Modelling required energy consumption with equivalence scales

Yuxiang Ye, Steven F. Koch**, and Jiangfeng Zhang****

ABSTRACT

Due to difficulties in accessing detailed energy modelling and usage data that are required to estimate energy needs that are responsive to local circumstances, we propose an equivalence scale approach to the determination of required energy consumption. Our method requires the estimation of energy equivalence scales that are used to rescale reference household energy consumption and, thus, yield household-specific energy requirements. We apply the method using readily available income and expenditure data, finding that estimated required energy is well above actual energy expenditure for low- and middle-income households, which is consistent with an expectation that basic energy needs for poor households may not be met. The proposed approach is general enough to incorporate other features that might be deemed relevant and available in other settings, and, therefore, can be used to examine the affordability component of SDG 7, undertake energy poverty analysis or design appropriate policies.

Keywords: Required energy consumption, Equivalence scale, Energy poverty, Semiparametric regression

http://www.example.com

*Corresponding author. Department of Economics, University of Pretoria, Private Bag X20, Hatfield, Pretoria 0028, Republic of South Africa. E-mail: yuxiang.ye@up.ac.za.

**Department of Economics, University of Pretoria, Republic of South Africa. E-mail: steve.koch@up.ac.za.

***Department of Automotive Engineering, Clemson University, Greenville, USA. E-mail: jiangfz@clemson.edu.

1. INTRODUCTION

Energy consumption is different from many other consumer goods, because it is often considered a basic need (Welsch and Biermann, 2017); its satisfaction is necessary for an acceptable quality of life. Furthermore, modern energy services, such as electricity and liquefied petroleum gas are instrumental for most of human basic needs (Chakravarty and Tavoni, 2013). However, energy consumption contributes to household expenses, and (especially) when households are income-poor, energy consumption may be compromised for other purchases. These concerns underpin Sustainable Development Goals (SDG) 7 – which seeks to "Ensure access to affordable, reliable, sustainable and modern energy for all" (United Nations, 2021). In other words, there is a concern that a non-negligible proportion of the world population is not able to purchase enough reliable energy for their needs, and are, therefore, energy poor. However, the determination of energy poverty, like any measure of poverty, requires an estimate of "need". For example, Boardman's (1991) energy poverty ratio requires "theoretical energy expenditure" or "required energy expenditure", which are measures of need. Intuitively, required expenditure focuses on the acquisition of adequate energy services (Liddell et al., 2012).[1]

Despite its intuitive appeal, or maybe because of it, as well as the fact that it cannot be observed, adequate energy services, and thus, required energy consumption (REC), are open to debate and interpretation. Three approaches are widely applied in the literature to measure required or adequate expenditure. One option models energy demand following engineering methods, which are based on detailed domestic energy usage (in kWh), appliances and/or building characteristics information. For instance, the United Kingdom's (UK) Building Research Establishment Domestic Energy Model (BREDEM) (BRE, 2015) requires extensive engineering calculations localized to the country to account for dwelling characteristics and energy usage; such data is not widely available, if at all, in many countries. Papada and Kaliampakos (2018, 2020) suggest a stochastic model for required energy estimation founded on the UK's BREDEM. However, their research does not account for household heterogeneities that affect need. Another option is to conduct a purposive survey incorporating the relevant aspects of energy usage (Ntaintasis et al., 2019); however, such

[1] In the literature, *energy poverty* is widely used to emphasize the lack of access to modern energy services in developing countries (Li et al., 2014), although in some instances it is also applied in developed country studies (e.g. Okushima, 2017; Kyprianou et al., 2019; Robinson, 2019; Bednar and Reames, 2020). *Fuel poverty* tends to refer to the affordability of energy in developed countries, especially in Europe (e.g. Boardman, 1991; Liddell et al., 2012; Legendre and Ricci, 2015; Heindl, 2015), and the U.S. (Mohr, 2018). *Energy insecurity* is also often used to describe the similar situation in the U.S. (Hernández, 2016; Memmott et al., 2021). In this paper, we talk about *energy poverty* as it covers both the lack of access to modern energy services and the lack of affordability of energy services.

surveys are expensive to conduct, and, therefore, difficult to replicate widely. A further approach uses actual energy expenditure, instead of required (Heindl, 2015; Legendre and Ricci, 2015; Mohr, 2018). Although actual expenditure is expected to capture localized conditions and differences across households, it is unlikely to correctly capture need, because some households will reduce energy consumption to fulfil other needs. We are not aware of any method for the determination of required energy consumption that: (i) is underpinned by readily available data, (ii) accounts for household heterogeneities, (iii) captures localized conditions and (iv) incorporates relevant aspects of energy usage.

In this paper, we propose such a method for estimating REC – a necessary parameter for the analysis of energy poverty. Our method is especially useful, when detailed engineering energy modelling and usage data are not available, as is the case in most developing country contexts. Instead, it is based on household expenditure surveys, which are collected nearly everywhere for the purpose of determining consumer price indexes and, thus, inflation. The method is also flexible enough to incorporate additional information about households, including information that correlates with household energy expenditure, as well as additional features that researchers or policymakers deem to be relevant, assuming such information is available. Specifically, we use semiparametric regression - a nonlinear multivariate regression - to incorporate a range of both household characteristics and energy usage factors. Finally, since the data is based on observed behaviour, it is able to capture localized information, such as that related to the weather.

Since the proposed approach makes use of widely available income and expenditure data and can be estimated from multivariate regression, it offers: (i) a clear and viable option for those interested in designing policies to meet SDG 7, (ii) a method that can be applied by those interested in energy poverty and policy analysis in otherwise limited data settings and (iii) a method that can be applied similarly across a variety of countries for comparability purposes. We apply the approach in a case study of South Africa to gauge its reasonableness finding intuitive, and, therefore, reasonable results. In particular, we estimated REC for low- and middle-income households to be well above actual energy expenditure. This result is intuitively appealing, because South Africa is an unequal country, where poverty is rife (Leibbrandt et al., 2016); thus, we would expect poor households to require more energy than they are currently using.

The remainder of this paper is organised as follows. Section 2 provides an overview of the existing literature in this field. Section 3 describes the methodology for required energy consumption modelling. We discuss the South African case study in Section 4 and present the results and discussion in Section 5. Section 6 concludes.

2. LITERATURE REVIEW

When it comes to required household energy consumption for energy poverty measurement, the literature incorporates a range of ideas. For example, the low-income-high-cost (LIHC) indicator (Hills, 2012) uses an energy consumption threshold determined by required household energy costs. Household required energy is underscored by engineering models of energy use that incorporate building specifications, such as insulation levels, heating systems, geographical location of the dwelling and construction type (BRE, 2015), amongst other things. Taking advantage of detailed dwelling and household information provided by the English Housing Survey (DCLG, 2009), the BREDEM calculates total household energy requirements for space and water heating (to meet defined standards), energy for lights and appliances (including requirements for pumps, fans and electric showers, and energy generated by renewables) and energy for cooking (BEIS and BRE, 2018). Once required energy usage has been determined, it is multiplied by the relevant energy price to derive required energy expenditure. Thus, required energy expenditure is often underpinned by detailed knowledge of the building stock and its energy efficiency (Rademaekers et al., 2016).

However, BREDEM is sensitive to the values of multiple parameters. Herrero (2017) shows that actual energy expenditure is well below (BREDEM) modelled energy expenditure, even in higher income deciles. Such results imply that wealthy households in the UK are energy poor, which is difficult to reconcile with its developed country status. These results also suggest that overestimation of household energy requirements is possible under the BREDEM model, limiting its value and generalizability in application. Furthermore, accessing such information may be problematic in many circumstances, especially in developing countries.

Similarly, Papada and Kaliampakos (2018, 2020) develop a model for Greek energy consumption taking into account all domestic energy uses (space heating, space cooling, electricity-cooking-lighting and domestic hot water), although it ignores a number of differences that exist across households, such as housing quality and the number of household members. Although that concern is partially addressed by Ntaintasis et al. (2019), who consider floor area, type of residence, age of residence, energy prices and annual specific electrical and thermal energy consumption required across different types of Greek residential buildings, their research is built upon their own survey, and, therefore, may be difficult to replicate in many developing country contexts.

The aforementioned REC models are relatively complex and require extensive data, or expensive data to collect. Apart from these difficulties, one must also address household heterogeneity, such that different types of households can be reasonably compared. Incorporating heterogeneity, as we do through the equivalization of household energy consumption, is uncommon. In the case of energy poverty, it is often ignored or assumed to be the same as income equivalence (Herrero, 2017; Hills, 2011). For example, Legendre and Ricci (2015) apply the OECD-modified income equivalence scales to adjust household income for the energy poverty ratio calculation, but there is no equivalization of energy expenditure.[2]

As acknowledged by Hills (2011), the OECD-modified scales do not reflect how energy requirements vary between households. Instead, Hills (2012) proposes an alternative underpinned by three years of the English Housing Survey.[3] Hills' equivalization factors are calculated from ratios between median required household energy expenditure within different household groups and median required energy expenditure in two-adult households. Heindl (2015) estimates German energy poverty rates using this median-to-median ratio scale; however, the scales are used to equivalize actual energy expenditure rather than required, which, even in a highly developed country like Germany are not observed. Even though actual expenditure is not expected to capture need, these ratio-to-ratio scales, which represent an adjustment factor, offer one convenient way to deal with heterogeneity.

Conceptually, our approach is most similar to Hills's (2012) equivalization factors, although we offer further generalizations and improvements. In most developing countries and even in many developed ones, clear estimates of energy need are not available; thus, rescaling actual expenditure makes sense. However, a median-to-median ratio, especially one that treats children and adults as equivalents, misses out on the fact that it is more than just the number of people in a household that determine need, even in a place as homogeneous as Germany. For example, domestic energy requirements vary with climates and regions (Pachauri and Spreng, 2004; Charlier and Legendre, 2019; Berkouwer, 2020) and housing energy efficiency (Boardman, 1991), as well as energy using appliances and living space (Ntaintasis et al., 2019). Thus, the ratio scale we develop below, more carefully controls for household heterogeneity, and, therefore, should allow researchers and policymakers in different locations to tailor the approach for locally relevant circumstances.

[2]The OECD-modified scale, first proposed by Hagenaars et al. (1994), assigns a value 1 to the household head, 0.5 to each additional adult member, and 0.3 to each child (under 14 year-old). A brief description of the equivalence scales used by the Organisation for Economic Co-operation and Development (OECD) is shown in `www.oecd.org/els/soc/OECD-Note-EquivalenceScales.pdf`.

[3]According to BEIS and BRE (2018), the UK's equivalization factors are based on three years of required energy costs data from the English Housing Survey (using the 2008, 2009 and 2010 datasets). The combined three year weights were used to arrive at the final equivalization factors. During the calculation, adults and children are treated equally in the equivalization of modelled bills - that is, a household with two adults and two children are treated the same as a household with four adults. In addition, the equivalization factors are not intended to be reviewed on an annual basis.

3. MODELLING REQUIRED ENERGY CONSUMPTION

Our approach takes two steps. In one step, we estimate an energy equivalence scale. In the other, we determine household required energy consumption. Our required expenditure estimate is the equivalence re-scaled reference (or reference household's) actual energy expenditure; thus, reference energy consumption is scaled up (or down), depending upon the household's equivalence factor, which we refer to as Λ_i to denote that it can differ by household.

3.1 Energy equivalence scales

Equivalence scales arose in welfare analysis in order to compare household well-being, although such scales are not perfect (Pollak and Wales, 1979). One common scale is the per capita scale; for example, country welfare is often inferred from gross domestic product (GDP) per capita comparisons. However, per capita income or expenditure comparisons ignore household economies of scale, the notion that some goods and services consumed by the household have public good characteristics, which offer benefits for household members apart from the primary consumer (Nelson, 1988). Electric light in a room is one such example (Lazear and Michael, 1980), as it can be shared. Sharing goods and services decreases the per capita cost of maintaining living standards (Nelson, 1988). Thus, the proportionate increase in energy expenditure for these services is less than the increase in household size (i.e. number of household members), although how much less should be examined within the relevant context.

The impetus for equivalence scales comes from Engel (1895), who argues that since richer households spend a smaller share of the household budget on food, the share of the household budget allocated to food is a reasonable measure of household welfare. An Engel income equivalence scale, derived from this argument, is defined as the ratio of two household incomes: each having the same food budget share, but each having different household sizes and structures (Lewbel and Pendakur, 2008). Typically, a single-person adult or two-person adult (no children) household will serve as the reference for the welfare comparisons.[4] For our analysis, we assume that household welfare can be proxied by the share of the budget devoted to energy consumption; in Figure 2, we show that the energy share is also decreasing in income, as Engel (1895) argued was true for food. Therefore, we believe it is reasonable to follow the share comparison approach.

[4] Nicholson (1976) argues, however, that Engel income equivalence scales are likely to overestimate the cost of a child, because they do not treat child goods separate from adult goods and because child costs, especially when young, are primarily food costs.

Scale estimation methodologies have been developed over decades and, as implied by the previous discussion, typically limit the focus to household size and age composition, i.e., the number of adults and children in the household; however, much work remains. We extend that focus, because the public goods nature of energy will be affected by the household's technology and circumstances. Therefore, as we describe below, analysis characteristics go beyond household adult-child composition. Our methodology also indirectly imposes two further constraints on energy shares. Firstly, shares, by definition, are expected to lie in the unit interval: a household can neither spend more on energy than they spend in total nor should they be paid to consume energy.[5] Second, equivalence scales should be base-independent (Pendakur, 1999).[6]

Empirically, base-independence requires the relationship between the budget share and (log) total expenditure to follow the same shape, allowing for either vertical or horizontal translations or both. To impose base-independence, we consider a variation on and extension of the nonlinear model proposed by Yatchew et al. (2003) more recently applied by Koch (2018), although with a specific focus on catastrophic health expenditures. We modify the method to focus attention on energy expenditure, rather than food expenditure, and we extend it through the introduction of additional controls related to household circumstances and technology, which could include the weather, the state of housing and ownership of domestic appliances, to name a few. Borrowing from Yatchew et al. (2003), our estimation is founded on equation (1), which equates household i's energy expenditure share to our reference household r's share. The point of the equation is to find Λ_i, which is determined indirectly. The solution is a ratio adjustment to household i's total (ln) expenditure, such that the shares would be equal.

An important feature of the model is that the first element of the coefficient vector - the coefficient on $\ln x$ in equation (1) - is equal to one. That is consistent with base independence and clarifies the identification intuition. The goal of the analysis is to estimate how nonreference household characteristics (d_j^i on the right hand side of equation (1)) can be used to determine the household's

[5]In South Africa, some households do receive subsidized electricity, which should also be counted as subsidized total consumption. However, in some cases, the data is not clear, and, therefore, we are forced to remove observations that do not make that disctinction clear.

[6]Base independence requires equivalence scales to be independent of the base level of utility (Pendakur, 1999); without such an assumption, the scales cannot be empirically identified (Blackorby and Donaldson, 1993). Given an expenditure function $C(\mathbf{p}, v, \mathbf{d})$, the relevant equivalence scale Λ, can be defined as:

$$\Lambda(\mathbf{p}, \mathbf{d}) = \frac{C(\mathbf{p}, v, \mathbf{d}^i)}{C(\mathbf{p}, v, \mathbf{d}^r)},$$

where $\mathbf{p}$ is the vector of prices and $\mathbf{d}$ is a vector of characteristics, described in the next section, and v denotes utility. Therefore, the equivalence scale is the ratio between the minimum expenditure to achieve utility for the household with characteristics $\mathbf{d}^i$ and the minimum expenditure of achieving the same level of utility for the reference household with characteristics $\mathbf{d}^r$. In this general form, the equivalence scale is a constant-utility, constant-price, cost-of-living index (Deaton and Muellbauer, 1980).

expenditure scaling factor, i.e., equivalence scale. Therefore, if we allow for additional adjustments to household expenditure (such as a separate parameter on ln expenditure) in the estimation routine, we would not reach a solution and the model would not be identified. Despite that, we are indirectly estimating an expenditure adjustment parameter – intuitively, it is a random expenditure parameter that is correlated with a range of household factors.

$$
\begin{aligned}
w(\mathbf{p}, x^r, \mathbf{d}^r) &= w(\mathbf{p}, x^i, \mathbf{d}^i) + \eta(\mathbf{p}) + \varepsilon \\
&= w\left(\mathbf{p}, \frac{x^i}{\Lambda_i(\mathbf{p}, \mathbf{d}^i)}\right) + \eta(\mathbf{p}) + \varepsilon \\
&= w\left(\frac{x^i}{\Lambda_i(\mathbf{d}^i)}\right) + \varepsilon \\
&= f\left(\ln x^i - \sum_j \lambda_j d_j^i\right) + \varepsilon.
\end{aligned}
\tag{1}
$$

In (1), $\mathbf{p}$ is the vector of prices, x refers to total household expenditure, and $\mathbf{d}$ is a vector of household characteristics. Λ_i is the energy equivalence scale to be estimated for household i, d_j^i denotes nonreference household i's characteristic j, λ_j is the coefficient for characteristic j in the semiparametric model, and ε is an error term assumed not to be correlated with the other variables in the model. We consider the likely effect of violations of that assumption following Dong (2010), which we describe in Section 3.2. f is a convolution of a reference group's budget share function w with the exponential function, while η is the elasticity of the budget share with respect to the price of energy. Unfortunately, we do not have information on prices, so we cannot incorporate price controls in the model – this explains the model change from line 2 to line 3.[7] In the last equality, we see an unknown function (f) of a linear index, which we estimate following Ichimura (1993). Given the way it is estimated, shares will stay within the unit interval.

Since the function f is not known, it will be estimated nonparametrically. Rather than using difference procedures for estimation, as suggested by Yatchew et al. (2003), we make use of bandwidth and leave-one-out cross-validation procedures, which are implemented via the np package (Hayfield and Racine, 2008) for R (R Core Team, 2021).[8] The log-linear index model within f, as well as the

[7] Although it would be preferable to have price data, this is a common data limitation in many developing countries. We also find, as described in Figure 2, where the curves are very close together implying minimal vertical or horizontal translation, that our $\eta(\mathbf{p}) = 0$ assumption is not unreasonable for our analysis.

[8] In its simplest form, nonparametric analysis averages data within an interval of the data space, and that interval is referred to as a bandwidth. As the bandwidth gets wider, more observations are used, and the average converges towards the full sample mean. However, smaller bandwidths contain fewer observations, and the averages within those intervals are expected to be more varied as well as more appropriate for the interval considered. Thus, bandwidth procedures trade-off the variability of the smaller interval with the reduced mean bias offered from focusing on the smaller interval.

binary nature of all of the control variables, which we describe below, yields equivalence scales that can be easily calculated. The scale will be the exponentiated value of the sum of the estimates of characteristics that differ from the reference household's characteristics.[9]

$$\Lambda_i(\mathbf{d}^i) = \exp\left(-\sum_j \lambda_j d_j^i\right). \tag{2}$$

3.2 Potential endogeneity

The final concern to be addressed is the potential for endogeneity, which would arise if any of the included variables were correlated with variables that were not included. For example, total expenditure would be endogenous, if it was measured with error or otherwise correlated with unobservables that might influence household formation, which is our main concern.[10]

One approach to examining endogeneity in the nonlinear setting is to follow Imbens and Newey (2009), based on a control function that requires both the estimation of a conditional cumulative distribution function and an instrument. Koch and Tshiswaka-Kashalala (2018) apply this to estimate the demand for contraceptive efficacy. Unfortunately, as noted in footnote 10, an instrument may not be readily available in the data. Therefore, we follow Dong (2010), which is similar in spirit to Imbens and Newey (2009), but does not require an instrument. Instead, it requires a continuous variable with a large support - we use income. Importantly, we are not assuming income is an instrument that meets the exclusion restriction, although we do assume it is correlated with our potential endogenous variable (expenditure). Rather, identification comes from nonlinearity; the extent of the real number line covered by income allows one to estimate the relevant nonlinear relationship.

Dong (2010) outlines a two-step procedure for addressing endogeneity within a binary response or probability model; although equation (1) is not a binary response model, expenditure shares lie in the unit interval, as do probabilities; thus, similar constraints apply to both.[11] In the first

[9] Characteristics are indicators, such as having a fridge or staying in rural area. For the analysis, we zero-out the reference household characteristics. In other words, $d_j^r = 0 \forall j$, such that, when adjusting for multiple household characteristics as required by the model, we sum the estimates for all of the relevant characteristics k (that are different from the reference household characteristics) in (2). As an example, if there were two characteristics, $j = \{1, 2\}$ – for the reference household r are $d_1^r = 0$ and $d_2^r = 0$, while the characteristics for a nonreference household i were $d_1^i = 0$ and $d_2^i = 1$ – the scale would simply be $\Lambda = \exp(\lambda_1 \cdot (d_1^r - d_1^i) + \lambda_2 \cdot (d_2^r - d_2^i)) = \exp(-\lambda_2)$.

[10] In earlier versions of this research (Ye et al., 2020), a linearized version of the model is also estimated, following Deaton (1987), which made it possible to apply two-stage least squares, using income as an instrument for expenditure. However, if endogeneity is driven by omitted variables that are correlated with total expenditure, those variables could also be correlated with income, such that income may not be an appropriate instrument. Thus, the heteroskedasticity instrumental variable procedure (Lewbel, 2012) was also considered, as a robustness check. In the linear setting, very limited evidence of endogeneity was uncovered - results available from the authors - and, more importantly, the effect of correcting the endogeneity made very little difference to the resulting scale estimates.

[11] An extensive literature has built up around the application of binary response models to shares; see, for example Papke

stage, a control function is estimated. In the second, the control function is incorporated into the model of interest, represented by equation (1). The control function is the regression error arising from the nonparametric regression of (ln) expenditure on (ln) income and all the other controls in equation (1). In the second step, Dong (2010) semi-parametrically estimates the binary response model following Klein and Spady (1993), where the second stage model includes the control function. Other than the fact that we have shares, rather than binary responses (probabilities), such that we estimate Ichimura's (1993) continuous version of Klein and Spady (1993), the approaches are identical.

For estimation in the first stage, we follow Li and Racine (2004) and Racine and Li (2004) in estimating a local linear regression with continuous and both ordered and unordered categorical variables. Estimates are implemented by the np package in R. We apply an epanechnikov kernel for the continuous variables and the Li and Racine (2007) kernel, which is an extension of Wang and van Ryzin's (1981) kernel that works well for both ordered and unordered discrete variables. As discussed in numerous places, the bandwidth is more important than the kernel (Li and Racine, 2007). Thus, we apply least squares leave-one-out cross-validation to estimate optimal bandwidths (Li and Racine, 2004). We then use the bandwidths to fit the regression and extract the residuals, which are used as the control function in the second stage. Code in R for all of the analysis is available from the authors, upon request.

3.3 Reference energy consumption

Once the scale is determined for a household, it is used to determine their required energy expenditure. Required energy consumption becomes the equivalence rescaled reference energy consumption, as in

$$E_{R,i} = \bar{E}_b \times \Lambda_i, \tag{3}$$

where $E_{R,i}$ is the required energy consumption for household i. As already discussed, the equivalence scale, Λ_i, adjusts for household attributes relative to the reference household.

In determining a reference energy requirement, $\bar{E}_b$, our goal is to find the energy consumption necessary to maintain a reasonable living standard. We assume that many analysts do not have access to the sort of heating, cooking and household efficiency data that would be used in, for example, the BREDEM. Instead, we assume that they are likely to have access to energy expenditure data, along with some appliance ownership information and a few other relevant pieces of data, such as total

and Wooldridge (1996).

income or expenditure, that are likely to correlate with energy consumption.

Although our method adjusts for the age and composition of household members, as well as many other items, we still require a reference for adjustment. Thus, we must define a reference household size, structure and living standard. For our purpose, we assume that a reasonable standard of living requires access to or use of electricity, and cooking with modern energy sources, i.e., the main energy source for cooking is either electricity from the grid, gas, or solar energy. We further assume that a reasonable standard of living requires cold food storage and the ability to communicate and be entertained. We also allow for differences across dwelling size (using small size living space as the reference), urban/rural locations (using urban-formal households for the reference), and differences in weather (spring and fall are the reference).

Thus, our reference households are single adult households that stay in a formal urban area, have access to or use electricity, cook with modern energy sources, own a refrigerator or freezer, have the ability to communicate (a cellphone) and entertainment options (such as a television (TV), radio or satellite TV). These reference households live in a small sized dwelling (between 30 and 59 m^2), which is in line with South Africa's National Building Regulations: the floor of any permanent building that is used as a "dwelling house" must be no less than 30 m^2.[12] Rao and Min (2018) also use 30 m^2, as it is recognised in at least a few national standards for public housing. Finally, in South Africa, the weather is generally mild to warm, while piped water access for each dwelling is often a luxury; furthermore, hot water may not be provided through individual geysers (electric water heaters), especially in apartment buildings. Therefore, we neither require geysers in our reference household nor adjust the equivalence scale for it, although we do use it to control for energy expenditure differences across households. We also set our reference using spring and fall, which are less extreme than summer and winter.

Even with the preceding reference rules, expenditure by these households covers an extensive range; thus we also need to define a reference energy expenditure level (to be rescaled). We use the 75^{th} percentile of the energy expenditure distribution for the reference households in the sample. We label this value $\bar{E}_b$, even though it is not the mean. We describe our percentile choice in the following section. However, the choice of consumption percentile and reference household attributes are expected to depend on the analyst's circumstances, as well as data availability.

[12] See `http://www.sans10400.co.za/size-dimensions-room-height/`. In our data, households in the selected dwelling units who occupy permanent structures are expected to respond to questions such as estimated floor areas. However, we find that data for this question for all renters are missing. Thus, we have removed renters from the analysis.

4. CASE STUDY DATA

We apply the above methods to a case study covering South Africa. Our analysis is based on data from a recent household expenditure survey, the Living Conditions Survey (LCS) 2014/2015 (Stats SA, 2017a). The LCS aims to provide data that will contribute to a better understanding of living conditions and poverty in South Africa and is meant to be used for monitoring levels of poverty over time. Importantly, the survey contains information on household expenditure, including energy expenditure, household size and structure and some information on household appliances, which fit our empirical models. The survey also provides information on dwelling space and, since the survey took place over a period of twelve months, it is possible to account for the season of the year. In different settings, we expect the available data to be different; however, the method we outline is general enough to accommodate more or less detail.

4.1 The data

The LCS uses classification of individual consumption by purpose (COICOP) categories. Under COICOP, energy expenditures lie in Category 04. All expenditures on COICOP Category 045, which includes spending/values on electricity (including conventional metering electricity and prepaid electricity), gas (including refilling gas and gas in cylinders), liquid fuels (including paraffin, petrol and diesel for household use, not transport use), solid fuels (including bought and fetched firewood, charcoal, candles, coal, bought and fetched dung, and crop waste), and other household fuel. In our analysis, total energy expenditure includes expenditure from all of these energy sources.

The energy budget share is the share of the household budget devoted to energy consumption. We use consumption expenditure instead of income for all the estimates in the analysis.[13] Household consumption expenditure comprises both monetary and in-kind payment on all goods and services, and the money value of the consumption of home-made products. Hence, the energy share equals the energy expenditure divided by total household expenditure. Given the collection timing, all reported expenditures were inflated/deflated to April 2015, the midpoint of the survey year, using the consumer price index.

The expenditure on COICOP category 0440 is coded as “water and electricity”, which is for

[13] In developing countries, it is more practical to use consumption rather than income. Due to the fact that formal employment is less common, many households have multiple and continually changing sources of income, while home production is more widespread. Hence, measured consumption tends to be ‘smoother’ than income (Deaton and Grosh, 2000), which implies that reported consumption expenditure is more accurate than reported income.

households with consolidated water and electricity bills. Since it is not possible to split electricity out of the bill, these households are ignored in the case study. We further limit the data to households with no more than seven adults and no more than five children, primarily because larger households are quite rare. We also drop observations missing information related to domestic appliances (refrigerator, freezer, cellphone, TV, radio, satellite TV and geyser), main energy source for cooking and living space (estimated area of the dwelling). In addition to this, we remove: household whose living space information is not applicable, which was captured by the survey to indicate a household either not living in a permanent structure or there are multiple households living in one permanent structure (Stats SA, 2017b). Unfortunately, within the survey data, the space not available category captures all renters in the data. Since renters often have electricity and water inclusive of their rent and may not have their own geyser for water heating, rather than trying to separate all of these different factors, we exclude them from the analysis. Finally, we remove any others whose energy expenditure cannot be separately determined, such as households, primarily individuals, who are borders and households that do not purchase any form of energy. Our analysis sample, therefore, includes 12 774 observations from the initial 23 380.[14]

4.2 Household characteristics

We present summary information related to household characteristics and expenditure in Table 1. As shown, the average household spends about 6% of its budget on energy, while average total expenditure in the household is ZAR 8 048 per month (≈USD 676; 1 ZAR = 0.084 USD, April 2015). On average, more than 93% of total energy expenditure is devoted to electricity consumption.

One of the main features we examine is the effect of household size and the number of children and adults – through binary values of these variables. On average, each household contains one child and 2.75 adults. The survey did not include temperature data, so we use *winter* (May-July) and *summer* (November-February) to incorporate seasonal variation in our analysis. As outlined above, we include the ownership of *fridge* (refrigerator, or combined fridge freezer, or freestanding deep freezer) and energy choice for *modern cooking* (main energy source for cooking is electricity from grid, other

[14]Eskom, the national electricity supplier, and municipalities, which are the local suppliers in much of South Africa, do offer free basic electricity (FBE) to some households, see Ye et al. (2018) for more information about the FBE policy in South Africa. In the initial LCS 2014/2015 data, 2 722 out of 23 380 (12%) households report positive values of FBE and 1 978 out of 23 380 (8%) record no spending on energy. Since it is hard to determine if FBE is directly affecting their expenditure behaviour, we also did not select households with positive FBE value recorded; however, since the data was limited to home owners, these considerations have little impact on the resulting data. In their recent work, Ye and Koch (2021) include the households who reported positive FBE values in the empirical analysis and estimate household-specific poverty line (i.e. household energy needs) accordingly. Their estimations of energy equivalence scale and required energy are not that different from what have been reported in our paper.

Table 1: Descriptive statistics of the LCS 2014/2015 data

Variable name and description	Mean	Standard deviation
Household size: total number of household members	4.01	2.22
Number of children (age less than 15-year old)	1.26	1.33
Number of adults	2.75	1.40
Urban formal: settlement type is urban formal	0.46	0.50
Urban informal: settlement type is urban informal	0.07	0.25
Traditional area: settlement type is traditional area	0.45	0.50
Rural: settlement type is rural	0.02	0.14
Modern cooking: main energy source for cooking is electricity from grid, other source of electricity (e.g. generator etc.), gas, or solar energy	0.80	0.40
Fridge: household owns a refrigerator/combined fridge freezer, or freestanding deep freezer	0.82	0.39
Cellphone: household owns a cellphone	0.92	0.27
Entertainment: household owns a TV, a radio or a satellite TV (e.g. DStv/TopTV)	0.89	0.31
Geyser: household owns an electric water heater (geyser)	0.21	0.41
Summer: survey month in November, December, January or February	0.35	0.48
Winter: survey month in May, June or July	0.25	0.43
Very small space: floor area less than 30 m^2	0.11	0.32
Small space: floor area between 30 and 59 m^2	0.28	0.45
Medium space: floor area between 60 and 119 m^2	0.39	0.49
Large space: floor area between 120 and 239 m^2	0.17	0.38
Very large space: floor area is 240 m^2 or more	0.05	0.21
Monthly energy expenditure (unit: ZAR)	306.11	361.21
Monthly total household expenditure (unit: ZAR)	8048.24	12580.63
Energy share (= energy expenditure/total expenditure)	0.06	0.05
Monthly electricity expenditure (unit: ZAR)	285.38	351.40
Monthly gas expenditure (unit: ZAR)	5.47	55.97
Monthly liquid fuel expenditure (unit: ZAR)	8.16	49.46
Monthly solid fuel expenditure (unit: ZAR)	7.10	56.47
Monthly paraffin expenditure (unit: ZAR)	7.21	33.03
Monthly wood expenditure (unit: ZAR)	2.31	38.77
Observation	12774	

Note: 1) Household energy expenditure includes spending on electricity, gas, liquid and solid fuels; 2) Electricity expenditure includes spending on conventional metering electricity and prepaid electricity; 3) Gas expenditure includes refilled gas and gas in cylinders (including gas for heating purposes); 4) Liquid fuel expenditure includes spending on paraffin, petrol and diesel (petrol and diesel for household use only, not transport use); 5) Solid fuel expenditure includes firewood bought and fetched, charcoal, candles, coal, dung bought and fetched, crop waste, and other household fuel. 6) Wood expenditure includes bought and fetched firewood values.

source of electricity, gas, or solar energy). In addition, basic equipment for social communication (cellphone) and self-entertainment (TV, radio or satellite TV) are reasonable appliances for households to achieve a reasonable standard of living. Table 1 shows us that more than 80% of households own a refrigerator or freezer for cold storage; cellphones, and entertainment equipment are also prevalent in South African households. Most of the households reside either in a urban formal (46%) or traditional area (45%) in South Africa, where traditional area refers to communally owned land under the jurisdiction of a traditional leader (Stats SA, 2017b). 80% of households use modern energy sources for daily cooking activities, in which electricity from national grid is the major fuel. We also see that more than 60% of households stay in medium to larger sized homes (>= 60 m^2).

4.3 The reference household

We have previously noted our reference household to be a single (adult) person household living in a small space (between 30 and 59 m^2) in an urban formal area, whose data was captured in spring or fall. This household is assumed to own a fridge/freezer and cook with modern energy sources. This individual is also able to communicate with a cellphone and access entertainment through at least a TV, radio or satellite TV. In Table 2, we see total and energy expenditure for this group (Panel A) compared to the entire sample (Panel C). Our reference group is relatively worse-off, as their total and energy expenditure is much lower. The primary reason for the energy expenditure difference is that the reference households do not own a geyser. Thus, for comparison, we also present a description of the five households who are similar to the reference group, other than their geyser ownership in Panel B. This difference is not unexpected. Due to South Africa's mild climate, water heating via electric geysers, rather than space heating, is the largest user of electricity in the domestic sector in the country (Meyer, 2000), accounting for 39% of household electricity use (DOE, 2018). Thus, as noted by Meyer (2000), when water heating consumed about 40% to 50% of monthly electricity for an average middle-to-upper income household, water heating remains an important expense for the economically advantaged in South Africa. In our data, geyser ownership is limited (21%), which also suggests that electric water heater (geyser) is not widely affordable for South African households. Hence, in our analysis, we focus on the more general reference group (Panel A) which does not own a geyser in their home.

Choosing the appropriate monthly electricity expenditure for our reference, however, requires some justification. For that reason, we plot the distribution of household energy expenditure in Figure 1; its skewness means that most South African households consume less energy than the mean, as the median energy expenditure is lower than the mean. According to Table 2, the 75^{th} percentile of energy expenditure for the reference group (Panel A) falls in between the median and mean of the entire sample (Panel C), therefore, we use that percentile for the reference energy requirement; using a lower value will lower required energy consumption, but will not lead to qualitative changes in the results we present below. Lastly, one might be concerned about the small sample size for our reference group. For comparison, the third quartile of energy expenditure for all single person

Table 2: Descriptive statistics of reference household group

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Panel A: Reference group without geyser (N = 44)						
Total expenditure (unit: ZAR)	709.09	1543.51	2109.58	3517.62	4312.04	16192.75
Energy expenditure (unit: ZAR)	89.92	99.74	179.84	196.58	231.10	802.40
Energy budget share	0.01	0.05	0.07	0.08	0.11	0.25
Panel B: Reference group with geyser (N = 5)						
Total expenditure (unit: ZAR)	2975.03	4134.13	15234.31	13320.70	16375.09	27884.95
Energy expenditure (unit: ZAR)	179.91	179.91	250.00	291.97	400.00	450.00
Energy budget share	0.01	0.03	0.03	0.04	0.06	0.06
Panel C: Entire sample (N = 12 774)						
Total expenditure (unit: ZAR)	213.14	1994.87	3689.53	8048.24	8150.78	160806.72
Energy expenditure (unit: ZAR)	2.17	99.95	199.59	306.11	359.69	6092.11
Energy budget share	0.00	0.03	0.05	0.06	0.08	0.57

households, a sample of 1683 observations, is ZAR 224.78; given the similarity with our reference sample, we are neither concerned with the sample size of the reference group nor the value of actual energy expenditure in our reference group.

Figure 1: Kernel density of monthly energy expenditure

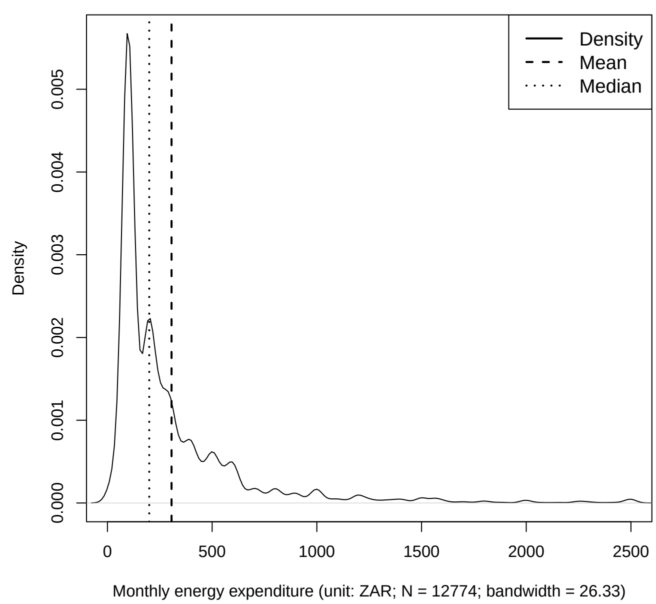
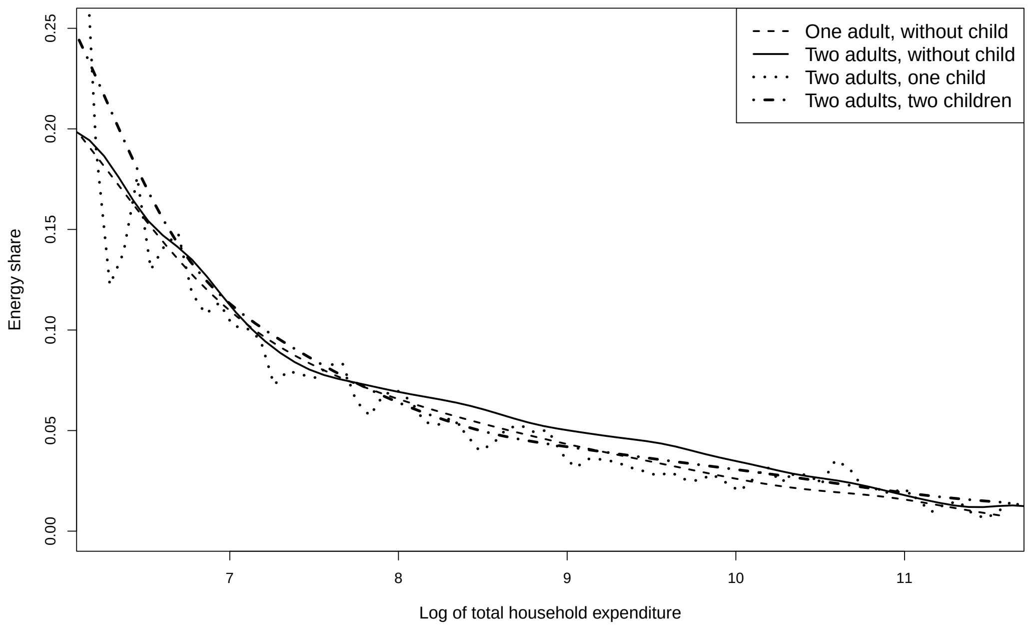


Figure 2: Household energy share and monthly total expenditure by family types



5. CASE STUDY RESULTS AND DISCUSSION

Our analysis focuses on two sets of results. Our primary interest is in the final required energy consumption estimates. To get there, energy equivalence estimates are needed, which are based on our semiparametric model. However, before we turn our attention to those, we illustrate the relationship between total expenditure and the share of energy expenditure, based on fitted nonparametric regressions for selected family types; see Figure 2. The relationship that we see is negative, suggesting that the extension of Engel's intuition to energy shares is realistic. We also see that the underlying relationship would not generally be described as linear, which suggests that the semiparametric equivalization method is appropriate in this setting. Finally, we see these curves lying fairly close to each other, which means that, even though we do not have price data, there does not appear to be much translation from one curve to the next; in other words, assuming $\eta(\mathbf{p}) = 0$, as in equation (1), is not unreasonable.

5.1 Semiparametric model estimates

As discussed in Section 3.1, we have applied a nonlinear model to indirectly estimate equivalence scales. The underlying empirical model estimates are available in Table 3, which contains two sets of parameter results, one for a model assuming expenditure exogeneity and one allowing for endogeneity. Since the reference group in the model is exactly the same as our reference households – which are single adult no child households living in a small sized home in an urban formal setting, with modern energy used for cooking, a refrigerator or freezer, cellphone and entertainment, whose consumption data was collected in either spring or autumn – the model controlled for the effect of not owning those particular appliances, adults in excess of one, children in excess of zero and spaces larger/smaller than the small size reference, as well as non-urban-formal locales, owning a geyser and non-spring/fall weather. Since the focus of the modelling is on determining the adjustments needed for nonreference household characteristics, we present the empirical results in this format.

The exogenous set of estimates, along with their standard errors are reported in the first two columns of Table 3. This initial model assumes that household expenditure is exogenous. In the last two columns, we present results that control for potential endogeneity in household expenditure. The first conclusion to draw from these results is that household expenditure is likely to be endogenous in this setting, as the control function is found to be statistically significant. The impact of that endogeneity, in most cases, is to bring the coefficient estimate closer to zero for the remaining variables. However, in one case, the sign switches while remaining statistically significant (*Very small space*); in others the estimate goes from statistically significant to not (*number of children* = 1 and = 5 or *number of adults* = 7) and in others, the estimate increases in absolute value (*number of adults* = 2, for example). Given these results, the second conclusion to draw is that the source of the endogeneity is correlated with most of the included variables. Regardless of the degree of exogeneity, the estimates show us that household age and size composition does matter, as do additional household characteristics.

In order to interpret the estimates with respect to the equivalence scales, the coefficients need to be related back to the point of the equivalence scaling process, which is to equate shares across reference and nonreference households. In Table 3, the reported estimates are the λ_j in equations (1) and (2). However, to calculate the scale, it is the opposite of the reported estimate that is required.[15] Thus, a positive semiparametric model estimate decreases the equivalence scale, while a negative estimate increases the equivalence scale. In other words, negative/positive model coefficients imply

[15]Although the semiparametric parameters, on their own, can be interpreted as in any other regression, such that the sign tells us the marginal effect of that characteristic on energy shares, that marginal effect is the opposite of the marginal effect on the scale. The scale model developed in equation (1), as well as the equivalence scale calculation described in equation (2) and footnote 9, include a negative sign to account for division within a logarithmic setting; thus, we must take the negative of our parameters to see the marginal effect on the equivalence scale. Further, note that equation (2) represents the exponentiated denominator of $\ln x$ in equation (1). For a positive coefficient, ($\lambda = 0.5$), for example, we have $\exp(-\lambda) < 1$, such that $x/\exp(-\lambda) > x$.

Table 3: Semiparametric index model parameter estimates

	(1) Exogenous expenditure		(2) Endogenous expenditure	
Variable	Estimated coefficient	Standard error	Estimated coefficient	Standard error
Log of household total expenditure[1]	1.0000^a	(0.000)	1.0000^a	(0.000)
Number of adults = 2	-0.0195^a	(0.001)	-0.0451^a	(0.002)
Number of adults = 3	-0.1547^a	(0.001)	-0.0451^a	(0.003)
Number of adults = 4	-0.1671^a	(0.001)	-0.0363^a	(0.003)
Number of adults = 5	-0.2794^a	(0.001)	-0.0447^a	(0.007)
Number of adults = 6	-0.1347^a	(0.002)	-0.0203^a	(0.006)
Number of adults = 7	-0.0660^a	(0.003)	0.0067	(0.016)
Number of children = 1	0.0860^a	(0.001)	-0.0033	(0.002)
Number of children = 2	0.0685^a	(0.001)	0.0348^a	(0.003)
Number of children = 3	0.1489^a	(0.001)	0.0929^a	(0.003)
Number of children = 4	0.0475^a	(0.001)	0.0147^c	(0.006)
Number of children = 5	0.1006^a	(0.004)	-0.0043	(0.013)
Urban informal	0.1375^a	(0.001)	0.0737^a	(0.005)
Traditional area	0.2109^a	(0.001)	0.2449^a	(0.002)
Rural	0.0418^a	(0.001)	0.0444^a	(0.009)
No modern cooking	0.3500^a	(0.001)	0.3631^a	(0.003)
No fridge	0.1319^a	(0.001)	0.1500^a	(0.004)
No cellphone	0.0541^a	(0.001)	0.0071	(0.005)
No entertainment	0.0030^a	(0.001)	0.0326^a	(0.004)
Geyser	-0.4328^a	(0.001)	-0.1911^a	(0.003)
Summer	-0.0031^a	(0.001)	-0.0117^a	(0.003)
Winter	-0.1181^a	(0.001)	-0.0747^a	(0.003)
Very small space	0.0034^a	(0.001)	-0.0244^a	(0.005)
Medium space	-0.0180^a	(0.001)	-0.0471^a	(0.002)
Large space	-0.0446^a	(0.001)	-0.0374^a	(0.004)
Very large space	-0.2980^a	(0.001)	-0.0784^a	(0.005)
Control function[2]			0.1238^a	(0.003)

Parameter estimates from semiparametric least squares applied to equation (1); see Ichimura (1993) for estimation details. The estimated coefficient is λ_j in equations (1) and (2). Significance levels: $^a \leq 0.005$, $^b \leq 0.01$, $^c \leq 0.05$, $^d \leq 0.1$. Additional notes: [1] - For identification, this parameter estimate is set to unity. [2] - Residual from first stage nonparametric regression of ln expenditure on ln income and the remaining model variables following Dong (2010), which does not require ln income to be an instrument, only that it be continuous and correlated.

that nonreference households have relatively more/fewer resources at their disposal than do reference households, and, therefore, the equivalence scales need to deflate/inflate nonreference households' resources in order to equivalize the energy shares across the groups.

With that understanding, the results are not entirely unexpected in this setting: larger spaces, geyser ownership, summer and winter lead to increased scales, while non-ownership of appliances and not living in a formal urban area lead to decreased scales; we look at adults and children, separately. The scale adjustments are found to be higher for characteristics that are expected to raise energy expenditure, and lower for characteristics expected to lower energy expenditure. The underlying explanation of those results can be derived from the understanding that in order to purchase more energy, a household must have the resources to do so. For example, although relatively larger spaces require more energy, they require more of many other things, and, therefore, are owned by those who are better-off; thus, their energy shares tend to be lower leading to larger equivalence scales. On the other hand, a household that does not own electric appliances is likely worse-off, and, therefore, has energy shares that tend to be higher. Similarly, more adults in the household could be a signal that adults are pooling resources, because they have to; because they are poorer, their energy shares are likely higher. The effect of children, however, is more nuanced, with the first and fifth child not having much influence. For the rest, children are associated with an increased scale.

5.2 Estimates of energy equivalence scales

Using the semiparametric model estimates, we are able to determine each household's equivalence scale. We summarize them in Table 4 and Figure 3. We can also calculate each household's required energy consumption, which we discuss in the following subsection. The results in Table 3, suggest that expenditure is endogenous, and controlling for that endogeneity matters. However, the summary breakdowns presented Table 4 suggest that the effect of endogeneity on the underlying scales is rather small, at least for smaller values of the scale; however, there is evidence that the endogeneity effect could be large for relatively larger scales. We explore that further in Figure 3, where we illustrate a smoothed plot of the estimated equivalence scales against the log of total household expenditure; the regression was smoothed via a cubic spline on (log) expenditure.

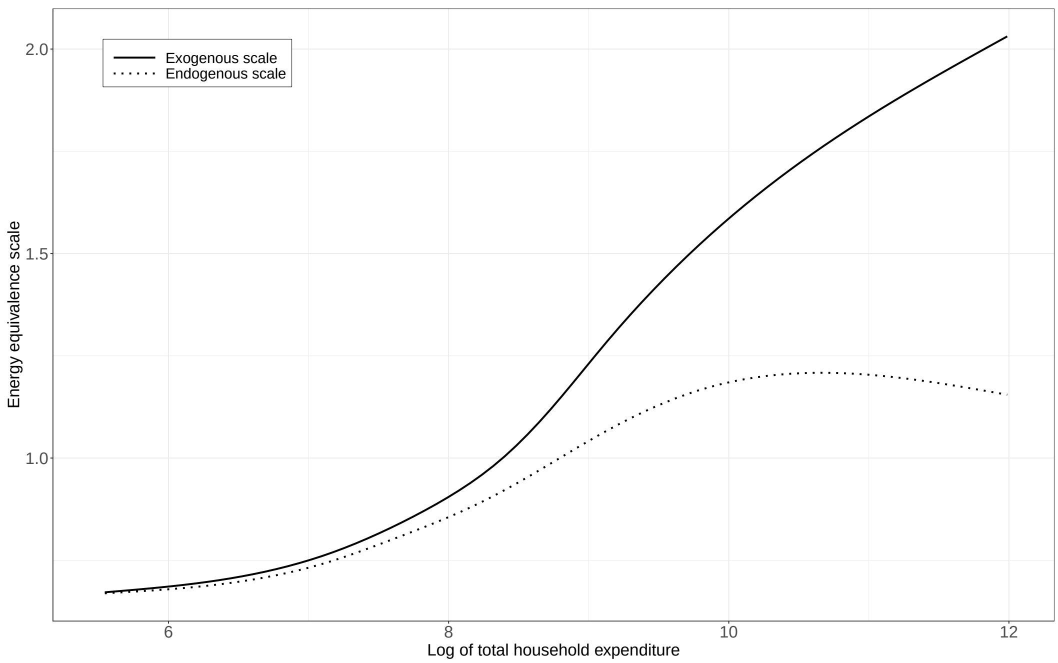
Table 4: Description of estimated equivalence scale and required energy expenditure

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Panel A: Expenditure assumed to be exogenous						
Equivalence scale	0.41	0.77	0.95	1.06	1.21	3.09
REC (unit: ZAR)	93.75	178.00	220.05	244.94	280.43	714.22
Panel B: Expenditure allowed to be endogenous						
Equivalence scale	0.39	0.76	0.92	0.92	1.10	1.44
REC (unit: ZAR)	91.22	174.72	211.62	213.57	253.42	332.01

REC (required energy consumption) is calculated from equation (3) as per household per month value.
N = 12 774.

Both Table 4 and Figure 3 suggest that the equivalence scale - expenditure gradient is positive, i.e., scales are relatively low for poor households and increase with expenditure. In fact, for very poor households, the scale tends to be less than one, which is intuitively appealing. The equivalence scale is a ratio adjustment to household expenditure, such that expenditure shares are equal across reference and nonreference households. Our results imply that poor household's income or expenditure should be divided by a number less than one; in other words, poor households do not have enough resources available to them; they are economizing on energy expenditures, at least in part, because they have other more pressing budgetary requirements. Finally, the illustration confirms that the endogeneity effect is not particularly big, except at the upper end. Furthermore, controlling for endogeneity flattens the gradient, i.e., the gradient is overestimated if we do not account for endogeneity.

Figure 3: Smooth plot of energy equivalence scales against total household expenditure



The most important take-away from these results is that the energy equivalence estimates are neither very similar to those reported for the UK nor similar to the OECD's income equivalence scales. Table 4 shows our estimated energy equivalence factors lies in between 0.41 and 3.09 (0.39 to 1.44 in the endogenous case), while the UK's energy equivalence factors range from (1.22-1.61) with single adult households as the reference group (BEIS and BRE, 2018). In other words, equivalence scales developed and used in one country may not be of use in another country, which is an important reason to develop a process that allows scales to be contextually estimated and in places where data might be limited, as is often the case in developing regions.

Although the reasons for the scale differences between South Africa and the UK or OECD are worthy of further investigation, the estimated differences are not entirely surprising. Firstly, as shown by the closeness of the budget share curves in Figure 2, our sampled South African households are not raising their energy spend share, when family size increases. In other words, in South Africa, there are large economies of scale in residential energy consumption: the scales are, on average, much less than actual household size. The endogeneity results suggest that part of this effect is due to an excluded factor that affects both total expenditure and household size; a limited range of opportunities or social capital are two such possibilities that might lead individuals to have very few resources, which they might attempt to pool with others, who similarly have very limited resources. There is some evidence that South African households form around sources of income (Schatz et al., 2015), and not necessarily around nuclear families, which could greatly influence the amount of money available for energy and other expenditure. Secondly, technology, weather, and, by extension, temperature and sunlight conditions in the UK and other OECD countries do not mimic those in South Africa. The former includes energy consumption inputs like insulation, which are not prevalent in South Africa, while the latter includes relatively milder weather for a larger swathe of the population for more of the year, reducing the need for indoor heating.

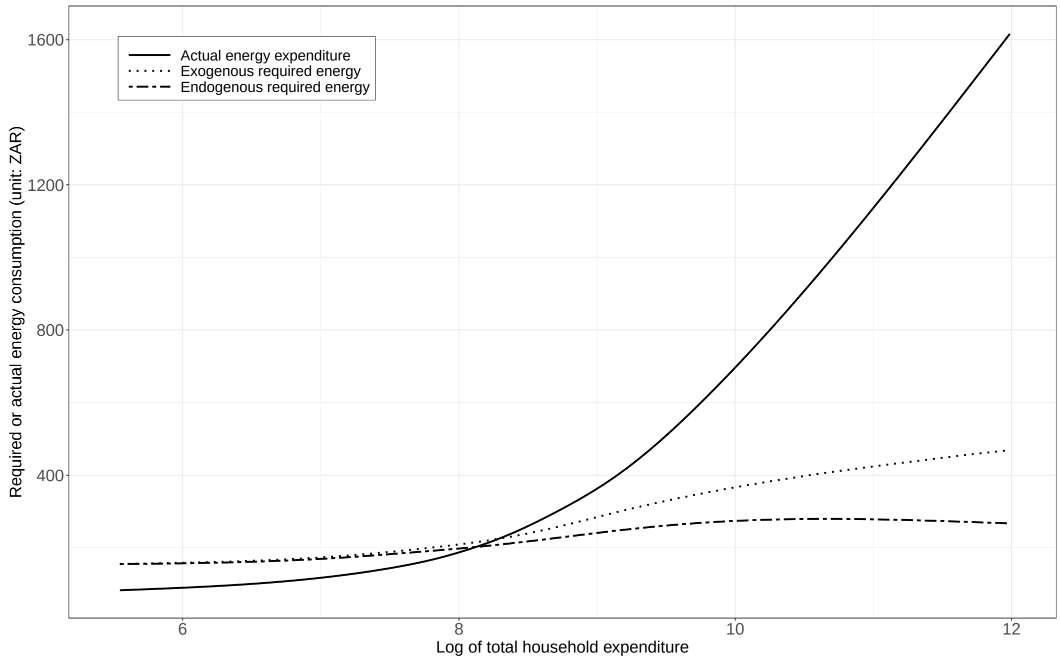
5.3 Estimates of required energy consumption

Given the scales, we turn our attention to household required energy. Recall that REC is the multiplication of reference energy consumption and the relevant energy equivalence scale (see equation (3)). As noted earlier, our reference, $\bar{E}_b$, was taken from the 75^{th} percentile of monthly energy expenditure in the reference group. Based on our data and estimates, the reference monthly energy requirement (in 2015) is ZAR 231.1. Given the structure of the analysis, using a higher (lower) percentile would increase (decrease) the estimated REC values. The REC is our estimate of how much energy expenditure a household requires for a reasonable standard of living, adjusting for differences in various characteristics relative to the reference household group. Hence, REC estimates differ by household, and, therefore, we present averages in the figure.

As was the case with the scales, we illustrate the average required energy (for exogenous and endogenous expenditure), along with average monthly energy expenditure, against (log) total household expenditure using a generalized additive model with a cubic spline over expenditure. The results are presented in Figure 4. The main take-away from this figure is that average actual energy expenditure rises more with income than the REC values. The result is both plausible and in line with an intuitive understanding of *required*, which should primarily be based on appliances, climate and family structure, rather than income.

The other main take-away is that poorer households are spending less than the estimated

Figure 4: Smooth plot of required and actual energy expenditure against total household expenditure



need (about half of their need at the low end), while richer households are consuming far more (in the region of four or five times more at the upper end) than their estimated needs. Our results, therefore, are quite different from those produced by the BREDEM model in the UK. The UK results suggested, implausibly, that many rich households were not meeting their energy needs; see Hills (2011) and Herrero (2017). Thus, the rescaling of the actual energy expenditure of a reference household group offers both a plausible and intuitive estimate of household energy requirements (REC) at different income levels.

In addition, according to descriptive statistics in Table 4, the estimated energy requirement lies in between ZAR 91.22 and ZAR 714.22 per household per month in 2014/2015. The mean REC value is lower than average actual energy expenditure (ZAR 306.11 in Table 2). Our required energy estimates are far below the modelled energy expenditure for the UK, which is between ZAR 1 947 and ZAR 2 876 per month, which can be partially explained by climactic factors.[16].

[16]The UK's estimated energy requirements were about GBP 107 to GBP 158 per month in 2009 (Hills, 2011) We convert the values to South African Rand by using the exchange rate 1 GBP ≈ 13 ZAR in 2009, and then inflate the ZAR value by 1.4 to match South African inflation from 2009 to 2015, such that the UK's estimation is comparable with the South Africa's in the year of 2015.

5.4 Discussion

We complete our analysis with a brief discussion of what the estimated REC implies, in terms of usage for a typical South African household, we describe what can be purchased by a two-adult and two-child household in an example using both the exogenous and endogenous REC. The average REC values for these two groups are: ZAR 228.19 for the exogenous REC and ZAR 214.9 for the endogenous. The average Eskom residential electricity price for 2014/2015 was 0.9806 ZAR/kWh (DOE, 2017), so these values equate to approximately 232 kWh and 219 kWh electricity consumption per household per month.[17]

The estimated REC is supposed to meet household basic energy needs for lighting, cooking, space heating and cooling, as well as social communication and entertainment. To have an in-depth understanding of the estimated REC, we conduct a further analysis on daily energy usage for home appliances. As shown in Table 5, each appliance's electricity consumption is the multiple of its rated power and duration of usage. Although the energy consumption patterns presented in Table 5 represent only one possible realisation out of various consumption patterns, it allows us to describe how the estimated REC could be utilised by family members for their basic energy needs. To summarise, required energy consumption covers a useful range of activities.

6. CONCLUSION AND POLICY IMPLICATIONS

This paper has developed an equivalence scale model for the determination of required energy consumption for households. The proposed approach rescales actual energy expenditure from a reference household group, accounting for differences in household structure, average weather, appliance ownership and dwelling size. The plausibility of the model has been examined using data from a recent South African household expenditure survey.

Two primary issues motivated this research. The first is that climate, appliance ownership and a number of other factors are likely to differ across countries, and, therefore, one would not expect either energy equivalence scales or required energy consumption from one country to necessarily be appropriate for another. The second is that data availability is often a problem, especially in developing countries, and, therefore, it might be difficult to develop requirements directly from actual energy usage (in kWh) patterns, dwelling energy efficiency and other details that are used in engineering

[17]This price is only applicable to Eskom's direct sales to the residential consumers. The distribution tariff applied by local authorities, who are resellers and use this activity for local revenue generation, is higher than the Eskom direct price.

Table 5: Example: Monthly electricity usage of household appliances in South Africa

Electrical appliance	Rated power (kW)	(1) Exogenous REC		(2) Endogenous REC	
		Duration (hours/day)	Power (kWh/day)	Duration (hours/day)	Power (kWh/day)
Lamp bulb	0.06	8.00	0.48	8.00	0.48
Electric stove	2.20	0.60	1.32	0.60	1.32
Microwave oven	1.00	0.20	0.20	0.23	0.23
Kettle	1.20	0.15	0.18	0.60	0.72
Refrigerator	0.40	6.00	2.40	6.00	2.40
Geyser	2.00	0.50	1.00	0.00	0.00
Washing machine	1.00	1.00	1.00	1.00	1.00
Iron	1.20	0.40	0.48	0.40	0.48
Television	0.15	4.00	0.60	4.00	0.60
Charger (for cell-phones)	0.01	7.00	0.07	7.00	0.07
kWh/day			7.73		7.3
kWh/month			232		219

Note: 1) REC refers to required energy consumption. 2) Monthly usage (kWh/month) is calculated as daily usage (kWh/day) multiplied by 30 days. 3) Information on rated power of each appliance is from Setlhaolo et al. (2014) and Setlhaolo and Xia (2015).

models to determine energy requirements in developed countries. Our approach addresses each of these concerns.

As expected, we find that our energy equivalence scales differ from both income equivalence scales and energy equivalence factors applied in developed country studies. The required energy consumption values that we derive from actual expenditure, a reference group and our estimated scales suggest that, on average, required energy consumption is well above actual energy expenditure for low- and middle-income groups, and well below actual for high-income groups. The results are consistent with an expectation that the basic energy needs of most poor households in South Africa have not been met, due to over-arching poverty in the country. These results stand in contrast to developed country studies using more complex engineering models, which have shown energy consumption by rich households to be less than what they need, a result that calls such methods into question.

Our study is the first to estimate required energy consumption following equivalence scale methods that can easily be applied to widely available household expenditure survey data. It can also be easily adapted to local data and circumstances. The scales and requirements that we are able to estimate offer value both to researchers and policymakers. For example, common energy poverty indicators use actual energy expenditure instead of required energy expenditure, because the modelling of required energy consumption has not generally been addressed. Our study, therefore, offers an important contribution, since it offers a way to determine required energy from actual consumption.

The suggested methods take into account family size and structure, as well as characteristics related to household energy needs. Such needs and characteristics can be adjusted for the context and the data at hand; therefore, our contribution is more general than the South African case study that we examine. Our method is expected to be especially beneficial, when detailed energy usage and housing energy efficiency data is not available, which is likely to be the case in the majority of settings.

With respect to policy, the proposed method is especially beneficial, because it offers policymakers a fairly easy way to determine household energy requirements, and potential energy subsidies/taxes that could be applied differentially. Importantly, those requirements would be based on local circumstances, rather than on the circumstances that were relevant for the UK or even South Africa. The method can offer policymakers more accurate comparisons, if they are interested in comparing domestic energy consumption across regions, since our adjustment allows for the incorporation of regional heterogeneity. Importantly, the definition of reference energy is flexible, and can be adjusted to suit policy goals; for example, one might be interested in subsidising solar geysers, as geysers are an important driver of energy consumption. Furthermore, since our method offers a way to estimate required energy, it offers a way forward, when it comes to energy poverty measurement, which is usually defined for required energy rather than actual. Such information can also help policymakers identify and more efficiently target subsidies to the benefit of households that are energy poor, and, thus, mitigate energy poverty.

Although the method is fairly general, our South African case study was limited in a few dimensions. For example, our analysis did not include any renters, primarily due to the broadly different underlying data concerns. For example, difficulties in splitting rent from utilities, as well as what seems to be a reporting error related to dwelling size across all renters; although there were supposed to be asked about estimated dwelling space, it appears that was not pursued adequately by the surveyors. In future research, we hope to be able to examine differences between renters and owners. Data limitations also reduced our ability to account for a wide range of domestic electrical appliances, their power usage and other drivers of domestic energy usage. Although the survey captured ownership of selected household appliances, detailed information related to the number and power of each appliance is not available. For that reason, we used common benchmarks in Table 5. Therefore, we believe there would be a benefit from a deeper survey that captured more information. Finally, in places where engineering models are available, a comparison of our model to that model would help develop a greater understanding of the pros and cons of each method.

ACKNOWLEDGEMENTS

We are grateful to the Editor, David C. Broadstock and three anonymous reviewers for their helpful comments. This research received funding from the *Centre of New Energy Systems* (CNES) at the University of Pretoria. We would also like to acknowledge the support of *Economic Research Southern Africa* (ERSA).

REFERENCES

Bednar, D. J. and T. G. Reames (2020). "Recognition of and response to energy poverty in the United States" *Nature Energy* 5(6): 432–439.

BEIS and BRE (2018). *Fuel poverty methodology handbook* Department for Business, Energy & Industrial Strategy (BEIS) and Building Research Establishment (BRE), London, UK.

Berkouwer, S. B. (2020). "Electric heating and the effects of temperature on household electricity consumption in South Africa" *The Energy Journal* 41(4).

Blackorby, C. and D. Donaldson (1993). "Adult-equivalence scales and the economic implementation of interpersonal comparisons of well-being" *Social Choice and Welfare* 10(4): 335–361.

Boardman, B. (1991). *Fuel poverty: from cold homes to affordable warmth* London, UK: Belhaven Press.

BRE (2015). *A technical description of the BRE domestic energy model, version 1.1* Building Research Establishment (BRE), UK.

Chakravarty, S. and M. Tavoni (2013). "Energy poverty alleviation and climate change mitigation: is there a trade off?" *Energy Economics* 40: S67–S73.

Charlier, D. and B. Legendre (2019). "A multidimensional approach to measuring fuel poverty" *The Energy Journal* 40(2): 27–53.

DCLG (2009). *English Housing Survey (EHS)* Department for Communities and Local Government (DCLG) London, UK https://www.gov.uk/government/collections/english-housing-survey.

Deaton, A. (1987). *The analysis of household surveys: a microeconometric approach to development policy* Washington, DC: World Bank.

Deaton, A. and M. Grosh (2000). "Consumption" in M. Grosh and P. Glewwe (editors), "Designing household survey questionnaires for developing countries," World Bank, Washington, DC 91–134.

Deaton, A. and J. Muellbauer (1980). *Economics and consumer behavior* Cambridge University Press.

DOE (2017). *South African Energy Price Report 2017* Department of Energy (DOE), Pretoria, South Africa.

DOE (2018). *South African energy efficiency label: appliance sales training for the retail sector* Department of Energy (DOE), Pretoria, South Africa https://www.savingenergy.org.za/wp-content/uploads/2018/05/Energy-Label-Learner-Guide.pdf.

Dong, Y. (2010). "Endogenous regressor binary choice models without instruments, with an application to migration" *Economics Letters* 107(1): 33–35.

Engel, E. (1895). "Die productions- und consumtionsverhältnisse des königreichs sachsen" in E. Engel (editor), "Die Lebenskosten belgischer Arbeiter-Familien, früher und jetzt," Dresden, Heinrich.

Hagenaars, A. J., K. de Vos, and M. A. Zaidi (1994). *Poverty statistics in the late 1980s: research based on micro-data* Eurostat (Statistical office of the European Communities).

Hayfield, T. and J. S. Racine (2008). "Nonparametric econometrics: the np package" *Journal of Statistical Software* 27(5): 1–32.

Heindl, P. (2015). "Measuring fuel poverty: general considerations and application to German household data" *FinanzArchiv: Public Finance Analysis* 71(2): 178–215.

Hernández, D. (2016). "Understanding 'energy insecurity' and why it matters to health" *Social Science & Medicine* 167: 1–10.

Herrero, S. T. (2017). "Energy poverty indicators: a critical review of methods" *Indoor and Built Environment* 26(7): 1018–1031.

Hills, J. (2011). *Fuel poverty: the problem and its measurement* Department of Energy and Climate Change (DECC), London, UK.

Hills, J. (2012). *Getting the measure of fuel poverty: final report of the fuel poverty review* Department of Energy and Climate Change (DECC), London, UK.

Ichimura, H. (1993). "Semiparametric least squares (SLS) and weighted SLS estimation of single-index models" *Journal of Econometrics* 58(1-2): 71–120.

Imbens, G. W. and W. K. Newey (2009). "Identification and estimation of triangular simultaneous equations models without additivity" *Econometrica* 77(5): 1481–1512.

Klein, R. W. and R. H. Spady (1993). "An efficient semiparametric estimator for binary response models" *Econometrica* 61(2): 387–421.

Koch, S. F. (2018). "Catastrophic health payments: does the equivalence scale matter?" *Health Policy and Planning* 33(8): 966–973.

Koch, S. F. and G. Tshiswaka-Kashalala (2018). "The demand for reproductive health care" *Journal of African Economies* 27(4): 405–429.

Kyprianou, I., D. Serghides, A. Varo, J. Gouveia, D. Kopeva, and L. Murauskaite (2019). "Energy poverty policies and measures in 5 EU countries: a comparative study" *Energy and Buildings* 196: 46–60.

Lazear, E. P. and R. T. Michael (1980). "Family size and the distribution of real per capita income" *American Economic Review* 70(1): 91–107.

Legendre, B. and O. Ricci (2015). "Measuring fuel poverty in France: which households are the most fuel vulnerable?" *Energy Economics* 49: 620–628.

Leibbrandt, M., A. Finn, and M. Oosthuizen (2016). "Poverty, inequality, and prices in post-apartheid South Africa" in

C. Arndt, A. McKay, and F. Tarp (editors), “Growth and Poverty in Sub-Saharan Africa,” Oxford University Press, Oxford, UK 393–417.

Lewbel, A. (2012). “Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models” *Journal of Business & Economic Statistics* 30(1): 67–80.

Lewbel, A. and K. Pendakur (2008). “Equivalence Scales” in S. N. Durlauf and L. E. Blume (editors), “The New Palgrave Dictionary of Economics: Volume 1 – 8,” Palgrave Macmillan UK, London 1832–1835.

Li, K., B. Lloyd, X.-J. Liang, and Y.-M. Wei (2014). “Energy poor or fuel poor: what are the differences?” *Energy Policy* 68: 476–481.

Li, Q. and J. Racine (2004). “Cross-validated local linear nonparametric regression” *Statistica Sinica* 14: 485–512.

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

Liddell, C., C. Morris, S. McKenzie, and G. Rae (2012). “Measuring and monitoring fuel poverty in the UK: national and regional perspectives” *Energy Policy* 49: 27–32.

Memmott, T., S. Carley, M. Graff, and D. M. Konisky (2021). “Sociodemographic disparities in energy insecurity among low-income households before and during the COVID-19 pandemic” *Nature Energy* : 1–8.

Meyer, J. P. (2000). “A review of domestic hot-water consumption in South Africa” *R & D Journal* 16(3): 55–61.

Mohr, T. M. (2018). “Fuel poverty in the US: evidence using the 2009 Residential Energy Consumption Survey” *Energy Economics* 74: 360–369.

Nelson, J. A. (1988). “Household economies of scale in consumption: theory and evidence” *Econometrica* 56(6): 1301–1314.

Nicholson, J. L. (1976). “Appraisal of different methods of estimating equivalence scales and their results” *Review of Income and Wealth* 22(1): 1–11.

Ntaintasis, E., S. Mirasgedis, and C. Tourkolias (2019). “Comparing different methodological approaches for measuring energy poverty: evidence from a survey in the region of Attika, Greece” *Energy Policy* 125: 160–169.

Okushima, S. (2017). “Gauging energy poverty: a multidimensional approach” *Energy* 137: 1159–1166.

Pachauri, S. and D. Spreng (2004). “Energy use and energy access in relation to poverty” *Economic and Political weekly* 39(3): 271–278.

Papada, L. and D. Kaliampakos (2018). “A stochastic model for energy poverty analysis” *Energy Policy* 116: 153–164.

Papada, L. and D. Kaliampakos (2020). “Being forced to skimp on energy needs: a new look at energy poverty in Greece” *Energy Research & Social Science* 64: 101450.

Papke, L. E. and J. M. Wooldridge (1996). “Econometric methods for fractional response variables with an application to 401(k) plan participation rates” *Journal of Applied Econometrics* 11(6): 619–632.

Pendakur, K. (1999). “Semiparametric estimates and tests of base-independent equivalence scales” *Journal of Econometrics* 88(1): 1–40.

Pollak, R. A. and T. J. Wales (1979). “Welfare comparisons and equivalence scales” *American Economic Review: Papers and Proceedings* 69(2): 216–221.

R Core Team (2021). *R: A Language and Environment for Statistical Computing* R Foundation for Statistical Computing Vienna, Austria.

Racine, J. and Q. Li (2004). "Nonparametric estimation of regression functions with both categorical and continuous data" *Journal of Econometrics* 119(1): 99–130.

Rademaekers, K., J. Yearwood, A. Ferreira, S. Pye, I. Hamilton, P. Agnolucci, D. Grover, J. Karásek, and N. Anisimova (2016). *Selecting indicators to measure energy poverty* Trinomics Rotterdam, The Netherlands.

Rao, N. D. and J. Min (2018). "Decent living standards: material prerequisites for human wellbeing" *Social Indicators Research* 138(1): 225–244.

Robinson, C. (2019). "Energy poverty and gender in England: a spatial perspective" *Geoforum* 104: 222–233.

Schatz, E., S. Madhavan, M. Collinson, F. X. Gómez-Olivé, and M. Ralston (2015). "Dependent or productive? A new approach to understanding the social positioning of older South Africans through living arrangements" *Research on Aging* 37(6): 581–605.

Setlhaolo, D. and X. Xia (2015). "Optimal scheduling of household appliances with a battery storage system and coordination" *Energy and Buildings* 94: 61–70.

Setlhaolo, D., X. Xia, and J. Zhang (2014). "Optimal scheduling of household appliances for demand response" *Electric Power Systems Research* 116: 24–28.

Stats SA (2017a). *Living Conditions Survey 2014-2015 [dataset]. Version 1.1* Statistics South Africa (Stats SA) Pretoria: Statistics South Africa (Stats SA) [producer]. Cape Town: DataFirst [distributor], 2017.

Stats SA (2017b). *Living Conditions Survey 2014/2015 metadata* Statistics South Africa (Stats SA) Pretoria, South Africa.

United Nations (2021). *The Sustainable Development Goals Report 2021* New York: United Nations.

Wang, M.-C. and J. van Ryzin (1981). "A class of smooth estimators for discrete distributions" *Biometrika* 68(1): 301–309.

Welsch, H. and P. Biermann (2017). "Energy affordability and subjective well-being: evidence for European countries." *The Energy Journal* 38(3): 159–176.

Yatchew, A., Y. Sun, and C. Deri (2003). "Efficient estimation of semiparametric equivalence scales with evidence from South Africa" *Journal of Business & Economic Statistics* 21(2): 247–257.

Ye, Y. and S. F. Koch (2021). "Measuring energy poverty in South Africa based on household required energy consumption" *Energy Economics* 103: 105553.

Ye, Y., S. F. Koch, and J. Zhang (2018). "Determinants of household electricity consumption in South Africa" *Energy Economics* 75: 120–133.

Ye, Y., S. F. Koch, and J. Zhang (2020). "Modelling required energy consumption with equivalence scales" Economic Research Southern Africa Working Paper No 809.