128, 104853. https://doi.org/https://doi.org/10.1016/j.worlddev.2019.104853

Chindarkar, N., Goyal, N., 2019. One price doesn't fit all: An examination of heterogeneity in price elasticity of residential electricity in India. Energy Econ. 81, 765–778. https://doi.org/https://doi.org/10.1016/j.eneco.2019.05.021

Choumert-Nkolo, J., Combes Motel, P., Le Roux, L., 2019. Stacking up the ladder: A panel data analysis of Tanzanian household energy choices. World Dev. 115, 222–235. https://doi.org/https://doi.org/10.1016/j.worlddev.2018.11.016

Christiaensen, L., Lanjouw, P., Luoto, J., Stifel, D., 2012. Small area estimation-based prediction methods to track poverty: Validation and applications. J. Econ. Inequal. 10, 267–297. https://doi.org/10.1007/s10888-011-9209-9

Chunekar, A., Sreenivas, A., 2019. Towards an understanding of residential electricity consumption in India. Build. Res. Inf. 47, 75–90. https://doi.org/(http://dx.doi.org/10.1080/09613218.2018.1506383

Churchill, A.S., Smyth, R., 2020. Ethnic diversity, energy poverty and the mediating role of trust: Evidence from household panel data for Australia. Energy Econ. 86, 104663. https://doi.org/https://doi.org/10.1016/j.eneco.2020.104663

Coase, R.H., 1946. The Marginal Cost Controversy. Economica 13, 169–182. https://doi.org/10.2307/2549764

Cooper, J.E., Benmarhnia, T., Koski, A., King, N.B., 2020. Cash transfer programs have differential effects on health: A review of the literature from low and middle-income countries. Soc. Sci. Med. 247, 112806. https://doi.org/https://doi.org/10.1016/j.socscimed.2020.112806

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/BF00994018

Cunha, J.M., De Giorgi, G., Jayachandran, S., 2019. The Price Effects of Cash Versus In-Kind Transfers. Rev. Econ. Stud. 86, 240–281. https://doi.org/10.1093/restud/rdy018

Currie, J., Gahvari, F., 2008. Transfers in cash and in-kind: Theory meets the data. J. Econ. Lit. 46, 333–383. https://doi.org/10.1257/jel.46.2.333

D'Odorico, P., Davis, K.F., Rosa, L., Carr, J.A., Chiarelli, D., Dell'Angelo, J., Gephart, J., MacDonald, G.K., Seekell, D.A., Suweis, S., Rulli, M.C., 2018. The Global Food-Energy-Water Nexus. Rev. Geophys. 56, 456–531. https://doi.org/10.1029/2017RG000591

Dang, H.A., 2012. Vietnam: A widening poverty gap for ethnic minorities. Indig. Peoples, Poverty, Dev. 304–343. https://doi.org/10.1017/CBO9781139105729.008

Das, H.P., Konstantakopoulos, I.C., Manasawala, A.B., Veeravalli, T., Liu, H., Spanos, C.J., 2019. A novel graphical lasso based approach towards segmentation analysis in energy game-theoretic frameworks. Proc. - 18th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2019 1702–1709. https://doi.org/10.1109/ICMLA.2019.00277

de Janvry, A., Sadoulet, E., 2011. Subsistence farming as a safety net for food-price shocks. Dev. Pract. 21, 472–480. https://doi.org/10.1080/09614524.2011.561292

de Vos, K., Garner, T.I., 1991. An Evaluation of Subjective Poverty Definitions : Comparing Results From The U.S. and The Netherlands. Reveiw of Income and Wealth Series 37.

Deaton, A., 2010. Price Indexes, Inequality, and the Measurement of World Poverty. Am. Econ. Rev. 100, 5–34.

Deaton, A., Drèze, J., 2009. Food and nutrition in India: Facts and interpretations. Econ. Polit.

Wkly. 44, 42–65.

Deaton, A., Dupriez, O., 2011. Purchasing Power Parity Exchange Rates for the Global Poor. Am. Econ. J. Appl. Econ. 3, 137–166. https://doi.org/10.1257/app.3.2.137

Deaton, A., Muellbauer, J., 1980. An Almost Ideal Demand System. Am. Econ. Rev. 70, 312–326.

Deaton, A., Subramanian, S., 1996. The Demand for Food and Calories. J. Polit. Econ. 104, 133–162.

Dong, X.-Y., Hao, Y., 2018. Would income inequality affect electricity consumption? Evidence from China. Energy 142, 215–227. https://doi.org/https://doi.org/10.1016/j.energy.2017.10.027

Dowd, J., 1989. The urban energy transition and interfuel substitution in developing countries: a review of the literature. World Bank. Unpubl. Rep.

Drehobl, A., Ross, L., 2016. Lifting the High Energy Burden in America's Largest Cities: How Energy Efficiency Can Improve Low Income and Underserved Communities.

Dubois, U., 2012. From targeting to implementation : The role of identification. Energy Policy 49, 107–115. https://doi.org/10.1016/j.enpol.2011.11.087

Easterlin, R.A., 1974. Does Economic Growth Improve the Human Lot? Some Empirical Evidence, in: DAVID, P.A., REDER, M.W. (Eds.), Nations and Households in Economic Growth. Academic Press, pp. 89–125. https://doi.org/https://doi.org/10.1016/B978-0-12-205050-3.50008-7

Easterlin, R.A., Connor, K.J.O., 2020. The Easterlin Paradox (No. 13923), Discussion Paper Series.

Ekholm, T., Krey, V., Pachauri, S., Riahi, K., 2010. Determinants of household energy consumption in India. Energy Policy 38, 5696–5707. https://doi.org/https://doi.org/10.1016/j.enpol.2010.05.017

Electricity of Vietnam Power Design Center, 2005. Ethnic Minority Development Plan (EMDP).

Embaye, W., Bergtold, J., Schwab, B., Zereyesus, Y.A., 2018. Modeling Farm Productivity for Heterogeneous Farm Households in developing countries, 2018 Annual Meeting, August 5-7, Washington, D.C. from Agricultural and Applied Economics Association. https://doi.org/10.22004/ag.econ.274226

Evenson, R.E., 1978. Time Allocation in Rural Philippine Households. Am. J. Agric. Econ. 60, 322–330. https://doi.org/10.2307/1240068

Fafchamps, M., 1992. Cash Crop Production, Food Price Volatility, and Rural Market Integration in the Third World. Am. J. Agric. Econ. 74, 90–99. https://doi.org/10.2307/1242993

Fang, P., Dong, S., Xiao, J., Liu, C., Feng, X., Wang, Y., 2010. Regional inequality in health and its determinants: Evidence from China. Health Policy (New. York). 94, 14–25. https://doi.org/https://doi.org/10.1016/j.healthpol.2009.08.002

Farrington, J., Slater, R., 2006. Introduction: Cash transfers: Panacea for poverty reduction or money down the drain? Dev. Policy Rev. 24, 499–511. https://doi.org/10.1111/j.1467-7679.2006.00344.x

Fauziah, M.A.T., Ruliana, 2022. Comparison of k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) Methods for Classification of Poverty Data in Papua. ARRUS J. Math. Appl. Sci. 2, 83–91.

Feeny, S., Trinh, T.-A., Zhu, A., 2021. Temperature shocks and energy poverty: Findings from Vietnam. Energy Econ. 99, 105310. https://doi.org/https://doi.org/10.1016/j.eneco.2021.105310

Filippini, M., Pachauri, S., 2004. Elasticities of electricity demand in urban Indian households. Energy Policy 32, 429–436. https://doi.org/10.1016/S0301-4215(02)00314-2

Finkelshtain, I., Chalfant, J.A., 1991. Marketed Surplus under Risk: Do Peasants Agree with Sandmo? Am. J. Agric. Econ. 73, 557–567. https://doi.org/10.2307/1242809

Fiszbein, A., Schady, N.R., 2009. Conditional cash transfers: reducing present and future poverty. World Bank Publications.

Fitzgerald, K.B., Barnes, D., McGranahan, G., 1990. Interfuel substitution and changes in the way households use energy: the case of cooking and lighting behavior in urban Java. The World Bank.

Flik, R.J., van Praag, B.M.S., 1991. Subjective poverty line definitions. Economist (Leiden). 139, 311–330. https://doi.org/10.1007/BF01423569

Frelat, R., Lopez-Ridaura, S., Giller, K.E., Herrero, M., Douxchamps, S., Djurfeldt, A.A., Erenstein, O., Henderson, B., Kassie, M., Paul, B.K., Rigolot, C., Ritzema, R.S., Rodriguez, D., Asten, P.J.A. van, Wijk, M.T. van, 2016. Drivers of household food availability in sub-Saharan Africa based on big data from small farms. Proc. Natl. Acad. Sci. 113, 458–463. https://doi.org/10.1073/PNAS.1518384112

Gang, I.N., Sen, K., Yun, M.-S., 2002. Caste, Ethnicity and Poverty in Rural India, Departmental Working Papers 200225, Rutgers University, Department of Economics. https://doi.org/10.2139/ssrn.358160

García, S., Saavedra, J.E., 2017. Educational Impacts and Cost-Effectiveness of Conditional Cash Transfer Programs in Developing Countries: A Meta-Analysis. Rev. Educ. Res. 87, 921–965. https://doi.org/10.3102/0034654317723008

Gaur, V., Gupta, E., 2016. The determinants of electricity theft: An empirical analysis of Indian states. Energy Policy 93, 127–136. https://doi.org/10.1016/j.enpol.2016.02.048

Gelan, A., 2006. Cash or food aid? A general equilibrium analysis for Ethiopia. Dev. Policy Rev. 24, 601–624. https://doi.org/10.1111/j.1467-7679.2006.00350.x

Gentilini, U., 2016. Revisiting the “cash versus food” debate: New evidence for an old puzzle? World Bank Res. Obs. 31, 135–167. https://doi.org/10.1093/wbro/lkv012

George, N.A., McKay, F.H., 2019. The public distribution system and food security in India. Int. J. Environ. Res. Public Health 16. https://doi.org/10.3390/ijerph16173221

Gertler, P.J., Shelef, O., Wolfram, C.D., Fuchs, A., 2016. The Demand for Energy-Using Assets

among the World's Rising Middle Classes. Am. Econ. Rev. 106, 1366–1401.

Ghatak, M., Kumar, C., Mitra, S., 2016. Cash versus Kind. Econ. Polit. Wkly. 51–60.

Gillespie, S., Harris, J., Kadiyala, S., 2012. The Agriculture-Nutrition Disconnect in India What Do We Know? IFPRI Discuss. Pap. 1187 1–56.

Goedhart, T., Halberstadt, V., Kapteyn, A., van Praag, B., 1977. The Poverty Line : Concept and Measurement. J. Hum. Resour. 12, 503–520.

Goldemberg, B.J., 1990. One kilowatt per capita. Bull. At. Sci. 46, 13–15.

Gorman, W.M., 1959. Separable Utility and Aggregation. Econometrica 27, 469–481. https://doi.org/https://doi.org/10.2307/1909472

Gundimeda, H., Köhlin, G., 2008. Fuel demand elasticities for energy and environmental policies: Indian sample survey evidence. Energy Econ. 30, 517–546. https://doi.org/10.1016/j.eneco.2006.10.014

Guo, Y., Zhou, Y., Liu, Y., 2019. Targeted poverty alleviation and its practices in rural China: A case study of Fuping county, Hebei Province. J. Rural Stud. https://doi.org/10.1016/j.jrurstud.2019.01.007

Gupta, S., Sunder, N., Pingali, P.L., 2020. Market Access, Production Diversity, and Diet Diversity: Evidence From India. Food Nutr. Bull. 41, 167–185. https://doi.org/10.1177/0379572120920061

Gustafsson, B., Shi, L., Sato, H., 2004. Can a subjective poverty line be applied to China? Assessing poverty among urban residents in 1999. J. Int. Dev. 16, 1089–1107. https://doi.org/10.1002/jid.1127

Guta, D.D., 2012. Application of an almost ideal demand system (AIDS) to Ethiopian rural residential energy use: Panel data evidence. Energy Policy 50, 528–539. https://doi.org/10.1016/j.enpol.2012.07.055

Han, H., Wu, S., Zhang, Z., 2018. Factors underlying rural household energy transition: A case study of China. Energy Policy 114, 234–244.

Handa, S., Daidone, S., Peterman, A., Davis, B., Pereira, A., Palermo, T., Yablonski, J., 2018. Myth-busting? Confronting six common perceptions about unconditional cash transfers as a poverty reduction strategy in Africa. World Bank Res. Obs. 33, 259–298. https://doi.org/10.1093/wbro/lky003

Hanna, R., Oliva, P., 2015. Moving up the energy ladder: the effect of an increase in economic well-being on the fuel consumption choices of the poor in India. Am. Econ. Rev. 105, 242–246.

Hasiner, E., Yu, X., 2016. Meat consumption and democratic governance: A cross-national analysis. China Econ. Rev. 59, 100950. https://doi.org/10.1016/j.chieco.2016.06.008

Haushofer, J., Shapiro, J., 2016. The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya. Q. J. Econ. 131, 1973–2042. https://doi.org/10.1093/qje/qjw025

Hausmann, J.A., Kinnucan, M., McFaddden, D., 1979. A two-level electricity demand model:

Evaluation of the connecticut time-of-day pricing test. J. Econom. 10, 263–289. https://doi.org/10.1016/0304-4076(79)90085-X

He, E., Brown, G., Lovekin, D., 2022. Transforming the Utility Business Model: options to improve services and opportunities for clean energy in remote communities. The Pembina Institute, Calgary.

Headey, D., Chiu, A., Kadiyala, S., 2012. Agriculture's role in the Indian enigma: Help or hindrance to the crisis of undernutrition? Food Secur. 4, 87–102. https://doi.org/10.1007/s12571-011-0161-0

Headey, D., Ecker, O., 2013. Rethinking the measurement of food security: From first principles to best practice. Food Secur. 5, 327–343. https://doi.org/10.1007/s12571-013-0253-0

Heath, A., Cheung, S.Y., 2007. The Comparative Study of Ethnic Minority Disadvantage. Unequal Chances Ethn. Minor. West. Labour Mark. https://doi.org/10.5871/bacad/9780197263860.003.0001

Henson, S.E., 1984. Electricity Demand Estimates under Increasing-Block Rates. South. Econ. J. 51, 147. https://doi.org/10.2307/1058328

Herrera, J., Razafindrakoto, M., Roubaud, F., 2006a. The determinants of subjective poverty: A comparative analysis between Madagascar and Peru, Développement Institutions and Analyses de Long Terme.

Herrera, J., Razafindrakoto, M., Roubaud, F., 2006b. The determinants of subjective poverty: A comparative analysis between Madagascar and Peru, Développement Institutions and Analyses de Long Terme.

Hidrobo, M., Hoddinott, J., Peterman, A., Margolies, A., Moreira, V., 2014. Cash, food, or vouchers? Evidence from a randomized experiment in northern Ecuador. J. Dev. Econ. 107, 144–156. https://doi.org/10.1016/j.jdeveco.2013.11.009

Hoddinott, J., Maluccio, J.A., Behrman, J.R., Flores, R., Martorell, R., 2008. Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults. Lancet 371, 411–416. https://doi.org/10.1016/S0140-6736(08)60205-6

Hoffmann, V., Gatobu, K.M., 2014. Growing their own: Unobservable quality and the value of self-provisioning. J. Dev. Econ. 106, 168–178. https://doi.org/10.1016/j.jdeveco.2013.08.006

Hofmann, M., 2006. Support Vector Machines—Kernels and the Kernel Trick. Univ. Bamb. 1–16.

Hosier, R.H., Dowd, J., 1987. Household fuel choice in Zimbabwe: an empirical test of the energy ladder hypothesis. Resour. energy 9, 347–361.

Houthakker, H.S., 1951. Electricity Tariffs in Theory and Practice. Econ. J. 61, 1–25.

Houthakker, H.S., Verleger, P.K., Sheehan, D.P., 1974. Dynamic Demand Analyses for Gasoline and Residential Electricity. Am. J. Agric. Econ. 56, 412–418. https://doi.org/10.2307/1238776

Htet, N.L., Kongprawechnon, W., Thajchayapong, S., Isshiki, T., 2021. Machine Learning Approach with Multiple Open-source Data for Mapping and Prediction of Poverty in

Myanmar. 18th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol.

Hung, M.F., Huang, T.H., 2015. Dynamic demand for residential electricity in Taiwan under seasonality and increasing-block pricing. Energy Econ. 48, 168–177. https://doi.org/10.1016/j.eneco.2015.01.010

IEA, 2021. India Energy Outlook 2021. India Energy Outlook 2021. https://doi.org/10.1787/ec2fd78d-en

ILOSTAT, 2017. Viet Nam - Household Living Standard Survey 2010.

Imai, K.S., Gaiha, R., Kang, W., 2011. Poverty, inequality and ethnic minorities in Vietnam. Int. Rev. Appl. Econ. 25, 249–282. https://doi.org/10.1080/02692171.2010.483471

Islam, S., Rana, M.J., Mohanty, S.K., 2021. Cooking, smoking, and stunting: Effects of household air pollution sources on childhood growth in India. Indoor Air 31, 229–249. https://doi.org/10.1111/ina.12730

Ismail, Z., Khembo, P., 2015. Determinants of energy poverty in South Africa. J. Energy South. Africa 26, 66–78.

Ito, K., 2014. Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing. Am. Econ. Rev. 104, 537–563. https://doi.org/10.1257/aer.104.2.537

Jain, A., Tripathi, S., Mani, S., Patnaik, S., Shahidi, T., Ganesan, K., 2018. Access to clean cooking energy and electricity: Survey of States 2018. CEEW, New Delhi.

Jain, R.K., Damoulas, T., Kontokosta, C.E., 2014. Towards Data-Driven Energy Consumption Forecasting of Multi-Family Residential Buildings: Feature Selection via The Lasso. 2014 Int. Conf. Comput. Civ. Build. Eng. 1675–1682.

Jamil, F., Ahmad, E., 2014. An Empirical Study of Electricity Theft from Electricity Distribution Companies in Pakistan. Pak. Dev. Rev. 53, 239–254.

Janvry, A. de, Fafchamps, M., Sadoulet, E., 1991. Peasant Household Behaviour with Missing Markets : Some Paradoxes Explained. Econ. J. 101, 1400–1417. https://doi.org/https://doi.org/10.2307/2234892

Jayasinghe, M., Selvanathan, E.A., Selvanathan, S., 2021. Energy poverty in Sri Lanka. Energy Econ. 101, 105450. https://doi.org/10.1016/j.eneco.2021.105450

Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science (80-.). 353, 790–794. https://doi.org/10.1126/science.aaf7894

Jenkins, K., McCauley, D., Heffron, R., Stephan, H., Rehner, R., 2016. Energy justice: A conceptual review. Energy Res. Soc. Sci. 11, 174–182. https://doi.org/10.1016/j.erss.2015.10.004

Jensen, R.T., Miller, N.H., 2010. A revealed preference approach to measuring hunger and undernutrition. NBER. https://doi.org/10.3386/w16555

Jensen, R.T., Miller, N.H., 2008. Giffen behavior and subsistence consumption. Am. Econ. Rev. 98, 1553–1577. https://doi.org/10.1257/aer.98.4.1553

Jones, A.D., 2017. On-farm crop species richness is associated with household diet diversity

and quality in subsistence- and market-oriented farming households in Malawi. J. Nutr. 147, 86–96. https://doi.org/10.3945/jn.116.235879

Kadiyala, S., Harris, J., Headey, D., Yosef, S., Gillespie, S., 2014. Agriculture and nutrition in India: Mapping evidence to pathways. Ann. N. Y. Acad. Sci. 1331, 43–56. https://doi.org/10.1111/nyas.12477

Kaicker, N., Gaiha, R., 2013. Calorie thresholds and undernutrition in India. J. Policy Model. 35, 271–288. https://doi.org/10.1016/j.jpolmod.2012.04.002

Kakwani, N., 1993. Statistical Inference in the Measurement of Poverty. Rev. Econ. Stat. 75, 632–639. https://doi.org/https://doi.org/10.2307/2110016

Kapteyn, A., Kooreman, P., Willemse, R., 1988. Some Methodological Issues in the Implementation of Subjective Poverty Definitions. J. Hum. Resour. 23, 222. https://doi.org/10.2307/145777

Khandker, S.R., Barnes, D.F., Samad, H.A., 2012. Are the energy poor also income poor? Evidence from India. Energy Policy 47, 1–12. https://doi.org/10.1016/j.enpol.2012.02.028

Khanna, N.Z., Guo, J., Zheng, X., 2016. Effects of demand side management on Chinese household electricity consumption: Empirical findings from Chinese household survey. Energy Policy 95, 113–125. https://doi.org/10.1016/j.enpol.2016.04.049

Khera, R., 2014. Cash vs. in-kind transfers: Indian data meets theory. Food Policy 46, 116–128. https://doi.org/https://doi.org/10.1016/j.foodpol.2014.03.009

Khosla, R., Agarwal, A., Sircar, N., Chatterjee, D., 2021. The what, why, and how of changing cooling energy consumption in India's urban households. Environ. Res. Lett. 16. https://doi.org/10.1088/1748-9326/abecbc

Khurana, M., Banerjee, S.G., 2015. Beyond Crisis: The financial performance of India's power sector. World Bank Group.

Kirimi, L., Gitau, R., Olunga, M., 2013. Household food security and commercialization among smallholder farmers in Kenya. 4th Int. Conf. African Assoc. Agric. Econ. Hammamet, Tunis.

Koppmair, S., Kassie, M., Qaim, M., 2017. Farm production, market access and dietary diversity in Malawi. Public Health Nutr. 20, 325–335. https://doi.org/10.1017/S1368980016002135

Krishnamurthy, P., Pathania, V., Tandon, S., 2017. Food price subsidies and nutrition: Evidence from state reforms to India's public distribution system. Econ. Dev. Cult. Change 66, 55–90. https://doi.org/10.1086/694033

Kshirsagar, V., Wieczorek, J., Ramanathan, S., Wells, R., 2017. Household poverty classification in data-scarce environments : a machine learning approach. arXiv:1711.06813.

Leach, G., 1992. The energy transition. Energy Policy 20, 116–123.

Lee, K., Miguel, E., Wolfram, C., 2016. Experimental evidence on the demand for and costs of rural electrification (No. 22292), NBER WORKING PAPER SERIES.

Leser, C.. E.. V.., 1963. Forms of Engel Functions. Econometrica 31, 694–703.

Lesmeister, C., 2015. Mastering machine learning with R : master machine learning techniques with R to deliver insights for complex projects.

Li, G., Cai, Z., Liu, X., Liu, J., Su, S., 2019. A comparison of machine learning approaches for identifying high-poverty counties: robust features of DMSP/OLS night-time light imagery. Int. J. Remote Sens. 40, 5716–5736. https://doi.org/10.1080/01431161.2019.1580820

Li, G., Cai, Z., Qian, Y., Chen, F., 2021. Identifying Urban Poverty Using High-Resolution Satellite Imagery and Machine Learning Approaches: Implications for Housing Inequality. Land 10. https://doi.org/https://doi.org/10.3390/land10060648

Lim, S.S., Vos, T., Flaxman, A.D., Danaei, G., Shibuya, K., Adair-Rohani, H., Amann, M., Anderson, H.R., Andrews, K.G., Aryee, M., Atkinson, C., Bacchus, L.J., Bahalim, A.N., Balakrishnan, K., Balmes, J., Barker-Collo, S., Baxter, A., Bell, M.L., Blore, J.D., Blyth, F., Bonner, C., Borges, G., Bourne, R., Boussinesq, M., Brauer, M., Brooks, P., Bruce, N.G., Brunekreef, B., Bryan-Hancock, C., Bucello, C., Buchbinder, R., Bull, F., Burnett, R.T., Byers, T.E., Calabria, B., Carapetis, J., Carnahan, E., Chafe, Z., Charlson, F., Chen, H., Chen, J.S., Cheng, A.T.A., Child, J.C., Cohen, A., Colson, K.E., Cowie, B.C., Darby, S., Darling, S., Davis, A., Degenhardt, L., Dentener, F., Des Jarlais, D.C., Devries, K., Dherani, M., Ding, E.L., Dorsey, E.R., Driscoll, T., Edmond, K., Ali, S.E., Engell, R.E., Erwin, P.J., Fahimi, S., Falder, G., Farzadfar, F., Ferrari, A., Finucane, M.M., Flaxman, S., Fowkes, F.G.R., Freedman, G., Freeman, M.K., Gakidou, E., Ghosh, S., Giovannucci, E., Gmel, G., Graham, K., Grainger, R., Grant, B., Gunnell, D., Gutierrez, H.R., Hall, W., Hoek, H.W., Hogan, A., Hosgood, H.D., Hoy, D., Hu, H., Hubbell, B.J., Hutchings, S.J., Ibeanusi, S.E., Jacklyn, G.L., Jasrasaria, R., Jonas, J.B., Kan, H., Kanis, J.A., Kassebaum, N., Kawakami, N., Khang, Y.H., Khatibzadeh, S., Khoo, J.P., Kok, C., Laden, F., Lalloo, R., Lan, Q., Lathlean, T., Leasher, J.L., Leigh, J., Li, Y., Lin, J.K., Lipshultz, S.E., London, S., Lozano, R., Lu, Y., Mak, J., Malekzadeh, R., Mallinger, L., Marcenes, W., March, L., Marks, R., Martin, R., McGale, P., McGrath, J., Mehta, S., Mensah, G.A., Merriman, T.R., Micha, R., Michaud, C., Mishra, V., Hanafiah, K.M., Mokdad, A.A., Morawska, L., Mozaffarian, D., Murphy, T., Naghavi, M., Neal, B., Nelson, P.K., Nolla, J.M., Norman, R., Olives, C., Omer, S.B., Orchard, J., Osborne, R., Ostro, B., Page, A., Pandey, K.D., Parry, C.D.H., Passmore, E., Patra, J., Pearce, N., Pelizzari, P.M., Petzold, M., Phillips, M.R., Pope, D., Pope, C.A., Powles, J., Rao, M., Razavi, H., Rehfuess, E.A., Rehm, J.T., Ritz, B., Rivara, F.P., Roberts, T., Robinson, C., Rodriguez-Portales, J.A., Romieu, I., Room, R., Rosenfeld, L.C., Roy, A., Rushton, L., Salomon, J.A., Sampson, U., Sanchez-Riera, L., Sanman, E., Sapkota, A., Seedat, S., Shi, P., Shield, K., Shivakoti, R., Singh, G.M., Sleet, D.A., Smith, E., Smith, K.R., Stapelberg, N.J.C., Steenland, K., Stöckl, H., Stovner, L.J., Straif, K., Straney, L., Thurston, G.D., Tran, J.H., Van Dingenen, R., Van Donkelaar, A., Veerman, J.L., Vijayakumar, L., Weintraub, R., Weissman, M.M., White, R.A., Whiteford, H., Wiersma, S.T., Wilkinson, J.D., Williams, H.C., Williams, W., Wilson, N., Woolf, A.D., Yip, P., Zielinski, J.M., Lopez, A.D., Murray, C.J.L., Ezzati, M., 2012. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. Lancet 380, 2224–2260. https://doi.org/10.1016/S0140-6736(12)61766-8

Lin, B., Liu, X., 2013. Electricity tariff reform and rebound effect of residential electricity consumption in China. Energy 59, 240–247. https://doi.org/10.1016/j.energy.2013.07.021

Liu, Y., Liu, J., Zhou, Y., 2017. Spatio-temporal patterns of rural poverty in China and targeted poverty alleviation strategies. J. Rural Stud. 52, 66–75. https://doi.org/https://doi.org/10.1016/j.jrurstud.2017.04.002

Longden, T., Quilty, S., Riley, B., White, L. V., Klerck, M., Davis, V.N., Frank Jupurrurla, N., 2021. Energy insecurity during temperature extremes in remote Australia. Nat. Energy 7. https://doi.org/10.1038/s41560-021-00942-2

Luckett, B.G., DeClerck, F.A.J., Fanzo, J., Mundorf, A.R., Rose, D., 2015. Application of the Nutrition Functional Diversity indicator to assess food system contributions to dietary diversity and sustainable diets of Malawian households. Public Health Nutr. 18, 2479–2487. https://doi.org/10.1017/S136898001500169X

MacKinnon, D.P., Lockwood, C.M., Hoffman, J.M., West, S.G., Sheets, V., 2002. A comparison of methods to test mediation and other intervening variable effects. Psychol. Methods 7, 83.

Mahmood, T., Yu, X., Klasen, S., 2019. Do the Poor Really Feel Poor? Comparing Objective Poverty with Subjective Poverty in Pakistan. Soc. Indic. Res. 142, 543–580. https://doi.org/10.1007/s11205-018-1921-4

Masera, O.R., Bailis, R., Drigo, R., Ghilardi, A., Ruiz-Mercado, I., 2015. Environmental Burden of Traditional Bioenergy Use. Annu. Rev. Environ. Resour. 40, 121–150. https://doi.org/10.1146/annurev-environ-102014-021318

Mathiassen, A., 2013. Testing Prediction Performance of Poverty Models: Empirical Evidence from Uganda. Rev. Income Wealth 59, 91–112. https://doi.org/10.1111/roiw.12007

Mayer, K., Banerjee, S.G., Trimble, C., 2015. Elite Capture: Residential Tariff Subsidies in India, World Bank Group. https://doi.org/10.1596/978-1-4648-0412-0

Mbuya, N.V.N., Atwood, S.J., Huynh, P.N., 2019. Persistent Malnutrition in Ethnic Minority Communities of Vietnam: Issues and Options for Policy and Interventions, World Bank Group. https://doi.org/10.1596/978-1-4648-1432-7

McFadden, D., Puig, C., Kirshner, D., 1977. Determinants of the long-run demand for electricity. Am. Stat. Assoc.

Mcneil, M.A., Letschert, V.E., 2010. Modeling diffusion of electrical appliances in the residential sector. Ernest Orlando Lawrence Berkely Natl. Lab. 12.

Mendoza, C.B., Cayonte, D.D.D., Leabres, M.S., Manaligod, L.R.A., 2019. Understanding multidimensional energy poverty in the Philippines. Energy Policy 133. https://doi.org/10.1016/j.enpol.2019.110886

Miller, M., Alberini, A., 2016. Sensitivity of price elasticity of demand to aggregation, unobserved heterogeneity, price trends, and price endogeneity: Evidence from U.S. Data. Energy Policy 97, 235–249. https://doi.org/10.1016/j.enpol.2016.07.031

Ministry of Power Government of India, 2020. Proposed amendment to Electricity Act 2003.

Ministry of Power Government of India, 2018. Amendment of 2018 to Electricity Act of 2003.

Ministry of Power Government of India, 2003. Electricity Act, 2003. Dehli.

Minority Rights Group International, n.d. Vietnam [WWW Document]. 2018.

Morales, L.E., Balié, J., Magrini, E., 2021. How has the minimum support price policy of India affected cross-commodity price linkages? Int. Food Agribus. Manag. Rev. 24, 355–369. https://doi.org/10.22434/IFAMR2020.0035

Muller, C., Yan, H., 2018. Household fuel use in developing countries: Review of theory and evidence. Energy Econ. 70, 429–439. https://doi.org/https://doi.org/10.1016/j.eneco.2018.01.024

Murthy, K.V.N., Sumithra, G.D., Reddy, A., 2001. End-uses of electricity in households of Karnataka state , India. Energy Sustain. Dev. 5, 81–94.

Nakajima, C., 1986. Subjective Equilibrium Theory of the Farm Household, Developments in Agricultural Economics, 3. Elsevier.

Narayanan, S., 2011. A case for refraining the cash transfer debate in India. Econ. Polit. Wkly. 46, 41–48.

Ngui, D., Mutua, J., Osiolo, H., Aligula, E., 2011. Household energy demand in Kenya: An application of the linear approximate almost ideal demand system (LA-AIDS). Energy Policy 39, 7084–7094. https://doi.org/10.1016/J.ENPOL.2011.08.015

Nguyen, T.T., Nguyen, T.-T., Hoang, V.-N., Wilson, C., Managi, S., 2019. Energy transition, poverty and inequality in Vietnam. Energy Policy 132, 536–548. https://doi.org/https://doi.org/10.1016/j.enpol.2019.06.001

Nguyen, Thanh Tung, Nguyen, Trung Thanh, Grote, U., 2020. Credit and Ethnic Consumption Inequality in the Central Highlands of Vietnam. Soc. Indic. Res. 148, 143–172. https://doi.org/10.1007/s11205-019-02202-z

Ni, Y., Li, X., Ye, Y., Li, Y., Li, C., Chu, D., 2020. An Investigation on Deep Learning Approaches to Combining Nighttime and Daytime Satellite Imagery for Poverty Prediction. IEEE Geosci. Remote Sens. Lett. 18, 1545–1549. https://doi.org/10.1109/lgrs.2020.3006019

Nordin, J.A., 1976. A Proposed Modification of Taylor ' s Demand Analysis : Comment. Bell J. Econ. 7, 719–721. https://doi.org/https://doi.org/10.2307/3003285

Nsabimana, A., Bali Swain, R., Surry, Y., Ngabitsinze, J.C., 2020. Income and food Engel curves in Rwanda: a household microdata analysis. Agric. Food Econ. 8. https://doi.org/10.1186/s40100-020-00154-4

Nussbaumer, P., Bazilian, M., Modi, V., 2012. Measuring energy poverty: Focusing on what matters. Renew. Sustain. Energy Rev. 16, 231–243. https://doi.org/https://doi.org/10.1016/j.rser.2011.07.150

Nussbaumer, P., Nerini, F.F., Onyeji, I., Howells, M., 2013. Global Insights Based on the Multidimensional Energy Poverty Index (MEPI). Sus 5, 2060–2076. https://doi.org/10.3390/su5052060

Ogutu, S.O., Gödecke, T., Qaim, M., 2020. Agricultural Commercialisation and Nutrition in Smallholder Farm Households. J. Agric. Econ. 71, 534–555. https://doi.org/10.1111/1477-9552.12359

Oishi, S., Kesebir, S., 2015. Income Inequality Explains Why Economic Growth Does Not Always Translate to an Increase in Happiness. Psychol. Sci. 26, 1630–1638. https://doi.org/10.1177/0956797615596713

Osunmuyiwa, O.O., Payne, S.R., Vigneswara Ilavarasan, P., Peacock, A.D., Jenkins, D.P., 2020. I cannot live without air conditioning! The role of identity, values and situational factors on cooling consumption patterns in India. Energy Res. Soc. Sci. 69, 101634. https://doi.org/10.1016/j.erss.2020.101634

Ozughalu, U.M., Ogwumike, F.O., 2019. Extreme Energy Poverty Incidence and Determinants in Nigeria: A Multidimensional Approach. Soc. Indic. Res. 142, 997–1014. https://doi.org/10.1007/s11205-018-1954-8

Pachauri, S., Jiang, L., 2008. The household energy transition in India and China. Energy Policy 36, 4022–4035. https://doi.org/10.1016/j.enpol.2008.06.016

Pachauri, S., Mueller, A., Kemmler, A., Spreng, D., 2004. On Measuring Energy Poverty in Indian Households. World Dev. 32, 2083–2104. https://doi.org/10.1016/j.worlddev.2004.08.005

Pachauri, S., Spreng, D., 2004. Energy Use and Energy Access in Relation to Poverty. Econ. Polit. Wkly. Vol. 39, 271–278.

Parikh, J., Parikh, K., 2011. India's energy needs and low carbon options. Energy 36, 3650–3658. https://doi.org/10.1016/j.energy.2011.01.046

Paudel, J., 2021. Why Are People Energy Poor? Evidence From Ethnic Fractionalization. Energy Econ. 102, 105519. https://doi.org/https://doi.org/10.1016/j.eneco.2021.105519

Pega, F., Liu, S.Y., Walter, S., Pabayo, R., Lhachimi, S.K., Saith, R., 2017. Unconditional cash transfers for reducing poverty and vulnerabilities: Effect on use of health services and health outcomes in low- and middle-income countries. Cochrane Database Syst. Rev. https://doi.org/10.1002/14651858.CD011135

Pelz, S., Chindarkar, N., Urpelainen, J., 2021. Energy access for marginalized communities: Evidence from rural North India, 2015–2018. World Dev. 137, 2015–2018. https://doi.org/10.1016/j.worlddev.2020.105204

Perez, A., Yeh, C., Azzari, G., Burke, M., Lobell, D., Ermon, S., 2017. Poverty Prediction with Public Landsat 7 Satellite Imagery and Machine Learning. 31st Conf. Neural Inf. Process. Syst. (NIPS 2017). arXiv1711.03654.

Pham Thai, H., Le Dang, T., Nguyen Viet, C., 2010. Poverty of the Ethnic Minorities in Vietnam : Situation and Challenges from the Poorest Communes, MPRA Paper. https://doi.org/https://mpra.ub.uni-muenchen.de/50372/

Piaggesi, S., Gauvin, L., Tizzoni, M., Adler, N., Verhulst, S., Young, A., Price, R., Ferres, L., Cattuto, C., Panisson, A., 2019. Predicting city poverty using satellite imagery. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. 2019-June, 90–96.

Pingali, P., Mittra, B., Rahman, A., 2017. The bumpy road from food to nutrition security – Slow evolution of India ' s food policy. Glob. Food Sec. 15, 77–84. https://doi.org/10.1016/j.gfs.2017.05.002

Pokhriyal, N., Jacques, D.C., 2017. Combining disparate data sources for improved poverty

prediction and mapping. Proc. Natl. Acad. Sci. U. S. A. 114, E9783–E9792. https://doi.org/10.1073/pnas.1700319114

Preacher, K.J., Hayes, A.F., 2004. SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav. Res. Methods, Instruments, Comput. 36, 717–731. https://doi.org/10.3758/BF03206553

Rafi, M., Naseef, M., Prasad, S., 2021. Multidimensional energy poverty and human capital development: Empirical evidence from India. Energy Econ. 101, 105427.

Rais, M., Pazderka, B., vanLoon, G.W., 2009. Agriculture in Uttarakhand, India—Biodiversity, Nutrition, and Livelihoods. J. Sustain. Agric. 33, 319–335. https://doi.org/10.1080/10440040902773152

Rammohan, A., Pritchard, B., 2014. The role of landholding as a determinant of food and nutrition insecurity in rural Myanmar. World Dev. 64, 597–608. https://doi.org/10.1016/j.worlddev.2014.06.029

Ranganathan, V., 1993. Rural electrification revisited. Energy Policy 21, 142–151. https://doi.org/https://doi.org/10.1016/0301-4215(93)90136-4

Razavi, R., Fleury, M., 2019. Socio-economic predictors of electricity theft in developing countries: An Indian case study. Energy Sustain. Dev. 49, 1–10. https://doi.org/https://doi.org/10.1016/j.esd.2018.12.006

Reames, T.G., 2016. Targeting energy justice : Exploring spatial, racial/ethnic and socioeconomic disparities in urban residential heating energy efficiency. Energy Policy 97, 549–558. https://doi.org/10.1016/j.enpol.2016.07.048

Reddy, A.A., 2010. Regional disparities in food habits and nutritional intake in Andhra Pradesh, India. Reg. Sect. Econ. Stud. 10, 125–134. https://doi.org/10.2139/ssrn.1611726

Reiss, P.C., White, M.W., 2005. Household electricity demand, revisited. Rev. Econ. Stud. 72, 853–883. https://doi.org/10.1111/0034-6527.00354

Ruel, M.T., Garrett, J.L., Hawkes, C., Cohen, M.J., 2010. The Food, Fuel, and Financial Crises Affect the Urban and Rural Poor Disproportionately: A Review of the Evidence. J. Nutr. 140, 170S-176S. https://doi.org/10.3945/jn.109.110791

Ryan, D.L., Plourde, A., 2009. Empirical modelling of energy demand, International Handbook on the Economics of Energy. https://doi.org/10.4337/9781849801997.00011

Sadath, A.C., Acharya, R.H., 2017. Assessing the extent and intensity of energy poverty using Multidimensional Energy Poverty Index: Empirical evidence from households in India. Energy Policy 102, 540–548. https://doi.org/10.1016/j.enpol.2016.12.056

Sadoulet, E., de Janvry, A., 1995. Household Models 6.1., in: Quantitative Development Policy Analysis. Baltimore, MD (USA) Johns Hopkins Univ. Press.

Saha, D., Bhattacharya, R.N., 2018. An analysis of elasticity of electricity demand in West Bengal, India: Some policy lessons learnt. Energy Policy 114, 591–597. https://doi.org/10.1016/j.enpol.2017.12.035

Saini, S., 2017. Social and behavioral aspects of electricity theft: An explorative review. Int. J. Res. Econ. Soc. Sci. 7, 26–37.

Sani, N.S., Rahman, M.A., Bakar, A.A., Sahran, S., Mohd, H., 2018. Machine Learning Approach for Bottom 40 Percent Households (B40) Poverty Classification. Int. J. Adv. Sci. Eng. Inf. Technol. 8, 1698–1705.

Sari, R., Soytas, U., 2007. The growth of income and energy consumption in six developing countries. Energy Policy 35, 889–898. https://doi.org/10.1016/j.enpol.2006.01.021

Schultz, T.W., 1980. The Economics of Being Poor. J. Polit. Econ. 88, 639–651.

Sedai, A.K., Nepal, R., Jamasb, T., 2021. Flickering lifelines: Electrification and household welfare in India. Energy Econ. 94, 104975. https://doi.org/10.1016/j.eneco.2020.104975

Sen, A., 1993. Capability and Well-Being, in: Sen, A.K., Nussbaum, M.C. (Eds.), The Quality of Life. Oxford University Press, Oxford. https://doi.org/10.1093/0198287976.003.0003

Sen, A., 1988. The standard of living, Cambridge University Press. https://doi.org/10.1093/oep/gpm015

Sen, A., 1976. Poverty: An Ordinal Approach to Measurement. Econometrica 44, 219. https://doi.org/10.2307/1912718

Sharma, T., Pandey, K.K., Punia, D.K., Rao, J., 2016. Of pilferers and poachers: Combating electricity theft in India. Energy Res. Soc. Sci. 11, 40–52. https://doi.org/https://doi.org/10.1016/j.erss.2015.08.006

Sharma, S. V., Han, P., Sharma, V.K., 2019. Socio-economic determinants of energy poverty amongst Indian households: A case study of Mumbai. Energy Policy 132, 1184–1190. https://doi.org/10.1016/j.enpol.2019.06.068

Shin, J., 1985. Perception of Price When Price Information Is Costly : Evidence from Residential Electricity. Rev. Econ. Stat. 67, 591–598.

Sibhatu, K.T., Krishna, V. V., Qaim, M., 2015. Production diversity and dietary diversity in smallholder farm households. Proc. Natl. Acad. Sci. U. S. A. 112, 10657–10662. https://doi.org/10.1073/pnas.1510982112

Sibhatu, K.T., Qaim, M., 2018. Review: Meta-analysis of the association between production diversity, diets, and nutrition in smallholder farm households. Food Policy 77, 1–18. https://doi.org/10.1016/j.foodpol.2018.04.013

Sibhatu, K.T., Qaim, M., 2017. Rural food security, subsistence agriculture, and seasonality. PLoS One 12, 1–15. https://doi.org/10.1371/journal.pone.0186406

Singh, J., Mantha, S.S., Phalle, V.M., 2018. Characterizing domestic electricity consumption in the Indian urban household sector. Energy Build. 170, 74–82. https://doi.org/https://doi.org/10.1016/j.enbuild.2018.04.002

Singh, M.R., Upadhyay, V., Mittal, A.K., 2005. Urban water tarrif structure and cost recovery opportunities in India. Water Sci. Technol. 52, 43–51. https://doi.org/10.2166/wst.2005.0424

Singh, P., Singh, G., Sodhi, G.P.S., 2019. Energy auditing and optimization approach for improving energy efficiency of rice cultivation in south-western Punjab, India. Energy 174, 269–279. https://doi.org/10.1016/J.ENERGY.2019.02.169

Singhal, S., Beck, U., 2015. Ethnic disadvantage in Vietnam, WIDER. https://doi.org/doi:10.35188/UNU-WIDER/2015/986-2

Sinha, D., Patnaik, B., 2016. Cash for Food : The NEED for caution. Oxfam Indian Policy Br. 3–6.

Smith, K.R., 1987. The biofuel transition Les énergie de la biomasse à une époque charnière. Pacific Asian J. Energy 1, 13–31.

Sobel, M.E., 1982. Asymptotic confidence intervals for indirect effects in structural equation models. Sociol. Methodol. 13, 290–312.

Statista and Ministry of Power (India), 2021. Number of households electrified under the Saubhagya campaign.

Storm, H., Baylis, K., Heckelei, T., 2020. Machine learning in agricultural and applied economics. Eur. Rev. Agric. Econ. 47, 849–892. https://doi.org/10.1093/erae/jbz033

Strotz, R.H., 1957. The Empirical Implications of a Utility Tree. Econometrica 25, 269–280. https://doi.org/https://doi.org/10.2307/1910254

Sun, C., Lin, B., 2013. Reforming residential electricity tariff in China: Block tariffs pricing approach. Energy Policy 60, 741–752. https://doi.org/10.1016/j.enpol.2013.05.023

Sun, C., Ouyang, X., 2016. Price and expenditure elasticities of residential energy demand during urbanization: An empirical analysis based on the household-level survey data in China. Energy Policy 88, 56–63. https://doi.org/10.1016/j.enpol.2015.10.012

Tang, B., Sun, Y., Liu, Y., Matteson, D.S., 2018. Dynamic Poverty Prediction with Vegetation Index, NIPS 2018 Workshop Spatiotemporal.

Tang, N., Mao, S., Wang, Y., Nelms, R.M., 2018. Solar Power Generation Forecasting With a LASSO-Based Approach. IEEE Internet Things J. 5, 1090–1099. https://doi.org/10.1109/JIOT.2018.2812155

Taylor, J.E., Adelman, I., 2003. Agricultural Household Models: Genesis, Evolution, and Extensions. Rev. Econ. Househ. 1, 33–58. https://doi.org/10.1023/A:1021847430758

Taylor, L.D., 1975. The Demand for Electricity. Bell J. Econ. 6, 74–110. https://doi.org/https://doi.org/10.2307/3003216

Taylor, P., 2008. Minorities at Large: New Approaches to Minority Ethnicity in Vietnam. J. Vietnamese Stud. 3, 3–43. https://doi.org/10.1525/vs.2008.3.3.3

Teschner, N., Sinea, A., Vornicu, A., Abu-Hamed, T., Negev, M., 2020. Extreme energy poverty in the urban peripheries of Romania and Israel: Policy, planning and infrastructure. Energy Res. Soc. Sci. 66, 101502. https://doi.org/10.1016/j.erss.2020.101502

Tian, X., Yu, X., 2015. Using semiparametric models to study nutrition improvement and dietary change with different indices: The case of China. Food Policy 53, 67–81. https://doi.org/https://doi.org/10.1016/j.foodpol.2015.04.006

Tibshirani, R., 2016. Regression Shrinkage and Selection via the Lasso. J. R. Stat. Soc. Ser. B 58, 267–288. https://doi.org/https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., Imai, K., 2014. Mediation: R package for

causal mediation analysis. J. Stat. Softw. 59, 1–38. https://doi.org/10.18637/jss.v059.i05

Tingzon, I., Orden, A., Sy, S., Sekara, V., Weber, I., Fatehkia, M., Herranz, M.G., Kim, D., 2019. Mapping poverty in the Philippines using machine learning, satellite imagery, and crowd-sourced geospatial information, in: AI for Social Good ICML 2019 Workshop.

Tiwari, S., Daidone, S., Ruvalcaba, M.A., Prifti, E., Handa, S., Davis, B., Niang, O., Pellerano, L., Quarles van Ufford, P., Seidenfeld, D., 2016. Impact of cash transfer programs on food security and nutrition in sub-Saharan Africa: A cross-country analysis. Glob. Food Sec. 11, 72–83. https://doi.org/https://doi.org/10.1016/j.gfs.2016.07.009

United Nations, 2021. The Sustainable Development Goals Report 2021, United Nations publication issued by the Department of Economic and Social Affairs.

United Nations, 2015. Transforming our wolrd: the 2030 agenda for sustainable development. https://doi.org/10.1201/b20466-7

Van de Walle, D., Gunewardena, D., 2001. Sources of ethnic inequality in Viet Nam. J. Dev. Econ. 65, 177–207.

Van der Kroon, B., Brouwer, R., Van Beukering, P.J., 2013. The energy ladder: Theoretical myth or empirical truth? Results from a meta-analysis. Renew. Sustain. Energy Rev. 20, 504–513.

Van Praag, B.M.S., Ferrer-i-Carbonell, A., 2008. A multidimensional approach to subjective poverty, in: Quantitative Approaches to Multidimensional Poverty Measurement. pp. 135–154. https://doi.org/10.1057/9780230582354

Vietnam General Statistics Officce, 2010. Result of the Vietnam's Household Living Standards Survey 2010 712.

von Braun, J., 1995. Agricultural commercialization: impacts on income and nutrition and implications for policy. Food Policy 20, 187–202. https://doi.org/10.1016/0306-9192(95)00013-5

Von Braun, J., Kennedy, E., 1994. Agricultural commercialization, economic development, and nutrition. Published for the International Food Policy Research Institute [by] Johns Hopkins University Press.

Wand, M.P., Coull, B.A., French, J.L., Ganguli, B., Kammann, E.E., Staudenmayer, J., Zanobetti, A., 2005. SemiPar 1.0. R package.

Wang, H., Maruejols, L., Yu, X., 2021. Predicting energy poverty with combinations of remote-sensing and socioeconomic survey data in India: Evidence from machine learning. Energy Econ. 102, 105510. https://doi.org/10.1016/j.eneco.2021.105510

Wang, H., Zhao, Q., Bai, Y., Zhang, L., Yu, X., 2020. Poverty and Subjective Poverty in Rural China. Soc. Indic. Res. 150, 219–242. https://doi.org/10.1007/s11205-020-02303-0

Wang, Q., Kwan, M.-P., Fan, J., Lin, J., 2021. Racial disparities in energy poverty in the United States. Renew. Sustain. Energy Rev. 137, 110620. https://doi.org/https://doi.org/10.1016/j.rser.2020.110620

Wijaya, D.R., Paramita, N.L.P.S.P., Uluwiyah, A., Rheza, M., Zahara, A., Puspita, D.R., 2020. Estimating city-level poverty rate based on e-commerce data with machine learning.

Electron. Commer. Res. https://doi.org/10.1007/s10660-020-09424-1

Wilder, R.P., Willenborg, J.F., 1975. Residential Demand for Electricity: A Consumer Panel Approach. South. Econ. J. 42, 212–217. https://doi.org/10.2307/1056769

Winkelmann, R., 2005. Subjective well-being and the family: Results from an ordered probit model with multiple random effects. Empir. Econ. 30, 749–761. https://doi.org/10.1007/s00181-005-0255-7

Working, H., 1943. Statistical Laws of Family Expenditure. J. Am. Stat. Assoc. 38, 43–56.

World Bank, 2021. GDP growth (annual %) – Vietnam [WWW Document]. World Bank Natl. accounts data, OECD Natl. Accounts data files.

Xu, J., Song, J., Li, B., Liu, D., Cao, X., 2021. Combining night time lights in prediction of poverty incidence at the county level. Appl. Geogr. 135, 102552. https://doi.org/https://doi.org/10.1016/j.apgeog.2021.102552

Yao, S., Zhang, Z., Hanmer, L., 2004. Growing inequality and poverty in China. China Econ. Rev. 15, 145–163. https://doi.org/10.1016/j.chieco.2003.09.002

Yawale, S.K., Hanaoka, T., Kapshe, M., 2021. Development of energy balance table for rural and urban households and evaluation of energy consumption in Indian states. Renew. Sustain. Energy Rev. 136, 110392. https://doi.org/10.1016/j.rser.2020.110392

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., Burke, M., 2020. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. Nat. Commun. 11, 1–11. https://doi.org/10.1038/s41467-020-16185-w

Yin, J., Qiu, Y., Zhang, B., 2021. Identification of Poverty Areas by Remote Sensing and Machine Learning : A Case Study in Guizhou , Southwest China. Int. J. Geo-Information 10. https://doi.org/https://doi.org/10.3390/ijgi10010011

You, J.S., Lim, S.Y., 2017. Welfare effects of nonlinear electricity pricing. Energy J. 38, 195–211. https://doi.org/10.5547/01956574.38.1.jyou

Yu, X., Abler, D., 2016. Matching food with mouths: A statistical explanation to the abnormal decline of per capita food consumption in rural China. Food Policy 63, 36–43. https://doi.org/10.1016/j.foodpol.2016.06.009

Yu, Y., Huang, J., 2021. Poverty Reduction of Sustainable Development Goals in the 21st Century: A Bibliometric Analysis. Front. Commun. 6, 1–15. https://doi.org/10.3389/fcomm.2021.754181

Zhou, D., Yu, X., 2014. Calorie elasticities with income dynamics: Evidence from the literature. Appl. Econ. Perspect. Policy 37, 575–601. https://doi.org/10.1093/aepp/ppu043

APPENDICES

Appendix 4.A: List of variables included in the LASSO model

Variable	Description	Mean	Std. Dev.	Min	Max
Outcomes					
Electricity consumption	Continuous: consumption of electricity per capita last month, in kWh	22.52	23.32	0.27	425
Electricity expenditure	Continuous: expenditure on electricity over the last 12 months, in thousand Vietnamese Dong	820.38	794.15	0	13740
Determinants					
Expenditure	Continuous: log of monthly real consumption per capita, in thousand Vietnamese Dong	9.35	0.57	7.35	12.53
Female	Binary: Gender of the household head, 1 if female, 0 if male	0.21	0.40	0	1
Married	Binary: Marital status of the household head, 1 if married, 0 if not	0.83	0.37	0	1
Age	Continuous: age of the household head in years	48	14	11	99
Low secondary	Binary: the household head completed lower secondary school as maximum education, 1 if yes, 0 if no	0.30	0.45	0	1
High secondary	Binary: the household head completed higher secondary school as maximum education, 1 if yes, 0 if no	0.11	0.31	0	1
College	Binary: the household head completed college or undergraduate as maximum education, 1 if yes, 0 if no	0.02	0.15	0	1
Graduate	Binary: the household head completed graduate school as maximum education, 1 if yes, 0 if no	0.00	0.03	0	1
Household size	Continuous: total number of members in the household	3.95	1.57	1	15
Number children	Continuous: Number of children below 6 years old in the household	0.38	0.61	0	4
Children	Binary: Presence of children under 6 years old, 1 if yes, 0 if no	0.31	0.46	0	1
Number aged members	Continuous: Number of adults above 65 years old in the household	0.26	0.54	0	3
Dependent	Binary: Presence of adult above 65 years old, 1 if yes, 0 if no	0.20	0.40	0	1
Wage	Binary: Someone in the household receives a wage or salary, 1 if yes, 0 if no	0.62	0.48	0	1
Wage 2	Binary: 2 or more persons in the household receive a wage or salary, 1 if yes, 0 if no	0.29	0.45	0	1
Household production	Binary: Someone in the household is self-employed in agriculture, 1 for yes, 0 for no	0.80	0.39	0	1

Household production 2	Binary: 2 or more persons in the household are self-employed in agriculture, 1 for yes, 0 for no	0.62	0.48	0	1
Business	Binary: Someone in the household is engaged in trade or business, 1 for yes, 0 for no	0.31	0.46	0	1
Business 2	Binary: 2 or more persons in the household are engaged in trade or business, 1 for yes, 0 for no	0.14	0.34	0	1
Poverty status	Binary: the household was classified as a poor one of the commune/ward, 1 for yes, 0 for no.	0.13	0.33	0	1
Kerosene expenditure	Continuous: Expenditures on kerosene last month, in thousand Vietnamese Dong	1758	236	0	43200
Squared meters	Continuous: indicates the size of the dwelling	66.14	38.05	4	410
Number appliances	Continuous: Number of appliances owned by the household	4.58	2.65	0	23
Fridge	Binary: Ownership of specific electric appliances, 1 for yes, 0 for no	0.29	0.45	0	1
Television	Binary: Ownership of specific electric appliances, 1 for yes, 0 for no	0.83	0.38	0	1
Air conditioning	Binary: Ownership of specific electric appliances, 1 for yes, 0 for no	0.02	0.122	0	1
Washing machine	Binary: Ownership of specific electric appliances, 1 for yes, 0 for no	0.07	0.24	0	1
Electric fan	Binary: Ownership of specific electric appliances, 1 for yes, 0 for no	0.75	0.43	0	1
Electric cooker	Binary: Ownership of specific electric appliances, 1 for yes, 0 for no	0.66	0.47	0	1
Heavy machine[37]	Binary: Ownership of specific electric appliances, 1 for yes, 0 for no	0.42	0.49	0	1
Music or video equipment	Binary: Ownership of specific electric appliances, 1 for yes, 0 for no	0.56	0.49	0	1
Computer	Binary: Ownership of specific electric appliances, 1 for yes, 0 for no	0.08	0.26	0	1
Office equipment [38]	Binary: Ownership of specific electric appliances, 1 for yes, 0 for no	0.29	0.45	0	1
Electricity price	Continuous: price calculated as the expenditure per capita last month over the consumption per capita last month	0.95	0.79	0	11.66
Northern	Binary: The household is located in Midlands or Northern Mountainous areas, 1 for yes, 0 for no	0.18	0.38	0	1
Central	Binary: The household is located in the Central Highlands, 1 for yes, 0 for no	0.07	0.25	0	1

[37] Electric generator, pumping machine.
[38] Landline, fax, printer.

Appendix 5.A: Examples of residential electricity rates and demand, 2012-2013

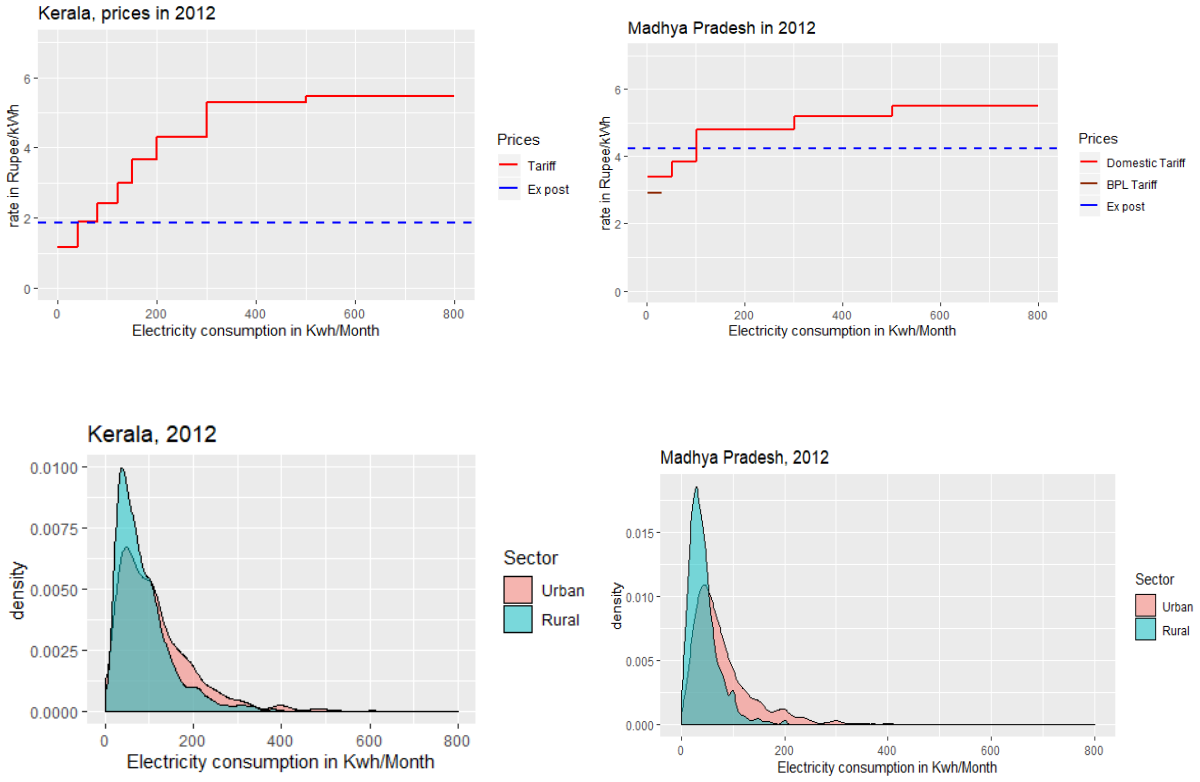

Figure 5.A. 1: Prices and Consumption density[39]

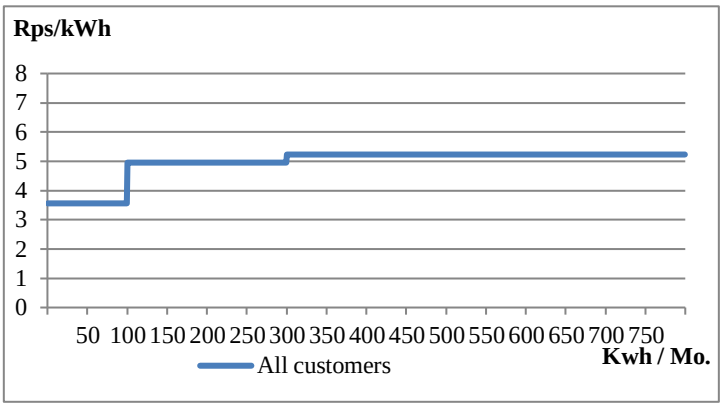

Figure 5.A. 2: Variable charge (IBP) for Punjab (2012-13)

[39] Figure 5.A.1: Madhya Pradesh offered a BPL price while Kerala offers very low rate for consumption up to 30 kWh/month. Less urban and less developed, Madhya Pradesh exhibits a higher density of electricity users around small volumes of monthly consumption than Kerala, and a more distinct gap between urban and rural patterns. In 2012, the *ex post* price (average price calculated from the survey responses) for Madhya Pradesh was 4.24 Rupee/kWh, one of the highest in India, while for Kerala it was 1.87 Rupee/kWh.

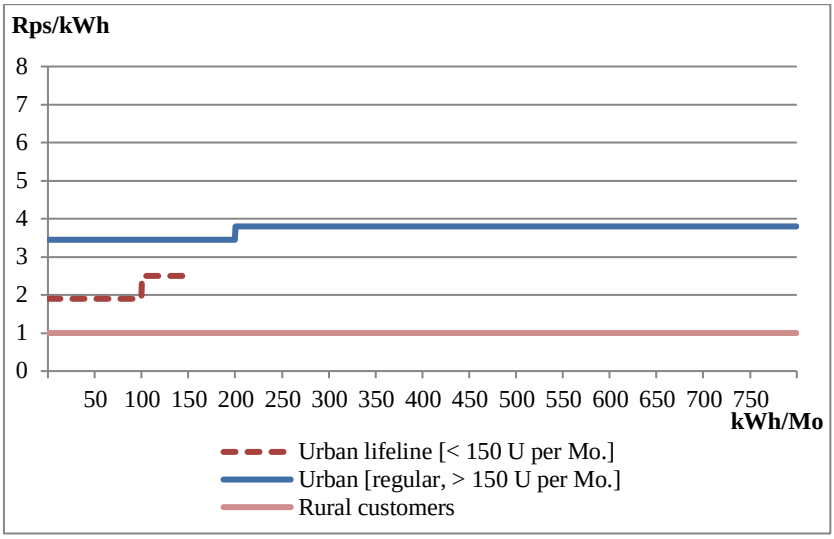

Figure 5.A. 3: Variable charge (IBP) for Uttar Pradesh (2012-13)

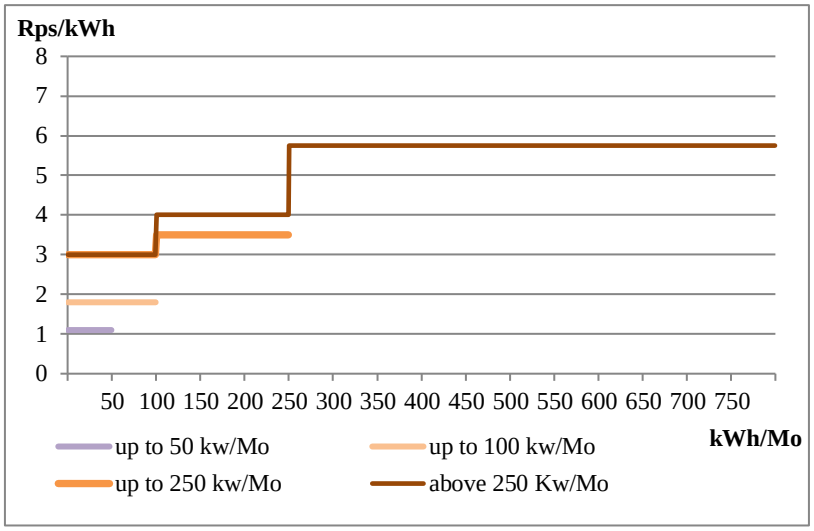

Figure 5.A. 4: Variable charge (IBP) for Tamil Nadu (2012-13)

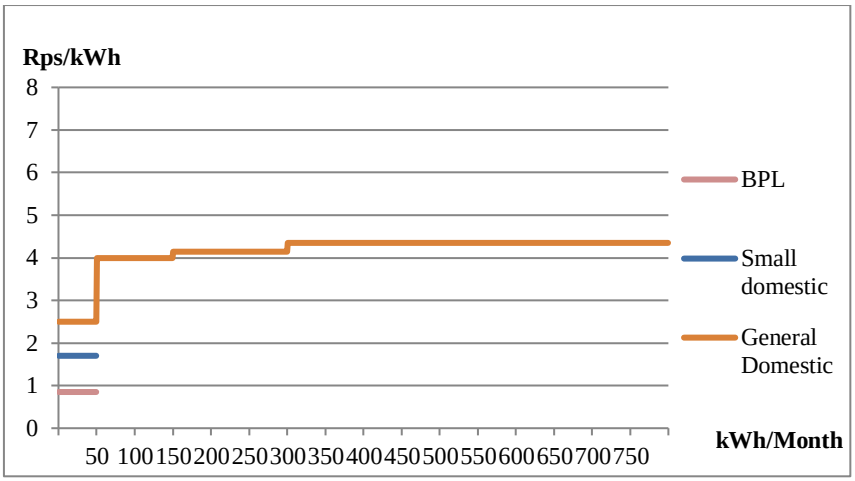

Figure 5.A. 5: Variable charge (IBP) for Rajasthan (2012-13)

Appendix 5.B: Summary statistics for electricity consumption per state, 2012-2013

State Name	Electricity users[40]	Share of households under 30 u./month	Average rupees/month for electricity	Share of electricity in energy budget	Average price of electricity (ex post)	Share of rural households
JAMMU & KASHMIR	96%	3%	205.24	29%	2.01	60%
HIMACHAL PRADESH	98%	2%	182.14	29%	1.31	82%
PUNJAB	99%	7%	628.30	50%	4.49	50%
CHANDIGARH	99%	10%	526.90	57%	3.92	21%
UTTARANCHAL	97%	14%	173.00	30%	2.36	58%
HARYANA	97%	5%	453.48	46%	3.69	54%
DELHI	99%	8%	627.14	60%	3.48	7%
RAJASTHAN	88%	16%	358.42	43%	4.31	58%
UTTAR PRADESH	65%	17%	236.01	32%	3.45	54%
BIHAR	50%	29%	155.54	25%	3.24	57%
SIKKIM	98%	22%	65.61	18%	1.36	79%
ARUNACHAL PRADESH	76%	54%	128.40	20%	3.71	57%
NAGALAND	99%	26%	159.64	21%	3.17	66%
MANIPUR	90%	8%	274.25	38%	3.79	51%
MIZORAM	94%	11%	202.00	25%	2.31	39%
TRIPURA	89%	33%	143.58	29%	3.05	68%
MEGHALAYA	90%	9%	194.59	30%	2.56	66%
ASSAM	71%	35%	178.63	31%	4.15	69%
WEST BENGAL	84%	32%	262.54	36%	3.88	52%
JHARKHAND	78%	21%	120.55	23%	2.00	57%
ORISSA	79%	13%	148.03	28%	1.81	70%
CHATTISGARH	92%	27%	184.32	30%	2.62	65%
MADHYA PRADESH	90%	30%	269.18	38%	4.24	55%
GUJARAT	97%	15%	353.73	46%	4.36	50%
DAMAN & DIU	100%	13%	213.29	36%	1.97	50%
D & N HAVELI	99%	11%	173.53	26%	1.60	51%
MAHARASTRA	96%	17%	345.82	43%	3.91	49%
ANDHRA PRADESH	98%	10%	184.32	40%	2.28	57%
KARNATAKA	97%	31%	177.27	30%	2.78	50%
GOA	100%	2%	262.11	39%	1.44	36%
LAKSHADWEEP	100%	5%	389.35	51%	1.58	33%
KERALA	98%	12%	198.74	32%	1.87	58%
TAMIL NADU	98%	12%	127.49	28%	1.18	50%
PONDICHERRY	99%	1%	170.57	36%	0.84	22%
A & N ISLANDS	95%	6%	255.95	36%	2.05	47%

[40] Electricity users among household with non-zero energy expenditure

Appendix 5.C: Summary statistics on residential energy consumption for 2012-2013

	Whole sample (n=*63563)*			**Urban sample** (n=*28358)*			**Rural sample** (n=*35205)*		
	Mean	*Min*	*Max*	*Mean*	*Min*	*Max*	*Mean*	*Min*	*Max*
Quantity, Consumption	69	1	1667	87	1	1667	55	1	1500
Share of households below 30 u/Mo	0.22	0	1	0.153	0	1	0.27	0	1
Price (average)	3.39	0.25	21	3.63	0.90	21	3.19	0.25	16
Marginal price for each household	2.5	0	9.62	2.83	0.001	9.62	2.23	0.001	6.25
Highest marginal price of schedule	4.81	1	9.62	5.11	1.91	9.62	4.57	1	8.1
Lowest marginal price of schedule	2.23	0.7	4.46	2.42	0.7	4.46	2.06	0.7	4.41
Marginal price for each household (instrument)	2.52	0	7.37	2.78	0.001	7.37	2.30	0.001	6.25
Stone Price index (base model)	12.38	0.65	89.3	14.98	0.65	81.85	10.28	0.65	89.3
Stone Price index (IBP model)	12.09	0.22	89.3	14.63	0.22	81.05	10.04	0.34	89.3
Expenditure, INR/Mo.	241	2	8000	323	4	7000	176	2	8000
Expenditure calculated at intramarginal rates	173	0	7624	240	0	6574	120	0	7624
Expenditure calculated at highest marginal rate	337	3	7845	444	5	7179	250	3	7845
Expenditure calculated at lowest marginal rate (i.e. benchmark expenditure)	156	1	5340	210	2	3489	113	1	5340
Total expenditure on energy	652	2	14400	704	4	12384	610	2	14400
Total expenditure of household (base model)	9186	206	295903	10265	340	289521	8317	206	295903
Tax, or rate structure premium	24	0	3531	40	0	3531	12	0	2284
Total expenditure of household (IBP model), residual after tax	9162	206	295903	10225	340	288675	8305	206	295903
Share of electricity in energy expenditures	0.35	0	1	0.44	0.01	1	0.29	0	1

Share of energy in total expenditure (base)	0.08	0	0.5	0.08	0	0.43	0.09	0	0.5
Share of electricity in total expenditure (base)	0.03	0	0.39	0.03	0	0.39	0.02	0	0.28
Share of energy in total expenditure (IBP model)	0.08	0	0.52	0.08	0	0.48	0.09	0	0.52
Share of electricity in total expenditure (IBP model)	0.03	0	0.45	0.03	0	0.45	0.02	0	0.3

Electricity prices in INR/kWh

Appendix 5.D: Progressivity of electricity rates 2012-2013, per state.

PANEL A:	States with single tariff (no discrimination based on customer characteristics)				
	Marginal price (in INR/kWh)			**% Price increase**	
State Name	at 50 kWh/Mo.	at 100 kWh/Mo.	at 200 kWh/Mo.	50 to 100 kWh/Mo.	100 to 200 kWh/Mo.
Jammu & kashmir	1.43	1.43	1.76	0%	23%
Punjab	4.09	4.09	5.49	0%	34%
Chandigarh	2.30	2.30	4.20	0%	83%
Haryana	4.50	4.50	4.50	0%	0%
Delhi	3.00	3.00	3.00	0%	0%
Sikkim	1.10	2.25	3.45	105%	53%
Nagaland	3.60	3.60	4.00	0%	11%
Chattisgarh	0.70	0.70	0.80	0%	14%
Daman & diu	1.60	2.25	2.25	41%	0%
D & n haveli	1.60	2.25	2.25	41%	0%
Andhra pradesh	1.45	2.60	3.60	79%	38%
Goa	1.20	1.60	1.60	33%	0%
Lakshadweep	0.75	1.00	2.00	33%	100%
Kerala	1.90	2.40	3.65	26%	52%
Pondicherry	1.91	1.91	1.91	0%	0%
A & n islands	1.70	1.70	3.20	0%	88%
Average	2.05	2.35	2.98	22%	31%
Median	1.65	2.25	3.10	0%	19%

PANEL B:	States that offer below-poverty line (BPL) rate						
	Marginal price (in INR/kWh)				**% Price increase**		
State Name	BPL rate	at 50 kWh/Mo.	at 100 kWh/Mo.	at 200 kWh/Mo.	BPL to standard 50 kWh	50 to 100 kWh/Mo.	100 to 200 kWh/Mo.
Himachal pradesh	2.5	3	3	3.9	20%	0%	30%
Uttaranchal	1.5	2.3	2.3	2.6	53%	0%	13%
Arunachal pradesh	2.65	4	4	4	51%	0%	0%
Manipur	1	2.2	2.2	2.7	120%	0%	23%
Mizoram	0.9	1.5	1.85	2.5	67%	23%	35%
Meghalaya	1.8	2.45	2.45	2.75	36%	0%	12%
Assam	2.5	3.25	3.25	4.3	30%	0%	32%
Orissa	0	2.2	3.9	3.9	n.a	77%	0%
Madhya pradesh	2.9	3.4	3.85	4.8	17%	13%	25%
Maharastra	0.78	2.47	2.47	4.37	217%	0%	77%
Average	*1.65*	*2.68*	*2.93*	*3.58*	*68%*	*11%*	*25%*
Median	*1.65*	*2.46*	*2.74*	*3.90*	*51%*	*0%*	*24%*