Income, household energy and health

Darby Jack*

November 16, 2004

1 Introduction

About 2.5 billion people rely on biomass for heating and cooking. As populations in in biomass dependent regions grow, this figure is expected to increase to 2.6 billion by 2030 even as some households switch to hydrocarbon-based fuels [1].

While burning solid fuels for warmth, security and cooking is as old as humanity, recent work on the epidemiology of respiratory infections reveals that burning solid fuels in poorly ventilated indoor spaces causes several serious health problems. Health researchers believe that approximately 1.6 million people die annually as a result of indoor air pollution from cooking with solid fuels [18]. Nearly all these deaths occur in developing countries. Women and children are much more likely to be affected. Adult women suffer disproportionately because they do most of the cooking and thus inhale more smoke than men. Children tend to be indoors near the fire with their mothers; they are also more likely to develop disease in response to a given exposure.

The central question of this study hinges on the health impacts of biomass energy. *Why do people use energy technologies that kill them?* The most obvious answer, from the point of view of economics at least, is that they are too poor to afford cleaner alternatives. On this view households invest in clean household energy technologies until the benefits of cleaner air would be more than offset by the loss of utility due to the additional cost of clean

---

*Doctoral Candidate, Kennedy School of Government, Harvard MA. email: djack@fas.harvard.edu. Support from the Repsol YPF Harvard Kennedy School Fellows Program is gratefully acknowledged.

fuels. This behavioral postulate-the but efficient hypothesis[8] -constitutes the theoretical foundation of the handful previous papers on the determinants of energy technology choice by poor households. I use household data from Peru to investigate this hypothesis. The central questions of this paper are, how does fuel choice depend on price and income? What does this dependence reveal about the value of risk reduction to Peruvian consumers?

Welfare arguments in favor of programs that promote indoor air quality are greatly strengthened if it turns out that household behavior diverges from optimality. Conversely, if we conclude that households are expected utility maximizers with reasonably good information (and if we believe that the environmental externalities associated with solid fuel use are small), then there is no particular reason to privilege indoor air quality programs over programs that increase income. Welfare concerns aside, understanding the details of household decision-making is necessary for the design of effective public policy aimed at indoor air in particular and clean energy in general.

An ideal test of the poor but efficient hypothesis would estimate household's willingness to pay for risk reductions implicit in fuel switching and compare this measure with other estimates of risk-money tradeoffs. If households are managing the risk of illness from indoor smoke in an efficient manner then the willingness to pay for risk reduction should accord with risk reduction expenditures in other domains. While I am not yet in a position to estimate the money-risk tradeoff decisions that helped generate my data, features of the data suggest that the poor are under-investing in clean indoor air.

This paper builds on previous literature on the role of income in household energy demand in two ways. First, I use nonparametric techniques to study the relationship between income and expenditure on clean fuels without forcing any structure. This analysis borrows from approaches used by Subramanian and Deaton [23] and others to study the relationship between income and calorie consumption. Second, I take advantage of the longitudinal structure of the Peruvian data to control for unobserved heterogeneity using household random effects.

I find that price- and income elasticities of demand for clean fuels are of the expected sign. In keeping with results elsewhere in the literature I find that urban dwellers are far more likely to use clean fuels and that electric lighting is *ceteris paribus* strongly predictive of clean fuel use. I compare my panel-data point estimates of price and income elasticities of demand for clean fuels with pooled estimates that correspond to approaches used elsewhere in the household fuel choice literature; this exercise suggests that error introduced by unobserved heterogeneity is likely to lead to overestimating the effects of price and income.

Assessing the meaning of these results poses interesting puzzles. In a paper that poses some of the questions considered here Larson and Rosen [14] estimate that switching from a moderately dirty cooking system to a moderately clean system will reduce annual mortality risk by 10% for each exposed member of the household. While the methods they use to calculate this figure have real problems–primarily stemming from the quality of the underlying dose-response fuctions–their result is strikingly large even if it overestimates risk by an order of magnitude.[1]

Overall perhaps the most surprising result is also one of the simplest: expenditures on clean fuels by the poor seem low given the risks associated with cooking with solid fuels. Households in the bottom 10% of the income distribution that cook only with clean fuels (kerosene and/or gas) allocate only 4% of their expenditure to energy. By way of reference, this only slightly exceeds the 3% each of budget share that dirty-fuel users among the poorest decile allocate to personal articles (toothpaste, deodorant, hand-cream, etc.) and to entertainment. It seems unlikely that the preference relations implied by these tradeoffs truly reflect underlying attitudes towards risk. The remainder of the paper proceeds as follows. In the following section I briefly review some of the relevant literature. Section 3 outlines the Peru data and uses tabulations and nonparametric regression techniques to explore patterns in the data. Section 4 presents early results from structural modeling household and Section 5 concludes.

2 Literature review

Since the oil shocks of the 1970's, energy scholars have devoted considerable effort to understanding household energy use in developing countries. Much of this work draws on the concept of an energy ladder with solid fuels on the lowest rung, and kerosene, LPG and electricity on successively higher rungs

---

[1] Indeed, the obvious biases in their estimate suggest that they may actually be underestimating the risk reduction from fuel switching. They use dose response curves from adults exposed to ambient air pollution; preliminary epidemiological evidence suggests that dose response curves are steeper for indoor pollutants and it is well known that they are steeper for children.

[15, 13]. An extensive literature drawing on evidence from South Africa [7], Zimbabwe [13, 5], India [2], Tanzania [12] and Pakistan [6] generally supports the notion that households climb the energy ladder as they grow wealthier (this literature is reviewed by Barnes [4]). In addition, several scholars find that urban dwellers are more likely to use cleaner technologies [5, 13, **?**] and at least one study finds that households that use electricity are more likely to use clean cooking fuels [7].

Spurred by growing concern about the health consequences of indoor air quality, the World Bank has recently embarked on a program of study aimed at developing a systematically documented set of stylized facts regarding household energy use patterns [26, 25]. Analysis of household cross sections carried out under the World Bank's Living Standards Measurement Surveys (LSMS) program form the core of the study. Data from Brazil, Ghana, Guatemala, India, Nepal, Nicaragua, South Africa and Vietnam exhibit patterns similar to those described above.

Though extensive, this research suffers from three significant problems. First, nearly all of the papers–important exceptions are discussed below–tabulate or run reduced form logit regressions on cross sectional data, and therefore fail to establish a causal relationship between income and energy choice. Second, omitted variable bias may be giving rise to a systematic overestimation of the effects of income on fuel choice. For example, if access to information about the health costs of dirty fuels is positively correlated with income but omitted from analysis, then the coefficient on income will have an upwards bias. Similarly, income and access to fuel may both be a function of infrastructure quality and/or proximity to markets. Third, the energy ladder literature is completely empirical, and fails to capitalize on insights that flow from even a modest use of economic theory.

Recent, more econometrically and theoretically sophisticated research suggests that the relationship between income and indoor air quality may more complex than the energy ladder metaphor indicates.

First, Bardhan et al. provide evidence from Nepal that suggests that households may not climb the energy ladder if supplies of cleaner fuels are lacking [3]. They relate intra-village variations in firewood collection to differences in consumption levels of Nepalese households, controlling for village fixed effects. They estimate a structural model of household consumption and find that that the quantity of firewood consumed is monotonically increasing in wealth. Further, they find no evidence of a higher shadow cost of collection time among wealthier households. They assert that "the absence of a negative wealth effect is unsurprising given the absence of affordable fuel substitutes for most households in Nepal" (p. 4). This finding is consistent with broader studies of forest use that show that demand for firewood generally increases as people become wealthier, including Rosenzweig and Foster's results from India [20] and Patel et al.s results in Kenya [19].

A second strand of research has examined how the use of multiple fuels can result in non-linear relationships between income and indoor air quality. Chaudhuri and Pfaff [6] estimate fuel-use Engel curves for natural gas, propane, kerosene and solid fuels using World Bank data from Pakistan. They find that as incomes rise, households switch from solid fuels to cleaner liquid fuels. This improvement does not, however, imply a monotonic increase in indoor air quality. Instead, initial increases in income appear to lead to increases in consumption of dirty fuels. Only after household income reaches a threshold level do households switch technologies and begin enjoying cleaner indoor air. Implicit in their findings–and in contrast with Bardan et al. [3]–is the fact a significant portion of household have access to a full set energy technologies. Indeed, they drop all households that live in villages in which some fuels are completely unused. Masera et al. [16] reach similar conclusions for Mexican households.

In sum, nearly three decades of research into the determinants of failed to advance our understanding beyond a fairly simple notion of a positive income elasticity of demand. Recent papers have introduced important methods, but the field still begs for better data and better models. Structural modeling using panel data represents one important direction for innovation, as I attempt to demonstrate below. Another direction that warrants serious attention is the exploitation of natural policy experiments or even experimental interventions

3 Data Description

The data used here come from rounds of the Peruvian *Estudio Nacional de Hogares* (ENAHO), a national household survey carried out in five successive novembers between 1998 and 2002. Table 1 reviews the structure of the panel. The first tranche of numbers shows the total number of households interviewed in each round of the survey and the number of households in each round that were also interviewed in another year. As the second tranche of numbers shows, households were rotated out of the panel; most households were interviewed only twice. This design makes it impossible to tell whether households disappear from the survey because they were intentionally dropped or because they moved away. A comparison of means across the panel subgroup and the balance of the observations failed to illuminate any significant differences.

The ENAHO survey is a general purpose survey modeled on the World Bank Living Standards Surveys. It collects detailed information on income and expenditures, eduction, health and a variety of other social and economic variables. Both the full surveys and the panel are systematic stratified samples based on 1993 census data. Both, however, deliberately oversample urban households.

Table 1: **Panel Structure**

Year	**Total Obs.**	**Panel Obs.**
1998	7551	3816
1999	4107	3898
2000	4169	3990
2001	16515	8687
2002	18598	6184
total	-	26575
period		
2001-2002	-	5103
1998-1999	-	2083
2000-2001	-	1470
1998-2001	-	1998
2000-2002	-	702
other patterns	-	823

Several features of the data make it particularly appropriate for the question at hand. First, as mentioned above, multiple observations on the same household allows better point estimates. As Table 3 illustrates, the period in question saw a marked decline in the use of kerosene and an increase gas. The relatively flat time path of wood in Table 3 masks the fact that households switch both ways as shown in Table 2, which reports transition probabilities for abandoning traditional fuels.

Table 2: **Transition matrix for Fuel Choice Dynamics; 1 ⇒ use only LPG & Kerosene** (Frequencies above, transition probabilities below.)

	0	1	Total
0	5,884 (0.81)	1,408 (0.19)	7,292 100
1	727 (0.11)	5,598 (0.89)	6,325 100
Total	6,611 48.55	7,006 51.45	13,617 100

Second, Peru exhibits unusual variation in fuel choice. Epidemiological studies of indoor air related illness show beyond a doubt that the problem is greatest among the very poor in Asia and Africa [18]. Unfortunately the lack of variation in household fuel choice that drives the health problems also makes it very difficult to study the economic antecedents of indoor air quality. A third virtue of the Peruvian data is that it encompasses considerable heterogeneity in geography, and thus affords insight into how the supply of fuelwood affects fuel-switching decisions. In particular, Peru comprises three distinct zones: the coastal desert, which is almost without exception inhospitable to woody plants; the Andean sierra, which harbor forests in the inhabited valleys; and the eastern selva, or rain forest, where wood is abundant and temperatures high.

Table 3: **Fuel Use Over Time. Numbers in columns give the portion of households using a given fuel in a given year.**

Year	**Electricity**	**Gas**	**Kerosene**	**Charcoal**	**Wood**	**Other**
1998	.01	.34	.37	.02	.48	.06
1999	.01	.38	.35	.03	.48	.07
2000	.01	.41	.31	.03	.47	.11
2001	.01	.41	.26	.04	.50	.07
2002	.01	.46	.20	.04	.47	.06
Total	.01	.40	.28	.03	.48	.07

Exposure is extremely difficult to measure, and even data on concentrations of indoor air pollutants are largely unavailable [10]. My approach is thus to use fuel choice as a proxy for indoor air quality, an approach that is validated by what limited data we have linking fuel choice to exposure [9] . In essence I am interested in P_{nj}, the probability that household n select fuel j. My approach is to treat j as a binary variable:

$$j = \begin{cases} 0 & \text{if the household uses any dirty fuels} \\ 1 & \text{if the household uses clean fuel only} \end{cases}$$

I define clean fuels as comprising liquified petroleum gas (LPG), electricity and kerosene; dirty fuels in my dataset include wood, animal dung and charcoal.[2] I ignore the possibility that households burn solid fuels in improved cookstoves that vent to the outdoors, for two reasons.

1. Data on cooking and ventilation arrangements simply is not available in the ENAHO dataset.
2. Interviews with approximately 100 households conducted by the author in the central sierra of Peru and with employees of several private and public sector groups in Lima involved with improved stove diffusion nationwide show that improved cookstoves are very rarely used (in the 100 interviews I conducted, one home used a chimney).

Perhaps the biggest problem is the paucity of reliable data regarding prices of fuels. Two issues arise. First, the data lack direct measures of local prices for modern fuels (kerosene and LPG) means that it is exceedingly difficult to estimate the price elasticity of demand. At present I am using annual price data from each of 25 regional capitals and assuming that village prices correlate with what I observe. The second problem is that I have no systematic data relating to the price of woodfuel and other biomass energy sources. Given that many households gather fuelwood, income is correlated with the shadow cost of labor and thus with the cost of wood. But this correlation depends on the availability of fuelwood, which I do not observe.

---

[2] Some might object framing the question in a way that promotes an expansion of societal dependence on fossil fuels. See Smith [22] for an interesting discussion of this issue; he concludes that a 0.5% annual increase in fuel efficiency would more than compensate for the additional demand for fossil fuels that would arise should all 2.5 billion solid-fuel users suddenly switch.

Table 4 shows the basic findings of the paper: fuel choice is strongly correlated with income, but exhibits significant variation across regions and between urban and rural settings. Note that the cut-points for the percentile bins are computed using the pooled data. The non-linearities suggested by Table 4 are apparent in Figure 1, which shows a locally weighted regression of $P_1 i$ on the log of per capita income. [3] The graphs all exhibit a sharp nonlinearity in the mid range of the income distribution.

Table 4: Percentage of households that report not using traditional fuels in the past year, by income decile, region and level of urbanization (urban ⇒more than 2000 inhabitants)

	Bottom 10%	**Bottom 50%**	**Top 10%**
national sample	5	21	93
rural coast	4	14	84
urban coast	75	77	99
rural sierra	1	4	90
urban sierra	34	42	93
rural selva	1	4	82
urban selva	12	28	84

4 Random utility modeling–theory and some preliminary results

The analysis thus far has refrained from imposing any assumptions about the nature of the decision process that gives rise to fuel choices. This section explores the insights that flow from analyzing the data within the framework of a discrete choice random utility model. The random utility model also provides a convenient approach to exploiting the panel structure of the data to address unobserved heterogeneity. The point of departure is a utility model that is composed two parts, one observed by the analysist, the other treated

---

[3] The weighting was performed using a lowess scatterplot smoothing procedure.

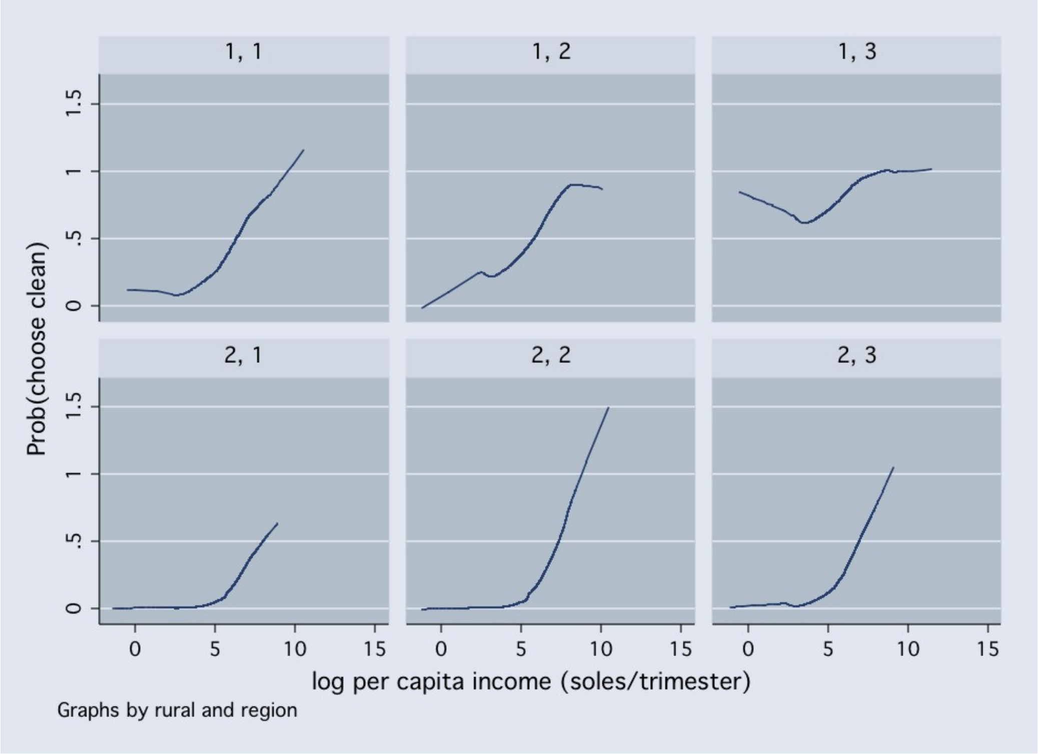

Figure 1: The graphs are labeled (rural, region) where 1 codes for rural and 2 codes for urban; and 1 codes for selva, 2 for sierra and 3 for the coast.

as random.[4] Household j's payoff from choice n is thus $U_{nj} = V_{nj} + \epsilon_{nj}$ where V_{nj} is observed and ϵ_{nj} is a random variable. Let the vector $\epsilon_n' = \epsilon_{n1}, ..., \epsilon_{nj}$ be distributed normal with a mean vector of zero and covariance matrix Ω.

The probability that household n choose option i is

$$\begin{aligned} P_{ni} &= \text{Prob}(U_{ni} > U_{nj} \forall j \neq i) && (1) \\ &= \int I(V_{ni} + \epsilon_{ni} > V_{nj} + \epsilon_{nj})\phi(\epsilon_n)d\epsilon_n && (2) \end{aligned}$$

where $I(\cdot)$ is the indicator function. Unlike other discrete choice models the probit choice probability does not have a closed form analytic solution and is estimated through simulation. The assumption of normal errors is motivated by two concerns. First, normal errors are suitable for panel data where errors (unobserved components of the utility function) are correlated over time. Normal errors are also attractive because they allow for random taste variation and free substitution patterns among alternatives (though this last virtue is irrelevant in the case of binary choice). The drawback with assuming that unobserved utility is normally distributed is that, given random taste variation, the normality assumption can lead to perverse forecasts–e.g., regarding density on both sides of zero of the distribution of the price coefficient. At present, given my sparse data, the assumption of normal errors seems appropriate. As I develop a richer behavioral model estimation using a random parameters logit approach may be preferable.

At present my utility model is parsimonious in the extreme. I decompose V_nj into a component s_n that varies across households and that affects the payoff from using each energy source. I also include modern energy prices, which vary across alternatives. Table 5 reports the regression results. I include two different specifications of the dependent variable: (M1) is a dummy that equals 1 when the household uses only modern fuels; and (M2) is a dummy that equals 1 when the household uses at least some modern fuels. A cursory review of the coefficients and standard errors shows that the model is not particularly sensitive to which specification is chosen. It is also important to note that I have chosen to report not the estimated coefficients but rather elasticities of the form $\partial y/\partial log(x)$. This gives the change in the probability that the household choose $j = 1$ given a 1% change in the dependent variable.

---

[4] The following draws heavily on Train [24], particularly Chapter 5 which focuses on probit models.

Table 5: Probit regression results.

	M1	SE	M2	SE	M3	SE
y per capita	0.323	(-0.014)***	0.430	(0.020)***	0.169	(0.005)***
male edu	0.062	(0.202)	-0.152	(0.189)	-0.025	(0.030)
female edu	0.050	(0.139)	-0.009	(0.125)	-0.021	(0.020)
gas prices	-0.064	(0.433)	1.689	(0.394)	0.365	(0.034)***
year 1999	-0.041	(0.010)***	-0.026	(0.010)***	-	-
year 2000	-0.070	(0.018)***	-0.085	(0.016)***	-	-
year 2001	0.170	(0.049)***	0.082	(0.045)*	-	-
year 2002	0.149	(0.025)***	0.057	(0.024)***	0.016	(0.002)***
central coast	0.118	(0.010)***	0.105	(0.011)***	0.006	(0.001)***
southern coast	0.054	(0.006)***	0.044	(0.007)***	-0.020	(0.001)***
northern sierra	-0.054	(0.010)***	-0.096	(0.009)***	-0.026	(0.002)***
central sierra	-0.084	(0.015)***	-0.134	(0.013)***	-0.025	(0.002)***
southern sierra	-0.029	(0.013)***	-0.123	(0.012)***	-0.041	(0.003)***
selva	-0.134	(0.018)**	-0.191	(0.016)***	0.035	(0.006)**
lima	0.229	(0.021)***	0.176	(0.035)***	-0.024	(0.003)***
large urban	-0.170	(0.014)***	-0.168	(0.018)***	-0.023	(0.002)***
medium urban 2000	-0.151	(0.010)***	-0.141	(0.011)***	-0.028	(0.001)***
small urban	-0.175	(0.009)***	-0.173	(0.010)***	-0.124	(0.004)***
rural	-0.747	(0.026)***	-0.713	(0.028)***	-0.019	(0.001)***
isolated	-0.123	(0.006)***	-0.115	(0.005)***	-0.078	(0.002)***
very isolated	-0.498	(0.017)***	-0.466	(0.016)***	-0.051	(0.002)***

Notes: ***:significant at 99%**: significant at 95%*: significant at 90 % **M1** is the preferred specification: Probit for panel data with random effects. Dependent variable = 1 iff households use only clean fuels. **M2** is also random effects probit but defines the dependent variable as 1 if households use at least some clean fuels. **M3** is a benchmark model: probit on the pooled data.

At this stage these results do little more than confirm the analysis presented above. The marginal effect of income is large and significant regardless of the specification. It is interesting and puzzling that probit on the pooled data significantly underestimates the effect of income and overestimates the effect of price.

5 Conclusions

The present results confirm that the basic stylized facts received from the energy ladder literature: the Engle curve for clean energy, and hence clean indoor air, is normal. The key question, which this paper has thus far failed to answer, is whether poor households in Peru are making welfare-optimizing choices regarding clean fuel consumption. The immediate challenge is to extend the modeling approach sketched out here to provide the appropriate welfare evaluation. The theory behind such evaluations is well developed [21, 17, 11]; the chief difficulty in this case is developing a model that works given the constraints of the available data.

References

[1] International Energy Agency. *World Energy Outlook 2002*, chapter Energy and Poverty. 2002.

[2] M. Alam, J. Sathaye, and D. Barnes. Urban household energy use in india: efficiency and policy implications. *Energy Policy*, 26:885–891, 1998.

[3] Pranab Bardhan, Jean-Marie Baland, Sanghamitra Das, Dilip Mookherjee, and Rinki Sarkar. The environmental impact of poverty: evidence from firewood collection in rural nepal. *mimeo*, 2002.

[4] Douglas F. Barnes and Willem M. Floor. Rural energy in developing countries: a challenge for economic development. *Annual Review of Energy and Environment*, 21:497–530, 1996.

[5] B. M. Campbell, S. J. Vermeulen, J. J. Mangono, and R. Mabugu. The energy transition in action: urban domestic fuel choices in a changing zimbabwe. *Energy Policy*, 31(6):553–562, 2003.

[6] Shubham Chaudhuri and Alexander S.P. Pfaff. Fuel-choice and indoor air quality: a household-level perspective on economic growth and the environment. *mimeo, Columbia University*, 2003.

[7] Mark Davis. Rural household energy consumption: the effects of access to electricity evidence from south africa. *Energy Policy*, 26:207–217, 1998.

[8] Esther Duflo. Poor but rational? *mimeo, MIT*, 2003.

[9] Majid Ezzati and Daniel M. Kammen. Evaluating the health benefits of transitions in household energy technologies in kenya. *Energy Policy*, 30:815–826, 2002.

[10] Majid Ezzati, Homayoun Saleh, and Daniel M. Kammen. The contributions of emissions and spatial microenvironments to exposure to indoor air pollution from biomass combustion in kenya. *Environmental health perspectives*, 108(9):833–839, 2000.

[11] W. Michael Hanemann and Barbara Kanninen. The statistical analysis of discrete response cv data. *mimeo, UC Berkeley*, 1996.

[12] R. H. Hosier and W. Kipondya. Urban household energy use in tanzania. *Energy Policy*, 21(5):454–473, 1993.

[13] Richard H. Hosier and Jeffrey Dowd. Household fuel choice in zimbabwe: an empirical test of the energy ladder hypothesis. *Resources and Energy*, 9:347–361, 1987.

[14] Bruce A. Larson and Sydney Rosen. Understanding household demand for indoor air pollution control in developing countries. *Social Science and Medicine*, 2002.

[15] Gerald Leach. The energy transition. *Energy Policy*, 20(2):116–123, 1992.

[16] Omar R. Masera, Barbara D. Saatkamp, and Daniel M. Kammen. From linear fuel switching to multiple cooking strategies: A critique and alternative to the energy ladder model. *World Development*, 28(12):2083–2103, 2000.

[17] K. E. McConnell. Consumer surplus from discrete choice models. *Journal of Environmental Economics and Management*, 1995.

[18] World Health Organization. *World Health Report 2002*. World Health Organization, Geneva, 2002.

[19] Sandeep H Patel, Thomas C. Pinkney, and William K. Jaeger. Smallholder wood production and population pressure in east africa: evidence of an environmental kuznets curve. *Land Economics*, 71(4):516–531, 1995.

[20] Mark Rosenzweig and Andrew Foster. Economic growth and the rise of forests. *Quarterly Journal of Economics*, 2003.

[21] Kenneth A. Small and Harvey S. Rosen. Applied welfare economics with discrete choice models. *Econometrica*, 1981.

[22] Kirk R. Smith. In praise of petroleum? *Science*, 298:1847, 2002.

[23] Shankar Subramanian and Angus Deaton. The demand for food and calories. *Journal of political economy*, 1996.

[24] Kenneth E. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York, 2003.

[25] UNDP and World Bank. Household energy use in developing countries a multicountry study. *mimeo, World Bank*, October 2003.

[26] UNDP and World Bank. Household fuel use and fuel switching in guatemala. June 2003.