The Applied Theory of Energy Substitution in Production

Henry Thompson

Energy Economics 2006

Auburn University

This paper reviews the applied theory of energy cross price partial elasticities of substitution, and presents it in a transparent fashion. It uses log linear and translog production and cost functions due to their economic properties and convenient estimating forms, but the theory applies other functional forms. The objective is to encourage increased empirical research that would deepen understanding and appreciation of energy substitution.

The economic outcome of decisions regarding energy policy often hinge on substitution between energy and other factors of production, but there is little consensus regarding the degree and even direction of energy substitution. As classic examples, Berndt and Wood (1975) find aggregate energy a substitute for labor but a complement with capital, while Griffin and Gregory (1976) find energy a substitute for both. Only continued theoretical and empirical investigation will be able to shed light on this issue, and the opaque dispersed literature on the theory of substitution must discourage new research. The empirical literature on energy cross price elasticities is thin relative to the economic impact of energy substitution.

The present paper presents a transparent theory of energy substitution using log linear (Cobb Douglas) and translog production and cost functions. The focus is on the derivation of cross price elasticities that describe the substitution of inputs due to a change in the price of an input. For instance, cross price elasticities describe the adjustment in capital and labor inputs to a higher price of energy, or the adjustment in fuel mix due to a higher price of oil. A wide range of policy issues and applications provide ample motivation for continued research in applied energy substitution.

1. A review of the literature on energy substitution elasticities

The economics of substitution has it roots in microeconomic production theory. Allen (1938) remains a fundamental source, supplemented by Varian (1984) and Takayama (1993). Beattie and Taylor (1985) and Chambers (1988) provide concise introductions to applied production theory. Ferguson and Pfouts (1962) and Berndt and Christensen (1973) develop the theoretical background of applied substitution in production. Sato and Koizumi (1973) clearly develop the link between Allen relative and cross price elasticities. Atkinson (1975) develops the properties of revenue shares of productive factors. The present paper synthesizes a good deal of this theoretical literature.

Estimates of cross price substitution are sensitive to the industries and regions of study. Cameron and Schwartz (1980), Field and Gerbenstein (1980), and Denny, Fuss and Waverman (1981) find differences in estimated energy substitution across industries and countries. Walton (1981) finds differences in substitution across US industries, and perhaps across regions. Burney and Al-Matrouk (1996) find substitution between energy and capital in electricity generation and water production in Kuwait. Caloghiro, Mourelatos, and Thompson (1997) find electricity a weak substitute for capital and labor in Greek manufacturing during the 1980s, implying electricity subsidies lowered the demand for capital and labor. Bamett, Reutter, and Thompson (1998) show that electricity is a weak substitute for both capital and labor in major Alabama industries, and note that regulatory constraints are binding due to the inelastic demand for electricity. Kemfert (1998) reports that aggregate energy, capital, and labor are substitutes in German manufacturing. In Pakistani manufacturing, Mahmud (2000) finds very little substitution between aggregate energy and other inputs but weak substitution between electricity and gas.

The choice of functional form may affect estimated cross price elasticities, but Chang (1994) finds little difference between translog and constant elasticity production functions in Taiwanese manufacturing, and reports that energy and capital are substitutes. The present paper uses log linear and translog production and cost function, but the theory generalizes to other functional forms.

The time period chosen and the dynamic model of substitution are critical. Yi (2000) finds different estimates of substitution with dynamic translog and generalized Leontief production functions across Swedish manufacturing industries. Urga and Walters (2003) show that the specification of dynamic translog functions has an effect on estimates of substitution, and find coal and oil substitutes in US industry. Kuper and van Soest (2003) show that the time period affects estimates of substitution due to path dependencies that arise given fixed costs of input adjustments.

Aggregation distorts the estimates of substitution although there has been no systematic study of its effects on estimated energy substitution. Clark, Hofler, and Thompson (1988) use separability tests to show that "labor" in US manufacturing has no fewer than nine distinct skill groups, and there are no estimates of energy substitution at this disaggregated level. Separability is an issue because energy might be a weak substitute for labor in an estimated model, but would be a complement with transport labor and a strong substitute for production labor in the estimate of a disaggregated model. As another example, energy and capital might be unrelated inputs in an estimated model, but capital would a complement for electricity and a substitute for fuel oil with disaggregated energy. The estimated substitution elasticity involving an aggregate is not necessarily a weighted or any other average of the disaggregated inputs. Aggregation also distorts substitution elasticities between inputs other than those involved in the aggregation.

2. The theoretical background for energy input and substitution

Energy input involves work that moves or transforms matter, and includes a range of fuels based on some natural resource. Homogeneous physical units such as the BTU can combine various energy inputs into an aggregate, or separate units can measure each type of energy for estimates of intra-energy substitution.

Produced energy E in general functional form is $E = E(K, L, N)$ where K is capital input, L is labor, and N is a natural resource input converted into energy. There is a good deal known about the physics of converting various natural resources to energy but this engineering structure is not incorporated into the economics of energy substitution which begins with the production function for subsequent output $x = f(K, L, E(K, L, N))$ embedding the production of energy. Studies of energy substitution typically start with $x = g(K, L, N)$ or simply as in the present paper $x = x(K, L, E)$. Estimates of energy production or the economics of the resource input would use the more primitive functions. The production function can expand to include various types of natural resources, energy, and capital or labor inputs.

The firm or industry is assumed to produce the profit maximizing output x^* hiring the optimal inputs of K, L, and E that minimize cost of production. The model assumes competitive price taking in the input markets, as well as the output market. Evidence of competition in factor markets is widespread, and introducing market power complicates but does not necessarily alter intuition. The basic results concern the comparative static substitution between energy and the other inputs, given cost minimization.

The present paper disregards material inputs and intermediate products that enter in fixed proportions. Production data generally report value added, netting out spending on intermediate products from revenue. Manufacturing data may not include energy input but energy spending and input may be available from other sources. Disaggregated data is much more suitable for estimation, and there is no substitute for knowing something about the industry under study.

Capital input K in theory is homogeneous machine hours. More disaggregated capital data is preferred due to the distorting effects of aggregation but there is a general lack of capital data. In practice, capital input is often a derived residual after subtracting labor and energy bills from value added. In such residual calculations, capital implicitly includes land and perhaps other inputs. In industrial data, there may be separate information on capital plant, equipment, and structures. In detailed industrial data, such as electricity generation and distribution, capital may include a variety of machinery and vintages of equipment.

Labor input L is the number of person hours hired by the firm or industry. Labor is homogeneous in the KLE model but in fact labor comes in a variety of skills and perhaps some types of labor specific to their industry. Labor aggregation affects estimates of substitution, and disaggregated labor data increases reliability of estimates.

3. Log linear production and cross price substitution elasticities

Cobb-Douglas log linear production (LLP) is a useful starting place because of its economic and estimating properties, and economic properties including concavity depend on the estimated parameters. The LLP function in exponential form is

$$x = A(t)K^{\alpha}L^{\beta}E^{\gamma}. \qquad (1)$$

The three parameters α, β, and γ would be less than one with diminishing marginal productivity.

The technology variable A(t) has a positive derivative $dA/dt > 0$ indicating that with a higher level of technology t the same inputs produce more output. In applications, there is typically no separate variable t and it is difficult to generate a measure of technology because capital machinery and equipment embed it. Estimates of total factor productivity arise from this ambiguity. The analysis of technical progress is very challenging and beyond the scope of the present paper, but technical progress relates to the issue of energy substitution as developed by Hunt (1986).

Efficient inputs minimize the cost of producing output on the three dimensional isoquant bowl of the KLE production function, and changes in input prices lead to input substitution. The shortcomings of LLP are that inputs are necessarily substitutes and the relative elasticities of substitution between the three pairs of inputs have unit value.

Transforming (1) into log linear form,

$$\ln x = \ln A(t) + \alpha \ln K + \beta \ln L + \gamma \ln E . \qquad (2)$$

Compared to a regression in levels, estimating with logarithms generally improves statistical properties. A zero constant term in (2) implies $A(t) = 1$. With a change in energy input, holding other inputs and technology t constant, the partial output elasticity of energy input is

$$\delta \ln x/\delta \ln E = \gamma . \qquad (3)$$

Competitive input markets are the rule for all but highly specialized industry inputs, and each input receives its marginal revenue product with competition. Assume the output market is also competitive with output price equal to marginal revenue, and for simplicity scale the price p of output to unit value. The marginal revenue product of energy is then $pMP_E = MP_E$, and

$$e = MP_E = \delta x/\delta E = \gamma AK^{\alpha}L^{\beta}E^{\gamma-1} , \qquad (4)$$

where e is the price of energy. Diminishing marginal productivity holds if $\gamma < 1$ since $\delta MP_E/\delta E = (\gamma - 1)\gamma AK^{\alpha}L^{\beta}E^{\gamma-2}$. The energy factor share is

$$\theta_E \equiv eE/x . \qquad (5)$$

Substituting (4) and (1) into (5),

$$\theta_E = (\gamma AK^{\alpha}L^{\beta}E^{\gamma-1})E/AK^{\alpha}L^{\beta}E^{\gamma} = \gamma . \qquad (6)$$

Similarly, $\theta_K = \alpha$ and $\theta_L = \beta$. Estimated coefficients are factor shares, a convenient property of LLP. If the null hypothesis $\alpha + \beta + \gamma = 1$ cannot be rejected, there are constant returns to scale (CRS) and $\Sigma_i\theta_i = 1$. With CRS, proportional changes in every input proportionally change output. Imposing CRS through parameter restrictions typically improves estimation properties.

There are increasing returns to scale (IRS) if the sum of the three exponents in (2) is greater than one. Proportional increases in inputs then cause more than a proportional increase in output, a pleasant situation to say the least. Industry would proportionally raise all inputs and cost, but output would increase by a larger percentage. Factor shares would be shares of cost rather than output, consistent with the excess profit. Regarding industrial organization, IRS is consistent with imperfect competition.

There are decreasing returns to scale (DRS) if the sum of the estimated coefficients in (2) is less than one. Industry could lower all inputs and total cost would fall by a lower percentage than output. An industry with DRS is producing too much, and changes in model specification might capture any negative externalities.

IRS and DRS are inconsistent with the present theory of competition and cost minimization, and evidence of variable returns requires more careful modeling of industrial organization. Production functions at the firm level may include industry output, and estimates of returns to scale may provide some hints about market structure and performance. On the face of it, CRS is a sensible expectation for mature free market firms and industries. In estimates of production or cost functions, imposing CRS typically improves statistical properties.

4. Relative substitution elasticities

The Allen (1938) relative substitution elasticity (RSE) measures the responsiveness of relative inputs to relative input prices. In the history of thought, the RSE has the predominant role in the economics of substitution because theory developed with two inputs and microeconomics stresses the importance of relative prices. The RSE between factors i and j is the percentage change in the relative input of factor i due to a change in the relative price of factor j,

$$\alpha_{ij} \equiv \delta \ln(a_i/a_j)/\delta \ln(w_j/w_i) \,, \tag{7}$$

where the a's represent cost minimizing inputs per unit of output and the w's factor prices.

With three or more inputs, the partial derivative α_{ij} is an artificial construction that holds all other inputs constant, while in fact all inputs adjust to any change in factor prices. In the KLE model with a focus on energy, the RSE has little relevance except perhaps in describing the short run with capital input and the price of capital constant. Even in that special circumstance, there might be little apparent reason to focus on the relative price of energy and labor. With more than three inputs, the RSE has no apparent economic interpretation.

In the short run model, hold K constant in (1) and scale AK^{α} to unit value. The short run production function is then $x = L^{\beta}E^{\gamma}$ with capital input fixed at $K = A^{-\alpha/A}$. The short run unit isoquant where $x = 1$ and $L = E^{-\gamma\beta}$ has slope $dL/dE = -\gamma/\beta E^{-(\gamma+\beta)/\beta}$. Along this short run unit isoquant, $0 = dx = MP_L dL + MP_E dE$ implying $dL/dE = -MP_E/MP_L = -\gamma L/\beta E$, the short run marginal rate of substitution of labor for energy.

The isocost surface shows all combinations of L and E with the same cost $c = rK + wL + eE$. Short run variable cost is $c_{SR} = wL + eE$. The wage w and energy price e are exogenous in competitive markets and the short run isocost line is $L = (c_{SR}/w) – (e/w)E$ with slope $-e/w$ and intercept term c_{SR}/w. To minimize cost, the firm chooses the lowest possible isocost line tangent to the unit isoquant $x = 1 = L^{\beta}E^{\gamma}$.

Solve this constrained short run cost minimization starting with the Lagrangian function

$$\Lambda = wL + eE + \lambda(1 - L^{\beta}E^{\gamma}) . \qquad (8)$$

The Lagrangian multiplier $\lambda = \delta\Lambda/\delta x^*$ is the marginal cost of output, and first order conditions are

$$\delta\Lambda/\delta L = w - \lambda(\beta L^{\beta-1}E^{\gamma}) = 0$$

$$\delta\Lambda/\delta E = e - \lambda(\gamma L^{\beta}E^{\gamma-1}) = 0 \qquad (9)$$

$$\delta\Lambda/\delta\lambda = 1 - L^{\beta}E^{\gamma} = 0 .$$

The first two equations in (9) imply the familiar first order condition $e/w = -\gamma L/\beta E = MRS$, and the third ensures unit production. Solve the system (9) for optimal short run labor and energy inputs. A higher relative price of energy e/w would increase labor relative to energy input. With LLP, input changes exactly offset input price changes and factor shares remain constant.

The Allen (1938) RSE α_{LE} holds K input constant relates to the short run cross price elasticities. Thompson and Taylor (1995) provide a glimpse into the role RSEs may play in describing energy substitution. The LLP has a constant RSE regardless of relative factor prices and output. If the price of energy e rises by some percentage holding w constant, the relative price of energy e/w rises by that same percentage. Since $L/E = -(\beta/\gamma)(e/w)$, L/E rises by the same percentage. With log linear production, $a_{LE} =$

1 anywhere along the unit isoquant and regardless of returns to scale. This condition might be overly restrictive for estimation, and the short run translog production function allows variation in the RSE.

The RSE reflects convexity of the production isoquant, and the theoretical extremes of no substitution and perfect substitution provide benchmarks. With fixed proportions, there is an L-shaped unit isoquant with two inputs and the hinge is the minimum input to produce a unit of output. Adding more of either L or E (but not both) would leave output unchanged due to the fixed proportions. With no substitution, the RSE is zero since the firm minimizes its cost of production with the same input ratio regardless of the relative input price.

At the other extreme, with perfect substitutes there is a linear isoquant. Let φ represent the slope of the unit isoquant. If $e/w < \varphi$ energy is relatively cheap and no labor is hired. If $e/w > \varphi$ no energy is used. The RSE is infinite in the neighborhood of p, not an issue in practice since complete switching does not occur in data. A more convex isoquant reflects less ability to substitute and convexity of the isoquant relates directly to convexity of the production frontier in general equilibrium, an issue explored graphically by Ford and Thompson (1997).

Moving to the long run, capital input would adjust to any change in e/w and there would be further adjustment in L/E. In the data, all inputs must adjust to a factor price change. Even with only three inputs in the KLE model, there is little motivation to report an RSE. Blackorby and Russell (1981) compare the various summary RSEs in the literature. Thompson (1997) shows RSEs can be misleading and proposes an alternative that at least preserves the sign of the underlying cross price elasticities. Even in the two input case, cross price elasticities provide more information than any summary RSE since flexibility in substitution may not be symmetric. The RSE is convenient for exposition but perhaps misleading and inadequate for issues in applied energy economics. Cross price elasticities, at any rate, would be estimated to build an RSE.

5. Energy cross price substitution elasticities

The firm or industry is assumed to hire inputs to minimize cost c = rK + wL + eE with a focus on unit production and average cost. For a competitive firm, its price p equals marginal cost MC and with unit price p = 1 = MC = c. The familiar first order conditions are that marginal rates of substitution equal input price ratios. The Lagrangian cost minimization problem subject to producing a unit of output introduces the Lagrangian multiplier λ and leads to the first order conditions $\lambda = r/x_K = w/x_L = e/x_E = 1$ where $x_i \equiv \delta x/\delta v_i$ is the marginal product of factor v_i. Effects of changes in factor prices on inputs then relate directly to the production function according to

$$\begin{aligned} dr &= d(\lambda x_K) = x_K d\lambda + \lambda dx_K = x_K d\lambda + dx_K \\ dw &= x_L d\lambda + dx_L \\ de &= x_E d\lambda + dx_E \,. \end{aligned} \tag{10}$$

Each marginal product is a function of all input levels, and any change in the marginal product of energy is due to input adjustment, $dx_E = x_{EK}dK + x_{EL}dL + x_{EE}dE$ where $x_{ij} \equiv \delta x_i/\delta v_j$. Similar expressions hold for dx_K and dx_L, and (10) becomes

$$\begin{aligned} dr &= x_K d\lambda + x_{KK}dK + x_{KL}dL + x_{LE}dE \\ dw &= x_L d\lambda + x_{LK}dK + x_{LL}dL + x_{LE}dE \\ de &= x_E d\lambda + x_{EK}dK + x_{EL}dL + x_{EE}dE \,. \end{aligned} \tag{11}$$

Along the unit isoquant $0 = dx = x_K dK + x_L dL + x_E dE$ which with (11) yields the cost minimizing comparative static system

$$\begin{pmatrix} 0 & x_K & x_L & x_E \\ x_K & x_{KK} & x_{KL} & x_{KE} \\ x_L & x_{LK} & x_{LL} & x_{LE} \\ x_E & x_{EK} & x_{EL} & x_{EE} \end{pmatrix} \begin{bmatrix} \delta\lambda \\ \delta K \\ \delta L \\ \delta E \end{bmatrix} = \begin{bmatrix} 0 \\ \delta r \\ \delta w \\ \delta e \end{bmatrix} . \tag{12}$$

The effects of changing input prices on cost minimizing input levels is solved directly in (12) using Cramer's rule. Focusing on energy input,

$$\delta E/\delta e = (-x_L^2 x_{KK} - x_K^2 x_{LL} + 2x_K x_L x_{KL})/H$$

$$\delta E/\delta r = \delta K/\delta e = (x_L^2 x_{KE} + x_K x_E x_{LL} - x_L x_E x_{KL} - x_K x_L x_{LE})/H \qquad (13)$$

$$\delta E/\delta w = \delta L/\delta e = (-x_K^2 x_{LE} + x_K x_E x_{KL} - x_L x_E x_{KK} - x_K x_L x_{KE})/H \ ,$$

where H is the determinant of the bordered Hessian matrix in (12). Partial derivatives in (13) are corresponding elements in the inverse of the system matrix. Diminishing marginal productivity implies $x_{ii} < 0$ for i = K, L, E. Symmetric cross partials x_{KL}, x_{KE}, and x_{LE} may be positive or negative, but additional inputs should typically raise other marginal products. Chang (1979) shows that a negative H is sufficient for concavity of the production function in the present model with three inputs. Own effects on marginal products such as x_{EE} must outweigh cross effects such as x_{LE} and x_{KE}, implying negative own substitution terms such as $\delta E/\delta e$ in (13). These conditions can be imposed in estimation and typical improve statistical properties.

By Shephard's lemma, $E = \delta c/\delta e$ and $\delta E/\delta r = \delta c^2/\delta e \delta r = \delta K/\delta e$. Similar relations hold between L and both K and E. If a higher input price lowers another input, the two are complements. If so, the key is how marginal products react to other inputs. For instance, firms facing a higher price of energy substitute away from energy input. Workers might have to take stairs or ride bicycles, becoming healthier and more productive. With the marginal productivity of labor increasing, firms would substitute away from capital making it a complement with energy.

Cross price elasticities can be used to derive Allen (1938) relative substitution elasticities, RSEs. In the two input model with energy and capital, the cross price partial elasticity is

$$\varepsilon_{KE} = (\delta K/\delta e)(E/K) = \delta \ln K/\delta \ln e \ . \qquad (14)$$

Percentage changes in relative inputs and relative input prices are

$$d\ln(K/E) = (EdK - KdE)/KE$$

$$\mathrm{dln}(e/r) = (r\mathrm{d}e - e\mathrm{d}r)/er\ . \tag{15}$$

Given competitive input markets and unit output price, $r = x_K$ and $e = x_E$, implying $dw = x_{LL}dK + x_{KE}dE$ and $de = x_{EK}dK + x_{EE}dE$. Due to cost minimization, $rdK + edE = 0$ implying $dL = -(e/w)dE = -(x_E/x_L)dE$. It follows from (15) that $\mathrm{dln}(L/E)= -(E(x_E/x_L)dE)/LE = -(x_EE+x_LL)dE/x_LLE = -xdE/x_LLE$ by Euler's theorem, and $\mathrm{dln}(e/w)\ (x_Lde - x_Edw)/x_Lx_E = (x_L(x_{EL}dL + x_{EE}dE) - x_E(x_{LL}dL + x_{LE}dE))/x_Lx_E = ((x_L^2x_{EE} + x_E^2x_{LL} + 2x_Ex_Lx_{EL})dE)/x_L^2x_E = -HdE/x_L^2x_E$ where H is the determinant in (12). Putting the two together, the Allen RSE is the familiar

$$\alpha_{LE} = -xx_L^2x_EdE/x_LHLE = -xx_Lx_E/HLE \tag{16}$$

that directly relates to the cross price elasticity according to

$$\alpha_{LE} = x_Lx_Ex/LEH = (x_Ex_L/H)(x/LE)$$

$$= (\delta L/\delta e)(x/LE) = [dL/L)(de/e)]/(eE/x) = \varepsilon_{LE}/\theta_E\ . \tag{17}$$

The RSEs are symmetric. By Young's theorem, $\delta L/\delta e = \delta(\delta c/\delta w)/\delta e = \delta^2c/\delta w\delta e = \delta(\delta c/\delta e)/\delta w = \delta E/\delta w$ and it follows that $\alpha_{LE} = (\delta L/\delta e)(x/LE) = (\delta E/\delta w)/(x/EL) = \alpha_{EL}$.

Shephard's lemma says the partial derivative of cost with respect to an input price is that input level, $v_i = \delta c/\delta w_i$. To prove Shephard's lemma in the KLE model, start with total cost $c = rK + wL + eE$. Any cost change is accounted for by the total derivative $dc = rdK + wdL + edE + Kdr + Ldw + Ede$. First order conditions from cost minimization are $dv_i/dv_j = -w_j/w_i$ which implies $rdK + wdL + edE = 0$ and $dc = Kdr + Ldw + Ede$. Shephard's lemma follows directly for each input holding other input prices constant: $\delta c/\delta r = K$, $\delta c/\delta w = L$, and $\delta c/\delta e = E$. The implication is that an incremental increase in an input price raises average cost by the level of the input. The approximate change in average cost due to a discrete change in a factor price is the level of input times the change in the factor price. This shift in the average cost curve assumes an underlying cost minimization.

6. Translog production and energy cross price substitution elasticities

Including interaction terms in the log linear production function improves empirical fit and allows pairs of factors to be complements in production. The translog production function (TPF) developed by Christensen, Jorgensen, and Lau (1973) introduces interaction terms and can be estimated in a symmetric system of derived factor share equations that improves estimation properties relative to a single equation. The TPF is

$$\ln x = \ln a_0 + \Sigma_i a_i \ln v_i + .5\Sigma_i \Sigma_j b_{ij} \ln v_i \ln v_j \ , \qquad (18)$$

and in the KLE model

$$\ln x = \ln a_0 + a_K \ln K + a_L \ln L + a_E \ln E + .5[b_{KK}(\ln K)^2 + b_{LL}(\ln L)^2 + b_{EE}(\ln E)^2$$
$$+ 2b_{KL} \ln K \ln L + 2b_{KE} \ln K \ln E + 2b_{LE} \ln L \ln E] \ . \qquad (19)$$

The first four terms in (19) are the LLP in (2) with a constant technology parameter $a_0 = A$. The remaining terms inside brackets allow for interaction between inputs. The TPF reduces to the LLP if the all of the b_{ij} coefficients equal zero, a null hypothesis to test. Generally, this null hypothesis is not rejected but the derived substitution elasticities from the LLP and the TPF differ very little. The reasons are that the estimated interaction terms in (19) tend to be small relative to the input coefficients and factor shares dominate the derivation of cross price elasticities.

The output elasticity of each input comes from the estimated TPF. For instance, the output elasticity of energy is

$$\delta \ln x/\delta \ln E = a_E + b_{EE} \ln E + b_{KE} \ln K + b_{LE} \ln L \ , \qquad (20)$$

making clear the reason for the .5 weight on the interaction terms in (18). This output elasticity depends on input levels unless $b_{EE} = b_{KE} = b_{LE} = 0$ in the estimate.

If a higher energy price raises the marginal product of labor, the two inputs may be complements. Less of one input should typically lower the productivity of another but the opposite may occur with more than two inputs, and interaction terms in (19) allow investigation of these second order effects.

If energy is paid its marginal product, $e = x_E$, the output elasticity of energy input is the energy factor share,

$$\delta \ln x/\delta \ln E = x_E E/x = eE/x = \theta_E \, . \qquad (21)$$

The same holds for capital and labor, leading to a system of factor share equations,

$$\theta_K = a_K + b_{KK}\ln K + b_{KL}\ln L + b_{KE}\ln E$$
$$\theta_L = a_L + b_{KL}\ln K + b_{LL}\ln L + b_{LE}\ln E \qquad (22)$$
$$\theta_E = a_E + b_{KE}\ln K + b_{LE}\ln L + b_{EE}\ln E \, .$$

Coefficients in (22) are symmetric across equations due to Young's theorem on partial derivatives applied to (19). Simultaneous estimation improves estimation properties over single equations, and imposing the symmetry constraints in (22) typically improves estimates further. It is possible to test the symmetry conditions in an unconstrained estimate as developed by Zellner (1962). Coefficients provide an indication of returns to scale, but imposing CRS typically improves the estimation properties.

Production functions are homothetic if unbiased toward any input as production expands. The production function in (22) is homothetic if $\Sigma_i a_i = 1$ and $\Sigma_i b_{ji} = 0$, a testable null hypothesis. Imposing this homothetic production typically improves estimation properties. If $\Sigma_i b_{ji} \neq 0$, a further direct test of homothetic production is to include output as an independent variable in the factor share equations,

$$\theta_j = a_j + \Sigma_i b_{ji}\ln v_i + \chi_j x \, . \qquad (23)$$

Failure to reject the null hypothesis $\chi_j = 0$ indicates homothetic production. If the coefficient χ_j is significant in the estimated system (23) but all of the b_{ji} terms are insignificant, factor shares are constant along production isoquants implying unit own factor price elasticities.

The price of energy relative to labor e/w is the absolute value of the slope of the LE tangent to the unit production isoquant, for given level of K input. Suppose the price of output increases with e/w unchanged and K constant. Output would increase as the industry increases inputs of L and E. Expansion paths illustrate how inputs expand holding relative input prices constant. With homothetic production, inputs expand along a ray from the origin. Production has a bias toward (away from) energy if the expansion path bends toward (away from) the energy axis. Homothetic production has implications for factor demand as output changes, and may be critical in policy issues.

A change in technology t shifts the isoquant structure through a change in A(t) in (1), or in the TFP in (19) with $\ln a_0$ expressed as a function of t. Factor share estimation of (22) hides such a shift in technology because it suppresses the intercept term $\ln a_0$. To allow technological change in the TFP factor share estimation, include an independent technology variable,

$$\theta_j = a_j + \Sigma_i b_{ji} \ln v_i + \tau_j t \ . \qquad (24)$$

In time series data, t can represent time although there are numerous time series issues to analyze prior to arbitrarily including a time trend variable

In cross section data, the null hypothesis $\tau_j = 0$ amounts to the hypothesis of identical technology across locations. There is an assumption that parameters of the production function do not drift and the null hypothesis that $\tau_j = 0$ tests this assumption in time series data. With detailed information, it is better to model specific technology shifts. For instance, the introduction of gas fired micro generators into electricity generation has specific dates and the availability of nuclear fuel has particular locations. If the null hypothesis $\tau_j = 0$ is not rejected in the estimated factor share system (24), technology can be considered constant.

The link between coefficients and cross price substitution elasticities begins with second order cross partial derivatives. Focusing on energy input,

$$\delta \ln x / \delta \ln E = x_E E/x = \theta_E = a_E + b_{KE} \ln K + b_{LE} \ln L + b_{EE} \ln E \ , \qquad (25)$$

from (22). Consider the cross partial derivative of (25) with respect to labor,

$$\delta^2 \ln x / \delta \ln E \delta \ln L = b_{LE} = \delta(\delta \ln x/\delta \ln E)/\delta \ln L = \delta(x_{EE}/x)/\delta \ln L = [\delta(x_E E/x)\delta L]L$$

$$= [(x_{EL}Ex - x_L x_E E)/x^2]L = (x_{EL}EL/x) - (x_L x_E EL/x^2)$$

$$= (x_{EL}EL/x) - (weEL/x^2) = x_{EL}EL/x - (wL/x)(eE/x) = x_{EL}EL/x - \theta_L\theta_E \qquad (26)$$

where $x_{EL} = \delta^2 x/\delta E\delta L$. Solving for x_{EL} in (26),

$$x_{EL} = (b_{EL} + \theta_L\theta_E)x/EL \;. \qquad (27)$$

To derive a value for x_{EL}, it is possible to use the mean, median, or most recent values of θ_L, θ_E, x, E, and L depending on the data and purpose of the estimate. Second order cross partial derivatives x_{EK} and x_{KL} come from the estimated coefficients b_{EK} and b_{KL} in (22). Second order cross partial derivatives are symmetric due to Young's theorem, $x_{ij} = x_{ji}$.

The own second derivatives x_{ii} are also derived directly. For energy,

$$\delta^2 \ln x/\delta \ln E^2 = b_{EE} = \delta(x_E E/x)/\delta \ln E = [\delta(x_E E/x/\delta E]E = [(x_{EE}Ex + x_E x - x_E{}^2 E/x^2]E$$

$$= x_{EE}E^2/x - x_E E^2/x^2 = x_{EE}E/x + eE/x - (eE/x)^2 = x_{EE}E^2/x + \theta_E - \theta_E{}^2 \;, \qquad (28)$$

and solving for x_{EE},

$$x_{EE} = (b_{EE} - \theta_E + \theta_E{}^2)x/E^2 \;. \qquad (29)$$

Similarly, x_{KK} and x_{LL} are derived using estimates of b_{KK} and b_{LL} from (22). To derive the Allen RSE, second order partials and cross partials of the production function then enter the bordered Hessian matrix H, and the cross price elasticities of interest then come from (17).

7. Translog cost functions and energy cross price substitution elasticities

Shephard's lemma states that input levels are derivatives of the cost function c(r, w, e; x) with respect to input prices, $E^* = \delta c/\delta e$ for instance. This link between cost minimizing inputs and input prices implies a correspondence between production functions and cost functions, although it may not be possible to derive one from the other.

Dual estimation of cross price elasticities developed by Fuss and McFadden (1978) and exemplified by Saicheua (1987) begins with the translog cost function (TCF) and includes the price of energy e and other inputs,

$$\ln c = c_0 + \Sigma_i c_i \ln w_i + .5\Sigma_i\Sigma_k c_{ik} \ln w_i \ln w_k = c_0 + c_K \ln r + c_L \ln w + c_E \ln e + \qquad (30)$$

$$.5[c_{KK}(\ln r)^2 + c_{LL}(\ln w)^2 + C_{EE}(\ln e)^2 + 2c_{KL}\ln r \ln w + 2c_{KE}\ln r \ln e + 2c_{LE}\ln w \ln e].$$

The TCF in (30) cannot be derived from the TPF in (19) or vice versa. In contrast, LLP implies a log linear cost function, a bit tedious to derive. Generally, estimation starts with specification of a production or cost function and deriving one from the other is problematic. The data or the issue under study may determine the choice between specifying a production versus a cost function.

The elasticity of cost with respect to the price of energy is the partial derivative of the TCF,

$$\delta \ln c/\delta \ln e = c_E + c_{KE}\ln r + c_{LE}\ln w + c_{EE}\ln e \,. \quad (31)$$

By Shephard's lemma, $E = \delta c/\delta e$ and $\delta \ln c/\delta \ln e = (\delta c/\delta e)(e/c) = E(e/c) = eE/c$. For a competitive firm, cost c equals revenue $c = px = x$. It follows that $\delta \ln c/\delta \ln e = eE/x = \theta_E$, the energy factor share.

Factor share equations for capital and labor are similar, leading to the cost share system

$$\theta_K = c_K + c_{KK}\ln r + c_{KL}\ln w + c_{KE}\ln e$$

$$\theta_L = c_L + c_{KL}\ln r + c_{LL}\ln w + c_{LE}\ln e \quad (32)$$

$$\theta_E = c_E + c_{KE}\ln r + c_{LE}\ln w + c_{EE}\ln e \,.$$

Factor shares and factor prices come from the data, and (32) provides estimates of the coefficients c_i and c_{ij}. Linear homogeneity or CRS implies $\Sigma_i c_i = 1$ and $\Sigma_i c_{ki} = 0$ in (32), conditions that may be tested or imposed. Similar to factor share estimation of (22) imposing CRS typically improves properties of the estimate. The discussion of homotheticity and technology apply to estimates of the TCF cost share system as well as TPF factor share system.

Cross price elasticities can also be derived from the estimated coefficients in the TCF cost share system (32). For instance, the second cross partial derivative of the TFC for e and w is

$$c_{EL} = \delta^2 \ln c/\delta \ln e \delta \ln w = \delta\theta_E/\delta \ln w = \delta(eE/c)/\delta \ln w = \delta(eE/c)/(\delta w/w) = ew[\delta(E/c)/\delta w] \quad (33)$$

since $\delta e/\delta w = 0$. It follows that

$$c_{EL} = ew[(cE_w - Ec_w)/c^2] \quad (34)$$

where $E_w \equiv \delta E/\delta w$. By Shephard's lemma, $c_w \equiv \delta c/\delta w = L$ and it follows that

$$c_{EL} = ew[(E_w/c - (EL/c^2) = [(e\varepsilon_{EL}E/c) - (eEwL/c^2)]$$

$$= (eE/c)\varepsilon_{EL} - \theta_E\theta_L = \theta_E\varepsilon_{EL} - \theta_E\theta_L \; . \quad (35)$$

Solving for the cross price elasticity,

$$\varepsilon_{EL} = (c_{EL} + \theta_E\theta_L)/\theta_E \; . \quad (36)$$

The derivations of the other five cross price elasticities ε_{KL}, ε_{KE}, ε_{LE}, ε_{LK}, and ε_{EK} are similar, as are the own price elasticities

$$\varepsilon_{ii} = (cii - \theta_i + \theta_i^2)/\theta_i \; , \; i = L, E \; . \quad (37)$$

Allen relative elasticities α_{ik} can be derived from the cross price elasticities ε_{ik} according to $\alpha_{ik} = \varepsilon_{ik}/\theta_k$.

The TCF approach allows direct derivation of cross price elasticities without inverting a Hessian production matrix. There is error in the parameter estimates that go into the Hessian matrix and inversion might magnify the error. Both the determinant and the cofactor contain error, however, that could cancel. Given available data, it is worthwhile to estimate substitution elasticities from both production and cost functions. There has been no systematic effort to compare the two sorts of estimates, and there is no presumption that either approach is superior.

Data generally include only input levels and spending, forcing derivation of "average" input prices. These average prices must distort estimation results, and more so with aggregated inputs. If capital payment is the residual of value added, there is no simple way to deduce the return to capital. There is particular capital machinery and equipment for particular types of energy input, complicating efforts to estimate energy substitution.

Capital payment rK is conceptually the return to capital times the stock of utilized capital. Data on physical K input is typically not be available, and separating r from rK is tenuous. Stock market returns offer a proxy for r, but stock prices depend partly on transient expectations. Real interest rates provide another imperfect proxy for r. Neither stock market returns nor real interest rates perform well in estimation. Occasional capital surveys may provide estimates of capital stocks, and utilization rates are available for some industries. Given enough history in a time series, a reasonable alternative is to build an estimate of capital input K from its investment series. Estimating capital input or return remains the most serious challenge for applied energy substitution. In the end, there is no substitute for knowledge about the industry under study.

8. Conclusion

The economics of energy substitution is important for a host of issues, including factor taxes, fuel taxes, environmental policy, tariffs on machinery, capital taxes, production subsidies, education subsidies, immigration policy, labor contracts, minimum wages, investment subsidies, depletion allowances, and international trade and investment policy. As an example, the success of electric industry restructuring and environmental regulations depends on how producers and buyers adjust to evolving energy prices, including substitution inside the industry between various fuels. General equilibrium models such as Chichilnisky and Heal (1993) and Thompson (1994, 2000) illustrate economy wide adjustments utilizing estimates of energy substitution.

The lack of empirical studies on energy substitution is partly due to the fragmented theoretical literature, and the present paper may prove useful to empirical researchers. Ideally, a wide range of energy cross price substitution elasticities would provide a catalog for reliable applications in energy economics.

References

Allen, R.G.D., 1938, Mathematical Analysis for Economists (St. Martin's Press, New York).

Atkinson, Lloyd, 1975, Empirical estimates of the elasticity of factor substitution: Some economic considerations, Southern Economic Journal 41, 395-402.

Barnett, Andy H., Keith Reutter, and Henry Thompson, 1998, Electricity substitution: Some local industrial evidence, Energy Economics 20, 411-419.

Beattie, Bruce and Robert Taylor, 1985, The Economics of Production (John Wiley & Sons, New York).

Bemdt, Ernst, and Laurits Christensen, 1973, The internal structure of functional relationships: Seperability, substitution, and aggregation, The Review of Economic Studies 55, 403-410.

Berndt, Ernst and D.O. Wood, 1975, Technology, prices and the derived demand for labor, Review of Economics and Statistics 57, 376-384.

Blackorby, Charles and Robert Russel, 1981, The Morishima elasticity of substitution: Symmetry, constancy, separability, and its relationship to the Hicks and Allen elasticities, Review of Economic Studies 48, 147-158.

Burney, Nadeem and Faisal Al-Matrouk, 1996, Energy conservation in electricity generation: A case study of the electricity and water industry in Kuwait, Energy Economics 18, 69-79.

Caloghiro, Yannis, Alexi Mourelatos, and Henry Thompson, 1997, Industrial energy substitution during the 1980s in the Greek economy, Energy Economics 19, 476-491.

Cameron, Trudy and S.L. Schwartz, 1980, Inflationary expectations and the demand for capital, labor, and energy in Canadian manufacturing industries, in W.T. Ziemba, et al, eds., Energy Policy Modeling: U.S. and Canadian Experiences (Matinus Nijhoff Publishing, Boston).

Chambers, Robert, 1988, Applied Production Analysis: A Dual Approach (Cambridge University Press, Cambridge).

Chang, Kuo-Ping, 1994, Capital-energy substitution and the multi-level CES production function, Energy Economics 16, 22-26.

Chang, Winston, 1979, Some theorems of trade and general equilibrium with many goods and factors, Econometrica 46, 709-726.

Chichilnisky, Graciela and Geofrey Heal, 1993, Energy-capital substitution: A general equilibrium Analysis, in G.M. Heal, ed., Critical Writings in the Economics of Exhuastible Resources (Edward Elgar, London).

Christensen, Laurits, Dale Jorgensen, and Lawrence Lau, 1973, Transcendental logarithmic production frontiers, The Review of Economics and Statistics 55, 28-45.

Clark, Don, Richard Hofler, and Henry Thompson, 1988, Separability of capital and labor in US manufacturing, Economics Letters 26, 197-201.

Denny, Michael, Melvyn Fuss, and Leonard Waverman, 1981, Substitution possibilities for energy: Evidence from US and Canadian manufacturing industries, in E.R. Berndt and B.C. Field, eds., Modeling and Measuring Natural Resource Substitution (MIT Press, Cambridge).

Ferguson, C.E. and Ralph Pfouts, 1962, Aggregate production functions and relative factor shares, International Economic Review 3, 328-337.

Field, Barry and Charles Grebensteinz, 1980, Capital energy substitution in US manufacturing, Review of Economics and Statistics 62, 207-212.

Ford, Jon and Henry Thompson, 1997, Global sensitivity of neoclassical and factor proportions models to production technology, International Economic Journal 11, 61-74

Fuss, Melvin and Daniel McFadden, 1978, Production Economics: A Dual Approach to Theory and Applications (North Holland, Amsterdam).

Griffin, James and Paul Gregory, 1976, An intercountry translog model of energy substitution responses, American Economic Review 66, 845-857.

Hunt, Lester, 1986, Energy and capital: Substitutes or complements? A note on the importance of testing for non-neutral technical progress, Applied Economics 18, 729-735.

Kemlert, Claudia, 1998, Estimated substitution elasticities of a nested CES production function approach for Germany, Energy Economics 20, 249-264.

Kuper, Gerard and Daan P. van Soest, 2003, Path-dependency and input substitution: Implications for energy policy modeling, Energy Economics 25, 397-407.

Mahmud, Syed, 2000, The energy demand in the manufacturing sector of Pakistan: Some further results, Energy Economics 22, 641-648.

Saicheua, Supavud, 1987, Input substitution in Thailand's manufacturing sector: Implications for energy policy, Energy Economics 9, 55-63.

Sato, Ryuzo, and Tetsunori Koizumi, 1973, On the elasticities of substitution and complementarity, Oxford Economic Papers 25, 44-56.

Takayama, Akira, 1993, Analytical Methods in Economics (University of Michigan Press, Ann Arbor).

Thompson, Henry, 1994, Do oil tariffs lower wages? Open Economies Review 5, 191-202.

Thompson, Henry, 1997, Substitution elasticities with many inputs, Applied Mathematics Letters 10, 123-127.

Thompson, Henry, 2000, Energy taxes and wages in general equilibrium, OPEC Review 24, 185-194.

Thompson, Peter and Timothy Taylor, 1995, The capital-energy substitutability debate: A new look, Review of Economics & Statistics 77, 565-569.

Urga, Giovanni and Chris Walters, 2003, Dynamic translog and linear logit models: A factor demand analysis of interfuel substitution in US industrial energy demand, Energy Economics 25, 1-21.

Varian, Hal R., 1994, Microeconomic Analysis (W. W. Norton & Company, New York).

Walton, Amy, 1981, Variations in the substitutability of energy and nonenergy inputs: The case of the middle Atlantic region, Journal of Regional Science 21, 411-420.

Yi, Feng, 2000, Dynamic energy-demand models: A comparison, Energy Economics 22, 285-297.

Zellner, Arnold, 1962, An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, Journal of the American Statistical Association 67, 348-68.