CHAPTER 3: REVIEW OF APPROACHES TO STUDYING SECTORAL ENERGY INTENSITIES AND ENERGY SWITCHING POLICIES

This chapter reviews energy economics as it relates to industrial production and household consumption, and how economists have studied policies aimed at encouraging a switch from one source of energy to another. Section 3.1 reviews the theory of production and how energy enters production relationships. It also discusses how technology may determine energy efficiency. Approaches to studying sectoral energy intensities and switching policies are reviewed in section 3.2. The section also discusses theoretical and empirical issues that arise in partial equilibrium and general equilibrium approaches. Section 3.3 introduces energy as a consumer good and section 3.4 reviews approaches to studying household energy demand and substitution possibilities. The section focuses on the household production and the random utility frameworks. Section 3.5 provides a synthesis of the reviewed literature and concludes the chapter.

3.1 Energy as a factor input

There are two basic definitions of energy that are relevant to economic theory of production. The first comes from physics, and describes energy as the capacity of matter or radiation to perform work. The second, and closest to everyday language usage, refers to energy as the power derived from physical or chemical resources to provide light and heat or to work machines (Oxford English Dictionary, 2001). Energy in the latter sense is often transformed into homogeneous physical units such as the British thermal units (Btu), combining various energy inputs into aggregate or separate units. Thompson (2006) refers to such a homogeneous physical unit of energy, E, as produced energy which required capital (K), labour (L) and a natural resource input (N) to convert into energy:

$$E = E(K, L, N) \qquad (3.1)$$

Further, energy is embodied in products through the generic production function of the form

$$y = f\left(z, E(K, L, N)\right) \tag{3.2}$$

Where $z = (z_1, z_2, \ldots, z_n)$ is a vector of primary and intermediate inputs other than energy, and y is gross output.

Presented as above, final energy consumption is mainly attributed to production activities, implying that an increase in final demand for goods and services would result in an increase in final energy consumption for any given technology (Ferng, 2002). However, if energy is a produced commodity available for final consumption, some energy would be demanded by households either separately or as a complement of some other commodity in household consumption. At macro level, the gross output of the energy sector must either be used up in the production of other goods and services (as intermediate input) or it must be absorbed by final demand sectors. Since the behavioural aspects of demand for produced energy is considered important for policy analysis, the next section outlines the theoretical aspects of energy demand by production sectors.

3.1.1 Energy in production

The standard production problem starts with an economic unit or entity (typically a firm) that transforms a set of different types of inputs into one or more outputs. The mapping from inputs to outputs is usually summarised using a production function which delimits the technical constraints of the representative firm. Generalizing equation 3.2 above, the technological constraint of a firm can be defined as follows:

$$0 \leq y \leq f(z, E); \quad z, E \geq 0 \tag{3.3}$$

Where z, and E are as defined earlier, except that E might have a different technology than the one implied by equation 3.1. Equation 3.3 states that a firm would, with positive inputs, produce at least some given level of positive output.

A technological delimitation that a firm might face is that of essentiality of energy. There is strict essentiality if it is impossible to produce output without using any form of energy, implying that:

$$f(z,0)=0 \tag{3.4}$$

Energy, just like all other factors of production has derived demand. Energy is demanded conditional on the firm's chosen output level and technology used. Assuming that the firm is a profit maximizing entity that faces exogenous input prices, the firm's cost minimization problem can be given by:

$$\min_{z_1,\cdots,z_n,E} C(w,e,y)=\sum_{i=1}^{n} w_i z_i + eE \quad s.t.\ (i)\ f(z_1,\cdots,z_n,E)\geq y$$
$$(ii)\ z_i \geq 0$$
$$(iii)\ E \geq 0 \tag{3.5}$$

Where w_i = unit price of input i

e = unit price of energy input

z_i = amount of input i

Since a profit maximizing firm would choose only that bundle of inputs which minimizes the total cost of producing a given level of output, the derived demand for inputs, including energy, depends on the level of output, the substitution possibilities among inputs implied by the production function, and the relative prices of all inputs (Berndt and Wood, 1975). Using Shephard's duality theorem (Humphrey and Moroney, 1975; Woodland, 1993), the partial derivative of $C(w,e,y)$ with respect to e gives the conditional energy demand, $E(w,e,y)$. Similarly, the partial derivative of $C(w,e,y)$ with respect to w_i gives the conditional input demand $z_i(w,e,y); \forall i=1,\cdots,n$.

The concern with the elasticity of substitution between capital and energy was considered important following the world oil crisis in the 1970s in view of uncertainty regarding future energy prices and availability. It was believed for instance that if capital and energy are complements, increases in prices would perhaps induce a reduction in the demand for capital goods, thereby stifling growth. On the other hand, if capital and energy are substitutes, rising energy prices would stimulate demand for capital (Thompson and Taylor, 1995; Berndt and

Wood, 1975). In general, the outcome of decisions regarding energy policy depends heavily on substitution between energy and other factors of production. However, literature on energy substitution offers no consensus regarding specification, size and direction of change due to relative prices (Thompson, 2006).

The Allen relative elasticity of substitution (RES), also called the Hicks-Allen elasticity of substitution, measures the responsiveness of relative inputs to relative input prices. The RES between inputs i and j is the percentage change in relative input factor i with respect to the change in the relative price of factor j (Thompson, 1997):

$$RES_{ij} = \frac{\partial \ln(z_i / z_j)}{\partial \ln(w_j / w_i)} \tag{3.6}$$

Where the z's are cost-minimizing inputs per unit of output and w's are input prices.

Allen (1938) showed for constant returns to scale production function $y = f(z)$ that the partial elasticity of substitution can be expressed as:

$$\pi_{ij} = \frac{yF_{ij}}{z_i z_j F} \tag{3.7}$$

Where F is the bordered Hessian matrix of partials and cross partials of the production function, and F_{ij} is the cofactor of the element i, j.

The Allen partial elasticity of substitution is inappropriate in energy studies because of the problem of economic interpretation. In particular, with three or more inputs, the percentage change in the relative input of factor i due to a change in the relative price of factor j is a meaningless statistic that holds all other inputs constant, when in fact all inputs adjust to any change in factor prices (Thompson, 2006). There are other reasons favouring alternative measures of elasticities to the Allen partial substitution elasticity in energy studies. Thompson and Taylor (1995) noted that for inputs such as energy that usually consist small cost shares, relatively small changes in the use of the input can induce large changes in Allen partial elasticity estimates. In addition, Allen partial elasticities are relatively less robust to levels of data aggregation in empirical applications (Shankar et al., 2003).

Welsch and Ochsen (2005) used the Morishima elasticity of substitution (MES) to measure substitutability between capital and labour, between capital and energy and between labour and energy. The MES measures the negative percentage change in the ratio of input i to input j when the price of input j alters. Blackorby and Russell (1989) define the MES between inputs i and j in a production function with many inputs as:

$$MES_{ij} = \frac{\partial \ln(z_i/z_j)}{\partial \ln w_j} \tag{3.8}$$

And cross-price Morishima elasticity of substitution (CMES) as:

$$CMES_{ij} = RES_{ij} - RES_{jj} \tag{3.9}$$

The MES is a generalization of the two-factor elasticity of substitution to the case of multiple (>2) inputs. An input j is a Morishima substitute (complement) for input i if $\mathrm{MES}_{ij} > (<) 0$ (Blackorby and Russel, 1989; Welsch and Ochsen, 2005). From (3.8) above, relative input price changes are not explicitly considered in the Morishima elasticity, although the cross-price elasticity shows a clear relationship between the Morishima and the Hicks-Allen relative price elasticity of substitution.

Thompson (1997) also considers the MacFadden elasticity measure in addition to the Allen (1938) and Morishima elasticities. The MacFadden elasticity allows for change in relative input price but holds cost constant (the cost-minimizing envelope). Taking the total differential of the cost function $C(w,e,y) = \sum_{j=1}^{n} w_j z_j + eE$, when only the prices of inputs i and E change, we get:

$$dC = 0 = \sum_{j=1}^{n} z_j dw_j + Ede = \theta_i \hat{w}_i + \theta_E \hat{e} \tag{3.10}$$

Where $-\frac{\theta_i}{\theta_E} = \frac{\hat{e}}{w_i}$, the percentage change in relative inputs, and the circumflex represents percentage change. The MacFadden shadow elasticity is expressed as half the weighted average of the two relevant Morishima elasticities

$$\phi_{iE} = \left.\frac{d\ln(z_i/E)}{d\ln(e/w_i)}\right|_{dC=0} = 0.5\frac{(\theta_i MES_{iE} + \theta_E MES_{Ei})}{(\theta_i + \theta_E)} \tag{3.11}$$

3.1.2 Energy intensity and efficiency

Apart from the implications of substitution possibilities between energy and non-energy factors, energy economics studies have also focused on efficiency of energy resource use by production activities. Energy efficiency is often defined in terms of energy intensity of a production activity. Energy efficiency improves if the energy intensity, i.e., the quantity of energy required per unit of output or activity, falls over time (Markandya et al., 2006). Energy intensity could therefore be interpreted as measure of single factor productivity similar to average output, since it is a ratio of output to the input of energy.

There is however some dissatisfaction with the quality of energy intensity indicators in literature. Freeman et al. (1997) quotes a US Department of Energy study which found that energy intensity in manufacturing had increased by 4.5 percent between 1988 and 1991 while when a value-based measure of output was used, energy intensity declined by 12.7 percent over the same period. Apart from differences in output measures used in literature, there are also differences in choice of unit of measurement of energy. For instance, the definition of energy intensity adopted by Markandya et al. (2006) uses tons of oil equivalent per 2000 purchasing power parity (PPP) dollar, while other studies measure energy intensity as Btu per unit of economic activity (value added or gross output).

Berndt (1978) proposed that energy efficiency should be analysed in the larger context of energy and non-energy inputs than just looking at energy-output ratios. Such a framework would allow analysis of issues such as the effect of energy price increases on tradeoffs between energy and labour in production. However, aggregating over a number of fuels to come up with one estimate of energy use per activity is unsatisfactory even after introducing non-energy inputs in the analysis. In particular, Berndt (1978) argues that aggregating over energy types to obtain the total Btu demand and supply forecasts is problematic because energy types are to some extent substitutable in end-use demands. In addition, the price per Btu of the various primary and secondary energy products is not equal among energy types.

Regardless of problem with the current energy efficiency measure, it is recognised in literature that energy efficiency is both an environmental and economic concern. From the environmental viewpoint, energy efficiency may be adopted as a policy goal in a bid to conserve or slow down the depletion of fossil fuel reserves. Complementary to the first goal is the reduction in greenhouse gas emissions related to fossil fuel use. From the economic point of view, energy efficiency may also be interpreted in terms of minimizing costs in the face of rising energy prices (Mukherjee, 2006). However, from the economic point of view, it is recognised that changes in energy intensity in production may not necessarily reflect underlying trends in technical efficiency, but rather changes in the structure of the industry (Freeman et al., 1997; Garbaccio et al., 1999). Further, the change in industrial composition may be as a result of international trade effects which induce energy saving on the economy (Welsch and Ochsen, 2005).

3.1.3 Technology as a determinant of energy intensity

If energy efficiency is interpreted as declining industrial energy intensity over time, there is a *prima facie* case for associating the state of technology with industrial energy intensity. According to projections from the International Energy Agency data, fossil based fuels will account for more than 90 percent of world primary energy demand up to 2010, and probably up to 80 percent in 2020 (IEA,2003). However, it is often assumed that technological advancement will generally lead to a reduction in some forms of energy use, especially fossil fuels because they are considered environmentally damaging, and/or economically wasteful.

Developing countries use fuels less efficiently than industrialized countries because of lack of state-of-the-art technology. According to Sathaye and Ravindranath (1998), fuel efficiency is also compromised because of the proportionately higher use of coal and biomass which produce more carbon dioxide per unit of energy than do petroleum products and natural gas. It is also suggested that capital intensive production activities in developing countries are the ones that demand proportionately more carbon-intensive fuels than labour intensive activities. It is therefore expected that energy policies would be key in determining not just energy market developments in developing countries but also economic growth and welfare (Solsberg, 1997).

There are several reasons for proposing that capital intensive sectors in developing countries are also energy intensive. The first reason is that at low levels of economic development, many of developing country plant and machinery are operated at excess capacity and are thus not energy efficient. Second, even where modern plant and machinery have been adopted, economic development may increase demand for goods and services to levels that erode the gains from adopting energy-efficient technologies. The second reason is called the rebound effect and it occurs when proliferation of energy-efficient technologies achieve substantial cost savings on energy services whose general equilibrium effects are increased demand for energy services and greater energy consumption as the savings are spent elsewhere in the economy (Jaccard and Associates, 2004; Boonekamp, 2007;Takase et al., 2005).

Policies aimed at stimulating energy efficiency in production may have one of two possible impacts on individual firms depending on whether or not a firm was producing at full employment. If a firm were operating below full employment, it could significantly reduce energy use without loss of output. Does this mean that it is possible for a firm to adjust employment of energy and other inputs at zero cost? If on the other hand production was already energy efficient, any policy designed to reduce energy use would necessarily raise the cost of producing a given level of output as energy prices are increased (Thompson, 2000; Smulders and de Nooij, 2003; Klepper and Peterson, 2006).

In the likely event that cost of production rise with the implementation of energy policies, energy studies quantify the magnitude by which costs rise. The direct impact of energy price changes would depend on the ease of substitution between energy and non-energy inputs, which in turn depends on the state of technology. Therefore, to avoid loss of output or to counteract rising production costs, a profit maximizing firm would either embark on a radical technological innovation (adoption of a completely new technology) or an incremental innovation to the existing technology. However, the former type of innovation is rarely observed in reality because of the presence of uncertainty (Jaccard and Associates, 2004).

In policy analysis, a distinction can be made between policies that reduce the level of energy use from those that reduce the growth rate of energy inputs. Although both policies may stimulate innovation, they have the unsavoury characteristic of reducing output levels. According to Smulders and de Nooij (2003) technical change should be viewed as an endogenous variable whose evolution is induced directly through changes in energy prices, or indirectly through innovation when a firm takes up energy saving technologies. A similar view to the one held by Smulders and de Nooij is presented in an endogenous growth theoretic framework by Otto et al. (2006) who developed a general equilibrium framework that links energy, the rate and direction of technical change and the economy.

The dichotomy between energy policies that reduce the level of energy use and those that reduce the growth rate of energy inputs is rather blurred in practice. According to Pindyck (1979), most energy studies have focused on isolating the substitutability of energy and other factors of production when examining the effect of GNP growth and changes in fuel prices on industrial demand for energy. However, one can also focus on substitutability of fuels within the energy aggregate (Mountain, 1989; Woodland, 1993; Jones, 1996). The distinction between elasticity of substitution among fuel types in the energy aggregate and elasticity of substitution between energy and non-energy inputs becomes important when firms generally use different production technologies.

The importance of both technology and elasticities in applied energy studies stem from the fact that elasticities determine the economic costs of technology adaptation under energy policy constraints. If energy and capital are substitutes, higher priced energy would *ceteris paribus*, increase demand for new capital goods. Also, limited substitutability between energy and non-energy inputs could be reflected in high adjustment costs by firms to higher energy prices as significant technical changes may be required (Berndt and Wood, 1975). Elasticities are also crucial in determining the rate of an environmental tax and subsidy that would attain a given environmental target (Pindyck, 1979; Klepper and Peterson, 2006; Kemfert and Welsch, 2000).

3.2 Approaches to studying sectoral energy intensities and switching policies

3.2.1 Nonparametric and parametric partial equilibrium models

Mukherjee (2006) used data envelopment analysis (DEA) to examine energy efficiency in manufacturing sectors for the period 1970 to 2001. DEA recognises that multiple inputs are used in the production of output, and thus allows input substitutions. Efficiency is measured based on an intertemporal production possibility frontier. With DEA, the concept of energy intensity is now replaced with that of a set of all possible input bundles that could produce a given level of output. Efficiency is therefore measured by comparing the actual level of either inputs or outputs against a minimum value implied by the inputs feasible set or maximum output value implied by the production possibility frontier. The input-oriented technical efficiency is defined as the ratio of optimal (minimum) input bundle to the actual input bundle of a decision making unit (DMU) for any given level of output, holding input proportions constant. The output-oriented technical efficiency is implicitly defined as the ratio of the observed output to the optimal (maximum) achievable output.

Garbaccio et al. (1999) used decomposition analysis to explain a 55 percent reduction in energy use per unit GDP in China between 1978 and 1995. The fall in energy use was decomposed into technical change and various structural changes including changes in quantity and composition of imports and exports. Technical change within sectors accounted for most of the fall in energy-output ratio while structural change actually increased energy use. It was also found that imports of energy-intensive goods lowered energy-GDP ratios. However, the level of aggregation for sectoral inputs and outputs was considered crucial for distinguishing the impact of technical and structural factors on energy-output ratios.

Descriptive decomposition studies are criticized for failing to identify sources of energy efficiency improvements and energy saving structural change. It is therefore not possible within the framework of descriptive decomposition to conduct a joint assessment of factor substitution and technological change. In the end, there is ambiguity as to whether changes in energy intensity are a result of technological factors (energy efficiency due to factor substitution and/or biased technological change) or structural factors (composition of aggregate output due to international trade effects). According to Welsch and Ochsen (2005) the alternative is to estimate factor share equations which in a way endogenize factor prices.

Estimates of interfuel elasticity of substitution have been empirically obtained using various specifications. The two most common specifications are the translog cost function and the linear logit cost share function. The translog function was developed by Christensen et al. (1973) and became popular over Cobb-Douglas specifications because it placed no *a priori* restrictions on Allen elasticities of substitution. It is however, Pindyck's (1979) translog model of capital-labour-energy aggregates that has been extensively adopted by various studies of energy demand.

Berndt and Wood (1979) interpreted and reconciled the contradictory evidence in literature regarding substitution possibilities between energy and capital. For instance, Berndt and Wood (1975) found complementarity between energy and capital in time series data while Griffin and Gregory (1976) and Pindyck (1979) found substitutability between energy and capital in pooled time series data. The conclusion by Berndt and Wood (1979) was that differences in results were partly due to differing data sets used, approaches to measuring input quantities and prices, treatment of excluded inputs and distinction between short-run and long-run elasticities. In addition, energy-capital complementarity based on time series data reflected short-run variations in capital utilization but the true long-run was one of energy-capital substitutability as found by Griffin and Gregory (1976) and Pindyck (1979). Thus, pooled cross-section time series elasticity estimation should be more realistic compared to elasticities estimated solely on time series data (Griffin and Gregory, 1976; Pindyck, 1979).

The issue of capital-labour-energy (KLE) substitution is however surrounded by uncertainty over the appropriate technological representation and numerical values for substitution elasticities (Kemfert and Welsch, 2000). The importance of both technology and elasticities in applied energy studies stem from the fact that elasticities determine the economic costs of technology adaptation under energy policy constraints. Elasticities are also crucial in determining the rate of an environmental tax that would attain a given level of environmental quality target (Pindyck, 1979; Klepper and Peterson, 2006; Kemfert and Welsch, 2000).

Pindyck (1979) used a translog cost function that is homothetically separable in the KLE aggregates. Although estimated at macro level, the cost function is consistent with microeconomic behaviour of cost minimization at two levels namely, the energy aggregation stage where the choice of fuel inputs minimize cost of energy input, and the output aggregation stage where the choice of KLE minimizes the cost of production. The model allows for cross-price effects of energy and non-energy inputs, as well as among individual fuels in the energy aggregate.

Earlier studies of aggregate input substitution like Berndt and Wood (1975), Pindyck (1979) and Griffin and Gregory (1976) relied heavily on separability assumption which is equivalent to placing restrictions on Hicks-Allen partial elasticity of substitution and price elasticities. According to Berndt and Christensen (1973a) use of capital and labour aggregates implies stringent separability restrictions on neoclassical production function or equivalently, that there exists a price aggregate for the weakly separable components of the aggregate inputs. Blackorby and Russell (1981) later developed equivalent restrictions for Morishima elasticity of substitution. According to Berndt and Christensen (1974), little information is lost by aggregating inputs if within each aggregate factors are highly substitutable for one another. Also, factor intensities can be optimized within each separate subset of a function on which certain equality restrictions on Allen partial elasticities of substitution hold (Berndt and Christensen, 1973b).

Other studies avoid the aggregation issue by including components of a subset in the estimation equation. Woodland (1993) for instance, used a translog system for coal, gas, electricity, oil, labour and capital as production factors. Unlike Pindyck's (1979) macroeconomic approach, Woodland used a repeated cross-section of companies observed from 1977 to 1985. Woodland also estimated separate translog functions for each observed energy patter (i.e., energy mix used by a company) on the assumption that the energy mix in a company was exogenously determined by technology.

There are however concerns about the appropriateness of the translog specification in energy studies. Compared with the linear logit model, the translog cost functional form has the potential to produce negative cost shares because it fails to satisfy regularity conditions (concavity) for negative own-price effects over the relevant range of fuel prices (Jones, 1996). Although the validity of concavity assumption depends both on functional form and the dynamic specification of the adjustment of producer behaviour, and could be tested ex-post, Urga and Walters (2003) found that the translog specification violated the concavity conditions in most cases. In particular, the translog specification does not guarantee positive cost shares and negative own-price effects. Also, unlike the linear logit model the translog cost function fails to meet the Le Chatelier principle, i.e., long-run direct price effects are never smaller than the short-run effects.

The dynamic linear logit model performs particularly well in applied energy studies. Jones (1999) used a dynamic linear logit model that estimates theoretically consistent fuel price elasticity, i.e., negative own-price effects and positive cross-price elasticities between fuels (for substitutes). The model also gives a direct estimate of the rate of dynamic adjustment to fuel price changes that is consistent with the Le Chatelier principle. The rate of adjustment is important as it relates to two main costs associated with energy policy changes. First, there are costs associated with the extra emissions during the transition from carbon-intensive fuels to cleaner fuels. Second, there are economic as well as investment costs that must be incurred as firms change their fuel technology.

There are other theoretical and empirical benefits from using the linear logit specification. In particular, the linear logit specification allows the estimation of nonlinear Engel curves, and partial adjustment mechanisms without placing undue restrictions on the input structure (Considine, 1990; Considine and Mount, 1984). The input shares satisfy the adding-up and non-negativity conditions consistent with neoclassical demand theory (Shui et al., 1993), and symmetry of the second partial derivatives of the cost function could be defined for each set of cost shares in a sample (Considine, 1990). In addition, the demand systems are continuous and thus subject to the same restrictions as the translog and CES cost functions (Brannlund and Lundgren, 2004; Atkinson and Halvorsen, 1976).

Despite the advantages that linear logit model has, there are econometric problems associated with the share demand formulation. In particular, the linear logit model leads to misleading inferences arising from the presence of prices on both sides of the equation (Hsiao and Mountain, 1989). Further, although the autoregressive nature of the error term of the logit model can be established ex-post (Chavas and Segerson, 1986; Considine, 1990), the distribution of the error term may not be consistent with the assumption of normality. Thus statistical hypotheses from linear logit models may be misleading (Mountain and Hsiao, 1989).

Thompson (2006) reviewed the applied theory of energy cross-price partial elasticities of substitution using regression analysis. The most important conclusions from the reviewed theory are that: (i) estimates of cross-price substitution are sensitive to the industries and regions of study, (ii) choice of functional form may affect estimated cross-price elasticities, (iii) time periods chosen and the dynamic model of substitution are critical due to path dependencies that arise given fixed cost of input adjustments and (iv) substitution involving an aggregate is not necessarily a weighted or other average of the disaggregated inputs.

Thompson (2006) also presented a duality theory based on log-linear (Cobb-Douglas) and translog specifications from which cross-price elasticities were specified and estimated.

Welsch and Ochsen (2005) estimated share equations for energy, capital, low-skilled labour, high skilled labour and materials. The focus of the study was on factor substitution between energy and capital in a translog cost function for aggregate gross output. The share equations were estimated using the method of iterated three stage least squares which is a special case of generalized method of moments (GMM). The study concluded that materials, capital, and low-skilled labour are Morishima complements to energy. They also concluded that energy is a Morishima substitute for all other inputs except materials, whereas all inputs are Morishima complements to energy.

The finding by Welsch and Ochsen (2005) that capital is a Morishima complement to energy differed significantly from previous findings in the 1970s and 1980s that capital is a Morishima substitute for energy (Thompson and Taylor, 1995). Welsch and Ochsen (2005) explained their result by noting that most of the earlier studies focused on manufacturing, whereas their study refers to overall production (aggregate data). Thus, while substitutability may prevail in manufacturing, the overall production function may be characterised by capital being a complement to energy. In addition, temporal differences in data coverage may have influenced the result. For instance, their energy data comprised a higher share of electricity than previously used data sets, which may actually imply that while capital might have been a substitute for fuels, capital was more likely to be a complement to electricity.

3.2.2 General equilibrium models of energy substitution

Leontief (1970) showed that economic systems and the environment are linked starting from natural inputs that enter production or consumption relationships. Leontief's idea was later extended to emissions that could feedback to the economy through production technologies and consumption functions (Mestelman, 1986). Others studies including Ferng (2002) and Kratena (2004) considered energy as fundamental to pollution analysis because biomass and fossil energy are the main sources of anthropogenic perturbations of the ecosystem carbon cycle. Kratena (2004) even suggested the use of energy as a 'numeraire' for ecosystems flows since energy is needed to drive the biogeochemical cycles in ecosystems.

The 1973 oil price shock provided the first impetus to the development of general equilibrium models for energy policy analysis. The first energy policy analyses focused on energy demand and supply options, but recently, the focus has shifted to environmental pollution (Bhattacharyya, 1996). With rising energy prices and uncertainty over future energy availability, energy policy issues that came to the fore were price formulation, output determination, income generation and distribution, consumption behaviour, government operation and reducing emission of greenhouse gases associated with energy use.

Within the framework of applied general equilibrium (AGE) modelling the major aim is to measure the overall economic impacts in any economy of changes in the energy sector. While the first studies concerned themselves with technological change and how to represent substitution between energy and non-energy inputs, the focus has shifted to problems associated with the supply of energy and the external effects associated with the use of energy, particularly fossil fuels at the beginning of the 1990s (Bergman and Henrekson, 2003).

There are at least three AGE modelling approaches discussed by Bhattacharyya (1996) that are relevant to energy economics. The first AGE modelling strategy due to Hudson and Jorgenson (1974) uses econometrics to estimate parameters of a general equilibrium system. The Hudson and Jorgenson (1975) energy study in particular, was aimed at examining how relative product and factor prices, and the allocation of resources might be affected by factors such as increasing energy costs, technological change in the energy sector or various energy policy changes. The paper assumed that capital and energy were substitutes other than complements, although the elasticities of substitution were less restrictive. Elasticities were econometrically estimated using constant returns to scale translog price possibility frontiers.

The most common specification of production technology in studies following the Hudson and Jorgenson (1974) approach is the ‘nested’ CES function that includes the KLE factors. These functions are estimated econometrically to obtain elasticities that are incorporated into general equilibrium models or other policy analyses. Kemfert and Welsch (2000) test three CES specifications, all with a neutral technical progress factor: (i) a two-level CES function with E/K composite substituting labour, (ii) a two-level CES function similar to the specification used by Manne and Richels (1992) with K/L composite substituting energy, and (iii) a two-level CES function with L/E composite substituting capital. From the empirical results Kemfert and Welsch (2000) conclude that a nested CES production with a composite of K/E seemed more appropriate for aggregate production function, although their disaggregated sectoral production functions had mixed results.

Bohringer (1998) used a simple separable nested CES functions to capture technology information on energy system in production. The purpose of the study was to compare and integrate elasticity based computable general equilibrium (CGE) models (top-down) and 'true' technology based activity analysis (bottom-up). The top-down CGE approach uses price dependent point-to-point continuously differentiable functions for which a Walrasian general equilibrium exists at which no firm earns excess profits and all output is allocated. To integrate bottom-up approaches, discrete Leontief technologies are specified for lower level activities. For energy economics studies, however, the top-down approach is more appropriate because it uses microeconomic models with detailed representation of the energy sector unlike the bottom-up approach which appeals to engineering search for different technical potentials for achieving set targets such as emission reductions (Klepper and Peterson, 2006).

The main advantage of the Hudson and Jorgenson (1974, 1975) approach is that endogenous relative energy price (response) functions are derived within a framework that allows for endogenous technological change. The model accommodates complementarity between two types of inputs as well as different partial elasticity of substitution between pairs of inputs, which are ruled out by technology constraints represented by CES and Cobb-Douglas production functions (Bergman, 1988).

The Hudson and Jorgenson approach requires annual time series data and thus the estimated elasticities are short run. The problem with short-run elasticities however, is that they understate the response capacity of agents when a longer adjustment period is considered. Also, the large number of parameters to be estimated would require long time series if the BLUE properties of the estimates are to be maintained. Structural changes during the time over which estimates are generated may also not be reflected in the parameters, and the parameters are generally not adequate because they are obtained without imposing the full set of general equilibrium constraints. In addition, lack of data, computational and conceptual difficulties in estimation and uncertainty concerning the validity of resulting estimates limits the applicability of the econometric approach in developing countries (Arndt et al., 2002).

The second approach due to Johansen (1960, 1974) follows the multisector growth model (MSG). The MSG assumes fixed input-output coefficients for intermediate inputs, log-linear or Cobb-Douglas production function for value added (mainly, labour and capital), and one representative household. Later variants of the original MSG introduce sectoral disaggregation and the Armington assumption for international trade. The model solution is found by calibrating the values to their base year.

Similar to the Hudson and Jorgenson (1975) approach, the MSG incorporates substitution possibilities between KLE and materials (M) aggregates. However, the substitution responses are represented by generalized Leontief cost functions interpreted as second order approximations to the underlying production structure (Bergman, 1988). Hence the MSG shares the same weakness as the Hudson and Jorgenson (1975) model. In addition, the MSG has the restrictive assumption of a representative household, which means that such a model would fail to account for impacts of energy policies on different sections of the population.

The third modelling approach is due to the works of Harberger (1962), Scarf (1967), and Shoven and Whally (1984). Harberger (1962) used a two sector general equilibrium model of tax and trade cast in the Walrasian and Heckscher-Ohlin traditions. Scarf (1967) on the other hand was the first to offer an algorithm for computing a Walrasian general equilibrium. Later, Shoven and Whalley (1984) implemented the Scarf algorithm to finding a general equilibrium with taxes (Bergman and Henrekson, 2003). The main characteristics of their approach are: (i) multiple households, each with initial endowment and set of preferences, (ii) detailed formulation of tax structures, and (iii) closely follow the Walrasian general equilibrium theory to analyse welfare effects of different policies. The model solution is found by calibration, just like the MSG model (Bhattacharyya, 1996).

Separately, Goulder is one of the most prominent authors applying the Harberger and Scarf models to energy studies (Borgess and Goulder, 1984, Goulder, 1994; Goulder, 1995a, Goulder, 1995b, Goulder et al., 1997; Goulder et al., 1999). Borgess and Goulder (1984) is a disaggregate model of 24 sectors developed for identifying direct, dynamic and terms of trade components of the impact of energy on the long-run growth. In addition, there were 12 household types and as the main feature of the model, production accounted for the possibility of substituting other factors for energy as relative prices changed.

The major criticism against the Harberger, Scarf, and Shoven and Whally (HSSW) type of models is their simplifying assumption of perfect competition and absence of rigidity and uncertainty. However, the HSSW models have become popular because of their intuitive appeal for 'putting numbers on theory'. In addition, the HSSW models are transparent and consistent with basic economic theory, and have proven useful for conducting welfare analyses focused on the efficiency and distributional effects of various economic policy measures (Bergman and Henrekson, 2003).

Many of the general equilibrium energy models could easily be redesigned for analysis of carbon taxation and other types of climate policies (Bergman and Henrekson, 2003). For example, Thompson (2000) analyses the theoretical link between energy taxes, production and income distribution. The study showed that energy taxes cause adjustment in production through two channels: (i) factor intensity, whereby the relative inputs of productive factors change across sectors and (ii) factor substitution, whereby firms switch between productive factors as relative prices change.

Thompson (2000) concluded that energy tax lowers the supply price of energy with the resulting income distribution among factors depending on factor intensity and income. In particular, the conclusion was reached based on two extreme cases: (i) if energy is an extreme factor in the factor intensity ranking, energy tax raises the return to other extreme factor(s) and lowers the return to the middle factor, while (ii) if energy is a middle factor, energy tax lowers the return to every factor. The case of small open economies is particularly interesting for economic growth implications as Thompson (2000) concludes that energy tariff lowers energy imports and has the potential of lowering wages.

Most studies find that environmental taxes typically aggravate pre-existing tax distortions by raising the cost of pollution abatement (Bovenberg and Goulder, 1996; Kim, 2002, Goulder, 1995a; Goulder, 1995b; Boyd and Ibarraran, 2002). In particular, when pollution costs are treated as extra expenditures necessary to produce the same level of valued output, but as income for the environmental regulator, outputs will become more expensive for consumers, hence the economy may experience declining real wages over time. Traditionally, declining real wage may imply declining productivity of labour when it has less capital to work with. Since savings are linked to income, the lower real wages result in less capital formation, and therefore sluggish economic growth (EPA, 1999).

Copeland and Taylor (1999) criticise the standard economic approach to trade and environment for failing to account for feedback effects between pollution and productivity in the economy. The result that environmental compliance is costly is usually driven by the assumption that pollution is harmful only because consumers suffer a disutility cost from pollution. Copeland and Taylor (1999) argue that if pollution also affects productivity, then it can jeopardize long-run sustainability and lower the competitiveness of environmentally sensitive industries. In a related argument, Mestelman (1986) demonstrated that when the negative effects of production are internalised with the use of a Pigouvian corrective tax, the optimal output of the representative firm in the polluting industry will be the same as the *status quo* if the firm's production function is homothetic. Hence, in the presence of other distortionary taxes, environmental regulatory instruments tend to compound those pre-existing distortions, a cost that is recognised as "tax interactions" or "interdependency effects" (Kim, 2002).

In response to the controversy surrounding the handling of feedback effects, Markandya (2001) suggests that the research issue is really one of adopting more sophisticated models to study the incidence effects of policy measures, especially when the policies affect a wide range of industries and result in a number of relative price changes. In such cases, a general equilibrium model is critical to accounting for the feedback effects of pollution even when such feedbacks are limited to the inter-industry dependence alone without considering the economy-environment nexus. In addition, such studies would in most cases conclude that in the presence of other taxes, the second-best optimal pollution tax lies below the Pigouvian level (Bovenberg and Goulder, 1996; Oates, 1995). Also, since pollution is highly correlated with the use of particular inputs (for instance biomass and fossil energy) in the production process, its abatement cost would depend on the substitution possibilities among inputs or other adjustments in production process (Kim, 2002).

Related to the issues raised by Markandya (2001) and Kim (2002) concerning appropriateness of modelling pollution abatement activities, Klepper and Peterson (2006) note that in most general equilibrium studies, abatement activities are ignored because of the gap that exists in scientific representation of, for example, carbon sequestration technologies. However, under certain conditions, and for selected emissions, it is still possible to define marginal abatement cost curves (MACCs) in general equilibrium where abatement level influences energy prices and in turn national MACCs. In their framework, Klepper and Peterson (2006) define marginal abatement cost (MAC) as the shadow cost that is produced by a constraint on carbon dioxide emissions for a given industry (or region) and a given time, or a tax that would have to be levied on emissions to achieve a target level, or a price of an emission permit in the case of emission trading.

There are however problems with political and economic implications of environmental taxes especially in developing countries. In particular, there are concerns that the introduction of an environmental tax would exacerbate existing distortions in the tax system (Bovenberg and Goulder, 1996). In addition, because of thin tax bases the introduction of environmental taxes in developing countries would necessitate revenue reforms aimed at eliminating distortionary taxes on income. In that regard, CGE modellers debate whether environmental taxes should or should not be revenue- neutral (i.e., reducing other tax rates so that the overall tax revenues remain constant). The related issue is whether or not there exists a “double dividend,” i.e., that environmental taxes result in not only a better environmental quality, but also a less distortionary tax system, thereby improving economic welfare.

Addressing environmental concerns in the context of a changing economy may also result in ambiguous projections of impacts. Most developing countries carried out significant structural reforms after the oil price crises of the 1970s and the subsequent debt crises of the late 1980s. Taeh and Holmoy (2003) found that trade reforms may cause a structural change in favour of heavy polluting export industries when exports prices increase over time. Environmental regulation may cause structural shifts due to changes in relative factor prices (costs to firms) and relative prices of output. These changes may lead to perverted environmental scenario, worse than the distortion the policy was meant to correct.

Thus, tax reforms aimed at incorporating environmental concerns would have to consider the efficiency and distribution effects of such reforms. The imposition of a tax on an activity will, in general, reduce welfare of the taxpayer. The issue that arises is how increases in marginal tax rate influence actions of economic agents. Some taxes are particularly distortionary because they impose a burden over and above the revenue that they are supposed to raise. Widmalm (2001) finds that the proportion of tax revenue raised by taxing personal income has a negative correlation with economic growth. As pointed out above, policy makers must contend with the finding in literature that environmental taxes typically aggravate pre-existing tax distortions by raising the cost of pollution abatement (Bovenberg and Goulder, 1996; Kim, 2002).

CGE models have variously been used in search for optimal taxation and in analysing tax reforms in the presence of externalities in a second-best framework (Mayeres and Regemorter, 2003; Bovenberg and Goulder, 1996; Jorgenson and Wilcoxen, 1993). The feedback effects of an environmental tax depend on how the tax affects households and firms. A progressive income tax is often imposed to correct the distortion caused by the initial distribution of wealth, and market power. In a CGE, the total welfare effect of a tax reform may be measured by the change in total utility. For example, Goulder (1994) tested the "double dividend" proposition of an environmental tax, i.e., that environmental tax not only improves the environment but also reduce the non-environmental costs (deadweight loss) of the tax system. The results from the study validated the theoretical insight that taxes on intermediate inputs cause larger welfare costs through distortions in labour and capital markets in addition to the effect on the input. The double dividend is examined from exploitation of existing tax wedges in the labour market and between consumption and saving. The size of the inefficiency costs in the existing taxes determines the prospect for a double dividend when an environmental tax reform is introduced (Bye, 2000).

Van Heerden et al. (2006) used a CGE for South Africa to assess the potential for triple dividend, i.e., reduction in carbon emissions, increase in GDP and reduction in poverty by recycling environmental taxes. The study focuses on energy-related emissions as about 94 percent of South Africa's electricity generation is coal-fired. In a related study, Blignaut et al. (2005) used a national energy balance to compile a greenhouse gas emission database using sector-by-sector consumption figures. The results showed that electricity generation sector contributes almost 51 percent of the emissions. South Africa's carbon emissions are between that of upper-middle income and the high income countries' at 7.4 metric tons per capita. However, South Africa is a non-annex I country according to the Kyoto Protocol on climate change.

In view of previous results that South African energy demand is complementary to capital while energy production is complementary to capital and labour (Blignaut and de Wet, 2001), Van Heerden et al. (2006) concluded that the absence of energy taxes provided an opportunity for exploring a double or even triple dividend. In particular, because of non-existence of energy taxes, a reduction in energy demand through the introduction of energy taxes would not lead to a fall in tax revenue directly. For South Africa however, a triple dividend was achieved when any of the proposed environmental taxes was recycled through reduction in food prices.

The "double dividend" hypothesis is criticized on several counts. First and foremost, environmental taxes have been shown to exacerbate, rather than alleviate pre-existing tax distortions (Bovenberg and Goulder, 1996). Second, the existence of a double dividend should not be taken as a principle, but rather left to empirical investigation. In most studies, the effects of tax and subsidy reforms are evaluated jointly by believing in advance that a double dividend exists. Third, while removing distorting subsidies and taxes may result in environmental and welfare gains, generalizations of the double dividend results are invalid to the extent that countries differ considerably in tax structure and factor markets (Miller et al., 2002).

3.3 Energy as consumer good

The standard neoclassical approach to explaining consumer behaviour can be used to study household demand for energy goods. In particular, consumers may be assumed to choose a fuel or a fuel-mix bundle that maximizes their utility subject to a bounded endowment set. However, energy consumed by households is a function of some underlying demand for a durable good service such a heating, lighting, refrigeration, or powering home equipment. Therefore, household energy demand and demand for energy-using household durable stocks such refrigerators, cookers, and entertainment units are weakly separable (Baker et al., 1989; Bernard et al., 1996).

Household energy demand can also be factored into two components representing efficiency of some type of energy-using capital equipment and the level of utilization of that equipment (Cameron, 1985; Biesiot and Norman, 1999). In general, a household's utility over energy and non-energy goods can thus be expressed as:

$$V = V\left[U\left(E\right), c\right] \tag{3.12}$$

Where E is a vector of energy goods, and c is a vector of all consumption goods, excluding energy-using stocks.

Utility is maximized subject to a budget constraint defined by household endowment of resources including labour, land and property. Baker et al. (1989) consider households that first allocate resources between energy and non-energy products, and then decide how to divide the total energy outlay among different fuels. Modern fuels such as electricity, kerosene and petroleum fuels have associated fixed costs (e.g., connection cost of electricity) and consumption-dependent charges. Biomass fuels collected or produced by the household itself carry the opportunity cost of time spent collecting fuelwood, or the opportunity cost of dung converted to energy that could have been used as manure to replenish soil nutrients (Heltberg et al., 2000; Heltberg, 2005). Households that obtain fuel from markets face market energy prices as a decision parameter, while those that collect or produce own biomass face a reservation price for biomass as determined by biomass availability and the opportunity cost of collection labour (Heltberg et al., 2000).

3.4 Approaches to studying household energy substitution

Various approaches have been used to study the substitution between different energy sources at household level. The most prominent approaches are the household production framework in which demand for fuel is a function of an underlying demand for services from household durables that use energy and the random utility framework in which fuel choices at household level are modelled using a multinomial logit model.

3.4.1 Household production framework

Household production satisfies basic services such as provision of food, shelter and clothing. Some of these services are produced using market goods (inputs) while others are produced using own labour and open access resources. In most cases the products of household production are tradable in nature although they are neither sold nor bought by members of the household. Households maximize utility by allocating optimal amounts of labour to different home production tasks and by purchasing market goods (inputs) subject to a broadly defined income constraint that includes own labour and endowments (Bandyopadhyay et al., 2006).

In the context of a developed country, Baker et al. (1989) specified consumer demand for fuels within a household production framework where the underlying demands are for services from energy-using capital equipment. In their model, Baker et al., allow the marginal rate of substitution across disaggregated energy demands to differ across households with different durable stocks, hence making energy demand non-separable from the stocks. In a related study, Vaage (2000) describes energy demand as a combination of discrete and continuous choice problems. In the first instance, household appliance choice is specified as a multinomial logit model with a mixture of appliance attributes and household's own characteristics. Then energy use is modelled conditional on the appliance choice. Thus, energy use depends on utilization of a given stock of energy-using appliances just like in Cameron (1985) and Biesiot and Norman (1999).

Boonekamp (2007) used a simulation model to analyse the relationship between historic energy prices, policy measures and household energy consumption. Household energy consumption was divided into seven energy functions: space heating, supply of hot water, cleansing (e.g. washing machines), cooling, cooking, lighting and other appliances. Like in Baker et al. (1989), demand for each energy function is met by one or more energy consuming systems or appliances, and for every system or appliance, total energy consumption is defined by three factors: ownership, intensity of use and efficiency of the system or appliance.

Although the household production framework is theoretically sound and quite useful in developed countries and other applications, the model has limited use for analysing energy demand in developing countries. In particular, most households' energy choices in developing countries have radically different structures than those presented by Baker et al. (1989), Vaage (2000), and Boonekamp (2007). In particular, because of widespread poverty, ownership of energy-using capital stock or appliances is low and hence would not explain much of households' energy demand. For instance, Vaage (2000) found that high income households tend to choose electricity as the only heating energy source while solid fuels such as fuelwood were unpopular. Thus, using these studies, one would conclude that low income households use fuelwood either because of lack of energy-using capital stock or because of low income.

Heltberg et al. (2000) also used a household energy production framework to estimate demand for fuelwood in rural India. The focus of the study, however, was on the substitution between non-commercial fuels a household obtains from open access sources (commons) and fuels obtained from the energy market. Elasticities were obtained from maximum entropy regression estimates of fuelwood collection, collection labour time and private energy consumption. The major result was that households respond to fuelwood scarcity and increased fuelwood collection time by substituting commercial fuels for forest fuelwood. However, the substitution rate was deemed too low to prevent current fuelwood collection from causing serious forest degradation. The other weakness of the model was that it was practically impossible to endogenize factors driving household choices between fuelwood from open access sources and commercial fuels in fulfilling a particular household function.

In a study of Zimbabwean households, Campbell et al. (2003) used two surveys of fuel use by low income households to describe energy transition from wood to electricity by means of a series of chi-square tests. Although the methodology is not similar to the household production framework, the underlying hypothesis is very close to assuming that households demand energy as a result of ownership of appliances. In Campbell et al. (2003) households were faced with an array of energy choices arranged in order of increasing technological sophistication. Using such an ordering, also called an “energy ladder”, households were hypothesized to make the transition from biomass fuels through kerosene to Liquid Petroleum Gas (LPG) and electricity, with the corresponding reduction of pressure on woody plant resources that form the bulk of biomass energy sources.

Campbell et al. (2003) accepted the energy ladder hypothesis, with income as the main determinant. About 3 percent of households switched to electricity from other fuels, citing as their main reasons the acquisition of a new appliance that required electricity and moving to new premises. However, other households did not use electricity because of lack of access (5percent) while the majority (51percent) of the households cited price as a deterrent. The use of wood for cooking ranged from 1.5 tons/year per household in 1994 to 0.7 tons/year in 1999. The study also concluded that most households use mixtures of fuels but failed to prove that the fuel stack varied over time. Fuel security was offered as an explanation to fuel stacking behaviour in response to insufficient or unreliable electricity supply. In addition, the proportions of fuels in household energy budgets were driven by price considerations for not only the fuels but also complementary appliances.

There are a number of challenges to the "energy ladder" hypothesis. First, widespread poverty in developing countries may lead to the conclusion that the energy ladder is nonexistent because proportionately large number of households are perpetually unable to afford other sources of energy apart from collecting biomass from open access sources and own fields. Second, households tend to use more than one fuel at a time, thus the transition process is not from exclusive use of one fuel to exclusive use of another, but from one fuel combination to another (Hosier and Dowd, 1987). Thus, the "energy ladder" hypothesis ought to be phrased in terms of proportion of biomass fuels in household energy compared with electricity over time. Also, there is need to identify the determinants of household fuel preferences and why households use one fuel or multiple fuels to fulfil a single household function.

3.4.2 Random utility framework and multiple fuels

The functional form of the utility from energy goods aggregate $U(E)$ in equation 3.12 may be specified as follows. Let U_{hj} be the indirect utility a household h obtains from acquiring fuel j. For a given set of K-energy sources (or just fuels), a typical household would consume zero or more fuels depending on the fuels' unique attributes which include the total economic cost of obtaining the fuel. Since some households obtain fuels from open sources, the total economic cost of energy consumed by a typical household is unobservable, but can be estimated from a random utility framework based on an indirect utility of the form:

$$U_{hj} = U\left(F_j^H, \xi_j, e, \tau_j; \theta\right) = \alpha_h y_h + F_j^H \beta - e_j \alpha + \xi_j + \mu_{hj}\left(F_j^H, e, \tau_j; \theta_2\right) \qquad (3.13)$$

Where α_h is household $h's$ marginal utility from income, F_j^H and ξ_j are observed and unobserved fuel characteristics, respectively, e is a vector of energy prices, τ_j is a vector of household characteristics influencing preferences over fuel j, and $\theta = (\theta_1, \theta_2)$ is a vector of unknown parameters. The last term $\mu_{hj}(\cdot)$ represents zero mean but heteroskedastic error term.

Following Nevo (2000), if the error term $\mu_{hj}(\cdot)$ is independently and identically distributed (i.i.d.) following a Type I extreme value distribution, equation 3.13 reduces to a standard logit model where the share of fuel j in household aggregate energy expenditure is:

$$s_j = \frac{\exp\left(F_j\beta - e_j\alpha + \xi_j\right)}{1 + \sum_{j=1}^{K}\exp\left(F_j\beta - e_j\alpha + \xi_j\right)} \tag{3.14}$$

Ouedraogo (2006) used a multinomial logit model to analyse factors determining household energy choices in urban Ouagadougou. The data and empirical analysis show that the actual (predicted) probability of a household adopting fuelwood as main cooking energy is 79.1 percent(92.2percent), and for kerosene is 2.7 percent(0.0percent). Household income was not significant for explaining demand for firewood probably because firewood users were the poorest households in Ouagadougou. It was also found that high costs of modern cooking energy and their capital stock requirements like cooking stoves are constraints for household fuel preferences.

Heltberg (2005) used the 2000 Guatemalan household survey to analyse patterns of fuel use, energy spending, Engel curves, multiple fuels (fuel stacking) and the extent of fuel switching. A significant share of fuelwood users were incurring more costs acquiring fuelwood from markets compared with the costs of modern fuels. The evidence also suggests that the widespread collection of firewood in rural areas is due to the low opportunity cost of labour time. Thus, rising labour cost may be the only factor capable of effectively regulating firewood supply from open access forests and commons.

Heltberg (2005) also estimated Engel curve regressions for LPG and firewood. It was found that prices were important for interfuel substitution although many households were using multiple fuels (fuel stacking) for cooking. Thus, for low income countries, fuel switching policies should be guided by determinants of not only fuel substitution but also factors that drive fuel complementarities. By employing a multinomial logit analysis of all possible fuel choices, Heltberg (2005) finds that education is a strong determinant of fuel switching from fuelwood to LPG while having electricity is associated with fuel switching by inter alia, being associated with smaller probability of using only wood, or only LPG.

The problem with the multinomial logit model used by Heltberg (2005) is that it excludes from the estimation households that collect firewood (sample selection bias), yet the opportunity cost of labour collection time is an important determinant of fuel substitution/complementarities for rural households. Econometrically, the main weakness of the multinomial logit model is the i.i.d. assumption. The assumption implies that the cross-price elasticities of demand do not depend on observed fuel differences (Besanko et al., 1989) and that own-price elasticities are proportional to own price (Nevo, 2000).

As a solution, Besanko et al. (1989) suggest using the generalized extreme value (GEV) or nested logit structure. The idea of nesting is to induce correlation among fuel options by grouping all fuels used by households into predetermined exhaustive and mutually exclusive sets (Nevo, 2000). The other solution is to use a segment specific dummy variable as one of the characteristics of the fuels under consideration. According to Nevo (2000), this is equivalent to estimating the multinomial system with the group specific dummy variable acting as one of the characteristics of the fuels.

3.5 Chapter summary

This chapter reviewed the literature on approaches to studying energy demand and fuel switching policies. The chapter also discussed both the perspective of energy as a factor input in production and as a consumer good that enters household utility functions either directly or indirectly.

For energy as an input in production, the reviewed studies can be categorized into two main groups: (i) those that focus on factor intensities and (ii) those that focus on energy switching, or factor substitution. In the first category, there are descriptive nonparametric approaches for which DEA is the main tool of analysis and parametric partial equilibrium (energy sector) approaches for which regression analysis is used. For the second category, both regression based approaches and AGE with or without regression estimates of substitution elasticities have been reviewed.

For sectoral energy intensities and related questions of energy efficiency, the DEA offers invaluable insights that could be used to foster energy efficiency as an environmental policy objective. However, the review has shown that technical progress is exogenous to the DEA system, hence limiting its use in energy switching studies. Similarly, for AGE models, the chapter has indicated that the focus of many energy studies is now shifting to examining the impact of emission reduction on energy prices following the Kyoto Protocol. Most of the literature on AGE impacts of Kyoto Protocol, such as the reduction in carbon dioxide emissions, is developing on the premise that meeting emission targets is the only policy objective that could be followed, although others, for example Otto et al (2006) and Smulders and de Nooij (2003) have tried to endogenize technical progress.

The literature review has therefore shown that energy intensity and factor substitution are important for the efficacy of energy policy. However the literature offers no consensus on the appropriate technological specification for substitution possibilities between energy and non-energy factor inputs. Similarly, the literature does not offer much agreement on the appropriate delineation of energy biased technical progress. From the foregoing, and considering objectives of this study, there is need to integrate approaches that focus on factor intensities and substitution possibilities on one hand, and those that seek to meet environmental targets.

For household energy demand, the literature from developed countries seem to advocate the household framework since household energy demand in those countries is intricately related to ownership and utilization of energy-using appliances. The review highlighted the inappropriateness of the household production framework similar to that of Baker et al. (1989) in studying household energy demand in developing countries where the majority of the households are rural based and poor. The random utility framework came out as a viable approach for analysing household fuel choices and for identifying and quantifying factors that determine the choice of one set of fuels from another in fulfilling a household function.