Essays on Energy Demand and Household Energy Choice

Amin Karimu

Department of Economics
Umeå School of Business and Economics
Umeå University
Doctorial thesis 2013

Copyright © Amin Karimu
Umeå Economics Studies no.864
ISBN: 978-91-7459-690-8
ISSN: 0348-1018
Electronic version available at http://umu.diva-portal.org/

Printed by Print & Media at Umeå University
Umeå, Sweden, 2013

To the N-yo Family

Abstract

This thesis consists of four self-contained papers related to energy demand and household cooking energy.

Paper [I] examine the impact of price, income and non-economic factors on gasoline demand using a structural time series model. The results indicated that non-economic factors did have an impact on gasoline demand and also one of the largest contributors to changes in gasoline demand in both countries, especially after the 1990s. The results from the time varying parameter model (TVP) indicated that both price and income elasticities were varying over time, but the variations were insignificant for both Sweden and the UK. The estimated gasoline trend also showed a similar pattern for the two countries, increasing continuously up to 1990 and taking a downturn thereafter.

Paper [II] studies whether the commonly used linear parametric model for estimating aggregate energy demand is the correct functional specification for the data generating process. Parametric and nonparametric econometric approaches to analyzing aggregate energy demand data for 17 OECD countries are used. The results from the nonparametric correct model specification test for the parametric model rejects the linear, log-linear and translog specifications. The nonparametric results indicate that the effect of the income variable is nonlinear, while that of the price variable is linear but not constant. The nonparametric estimates for the price variable is relatively low, approximately −0.2.

Paper [III] relaxed the weak separability assumption between gasoline demand and labor supply by examining the effect of labor supply, measured by male and female working hours on gasoline demand. I used a flexible semiparametric model that allowed for differences in response to income, age and labor supply, respectively. Using Swedish household survey data, the results indicated that the relationship between gasoline demand and income, age and labor supply were non-linear. The formal separability test rejects the null of separability between gasoline demand and labor supply. Furthermore, there was evidence indicating small bias in the estimates when one ignored labor supply in the model.

Paper [IV] investigated the key factors influencing the choice of cooking fuels in Ghana. Results from the study indicated that education, income, urban location and access to infrastructure were the key factors influencing household's choice of the main cooking fuels (fuelwood, charcoal and liquefied petroleum gas). The study also found that, in addition to household demographics and urbanization, the supply (availability) of the fuels influenced household choice for the various fuels. Increase in household income was likely to increase the probability of choosing modern fuel (liquefied petroleum gas and electricity) relative to solid (crop residue and fuelwood) and transition fuel (kerosene and charcoal).

Key words: Choice Probability, derivatives, energy policy, gasoline demand, propensity score, tax simulation, unobserved trend

Acknowledgements

The journey of my PhD program has reach its final destination and I would like to say thank you to all the people who in one way or the other contributed in diverse ways in making this journey a successful one. Firstly, I would like to say a big thank you to the Almighty Allah for all the blessing and guidance over the years. It has been a wonderful and productive 3 years. Secondly, I would like to express my sincere gratitude and appreciation to all the support I had from my principal supervisor, Professor Runar Brännlund. You have provided me with all the support I needed, with your prompt reading of the manuscripts with constructive suggestions kept me on my toes. I would not have completed this work without your criticism and suggestions. You also kept your office open for me to pop in every now and then, with all my questions whether trivial or not. You were always ready to give your expert comments and suggestions, Thank you for all the comments, criticism and above all the trust and support during the past three years, I say Ayikoo in the Ashanti language (meaning welldone).

My sincere appreciation and gratitude also goes to Professor Kurt Brännäs, in his capacity as my second supervisor. Kurt has been very instrumental in my PhD education. You have over the three years of my PhD education supported me in diverse ways and provided comments, suggestions and even possible topics to consider in writing towards my PhD thesis. I have learned a lot of econometric logic from you and how to best market my reach papers for possible publication, this I gladly acknowledge by saying thank you.

I would also like to thank all members of staff, both in the department of economics and CERE for their support during my PhD studies, especially Marie Hammarstedt for all the administrative support and also checking each of the papers for possible errors, Jurate Jaraite, Andrius Kazukauskas, Chandra Kiran, Niklas Hanes, Tommy Lundgren and Göran Bostedt, for their comments and suggestions. Your comments have helped to improve the quality of my work.

Also Acknowledge here are my PhD year group (Tomas Raattamaa, Emma Zetterdahl, Elon Strömbäck and Mathilda Eriksson), you have all been supportive during the meandering road to my PhD title. A special thank goes to Catia Cialani, Tomas Raattamaa (formal office mate), Elena Kotyrlo, Liu Yuna Erik Brockwell and

Sevil Acar for all their support. Catia, Tomas and Elena made my life in Umeå much easier than it would have been. Thank you for all our conversations across all topics, from religion to economics, it helped me keep part of my brain open to the broader perspective of life while concentrating the other part on my economic research.

I would also like to show my deep appreciation to the JJC group (Manex Yonis a.k.a the King, Vicky Said a.k.a the Secretary, Reuben Nyong a.k.a the Spokesman and Nice Muhanzu a.k.a The Chair Lady) for the diverse support I received from you, especially our dinners and movie times, which helped kept a sound mind and kept me away from becoming all work and no play individual. Contrary to the old adage "keep your friends close and enemies much closer" I would say "keep your enemies close and friends much closer" and this I had in my heart with the JJC group. My appreciation and gratitude also goes to the Ghanaian community in Umeå, especially Mr. Ali and the family, and Bright Boye Kumordzi (PhD Candidate, SLU, Umeå) for their support.

I am also grateful to the following people whom I have discussed research ideas and academic life in general-Mahama Duwiejua (Professor,KNUST,Ghana), George Adu (Lecturer, Department of Economics, KNUST, Ghana), Paul Alagidede (Professor, University of the Witwatersrand, South Africa), Sunday Azagba (Postdoc, Dalhousie University, Canada), Yves Tehou (PhD candidate, Concordia University, Canada), Sharif Faisal (PhD candidate, Concordia University), Shadi El Ramli (PhD candidate, Concordia University) and Samuel Fosu(PhD candidate,University of Leicester).

Finally, I would like to thank the entire N-yo family, especially my Mum and Dad for all their support and the good upbringing they gave me. They stood by me during both the good and bad times. When things get really difficult, the N-yo family is always there to give all the support I needed. To my brothers and sister, I say thank you for being there for me and always treating me as the baby of the family, which is fun, but not being called "the baby of the family". I enjoy the attention associated with being the baby of the family but not the name. Allah blesses you all.

Umeå, August 2013

Amin

This thesis consists of a summary and the following four papers:

[I] Amin Karimu, 2011. Impact of economic and non-economic factors on Gasoline demand: a time varying parameter model for Sweden and the UK.

[II] Amin Karimu and Runar Brännlund, 2013, Functional form and aggregate energy demand elasticities: A nonparametric panel approach for 17 OECD countries, *Energy Economics*, 36, 19-27. (Reprinted with permission).

[III] Amin Karimu, 2012. Effects of demography and labor supply on household gasoline demand in Sweden: a semi-parametric approach.

[IV] Amin Karimu, 2013. Cooking fuel preferences among Ghanaian Households: an empirical analysis

1. Introduction

This thesis contains four single papers that address four key questions on energy demand that have not been fully addressed in the literature. The four questions are; (1) do gasoline price and income elasticities vary over time, and what are the effects of non-economic factors on gasoline demand? (2) Are the usual linear, log-linear and translog[1] functional specifications for aggregate energy demand models appropriate for the data generating process? (3) What are the effects of labor supply on household gasoline demand, especially on the income elasticity if labor supply is not weakly separable from gasoline demand? (4) What are the key potential factors that influence cooking fuel choice in Ghana?

The above questions are the focus of the thesis and we will discuss shortly why these questions are very important and why there is a need to undertake research that seeks to find answers to them. It is of specific importance in this age of environmental concerns due to rising economic growth and the high dependence on fossil fuels to propel this growth of the world economy.

In the literature on gasoline demand, the commonly used econometric models for estimating both price and income elasticities assume that elasticities are constant over time. This implicit

[1] The translog function represents a class of flexible functional forms. It is a second order approximation (in logarithms) for an arbitrary non-linear function. The translog function, an abbreviation of "transcendental logarithmic function" "was first proposed by Christensen, et al. (1971, 1973). Their motivation for the translog function was the problems of strong separability and homogeneity of the Cobb-Douglas and CES production functions.

assumption is only realistic if and only if the behavioural changes to both price and income take a long time to evolve; hence changes on an annual basis may therefore be insignificant and hence reasonable to assume constancy over time. For example, Bohi (1981), Drollas (1984), Sterner (1990), Goodwin (1992), Bentzen (1994), Samimi (1995), Eltony and Al-Mutairi (1995) all assumed constancy for the estimated parameters for a given period, classified into either a short-run or a long-run (only allow variation between a short and a long-run) based on either a dynamic model (with a lag-dependent variable as a right-hand side variable and with or without lags in the other explanatory variables) or within a cointegration framework. However, whether the variations in price and income elasticities are insignificant or not, and hence the validity of the constant parameter assumption, is an empirical issue that must be studied before making conclusions on any specific study.

Moreover, most of the literature on gasoline demand has only relied on direct economic drivers of gasoline demand, with most of the studies focusing on the response of gasoline demand to price and income changes. There is arguably the need to also try to quantify the impact of "non-economic" or indirect factors in order to aid our understanding and policy prescription in curtailing the demand for gasoline, and hence associated CO_2 emissions, rather than relying only on the direct economic drivers for policy prescription. For instance, a study by Allcott (2009), on consumer behaviour and electricity consumption in Minnesota (USA), found that information, attention and social norms are important factors that influence consumer behaviour regarding electricity consumption in Minnesota. We argue that this is not only limited to electricity but also to other forms of energy.

The question of the appropriate functional form for energy demand models is even more important now than before as there is an increasing demand on how to design appropriate policies to curb the reliance on fossil fuels and how to "decouple" economic growth and energy use. An obvious benefit for using the appropriate functional specification for empirical energy demand models is that it will help reduce the possible bias associated with using wrong specifications for estimating energy demand models, and consequently improve energy policy effectiveness. The choice of the "appropriate" functional form is more of an empirical issue, as theory does not provide sufficient guidance as to the correct functional form. Hence a priori assigning a specific functional form for any econometric model may have serious implications for the estimates if the specification is wrong. For instance, if the assumed functional form is not the correct one, the estimates may be biased and inconsistent as indicated in Li and Racine (2007).

In a review by Blum et al (1988) on aggregate time series gasoline demand studies for West Germany and Austria, they found a wide range in the estimated price (-0.25 to -0.83) and income (0.86 to 1.90) elasticities. The authors attributed most of the variations to the restrictive functional forms used for the estimations. Furthermore, a study by Zarnikau (2003) on functional forms in energy demand modelling found that the three most used parametric functional forms for energy demand models (linear, log-liner and translog) all performed poorly using a specification test on United States household data. Xiao et al. (2007) used a Bayesian approach to assess the accuracy of using linear, log-linear, translog and AIDS[2] specification for analysing the underlying relations for household electricity demand for the United States. The result from Xiao et al. (2007) indicated that the AIDS model is slightly better than the translog model, which is superior to log-linear, which in turn is better than the linear model. The above suggests the need to further study the issue of functional form, especially on panel data sets, as the studies so far on functional forms have only used time series or cross-sectional data that do not allow analysis across sections and time periods, which paper II in this thesis, seeks to address.

The third question is important due to the fact that in the demand literature, most studies assume that labor supply and demand for commodities are weakly separable, implying that our decisions on labor supply do not influence our consumption choice for goods and services, which on casual observation seems to be at odds with reality. Studies by Browning and Meghir (1991), and Brännlund and Nordström (2004) found that labor supply and demand for goods are not weakly separable, using a system of demand model (AIDS model and versions of it). Previous studies on goods demand and labor supply suggests that labor supply matters in the demand for goods.

---

[2] AIDS is an abbreviation for "Almost Ideal Demand System". The AIDS model was proposed by Deaton and Muellbauer (1980). AIDS gives an arbitrary first-order approximation to any demand system and possesses most of the properties that are required for a demand system· Only some of these properties are inherent in the translog model (see Deaton and Muellbauer, 1980)

Therefore, investigating this further on gasoline demand by relaxing the weak separability assumption between gasoline demand and labor supply will hopefully advance our knowledge in this area. Besides, it is also important to try to understand the factors that influence gasoline demand at the household level as emissions from transport constitute a major source of global CO_2 emission and a better understanding of this will help in the design of the appropriate policies to help in reducing emissions generated from this sector.

A concern among policy analysts, especially those in the developing world, is the high dependence on fuelwood as a choice of energy for cooking, and the implication of this dependence on the health and poverty levels, especially for women. A World Bank report (1993) indicates that the high use of fuelwood significantly contribute to health problems via the smoke generated from biomass fuels, especially fuelwood. The health problems in turn contribute to the consequences of poverty in the developing world via the health cost and also the opportunity cost associated with the gathering and use of fuelwood, especially on children and women. High dependence on fuelwood for cooking is a common feature of the Ghanaian economy. A report by the Ghana Statistical service (2008) indicates that, approximately 69% and 58% of Ghanaian households in 1990 and 2005 used fuelwood as the main cooking fuel, respectively. Irrespective of the decreasing trend, 58% is still very high and therefore a clear understanding of the key variables driving the high dependence on fuelwood is necessary for a better understanding of how energy policy should be designed in Ghana.

The first paper in this thesis investigates issues such as the impact of non-economic factors on aggregate gasoline demand, and whether the price and income elasticities vary over time for Sweden and the UK. The second paper investigates the issue of functional form and the impact of this on price and income elasticities by applying a non-parametric approach on 17 OECD countries. The third paper on the other hand addresses the question relating to labor supply and gasoline demand and whether they are weakly separable. The last paper addresses the factors that influence the choice of cooking fuel in Ghana, using survey data for Ghanaian households to aid our understanding about what to include in the energy policy package in Ghana to promote the use of less polluting and less time consuming fuels such as liquefied petroleum gas (LPG).

2. Energy system

Before explaining what an energy system is, it is important to give a summary of two key classifications of energy that are commonly used in the literature as well as in this thesis. The two classifications are;

a. Primary and secondary energy.
b. Renewable and non-renewable energy.

There are other classifications such as commercial/non-commercial, and conventional/non-conventional energy, but these latter classifications are not used in this thesis, hence we will not elaborate on this but refer the reader to Bhattacharyya (2011) for details on the other two classifications.

Primary and secondary energy

Primary energy refers to an energy source that is extracted from a stock of natural resources or captured from a flow of resources, which has not undergone any transformation or conversion other than separation and cleaning (IEA, 2004). Secondary energy on the other hand refers to an energy source derived from primary energy via a transformation or conversion process. Based on the above definitions, crude oil, coal and natural gas are good examples of primary energy, while electricity and oil products such as gasoline and diesel are good examples of secondary energy.

Renewable and non-renewable energy

Non-renewable energy comes from sources where the energy resource is consumed much faster than nature can create it (examples include crude oil and coal). This is a more general definition that applies to all stocks of resources for which the consumption rate of the resource is higher than the natural rate of growth of the resource to replenish the stock of the resource. For instance, if the consumption of fuelwood exceeds the rate of growth of forests, fuelwood in such a situation can be classified as a non-renewable. Renewable energy on the other hand is energy derived from resources that are continually replenished such as wind, solar energy, and tides.

In order to have a clear picture of the dynamics of the energy sector, it is important to outline the various activities that are involved in the energy supply chain that we refer to as the “energy system”. Three key sectors are involved in the energy system and these are: the energy supply sector, the transformation sector, and the consuming sector. The three sectors are interlinked to ensure the supply of and demand for energy. The supply sector involves domestic production (this is a process that includes the discovery and extraction of energy resources and the capture of energy in the case of energy resources such as solar energy), the import or export of energy resources and stock changes via stock acquisition and drawdown. The transformation sector includes the transportation of energy resources to conversion centres for processing and onward to the final consumer, and the processing of primary energy such as crude oil to secondary energy such as gasoline, diesel and other forms of oil products. The consuming sector on the other hand is made up of the final users of energy and involves demand for energy for various purposes such as heating, running industrial machines, lighting needs and cooking, amongst other things, and hence comprises the final consumer of energy. Consumers of energy are commonly categorised into broad groups such as industrial, transport, residential, services and agricultural sectors. The grouping is to allow for easy analysis, and also because there are supposed to be differences in demand responses across the various categories, partly due to how energy is used and the type of energy appliances or equipment used by each category. However, this thesis only focuses on the consuming sector in the energy system as all the papers are related to energy demand; hence we will expand more on this sector of the energy system. Figure 1 presents a broad overview of the energy system (energy supply system) with the linkages within the system.

The energy consumption sector

The total final energy consumption (TFEC) can be analyzed via the various energy carriers (the various petroleum products, natural gas, electricity, nuclear, coal and the various renewable energy carriers) or through the broad user categories such as industrial, transport, residential, commercial and public service, and agricultural sector. If one considers the total energy consumption based on share of fuels, three fuels (oil, natural gas and coal) have dominated in share volumes since 1973, with a combined share of 75.1% in 1973 (IEA, 2012).

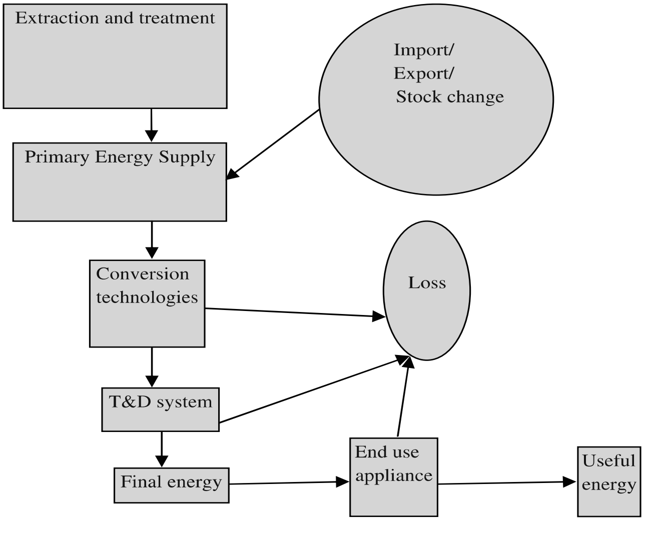
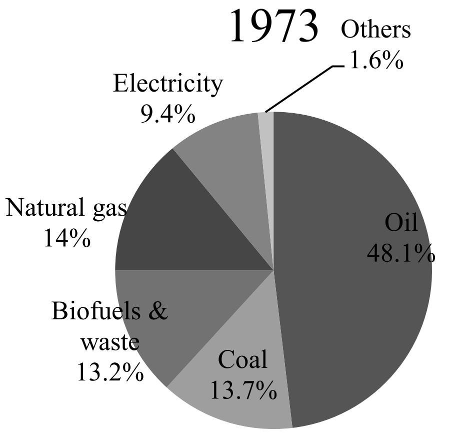


Figure 1: Energy supply system, source: Bhattacharyya (2011)

The dominance of oil, natural gas and coal is gradually declining, as the combined share of total final energy consumption in 2010 was 66.2%, which is still high considering the environmental impact of these fuels. For instance the combined share of global CO_2 emissions (based on fuel combustion only) from the above three fuels in 1973 and 2010 was 100% and 99.6%, respectively (IEA, 2012). The contribution of the various fuels to total final energy consumption is presented in Figure 2 below, where others include geothermal, solar, wind and heat. If one considers regional shares of TFEC, OECD countries as a region constitute the biggest consumer, consuming 60.3% and 42.5 of TFEC in 1973 and 2010, respectively. China has replaced non-OECD Europe as the second biggest consumer of final energy in 2010, with a share of 17.5% of TFEC (IEA, 2012).

Furthermore, considering the sectoral shares of world oil consumption, the transportation sector is the largest consumer, accounting for 45.4% and 61.5% of the total oil consumption in 1973 and 2010, respectively. Irrespective of all the environmental concerns, oil is still a dominant fuel in the world and will continue to be so in the near future due to many factors, including the increasing use of personal vehicles in the developing and middle income countries, as well as the small market share of renewable energy in total energy consumption, which is due to the high cost of renewable energy and technological constraints.



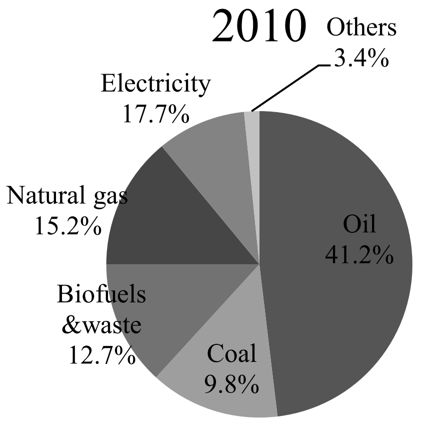


Figure 2: Share of fuels in total final energy consumption in 1973 and 2010, source: IEA, 2012

3. Summary of the Papers

The specific issues addressed in this thesis and within each of the four papers are summarized below with details appended for interested readers. The summaries are based on methodology, contribution and findings from each of the papers included in the thesis.

Paper [I]: Impact of economic and non-economic factors on gasoline demand: a time varying parameter model for Sweden and the UK

This paper investigates the impact of non-economic factors, such as technical progress and preferences (taste) on gasoline demand, and also whether both price and income elasticities vary over time. Furthermore, the impact of various tax scenarios on long-term gasoline demand in Sweden and the UK is forecast. The theoretical background for the study is cast in the utility maximisation principle, which we incorporate in a multi-stage decision-making process as in Edgerton (1997). This means that in order to derive the uncompensated demand function for gasoline an assumption of weak separability in consumer preferences is imposed. The derived gasoline demand, used in the empirical estimation is then a function of real gasoline price, real diesel price, real income, and an unobserved gasoline demand trend that is assumed to be a proxy for technological progress and preferences ("taste" for gasoline).

In this study, unlike many other studies (reviewed by Bohi (1981), Dahl and Sterner (1991), Espey (1998) and Dahl (2012)), we model gasoline demand in a structural time series framework, following the work of Broadstock and Hunt (2010). This approach allows the gasoline trend to be estimated more flexibly via a stochastic process rather than the usual deterministic time trend that is often used in the literature to account for technological progress and taste (taste refers to socio-economic factors such as lifestyles and effects of values and norms). The structural time series modelling approach is based on Harvey (1989), in which a maximum likelihood estimation procedure is combined with Kalman filtering to estimate the "unobserved" trend in a more flexible way. The approach allows both the level and slope of the trend to follow a stochastic process, without a priori assigning numbers such as 1, 2, 3, etc. as data points for estimating the trend.

The econometric approach for the study consists of three steps; in the first step we estimate a constant parameter model (CPM) with an autoregressive structure for both Sweden and the UK. The reason for this is to identify the parameters that are likely to be significant in the time varying parameter-structural time series model (TVP-STSM) and also to assess if the estimates from the TVP-STSM are significantly different from those from the CPM. In the second step, we allow all variables in step one to vary over time and we assess the pattern of the unobserved gasoline demand trend for both Countries. In the final step we make a long-term forecast for gasoline demand in each of the countries using three tax scenarios based on the time varying parameter model.

The results indicated that non-economic factors did have an impact on gasoline demand and also one of the largest contributors to changes in gasoline demand in both countries, especially after the 1990s. The study further indicated that both price and income varies over time but the variations are insignificant on yearly basis. Estimated gasoline trend for both countries depicts similar pattern, rises until 1990 and decreases thereafter. The results from the simulation indicated that the projected gasoline consumption decreases over time with the highest reduction occurring with the high tax scenario in the high world oil price case for both countries, but the relative impact of taxes is rather high with the low world oil price case compare to both the reference and high world oil price cases. Furthermore, the simulation indicated that gasoline tax is likely to be a more effective gasoline reduction policy tool in Sweden than in the UK due to the stronger response to price changes in Sweden relative to that of the UK.

Paper [II]: Functional form and aggregate energy demand elasticities: A nonparametric panel approach for 17 OECD countries

In the empirical literature on aggregate energy demand the usual functional specification assumed is that of linear, log-linear and translog. However, in most cases whether the a priori specification is appropriate or not for the data generating process is not tested for, and therefore conclusions drawn on the estimates only hold if the assumed specification is indeed the correct one for the data. If the assumed specification is wrong, the estimates may be biased and this can lead to incorrect inferences and hence conclusions will be invalid and therefore may have serious consequences on policies prescribed based on the biased estimates. Getting the correct functional specification is therefore extremely important in this age of great concern about world energy demand and how policy can influence this demand in the right direction.

This paper therefore contributes to the literature on testing the functional form on energy demand by extending the work by Zarnikau (2003) and Xiao et al. (2007) in three key areas; firstly, Zarnikau

(2003) and Xiao et al. (2007) did not consider issues such as cointegration and error correction that distinguishes short-run and long-run estimates, and given the long time period in our data, the issue of spurious regression cannot be underestimated. Hence a cointegration econometric approach becomes important for this study. The second area of extension is that the previous papers only investigated the issue based on time series data. In this paper, we extended that into a panel data setting that incorporates both a time and cross-section dimension of the data. Finally, this paper also considers a nonparametric estimation of the model if the nonparametric test of each of the usually assumed parametric specifications is rejected. This was not done in the previous papers.

The paper combines the standard utility maximisation theory and applied econometric model to address the main issues of the paper. This means that the framework underlying the econometric analysis in this study is a derived demand framework. Since the empirical analysis concerns aggregate energy use, demand comes from final demand from households and from demand from firms for energy as an input. Implicitly we assume that firms choose the amount of energy in order to minimize costs (or maximize profits), and that households are utility maximizers with respect to energy consumption (and other consumption goods). We assume a multistage decision process for households and firms, where in the first stage they decide between energy and non-energy goods, and in the second stage they decide how much energy they will consume given budgetary and technological constraints. Based on this we express energy demand as a function of the real price of energy, real income and a vector of demand shifters such as technology and climate effects. The econometric procedure used in estimating the derived energy demand is as follows. We first determine the unit root properties of the variables in the model, and if they are not stationary then in the second step we estimate the model using the error–correction approach of Westerlund (2007). The next step is to test whether each of the assumed specifications is appropriate using a nonparametric specification test. If the test rejects all of the parametric specifications, we estimate the demand model nonparametrically.

The panel consists of 17 OECD countries over the period 1960 – 2006. All the variables in the demand model are found to be non-stationary, and therefore a cointegration approach is necessary to ascertain that a long run relationship do exist between energy demand and the covariates in the demand model. Furthermore, the cointegration test confirmed a long run relationship exist between energy demand, price, income and demand shifters. It also confirmed that both price and income are weakly exogenous in the demand model. The functional specification test rejected each of the parametric specification and a further nonparametric estimation, using the approach developed in Li and Racine (2007), revealed that the demand response to price is linear but not constant, while income response is non-linear and varies (decreases) as income increases. This implies that energy tend to be a luxury good at low-income levels, whereas energy moves towards a normal good, or a necessity as income rises.

Paper [III]: Effects of demography and labor supply on household gasoline demand in Sweden: a semi-parametric approach

The usual approach taken in empirical demand estimation is to impose a weak separability assumption between demand for goods and labor supply. Whether or not this is appropriate is not normally tested. Studies by Browning and Meghir (1991) and Brännlund and Nordström (2004) found the separability assumption to be invalid for their respective system of demand equations. This paper builds on the idea of Browning and Meghir (1991), Brännlund and Nordström (2004) by focusing on gasoline demand and extending the parametric case of the stated papers to a flexible specification where both male and female working hours as well as income and age of the "household head" is included. This further extends the work of Hausman and Newey (1995, 1998), Schmalensee and Stoker (1999), Yatchew and No (2001), and Wadud et al. (2010). The main idea is to relax the separability assumption by allowing for the possible effects of male and female working hours on gasoline demand as well as allowing for the effects of demographic variables.

The theoretical background for this paper is that of the standard household consumer theory, and within this framework we can derive the gasoline demand function for households to depend on gasoline price, income, a vector of household characteristics, location of the household, and labor supply. The latter approximated by male and female working hours. Using data from Swedish family expenditure surveys for (1985, 1988 and 1992) we estimate a household demand model using a semiparametric econometric model. The reason for using a semi-parametric model was to relax the parametric linear restriction of some of the key variables such as labor supply, income, and age of the “household head”. In addition, due to the occurrence of zeros in the gasoline expenditure recording, we applied a semiparametric selection model proposed by Das et al. (2003), which generalises the Heckman selection model that is usually used to correct for selection bias. This flexible modelling approach further allows for various levels of income to have different effects on gasoline demand. Similarly, it also allow different working hours and age to have different effects on gasoline demand, which is more realistic with sound intuition than constant response effects to these variables. This paper therefore makes two key contributions to the literature; the first contribution is in the estimation of a flexible semi-parametric gasoline demand model that allows for possible non-linearity and interactions between income, age, and labor supply of the households. The second contribution of the paper is in the modelling and correction for selection bias in a more flexible framework that reduces misspecification errors and is likely to lead to more reliable estimates.

Our study revealed the following findings. Gasoline demand and labor supply are not separable, implying the weak separability assumption is rather restrictive and therefore complements the finding in Browning and Meghir (1991), and Brännlund and Nordström (2004). The second key finding is that both male and female working hours have a positive effect on gasoline demand for working hours of less than 38 per week, while turning to a negative impact for working hours greater than 38 for female. The negative effect for male working hours only occurs within the working hours bracket of 38 to 55 hours, and turns into a positive effect for hours greater than 55 hours per week. This finding on the dynamics of labor supply on gasoline demand can be explained by the amount of leisure time and the possible activities that males and females engage in during this time. The other key finding is that the effect of income is non-linear and varies across income level. Besides, age of "household head" over 40 years old tends to reduce gasoline consumption and this is consistent with other studies in the literature on the negative effect of old age on gasoline demand.

Paper [IV]: Cooking fuel preferences among Ghanaian Households: an empirical analysis

In most developing countries, fuelwood is a major household energy source especially for cooking. However, the use of fuelwood in developing countries has serious health, environmental, and economic consequences, such as loss of productivity, either due to poor health as a result of polluted air, or time spent in gathering fuelwood at the expense of working or studying. It is therefore important to better understand the factors that influence the choice of this source as well as cleaner alternatives such as liquefied petroleum gas (LPG) to aid energy policy formulation. In Ghana fuelwood accounts for approximately 57.8% of total household cooking fuels used in 2005. This heavy dependence on fuelwood is making it difficult for Ghana to achieve the United Nations millennium development program due to the indirect effect of this dependence on health and poverty in general. Based on the so-called “energy ladder hypothesis”[3] the study examines the key factors influencing the choice of modern (LPG and electricity), transition (kerosene and charcoal) and solid fuels (crop residue and fuelwood). Moreover, we also study the factors influencing the most used cooking fuel in each of the three groups based on the “energy ladder hypothesis” (the most used fuels are fuelwood, LPG and charcoal).

The econometric strategy is a two-step procedure. In the first step we estimated a multinomial probit model for the three groups of fuels, using transition fuel as the reference fuel. In the second step we estimate a multinomial probit model on the most used fuel in each group, using the other fuels (all fuels that are not the most used fuel in each of the groups) as reference fuels. As a check as to whether grouping all the fuels that are not the most used cooking fuel as a reference group could influence the result in a certain direction, we also estimated a multinomial probit model on all the fuels constituting the three groups. The results from this check indicated no significant differences in the estimates, we did not report this result in the thesis but it can be provided on request.

The results indicated that household characteristics, such as age, gender, size of the household and education significantly influenced the choice probabilities for the three groups of fuels. Moreover, focusing on the three most used fuels; we found similar effects from

[3] The energy ladder hypothesis states that at low level of income, households tend to use biomass fuel as the main cooking fuel. However, as income level increases, households replace biomass fuel with “transition” fuels such as kerosene and further switch to “modern” fuels such as LPG and electricity, as income further increases. In this hypothesis, biomass fuel is considered to be at the bottom of the ladder, while LPG and electricity at the top of the ladder.

the same household characteristics, which is consistent with previous studies by Heltberg (2004, 2005) and Ouedraogo (2006). In addition, the results further indicated that availability of the various fuels has significant influence on the choice probabilities. Specifically, whereas the availability of LPG increases the choice probability for LPG, the availability of kerosene on the other hand decreases the adoption probability for LPG, while the availability of charcoal and fuelwood has no significant effect on the choice of LPG. In the case of fuelwood, both the availability of LPG and charcoal significantly reduces the choice probability for this fuel, while the availability of fuelwood and kerosene tends to increase fuelwood adoption probability. Moreover, the availability of charcoal, as expected, increases its choice probability. The findings also indicated that when we grouped the fuels into modern, transition and solids, availability of the various fuels still has significant impact on the choice probabilities.

Another interesting finding from this paper is that urbanisation tends to increase the adoption probability for modern fuel, while it decreases the adoption probability for solid fuel. Access to modern infrastructure also increases the adoption probability for modern fuel and decreases the adoption probability for solid fuel. Overall the results clearly show that fuel choice is very strongly connected to factors on both the demand and supply side, which in turn are very strongly related to economic development in a broad sense.

4. References

Allcott, H., 2009. Social norms and energy conservation, MIT Centre for Energy and Environmental policy Research (MIT CEEPR), WP-2009-014.

Bentzen, J., 1994. An empirical analysis of gasoline demand in Denmark using cointegration techniques, *Energy Economics*, 16, 139-143.

Bhattacharyya, S.C., 2011. *Energy Economics: Concepts, Issues, Markets and Governance*, Springer-Verlag London Limited.

Blum, U., Foos, G. and Guadry, M., 1988. Aggregate time series gasoline demand models:review of the literature and new evidence for West Germany, *Transportation search A*, 22A, 75-88.

Bohi, D.R., 1981. *Analysing demand behaviour: a study of energy elasticities*, Johns Hopkins University Press: Baltimore, U.S.

Brännlund,R., and Nordström,J.,2004. Carbon tax simulations using a household demand model. *European Economic Review,* 46, 211 – 233.

Broadstock, D.C., and Hunt, L.C., 2010. Quantifying the impact of exogenous non-economic factors on UK transport oil demand, *Energy Policy*,38(3), 1559-1565

Browning, M. and Meghir, C., 1991. The effects of male and female labor supply on Commodity demands, *Econometrica,* 59,925 – 951.

Christensen, L., Jorgenson, D. and. Lau, L., 1971. Conjugate Duality and the Transcendental Logarithmic Production Function, *Econometrica,* 39, 255- 256.

Christensen, L.R, Jorgenson, D.W. and Lau, L.J., 1973. Transcendental Logarithmic Production Frontiers, *The Review of Economics and Statistics,* February, 28-45.

Dahl, C. and Sterner, T., 1991. A survey of econometric gasoline demand elasticities, *International Journal of energy Systems*, 11, 53-76

Dahl, C.A., 2012. Measuring global gasoline and diesel price and income elasticities. *Energy Policy*, 41, 2–13.

Das, D., Newey, W.K., and Francis, V., 2003. Nonparametric estimation of sample selection Models, *Review of Economics Studies,* 70, 33 – 58.

Deaton, A. and Muellbauer, J., 1980. An Almost Ideal Demand System, *The American Economic Review*, 70(3), 312-326.

Drollas, L., 1984. The demand for gasoline: further evidence, *Energy Economics*, 6, 71-82.

Edgerton, D.L., 1997. Weak separability and the estimation of elasticities in multistage demand system, *American Journal of Agricultural Economics*, 79(1), 62-79.

Eltony, M. and Al-Mutairi, N., 1995. Demand for gasoline in Kuwait: an empirical analysis using cointegration techniques, *Energy Economics*, 17, 249-53.

Espey, M., 1998. Gasoline demand revisited: an international meta-analysis of elasticities, *Energy Economics*, 20, 273-295.

Ghana Statistical Service, 2008. *Ghana Living standard survey: report of the fifth round (GLSS5)*.

Goodwin, P., 1992. A review of new demand elasticities with special reference to short and long run effects of price changes, *Journal of Transport Economics and Policy*, 26, 155-63.

Harvey, A.C., 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*,Cambridge University Press: Cambridge.

Hausman, J. and Newey, W.K., 1995. Nonparametric estimation of exact consumer surplus and deadweight loss. *Econometrica*, 63 (6), 1445–1476.

Hausman, J. and Newey, W.K., 1998. *Nonparametric estimation of exact consumer surplus and deadweight loss,* Econometrics and economic theory in the 20th century. Cambridge: Cambridge University Press, 111 – 146.

Heltberg, R., 2004. Fuel switching: Evidence from eight developing countries, *Energy Economics* 26:869–887.

Heltberg, R., 2005.Factors determining household fuel choice in Guatemala,*Environment and Development Economics*, 10, 337-361

IEA, 2004. Energy statistics manual, International Energy Agency, Paris(seehttp://www.iea.org/textbase/nppdf/free/2005/ statistics_manual.pdf).

IEA, 2012. Key World Energy Statistics (see http://www.iea.org /publications/freepublications/publication/kwes.pd)

Li, Q. and Racine, J.S., 2007. *Nonparametric econometrics: theory and practice*, Princeton University Press.

Ouedraogo, B., 2006. Household energy preferences for cooking in urban Ouagadougou, Burkina Faso, *Energy Policy*, 34, pp.3787-3795.

Samimi, R., 1995. Road transport energy demand in Australia: a cointegrated approach, *Energy Economics*, 17, 329-39.

Schmalensee, R., Stoker, T.M., 1999. Household gasoline demand in the United States. *Econometrica*, 67 (3), 645–662.

Sterner, T., 1990. The Pricing of and Demand for Gasoline, Swedish Transport Research Board: Stockholm.

Wadud, Z., Graham, D.J., Noland, R.B., 2010. A semiparametric model of household gasoline demand. *Energy Economics*, 32, 93-101.

Westerlund, J., 2007. Testing for error correction in panel data, *Oxford Bulletin of Economics and Statistics*, 69,709-748.

World Bank, 1993. *World Development Report*, Oxford University Press, New York.

Xiao, N., Zarnikau, J. and Damien, P., 2007. Testing Functional Forms in Energy Modeling: An application of the Bayesian Approach to U.S. Electricity Demand, *Energy Economics*, 29(2), 158-166.

Yatchew, A, No, JA, 2001. Household gasoline demand in Canada, *Econometrica,* 69(6), 1697–1709.

Zarnikau, J., 2003. Functional forms in energy demand modelling, *Energy Economics,* 25, 603-613.

I

Impact of economic and non-economic factors on gasoline demand: a varying parameter model for Sweden and the UK.

Amin Karimu
Department of Economics
Umeå School of Business and Economics
Umeå University

Abstract

This paper investigated the impact of price, income and non-economic factors on gasoline demand using a structural time series model. The results indicated that non-economic factors did have an impact on gasoline demand and also one of the largest contributors to changes in gasoline demand in both countries, especially after the 1990s. The results from the time varying parameter model (TVP) indicated that both price and income elasticities were varying over time, but the variations were insignificant for both Sweden and the UK. The estimated gasoline trend also showed a similar pattern for the two countries, increasing continuously up to 1990 and taking a downturn thereafter.

Keywords: Gasoline demand, Time Varying Parameter, Tax Simulation, Unobserved Trend

JEL Classification: C22, Q41, Q4

1. Introduction

In the past decades both the UK and Sweden have experienced an increasing share of transport oil in their total energy mix, irrespective of the increasing campaign and polices targeting the reduction of transport oil in favor of less polluting sources of fuel such as biofuels, coupled with the emphasis in both countries on energy efficiency as a whole and fuel efficiency in particular. Besides these similarities, there are remarkable differences in transportation culture between the two countries, which make it interesting to study, especially the importance of smart driving choices by consumers for the possible reduction in road transport gasoline consumption.

According to the International Energy Agency (IEA, 2010) oil use in the transport sector in the UK increased from just 66% in 2000 to around 70% in 2009. In Sweden, oil use in road transport was around 95% of total fuel use in this sector in 2007, (IEA, 2008). Gasoline is still the dominant transport oil used in both Sweden and the UK (for personal passenger transport), although in recent years its share on total transport oil demand in these two countries has been gradually declining. The future forecast for fossil fuel use in the transport sector is expected to increase due to increasing demand in both countries but much of the increases are likely to be due to increases in the demand for diesel rather than gasoline, partly due to policies favoring diesel and also due to the efficiency of diesel engines compared to petrol engines.

It is therefore clear that for both countries to achieve significant reductions in their respective CO_2 emissions from fossil fuels, more reductions have to come from gasoline both in terms of volumes and efficiency. As indicated earlier, gasoline demand is one of the major contributing sources of aggregate energy demand in these two countries and hence there is a need for policy to shift some of the demand from this source of fuel to less polluting sources of fuel. For any policy to be effective in reducing demand for gasoline and the associated CO_2 emissions there is the need to have knowledge about how consumers respond to changes in prices, taxes and income. Thus, there is a need for accurate estimates of both the price and income elasticities for gasoline demand. This is because for any policy to have the desired impact on gasoline demand and CO_2 emissions, the responsiveness of demand to price changes and incomes changes or both, will be the main determining factor on the level of success of the policy (assuming there is a political will to implement the policy).

The response on the supply side is also important for the success of any policy to reduce gasoline consumption via, for example, taxes, but the degree of importance depends on the magnitude of demand elasticity. In this study, the supply side of the market is not considered; the emphasis is rather on modeling gasoline demand and deriving its elasticity with respect to price and income in particular and to assess how non-economic factors such as technology and preference (taste) affect gasoline demand. Given that I study separate national markets that only account for a small fraction of world demand, it is reasonable to assume that the price is exogenous and hence do not necessarily have to consider the supply side.

In the literature various econometric models have been implemented over the years in estimating gasoline demand elasticities and some of the popular models include; static models, dynamic models, vehicle stock models, vehicle characteristic models, lagged endogenous models, inverted-v models and recently cointegration and structural time series models. These models have produced price and income elasticities for individual countries and groups of countries using time series, cross sectional and panel data sets. In addition there is a group of models using micro data within countries, thus focusing on differences between different types of households within a country or region (see for example Archibald & Gillingham, 1980, Schmalensee & Stoker, 1999, Yatchew & No, 2001, Brännlund & Nordström, 2004, Wadud et al, 2010). There are a few meta-analyses as well such as the Espey (1998) meta-analysis that aims to determine factors that significantly influence variations in gasoline demand elasticities. But it is still not that clear regarding the factors that account for the variations in gasoline demand elasticities, at least not a general consensus in this regard.

The main objective of this paper is to estimate the underlying gasoline demand trend that is referred to in the paper as the unobserved energy demand trend (UEDT)[1] for gasoline which is assumed to capture technology and other factors such as taste, and the impact of this on gasoline demand. The other objectives include; assessing whether price and income elasticity vary over time, and

[1] In this study, non-economic factors refer to UEDT. The assumption is that the UEDT capture factors such as taste and changes in technology, which are not direct economic factors for instance, when dealing with demand for gasoline. Price and Income on the other hand are classified as economic factors in this study.

finally to investigate the impact of various gasoline tax scenarios on forecast gasoline demand up to the year 2020. In addition to the projection on the level of gasoline demand, the simulations also project future CO_2 emissions, originating from gasoline consumption, in Sweden and the UK.

In achieving the above objectives, I apply a time varying parameter model within a structural time series-modeling (STSM) framework (Harvey, 1989) on a data set from Sweden and the UK. In the STSM framework, the UEDT is estimated rather than just included through a deterministic time trend. This method will accurately account for technological change as well as changes in the preferences for gasoline. The time varying nature of the parameters makes it possible to estimate elasticities for each year and therefore it contains more information, which in turn will enhance the predictive power of the model (for details on the time varying parameter model, see Chow (1984), Park and Hahn (1999), Park and Zhao (2010)). In the gasoline demand literature, the only papers to the best of my knowledge that have applied time varying parameter models are Park and Zhao (2010), and Dilaver (2010).

The rest of the paper is organized as follows; Section 2 contains a short literature review of gasoline demand elasticities, Section 3 deals with the model for the study and the application of the STSM to the data set. Section 4 contains the data and econometric analysis of the paper as well as the demand forecast and tax simulations. Finally, Section 5 contains some concluding remarks and some thoughts about future research.

2. Literature review of gasoline demand models and elasticities

The following gasoline demand surveys exist in the literature; Bohi (1981), Drollas (1984), Dahl (1986), Dahl and Sterner (1991), Espey (1998), Sterner and Dahl (1992), Goodwin et.al (2004), and Dahl (2012). All these surveys review previous empirical studies on gasoline demand, which cut across different methodologies, countries and data sets. Summaries of the above surveys on estimated price and income elasticity are presented in Table 1. Also included in Table 1 are UK studies (not surveys) based on STSM, since I could not find in the literature studies using the STSM approach on Sweden. An interesting feature from the surveys in Table 1 is that, most of the studies focus only on economic factors such as price and income in their empirical investigation, without any effort to investigate the possible contribution of non-economic factors such as taste on gasoline demand, except the individual studies on the UK by Hunt and Nimamoya (2003), Dimitropoulos et al. (2005), and Broadstock and Hunt (2010). However, in order to have a clearer understanding of the key drivers of gasoline demand/consumption, there is arguably the need to quantify, not only the economic factors, but also the impact of non-economic factors.

Table 1: Summary of the literature on gasoline demand elasticities

Surveys	Price Elasticity		Income elasticity		Data
	SR	LR	SR	LR	
Dahl & Sternar(1992)	-0.24	-0.79	0.41	1.17	Time
Goodwin (1992)	-0.27	-0.73			Time/cross
Espey(1998)	-0.26	-0.58	0.47	0.88	Time/cross
Graham&Glaister(2002)	-0.2 -0.3	-0.6 -0.8	0.35 0.55	1.1 1.3	Time/cross
Goodwin et. al (2004)	-0.01 -0.57	0 -1.81	0 0.89	0.27 1.71	Time series
Goodwin et. al (2004)		-0.28 -0.77		0.02 1.34	Time series
Dahl (2012)		-0.11 -0.33		0.66 1.26	Time/cross
		STSM For UK			
Hunt&Nimamoya(2003)		-0.12		0.80	Time series
Broadstock&Hunt(2010)		-0.10		0.57	Time series
Dimitropoulos et al.(2005)		0.11		0.81	Time series

Note: All implies all models including static and dynamic specification but excluding STSM; Time/cross present both time series and cross sectional data, reviews with both short and long run estimates are based on dynamic model and also include studies based on cointegration.

Another feature of the literature surveys as presented in Table 1 is that the dynamic models produced larger long-run elasticity estimates relative to the static model. Furthermore, evidence from previous studies (Goodwin et al., 2004) indicated that price elasticity tend to be negatively related to income and would tend to fall over time as income rises. Both price and income elasticity also tends to vary based on model and method used for the estimation and period for the analysis (pre 1974, 1974-81 and post 1981). A study by Foo (1986) indicated that variables such as the level of economic activity (employment, retail sales and industrial activity), had a significant impact on the income elasticity for West Germany as the reported income elasticity with the inclusion of the level of activity variables reduced the income elasticity to 0.25, but adding the elasticities of the level of activity variables to that of income gives a total elasticity of 1.22 (Blum et al., 1988).

Espey (1998) survey is quite novel in the gasoline literature, instead of further documenting the variability in the elasticity estimates for gasoline demand (as done in the other surveys), she carried out a "meta-analyses" that seeks to answer the question regarding the causes of the variations in the various estimates in the literature. In the study, Espey reviewed extensive articles published between 1966 and 1997. In order to answer her research question, she hypothesized four channels via which the variation in the estimates could potentially come from; thus from the demand specification, data characteristics, "environmental" characteristics (the level of the data, the setting and time span analyzed) and the estimation method.

Finding from the study indicated that elasticity estimates are sensitive to different aspect of the model structure. In regard to price elasticity, static models tend to produce lower long run price elasticity estimates relative to dynamic models. No differences in long run price elasticity across different dynamic specifications and likewise across different data types (time series, cross-sectional and cross-sectional-time series). The study however found that different time span affected the variation of the price elasticity estimates, especially pre 1974, 1974 to 1981 and post 1981. For income elasticity, vehicle characteristics and vehicle ownership when included in the model tend to lower both the short run and long run elasticity estimates. Furthermore, time span appear to only affect the variation for the long run estimates, with a tendency of declining over time, while the short run income elasticity appear constant cover time.

3. Model

The demand model that will be finally estimated is based on a standard utility-maximizing framework in which I assume a multistage decision[2] process (see for example Edgerton, 1997) for the consumer. In the first stage the consumer maximizes utility by allocating his/her budget for consumption on aggregates of goods[3]. Further, assume that in the first stage a utility function for a representative agent can be expressed as a function of three goods, gasoline (Q_g), diesel (Q_d) and a non-transport related good (Q_n). Given the utility function and the budget constraint of the agent, the unconditional demand functions for these main aggregates are found by maximizing the utility function subject to the budget constraint as follows;

$$\begin{aligned} &\max_{Q_g, Q_d, Q_n} U(Q_g, Q_d, Q_n) \\ &s.t\ P_g Q_g + P_d Q_d + P_n Q_n \leq Y \end{aligned} \tag{1}$$

The solution to this gives us the unconditional demand function for the respective aggregate goods;

$$Q_g = f(P_g, P_d, P_n, Y) \tag{2}$$
$$Q_d = f(P_g, P_d, P_n, Y) \tag{3}$$
$$Q_n = f(P_g, P_d, P_n, Y) \tag{4}$$

[2] For the multistage decision process to be appropriate, one has to assume that preferences are weakly separable in the sense that the quantity consumed of transport fuel (gasoline and diesel) does not affect the marginal rate of substitution between other goods (for example chicken and beef).
[3] Here I simply assume that the saving-consumption choice has already been made.

In the second stage the consumer maximizes utility by selecting goods consumption within each aggregate, given the predetermined expenditure for each aggregate, but since the focus is on gasoline, I will not outline the second stage choice process but rather focus on equation(2). Moreover, equation (2) is further modified by including an unobserved gasoline trend, which is assumed to capture exogenous technological progress and taste[4]. Since the demand function is homogenous of degree zero in all prices and income, one can divide through by P_n and hence obtain demand as a function of real prices and real income, i.e.

$$Q_{gt}^{d} = f(P_{gt}^{r}, P_{dt}^{r}, \mathrm{Y}_{t}^{r}, \mu_{t}) \tag{5}$$

Where $P_i^r = P_i / P_n$ and $Y^r = Y / P_n$, μ_t denote the trend, Q_g^d is the aggregate consumption of gasoline per capita, P_g is the nominal price of gasoline, P_d is the nominal price of diesel, P_n is a price index for "other goods", here approximated by CPI, and Y is total expenditure on gasoline, diesel and all other goods, here approximated by GDP per capita.

3.1 Econometric model

To estimate the model, I adopt Harvey's (1989) structural time series approach. This approach allows more flexibility in modeling trends (UEDT) than other methods such as the cointegration approach, which does not allow for flexible modeling of the trend components. Since

[4] Taste refers to socio-economic factors such as lifestyles and the effects of values and norms.

the objective of the paper is to assess the impact of non-economic factors such as technology and taste on gasoline elasticities and the time varying aspects of these elasticities, it is therefore important to consider models that are more flexible in capturing these effects.

Assuming a log-linear functional form, a structural framework for estimating gasoline demand trend in a constant parameter model (CPM) from equation (5) can be presented as follows[5];

$$q_{gt}^{d} = \alpha_1 p_{gt} + \alpha_2 p_{dt} + \alpha_3 y_t + \mu_t + \varepsilon_t \tag{6}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + v_t, \quad v_t \sim NID\,(0, \sigma_V^2) \tag{7}$$

$$\beta_t = \beta_{t-1} + \xi_t, \quad \xi_t \sim NID\,(0, \sigma_\xi^2) \tag{8}$$

where μ_t, the gasoline demand trend (UEDT), is estimated via the transition equation (7) and (8). The random error term in equation (6) is represented by ε_t which is referred to as an irregular term[6]. The slope parameter of the trend (β_t) is estimated via equation (8). The variables, v_t and ξ_t are random error terms that are uncorrelated and responsible for high and low movements of the trend. The variance of the random error terms in equations (7) and (8), σ_V^2 and σ_ξ^2 are the hyper-parameters. Given that the hyper-parameters are both zero, the above stochastic trend collapses to a deterministic trend. Equation (6) and (9) are estimated by the maximum likelihood method, while that

[5] A lower case variable denotes that the variable is in natural logarithms and this applies to the rest of the paper.
[6] Irregular term refers to the random error term in a given model that is assumed to be normally distributed with zero mean and constant variance.

of equations (7) and (8) by the Kalman filter (see Harvey (1989) for details).

3.2 Time Varying Parameter Model (TVP)

Equation (6) is basically a constant parameter model and therefore cannot address one of the main questions that the paper seeks to answer. That is, whether gasoline price and income elasticities are varying over time. To address this, a TVP model is adopted and it is expressed as follows;

$$q_{gt}^{d} = \alpha_{1,t} p_{gt} + \alpha_{2,t} p_{dt} + \alpha_{3,t} y_t + \mu_t + \varepsilon_t \tag{9}$$

$$\begin{aligned} &\alpha_{i,t} = \alpha_{i,t-1} + \delta_t, \quad \delta \sim NID(0, \sigma_\delta^2) \\ &i = 1, 2, 3. \end{aligned} \tag{10}$$

Where $\alpha_{i,t}$ denotes the coefficients of the parameters that are allowed to vary with time via an autoregressive process as expressed in equation (10), the subscript i stands for the variables (p_g, p_d, y) in the model and δ is a random error term. This specification therefore allows the estimated coefficients to vary from year to year, and with this one can determine if the variations are significantly different from one year to the next.

The contributions of the independent variables to annual change in gasoline demand is calculated using the fitted version of equation (9) and taking annual difference given by;

$$\Delta \hat{q}_{gt}^{d} = \hat{\alpha}_{1,t} \Delta p_{gt} + \hat{\alpha}_{2,t} \Delta p_{dt} + \hat{\alpha}_{3,t} \Delta y_t + \Delta \hat{\mu}_t \tag{11}$$

where Δ denote a change and therefore $\hat{\alpha}_{1,t}\Delta p_{gt}$, $\hat{\alpha}_{2,t}\Delta p_{dt}$, $\hat{\alpha}_{3,t}\Delta y_t$ and $\Delta\hat{\mu}_t$ are the contributions of price of gasoline, price of diesel, income and UEDT to annual change in fitted gasoline demand ($\Delta\hat{q}^d_{gt}$), respectively.

4. Empirical Analysis

The estimation strategy was to first estimate equation (6) to identify the parameters that are likely to be significant in the TVP-STSM and also to assess if the estimates from the TVP-STSM are significantly different from those from the CPM. Equation (9) would be estimated using all the variables as specified in equation (6), irrespective of whether they are statistically significant or not in the CPM. The TVP-STSM would be used to forecast 11 years head gasoline demand, based on three oil price and tax scenarios.

4.1 Data and data sources

The data series for each of the countries under study comprised annual data that covered the period from 1970 to 2009 and they were expressed in the local currency of each of the countries under study. The Swedish data set was taken from two main sources, the Swedish Petroleum Institute and Statistics Sweden. Data on gasoline consumption, gasoline price and diesel price were taken from the Swedish Petroleum institute, and data on GDP, consumer price index (CPI) and population were taken from the Statistics Sweden. Gasoline consumption was in cubic meters, which is converted to tons of oil equivalent (toe). The nominal prices of gasoline and diesel were both converted to real prices by deflating the nominal series by the CPI in order to obtain the real series. Similarly, the nominal GDP was also deflated by the CPI in order to obtain the real GDP series.

The United Kingdom (UK) data set was also taken from two data sources, that of the Digest of UK Energy Statistics (DUKES) website, (http://www.decc.gov.uk/) and the Office of National Statistics (ONS) website (http://www.statistics.gov.uk/hub/index.html). The gasoline consumption series (in toe), gasoline price series and the price series for diesel were all taken from DUKES website. CPI, population and GDP series on the other hand were taken from the ONS website. As in the case of Sweden, the nominal series for the UK was deflated by the CPI series in order to obtain the real series for GDP and prices. Graphs of the growth rates of aggregate gasoline consumption per capita for both Sweden and the UK are presented in Figure 1. From the Swedish gasoline consumption graph, it revealed a sharp decline in 1974, and 1993, which corresponded to the 1973-1974 oil crisis and the cumulative effects of the 1990 oil crisis and the 1990's recession respectively. The 1979-1980 oil crises seemed to have had a less dramatic effect on the reduction in gasoline consumption. The graph for the UK's gasoline consumption growth rate did indicate declines during the various oil crisis years, but the declines were less dramatic, except for the 1979-1980 oil crisis and 1990's recession, which had a more significant impact in the year 1994.

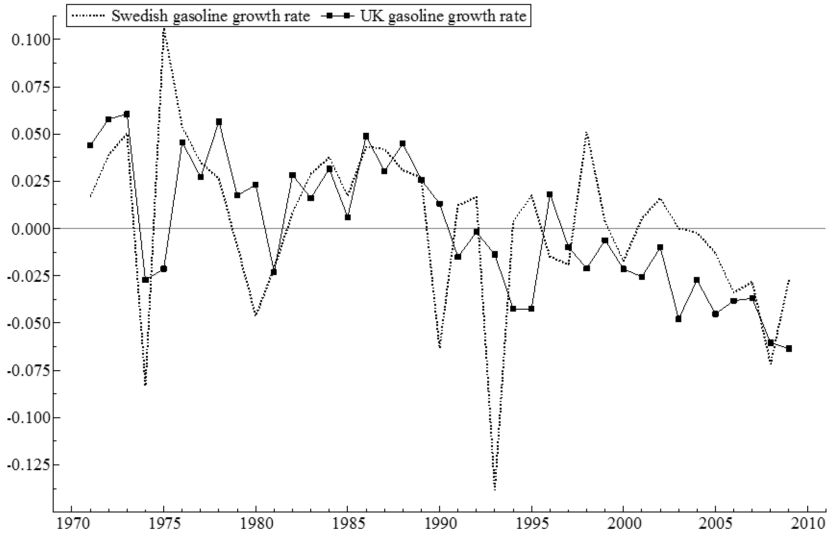


Figure 1: Aggregate gasoline consumption per capita growth rate for Sweden and the UK.

4.2 Results

I first run a regression (CPM) on equation (6) for both countries. The results from equation (6) are reported in Table 2, where a level intervention was included in the model for the UK to capture structural break in the data to ensure normality of the residuals. The interventions according to Harvey and Koopman (1992) ensure normality of the auxiliary residuals.

4.2.1 Constant Parameter model (CPM)

Sweden

Since the variables were all in logarithm, one could interpret the estimated parameters as elasticities. The estimated gasoline price elasticity was -0.45, which had the expected sign and was highly significant. Furthermore, it lay close to the long-run price elasticities of -0.37 as documented in the review by Sterner et al (1992) for Sweden. The estimated coefficient for the diesel price was significantly positive but small in magnitude. One could therefore reasonably conclude that the price of diesel had a small positive effect on gasoline consumption. This was quite reasonable, at least in the long run because of the possible effects from the new European Union directives on new car registration that favor diesel cars, couple with the efficiency level of diesel engines relative to gasoline engines. The results implies that in the long run, diesel serves as a possible fuel substitute to gasoline via switching from gasoline cars to diesel cars, but this is likely to be the case, for instance when the price of gasoline was so high relative to diesel price that the savings from switching from gasoline engines to diesel engines outweighed the cost associated with the switching. Since the possibility of meeting the condition that could result in the significant positive diesel price effect on gasoline demand was small, the finding from the model was therefore reasonable and satisfactory.

The income elasticity estimated from the model had the expected sign but the magnitude (0.54) was quite low compared to the 0.99 found in Sterner et al. (1992) and the approximately unity found in other surveys presented in Table 1 above. There are two possible reasons for this low-income elasticity found in this study relative that found in Sterner et al (1992). The first is that, sterner et al (1992). used data from 1960 to 1985, a period where income levels are low relative to the late 1990s and 2000s. The possible effect of this is that people tend to be more responsive to income changes due to the relative low-income levels.

Table 2: Estimated results a Constant Parameter Model (CPM) for Sweden and the UK

Sweden		UK	
Pg	-0.446***	Pg	-0.108***
	(0.050)		(0.036)
Pd	0.068***	Pd	0.049
	(0.024)		(0.042)
Y	0.541***	Y	0.577***
	(0.126)		(0.120)
Intervention		Intervention	
NA		Level break	-0.061***
		(1994)	(0.016)
Trend		Trend	
Growth rate at end of 2009	-2.92 p.a	Growth rate at end of 2009	-4.98 p.a
Diagnostics (residuals)		Diagnostics (residuals)	
Std. error	0.022	Std. error	0.017
Normality	1.336	Normality	1.522
H (11)	0.834	H (11)	0.262
r (1)	0.034	r (1)	-0.462
r (6)	-0.118	r (6)	0.152
DW	1.731	DW	2.692
Predictive Test		Predictive Test	
Failure test $\chi^2(9)$	9.159 [0.422]	Failure test $\chi^2(9)$	3.296 [0.951]
Cum sum t(9)	-1.189 [1.735]	Cum sum t(9)	-0.738 [1.521]
Goodness of Fit		Goodness of Fit	
P.e.v	0.0005	P.e.v	0.0003
R^2	0.951	R^2	0.987
AIC (a)	-7.280	AIC (a)	-7.724
AIC (b)	-4.902	AIC (b)	-4.072
Hyper Parameters (std)		Hyper Parameters (std)	
Level	0.018	Level	0.001
Slope	0.006	Slope	0.006
Irregular	0.002	Irregular	0.009
Auxiliary Residual Normality test		Auxiliary Residual Normality test	
Level	2.248 [0.324]	Level	3.084 [0.213]
Slope	0.354 [0.837]	Slope	1.208 [0.546]
Irregular	0.858 [0.651]	Irregular	0.093 [0.954]

Notes:

- ***, and** indicated 1%, and 5% levels of significance respectively with standard errors in brackets and p-values in square brackets.

- R^2 is the coefficient of determination, AIC (a) is the Akaike information criteria for the model with stochastic trend and AIC (b) is the Akaike information criteria for the model with a deterministic (linear) trend.
- P.e.v is the predictive error variance. Normality test is conducted using corrected Bowman-Shenton test, which is χ^2 distributed. DW is the Durbin-Watson statistic for test of autocorrelation; r (1) and r (6) are serial correlation coefficients at lag 1 and 6 respectively.
- H (11) is test statistic for heteroskedasticity, which is approximately distributed as F (11, 11).
- Predictive failure test is the post-sample predictive failure test for 2000-2009, distributed as χ^2 (9), while cusum is a mean stability statistic.

The second reason is that Sterner et al. (1992) used a dynamic model without controlling for changes in the structure of the economy and taste. Using the pre 1985 period and dynamic model as done in sterner et al. I found similar income elasticity, but using the same dynamic model for the whole sample (1970-2009) resulted in income elasticity that is close to that found with the STSM (this result is not reported but available on request). This is consistent with the findings from Espey(1998), that indicated that time span had a significant effect on the long run income estimates and that the tendency is that, the income estimates decreases over time. The conclusion therefore is that the difference in income estimates is most likely due to the different time span and rather than the econometric model.

Diagnostic tests for the model were also very satisfactory. I also performed a prediction failure test by saving the last nine observations of the data in order to assess the stability of the estimated model and the results confirmed that the model was stable at the 5% significance level. This result was confirmed by the cusum test, which could not also reject the stability of the model. Furthermore, I also tested whether the deterministic trend or the more flexible stochastic trend was the most appropriate in fitting the data, by using the AIC to discriminate between the two. The result supported the stochastic trend structure.

United Kingdom

Applying the CPM for the UK data produced a gasoline price elasticity of -0.11 and it was significant at the 1% level. This value is out of range of most of the price elasticities reported in the various surveys presented in Table 1. However, the value was approximately the same as that found in Hunt and Ninomiya (2003), Dimitropoulos et al (2005), and Broadstock and Hunt (2010). Unlike the various surveys, these three studies all applied STSM in analyzing the data with different time coverage. The estimated coefficient for diesel price for the UK was positive but not significant, unlike the case of Sweden, where it was significant.

The estimated income elasticity from the model was approximately 0.57. This value was highly significant and had the expected sign. The magnitude however was relatively low compared to the 0.80 and 0.81 found in Hunt and Ninomiya (2003), and Dimitropoulos et al (2005) respectively. However my value was the same as the 0.57 found in Broadstock and Hunt (2010). Unlike my model, Broadstock and Hunt (2010) included car fuel efficiency as an additional control variable in their model. A possible explanation of the slight differences in the estimates is due to the different time span of the data. Furthermore, the estimated model for UK passed all the model diagnostics and also the stability test. The fit of the model was highly satisfactory, and the AIC supported the stochastic trend relative to the deterministic linear trend for the data generating process.

4.2.2 Time Varying Parameter Model (TVP-STSM)

Sweden

The diagnostics for the TVP-STSM regression are presented in Table 3, and the estimated elasticities and the trend are presented in Figure 2 and 3 respectively. The estimated price elasticity for 2009 was -0.45 and within the range of the long-run price elasticity for Sweden that currently exists in the literature (see, Dahl, 2012 and Sterner et al 1992). An interesting finding from the TVP-STSM was that, the price elasticity did indeed vary overtime, but that the variations were infinitesimal and therefore the constant parameter model was a good approximation and could model the data relatively well.

The estimated results from the TVP-STSM also indicated that the income elasticity varied over time, but similar to the price elasticity, the variations were very small. The conclusion one could draw regarding these results was that the price and income elasticities were stable overtime and relatively small in magnitude and that non-economic factors such as taste had a significant impact on gasoline demand, for instance in the case of Sweden, the estimated level of the trend at the end of 2009 indicated that trend accounted for 0.2% reduction in gasoline demand. From this it followed that effects from policy, such as gasoline taxes, would be limited. Furthermore, the stability of the elasticities suggested that policy changes had not affected basic consumer behavior, expressed here in terms of price and income elasticity significantly over time.

Table 3: Diagnostic test for the TVP model for Sweden

Sweden		UK	
Level of trend	-0.201	Level of trend	-1.435
Slope of trend	-0.029	Slope of trend	-0.048
Intervention		Intervention	
NA		Level break	-0.062***
		(1994)	(0.017)
Diagnostics (residuals)		Diagnostics (residuals)	
Std. error	0.010	Std. error	0.017
Normality	2.282	Normality	2.059
H (11)	0.643	H (11)	0.200
r (1)	0.058	r (1)	-0.291
r (6)	-0.169	r (6)	-0.002
DW	1.704	DW	2.377
Predictive Test		Predictive Test	
Failure test (9)	8.715	Failure test (9)	2.833
	[0.463]		[0.970]
Cum sum (9)	-1.194	Cum sum (9)	-0.828
	[1.737]		[1.571]
Goodness of Fit		Goodness of Fit	
P.e.v	0.0001	P.e.v	0.0003
R^2	0.988	R^2	0.987
AIC (a)	-8.743	AIC (a)	-7.736
AIC (b)	-4.102	AIC (b)	-4.008
Hyper Parameters (std)		Hyper Parameters (std)	
Level	0.0018	Level	0.0007
Slope	0.0064	Slope	0.0066
Y	0.0008	Y	0.0001
Pg	0.0021	Pg	0.0002
Pd	0.0002	Pd	0.0002
Auxiliary Residual Normality test		Auxiliary Residual Normality test	
Level	2.323 [0.312]	Level	2.588 [0.274]
Slope	0.435 [0.804]	Slope	1.320 [0.516]
Irregular	0.867 [0.648]	Irregular	0.139 [0.932]

The estimated gasoline trend for Sweden, displayed in Figure 3, increased (in log term) until 1990 and thereafter decreased. This meant that growth rate in gasoline consumption per capita increased (though at a decreasing rate) up till 1990 and decreased thereafter, implying, among other things, less efficient use of gasoline for the period up till 1990 and more efficient use of gasoline after 1990 if taste is assume to be constant or moving in the direction that favor diesel consumption. The estimated growth rate of the trend in 2009 was approximately -2.9%.

United Kingdom

The UK estimates also revealed that the price elasticity varied over time, but as in the Swedish case the variation was very small. The estimated price elasticity in year 2009 was -0.11. Figure 2 depicts the estimated price and income elasticities for the UK, while Figure 4 depicts the estimated trend for the UK from the TVP-STSM. The income elasticity also showed very small variations from year to year but the variations are insignificant. The estimated level of gasoline trend for UK was -1.44, indicating a higher reduction effect of trend on gasoline demand in the UK relative to Sweden. Comparing the fit of the two models, it was clear that TVP-STSM with stochastic trend did fit the UK data relatively better than that with a deterministic trend, based on the AIC value. Stability of the model was also checked by performing a prediction failure test, were the nine last observations were used for the test. The results from the test indicated that the model was stable.

The gasoline trend for the UK showed a similar pattern as that of Sweden. It increased continuously from 1970 to 1990 in log scale and thereafter took a downturn, but much smoother and with a higher rate of reduction (-4.9%) than Sweden, implying, among other things, a higher level of gasoline efficiency or a more response to policies favoring diesel engines or both in the UK than in Sweden.

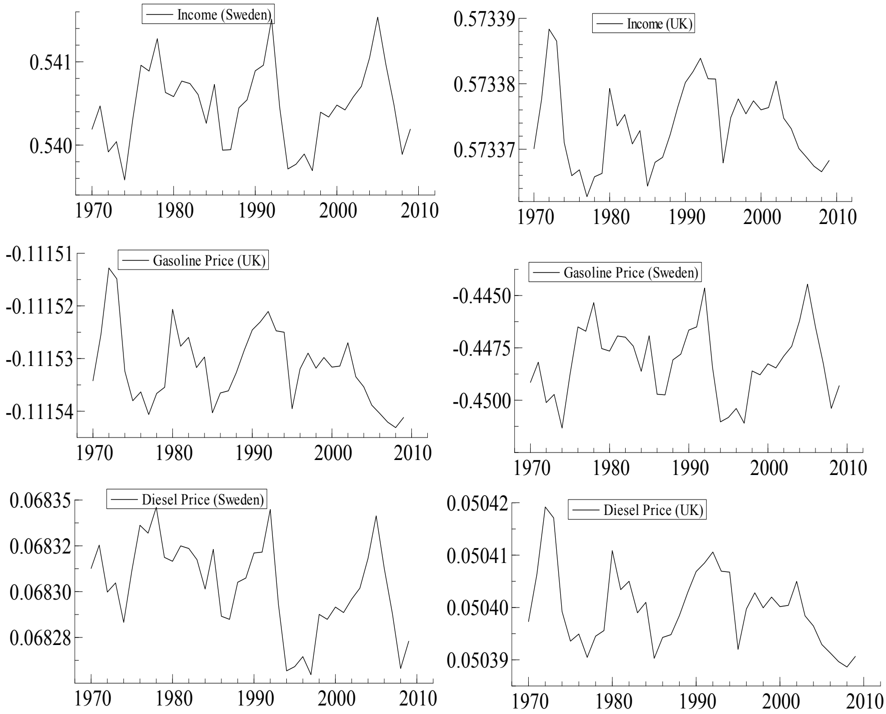


Figure 2: Estimated TVP income and price elasticities for Sweden and UK

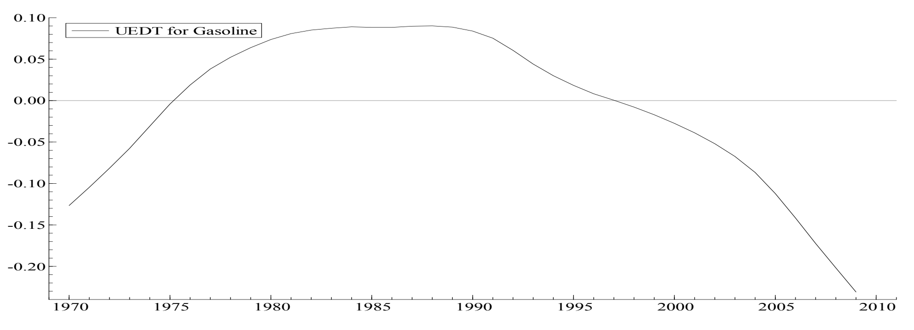


Figure 3: Estimated gasoline trend (UEDT) for Sweden in log scale.

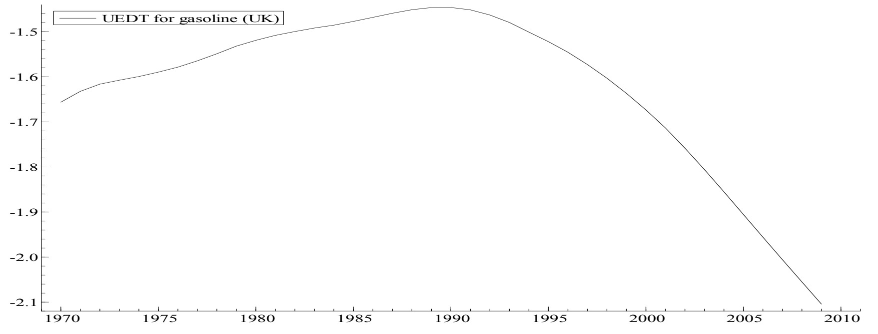


Figure 4: Estimated gasoline trend (UEDT) for UK in log scale.

Contributions of economic and non-economic factors to annual gasoline demand changes

The contributions of gasoline price, diesel price, income and UEDT to annual change in fitted gasoline demand for Sweden and UK are presented in Figure 5 and 6, respectively. As indicated in the empirical model section, the calculated contributions are done using equation (11). The results indicated that non-economic factors proxy by UEDT, significantly contributed to annual changes in fitted gasoline demand. Furthermore, the result also indicated that, in most of the years for both countries, UEDT is one of the dominant contributors to the annual changes in the fitted gasoline demand. This result is consistent with the finding in Broadstock and Hunt (2010). It is therefore important to incorporate non-economic factors as much as possible in estimating gasoline demand models as they contribute a significant proportion to gasoline demand changes for both Sweden and the UK.

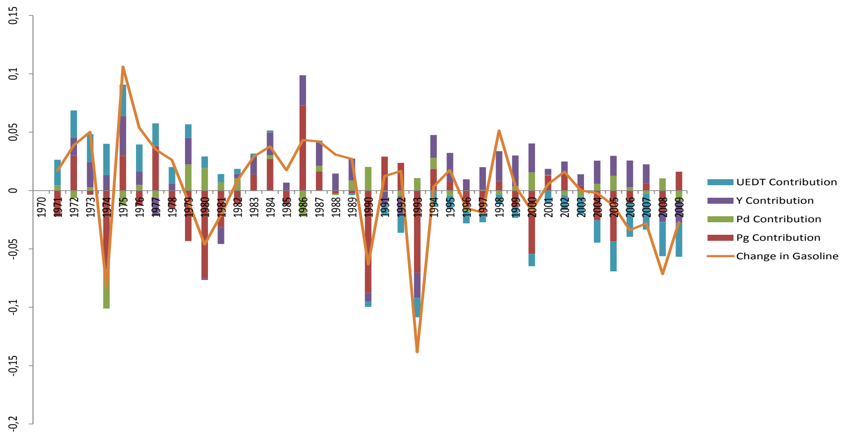


Figure 5: Contributions of economic and non-economic factors to annual change in Sweden gasoline demand ($\Delta\hat{q}_{gt}^{d}$).

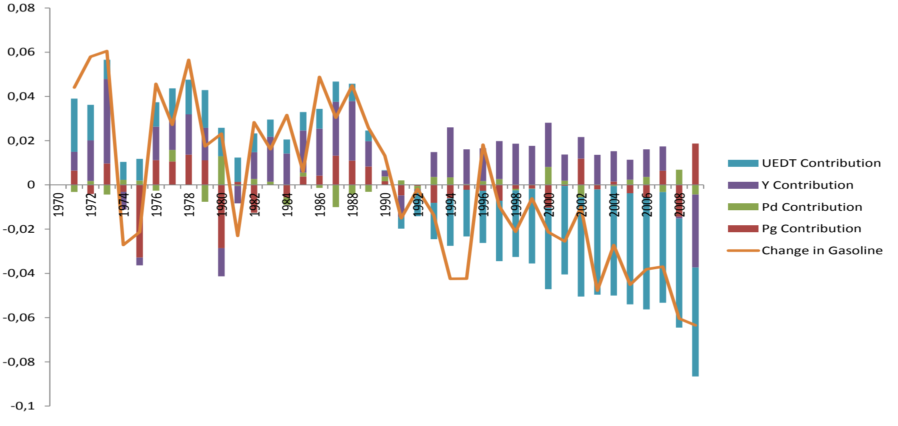


Figure 6: Contributions of economic and non-economic factors to annual change in UK gasoline demand ($\Delta\hat{q}_{gt}^{d}$).

Robustness Check

To check how robust the results were with regard to the assumption made and the modeling specification for the trend in the model, I also modeled a deterministic trend as is usually done in the literature but specified it more flexibly with regard to the functional relationship between the trend and gasoline demand. Using a semi-parametric approach where all the other variables entered the model linearly while the trend was estimated using spline regression that allowed for possible non-linearity. The estimated parameters for price and income were not significantly different from the CPM both for Sweden and the UK, and the estimates for the trend were remarkably similar in pattern as the TVPM case for both countries, detailed results are reported in Table A and Figure A in the Appendix.

Furthermore, I investigated the forecasting accuracy of the TVP-STSM relative to other popular econometric time series models (dynamic model, ARDL model, and autoregressive integrated moving average (ARIMA) model), based on in-sample forecasting (I used the last nine observation of the sample, 2001 to 2009). Based on both the root mean square prediction error (RMSPE) and mean absolute prediction error (MAPE), the TVP-STSM was ranked first for Sweden for each of the periods. In the case of UK, the TVP-STSM was ranked second for the first four periods (ARDL was ranked first for these periods) and first for the entire nine periods, results are reported in Appendix B. The conclusion from this was that, the TVP-STSM had relatively impressive forecasting ability compare to some of the popular econometric models, hence out-of sample forecasting from this model is likely to be reasonable and reliable.

5. Forecasting results

Based on three tax scenarios and certain assumptions regarding key variables, 11-years ahead forecasting was done using TVP-STSM. The scenarios are; a reference scenario, a medium tax rate, and a high tax rate scenario. In the reference case, the current gasoline tax rate is applied along with the growth rate of the key macroeconomic variables such as growth rate of the GDP, real price of gasoline, real price of diesel, growth rate of the trend and population growth rate. The medium tax scenario is a scenario in which I consider a 5% increase in the gasoline tax rate above the reference tax rate with the same growth rates for the other variables in the model as in the reference scenario. In the high tax scenario, I consider a 10% increase in gasoline tax above the reference tax rate. Here I again assume the same rates as in the reference case for the key variables. In the case of real gasoline and diesel price assumptions, I considered three projected assumptions for real world oil price; a reference, low and high oil price projections.

Sweden

In the reference case scenario, the current tax rate in addition to the projected real GDP growth rates presented in Table 4 were assume for the entire forecasting horizon for Sweden. The projected real GDP growth rates are retrieved from the Eurostats website[7]. Regarding real gasoline and diesel prices, I assume three scenarios based on forecast

[7] http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tec00115.

world oil prices from the Energy Information Administration (EIA). The growth rate assumption of real oil prices is also presented in Table 4. For the same forecasting period, I assume the growth rate of population to be approximately the same rate as the 2009 rate, while the gasoline consumption trend is assumed to continue to decline each year at the same rate as the 2009 rate of -2.9% for the forecasting period. In the medium tax scenario, I consider the same growth rates in real GDP, the growth rate of gasoline trend, population growth rate, real gasoline and diesel prices as that of the reference scenario, but with a different tax rate that is higher than the reference tax rate. The medium tax rate is 5% above the reference tax rate. The high tax scenario on the other hand is 10% higher than the reference case but with the same growth rates for real GDP, trend, population and real prices. The resulting forecasts are presented in Table 5.

In the reference oil price case, the 2020 forecast value represents a 26.9%, 28.1% and 29.2% reduction of gasoline consumption per capita over the 2009 value for the reference tax, 5% tax and 10% tax scenarios respectively. The reductions based on low oil price assumption are approximately 10 points lower for each of the tax scenarios compared to the reference oil price case, while the reductions are approximately 3 points higher based on high oil price assumption relative to the reference case. I further broke down the reduction in gasoline consumption to assess the impact of the 5% and 10% tax increases relative to the reference tax. The result indicates that the impact of taxes are higher with the low oil price case, while the lowest impact is from the high oil price case. A possible reason for this is that prices are already high in the high oil price case and therefore further increases due to taxes only result in diminishing marginal effects.

Table 4: Real oil price and real GDP growth rate assumptions (% per annum).

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Real Price											
Low	rv	rv	-5.4	-1.5	-0.94	-0.91	-0.87	-0.84	-0.79	-0.76	-0.72
Reference	rv	rv	4.2	4.1	3.7	3.4	3.1	2.9	2.7	2.4	2.2
High	rv	rv	13.8	7.3	4.1	3.8	3.5	3.2	2.9	2.7	2.5
Real GDP											
UK	1.8	0.9	-0.3	0.9	2	2.2	2.2	2.2	2.2	2.2	2.2
Sweden	6.6	3.7	1.1	1.9	2.5	2.8	2.8	2.8	2.8	2.8	2.8

Note:

- rv represents the observed real growth rates and therefore the same for each of the three scenarios.

United Kingdom

Similarly, the UK's real GDP growth rates were taken from Eurostats' projections, while the population growth rate is assumed to be at the same rate as that of the 2009 rate for each year. The gasoline trend on the other hand is assumed to decline at a rate of -4.9% (2009 trend growth rate) per annum for the forecasting period. For real gasoline and diesel prices, I made the same three scenario assumptions regarding world real oil prices in a similar way as in the case of Sweden. The same tax rate scenarios as for Sweden also apply to the case of the UK. The results are presented in Table 6.

The forecast results indicate a similar pattern of reductions of gasoline consumption per capita. The reductions based on the low oil price case are approximately 4 points lower than the reference oil price case for all tax scenarios, while the reductions from the high oil price case are approximately 1 point higher than the reference oil price reduction.

Table 5: Forecast gasoline consumption per capita (toe) for Sweden

	Ref. oil price scenario			**Low oil price scenario**			**High oil price scenario**		
year	*Ref.tax*	*5% tax*	*10%tax*	*Ref.tax*	*5% tax*	*10%tax*	*Ref.tax*	*5%tax*	*10%tax*
2010	0.5077	0.4997	0.4918	0.5077	0.4997	0.4918	0.5077	0.4997	0.4918
2011	0.4895	0.4818	0.4742	0.4895	0.4818	0.4742	0.4895	0.4818	0.4742
2012	0.4711	0.4637	0.4564	0.4857	0.4781	0.4706	0.4581	0.4509	0.4438
2013	0.4553	0.4481	0.4411	0.4778	0.4703	0.4629	0.4385	0.4316	0.4248
2014	0.4419	0.4349	0.4281	0.4705	0.4631	0.4558	0.4251	0.4184	0.4118
2015	0.4299	0.4231	0.4164	0.4639	0.4566	0.4494	0.4130	0.4065	0.4001
2016	0.4186	0.4120	0.4055	0.4574	0.4502	0.4431	0.4016	0.3953	0.3891
2017	0.4078	0.4014	0.3951	0.4509	0.4438	0.4368	0.3910	0.3848	0.3788
2018	0.3976	0.3913	0.3852	0.4444	0.4374	0.4305	0.3809	0.3749	0.3690
2019	0.3880	0.3819	0.3759	0.4380	0.4311	0.4243	0.3714	0.3655	0.3598
2020	0.3788	0.3729	0.3670	0.4316	0.4248	0.4182	0.3623	0.3566	0.3510
% change (2009-2010)									
	26.9	28.1	29.2	16.7	18.0	19.3	30.1	31.2	32.3
% change due to taxes relative to the ref.tax									
		4.3	8.5		7.8	15.5		3.7	7.3

Table 6: Forecast gasoline consumption per capita (toe) for the UK

	Ref. oil price scenario			**Low oil price scenario**			**High oil price scenario**		
year	*Ref.tax*	*5% tax*	*10%tax*	*Ref.tax*	*5%tax*	*10%tax*	*Ref.tax*	*5%tax*	*10%tax*
2010	0.2549	0.2549	0.2549	0.2549	0.2549	0.2549	0.2549	0.2549	0.2549
2011	0.2426	0.2415	0.2404	0.2464	0.2452	0.2441	0.2426	0.2415	0.2404
2012	0.2294	0.2284	0.2273	0.2338	0.2327	0.2317	0.2294	0.2284	0.2273
2013	0.2176	0.2166	0.2156	0.2237	0.2227	0.2217	0.2159	0.2149	0.2139
2014	0.2089	0.2079	0.2069	0.2158	0.2148	0.2138	0.2066	0.2057	0.2047
2015	0.2012	0.2003	0.1994	0.2088	0.2078	0.2068	0.1989	0.1980	0.1971
2016	0.1935	0.1926	0.1917	0.2015	0.2006	0.1997	0.1912	0.1904	0.1895
2017	0.1859	0.1851	0.1843	0.1944	0.1935	0.1926	0.1837	0.1829	0.1821
2018	0.1788	0.1780	0.1771	0.1875	0.1867	0.1858	0.1766	0.1758	0.1750
2019	0.1719	0.1711	0.1703	0.1809	0.1801	0.1792	0.1698	0.1690	0.1682
2020	0.1653	0.1646	0.1638	0.1745	0.1737	0.1729	0.1633	0.1625	0.1618
% change(2009-2020)									
	37.6	38	38.2	34.0	34.3	34.6	38.4	38.7	39.0
% Change due to taxes relative to the ref.tax									
		0.8	1.5		0.9	1.8		0.7	1.5

The results further indicate that the impact of taxes is higher with the low oil price case compared to the reference and high oil price cases. It is important to note that, although the pattern of the effects of taxes is similar for both Sweden and the UK, the magnitude of the impact is very different, due to the different price elasticities in the two countries. For instance, the percentage change due to a 10% increases in taxes for the reference world oil price case in Sweden is 8.5, while it is only 1.5 for the UK.

6. Conclusion

The paper examined the impact of non-economic factors (UEDT) on gasoline demand for Sweden and the UK. More importantly I studied whether gasoline demand elasticities were constant or significantly vary over time for each of the countries under study, using Harvey's (1989) STSM. The results indicated that, for both countries, in addition to price and income, non-economic factors significantly influenced gasoline demand non-linearly. Both price and income elasticities did vary over time for each of the countries, but the variations were extremely small.

The estimated trend for both countries showed a similar pattern, reflecting similar technological progress across the two countries, which was consistent with the concept of diffusion of technology and its adaptation across the developed world. The policy implication of the results from the tax simulations was that, given the assumptions made on the key variables, although the total reduction was greater for the higher tax scenario for each of the countries across the three oil price assumptions, the impact of taxes on the other hand was rather higher for the low oil price case relative to both the reference and the high tax cases. The implication of this was that the marginal impact of taxes was higher with low price levels relative to high price levels. Furthermore, the study found that the percentage change due to taxes was higher in Sweden than in the UK, which, among other things, implied that taxes were more likely to be an effective policy tool in Sweden than in the UK. Another interesting finding was that, irrespective of the fact that the impact of taxes was much higher in Sweden relative to the UK; the total reduction across all the oil price assumptions and tax scenarios was higher in UK than in Sweden. This was partly due to the high income elasticity combined with a higher trend reduction (a combination of technology and taste) in the UK relative to that of Sweden.

The results also indicated a higher response rate to income in the UK than in Sweden, implying, among other things, the differences in private transport culture for the two countries. As an average person became richer in the UK, he or she was much more likely to drive more than an average rich Swede. But whether these differences were mainly due to economic factors or culture was not completely clear. It was therefore important to investigate this further, using micro-base studies on each of the countries, controlling for demographics and locational differences as well as past policies on energy in general or gasoline in particular.

Furthermore, the study was limited to two countries but it will be interesting to extend similar analysis to all OECD countries to assess if the finding from the two countries applies to other OECD countries and to assess which of the countries are likely to achieve greater impact from non-economic factors on changes in gasoline demand. Also in the tax simulations, I did not consider other forms of taxes such as income tax, fuel subsidies and various forms of tax allowances and therefore incorporating such factors in the simulations will be a valuable contribution to the literature, which I hope to consider in my next paper.

References

Archibald, R. and Gillingham, R., 1980.An analysis of short-run consumer demand for gasoline using household survey data. *The review of Economics and Statistics* 62(4), 622-628.

Blum, U., Foos, G. and Guadry, M., 1988. Aggregate time series gasoline demand models:review of the literature and new evidence for West Germany, *Transportation search A*, 22A, 75-88.

Bohi, D., 1981. *Analysing demand behaviour*: *a study of energy elasticities*", Johns Hopkins University Press: Baltimore, U.S.

Brännlund, R and Nordström, J., 2004. Carbon tax simulations using a household demand model. *European Economic Review*, 46(1), 211-233.

Broadstock,D.C. and Hunt,L.C., 2010 .Quantifying the impact of exogenous non-economic factors on UK transport oil demand. *Eenergy Policy*, 38(3), 1559-1565

Chow, G.C., 1984.Random and changing coefficient models, in: Z. Griliches and M.D. Intriligator, Editors, *Handbook of Econometrics*, North Holland, Amsterdam, Chapter 21

Dahl, C., 1986. Gasoline Demand Survey. *The Energy Journal*, 7, 67-82

Dahl, C. and Sterner 1991a.Analysing gasoline demand elasticities: a survey. *Energy Economics* 13, 203-210.

Dahl,C. and Sterner, T., 1991b. A survey of econometric gasoline demand elasticities. *International Journal of energy Systems*, 11, 53-76

Dahl,C., 2012. Measuring global gasoline and diesel price and income elasticities. *Energy Policy*, 41, 2–13

Dimitropoulos, J., Hunt, L.C. and Judge,G.,2005. Estimating underlying energy demand trends using UK annual data, *Applied Economics Letters*, 12(4), 239-244

Drollas, L.,1984. The demand for gasoline: further evidence. *Energy Economics*, 6, 71-82.

Espey,M, 1998. Gasoline demand revisited: an international meta-analysis of elasticities. *Energy Economics*, 20, 273-295

Edgerton, D.L., 1997. Weak Separability and the Estimation of Elasticities in Multistage Demand Systems. *American Journal of Agricultural Economics*, 79(1), 62-79.

Foos, G., 1986. Die determinanten der verkehrnachfrage', Karlsruher Beiträge zur Wirtschaftspolik und Wirschaftsforschung, 12, Loper Verlag: Karlsruhe.

Gately,D. and Huntington, H.G., 2002. The asymmetric effects of changes in price and income on energy and oil demand. *The Energy Journal*, 23, 19-55

Goodwin, P.,1992. A review of new demand elasticities with special reference to short and long-run effects of price changes. *Journal of Transport economics and Policy*, 2, 155-169.

Goodwin et al, 2004. Elasticities of road traffic and fuel consumption with respect to price and income: a review. *Transport Reviews*, 24 (3), 275–292

Graham,D.J. and Glaister, S., 2002. The demand for automobile fuel: a survey of elasticities. *Journal of Transport Economics and policy*, 36, 1-26.

Harvey, A.C., 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*.Cambridge University Press: Cambridge.

Harvey, A.C. and Koopman, S.J., 1992. Diagnostic checking of non-economic-components in time series models. *Journal of Business and Economics Statistics*, 10, 377-389

Hunt,L.C. and Ninomiya,Y., 2003.Unraveling Trends and Seasonality: a Structural Time Series Analysis of Transport Oil Demand in the UK and Japan. *The Energy Journal* 24(3), 63-96.

IEA, 2010.Oil and Gas Emergency Policy-UK, retrieved at: http://www.iea.org/papers/security/uk_2010.pdf

IEA, 2008.Energy Policies of IEA Countries, retrieved at: http://www.iea.org/textbase/nppdf/free/2008/Sweden2008.pdf

Park,J.Y.and Hahn,S.B.,1999.Cointegration regressions with time varying coefficients. *Econometric Theory*, 15, 664-703

Park, S.Y. and Zhao,G., 2010. An estimation of US gasoline demand: a smooth time varying Cointegration approach. *Energy Economics*; 32, 110-120

Schmalensee,R. and Stoker,T.M., 1999. Household gasoline demand in the United States. *Econometrica*, 67,645-662

Sterner,T. and Dahl,C., 1992. Modelling transport fuel demand, in: T. Sterner (Ed.). *International Energy Economics*, 65-79

Yatchew,A. and No,J.A., 2001.Household gasoline demand in Canada. *Econometrica* 69(6), 1697-1709

Wadud, Z.,Graham,D.J. and Noland,R.B, 2010. Modelling gasoline demand with heterogeneity in household responses. *Energy Journal*, 31(1), 47-73.

Appendix

Table A: Semi-parametric model results

	Sweden		UK	
Parameters	Coefficient	*P*-value	Coefficient	*P*-value
Pg	-0.414	0.000	-0.120	0.012
Pd	0.052	0.068	0.034	0.538
y	0.553	0.000	0.547	0.001
R^2	0.972		0.992	
GCV score	0.0003		0.0002	

UK Sweden

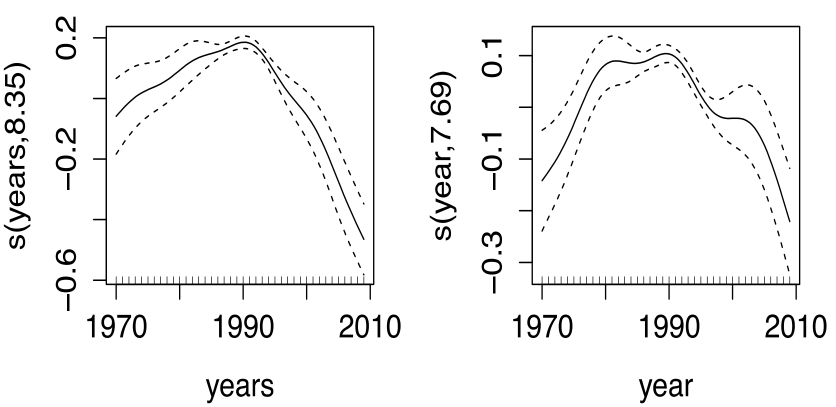


Figure A: Plot of Trend for UK and Sweden, respectively from Semi-parametric model

Table B: In-sample prediction accuracy for different models for Sweden and the UK

Sweden								
	TVP-STSM		DYNAMIC		ARDL		ARIMA	
Year	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE
1	0.008[1]	0.008[1]	0.022[2]	0.021[2]	0.023[3]	0.023[3]	0.032[4]	0.032[4]
2	0.005[1]	0.006[1]	0.021[4]	0.021[3]	0.020[3]	0.020[2]	0.019[2]	0.023[4]
3	0.003[1]	0.001[1]	0.028[4]	0.029[4]	0.021[3]	0.021[3]	0.017[2]	0.020[2]
4	0.004[1]	0.005[1]	0.031[4]	0.025[4]	0.019[3]	0.019[3]	0.014[2]	0.018[2]
9	0.017[1]	0.023[1]	0.050[4]	0.071[4]	0.043[3]	0.052[3]	0.026[2]	0.033[2]

UK								
	TVP-STSM		DYNAMIC		ARDL		ARIMA	
Year	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE	MAPE	RMSPE
1	0.011[2]	0.011[2]	0.016[3]	0.016[3]	0.008[1]	0.008[1]	0.085[4]	0.085[4]
2	0.010[2]	0.010[2]	0.012[3]	0.013[3]	0.009[1]	0.009[1]	0.055[4]	0.063[4]
3	0.014[2]	0.016[2]	0.022[3]	0.027[3]	0.013[1]	0.013[1]	0.056[4]	0.061[2]
4	0.013[2]	0.014[2]	0.024[3]	0.027[3]	0.010[1]	0.012[1]	0.041[4]	0.056[2]
9	0.009[1]	0.011[1]	0.042[3]	0.047[3]	0.011[2]	0.014[2]	0.052[4]	0.054[2]

Note:

- Numbers in the square brackets are the ranking of the model relative to the other models based on the prediction accuracy value.
- Dynamic model denotes a model with a lag dependent variable as one of the regressors.
- A lag length of 1 was used for the ARDL model. The choice of lag 1 is due to the sample size.
- ARIMA (1,0,1) was used for the ARIMA model.

II

Energy Economics 36 (2013) 19–27

Contents lists available at SciVerse ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneco

Functional form and aggregate energy demand elasticities: A nonparametric panel approach for 17 OECD countries

Amin Karimu *, Runar Brännlund

Department of Economics, Umea University, 90187, Umea, Sweden
Centre For Environmental and Resource Economics (CERE), Umea, Sweden

ARTICLE INFO

Article history:
Received 2 June 2012
Received in revised form 8 October 2012
Accepted 24 November 2012
Available online 16 December 2012

JEL classification:
Q4
C3
C4

Keywords:
Cointegration
Log-linear
Panel data and specification test

ABSTRACT

This paper studies whether the commonly used linear parametric model for estimating aggregate energy demand is the correct functional specification for the data generating process. Parametric and nonparametric econometric approaches to analyzing aggregate energy demand data for 17 OECD countries are used. The results from the nonparametric correct model specification test for the parametric model rejects the linear, log-linear and translog specifications. The nonparametric results indicate that the effect of the income variable is nonlinear, while that of the price variable is linear but not constant. The nonparametric estimates for the price variable is relatively low, approximately −0.2.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

There has been an increased awareness and interest on the impact of human activities on the world's climate in recent years, especially through emission of greenhouse gases. A lot of focus in the discussions of curbing climate change has been on issues related to energy use and clean production. For example, there has been an intense debate among energy policy analysts on how policies can promote efficient energy use, less dependence on fossil fuels, and how these have contributed to reduce emissions of greenhouse gases. The central point in this discussion concerns the behavioral response among households and firms to various policy reforms. That is, how do households and firms change their energy consumption as a result of changes in, for instance energy taxes and/or incomes? In order to answer such questions, which are a prerequisite for designing an efficient policy, knowledge about behavioral response is needed. In other words, we need knowledge about price and income elasticities.

Most of the elasticities that are used come from econometric models of energy demand. A look at the literature on energy demand reveals that there exist numerous econometric models, of different types, starting with simple static models to more general dynamic ones. It is interesting to note that most of these models are linear or log-linear models. However, whether the assumed functional form is appropriate or not for the underlying energy demand data generating process (DGP) is usually not tested for. Hence the conclusions drawn must be viewed as conditional on the assumption that energy demand is linear. Aggregate Energy demand studies can be found in the following surveys; Hartman (1979), Bohi (1981), Bohi and Zimmerman (1984), Dahl (2005) as well as Atkinson and Manning (1995).

The consequences of a mis-specified model on the parameter estimates cannot be over emphasized, as it leads to bias and inconsistent estimates. To the best of our knowledge, only two studies exist that study the choice of functional form for energy demand models, Zarnikau (2003) and Xiao et al. (2007). Zarnikau (2003) focuses on electricity demand at the household level for the USA. He studied three functional forms for household electricity demand (linear, log-linear and trans-log) and uses a nonparametric specification test developed by Härdle and Manmen (1993) and Zheng (1996) to test the parametric models. The test results from his study reject each of the three functional forms as the correct specification for the household electricity demand (regressors used are price of electricity, price of natural gas, income and heating degree days). Xiao et al. (2007) on the other hand adopted a Bayesian model selection criterion (that of the Deviance information criterion (DIC)) proposed by Spiegehalter et al. (2002) and applied it to the same data set as in Zarnikau (2003). The result from their study indicates that the

* Corresponding author at: Department of Economics, Umea University, 90187, Umea, Sweden. Tel.: +46 90 7867794.
E-mail address: amin.karimu@econ.umu.se (A. Karimu).

0140-9883/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.eneco.2012.11.026

AIDS[1] and trans-log models are superior to the log-linear model, which in turn is better than the linear model. However, none of these studies makes extensions to panels nor uses aggregate data on energy demand using fully nonparametric modeling that allows for all possible nonlinearities and interactions among the regressors, while controlling for boundary bias at the same time. There is thus a scope for further investigations into the correct functional forms in estimating energy demand with aggregate data. Further, Zarnikau (2003) and Xiao et al. (2007) did not consider cointegration relationships between the dependent variable and the regressors and hence implicitly assume a long run relationship to exist without testing for it. In this paper we will test for cointegration in panels, using the Westerlund (2007) error-correction approach.

Getting the correct functional form is a very difficult task to do, since economic theory does not provide a guide regarding the correct functional specification. In this paper we therefore model energy demand more flexibly by applying a fully nonparametric approach. Particularly the local linear kernel estimator is used in estimating energy demand for 17 OECD countries from 1960 to 2006. Modeling energy demand in a fully nonparametric sense will avoid the issue of functional mis-specification since we do not restrict the functional form a priori.

The main objective of this paper is thus to assess the commonly used log-linear functional form for energy demand models (our focus here is on long run energy demand model) which is the appropriate functional form for such models. We applied the panel error correction model to study if there is cointegration between energy demand, real price of energy, real income per capita and the effect of climate. This paper is organised as follows: Section 2 deals with the literature review for the study; Section 3 is the modeling of aggregate energy demand for 17 OECD countries. The data and the econometric analysis are presented in Section 4, and Section 5 concludes the paper with a summary of the findings and some proposals for future research.

2. Literature review

As mentioned in the introduction, there exists a large body of empirical research in estimating price and income elasticities for energy demand. However, most of the works have an individual country time series perspective or cross-sectional dimension of countries, but few exist in estimating the elasticities using panel data. Most of the panel data research on this topic rather focuses on sub-groups of aggregate energy demand such as industrial, residential and transport, or on specific energy sources such as electricity, gasoline, coal or gas. Our focus here is to review the existing literature on aggregate energy demand that uses panel data, but we will also consider the literature that are using aggregated OECD country data using a time series approach.

Balestra and Nerlove (1966) studied the demand for natural gas for US states between the periods 1950 to 1962. In their estimation of demand for natural gas, they assumed a linear specification in a dynamic panel model. The results from their study indicate that the long run price and income elasticities for the unconstrained model are −0.63 and 0.62, respectively, and −0.63 and 0.44 for the constrained model, where the depreciation rate is assumed to be 11% for all fuel-using appliances.

Using annual data for seven European Economic Community (EEC) Countries from 1955 to 1970, Kouris (1976) adopted both a time series model and a pooled model to determine the main factors affecting energy demand in EEC areas. The author included price, income and temperature as the main determining factors of aggregate energy demand in the seven EEC countries. In the empirical analysis, Kouris favored the estimates from the pooled model compared to the estimates from the time series model. One reason for this is that the price elasticity from the time series model for four of the countries was of the wrong sign, compared with what was expected from theory. Of the three remaining countries, only the Danish price elasticity was statistically significant. The pooled model resulted in a long run price elasticity of −0.76, and an income elasticity of 0.84. The author also investigated the variation of elasticities over time by dividing the entire sample into six-year overlapping intervals. The conclusion from this is that, all elasticities do vary from period to period.

Nordhaus (1977), on the other hand, studied the demand for energy in seven different western countries over the period 1955 to 1972. In this study, both individual country estimation and pooling was done for four major sectors of the economy, namely transport, domestic, industry and energy sectors. In addition to the four sectors, he also estimated energy demand at the aggregate (over sectors) level. The results from the individual country aggregate model shows that the short run price elasticities ranges from −0.03 to −0.68, while the income elasticity ranges from 0.29 to 1.11. The long run price elasticities however were estimated to be between −1.94 and 1.45, and the long run income elasticities were ranging from 0.26 to 1.42. The estimated price elasticity from the pooled model with country dummies was −0.85, and −1.15 for the model without country dummies. The income elasticities were 0.79 and 0.87 for the model with and without country dummies respectively. All the estimations were done with the assumption of a linear specification.

The standard approach used in the literature for accounting for technological progress is to include a time trend, especially in a dynamic model. In this regard, Beenstock and Willcocks (1981) included a time trend in an error correction model for annual aggregate energy demand model for industrialised OECD countries, using aggregated annual data from 1950 to 1978. In their study they found the short run price and income elasticities to be −0.01 and 0.37, respectively, and that of the long run to be −0.06 and 1.78, respectively. The authors also estimated a model without a time trend and found the long run price elasticity to increase to −0.13, and the long run income elasticity to decrease to 0.88, a value that is consistent with the, "around unity income elasticity estimates consensus" for econometric energy demand models.

Kouris (1983) estimated price and income elasticity for OECD countries by excluding the time trend as an explanatory variable in his aggregate energy demand model. Income elasticity was restricted to only account for short run dynamics, while allowing the price variable to have both a short run and long run effects. Estimates from the various sub-samples of OECD data set from 1961 to 1981 indicate that the short run price elasticity ranges from −0.14 to −0.32, while that of the long run ranges from −0.26 to −0.84. Prosser (1985), using annual data set for seven of the most industrialised OECD countries from 1961 to 1982, found similar income elasticities as in Kouris (1983) but different price elasticities both for the short run and the long run. The long run estimated price elasticity in Prosser (1985) ranges from −0.28 to −0.42. Welsch (1989), using both a pooled and country specific annual data on eight OECD countries, found the long run price elasticity to be −0.34 and the long run income elasticity to be 0.64 for the preferred pooled dynamic model without a time trend.

Jones (1994), on the other hand, used a much longer data set than that from Kouris and Prosser on the same seven OECD Countries. The idea of the study was to access the validity of Kouris and Prosser model, which is a restricted version of a general autoregressive distributed lag model without a time trend. Results from the study indicated that the restriction imposed by Kouris–Prosser is valid. The estimated long run price elasticity from Kouris–Prosser model without trend was −1.73, while adding a time trend this reduces the long run price elasticity to a reasonable value of −0.7.

[1] Almost ideal demand system proposed by Deaton and Muellbauer (1980) and extended quadratic form by Banks et al. (1997) to handle aggregated non-linear expenditure effects.

Adeyemi and Broadstock (2009) studied the contribution of consumer preferences to energy demand for 17 OECD countries from 1960 to 2004. In their preferred model, they found no evidence of price asymmetries and the estimated long run income elasticity was 0.5 and −0.1 for the long run price elasticity. A study by Filippini and Hunt (2011) on energy demand and energy efficiency for 29 OECD countries, using an aggregate annual panel data set, found the estimated price and income elasticities to be −0.45 and 0.81, respectively for the preferred pooled model with time dummies. The results from our parametric long rum model are consistent with that of Welsch (1989) and Filippini and Hunt (2011) in terms of sign and magnitude for both price and income elasticity estimates and consistent with the income elasticity from Kouris (1976) and Nordhaus (1977).

3. Energy demand model

The framework underlying the econometric analysis in this study is a derived demand framework. Since the empirical analysis concerns aggregate energy use, demand comes from final demand from households and from firms' demand for energy as an input. Implicitly we assume that firms choose the amount of energy in order to minimize costs (or maximise profits), and that households are utility maximizers with respect to energy consumption (and other consumption goods). The basic modeling strategy can be illustrated as follows for a representative household. Given an income, the household, or individual (consumer), chooses a bundle of goods that maximises his/her utility (U), i.e.:

$$\max_{Q} = U(Q) \quad s.t. PQ \leq Y \quad , \tag{1}$$

where $Q = [Q_1, \ldots, Q_K]$ is a vector of K goods and $P = [P_1, \ldots, P_K]$ is the corresponding price vector. The solution to the above consumer utility maximization problem is K demand functions with all prices and income as arguments:

$$Q_k^d = f(P_1, \ldots, P_K, Y), \quad k = 1, \ldots, K \tag{2}$$

A similar system of input demand functions for a typical firm can be derived by cost minimization. The demand function as presented in the above has the following implication for empirical estimation: the demand for any good (for example energy) will depend on all prices including energy. This makes the empirical estimation very challenging, as it requires data on prices on each and every good in the consumption set even if we are interested in only one of the goods, energy in this case. To get around this we assume that preferences are weakly separable in the sense that, the demand for energy for instance does not depend on the marginal rate of substitution between any two non-energy goods (for instance butter and margarine). Given this assumption we can view the households' consumption decision as a multistage process. In the first stage the consumer decides how much energy and non-energy goods she will consume, and given this the consumer decides in the second stage the composition of non-energy consumption. As a result, demand for energy from the first stage will not depend on individual non-energy commodity prices, but rather on the composite price of non-energy goods. Formally we can express the first stage problem as follows:

$$\max_{Q_E, Q_N} U(Q_E, Q_N) \quad s.t. \; P_E Q_E + P_N Q_N, \tag{3}$$

where Q_E represents the energy good and Q_N the non-energy good, while P_E and P_N are the corresponding nominal prices, respectively. The solution to the maximization problem in Eq. (3) gives the unconditional demand functions for the two aggregate goods Q_N and Q_E, respectively:

$$Q_N^d = f(P_N, P_E, Y) \tag{4}$$

$$Q_E^d = f(P_N, P_E, Y) \tag{5}$$

Since the demand function is homogeneous of degree zero in all prices and income, we can divide each of the variables in Eqs. (4) and (5) by P_N to get demand as a function of the real price and real income, i.e.:

$$Q_E^d = f(P_E^r, Y^r) \tag{6}$$

where $P_E^r = P_E / P_N$ and $Y^r = Y / P_N$. In the empirical model the composite price P_N will be approximated by the consumer price index (CPI), and expenditure, or income (Y), by GDP. Since in the empirical section, we will use time series country level data, i.e. panel data, Eq. (6) is further modified to accommodate for a vector of demand shifters as presented in Eq. (7):

$$Q_{E,it}^d = f\left(P_{E,it}^r, Y_{it}^r, Z'_{it}\right), \quad i = 1, \ldots M, \quad t = 1, \ldots, T \tag{7}$$

where i and t are country and time subscripts, and u_{it}denotes an unobserved term that is decomposed into an individual effect term (δ_i) and a pure random error term (ε_{it}), z'_{it} is a vector of demand shifters that include a cold climate dummy (CD) and time dummies to proxy the unobserved energy demand trend (UEDT)[2] that is expected to capture the effect of technology and structural changes in the economy. We get a similar expression as in Eq. (7) if we instead consider a representative firm's cost minimisation problem, where we assume a Cobb–Douglas production function with output being a function of labor, capital and energy as inputs in the production process. The energy input demand function can be shown to depend on output, price of energy, wages and rental rate of capital. If we consider rental rate of capital and wages as the price of other inputs, we can express the energy input demand as a function of real output and real price of energy. Both output and price of energy are divided by the price of other inputs to convert into real terms.

Our focus here is to investigate the functional forms and hence the variables that are mostly used in the empirical literature on aggregate energy demand. We therefore make the assumption that real price of energy, real income per capita and a vector of demand shifters are the only regressors that explicitly enter our energy demand model.

3.1. Econometric model

3.1.1. Parametric model for aggregate energy demand

The parametric energy demand models frequently used in the literature can be grouped into three functional forms: log-linear, linear and trans-log. Between the three functional forms, the log-linear form dominates in the literature, partly due to the simple economic interpretation and the convenience concerning formulation and estimation. We therefore focus on this functional specification for our parametric estimation, and henceforth all lower case variables denote variables that are in natural logarithm form.

The econometric strategy for this study is essentially divided into four steps: in the first step, we run a panel unit root test to determine

[2] Filippini and Hunt (2011) used UEDT to proxy for similar unmeasured variables as we did in this paper and we also borrowed the idea of using the Köppen–Geiger climatic classification to create our climate dummy variable. Note that the climate dummy used in this paper is prone to issues of aggregation just as in the case of using heating degree-days as a proxy for the effect of climate on a large country such as the USA. The USA is classified as a cold climate country as the predominant climate type is that of cold and temperate climate.

if the series are $I(0)$ or $I(1)$. In this study we will implement the Breitung (2000) unit root test and complement that with the Hadri (2000) unit root test. The Breitung unit root test follows the augmented- Dickey-Fuller (ADF) type of unit root testing but is extended to handle panel data with the restriction that rho is identical across individual units as presented in Eq. (8), while allowing the lag order of the first difference term to vary across individual units in the panel:

$$y_{it} = c_i + \lambda_i t + \rho y_{it-1} + \sum_{j=1}^{k} d_{ij}\Delta y_{it-j} + \varepsilon_{it} \tag{8}$$

where c_i is the constant term and t is the deterministic time trend. The null and alternative hypothesis of the test are: H_0: $\rho = 0$ and H_1: $\rho < 0$.

The Hadri (2000) unit root test on the other hand follows the KPSS[3] approach but is extended to handle panel data. The null hypothesis of the Hadri (2000) unit root test is that the series is stationary against the alternative of a unit root and therefore can serve as a check on the Breitung unit root test that has its null hypothesis as a unit root. Based on the outcome of the unit root test, we run a panel cointegration test, using the statistics proposed by Westerlund (2007). The choice of the Westerlund approach to the residual-based approach is due to its robustness in regards to size accuracy and power in a small sample. Westerlund (2007) approach is therefore appropriate for our case.

The basic idea of the Westerlund approach is that, given that the variables are $I(1)$, then we can test for cointegration by formulating an error correction model (ECM) and testing whether there is an error correction in the model or not. Given that there is error correction, it implies that there is cointegration while the opposite is true if no error correction is found in the model. The model for testing for cointegration can be expressed as:

$$\Delta e_{it} = \delta'_i d_t + \alpha_i e_{i,t-1} + \theta'_i x_{i,t-1} + \sum_{j=1}^{m_i} \alpha_{ij}\Delta e_{i,t-1} + \sum_{j=-q}^{m_i} \gamma_{ij}\Delta x_{i,t-j} + v_{it} \tag{9}$$

where the deterministic component is represented by d_t and comprises of a constant and a time trend, $t = 1,\dots\dots,T$ which is the index for time and $i = 1,\dots\dots, N$, the index for the individual panel units (countries in this study). The index $j = 1,\dots\dots, m$ is the lag term of the model, while $j = -q,\dots\dots, m$ is the leading term, x_{it} is a vector of weakly exogenous regressors[4] (in this paper the regressors are y^r_{it}, p^r_{it} and z'_{it}) and v_{it} is the random error term. Testing for cointegration therefore implies testing for error correction in the model, if $\alpha_i<0$. Given that $\alpha_i<0$, we have an error correction and it implies a cointegration relationship between $e_{it}$and x_{it}, while if $\alpha_i = 0$ we have no cointegration relationship. Westerlund proposed four test statistics, which are grouped into two, namely the group-mean statistics (G_t and G_a) and panel statistics (P_t and P_a). The null hypothesis of no cointegration for each of the four statistics is that H_0:$\alpha_i = 0$ for all panel units (details on the alternative hypothesis specification for each of the two groups of test statistics, derivation and asymptotic distribution are found in Westerlund (2007)).

In the case of cointegration, the next step is to estimate the conditional long run model from the reduced form solution from Eq. (9), by setting $\Delta e_{it} = \Delta e_{i,t-1} = \Delta x_{i,t-j} = 0$:

$$e_{it} = \sigma_i + \phi_1 p^r_{it} + \phi_2 y^r_{it} + \phi_3 z'_{it} + u_{it} \tag{10}$$

[3] KPSS denotes the Kwiatkowski–Phillips–Schmidt–Shin test for testing for stationarity of a series around a deterministic trend in time series literature.

[4] The main assumption of the Westerlund (2007) approach is that the regressors are weakly exogenous, though restrictive, it is a reasonable assumption to make in aggregate energy demand estimation, especially with the income variable.

where $\phi_1 = \frac{-\theta_1}{\alpha_i}, \phi_2 = \frac{-\theta_2}{\alpha_i}, \sigma_i$ is the individual specific effect term, which is the deterministic term as in Eq. (9) but restricted to the case where we do not have individual specific time trend. The individual effect is assumed to be uncorrelated with $p^r_{it}, y^r_{it}\,and\,z'_{it}$ for a random effect (RE) model, but correlated with $p^r_{it}, y^r_{it}\,and\,z'_{it}$ for a fixed effect (FE) model. Economists generally favor the FE model as being the model that is more likely to be supported by the data compared to the RE model. The RE model assumes that the unobserved individual heterogeneity is distributed independently of the regressors, which is a very restrictive assumption that most economic data are unable to meet. If the DGP is that of the individual-specific fixed effect model, estimates from the RE model are inconsistent (Cameron and Trivedi, 2005). Our *a priori* believe is that the DGP supports the FE model but we will formally test this, using the Hausman test as explained in Cameron and Trivedi (2005), and in Wooldridge (2010).

From Eq. (10), the long run price and income elasticities are calculated via the expressions in Eqs. (11) and (12) respectively:

$$\eta_p = \frac{\partial e_{it}}{\partial p_{it}} = \phi_1 \tag{11}$$

$$\eta_y = \frac{\partial e_{it}}{\partial y_{it}} = \phi_2 \tag{12}$$

where η_p and η_y are price and income elasticity, respectively. In empirical work on energy demand modeling, price responses are assume to be symmetric but there exists other works such as Gately and Huntington (2002) and Griffin and Schulman (2005), in which energy demand is modeled in the presence of asymmetric price responses. In this paper, the expression in Eqs. (9) and (10) assumes symmetry of prices in modeling aggregate energy demand. We will estimate Eq. (10) using FE or RE, depending on which of them that support the DGP.

3.1.2. Nonparametric model for aggregate energy demand

The last step in our estimation strategy is to test the linear specification of the parametric model using nonparametric the correct model specification test (the I_n test, see Appendix A for details) proposed by Härdle and Manmen (1993) and Zheng (1996) and also extended by Hsiao et al. (2007) to account for the presence of discrete regressors, to assess the accuracy of using the log-linear functional form for the data. If the model fails the test, we will apply the nonparametric approach to re-examine the relationship between energy demand, price and income to identify the functional relationships between the dependent variable and the regressors in the model. The non-parametric aggregate energy demand model is expressed as:

$$e_{it} = m(X_{it}) + u_{it}, i = 1, \dots\dots, N\,and\,t = 1, \dots\dots, T \tag{13}$$

where X_{it} is a vector of explanatory variables (price, income, CD and UEDT), and all other variables in Eq. (13) assume the same definitions as in Eq. (9), while $m(.)$ is an unknown smooth function and it is estimated using kernel methods. The main idea of the nonparametric approach is to allow the data to determine the functional relationships between variables and it is therefore a data driven approach that does not specify the functional relation a priori. The major drawback of this approach is that the rate of convergence to the true model is slow compared to a correctly specified parametric model, referred to in the literature as “the curse of dimensionality”. It is true that an incorrect parametric specification will not even converge in the first place and therefore in such a case, the slow to converge nonparametric estimation is better. If the functional specification for the parametric model is the correct functional relation, the parametric model is preferred; if not, the nonparametric model is preferred, as it at least avoids issues related to functional misspecification such as biasness and inconsistency of the estimates.

The estimation strategy is in three steps: in the first step we will determine the optimal bandwidth for the chosen kernel estimator that will be implemented in estimating $m(.)$. In the second step we implement the optimal bandwidth found in the first step to estimate Eq. (13) and in the final step we obtain the partial regression plot and partial gradient plots for each of the regressors, adapting a wild bootstrap to construct heteroskedastic consistent standard errors for the confidence bands for the partial regression and partial gradient plots. Eq. (13) is estimated using the local linear kernel[5] estimator proposed by Stone (1977) and Cleveland (1979) and extended by Li and Racine (2007) to handle panel data with both continuous and discrete variables, and implemented in Gyimah-Brempong and Racine (2010). For a detailed account of this estimator, see Appendix B.

The choice of the local linear kernel estimator instead of the most popular kernel estimator used in empirical work (the local constant kernel estimator) is due to the ability of the former to correct boundary bias. Another reason is that, the cross-validated local linear kernel estimator has a faster convergence rate when the underlying relationship in the DGP is relatively linear compared to other nonparametric kernel estimators. It achieves the same rate of convergence as the case of a truly specify parametric model when the correct relationship in the data is linear (Li and Racine, 2004). Besides, this estimator also allows for all possible nonlinearity and interactions between the regressors. Eq. (13) is estimated using the leave-one-out local linear kernel estimator to avoid over-fitting.

4. Data and results

The consumption, price and income data series for this paper were all taken from Broadstock et al. (2010), an annual data set for a panel of 17 OECD countries covering the period 1960 to 2006. The variables comprise aggregate energy consumption (E) for each country expressed in thousand tonnes of oil equivalent (ktoe), GDP (Y) in billions of 2,000 US$ using PPP for the entire period and the real energy price index (P) at 2,000 US dollars. Both Y and E are converted into per capita terms by dividing each country's Y and E by their respective populations. A cold climate dummy (CD) variable that takes a value of 1 for a country characterised as belonging to a cold climatic zone and zero otherwise, based on Köppen–Geiger classification (details of the classification is in Peel et al., 2007) and a series of time dummies (UEDT). All the key variables and their descriptive statistics are presented in Table 1, the Countries in the study are: Austria, Belgium, Canada, Denmark, France, Greece, Ireland, Italy, Japan, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, the UK and the USA.

The plots in Appendix D reveal the pattern of trends in the individual series for each of the countries in the study. The energy consumption per capita series (henceforth referred to as energy consumption) shows a positive increasing trend for all the countries except that of Denmark, Sweden and the USA, where it increases up to 1975 and then remains relatively constant thereafter for Denmark and Sweden. The USA had an increasing trend up to 1975 and thereafter a downward trend. Real income per capita series (henceforth referred to as income) on the other hand shows a positive trend for all countries, while that of the real price series (referred to as price for rest of the paper) for seven of the countries is upward trending, the ten remaining countries have a trend pattern for the price series that is increasing at certain periods and decreasing at other periods.

[5] There exist several kernel estimators and they include the local constant kernel estimator, local linear kernel estimator and local polynomial kernel estimator. See Fan and Gijbels (1996) for an in-depth treatment of local linear and polynomial kernel estimators.

Table 1
Descriptive statistics.

Variables	Mean	Std. Dev.	Minimum	Maximum
Energy consumption per capita (E)	2.768	1.391	0.239	6.346
Real price of energy (P)	95.340	21.174	44.752	170.300
Real GDP per capita (Y)	19.000	7.154	3.776	39.641
Climate dummy (CD)	0.470	0.499	0	1

4.1. Interpretation of parametric results

4.1.1. Results from unit root and Cointegration test

Results from the Breitung (2000) unit root test could not reject the null of unit root for each of the variables: price, income and energy consumption, implying they are all trending and therefore follow a unit root process.

As a check on the Breitung unit root test, we applied the Hadri unit root test with the null hypothesis being a stationary process. The results from the Hadri unit root test rejected the null of a stationary process for each of the series, implying unit root for the five series. Test results for the two panel unit root tests implemented in this paper are reported in Table 2. The general conclusion is that, all the series are $I(1)$ and hence we can proceed to apply the Westerlund (2007) cointegration test.

Applying Eq. (9) to the data set produces the results presented in Table 3 below. The two panel cointegration statistics rejected the null of no error correction in favor of the alternative, implying cointegration on the whole panel. One of the two group statistics (G_a), however, could not reject the null hypothesis based on the assumption of cross-sectional independences of the residuals. We also calculated the p-values using a bootstrap approach to allow for possible cross-sectional dependence and the results from the bootstrap calculated p-values indicate clear rejections of the null for both the panel (P_t and P_a) and group (G_t and G_a) test statistics at the conservative 1% level of significance. Our broad conclusion is that, there is cointegration between energy consumption, price and income, and therefore there is a long run equilibrium relationship such as the expression in Eq. (10).

To check for the validity of the weakly exogenous assumption for the regressors, we run two separate regressions with price and income as the dependent variables and energy consumption as the regressor. The idea is that if a variable is weakly exogenous in a model, using such a variable as the dependent variable in a regression should not generate any cointegration relationship on the endogenous variables in the model. The results from the reversed regressions indicate, both income and price variables are weakly exogenous. Since the null of no cointegration is not rejected at the 1% level based on the bootstrap p-values, except in the case of the price variable that the null was rejected for one of the group test statistic (G_t) (results are reported in Table E1, Appendix E).

4.1.2. Results from long run parametric aggregate energy demand model

We estimated both the FE and RE models using Eq. (10), but the Hausman test statistic supported the FE model as the appropriate parametric model. Henceforth, we will base our analysis for the long run parametric model on the FE model. The estimated long run own price elasticity and the income elasticity both have the expected sign and significance at the 1% level and the size is within the range of values usually found in the literature for price elasticity. For instance Filippini and Hunt (2011) found the own price elasticity to be approximately −0.3 for 29 OECD countries. The estimated long run income elasticity is also significant at the 1% level and has the expected sign: the size (0.87) is very close to the income elasticity (0.9) found in Filippini and Hunt (2011).

However, the I_n test results indicate that the log-linear applied in estimating the long run model is not the correct functional specification, hence we cannot make any meaningful inferences from the long

Table 2
Results of Breitung and Hadri Unit Root Tests.

Variables	Breitung				Hadri			
	Constant	*p*-Value	C&T	*p*-Value	Constant	*p*-Value	C&T	*p*-Value
e	2.955	0.998	2.017	0.978	82.677	0.000	58.751	0.000
p	−0.635	0.262	−0.063	0.474	45.286	0.000	29.836	0.000
y	6.010	1.000	2.675	0.996	115.882	0.000	63.659	0.000

Notes: C&T are for constant and trend. The Null for Breitung tests is that of unit root. Null for Hadri is that of stationarity.

Table 3
Results from Westerlund ECM panel cointegration tests.

Statistic	Value	*Z*-value	*p*-Value	p^*-value
G_t	−2.645	−1.802	0.036	0.000
G_a	−8.477	1.459	0928	0.000
P_t	−9.565	−1.558	0.060	0.000
P_a	−8.574	−0.700	0.242	0.000

Notes: p^* is the *p*-value based on 100 bootstrap replications, $G_{t,}$ and G_a denotes the test statistics for group panel Cointegration test, while P_t and P_a is the panel cointegration statistics. The optimal lag (1.18) and lead (0.29) lengths are chosen based on AIC. The H_0: no cointegration.

run estimates as reported in Table 4. One major concern is that of omitted variables and the possible implication of that on the test results and the estimated parameters. However, given that we included the most important variables that influence energy consumption, it is very much likely that even if there are still some omitted variables, their effects on the results are minuscule. For instance variables, such as government policy, are not included explicitly in our model but we believe its impact on energy consumption per capita will be picked up by the proxy for technology due to the fact that, most if not all of government policies on energy are directed towards conservation and therefore efficient use of energy, but there is a close relationship between technical progress, efficiency and conservation. Another important variable we could not explicitly control for is the size of the industrial sector to proxy the effect of relocation of manufacturing industries to other countries outside the group of OECD countries.

We check for the impact of not including the size of the industrial sector, we tried two proxy variables such as the value added by the industrial and service sectors[6] and their effects on the I_n test outcome. The results indicate no effect on the test outcome (details are in the next section). We also tried different specification of Eq. (10) by including the proxy for the industrial and service sectors of each of the countries, but the test results from this specification could not also pass the I_n test (results are reported in Table E2 in Appendix E). We further tried two linear parametric and translog specifications, one with the two proxy variables for the size of the industrial and service sectors and another without, but none could pass the I_n test. These results are reported in Table E3 and E4 in the appendix. The conclusion therefore is that, our results from the log-linear model are robust for the non-inclusion of the size of both the industrial and service sectors and that linear parametric and translog specifications are not also appropriate, irrespective of including proxies for the size of both the industrial and service sectors.

4.2. Interpretation of non-parametric result

4.2.1. Optimal bandwidth

The I_n test for the parametric model indicated that the log-linear form is not an appropriate form to capture the true relationship between aggregate energy consumption, price of energy, income, impact of climate and the size of both industrial and service sectors. As a result, a fully nonparametric model as presented in Eq. (13) is applied to the data set, where the functional form is not specified a priori. The estimation is done using the local linear kernel method with the Hurvich et al. (1998) AIC approach in the np package in R-statistical software. The optimal bandwidth is selected using Kullback–Leibler cross-validation approach and the results are reported in Table 5. The bandwidth for the price variable is approximately 0.769 with a scale factor of 13.260, while the optimal bandwidth for income is 0.122 with a scale factor of 0.883. Based on the optimal bandwidth results, we expect the price variable to be linear while that of income to be non-linear; this is the case because for a non-linear continuous variable, the scale factor is expected in theory to be small and large for a linear variable. The bandwidth for the climate variable is approximately 1, while that of UEDT is about 0.7.

4.2.2. Nonparametric regression using local linear kernel estimator

4.2.2.1. Partial regression and partial gradient plots. The partial regression estimates[7] measure the relationship between aggregate energy consumption and each of the explanatory variables, holding the others constant at their mean/mode, while the partial gradient measures the slope of the partial regression line. The plots for both the partial regression and gradient lines are presented in Figs. 1 and 2, respectively. The plot in Fig. 1 indicates a negative relationship between energy consumption and price of energy where the relationship appears to be a linear one, which is consistent with the large-scale factor for the optimal bandwidth for the price variable as reported in Table 4. The gradient plot for energy and price relationship as presented in Fig. 2 confirms the negative relationship between these two variables as well as the linear relationship between them. The gradient plot can be interpreted as the estimated price elasticity, which is linear and varies slightly from 0.18 to 0.19 as the price level increases from 3.6 to 5. The graph in Fig. 1 for the income variable depicts a positive but non-linear relationship with aggregate energy consumption:

The income gradient as shown in Fig. 2 confirms the positive but non-linear relationship between income and aggregate energy demand. It varies between 0.4 and 2.2 for the nonparametric model. From the nonparametric estimates, the gradient plot indicates that income elasticity ranges from 1.5 to 2.2 for income levels that ranges from 1.5 to 2.18 per capita and lies between 0.4 and 1.5 for income levels above 2.18. The estimated gradient for the climate variable is approximately 0.04 and that of UEDT ranges from −0.0175 to 0.0024. To avoid interpreting insignificant results, we computed a 5% confidence band for the gradients using "wild" bootstrap methods (Härdle and Marron, 1991) and the results as presented in Fig. 2 above, indicate that both the price and income elasticity estimates lies within the band and therefore we conclude that they are significant. Interestingly, the results from the non-parametric model indicates that

[6] We did not include the two proxies for the size of the industrial sector in our preferred model due to lack of data and this could generate more serious consequences on the estimates than the possible omitted bias it might cause for non-inclusion. A number of countries in our analysis only have data starting from the 1990s. For purposes of the functional form test, we interpolated the data to fill in the missing data to perform the functional form test and therefore the results should be treated with caution.

[7] The partial regression estimates and the partial gradient measures how the dependent variable and its response surface change in response to changes in explanatory variables, holding all other variables constant at the mean/mode. While partial regression estimates refer to $\hat{m}(x_{it}^c, x_{it}^d)$, while that of the partial gradient refers to $\hat{\beta}(x_{it}^c)$.

Table 4
Estimates from long run FE model.

Variable	Estimates	*p*-Value
p	−0.495	0.000
y	0.874	0.000
CD	0.542	0.000
Hausman test statistic	47.446	0.000
I_n-test statistic	14.14554	0.000
UEDT	Yes	
Adjusted *R*-squared	0.942	

Notes: number of cross sectional units is 17, time period is 47 and a total of 799 observations. The null hypothesis for the I_n test is; correct specification.

Table 5
Optimal bandwidth selection and local-linear regression estimates.

Variable	Optimal bandwidth	Scale factor/lambda max.
p	0.769	13.260
y	0.122	0.883
CD	0.98	9.098
UEDT	0.731	6.789
Factor (country)	0.0005	0.005
R^2	0.998	
Residual standard error	0.0004	
Regression type: local-linear		

Note: The Bandwidth selection method is that of the Expected Kullback–Leibler cross-validation, Factor (Country) denote dummy for the countries under study, CD is climate dummy and UEDT is time dummies.

energy tends to be a luxury good at low income levels, whereas energy moves towards a normal good, or a necessity as income rise. This is an important finding, not the least when we want to analyse the effects of different energy policies under different growth scenarios.

5. Conclusion

In this paper we applied a nonparametric approach to investigate whether the log-linear functional form usually assumed in the empirical literature in estimating aggregate energy demand models is appropriate. To do this we used a panel data on aggregate energy consumption in 17 OECD countries. The log-linear parametric model could not pass the nonparametric correct model specification test. Since the test itself does not give any guidance concerning what parametric specification is to be chosen. One possible option is to use a less restrictive modeling approach, such as the nonparametric approach, which provides an excellent option in this regard. The results from the nonparametric estimation indicate a non log-linear relationship between income and consumption, but with a log-linear relationship between the price of energy and energy consumption. Furthermore, it turns out that the coefficient of determination is higher in the non-parametric than in the log-linear parametric model. The estimates for the price variable from the nonparametric estimation appear to be log-linear but vary over time; the price elasticity is −0.18 at low price levels and decreases only slightly (increases in absolute terms) to −0.19 at higher price levels. The income elasticity on the other hand is fluctuating relatively much with respect to the income level, although the trend is decreasing in the sense that the income elasticity falls as income rises.

In a nutshell we can say that for a panel of developed countries the simple log-linear functional specification appears to be an incorrect functional form to use for aggregate energy demand modeling, more so with respect to the income variable. The log-linear parametric model performed badly in capturing the price elasticity when the benchmark is the estimates from the more flexible nonparametric model. For the income elasticity, the log-linear parametric specification performed almost equally as poorly, if not worse, using the nonparametric estimates as the benchmark. An alternative linear and translog specification also fails the functional specification test, which implies that both the linear and the translog specification are also inappropriate. It will be very interesting to investigate this on individual country basis using time series data, but the nonparametric approach is data driven and therefore requires a long time series data, which we do not have at the moment. An alternative may therefore be to apply a semi-parametric approach, where we assume a log-linear parametric relation between energy consumption, price

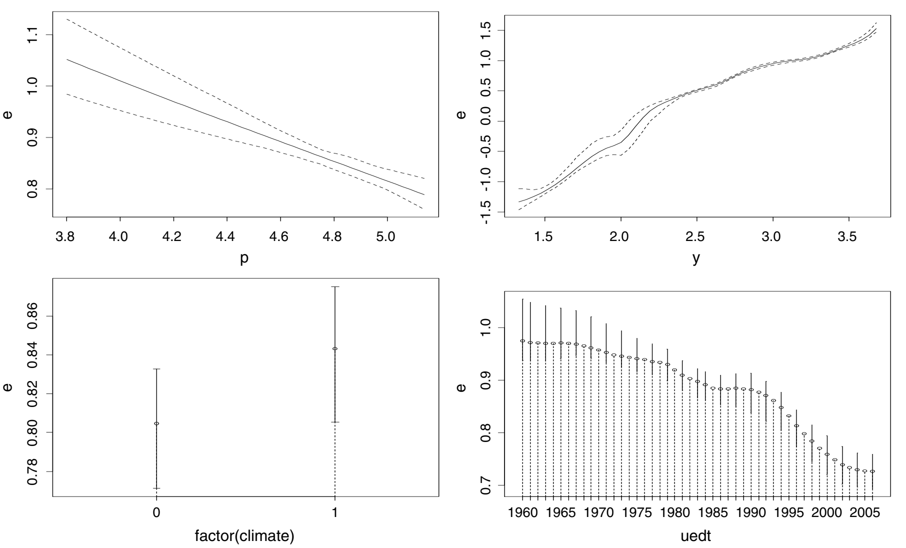


Fig. 1. Nonparametric partial plots of the estimates for price and income for energy demand.

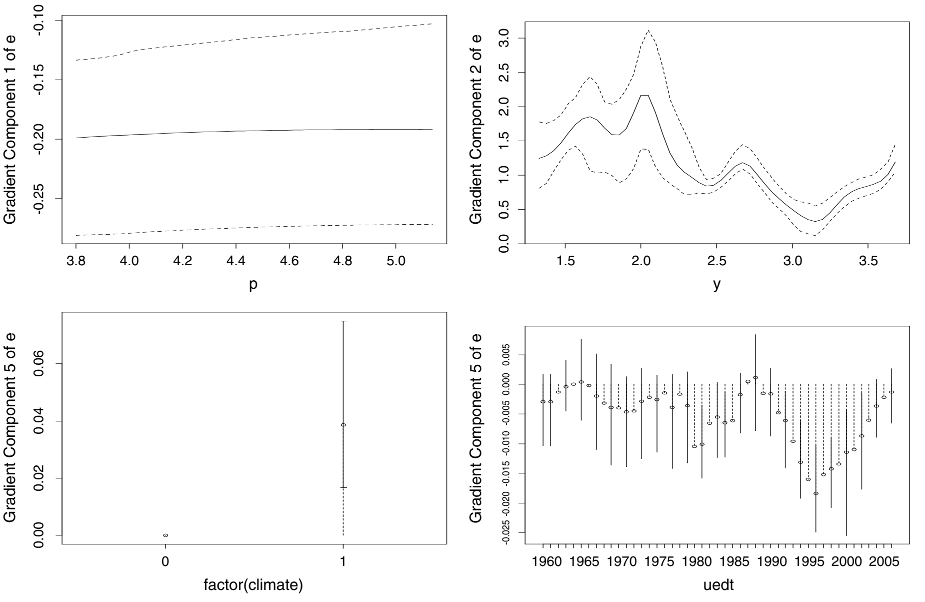


Fig. 2. Nonparametric plot of the gradients (elasticities) for price and income for energy demand.

and other important covariates, while we model the income variable nonparametrically.

As regards to policy implications, the results from the nonparametric model shows that the response of energy consumption to price changes is relatively low, compared to the values usually used in energy policy discussions, which are based on results from parametric models. One implication is that the potential policy impact on energy consumption from, say, an energy tax may be exaggerated. Furthermore, since the response to income seems to vary across income levels with both a trend and a fluctuation around the trend it should be clear that this may have serious consequences on the success of energy policy. Again, we may very well over- or underestimate the effect of policy, especially the long run effects depending on what assumptions we make concerning economic growth. Finally, it is important to treat the results here with caution because the nonparametric approach is also prone to most of the problems associated with the parametric model, except that of functional misspecification.

Acknowledgments

Insightful comments and suggestions from Kurt Brännäs, Andrius Kazukauskas, Jurate Jaraite, the editor of the journal and an anonymous referee whose comments enrich the quality of this paper. Financial support from the Swedish Energy Agency is also gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.eneco.2012.11.026.

References

Adeyemi, Broadstock, 2009. Underlying consumer preferences and their contribution to energy demand. OPEC Energy Rev. 33 (3–4), 198–204.

Atkinson, Manning, 1995. A survey of international energy elasticities. In: Barker, Terry, et al. (Ed.), Chapter 3 in *Global Warming and Energy Demand*. Routledge 11 New Fetter Lane, London.

Balestra, Nerlove, 1966. Pooling cross-section and time-series data in the estimation of a dynamic model: the demand for natural gas. Econometrica 34, 585–612.

Banks, Blundell, Lewbel, 1997. Quadratic Engel curves and consumer demand. Rev. Econ. Stat. 79 (4), 527–539.

Beenstock, Willcocks, 1981. Energy consumption and economic activity in industrialised countries. Energy Econ. 3, 225–232.

Bohi, 1981. Analyzing Demand Behaviour: A Study of Energy Elasticities. John Hopkins University Press, Baltimore.

Bohi, Zimmerman, 1984. An update on econometric studies of energy demand behavior. Annu. Rev. Energy 9, 105–154.

Breitung, 2000. The Local Power of Some Unit Root Tests for Panel Data. In: Baltagi, B. (Ed.), Nonstationary Panels, Panel Cointegration and Dynamic Panels. : Advances in Econometrics, vol. 15. JAI, Amsterdam.

Broadstock, Hunt, 2010. Quantifying the Impact of Exogenous Non-Economic Factors on UK Transport Oil Demand. Energy Policy 38 (3), 1559–1565.

Cameron, Trivedi, 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

Cleveland, 1979. Robust locally weighted regression and smoothing scatter plots. J. Am. Stat. Assoc. 74, 829–836.

Dahl, 2005. A survey of energy demand elasticities for the developing world. J. Energy Dev. 18 (1), 1–47.

Deaton, Muellbauer, 1980. An almost ideal demand system. Am. Econ. Rev. 70, 312–326.

Fan, Gijbels, 1996. local polynomial modelling and its applications. Chapman and Hall, London.

Filippini, Hunt, 2011. Energy demand and energy efficiency in the OECD countries: a stochastic demand frontier approach. Energy J. 32 (2), 59–80.

Gately, Huntington, 2002. The asymmetric effects of changes in price and income on energy and oil demand. Energy J. 23 (1), 19–55.

Griffin, Schulman, 2005. Price asymmetry in energy demand models: a proxy for energy-saving technical change? Energy J. 26 (2), 1–21.

Gyimah-Brempong, Racine, 2010. Aid and investment in LDCs: A robust approach. J. Int. Trade Econ. Dev. 19 (2), 319–349.

Hadri, 2000. Testing for stationarity in heterogeneous panel data. Econ. J. 3, 148–161.

Härdle, W., Manmen, E., 1993. Comparing nonparametric vs. parametric regression fits. Ann. Stat. 21 (4), 1926–1947.

Härdle, Marron, 1991. Bootstrap simultaneous error bars for nonparametric regression. Ann. Stat. 19 (2), 779–796.

Hartman, 1979. Frontiers in energy demand modelling. Annu. Rev. Energy 4, 433–466.

Hsiao, Li, Racine, 2007. A consistent model specification test with mixed categorical and continuous data. J. Econom. 140 (2), 802–826.

Hurvich, Simonoff, Tsai, 1998. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. J. R. Stat. Soc. Ser. B 60, 271–293.
Jones, 1994. Accounting for technical progress in aggregate energy demand. Energy Econ. 16, 245–252.
Kouris, 1976. The determinants of energy demand in the EEC area. Energy Policy 4, 343–355.
Kouris, 1983. Fuel consumption and economic activity in industrialised economies: a note. Energy Econ. 5, 207–212.
Li, Racine, 2004. Cross-validated local linear nonparametric regression. Stat. Sin. 14 (2), 485–512.
Li, Racine, 2007. Nonparametric Econometrics: Theory and Practice. Princeton University Press.
Nordhaus, 1977. The Demand for Energy: An International Perspective. In: Nordhaus, W.D. (Ed.), International Studies of the Demand for Energy. North-Holland, Amsterdam.
Peel, Finlayson, McMahon, 2007. Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644.
Prosser, 1985. Demand Elasticities in OECD: Dynamic Aspects. Energy Econ. 7, 9–12.
Spiegehalter, Best, Carlin, Linde, 2002. Bayesian measures of model complexity and fit (with discussion). J. R. Stat. Soc. 64, 583–639.
Stone, 1977. Consistent nonparametric regression. Ann. Stat. 5, 595–645.
Welsch, 1989. The reliability of aggregate demand functions: an application of statistical specification error tests. Energy Econ. 11, 285–292.
Westerlund, 2007. Testing for error correction in panel data. Oxf. Bull. Econ. Stat. 69, 709–748.
Wooldridge, 2010. Econometric Analysis of Cross Section and Panel data. Massachusetts Institute of Technology Press.
Xiao, Zarnikau, Damien, 2007. Testing functional forms in energy modeling: an application of the Bayesian approach to U.S. electricity demand. Energy Econ. 29 (2), 158–166.
Zarnikau, 2003. Functional forms in energy demand modelling. Energy Econ. 25, 603–613.
Zheng, 1996. A consistent test of functional form via non-parametric estimation functions. J. Econ. 75, 263–290.

III

Effects of demography and labor supply on household gasoline demand in Sweden: a semiparametric approach

Amin Karimu
Department of Economics
Umeå School of Business and Economics
Umeå University

Abstract

A typical gasoline demand model generally assumes that demand and labor supply are weakly separable. In this study, I relaxed the weak separability assumption by examining the effect of labor supply, measured by male and female working hours, on gasoline demand. I used a flexible semiparametric model that allowed for differences in response to income, age and labor supply, respectively. Using Swedish household survey data, the results indicated that the relationship between gasoline demand and income, age and labor supply were non-linear. The formal separability test rejects the null of separability between gasoline demand and labor supply. Furthermore, there was evidence indicating small bias in the estimates when one ignored labor supply in the model.

Keywords: Derivatives, gasoline demand, labor supply and propensity score.

JEL Classification: Q4, C3, C4.

1. Introduction

Understanding the factors that influence gasoline demand has been of great interest among economists for the past four decades, and this interest is still growing due to the fact that gasoline constitutes a great portion of total world energy use, and more importantly to the share of global carbon dioxide (CO_2) emissions generated from energy sources. Early studies of gasoline demand were mainly focused on concerns relating to the depletion of oil resources and its implications for national security. In recent years, this focus on resource depletion and national security has shifted to environmental consequences of gasoline consumption, particularly in the area of the emission of greenhouse gases.

The fact that policies related to gasoline consumption are a cornerstone of the overall energy policy in most countries, especially in the developed world, has led to many studies in this area. However, the studies in this area vary considerably; in the selection of sample country/countries, type of data, time period, and model specification. Dahl (2012) contains a recent comprehensive survey of the literature on gasoline demand. Many of the gasoline demand studies utilise aggregate data to estimate the price and income elasticities, while only few of them utilise data at the household level (Archibald and Gillingham, 1980, 1981; Greening et al., 1995; Hausman and Newey, 1995, 1998; Schmalensee and Stoker, 1999; Kayser, 2000; Yatchew and No, 2001; Nicol, 2003; Brännlund and Nordström, 2004; and Wadud et al., 2010). Irrespective of the fact that studies based on aggregate data do provide important information on average price and income elasticity for a country or region and are therefore important for nationwide pricing or gasoline tax policy, they fail to account for some important household specific responses to price and income changes. For example, effects of rural or urban location, as well as the response of individuals or households in different income groups to both price and income changes. Recent studies have found heterogeneous responses by households to both gasoline price and income changes (Nicole, 2003; West and Williams, 2004; Brännlund and Nordström, 2004; Wadud et al, 2007, 2010) and therefore using household data will further aid our understanding of gasoline demand. In addition, most empirical studies of gasoline demand assume that preferences over goods are separable from that of labor supply. But as Browning and Meghir (1991) pointed out, "casual observation suggests that this may be a poor assumption; obvious examples are the (travel and child care) costs of going to work and, say, heating needs in the day". In this study, the weak separability assumption between gasoline demand and labor supply is relaxed and is further tested, using the conditional approach[1], where male and female hours of work are used as right-hand side variables to proxy for labor supply.

This paper makes two key contributions to the literature, the first contribution is in the estimation of a flexible semiparametric gasoline demand model that allows for possible non-linearity between income, age, and labor supply of the households, while controlling for

[1] This approach is in contrast to the unconditional approach that requires joint system of demand and labor supply estimation, implying that conclusions are based on having a correct model of labor supply determination, which is not necessary in the case of the conditional approach. Besides, since this study used a few cross sections, and hence has a very small price variation, the focus here is on income and other household characteristics and locational variables.

locational and household characteristics. The second contribution of the paper is in the modelling and correction for selection bias. Instead of using the Heckman parametric selection model I extended the nonparametric selection model proposed by Das et al. (2003) to a gasoline demand model. The reason for doing this was to avoid the issue of misspecification of functional form in the selection equation, instead of a priori specifying the functional form as in the standard Heckman selection model. It is true that there are other studies in the literature on the demand for gasoline that have applied semiparametric methods in estimating gasoline demand model. For instance Hausman and Newey (1995, 1998) used the partial linear model to study household gasoline price and income elasticity for the U.S. The results indicate that welfare effects of tax changes are remarkably different if price effects are estimated non-parametrically rather than parametrically[2].

Schmalensee and Stoker (1999) applied the partial linear model to study household gasoline demand in the U.S. focusing more on demographic effects. They modelled income and age non-parametrically while the other variables entered their gasoline demand model parametrically. They authors concluded that, the inclusion of demographic variables such as number of licensed drivers and age had a great impact on gasoline consumption and should be included in gasoline demand models. But they could not estimate price effect as

[2] Refers to total welfare effect of a gasoline tax of \$0.30 and \$0.50 and therefore includes the deadweight loss. The difference in welfare is mainly due to differences in the estimated deadweight loss from the log-linear parametric and the nonparametric models, which is in the range of 40-50%. The difference may be explained by the constant restriction on demand response to price changes in a log-linear model, while the nonparametric model does not restrict demand response to price changes to being constant.

they did not have reliable data on gasoline prices. They further indicated that including demographic variables reduced their income estimate by half. Hausman and Newey, who included both price and income variables in their gasoline demand model, did not however include demographic variables, and one may argue that including demographic variables may alter their estimated results. Schmalensee and Stoker on the other hand included demographic factors in their model but excluded the price variable, which could also alter their finding. To reconcile this, Yatchew and No (2001) included both demographic and price variables in their gasoline demand model, and they modelled the relationship between gasoline demand, gasoline price, and age with a flexible specification, while income, demographic and locational variables entered the model parametrically. Using Canadian data, they found that price was orthogonal to demographic variables and concluded that; the absence of demographics in Hausman and Newey and non-availability of reliable price data in Schmalensee and Stoker "would appear not to compromise either set of results".

Wadud et al. (2010) also adopted a semiparametric approach to model gasoline demand in a panel data set for the U.S. In their study, they modelled both price and income using a more flexible specification, while modelling the other variables in a linear parametric specification. Their study results indicated that price responses do vary with demographic factors such as multiple vehicle holding, presence of multiple wage earners and rural or urban location. They also found that households' response to price change decreases with higher income, but they found no significant differences between the results from the semiparametric model and a parametric translog model.

In all the above literature on the use of the semiparametric approach in analysing household gasoline demand, none of them jointly modelled income, age, and labor supply using a flexible specification. Moreover, the data set used in this paper is different from the data sets used by the cited literature, since it is a Swedish household data set. One could argue that the market for gasoline differs substantially between North America and Sweden. In North America prices vary regionally and over time, whereas in Sweden prices between regions are almost perfectly correlated over time. Hence one could argue that omitting prices is less severe in the Swedish case. More importantly, none of the above studies considered the issue of selection. Either they do not have such an issue in their respective data sets, or they have a selection issue but decided to be silent about it. The only paper that considers labor supply effects on household gasoline demand and also correcting for selection is that of Brännlund and Nordström (2004), but they did so in a parametric model that a priori specifies the functional relationship in both the selection and outcome equations. In this paper, there is also an issue of selection and I propose to use a nonparametric selection approach to correct for the possible selection bias in the gasoline demand model.

The main objective of this paper is thus to estimate the relationship between gasoline demand, age, household income, and labor supply in a flexible specification that allows the appropriate identification of the relationships between these variables and gasoline demand, while also controlling for other demographic and locational variables in a linear parametric specification. Specifically the separability[3] assumption regarding gasoline consumption and labor supply is relaxed. I also relaxed the linear assumption regarding age and income. The rest of the paper is organised as follows; theoretical consideration and econometric model is in section 2, section 3 contains the data and results of the study with discussions, while in section 4, I present the conclusion for the study and ideas for future work.

2. Theoretical consideration

The conditional demand model that is used in the empirical section for the analysis of household gasoline demand is based on the idea of conditional cost function and the associated conditional demand following the works of Pollak (1969, 1971), Browning (1983), Browning and Meghir (1991). In this set-up, one can classify all goods into two exclusive classes, a “good of interest” and a “conditioning good” where q and p denote the quantity and price of the good of interest, respectively, while h and w are the quantity and price vectors for the conditioning good, respectively. Finally, in addition to the two

[3] The general modelling of gasoline demand makes the explicit assumption that demand for gasoline is weakly separable from labor supply and therefore the derived individual or aggregate gasoline demand equation do not include labor supply as a right-hand side variable. A study by Browning and Meghir (1991) found that labor supply proxy by male and female working hours as conditioning variables is not separable from commodity demand for the UK. Brännlund and Nordström (2004) also found similar results for Sweden.

classes of goods, "demographic"[4] (z) and locational (d) variables may also affect preferences over the good of interest. In this paper, the conditioning good is restricted to labor supply[5].

If one assumes that preferences over all goods can be represented by the utility function $U(q, h, z, d)$, the conditional cost function for the good of interest can be defined as:

$$C(p,h,z,d,u) = \min_{q} p \cdot q \,\middle|\, U(q,h,z,d) = u, \tag{1}$$

where u is a given utility *level*.

The choice of the conditional cost function is due to the fact that it is easy to relate the structure of the conditional cost function to the structure of the direct utility function, but this is not the case for other preference representation. This property of the conditional cost function, that allows one to relate the cost structure to the structure of the direct utility function, makes it an appropriate channel for testing the weak separability assumption between the good of interest and the conditioning good. From the above conditional cost function, one can derive the conditional demand function as follows; the conditional compensated demand function can be derived by taking the derivative of the conditional cost function with respect to the price of the good of interest, which can be expressed as:

---

[4] Demographic variables in this study refers to household characteristics such as children under 18 years old, children aged 18 or over, number of adults in the household and age of the household head. This definition is consistent with other previous studies such as Schmalensee and Stoker (1999).

[5] The choice of labor supply as conditioning good is due to the notion that labor supply may have a significant influence on gasoline demand via the demand for leisure, and some of the activities engaged in during leisure time may have a positive impact on gasoline.

$$q^{c}(p,h,z,d;u)=\frac{\partial[C(p,h,z,d,u)]}{\partial p} \tag{2}$$

For the conditional uncompensated demand function, let $C(p, h, z, d, u) \leq y$ (y is total income for the household). Substituting this into the conditional compensated demand function gives the uncompensated demand function:

$$q=q^{uc}(p,h,z,d,y) \tag{3}$$

If one imposes the weak separability assumption on the conditioning good and the good of interest, as in Pollak (1971), the conditional uncompensated demand function, as expressed in equation (3), becomes:

$$q=f(p,z,d,y) \tag{4}$$

In this study, the good of interest is gasoline (q), while the conditioning good is that of labor supply (h), proxied by male working hours (h^{m}) and female (h^{f}) working hours. Based on equation (3), the derived conditional uncompensated gasoline demand function for household i is expressed as:

$$q_i=q(p_i,y_i;\mathrm{z}_i,d_i,h_i^{m},h_i^{f},z_{age,i},t) \tag{5}$$

As before, q_i denotes gasoline consumption, p_i is the price of gasoline, y_i is household income. Moreover, z_i is a vector of household characteristics (excluding age), d_i is a location variable, t is a time variable to capture heterogeneous effects such as changes in the car stock over the pooled years (a timed specific dummy variable), h^m and h^f is male and female working hours respectively. Finally, $z_{\text{age,i}}$ is age (years old).

3. Econometric Model

For the parametric specification of the gasoline demand function in equation (5) we use a log-linear specification. If gasoline consumption, gasoline price, and income are in logarithms, the parametric specification can be expressed as:

$$\dot{q}_i = \beta_0 + \beta_1 p_i + \beta_2 z_i + \beta_3 d_i + \beta_4 t + \beta_5 y_i + \beta_6 h_i^f + \beta_7 h_i^m + \beta_8 z_{age,i} + \varepsilon_i \quad (6)$$

Where ε_i is the error term, however, since the model in (6) is based on a sample of the population it may be the case that some of the right-hand side variables are correlated with the error term and hence the estimates will be biased. This can be corrected for in the parametric model using the Heckman's selection model if the unobserved dependent variable is due to misreporting and other factors rather than a corner solution (Heckman, 1979). Following Das et al. (2003), I extended this parametric selection model to a semiparametric selection model for gasoline demand as follows:

$$\dot{q}_i = \beta_0 + \beta_1 p_i + \beta_2 z_i + \beta_3 d_i + \beta_4 t + \sum_{l=1}^{4} m_l(\tilde{x}_{li}) + \varepsilon_i \quad (7)$$

$$q_i = I_i \cdot \dot{q}_i \quad (8)$$

Where $\dot{q}_i$ is a latent variable, I_i is a binary selection indicator variable that takes the value 1 for non-zero observation, and 0 otherwise, $\tilde{x}_1 = y$, $\tilde{x}_2 = h^f$, $\tilde{x}_3 = h^m$, $\tilde{x}_4 = d_{age}$ and $m(.)$ is an unknown smooth function that is continuously differentiable and additive in its components. If I_i and ε_i are correlated, one needs to correct for selection bias in order to estimate both the β's and the smooth function m. To correct for selection, let η denote a vector of variables that determine selection and x a vector of all covariates as specified in equation (7). Also, define, $P \equiv E(I|x,\eta)$ where P is the propensity score (PS). For the PS to correct for selection, Das et al. (2003) proposed the following assumption for a semiparametric model.

Assumption 1:

(i) $E(\varepsilon_i|\mathrm{x},\eta, I=1) = \lambda(P)$

(ii) For any random variables $l(x)$ and $b\,(P)$ $pr[l(x)+b(P)=0|I=0]=0$, this implies that $l(x)$ is constant, where $l(x)$ and $b\,(P)$ are deviations from the true values of x and P respectively.

In assumption 1, (i) means that the conditional expectation of the disturbance given selection depends only on the PS, and condition (ii) is an identification condition, which implies that $\lambda(P)$ depends only on variables in η that allow for identification of $m(.)$ up to an additive constant, similar to those in Ahn and Powell (1993). A Proof of the above two conditions can be found in Das et al. (2003). Incorporating the above assumption into the semiparametric model, the conditional expectation for the selected data is given by:

$$E\left(q_i \middle| x,\eta,I=1\right)=\beta_0+\beta_1 p_i+\beta_2 z_i+\beta_3 d_i+\beta_4 t+\sum_{l=1}^{4} m_l(\tilde{x}_{li})+\lambda(P_i). \quad (9)$$

Where I imposed the restriction that both $m(.)$ and $\lambda(.)$ are smooth functions and that equation (9) identifies $m(.)$ up to an additive constant. The vector of characteristics, z, is defined as z = [*children under 18 years old*, *children aged 18 years or over*, *number of adults*], while the location variable, *d*, is *d* = [*Stockholm*, *Gothenburg/Malmö*, *Major towns*, *Southern area*, *Major towns in Northern area*].

Following Das et al. (2003), I proposed a two-step estimation method. In the first step, one estimates $P \equiv E(I|x,\eta)$ using a semiparametric series base approach proposed in Das et al. (2003). In the second step, I regress the observed *q* on *x* and the estimated *P* from the first step with the additive restriction imposed. The reason for choosing the series estimator for the estimation is because of its ability to accommodate the additive restriction imposed on both the selection and outcome equations in the above (for details on the properties of the series estimator, see Li and Racine, 2007). It also allows for identification of the coefficients of the variables in the parametric part of the model.

Estimation of the semiparametric model

The semiparametric model presented in equation (9) is estimated using the mgcv-package in the R-statistical platform. The smooth terms (non-parametric terms) are estimated using penalised regression splines (a series regression approach). That is, the smooth functions are estimated using a chosen basis function (the basis function employed by the mgcv-package is that of the cubic Hermite polynomials, known for its computational convenience), each with an associated penalty to control the trade-off between “fidelity” to the data and “wiggliness” of the functions in order to control over (under) fitting the smooth functions. The key idea of penalised regression splines is that, in order to represent the smooth curve, it first over-fits the spline function, but to balance the over-fitting, a penalty term is subtracted from the model’s likelihood function.

The penalised regression spline approach can be outline by considering equation (9) in a compact form as follows:

$$E(q_i \mid x, \eta, I = 1) = A\theta + \sum_{l=1}^{4} m_l(\tilde{x}_{li}) + \lambda(P_i) \tag{10}$$

Where $m_l(.)$ and $\lambda(.)$ are smooth functions, $A\theta$ is the parametric part of the model with A being the model matrix for the parametric terms, θ is the vector of parameters including an intercept. The smooth functions $m_l(.)$ and $\lambda(.)$ are represented using penalised cubic regression splines so that:

$$m_1(\tilde{x}_1) = \sum_{j=1}^{q_1} \beta_j b_{1j}(\tilde{x}_1) \qquad m_2(\tilde{x}_2) = \sum_{j=1}^{q_2} \beta_{j+q_1} b_{2j}(\tilde{x}_2)$$

$$m_3(\tilde{x}_3) = \sum_{j=1}^{q3} \beta_{j+q_2} b_{3j}(\tilde{x}_3) \qquad m_4(\tilde{x}_4) = \sum_{j=1}^{q4} \beta_{j+q_3} b_{4j}(\tilde{x}_4)$$

$$\lambda(P) = \sum_{j=1}^{q_5} \beta_{j+q_4} b_{5j}(P)$$

Where b_{1j}, b_{2j}, b_{3j}, b_{4j}, b_{5j} are cubic spline basis functions for $m_1(.)$, $m_2(.)$, $m_3(.)$, $m_4(.)$, $\lambda(.)$, respectively. Furthermore, before fitting the model, one needs to specify the structure of the penalty since the underlying approach is a penalised regression spline. The penalty parameter for penalising the wiggliness of each of the smooth functions can be defined as:

$$J(m) = \int \left(\frac{\partial^2 m(\tilde{x})}{\partial \tilde{x}^2} \right)^2 d\tilde{x} \qquad\qquad J(\lambda) = \int \left(\frac{\partial^2 \lambda(P)}{\partial P^2} \right)^2 dP$$

Where the penalty on $m(.)$ or $\lambda(.)$ varies with the wiggliness of each of the functions, it is large when the function is very wiggled and small if the function is almost flat. If one defines a model matrix $\mathrm{X} = [A; \tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \tilde{x}_4, P]$ and a penalty matrix $H = [H_1, H_2, H_3, H_4, H_5]$, where $\int \left(\frac{\partial^2 m_i(\tilde{x}_i)}{\partial \tilde{x}_i^{\,2}} \right)^2 d\tilde{x}_i = \beta' H_i \beta$ for $i = 1,..4$ and $\int \left(\frac{\partial^2 \lambda(P)}{\partial P^2} \right)^2 dP = \beta' H_5 \beta$ (details on the derivation of the penalty matrix are in Wood, 2000), equation (10) can be expressed as:

$$E(q_i | x, \eta, I = 1) = f(\mu) = X\beta \tag{11}$$

where f is a smooth monotonic "link" function, and let μ and V be the mean and variance of q ,respectively. The generalised additive model (GAM) as expressed in equation (11) is fitted by minimising the negative log likelihood:

$$-l(\beta)+\frac{1}{2}(\alpha_1\beta' H_1\beta+\alpha_2\beta' H_2\beta+\alpha_3\beta' H_3\beta+\alpha_4\beta' H_4\beta+\alpha_5\beta' H_5\beta). \qquad (12)$$

Where $\alpha 's$ are smoothing parameters that control for the trade-off between a good fit and model smoothness (Wood, 2000) .The $\beta 's$ obtained are updated by the iterated re-weighted least square problem. For instance $\beta^{[k+1]}$(the superscript k denotes the estimate at the k^{th} iteration) is obtained by solving a penalised least squares problem expressed as:

$$\begin{aligned}&\min\left\{\left\|W^{[K]}(\mathrm{X}\beta-\tilde{\mathrm{q}}^{[k]})\right\|^2+\alpha_1\beta' H_1\beta+\alpha_2\beta' H_2\beta+\alpha_3\beta' H_3\beta+\alpha_4\beta' H_4\beta+\alpha_5\beta' H_5\beta\right\} \\ &s.t.\, C\beta=0.\end{aligned} \qquad (13)$$

Where $\tilde{q}^{[k]}=\mathrm{X}\beta^{[k]}+\Gamma^{[k]}(q-\mu^{[k]})$ is a pseudo data, $\Gamma^{[k]}=f'(\mu^{[k]})$ is a diagonal matrix, $W^{[\mathrm{k}]}=\left[\Gamma^{[k]}\sqrt{V(\mu^{[k]})}\right]^{-1}$ and C is a linear equality constraint matrix and it is constructed based on the "natural" spline constraint[6] (See Hastie and Tibshirani, 1990, Wahba, 1990, Green and Silverman, 1994 and Wood, 2000 for information on penalised regression spline).

---

[6] A "natural" spline constraint implies that the spline has a zero second derivative outside the interval that it is evaluated, for instance given a cubic spline function $f(x)=\sum_{i=1}^{m+2}\partial_i b_i(x)$, the "natural" spline constraint for the above cubic function is that outside the interval $\left[x_i, x_m\right]$, the spline has a zero second derivative.

Econometric concerns

The following econometric concerns crop up when one tries to apply the model outlined above on pooled survey data. The first concern is on the issue of the homogeneity assumption imposed by a pooled model and the consequences of such a restriction on the model estimates. An argument usually provided in support of choosing a pooled model is that the efficiency gain for pooling the data outweighs the cost for ignoring the possible heterogeneity in the *estimates* across sections for the different time periods and this argument is also likely to hold in this case given the small time dimension over which the data are pooled and also the fact that the effects of most of the covariates on gasoline demand are likely to be constant. For instance, the impact of location on gasoline demand is likely to be the same across the pooled years, likewise the impact of children in the households, it is therefore very likely that even if there is variability in the estimates across sections, the variations will be insignificant to generate any serious concerns. Also, to account for the aggregate time effect, I included time dummies that may account for some of the effects of car stock across the time period over which the data are pooled.

Another concern is the omission of gasoline price in the final empirical demand model that is presented in the results section *below*, and the possible bias that this may generate on the estimates. To deal with the issue of the omitted price variable, I relied on the argument put forward by Yatchew and No (2001) that price is orthogonal to both demographic and locational factors and therefore no omission bias in these estimates. The only variable that is likely to suffer from omission bias is that of income, but this bias is likely to be minimal if the covariance between income and price is close to zero. I propose to investigate the effect of the omitted price variable via a sensitivity analysis where I will include the “infrequent” price variable in the gasoline demand model to assess the price effects on the other estimates (details are under the subsection, sensitivity analysis).

Finally, another potential concern is the possible heteroskedasticity and serial correlation as a result of pooling across sections and time periods. Hence the usual formula for calculating the variance-covariance matrix and consequently the standard errors are not appropriate, which means that inferences cannot be made. In addressing this concern, I opted for the bootstrap approach as described in Wooldridge (2002) that resamples the data a number of times based on the cross sectional units to estimate the standard errors. These are robust to heteroskedasticity and serial correlation in a data set with large cross sectional units, but with few time periods.

Outline of Empirical methodology

The empirical approach can be outlined as follows; In the first step, I will test the weak separability assumption between gasoline demand and labor supply using the likelihood ratio test approach proposed by Browning and Meghir (1991) This will help in determining whether or not to condition the final empirical gasoline demand model on labor supply. In the second step, I will estimate the propensity score using the semiparametric approach for the selection equation (the first step estimation procedure). The propensity score from the first step is then included as one of the regressors in estimating the equation of interest, the gasoline demand equation (a semiparametric series estimator-penalised regression spline will be used for estimating the selection and outcome equations). The third step in the methodology is focused on testing how well the semiparametric model is able to fit the data relative to a parametric OLS model by using a nonparametric correct model specification test. In the final step, I will estimate the average derivative of the smooth functions in the gasoline demand model to formalise the estimated smooth functions and undertake sensitive analysis of the model with regards to gasoline price and the effect of imposing the weak separability assumption between gasoline demand and labor supply. Before applying the approach outlined above, it is important to consider the nature of the data under study, which is presented in the next section.

4. Data

The data set for this study is the Swedish Family Expenditure Survey (FES) for 1985, 1988 and 1992 and it is the same data set that Brännlund and Nordström (2004) used in their study. It is a micro-data set that is pooled for three time periods, which result in 11922 observations. The sample frame for each of the three surveys was the Swedish register of the total population and consisted of all persons living in non-institutional households in Sweden during the survey years. A simple random sampling approach was used to obtain the sampled families from the population register. Further, the sampled families were distributed into 26 subsamples each with a pre-specified starting date for a four-week diary-keeping period. Data were collected using personal interviews, diaries and register information; details on the sample design, organisation of the surveys, estimation and precision can be found in Lindström et al. (1989) and SCB (1994).

In the survey, households were asked to report their expenditure on non-durable goods in a four week time period, and the resulting expenses are then inflated into an annual average. However, in the case of gasoline the households reported their annual expenditures. The data set also contains information on gasoline price, price for taxi and price for public transport. Whereas prices of taxi and public transport are in quarterly frequency, gasoline price is in annual frequency.

Additionally, the data set also contains information on the household's regional location (Stockholm, Gothenburg/Malmö, Major towns, Southern areas, Major towns in northern areas, and northern rural areas. The data set also contains information on households' characteristics, such as age, number of adults in the household and age of children from under 1 year old to 17 years old, and 18 or more years old, which helps in constructing two variables to account for the effects of having children in the household, one for children under 18 years old and the second variable for children aged 18 or more years old. Finally, I have information on household total expenditures as well as working hours for males and females. Descriptive statistics of the relevant variables in the data set are presented in Appendix A.

The dependent variable in the model is gasoline expenditure (q), which is used as a proxy for gasoline consumption. For gasoline expenditure, there are 9829 non-zero observations, meaning that there are 2093 observations with zero expenditure, which is approximately 18% of the total observations on gasoline expenditure. Hence there is a need to deal with the occurrence of the zeros in the data set. The zeros in the diary recording on gasoline expenditure constitute an issue of selection rather than a corner solution due to the following reasons. Selection can be an issue either due to a non-random sample as a result of the sample design or the behaviour of the units being sampled (Wooldridge, 2002). In this study, selection is an issue not because of the survey design but rather the behaviour of the sampled households such as non-responses to some of the survey questions. But also due to the postponement of the sampled dairy keeping date to a later period due to travelling, a lot to do at work, illness of a family member and change of address. One effect of postponement of the dairy keeping period may be to underestimate expenditures on non-regularly purchased items, and overestimate expenditures made on regularly purchased items (Lindström et al., 1989). In addition, since households were asked to report their annual expenditure on gasoline, it is very likely that households with infrequent purchases may not record a positive amount, irrespective of the fact that they may actually have a positive demand for gasoline. Therefore, restricting the sample to only the positive gasoline recording may generate a situation where the expectation of the errors from the selected model conditional on the covariates and the selection indicator will not be equal to zero. Lastly, following the argument put forward by Kayser (2000), Brännlund and Nordström (2004), the fact that the decision to own a car and the decision on the amount of gasoline consumption are made jointly, therefore not accounting for this joint decision in the estimation process, may also result in bias estimates.

The above reasons suggest using a two-stage estimation approach. In this case I adopted a semiparametric version of the Heckman selection model to account for the zeros in gasoline expenditure recording. One main concern regarding the data is that of the infrequent nature of the price variable that could lead to imprecise estimates. I propose to follow the approach taken in Schmalensee and Stoker (1999) by dropping the price variable in the estimation. This is motivated by the empirical evidence in Yatchew and No (2001) that the price of gasoline appears to be orthogonal to demographic variables. I therefore expect the possible consequences from excluding the price variable on the estimates (bias) not to be severe, but will investigate this further in a sensitivity analysis by including the infrequent price to check its effects on the estimates.

5. Results

Before going into the details of the results reported in Table 2, it was important to first test for whether gasoline demand and labor supply were separable.

The separability test[7] was simply a test on whether the variation in gasoline demand could be "explained" solely by the variation in non-labor variables such as household demography variables, location variables and income. This was achieved by performing a likelihood ratio test of a restricted gasoline demand model against a non-restricted gasoline demand model. The results from the separability test, as reported in Table 1 should be interpreted as follows; in the first column and the first row, the chi-square statistic of 97.44 with a *P*-value of 0.000 meant that including male hours in the gasoline demand model without PS significantly improved the model, compared to the model without male hours, at the 1% significance.

Table 1: Separability test

Model	Male hours Test Stats.	Female hours Test Stats.	Total hours Test Stats.
OLS no PS.	97.440	7.488	113.330
	[0.000]	[0.023]	[0.000]
OLS with PS.	34.716	6.757	45.255
	[0.000]	[0.034]	[0.000]

Note: OLS no PS represents an OLS model without propensity score, and with PS denotes with propensity score. Values under male, female and total hours are the chi-square statistics for separability test for gasoline demand and labor supply (male and female hours of work) and those in square brackets are the *P*-values. The null is that labor supply does not significantly influence gasoline demand.

The general conclusion from Table 1 was that, in all cases labor supply was not separable from gasoline demand at the 5% significance

[7] For purposes of the test, I used an OLS parametric model for gasoline demand as specified in equation (5) with square terms included for age, income, mwh, fwh, and PS to capture possible non-linearity of these variables. The reason for using a parametric model rather than the semiparametric model was because it was easy to perform the test in a parametric set-up, as in the semiparametric case, one had to perform the test on a transform variable on fwh and mwh as they entered the model nonparametrically and therefore conclusions would be based on the transformed variable rather than the original.

level. The result from the likelihood ratio test implied therefore that labor supply should be included in the gasoline demand model as a covariate.

In estimating the gasoline demand model as specified in equation (5), I considered two different models, as presented in Table 2. In model 1 I estimated a parametric OLS model with the inclusion of the estimated PS, PS squared, age, age squared, h^m, h^m squared, h^f, h^f squared, and log income, log income squared. Model 2 was the semiparametric model with the estimated PS. The idea was simply to test the semiparametric model against a fully parametric model to evaluate the fit of the flexible semiparametric model relative to the parametric model. The results from each of the two models indicated that the estimated coefficient for the propensity score was significant, and therefore correcting for selection bias appeared to support the data. Testing the semiparametric specification against the parametric specification was performed by using the ´mixed` residual approach of Li and Wang (1998). The result of this test, the I_n-test[8], (derivation of the test-statistic is in Appendix B) rejected the parametric specification in favour of the semiparametric specification of model 2 even at the 1% significance level. For that reason I shall henceforth focus on the estimates from model 2.

[8] A nonparametric test for correct model specification of a parametric model (details in Li and Racine, 2007)

Location variables

The estimated parameters for the locational dummies could be interpreted in comparison to a reference location. In the estimation, the reference location was Northern areas. Households that lived in Stockholm on average demanded approximately 30% less gasoline than those living in Northern areas, while those living in Gothenburg and Malmö demanded approximately 14% less relative to the reference location. The results further indicated that households living in major towns on average demanded 13% less gasoline than the reference location, and those in Southern areas demanded approximately 9% less gasoline than those in Northern areas. Finally, households located in major towns in Northern areas, demanded approximately 7% less gasoline than households in the reference location.

These estimates were consistent with results from previous studies on the impact (direction) of rural and urban locations on gasoline demand. That is, households located in urban areas did on average consumed less gasoline than those living in rural areas. Studies by Schmalensee and Stoker (1999), Kayser (2000), Yatchew and No (2001), and Wadud et al. (2010) all found rural location to significantly increase gasoline consumption. One possible explanation for this was that, the choice set for alternative transport modes in rural areas was generally limited compared to the choice set for urban areas, and as such households in rural areas depend more on personal transport for their daily commuting demand and other activities than those in the urban areas.

Table 2: Gasoline demand estimates

Variables	Model 1 OLS	Model 2 Semiparametric
Intercept	4.578(4.300)	8.524 (0.071)
D88	-0.015(0.018)	-0.033(0.030)
D92	0.279(0.023)	0.297(0.041)
Children age<18	-0.004 (0.002)	-0.001 (0.002)
Children age18 or more	0.118(0.014)	0.113 (0.017)
Number of Adults	0.141 (0.021)	0.136 (0.020)
Stockholm	-0.237 (0.057)	-0.299 (0.103)
GB/Malmö	-0.103 (0.040)	-0.141 (0.067)
Major Towns	-0.102(0.028)	-0.125 (0.038)
Southern areas	-0.080 (0.025)	-0.093 (0.031)
Major Towns in N.A.	-0.056 (0.032)	-0.070 (0.031)
Age	0.012 (0.003)	Smooth
Age square	-0.0002(0.00004)	…………
h^f	0.001 (0.002)	Smooth
h^f square	-0.0001(0.00003)	…………
h^m	0.008 (0.003)	Smooth
h^m square	-0.0002(0.00004)	…………
Log income	0.224 (0.733	Smooth
Log income2	0.003(0.030)	…………
Propensity score	1.366(0.451)	Smooth
Propensity score2	-0.570(0.268)	…………
GCV-score	…………	0.371
R-Square (Adj.)	0.216	0.222
I_n-test	7.31 [0.000]	

Note: N.A. denotes Northern area, GCV is Generalised Cross Validation, GB is Gothenburg and values in parenthesis are bootstrap standard errors with1000 replications, while smooth denote a nonparametric variable. The null hypothesis of the I_n-*test* was that the parametric model was the correct specification against the alternative being the semiparametric specification with the *p*-values in square brackets. D88 and D92 were time dummies for the years 1988 and 1992 with 1985 as the reference year. The estimates were made using the R statistical platform, and with Wood (2007) mgcv-package for the semiparametric model.

"Demographic" variables

The estimated coefficient for the number of children in the household with age under 18 years old was negative and significant at the 5% level. This finding was consistent to other findings in the literature, such as Kayser (2000) and Wadud et al. (2010). Number of Children within the age bracket 18 and above on the other hand, significantly increased gasoline demand by 11%, while the numbers of adults in the household increased gasoline demand by 14%.

Nonparametric variables

The nonparametric variables in the model consisted of income, age of "household head", male working hours, and female working hours. The relationship between gasoline demand and income appeared to be non-linear. Gasoline demand increased continuously for each increase in log income (annual income) until an approximate annual income of SEK 442 000 (13 in log income). Thereafter further increases in annual income decreased gasoline demand. The graph as presented in Figure 1 appears to indicate a quadratic income relationship with gasoline demand with a large 95 confidence interval for high income levels depicting uncertainty of the estimates within the high income levels due to the fewer data points for this income bracket (the dash lines on the x-axis for each plot are the data points for the covariate).

The plot of gasoline demand response to age appeared to be almost flat between the ages of 20 to 40 and then a negative slope thereafter, as shown in Figure 1. This relationship was consistent with the finding in Schmalensee and Stoker (1999) and also consistent with the finding in Yatchew and No (2001), except for the age dummies between 20 to 30 years, where unlike my finding as well as Schmalensee and Stoker, gasoline demand declined sharply. It is important to note that in this study, just as in Schmalensee and Stoker, age was a continuous variable, while it was a series of dummies in Yatchew and No.

The other variables that entered the gasoline demand model in a more flexible specification were those of male and female working hours. The graph for female working hours indicated a slightly increasing phase between 10 and 38 working hours implying a positive impact of female working hours within this bracket, while working more than 38 hours tended to have a negative impact as indicated in Figure 1.

The increasing phase of the curve for female working hours and gasoline demand when working fewer hours could be explained by the availability of more leisure time, and some of this time was spent on activities that required driving and hence demand for more gasoline. But when leisure time was further reduced due to more working hours, they were likely to sacrifice some of the time spent on being on the road to activities that did not require more deriving such as being at home with the kids.

The relationship between gasoline demand and male working hours was rather complex. It increased initially up to 38 working hours, declined on average between 38 and 55 working hours. Thereafter it was rather flat with a tendency to increase. But these changes were not dramatic, as was evident from the estimated slope of the curve presented in Figure 1. The increasing impact of male working hours on gasoline demand might be explained by the fact that working fewer hours implied more free time, which could be spent on driving for pleasure, visiting friends, road trips and picking up the children from school, while the declining phase might be explained by more time spent at work that reduced the amount of time spent on all the other activities that could lead to more driving.

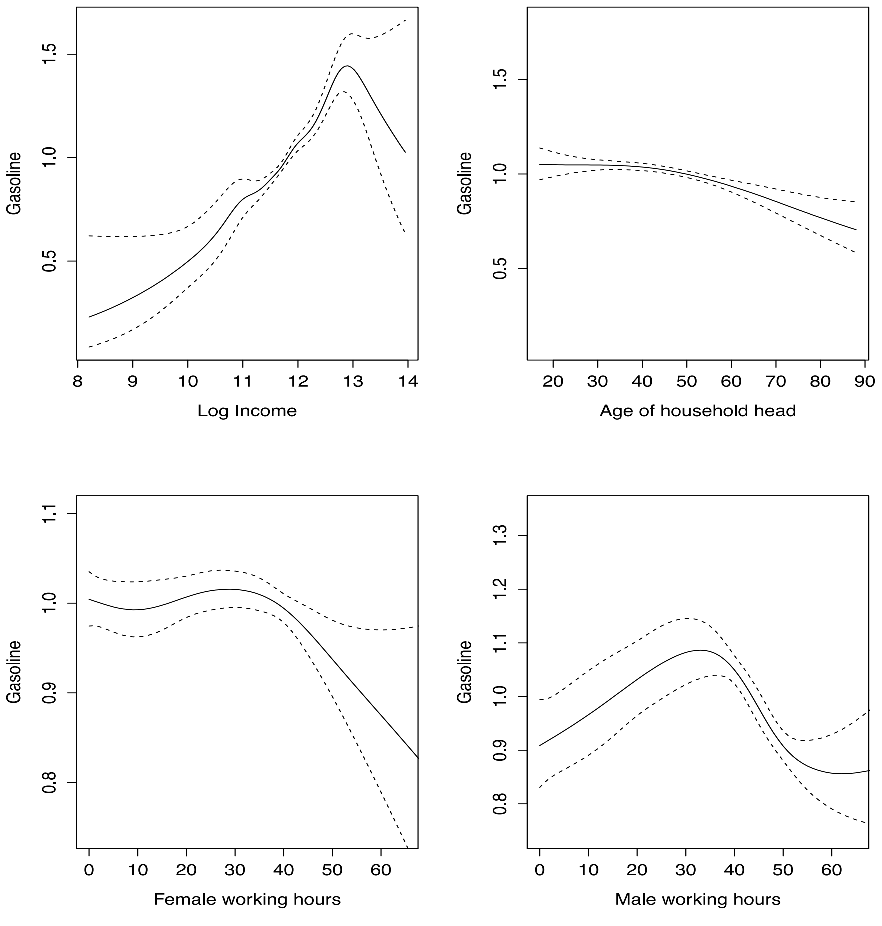


Figure 1: Fitted smooth function between gasoline demand and income, age, female and male working hours, respectively with 95% Bayesian confidence interval (dash lines are the confidence intervals), smoothing parameters were chosen by GCV.

This changing impact of male working hours on gasoline demand, as displayed in Figure 1, could not be revealed in the usual log-linear parametric model. This meant that the flexible specification became very valuable in revealing such important information that otherwise would be hidden in the parametric model.

The estimated relationships presented so far between gasoline demand, income, age of household head, h^m and h^f were purely descriptive. To convert this to specific values, the slopes (derivatives) of the various smooth curves across the whole spectrum of the curves were estimated. The estimated derivatives and their 95% confidence interval are presented in Table 3. The estimated average derivative[9] for income was approximately 0.3. The estimated derivative with respect to age was 0.011 and that of h^m was -0.014, while for h^f it was 0.008.

Table 3: Estimated derivative for the smooth functions

Variables	Average derivatives	95% Confidence Interval
Log income	0.257	[0.736 -0.232]
Age	0.011	[0.049 -0.027]
h^m	-0.014	[0.024 -0.053]
h^f	0.008	[0.047 -0.030]

[9]The derivatives were calculated as follows: $E[\partial m(.)\left|\partial y\right.] = \alpha_y$, $E[\partial m(.)\left|\partial z_{age}\right.] = \alpha_{z_{age}}$, $E[\partial m(.)\left|\partial h^m\right.] = \alpha_{h^m}$ and $E[\partial m(.)\left|\partial h^f\right.] = \alpha_{h^f}$ where h^m and h^f denoted male and female working hours, respectively.

Sensitivity analysis

As a robustness check on the influence of the non-inclusion of price on the estimates, I further ran a semiparametric model with the infrequent price variable. The result from the model with the infrequent price variable is reported in Appendix C and it indicated essentially no changes in the results with respect to the effects from the demographic and locational variables, neither on the sign nor magnitude, in comparison to the results from the semiparametric model without the price variable. This finding was consistent with the finding in Yatchew and No (2001). Furthermore, the inclusion of the price variable did not alter the functional relationship for each of the variables that entered the semiparametric model (graphs are in Appendix D). I also made a sensitivity analysis of the effect of excluding labor supply on the estimates of the included covariates by adopting Browning and Meghir (1991) approach. The results as reported in Appendix C indicated a slight decrease in the estimates for number of adults in absolute terms when labor supply was excluded from the model, compared to the model with labor supply. The estimates for all the location variables also decreased in the model without labor supply, relative to the model with labor supply, but the magnitude of the changes were larger in this case compared to the results on the number of adults. The average derivatives for income and age of the "household head" both increased slightly in the model without labor supply relative to the model with labor supply.

6. Conclusion

In this paper, I estimated household demand for gasoline using semiparametric techniques to reduce specification errors. Furthermore, the assumption that labor supply was separable from gasoline demand was relaxed in the empirical analysis. The results from the preferred model revealed that labor supply, measured by male and female working hours, significantly influenced gasoline consumption at the household level. This was also confirmed by a formal separability test. This finding was very intuitive as working more hours reduced the leisure time that could be spent on all the possible activities that could lead to more driving.

The results further revealed that the effect of labor supply on gasoline consumption differed slightly for men and women. Whereas female working hours appeared to have a positive effect on gasoline demand when working hours per week were between 10 and 38 hours, their effect on gasoline demand tended to have a negative effect when working hours increased beyond 38 hours per week. The effect of male hours on gasoline demand increased as working hours increased from 0 to 38 per week, decreased as working hour's increased beyond 38 but less than 55 hours, and then gradually began to increase as hours increased beyond 55 per week. It was important to note that there were few observations in the data with more than 60 hours per week and therefore the estimates for hours of work greater than 60 were not very reliable. The estimated slope for male working hours indicated a non-linear slope with an estimated average of -0.014, while the estimated average derivative for female working hours was 0.008.

The sensitivity analysis with respect to the non-inclusion of the gasoline price in the analysis revealed that the exclusion of price appeared not to have serious consequences on the estimates of the other parameters. However, with regard to this, there was room for further study as the price variable in this data set only had three levels of variations and was therefore not very reliable.

The policy implication of the results was that, with regard to income, households tended to react differently to income changes depending on their level of income. This suggested that any energy policy, specifically gasoline policy that ignored this aspect of the income effect would be likely to have serious consequences on the success of the policy. Irrespective of the relatively small labor supply effects on gasoline demand, gasoline policy that completely ignored them could once again have serious consequences for the success of the policy. It was also clear that in addition to the price and income effects, demography and location also had significant influence on gasoline demand responses, especially at the household level, but many of the macro-level studies completely ignored these effects and therefore policies formulated based on macro-level estimates might well overestimate or underestimate the effects of income.

There were other factors that were uncontrolled for in this study; such as vehicle type used by the household, a more frequent price variable, such as weekly, monthly or even quarterly gasoline prices and if possible data for several years in order to analyse across households and time that allowed for tracing both across section and time effects. Besides this, it would be of interest to estimate the welfare effects of different level of taxes using a household gasoline demand model with all the stated variables included, using either semiparametric or nonparametric approach. This would be a significant extension of the work by Hausman and Newey (1995, 1998), since they did not include demographic variables in their study.

Further, the approach taken in this paper assumed an additive restriction for the smooth variables, but it is possible that there was an interaction between some of the smooth variables, especially that of income and labor supply. It would be interesting to allow for this kind of interaction, in order to check how it affected the functional relations between the interaction term and gasoline demand.

References

Ahn, H. and Powell, J.L., 1993. Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics* 58, 3 – 29.

Archibald, R. and Gillingham, R., 1980. An analysis of short-run consumer demand for gasoline using household survey data. *The Review of Economics and Statistics* 62 (4), 622–628.

Archibald, R. and Gillingham, R., 1981. A decomposition of the price and income elasticities of the consumer demand for gasoline. *Southern Economic Journal* 47, 1021–1031.

Browning, M.,1983. Necessary and sufficient conditions for conditional cost functions. *Econometrica*, 51, 851–856.

Browning, M. and Meghir, C.,1991. The effects of male and female labor supply on commodity demands. *Econometrica* 59, 925–951.

Dahl, C.A., 2012. Measuring global gasoline and diesel price and income elasticities. *Energy Policy* 41, 2–13.

Das, D., Newey, W.K. and Francis, V., 2003. Nonparametric estimation of sample selection Models. *Review of Economics Studies* 70, 33–58.

Green, P.J. and Silverman, B.W., 1994. *Nonparametric Regression and Generalized Linear Models*. London: Chapman and Hall.

Greening, L.A., Jeng, H.T., Formby, J.P. and Cheng, D.C., 1995. Use of region, life-cycle and role variable in the short-run estimation of the demand for gasoline and miles travelled. *Applied Economics* 27, 643–656.

Hastie, T.J. and Tibshirani, R.J.,1990. *Generalized Additive Models.* London: Chapman and Hall.

Hausman, J. and Newey, W.K., 1995. Nonparametric estimation of exact consumer surplus and deadweight loss. *Econometrica* 63 (6), 1445–1476.

Hausman, J. and Newey, W.K., 1998. *Nonparametric estimation of exact consumer surplus and deadweight loss*. Econometrics and economic theory in the 20th century. Cambridge: Cambridge University Press, 111–146.

Heckman, J., 1979. Sample selection bias as a specification error. *Econometrica* 47, 153–161.

Kayser, H.A., 2000. Gasoline demand and car choice: estimating gasoline demand using household information. *Energy Economics* 22, 331–348.

Li, Q. and Wang, S., 1998. A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics* 87, 145–165.

Li, Q. and Racine, J.S., 2007. *Nonparametric econometrics: theory and Practice*, Princeton University Press.

Lindström, H.L., Lindkvist, H., and Näsholm, H., 1989. Design and quality of the Swedish family expenditure survey. Statistics Sweden, R&D report 1989:7.

Nicol, C.J., 2003. Elasticities of demand for gasoline in Canada and the United States. *Energy Economics* 25, 201–214.

Pollak, R.A., 1969. Conditional demand functions and consumption theory. *Quarterly Journal of Economics*, 83, 70–78.

Pollak, R.A., 1971. Conditional demand functions and implications of separability. *Southern Economic Journal*, 37, 423–433.

Brännlund,R. and Nordström, J., 2004. Carbon tax simulations using a household demand model. *European Economic Review* 46, 211–233.

SCB. 1994. The Family Expenditure Survey1992 - Final Report, Statistics Sweden.

Schmalensee, R. and Stoker, T.M., 1999. Household gasoline demand in the United States. *Econometrica* 67 (3), 645–662.

Wadud, Z., Graham, D.J. and Noland, R.B., 2007. Modelling fuel demand for different socioeconomic groups. Proceedings of the 86th Annual Meeting of the Transportation Research Board, Washington DC.

Wadud, Z., Graham, D.J. and Noland, R.B., 2010. A semiparametric model of household gasoline demand. *Energy Economics* 32, 93–101.

Wahba, G., 1990. Spline models for observational data. Regl Conf.Ser.APPL.Math., 59.

West, S.E. and Williams III, R.C., 2004. Estimates from a consumer demand system: implications for the incidence of environmental taxes. *Journal of Environmental Economics and Management* 47 (3), 535–558.

Wood, S.N., 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society*. Series B 62(2), 413–428.

Wood, S.N., 2007 [online]. The mgcv package, available at:

(hhtp:cran.r-project.org/doc/packages/mgcv.pdf).

Wooldridge, J.M., 2002. *Econometric Analysis of cross section and panel data*. Cambridge, MA, MIT Press.

Yatchew, A. and No, J.A., 2001. Household gasoline demand in Canada. *Econometrica* 69 (6), 1697–1709.

Appendix:

Appendix A:

Descriptive statistics of variables used for the model on the selected sample

Variables	Mean	SD
Age	44.373	13.274
Stockholm	0.141	0.348
GB/Malmö	0.127	0.333
Major Towns	0.347	0.476
Southern areas	0.231	0.422
Major Towns in N.A.	0.083	0.275
Gasoline expenditure	7671.311	6037.23
Female working hours	25.078	16.211
Male working hours	33.546	17.008
Adult	1.869	0.393
Children age 18 or more	0.186	0.463
Children age less than 18	3.751	3.503
Log income	11.811	0.459
Total expenditure	149211.5	70664.73
Price of gasoline	88.201	2.903
Propensity score	0.858	0.137
N	9829	

Note: Both gasoline and total expenditures are inflated to yearly levels and are in local currency and SD denotes standard deviation.

Appendix B:

The I_n-*test*, which is a nonparametric consistent specification test for a parametric model, the null hypothesis is that a parametric model is correctly specified and stated below;

$$H_0 : \Pr[E(g_i \mid x_i, s_i) = \theta(x_i, s_i; \beta, \gamma)] = 1 \qquad (b1)$$

$$H_1 : \Pr[E(g_i \mid x_i, s_i) = \theta(x_i, s_i; \beta, \gamma)] < 1 \qquad (b2)$$

Where $\theta(.)$ is an unknown function and β and γ are vectors of unknown parameters, x_i and s_i are vectors of random variables, and to test the parametric specification against the semiparametric specification, the null hypothesis as in equation (*b1*) implies that:

$$\theta(x_i, s_i; \beta, \gamma) = x_i'\beta + s_i'\gamma + \varepsilon_{1,i}$$

The alternative hypothesis of a semiparametric specification implies that:

$$\theta\left(x_i, s_i; \beta, \gamma\right) = x_i'\beta + m(s) + \varepsilon_{2,i} < 1$$

Where $m(.)$ is an unknown smooth function that is twice differentiable, the test statistic for the parametric specification against the semiparametric specification using Li and Wang (1998) ´mixed` residual approach can be expressed as:

$$I_n = n^{-1}\sum_{i=1}^{n} \hat{u}_i \hat{E}_{-i}(u_i \mid x_i, s_i)\hat{f}_{-i}(x_i, s_i)$$

$$\hat{u} = g_i - x_i'\hat{\beta} - s_i'\hat{\gamma}$$

where $\hat{\beta}$ is the estimated β from the semiparametric model, while $\hat{\gamma}$ is estimated from the parametric model, $f(.)$ is the joint PDF, while $\hat{E}_{-i}(u_i \mid x_i, s_i)\hat{f}_{-i}(x_i, s_i)$ is a leave-one-out kernel estimator. In this paper $x = (p_i, \mathbf{z}_i, d_i, t_i)$ and $s = (h_i^m, h_i^f, z_{age,i}, y_i)$.

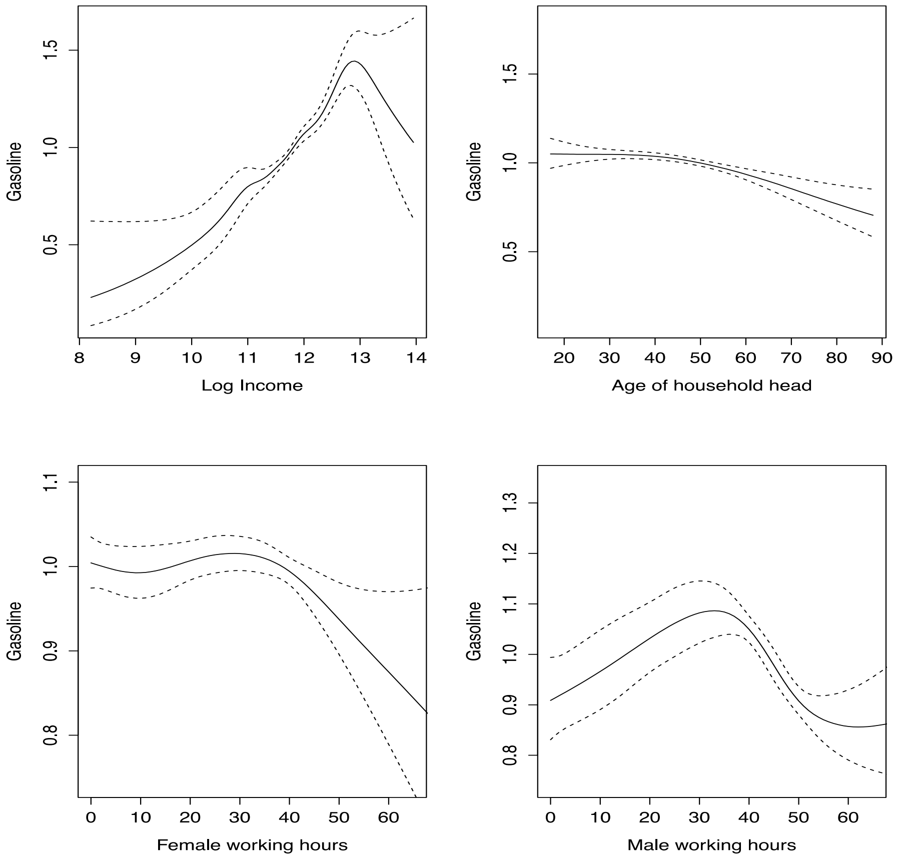
Appendix C:

Robustness check for the effect of price and labor supply on the semiparametric estimates

Variables	Semiparametric Model A	Semiparametric Model B	Semiparametric Model C
Intercept	0.466 (0.006)	8.524 (0.071)	8.478 (0.004)
D88	0.064(0.030)	-0.033(0.030)	-0.003(0.033)
D92	-0.334(0.044)	0.297(0.041)	0.297(0.035)
Children age<18	-0.001 (0.002)	-0.001 (0.002)	-0.003 (0.002)
Children age 18 or more	0.113 (0.015)	0.113 (0.017)	0.113 (0.018)
Number of Adults	0.136(0.021)	0.136 (0.020)	0.117 (0.082)
Stockholm	-0.299 (0.115)	-0.299 (0.103)	-0.140 (0.130)
GB/Malmö	-0.141 (0.071)	-0.141 (0.067)	-0.044 (0.079)
Major Towns	-0.125 (0.040)	-0.125 (0.038)	-0.077 (0.043)
Southern areas	-0.093(0.031)	-0.093 (0.031)	-0.067 (0.031)
Major Towns in N.A.	-0.070 (0.036)	-0.070 (0.031)	-0.042 (0.038)
Log price	1.885(0.015)	……………..	………….
Age	Smooth	Smooth	Smooth
h^f	Smooth	Smooth	………….
h^m	Smooth	Smooth	………….
Log income	Smooth	Smooth	Smooth
Propensity score	Smooth	Smooth	Smooth
GCV-score	0.371	0.371	0.373
R-Square (Adj.)	0.222	0.222	0.215
Average derivative (Income)	…………….	0.257	0.293
		[0.736 -0.232]	[0.792 -0.204]
Average derivative (Age)	…………….	0.011	0.014
		[0.049 -0.027]	[0.057 -0.029]

Note: N.A. denotes Northern area, GB is Gothenburg and values in parenthesis are bootstrap standard errors with 1000 replications, which are robust to heteroskadasticity and serial correlation, values in square brackets are the 95% Bayesian confidence band, while smooth denotes a nonparametric variable. Model A contains all the variables as in model 2 in addition to price of gasoline, while model C contains all the variables as in model 2, except the labor supply variables. Model B is the same as model 2 in Table 2.

***Appendix D*:**



Fitted smooth functions between gasoline demand and income, age, female and male working hours (dash lines are the 95% Bayesian confidence intervals), smoothing parameters are chosen by GCV

IV

Cooking fuel preferences among Ghanaian Households: an empirical analysis

Amin Karimu
Department of Economics
Umeå School of Business and Economics
Umeå University

Abstract

This paper investigated the key factors influencing the choice of cooking fuels in Ghana. Results from the study indicated that education, income, urban location and access to infrastructure were the key factors influencing household's choice of the main cooking fuels (fuelwood, charcoal and liquefied petroleum gas). The study also found that, in addition to household demographics and urbanization, the supply (availability) of the fuels influenced household choice for the various fuels. Increase in household income was likely to increase the probability of choosing modern fuel (liquefied petroleum gas and electricity) relative to solid (crop residue and fuelwood) and transition fuel (kerosene and charcoal). I therefore proposed that poverty reduction policies, provision of education and modern infrastructure, as well as provision of reliable supply of modern fuels should be part of the policy framework in promoting the use of modern fuels in Ghana.

Keywords: Choice Probability, Energy policy, Multinomial probit.
JEL Classification: D12; Q41; O13

1. Introduction

In most developing countries, fuelwood is the major energy source for the household (mainly for cooking), irrespective of the health implications that this source of energy can potentially involve, especially when used indoors. It is estimated that over 2.5 billion people in the developing world depend on biomass as their primary energy source for cooking (IEA, 2006). Air pollution is increasingly becoming a major contributing factor for poor health in the world, especially respiratory diseases, of which "dirty fuel"[1] is one of the major contributing factors. For instance a study by World Health Organization (WHO, 2009) indicates that the burden of diseases attributable to indoor smoke from solid fuels for developing countries is about 1.94 million premature deaths per year. The health consequences of using dirty fuels in homes cannot be overemphasized as it contributes massively to indoor air pollution, which has both direct and indirect health consequences, which women and children are the most exposed in society.

Besides the health concerns from the high usage of "dirty fuels", especially fuelwood, there are economic consequences such as loss of productivity, either due to poor health as a result of polluted air, or time spent in gathering fuelwood at the expense of working or studying. The loss of economic opportunities via the use of fuelwood

[1] Dirty fuel in this paper refers to biomass fuels, especially cow dung, crop residue and fuelwood, Charcoal on the other hand is in this paper classified as a transition (partial) fuel in the sense that it is not as dirty as fuelwood, but not as clean as liquefied petroleum gas (LPG). Kerosene is also classified as a transition fuel in this paper. Modern fuels or "clean fuel" refers to LPG and electricity.

falls heavily once again on women and children, as they are responsible for gathering fuelwood for the household in most developing countries. Biomass collection is also one of the factors that contribute to deforestation in developing countries, especially near cities and major roads (Heltberg, 2001).

The proportion of households in developing countries using biomass energy (especially fuelwood) is very high compared to rich-industrialized countries. According to Byer (1987) and Leach (1987) biomass accounts for approximately 60-95% of the total energy use in the poorest developing countries, whereas it only accounts for 25-60% for middle income countries, and less than 5% for rich-industrialized countries. For instance, in some urban cities such as Ouagadougou, 70% of the households use fuelwood as the main cooking fuel (Ouedraogo, 2006) and it is similar in the case of Ghana.

Switching to modern fuels therefore provides many potential benefits such as less time required for cooking and cleaning pots. It also increases the productivity of the poor as it allows them to redirect labor and land resources from fuelwood collection and production to activities that generate income (Heltberg, 2004). Switching into modern fuels also improves the welfare of women by providing them with the opportunity to engage in income-earning activities as a consequence of the efficiency and reduced time required for cooking.

Despite the disadvantages outlined above in the use of biomass fuels, such as fuelwood, it is still the major cooking fuel in Ghana. In 1990 approximately 69% of Ghanaian households used fuelwood as the main cooking fuel, and this figure decreased to 57.8% in 2005 (Ghana statistical service report, 2008). The reduction in 2005 indicates a remarkable progress as a result of the efforts made by the government of Ghana with the support of the United Nations Development Program (UNDP) to promote the use of modern fuels in Ghanaian households, especially liquefied petroleum gas (LPG). Various policies have been undertaken, including the national LPG promotion Program[2] and the West African Gas Pipeline (WAGP) project to aid the supply of LPG from Nigeria. Despite the efforts made by the Ghanaian government over the years, the percentage of households using fuelwood in Ghana is still very high. Within the same period (1990-2005), LPG usage increased from 0.8% in 1990 to 6.4% in 2005, and that of electricity also increased from 0.5% in 1990 to 1.1% in 2005. Irrespective of the progress made over the years to influence households to switch to modern fuels, fuelwood and charcoal are still the preferred household fuel choices for cooking in Ghana. It is therefore important to understand the main factors influencing household preferences regarding cooking fuels in order to develop appropriate policies to aid the penetration of "clean" fuels in Ghana as principal cooking fuels.

In this study, fuels are classified into three groups; modern, transition and solid fuels (traditional fuel). This classification is based on the so-called energy ladder hypothesis. The energy ladder hypothesis states that at a low level of income, households tend to consume fuels that is at the bottom of the ladder and regarded as "dirty" such as biomass fuel. As income level rises, households tend

[2] This program includes expanding the capacity of the Tema oil refinery in the production of LPG to increase domestic supply, instituting the uniform petroleum price fund (UPPF) that uses sales from petrol to cross-subsidize LPG and providing financial incentives for LPG sales occurring in places more than 200km from Tema refinery.

to move up the ladder by replacing biomass fuels with “transition” fuels such as kerosene and further to “modern” fuels such as LPG and electricity as income rises still further. Given this classification it will be possible to investigate the energy ladder hypothesis and its relevance in the case of Ghana.

In the literature, high cost of equipment and the high price of modern fuels, among other factors, are cited as the main constraints to the adoption of modern fuels. Furthermore, Leach (1987) states that income, cost of appliances, relative fuel prices and the availability of commercial fuels are the most important variables influencing household fuel preferences in South Asia. Soussan (1988) found that, both multiple fuels and fuel switching were common in poor households due to specific budgeting strategy. Reviewing a larger number of energy surveys, Leach and Gowan (1987) found that income, household size, climate, cultural factors and cost of appliances were the key demand-side variables influencing fuel choice. In cases of insecure energy supplies, fuel security rather than fuel switching dominates in the household energy plan (O’Keefe and Munslow, 1989). Other works that based their analysis on the energy ladder model include, Hosier and Dowd (1987), Reddy & Reddy (1994), Barnes et al. (2002), Heltberg (2004), Gupta and Köhlin (2006) and Ouedraogo, 2006. Contrary to the energy ladder model, Masera et al. (2000) found that in rural Mexico, fuel switching is actually a step toward “multiple fuel cooking” or “fuel stacking” for both fuelwood and LPG.

The aim of this paper is to determine the key factors that induce the choice between modern, solid and transition fuels, and to investigate the energy ladder hypothesis. Given the benefits of switching to modern fuels, and the challenges that high dependence on the use of biomass fuel poses on poverty alleviation Program such as the United Nations millennium development Program, it is imperative to have a clear understanding of the key variables that influence household decisions regarding the choice of cooking fuel. This will help in the designing of the appropriate policies towards efficient and sustainable cooking energy consumption. To reach the objectives I will adopt a multinomial probit regression (MNP) approach to try to answer the question relating to the factors that determine the choice of a particular group of fuels (modern, solid, and transition). I will also decompose the groups into their specific fuels and investigate the factors influencing the probability of choosing each of the three main cooking fuels in Ghana (fuelwood, LPG and charcoal).

In the literature, to the best of my knowledge, the only published work on Ghana in the area of household cooking fuel is that of Heltberg (2004) and Akpalu et al.(2011). Heltberg (2004) studied eight developing countries (Ghana as one of the countries). In the paper, Heltberg used the 1998/99 survey data for each of the countries. I argue that a lot has happened since then, especially in the area of energy policy aimed at increasing LPG penetration in the domestic fuel mix from the 0.8% in 1989 to 50% by 2020. Therefore, by using new data, new light will be shed on the possible progress made, and we can also assess the impact of the availability/non-availability of the fuels on the fuel-choice process. In addition, the 2004/05 survey is more extensive in terms of coverage (increase in the number of households and additional variables such as availability of the fuels) than the 1998/99 survey, and will contribute to the literature on Ghana as the factors influencing choice of fuels are context-specific. Akpalu et al. on the other hand studied the extent to which preference matter regarding four cooking fuels (fuelwood, charcoal, kerosene and LPG), which is a different focus in comparison to that of this study. Besides, they also used the 1998/99 survey data that did not capture the second phase of the LPG promotion program (Rural LPG Challenge program) that was launched in 2004.The rest of the paper is organised as follows; section 2 contains the theoretical considerations and econometric model, section 3 presents the data. The results of the study are presented in section 4, in section 5, I present the conclusion of the study and ideas for future work.

2. The model

In this section, I will outline the theoretical model for household fuel demand and consequently the indirect utility function that will be used in the empirical section. The theoretical model for household demand for cooking fuels can be derived from the household utility maximisation principle. Assume that household utility depends on food consumption (C) and on the consumption of other goods and services (OG). The utility function can then be expressed as;

$$U = u(C, OG) \tag{1}$$

Further, assume that food consumption is a function of cooking fuel (F) and groceries (G), conditional on the cooking technology:

$$C = c(F_j, G),\ j = type\,/\,alternatives\ of\ cooking\ fuel \tag{2}$$

Substituting equation (2) into equation (1) results in a household utility function that has cooking fuel as one of its arguments via food consumption:

$$U = u[c(F_j, G), OG] \tag{3}$$

The utility function, as expressed in equation (3), permits cooking fuel to enter the household utility function indirectly via food consumption. The optimization problem for the household is therefore formulated as follows;

$$\begin{aligned} &\max_{F,G,OG} u[c(F_j, G), OG] \\ &st\ \ P_{F,j}F_j + P_G G + OG \le Y \end{aligned} \tag{4}$$

where $P_{F,j}$ and P_G correspond to the price for cooking fuel and price of groceries, respectively and Y is the household income. The price of other goods and services is normalized to 1.The solution to the above maximization problem gives the household demand for cooking fuel expressed as:

$$Q_j = q(P_{F,j}, P_G, Y). \tag{5}$$

Since the demand function in equation (5) is homogeneous of degree 0 in prices and income, one can express fuel price as relative to the price of groceries by dividing $P_{F,j}$ by P_G. Incorporating the relative price modification, equation (5) becomes:

$$Q_j = q(P^r_{F,j}, Y), \; where \; P^r_{F,j} = \frac{P_{F,j}}{P_G}. \tag{6}$$

The demand equation, as expressed in equation (6), is further augmented with a vector of household characteristics, Z , in addition to availability of the fuels A. The variables A and Z act as demand shifters. The augmented demand equation for fuel j can then be expressed as:

$$Q_j = q(P^r_{F,j}, Y; A, Z) \tag{7}$$

From the fuel demand function, the indirect utility function becomes:

$$V_j = v(Q_j) = v(P^r_{F,\mathrm{j}}, Y; A, Z) \tag{8}$$

3. Econometric model

Suppose a representative household i faces j alternatives of cooking fuels (where $j = 1, 2, 3$), and that the indirect utility derived from each of the j alternatives is defined by V_{ij} . The indirect utility function is broken down into an observed part, $x_i'\beta_j$ and an unobserved part ε_{ij} , where x_i is a vector of all the variables in equation (8). The indirect utility for alternative j for household i can then be expressed as:

$$V_{ij} = x_i'\beta_j + \varepsilon_{ij}. \tag{9}$$

The unobserved part, ε_{ij} , is assumed to be jointly normally distributed with $\varepsilon \sim N[0, \Sigma]$.

The probability that household i chooses the first alternative is now:

$$\begin{aligned} P_{i1} &= pr(\varepsilon_{i2} - \varepsilon_{i1} < x_i'\beta_1 - x_i'\beta_2 \;\; and \;\; \varepsilon_{i3} - \varepsilon_{i1} < x_i'\beta_1 - x_i'\beta_3) \\ &= pr[\dot{\varepsilon}_{i,21} < x_i'(\beta_1 - \beta_2) \;\; and \;\; \dot{\varepsilon}_{i,31} < x_i'(\beta_1 - \beta_2)] \end{aligned} \tag{10}$$

where $\dot{\varepsilon}_{i,21} = \varepsilon_{i2} - \varepsilon_{i1}$ and $\dot{\varepsilon}_{i,31} = \varepsilon_{i3} - \varepsilon_{i1}$. Similar expressions can be obtained for P_{i2} and P_{i3}. It is assumed that ε_{ij} has a joint normal density function defined as $f(\varepsilon_{ij}) = f(\varepsilon_{i1,}\varepsilon_{i2}, \varepsilon_{i3})$, and let y_{ij} denote a discrete choice outcome variable that takes a value 1 if household i chooses fuel j and 0 otherwise. The cumulative probability for the choice of the first alternative fuel by household i can now be expressed as:

$$P_{i1} = pr[y_i = 1] = \int_{-\infty}^{\dot{V}_{i,12}} \int_{-\infty}^{\dot{V}_{i,13}} f(\dot{\varepsilon}_{i,21}, \dot{\varepsilon}_{i,31}) d\dot{\varepsilon}_{i,21} d\dot{\varepsilon}_{i,31.} \; . \tag{11}$$

Where $\dot{V}_{i,12} = x_i'(\beta_1 - \beta_2)$ and $\dot{V}_{i,13} = x_i'(\beta_1 - \beta_3)$, the expression in equation (11) is specific to the first fuel. In a more general case, the choice probability for household i choosing alternative j is given by $P_{ij} = pr[y_i = j] = m_j(x_i'\beta_j)$. Where $m_j(x_i'\beta_j)$ takes a similar expression as in equation (11). The log likelihood function for a sample of N independent households with J alternatives can then be expressed as:

$$\ell = Ln\, L = \sum_{i=1}^{N} \sum_{j=1}^{J} y_{ij} Ln(\hat{P}_{ij}) \tag{12}$$

where $\hat{P}_{ij}$ is estimated via a similar expression as in equation (11) using simulation methods and substituted into the log likelihood function, which is then maximised to obtain the parametric estimates for the $\beta's$.

4. Data

The data set for this paper is the fifth round of the Ghana living standards survey (GLSS 5, 2005/06) conducted in the year 2005/06. The sample consists of 8687 households of which 8262 contain information regarding household energy use. In this paper the analysis is based on the 8262 households as the focus of the paper is on household choice of cooking fuel. The sample frame for the GLSS 5 was defined as the population living within private households in Ghana and was divided into two units, the primary and secondary sampling units. The primary sampling unit was defined as the census enumerated areas (EAs), and the household within each EA constituted the secondary sampling unit. The EAs were first stratified into the ten administrative regions in Ghana, based on proportional allocations using the population in each of the regions as the basis for the allocation. Furthermore, each EA in each region was further subdivided according to rural and urban area of location. In order to achieve a reliable and comprehensive coverage, the Ghana statistical service (GSS) adopted a two-stage stratified random sampling design, where 550 EAs were considered at the first stage of sampling. In the second stage of sampling, 15 households per EA were considered. Combining both stages of sampling resulted in an overall sample size of 8700 households nationwide. In the end, however, 8687 households were successfully interviewed representing a 99.85 % response rate.

Based on the survey design outlined above, in-depth data was collected on the following key variables; household income, consumption, expenditure, education, energy use for cooking, demographic characteristics, and type of housing. The data set therefore contains information on household living characteristics such as education, employment, main fuel for lighting, household location, income, and availability of the various cooking fuels. There is also information on whether the "household head" is a female or a male, age of the household head, size of the household. The education variable has three levels; basic, secondary and tertiary (university education). In addition, there is information on whether the household is located in a rural or urban area and the main source of lighting for the household. The latter can serve as a proxy for the level of access to modern infrastructure, although it may be correlated with household income and fuel expenditure.

The data set also contains information on the main fuel sources used by each of the households. This cover fuels such as crop residue, fuelwood, charcoal, kerosene, liquefied petroleum gas (LPG), and electricity. For each of the fuels, the information in the data set only refers to choice of fuel, not the quantity. Therefore, our analysis is restricted to discrete choices and probabilities related to the choice of these fuels.

The GLSS data set is the main source of information for energy sector reforms in Ghana due to the information it contains. It allows for the identification of the major fuel use among households, the identification of major sources of fuel for the poor and rich households, as well as a comparison across the whole country in order to assess how well the country is doing in its energy reforms. A drawback of the GLSS data is that they do not contain information regarding energy prices, which limits the analysis. Other purposes of energy use in the household are not captured either, and it does not allow for analysis related to fuel stacking (combining different fuels to satisfy household cooking needs), which is a common feature especially among urban households in developing countries. A list of the variables in the data set and their definitions are provided in Table A in the appendix.

In the sample data, fuelwood is the most used cooking fuel in Ghana (57.6%), followed by charcoal (30.0%), LPG (9.3%), crop residue (2.2%), kerosene (0.6%), and electricity (0.3%). The details are presented in Table 1. Irrespective of the various policies implemented over the years to promote a switch from fuelwood use into modern fuels, fuelwood still remains the dominant choice of fuel for cooking. There are many factors that could explain the slow response to government policy targeting switching from fuelwood into modern fuels. According to the energy ladder hypothesis, a major factor for moving up the ladder is that of income and therefore low levels of income could be one of the reasons for fuelwood still being the dominant cooking fuel in Ghana. Other factors include luck of proper education on the health impact and the consequent opportunity cost associated with poor health in the use of fuelwood (especially indoor), the relatively high unit cost of modern fuels and cooking stoves appropriate for the use of modern fuels, and the supply of modern fuels.

Table 1: Main cooking energy sources and their percentages

Main Energy Source	Frequency	Percentage
Crop residue/sawdust	185	2.24
Fuelwood	4,762	57.64
Charcoal	2,477	29.98
Kerosene	49	0.59
LPG (Gas)	766	9.27
Electricity	23	0.28
Total	8,262	100.00

5. Results

The empirical strategy for the study was as follows: I first ran a MNP model to determine the factors that influenced household choice for each of the three groups of fuels; modern, solid and transition fuel. Secondly, I broke down the group of fuels into their respective constituent fuels and analysed the choice process for each of the three main cooking fuels in Ghana (fuelwood, LPG and charcoal) in order to obtain a clear picture of the choice process for these specific fuels.

Factors influencing choice for modern, solid and transition fuels among Ghanaian Households

I applied a MNP regression model to determine the factors influencing the choice for modern, solid and transition fuels, controlling for fuel supply (availability of the fuels). The choice of MNP model was due to its ability to relax the IIA (Independence of Irrelevant Alternatives) property. The IIA property states that the relative probabilities of choosing between two alternatives is unaffected by the presence of additional alternatives. Given the data set it was very unlikely that this property would hold. For instance, it was very likely that presence of modern fuel would have an impact on the probability of choosing between solid and transition fuel.

The estimated marginal effects are presented in Table 2. Modern fuel in this paper referred to LPG and electricity, solid fuels on the other hand comprised crop residue and fuelwood, whereas transition fuel referred to kerosene and charcoal[3]. In Table 2, I only reported the marginal effects, as the estimated $\beta's$ were difficult to interpret directly (the $\beta's$ are reported in Table B in the Appendix). The results from Table 2 indicated that all the marginal effects were significant, except age of household head, availability of kerosene and charcoal that were not significant in influencing the choice probability for modern fuel at the 5% significance level. Whereas for solid fuel, all the marginal effects were significant except the availability of LPG, while for transition fuel, the results indicated that the availability of kerosene, secondary and tertiary education were the variables that did not significantly influence the choice probabilities. The marginal effects for an explanatory variable could be interpreted as an increase (if sign was positive) or a decrease (if sign was negative) in the adoption probability for a given fuel, for example, the estimated marginal effect for household size was -0.6 for modern fuel. The interpretation of the value (-0.6) is as follows; increases in household size, decreased the adoption probability for modern fuel by 0.6%, implying that a large household was less likely to adopt modern fuel as a cooking fuel in relation to a small household.

[3] This classification is based on the energy ladder hypothesis that considers kerosene and charcoal as transition fuels, while crop residue and fuelwood are at the bottom of the ladder. LPG and electricity are at the top of the ladder.

The results indicated that income was a significant factor in determining the probability of choosing modern, transition and solid fuels. For instance, income increased the adoption probability for modern fuel by 2.6%, while it decreased the adoption probability for solid and transition fuels by 1.4% and 1.2%, respectively. The implication of this was that, the higher the household income, the more likely it was that it would choose modern fuel in relation to both solid and transition fuels. The choice of modern fuel as the main cooking fuel therefore displayed characteristics of a normal good while that of solid and transition fuel displayed those of an inferior good. The negative income effect for both solid and transition fuel and the positive effect for modern fuel also appeared to validate the energy ladder hypothesis, since at higher income levels, households tended to increase their probability in choosing modern fuels as postulated by the energy ladder hypothesis.

The results also indicated that household characteristics such as household size, male head and age had significant influence on the choice probability for modern, transition and solid fuels. Whereas household size and male head had a negative effect on the choice probabilities for both modern and transition fuels, it had a positive effect in the choice probability for solid fuel. Age of the household head only had significant impact on the probability of choosing both solid and transition fuel, it was positive for solid fuel and negative for transition fuel. In the case of modern fuel, both household size and male head decreased the adoption probability by 0.6% and 1.7%, while they increased the adoption probability for solid fuel by 1.4% and 4.8%, respectively.

The results also showed that education was a significant determinant on the choice of fuel. All the three levels of education had a significant positive effect on the choice probability of modern fuel. The estimated marginal effects on basic, secondary and tertiary levels of education indicated an increase in the choice probability for modern fuel by 5.2%, 11.9% and 15%, respectively. Each of the three levels of education on the other hand had a negative effect on the choice of solid fuel. A household head with a basic level of education decreased the adoption probability for solid fuel by 7.6, while with a tertiary level of education, the adoption probability for solid fuel decreased by 17.8%. The implication of this was that when the head of the household head was educated, the likelihood of choosing modern fuel increased. In the case of transition fuel, only basic education had a significant (positive) effect on the choice probability, implying that a higher level of education beyond the basic level had no effect on the adoption probability for transition fuel. In general, education appeared to increase the probability of choosing both modern and transition fuels, but decreased the probability of choosing solid fuel. One possible explanation for this was that, education tended to inform the households about the opportunity cost associated with using solid fuels such as the health cost, storage cost, and cost of time. When the opportunity cost was taken into consideration, the solid fuel was likely to be an expensive alternative compared to modern or transition fuel.

The results as presented in Table 2 also revealed that the availability of the various cooking fuels significantly influenced the choice of cooking fuel. Specifically, both the availability of LPG and fuelwood had a significant impact on the choice of modern fuel.

Whereas the availability of LPG increased the choice probability for LPG, the availability of fuelwood on the other hand decreased the probability of choosing LPG. In the case of solid fuel, the availability of kerosene, charcoal and fuelwood had a significant positive, negative and positive effect on the choice probability, respectively.

Table 2: Marginal effects for choice of modern, solid and transition cooking fuels in Ghana.

	Modern Fuels *Marginal Effect*	Solid Fuels *Marginal Effect*	Transition Fuels *Marginal Effect*
hhsize	-0.6***	1.4***	-0.7***
	(0.000)	(0.000)	(0.001)
Male head	-1.7**	4.8***	-3.1***
	(0.005)	(0.000)	(0.001)
Age	-0.0	0.2***	-0.2***
	(0.756)	(0.000)	(0.000)
Log income	2.6***	-1.4***	-1.2***
	(0.000)	(0.000)	(0.001)
Basic educ.	5.2***	-7.6***	2.4*
	(0.000)	(0.000)	(0.017)
Secondary educ.	11.9***	-14.4***	2.5
	(0.000)	(0.000)	(0.081)
Tertiary educ.	15.0***	-17.8***	2.8
	(0.000)	(0.000)	(0.221)
Av. of LPG	15.7***	-4.7	-9.9***
	(0.000)	(0.134)	(0.001)
Av. of Kerosene	-4.2	6.7**	-2.5
	(0.050)	(0.014)	(0.421)
Av. of Charcoal	-1.9	-25.6***	27.4***
	(0.385)	(0.000)	(0.000)
Av. of Fuelwood	-3.4*	18.0***	14.6***
	(0.016)	(0.000)	(0.000)
Urban	5.7***	-28.0***	22.3***
	(0.000)	(0.000)	(0.000)
Electricity	6.2***	-15.1***	8.9***
	(0.000)	(0.000)	(0.000)
Sample size (N)	8262		
LR Chisq.	3563.6***		
Log likelihood	-4204.3		

Note: The estimates are obtained using MNP regression with kerosene and charcoal as the reference case. Where * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.The marginal effects are multiplied by 100 to convert in to percentages. Av. Stand for availability.

The result also identified the availability of LPG, charcoal and fuelwood to have a significant negative, positive and positive effect, respectively on the choice probability for transition fuel.

Access to modern infrastructure (electricity) was also a significant factor and it decreased the choice probability for solid fuel by 15.1% but increased the choice probability for modern and transition fuels by 6.2% and 8.9%, respectively. The implication of this result was that a well-developed infrastructure that aided the use of modern fuels such as electrification, availability of LPG stations, accessible road network connecting production and delivery points for LPG and the easy accessibility of such infrastructure was likely to promote the use of modern fuels, especially LPG. In summary, the results suggested that for any energy policy in Ghana to be successful, especially policy targeting promoting the use of modern fuels, provision of good education and access to modern infrastructure should be important parts of the policy mix. Besides, the policy mix should also include measures that would lead to reliable supply of modern fuels, especially LPG supply and the affordability of these fuels via poverty reduction measures.

The theoretical model indicated that the relative price of fuel was an important variable in the indirect utility function for the household regarding the choice of fuel, but in the empirical results, this variable was omitted due to lack of data. A potential effect of the omission was the bias it created for the estimated coefficients for the variables included in the model. However, as shown in Wooldridge (2002, pp.470) this bias did not carry over to the marginal effects in probit models, and since my interest was in obtaining the relative (marginal) effects of the explanatory variables, the omitted variable bias did not have serious consequences for my analysis.

Factors influencing the choice of the three main cooking fuels in Ghana

The focus in this section was to get a clear understanding of the factors that influenced the choice of the three main cooking fuels (fuelwood, charcoal and LPG) within the groups that have been analysed so far. The choice of the above three fuels was because they were major fuels in each of three groups of fuels (modern, solid and transition), for instance LPG was the major modern cooking fuel in Ghana, while fuelwood and charcoal were the major solid and transition fuels, respectively. The rational for analysing the individual fuels was to avoid the aggregation problem in order to assess the major fuels in each group. This is because it was likely that, irrespective of the fact that each of the fuels in a particular group shared common characteristics, households did not necessarily view them as the same, and might therefore prefer one fuel in the same group over the other, for instance LPG over electricity, or vice versa.

The estimated result for the three main cooking fuels is in Table 3 while the estimated $\beta's$ are reported in Table C in the Appendix, where other fuels (electricity, kerosene and crop residue) is used as the reference fuel for the MNP estimation. The marginal effects showed that household size and male head had a significant negative influence on the adoption probability for LPG. Income and education, on the other hand, had a positive effect on the probability of choosing LPG. Households located in urban areas were more likely to choose LPG as the main cooking fuel than those located in rural areas. A possible explanation for the positive urban effect was that, the opportunity cost of using fuelwood was very high due to the lack of cheaply available fuelwood in urban areas and lack of space for storing fuelwood.

Table 3: Marginal effects for choice of LPG, Fuelwood and Charcoal in Ghana (MNP)

	LPG *Marginal Effect*	Fuelwood *Marginal Effect*	Charcoal *Marginal Effect*
Hhsize	-0.6***	1.2***	-0.6**
	(0.000)	(0.000)	(0.003)
Male head	-1.9**	3.4***	-3.5***
	(0.001)	(0.000)	(0.000)
Age	-0.0	0.1***	-0.2***
	(0.889)	(0.000)	(0.000)
Log income	2.6***	0.0	-1.3***
	(0.000)	(0.771)	(0.000)
Basic educ.	4.9***	-5.8***	2.8**
	(0.000)	(0.000)	(0.007)
Secondary	11.6***	-15.1***	1.9
	(0.000)	(0.000)	(0.194)
Tertiary educ.	14.6***	-19.2***	1.2
	(0.000)	(0.000)	(0.587)
Av. of LPG	14.2***	-7.7*	-12.1***
	(0.000)	(0.025)	(0.000)
Av. of Kerosene	-4.7	10.0**	-4.4
	(0.026)	(0.001)	(0.172)
Av. of Charcoal	-2.2	-29.5***	27.4***
	(0.301)	(0.000)	(0.000)
Av. of Fuelwood	-2.1	20.5***	-13.6***
	(0.117)	(0.000)	(0.000)
Urban	5.8***	-29.0***	22.5***
	(0.000)	(0.000)	(0.000)
Electricity	6.1***	-14.3***	9.2***
	(0.000)	(0.000)	(0.000)
Sample size (N)	8262		
LR Chisq.	3615.4***		
Log-likelihood	-5238.2		

Note: *p*-values are in parentheses where * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Av. denote availability.

Consistent with previous studies, the results indicated that households with access to modern infrastructure had a higher choice probability for LPG as the main cooking fuel than households without access. The same results were found in Barnes et al. (2004) and Heltberg (2004, 2005). The results also indicated that the availability of LPG significantly increased the probability of choosing LPG, similar to the results in Gupta and Köhlin (2006), while the availability of kerosene significantly reduced the choice of LPG as a cooking fuel. With regard to the availability of both charcoal and fuelwood, I found no significant influence of these two on the choice of LPG.

In the case of fuelwood, the results revealed that the availability of each of the fuels (charcoal, fuelwood, kerosene and LPG) had a significant impact on the probability of choosing fuelwood. As expected, the availability of fuelwood increased the probability of choosing fuelwood, and this was also true for the availability of kerosene. On the other hand, the availability of both LPG and charcoal decreased the choice probability for fuelwood by 7.7% and 29.5%, respectively. Head of the household with education, urban location and access to infrastructure decreased the likelihood of using fuelwood for cooking. The reason for the negative impact of access to modern infrastructure on the choice of fuelwood was not very clear. Heltberg (2004) found similar result and suggested that, it could be because having electricity provided the basic infrastructure for easy adoption of modern fuels such as LPG, and therefore influenced the choice of cooking fuel away from traditional fuels such as fuelwood. Income did not have a significant impact on the choice probability for fuelwood, contrary to my a priori expectation.

Furthermore, the results showed that the availability of each of the fuels had a significant impact on the probability of choosing charcoal, except that of kerosene. Whereas age, male head and income decreased the adoption probability for charcoal by 0.2%, 3.5% and 1.3%, respectively, they increased the choice probability in the case of fuelwood, except income that was insignificant. Unlike the case of fuelwood and LPG, where each of the three levels of education had a significant effect on the choice probabilities, in the case of charcoal only basic education had a significant effect. Whereas household head with basic education increased the probability of using charcoal by 2.8% as the main cooking fuel, those with either secondary or tertiary levels of education had no significant effect. An interesting finding was that, whereas a male head of household reduced the choice probability for both LPG and charcoal, it increased the probability of choosing fuelwood. This was because in most Ghanaian households, women did most of the cooking, and hence most of the burden in collecting and the use of fuelwood fell on the woman. This simply meant that the male head preferred the cheaper fuel, fuelwood, since his opportunity cost for using fuelwood was low.

6. Conclusion

In this paper, I have analysed the factors responsible for the choice of modern, transition, and solid cooking fuels in Ghanaian households. I also broke down the fuel groups into their respective constituents and analysed the factors influencing the choice of each of the main cooking fuels (fuelwood, LPG and charcoal). The results from the study identified household characteristics in addition to the availability of the various fuels as significant factors that influenced the choice probabilities for modern, transition and solid fuels. In the case of modern fuels, the results showed that education, availability of LPG, access to modern infrastructure, household location, and income were the key factors that influenced the choice of this fuel. Whereas in the case of solid fuel, the results identified household location, education, access to modern infrastructure, male head, and availability of kerosene, fuelwood, and charcoal, as the key factors influencing the choice of solid fuel for cooking.

The results therefore supported policies that could increase the supply of modern fuels such as LPG, to promote the use of such clean fuels in Ghana. It would be important to emphasise that increasing the supply of modern fuels alone would not be enough to encourage households (especially poor households) to increase their use of modern fuels. Supply of these fuels would have to be complimented with efforts in educating households on the opportunity cost associated with using fuelwood, for example the dangers of using dirty fuels. Maybe more importantly, the monetary cost associated with the use of modern fuels must reduce significantly. The costs included not only the price of the fuels, but also the capital cost of stoves that were needed for modern fuels, as well as the cost of handling and storing the fuels. Finally, the results also identified household location and access to modern infrastructure as important factors determining the choice of cooking fuel, in the sense that households in urban areas with access to modern fuels tended to choose modern fuels, such as LPG, over traditional fuels. This does not imply a complete switch from traditional fuels to modern fuels but rather modern fuels become the dominant energy source for daily cooking needs, while traditional fuels such as fuelwood could still be used less frequently for particular services (such as for baking and cooking traditional staple food).

Furthermore, results based on the three main fuels also identified in general the availability of the various fuels as significant determinants in the choice of LPG, fuelwood and charcoal as cooking fuels. Income on the other hand appeared to significantly influence the choice probabilities for LPG and charcoal as cooking fuels but was insignificant in the case of fuelwood. The result should however be interpreted with some caution, as there may have been other important variables that we could not control for in this study. For example, differences in relative fuel prices, unit cost of cooking equipment, and awareness about health effects varied between households (which the education variable may not adequately capture). These and other factors could shed further light on the key factors influencing choice of fuel. For future work, it would be important to investigate these factors further in a panel framework, and furthermore, it would be interesting to investigate the factors that influence a switch between fuels by households. This would obviously be a significant contribution to this topic on Ghana.

References

Akpalu, W., Dasmani, I., and Aglobitse, P.B., 2011.Demand for cooking fuels in a developing country: To what extent do taste and preferences matter? *Energy Policy*, 39(10), 6525-6531.

Barnes, D.F., Krutilla, K., and Hyde, W., 2002. *The Urban Energy Transition: Energy, Poverty and the Environment in the Developing World*. World Bank.

Barnes, D.F., Krutilla, K., and Hyde, W., 2004. *The Urban Energy Transition: Energy, Poverty, and the Environment*. Forthcoming book volume, in press.

Byer, T., 1987. *Review of household energy issues in Africa*. World Bank internal memorandum, Washington D.C., USA.

Ghana Statistical Service, 2008. *Ghana Living standard survey: report of the fifth round (GLSS5)*.

Gupta, G. and Köhlin, G., 2006. Preferences for domestic fuel: Analysis with socio-economic factors and rankings in Kolkata, India. *Ecological Economics*, 57, pp. 107-121.

Heltberg, R., 2001. Determinants and impact of local institutions for common resource management. *Environment and Development Economics*, 6 (2), pp. 183-208.

Heltberg, R., 2004. Fuel switching: Evidence from eight developing countries. *Energy Economics*, 26, pp. 869-887.

Heltberg, R., 2005. Factors determining household fuel choice in Guatemala,*Environment and Development Economics*, 10, 337-361.

Hosier, R.H., and Dowd, J., 1987. Household fuel choice in Zimbabwe: an empirical test of the energy ladder hypothesis. *Resources and Energy*, 9, pp. 347-361.

IEA, 2006. Energy for cooking in developing countries. World Energy Outlook 2006. IEA:Paris, France.

Leach, G., 1987. *Household Energy in South Asia*. London and New York: Elsevier Applied Science.

Leach, G., and Gowan M., 1987. Household Energy Handbook: An Interim Guide and Reference Manual. *World Bank Technical Paper No. 67*. Washington D.C.

Masera, O.R., Saatkamp, B.D., and Kammen, D.M., 2000. From linear fuel switching to multiple cooking strategies: a critique and alternative to the energy ladder model. *World Development*, 28 (12), pp. 2083-2103.

O'Keefe, P., and Munslow, B., 1989. Understanding fuelwood. *Natural Resources Forum*, 13 (1), pp. 2-19.

Ouedraogo, B., 2006. Household energy preferences for cooking in urban Ouagadougou, Burkina Faso. *Energy Policy*, 34, pp. 3787-3795.

Reddy, A.K.N., and Reddy, B.S., 1994. Substitution of energy carriers for cooking in Bangalore. *Energy-The international Journal*, 19 (5), pp. 561-572.

Soussan, J., 1988. *Primary Resources and Energy in the Third World*. Routledge, London.

Wooldridge, J.M., 2002. *Econometric Analysis of Cross Section and Panel Data*. Mass: MIT Press, Cambridge.

WHO, 2009. The Energy Access Situation in Developing Countries

(http://www.who.int/indoorair/publications/PowerPoint_Energy_Access_paper-lr.pdf).

Appendix:

Appendix A;

Table A: List of variables and definitions

Variables	Type of Variable
Sex of household head	Dummy (Male=1, Female=0)
Age of household head	Continuous
Log annual household income	Continuous
Household size (HHS)	Continuous
Basic education	Dummy (Basic=1,No formal education=0)
Secondary education	Dummy (Secondary=1,No formal education=0)
Tertiary education	Dummy (Tertiary=1,No formal education=0)
Urban location	Dummy (Urban=1, Rural =0)
Electricity as the main lightening	Dummy (electricity=1, Others=0)
Charcoal	Dummy (charcoal=1, Others=0)
Fuelwood	Dummy (Fuelwood=1,Others=0)
LPG	Dummy (LPG=1, Others=0)
Modern fuel	Dummy (LPG or Electricity=1,Others=0)
Transition fuel	Dummy (Kerosene or Charcoal=1, Others=0)
Solid fuel	Dummy (Crop residue or fuelwood=1,Others=0)
Availability of LPG	Dummy (Yes=1, No=0)
Availability of kerosene	Dummy (Yes=1, No=0)
Availability of charcoal	Dummy (Yes=1, No=0)
Availability of fuelwood	Dummy (Yes=1, No=0)

Table B: Estimated $\beta's$ for modern and solid fuels from multinomial probit model

Variables	Solid $\beta's$	*P*-value	Modern fuel $\beta's$	*P*-value
hhsize	0.0885^{***}	(0.000)	-0.0604^{**}	(0.003)
Male head	0.318^{***}	(0.000)	-0.146	(0.077)
Age	0.0113^{***}	(0.000)	0.002	(0.472)
Log income	-0.0533^{*}	(0.020)	0.340^{***}	(0.000)
Basic education	-0.462^{***}	(0.000)	0.578^{***}	(0.000)
Secondary educ.	-0.840^{***}	(0.000)	1.393^{***}	(0.000)
Tertiary education	-1.028^{***}	(0.000)	1.757^{***}	(0.000)
Av. of LPG	-0.0746	(0.739)	1.951^{***}	(0.000)
Av. of kerosene	0.416^{*}	(0.034)	-0.464	(0.113)
Av. of Charcoal	-1.904^{***}	(0.000)	-0.723^{*}	(0.013)
Av. of fuelwood	1.254^{***}	(0.000)	-0.141	(0.452)
urban	-1.946^{***}	(0.000)	0.288^{**}	(0.005)
electric	-0.990^{***}	(0.000)	0.589^{***}	(0.000)
Intercept	1.900^{***}	(0.000)	-7.600^{***}	(0.000)
Sample size (N)	8262			

Table C: Estimated $\beta's$ for LPG, Fuelwood and Charcoal from multinomial probit model

Variables	LPG $\beta's$	*P*-value	Fuelwood $\beta's$	*P*-value	Charcoal $\beta's$	*P*-value
hhsize	-0.118***	(0.000)	0.0197	(0.255)	-0.0552**	(0.004)
Male head	-0.729***	(0.000)	-0.272**	(0.005)	-0.566***	(0.000)
Age	-0.008*	(0.013)	-0.001	(0.660)	-0.011***	(0.000)
Log income	0.584***	(0.000)	0.243***	(0.000)	0.250***	(0.000)
Basic educ.	1.158***	(0.000)	0.215*	(0.044)	0.614***	(0.000)
Secondary edu	1.458***	(0.000)	-0.671***	(0.000)	0.0848	(0.567)
Tertiary educ.	1.587***	(0.000)	-1.081***	(0.000)	-0.175	(0.424)
Av. of LPG	0.768**	(0.003)	-1.211***	(0.000)	-1.206***	(0.000)
Av of kerosene	-0.664	(0.075)	0.396	(0.137)	-0.172	(0.547)
Av of Charcoal	-0.343	(0.370)	-1.464***	(0.000)	0.446	(0.148)
Av of fuelwood	0.168	(0.510)	1.343***	(0.000)	0.171	(0.410)
urban	1.356***	(0.000)	-0.802***	(0.000)	1.084***	(0.000)
electric	1.355***	(0.000)	-0.141	(0.153)	0.791***	(0.000)
Intercept	-9.853***	(0.000)	-1.212*	(0.012)	-2.388***	(0.000)
Sample size (N)	8262					

Avhandlingar framlagda vid Institutionen för nationalekonomi, Umeå universitet

List of dissertations at the Department of Economics, Umeå University

Holmström, Leif (1972) Teorin för företagens lokaliseringsval. UES 1. PhLic thesis

Löfgren, Karl-Gustaf (1972) Studier i teorin för prisdiskriminering. UES 2. PhLic thesis

Dahlberg, Åke (1972) Arbetsmarknadsutbildning - verkningar för den enskilde och samhället. UES 3. PhD thesis

Stage, Jørn (1973) Verklighetsuppfattning och ekonomisk teori. UES 4. PhLic thesis

Holmlund, Bertil (1976) Arbetslöshet och lönebildning - kvantitativa studier av svensk arbetsmarknad. UES 25. PhD thesis

Löfgren, Karl-Gustaf (1977) En studie i neokeynesiansk arbetslöshets- och inflationsteori. UES 34. PhD thesis

Lundberg, Lars (1976) Handelshinder och handelspolitik - Studier av verkningar på svensk ekonomi. Industriens Utredningsinstitut, Stockholm. PhD thesis

Johansson, Per-Olof (1978) Sysselsättning och samhällsekonomi - En studie av Algots etablering i Västerbotten. UES 53. PhD thesis

Wibe, Sören (1980) Teknik och aggregering i produktionsteorin. Svensk järnhantering 1850-1975; en branschanalys. UES 63. PhD thesis

Ivarson, Lars (1980) Bankers portföljvalsbeteende. En teoretisk studie. UES 64. PhD thesis

Batten, David (1981) Entropy, Information Theory and Spatial Input-output Analysis. UES 92. PhD thesis

Hårsman, Björn (1982) Housing Demand Models and Housing Market Models for Regional and Local Planning. Swedish Council for Building Research, D13:1981. PhD thesis

Holm, Magnus (1983) Regionalekonomiska modeller för planering och samordning i en decentraliserad ekonomi. Byggforskningsrådet, R118:1981 and R5:1983. PhD thesis

Ohlsson, Henry (1986) Cost-Benefit Analysis of Labor Market Programs - Applied to a Temporary Program in Northern Sweden. UES 167. PhLic thesis

Sarafoglou, Nikias (1987) A Contribution to Population Dynamics in Space. UES 179. PhD thesis

Ohlsson, Henry (1988) Cost-Benefit Analysis of Labor Market Programs - Applied to a Temporary Program in Northern Sweden. UES 182. PhD thesis

Anderstig, Christer (1988) Applied Methods for Analysis of Economic Structure and Change. CERUM 1988:2, Umeå University. PhD thesis

Karlsson, Charlie (1988) Innovation Adoption and a Product Life Cycle. UES 185. PhD thesis

Löfström, Åsa (1989) Diskriminering på svensk arbetsmarknad - En analys av löneskillnader mellan kvinnor och män. UES 196. PhD thesis

Axelsson, Roger (1989) Svensk arbetsmarknadsutbildning - En kvantitativ analys av dess effekter. UES 197. PhD thesis

Zhang, Wei-Bin (1989) Theory of Economic Development - Nonlinearity, Instability and Non-equilibrium. UES 198. PhD thesis

Hansson, Pär (1989) Intra-Industry Trade: Measurements, Determinants and Growth - A study of Swedish Foreign Trade. UES 205. PhD thesis

Kriström, Bengt (1990) Valuing Environmental Benefits Using the Contingent Valuation Method: An Econometric Analysis. UES 219. PhD thesis

Aronsson, Thomas (1990) The Short-Run Supply of Roundwood under Nonlinear Income Taxation - Theory, Estimation Methods and Empirical Results Based on Swedish Data. UES 220. PhD thesis

Westin, Lars (1990) Vintage Models of Spatial Structural Change. UES 227. PhD thesis

Wikström, Magnus (1992) Four Papers on Wage Formation in a Unionized Economy. UES 287. PhD thesis

Westerlund, Olle (1993) Internal Migration in Sweden - The Role of Fiscal Variables and Labor Market Conditions. UES 293. PhLic thesis

Bergman, Mats A. (1993) Market Structure and Market Power. The Case of the Swedish Forest Sector. UES 296. PhD thesis

Johansson, Per (1993) Count Data Models - Estimator Performance and Applications. UES 315. PhD thesis

Roson, Roberto (1994) Transport Networks and the Spatial Economy - A General Equilibrium Analysis. UES 340. PhD thesis

Li, Chuan-Zhong (1994) Welfare Evaluations in Contingent Valuation - An Econometric Analysis. UES 341. PhD thesis

Østbye, Stein (1994) Regional Labour and Capital Subsidies - Theory and Evidence of the Impact on Employment under Wage Bargaining. UES 344. PhLic thesis

Westerlund, Olle (1995) Economic Influences on Migration in Sweden. UES 379. PhD thesis

Mortazavi, Reza (1995) Three Papers on the Economics of Recreation, Tourism and Property Rights. UES 396. PhLic thesis

Østbye, Stein (1995) Regional Labour and Capital Subsidies. UES 397. PhD thesis

Hussain-Shahid, Imdad (1996) Benefits of Transport Infrastructure Investments: A Spatial Computable General Equilibrium Approach. UES 409. PhD thesis

Eriksson, Maria (1996) Selektion till arbetsmarknadsutbildning. UES 410. PhLic thesis

Karlsson, Niklas (1996) Testing and Estimation in Labour Supply and Duration Models. UES 413. PhD thesis

Olsson, Christina (1996) Chernobyl Effects and Dental Insurance. UES 428. PhLic thesis

Vredin, Maria (1997) The African Elephant - Existence Value and Determinants of Willingness to Pay. UES 441. PhLic thesis

Eriksson, Maria (1997) To Choose or not to Choose: Choice and Choice Set Models. UES 443. PhD thesis

Widerstedt, Barbro (1997) Employer Change and Migration. Two Papers on Labour Mobility in Sweden. UES 444. PhLic thesis

Lundberg, Sofia (1997) The Economics of Child Auctions in 19th Century Sweden. UES 445. PhLic thesis

Lundberg, Johan (1997) Two Papers on Revenue and Expenditure Decisions in the Swedish Local Public Sector. UES 454. PhLic thesis

Widerstedt, Barbro (1998) Moving or Staying? Job Mobility as a Sorting Process. UES 464. PhD thesis

Bask, Mikael (1998) Essays on Exchange Rates: Deterministic Chaos and Technical Analysis. UES 465. PhD thesis

Löfgren, Curt (1998) Time to Study Students: Two Essays on Student Achievement and Study Effort. UES 466. PhLic thesis

Sjögren, Tomas (1998) Union Wage Setting in a Dynamic Economy. UES 480. PhD thesis

Mortazavi, Reza (1999) Essays on Economic Problems in Recreation, Tourism and Transportation. UES 483. PhD thesis

Rudholm, Niklas (1999) Two Essays on Generic Competition in the Swedish Pharmaceuticals Market. UES 485. PhLic thesis

Olsson, Christina (1999) Essays in the Economics of Dental Insurance and Dental Health. UES 494. PhD thesis

Marklund, Per-Olov (1999) Environmental Regulation and Firm Efficiency. UES 504. PhLic thesis

Berglund, Elisabet (1999) Regional Entry and Exit of Firms. UES 506. PhD thesis

Hellström, Jörgen (1999) Count Data Autoregression Modelling. UES 507. PhLic thesis

Nordström, Jonas (1999) Tourism and Travel: Accounts, Demand and Forecasts. UES 509. PhD thesis

Johansson Vredin, Maria (1999) Economics Without Markets. Four papers on the Contingent Valuation and Stated Preference Methods. UES 517. PhD thesis

Schei, Torbjørn (2000) Natural recreation resources: production and a diversity of interests related to the management of grouse as an outfield resource in Finnmark, Norway, in the Euro-Arctic Barents region. UES 523. PhLic thesis

Backlund, Kenneth (2000) Welfare Measurement, Externalities and Pigouvian Taxation in Dynamic Economies. UES 527. PhD thesis

Andersson, Linda (2000) Job Turnover, Productivity and International Trade. UES 530. PhLic thesis

Ylvinger, Svante (2000) Essays on Production Performance Assessment. UES 531. PhD thesis

Bergkvist, Erik (2001) Freight Transportation. Valuation of Time and Forecasting of Flows. UES 549. PhD thesis

Rudholm, Niklas (2001) The Swedish Pharmaceuticals Market - Essays on Entry, Competition and Antibiotic Resistance. UES 552. PhD thesis

Lundberg, Johan (2001) Local Government Expenditures and Regional Growth in Sweden. UES 554. PhD thesis

Lundberg, Sofia (2001) Going Once, Going Twice, SOLD! The Economics of Past and Present Public Procurement in Sweden. UES 557. PhD thesis

Eliasson, Kent (2001) University Enrollment and Geographical Mobility: The Case of Sweden. UES 558. PhLic thesis

Samakovlis, Eva (2001) Economics of Paper Recycling. Efficiency, policies, and substitution possibilities. UES 563. PhD thesis

Daunfeldt, Sven-Olov (2001) Essays on Intra-Household Allocation and Policy Regime Shifts. UES 570. PhD thesis

Hellström, Jörgen (2002) Count Data Modelling and Tourism Demand. UES 584. PhD thesis

Andersson, Linda (2002) Essays on Job Turnover, Productivity and State-Local Finance. UES 586. PhD thesis

Rashid, Saman (2002) Invandrarinkomster, förvärvsdeltagande och familj. UES 588. PhLic thesis

Hanes, Niklas (2003) Empirical Studies in Local Public Finance: Spillovers, Amalgamations, and Tactical Redistributions. UES 604. PhD thesis

Stenberg, Anders (2003) An Evaluation of the Adult Education Initiative Relative Labor Market Training. UES 609. PhD thesis

Stage, Jesper (2003) Mixing Oil and Water. Studies of the Namibian Economy. UES 611. PhD thesis

Marklund, Per-Olov (2004) Essays on Productive Efficiency, Shadow Prices, and Human Capital. UES 621. PhD thesis

Rashid, Saman (2004) Immigrants' Income and Family Migration. UES 625. PhD thesis

Sandberg, Krister (2004) Hedonic Prices, Economic Growth, and Spatial Dependence. UES 631. PhD thesis

Sjöström, Magnus (2004) Factor Demand and Market Power. UES 633. PhD thesis

Nilsson, William (2005) Equality of Opportunity, Heterogeneity and Poverty. UES 652. PhD thesis

Quoreshi, Shahiduzzaman (2005) Modelling High Frequency Financial Count Data. UES 656. Ph Lic thesis

Ankarhem, Mattias (2005) Bioenergy, Pollution, and Economic Growth. UES 661. PhD thesis

Quoreshi, Shahiduzzaman (2006) Time Series Modelling of High Frequency Stock Transaction Data. UES 675. PhD thesis

Ghalwash, Tarek (2006) Income, Energy Taxation, and the Environment. An Econometric Analysis. UES 678. PhD thesis

Westerberg, Thomas (2006) Two Papers on Fertility – The Case of Sweden. UES 683. Ph Lic thesis

Simonsen, Ola (2006) Stock Data, Trade Durations, and Limit Order Book Information. UES 689. PhD thesis

Eliasson, Kent (2006) College Choice and Earnings among University Graduates in Sweden. UES 693. PhD thesis

Selander, Carina (2006) Chartist Trading in Exchange Rate Theory. UES 698. PhD thesis

Humavindu, Michael N (2007) Essays on Public Finance and Environmental Economics in Namibia. UES 705. Ph Lic thesis

Norberg-Schönfeldt, Magdalena (2007) The Phase-Out of the Nuclear Family? Empirical Studies on the Economics and Structure of Modern Swedish Families. UES 708. PhD thesis

Granlund, David (2007) Economic Policy in Health Care: Sickness Absence and Pharmaceutical Costs. UES 710. PhD thesis

Jonsson, Thomas (2007) Essays on Agricultural and Environmental Policy. UES 719. PhD thesis

Broberg, Thomas (2007) The Value of Preserving Nature – Preference Uncertainty and Distributional Effects. UES 720. PhD thesis

Witterblad, Mikael (2008) Essays on Redistribution and Local Public Expenditures. UES 731. PhD thesis

Thunström, Linda (2008) Food Consumption, Paternalism and Economic Policy. UES 739. PhD thesis

Humavindu, Michael N (2008) Essays on the Namibian Economy. UES 745. PhD thesis

Persson, Lars (2008) Environmental Policy and Transboundary Externalities –Coordination and Commitment in Open Economies. UES 755. PhD thesis

Sahlén, Linda (2008) Essays on Environmental and Development Economics – Public Policy, Resource Prices and Global Warming. UES 762. PhD thesis

Lönnbark, Carl (2009) On Risk Prediction. UES 770. PhD thesis

Norin, Anna (2009) Worker Safety and Market Dynamics. UES 772. PhLic thesis

Holmlund, Linda (2009) Essays on Child Care and Higher Education. UES 783. PhD thesis

Landström, Mats (2009) Two essays on Central Bank Independence Reforms. UES 792. PhLic thesis

Åström, Johanna (2009) Marriage, Money and Migration. UES 790. PhD thesis

Birkelöf, Lena (2009) Spatial Interaction and Local Government Expenditures for Functionally Impaired in Sweden. UES 798. PhD thesis

Mannberg, Andrea (2010) Risk and Rationality – Effects of contextual risk and cognitive dissonance on (sexual) incentives. UES 806. PhD thesis

Andersson, Camilla (2010) Changing the Risk at the Margin: Smallholder Farming and Public Policy in Developing Countries. UES 810. PhD thesis

Sörensson, Robert (2010) Marshallian Sources of Growth and Interdependent Location of Swedish Firms and Households. UES 815. PhD thesis

Soultanaeva, Albina (2011) Back on the Map - Essays on the Financial Markets in the three Baltic States. UES 820. PhD thesis

Holmberg, Ulf (2012) Essays on Credit Markets and Banking. UES 840. PhD thesis

Lundgren, Jens (2012) Market Liberalization and Market Integration – Essays on the Nordic Electricity Market. UES 852. PhD thesis

Karimu, Amin (2013) Essays on Energy Demand and Household Energy Choice. UES 864. PhD thesis