Equivalence scales with endogeneity and base independence

Steven F. Koch

ERSA working paper 874

February 2022

Equivalence scales with endogeneity and base independence*

Steven F. Koch†

05 December, 2021

Abstract

This study estimates food budget share equations to calculate household equivalence scales that address both base-independence and potential endogeneity, even though an instrument that satisfies the usual exclusion restriction may not be available. The application incorporates semi-parametric methods, control functions and heteroscedasticity instrumentation. The application is founded on the most recent income and expenditure data that is available for South Africa. We find that endogeneity matters, and that failing to account for it leads to overstated equivalence scales in nearly every circumstance. When we fit our calculated scales to a typical $(A + \kappa K)^{\psi}$ equivalence structure via non-linear least squares, we find values of κ near unity and values of ψ mostly below 0.5. Thus, our analysis suggests that a square-root scale is more appropriate than other scales that have been used to examine poverty and inequality in South Africa.

1 Introduction

In this paper, we present estimates of equivalence scales in a developing country with extensive inequality – South Africa – making use of the 2014/15 Living Conditions Survey (Stats SA 2017), which is the most recent survey capturing the necessary information. Our estimates are underscored by base-independence, which we incorporate through a linear index semi-parametric model of each household's food budget share. We further extend our nonlinear share model to account for the potential endogeneity of household expenditure. Neither base-independence nor endogeneity corrections feature in the developing country literature on equivalence scales; in reality, the estimation of equivalence scales in those settings is also rare, due to the fact that scales are estimated indirectly. Our models include categorical household structure variables, rather than continuous measures, as outlined by Deaton (1987). We also incorporate a broad range of location and household-specific information. Thus, from our results, we can calculate scales that vary across all regions

*The author would like to thank Jeff Racine and Yuxiang Ye for their thoughts and constructive discussions related to earlier work on this topic without implicating them for any mistakes that may exist. The author would also like to thank Adonis Yatchew for his coding help, when the author first considered this research.

†Corresponding author: Department of Economics, University of Pretoria, Private Bag X20, Hatfield, South Africa; +27-12-420-5285; steve.koch@up.ac.za

of the country and account for many other household preference differences not previously addressed in the literature; we limit our presentation of scales in this analysis, however, in order to save space.

Equivalence scales are used widely in the applied literature to adjust household income or expenditure to account for the number of dependents in the household, and allow both researchers and policymakers to compare income or expenditure across households, between regions and through time (Aaberge and Melby 1998; Economic Co-operation and Development 2011). Equivalence scales also underpin means-tests for government transfers, tax policy – especially exemptions for dependents – and subsidy policy, such as those that might focus on support for young children. Equivalence scales are also used in the estimation of catastrophic health expenditures (Ataguba 2012; Koch and Setshegetso 2020, 2021), although Koch (2018) suggests that the choice of scale does not make much difference in the underlying estimation of catastrophe, when following the World Health Organization method (Xu 2005). Ye and Koch (2021) and Ye, Koch, and Zhang (forthcoming) adopt the concept to determine household energy requirements and measure energy poverty.

Although widely used, much of the applied literature uses (seemingly) arbitrary values (May, Carter, and Posel 1995; May et al. 1995; Woolard and Leibbrandt 2001; Posel, Casale, and Grapsa 2016), is underscored by estimates from quite a number of years ago (Woolard and Leibbrandt 1999; Woolard 2002; Yatchew, Sun, and Deri 2003) or does not use the most recent data (Posel, Casale, and Grapsa 2020; Daley et al. 2020), at least in the case of South Africa. One commonly applied arbitrary equivalence scale is the "per capita" scale. It implies that a household with four adults is equivalent in its "needs" to a household with one adult and three young children. Similarly, it implies that a female-headed household with one adult and two children is equivalent to a male-headed household with two adults and one child. However, Posel, Casale, and Grapsa (2016) argue that male-headed and female-headed households are different, primarily because female-headed households have more children, on average, than male-headed households and tend to be poorer. Research across approximately 90 countries, although not focused on equivalence scale estimation, supports this view (Munoz Boudet et al. 2018, 2021). Similarly, Posel, Casale, and Grapsa (2020) argue that there are ethnic differences in household structure and the distribution of income/expenditure that justify further variation in equivalence scales by ethnicity. In what follows, we extend these arguments further, suggesting that, very broadly, household preferences are likely to be heterogeneous, and, therefore, equivalence scales should be allowed to vary by more than just the composition of children and adults in the household.

Furthermore, research in developed countries has not generally considered base independence − Yatchew, Sun, and Deri (2003) is a notable exception. However, their analysis is based on data from 1993.[1] Base

[1] Hasan (2016) estimates Engel curves for Bangladesh applying some of the methods in Yatchew, Sun, and Deri (2003);

independence arises from the early theoretical literature on duality – see Pollak and Wales (1979), for example – although Blundell and Lewbel (1991) clarified that it is only possible to determine relative equivalence scales, if one is willing to make a functional form assumption.[2] Their assumption is that the cost function is separable in household characteristics and utility, while Blackorby and Donaldson (1993) force any monotonic transformation of utility to not include a measure of the demographic structure of the household. Importantly, Blundell and Lewbel (1991) show that base-independent equivalence scales can not be estimated within the context of the almost ideal demand system, which has Working-Leser share functions. However Working-Leser share functions often lie behind the calculation of equivalence scales (Deaton 1987; Posel, Casale, and Grapsa 2020).[3] In our research, we estimate equivalence scales assuming base independence, and we compare those scales to estimates that would arise, if it was not assumed.

Finally, as noted by Summers (1959), household expenditure choices are part of total expenditure, and, therefore, expenditure is likely to be endogenous in any engel curve model, due to simultaneity. However, the developing country literature on equivalence scales has not adjusted for potential endogeneity; as noted previously, equivalence scale estimation in developing countries is also not that common. Although instrumentation is the standard correction, one may struggle to find instruments. If one is willing to make a two-stage budgeting decision, income will be correlated with expenditure, but should not affect shares directly (Banks, Blundell, and Lewbel 1997). If the two-stage budgeting assumption is not reasonable, income will not be an appropriate instrument. However, recent advances in the literature open the door to different identification assumptions (Lewbel 2012; Dong 2010). Thus, we also contribute an application to the literature that corrects for endogeneity in expenditure, without assuming that income is purely exogenous.

Our results suggest that (a) scale estimates underpinned by base-independence are lower than those underpinned by the Working-Leser method outlined in Deaton (1987), as well as previous estimates that follow it (Posel, Casale, and Grapsa 2020); (b) correcting for endogeneity leads to (almost exclusively) lower scale estimates, and that the reductions can be both economically and statistically significant; (c) there are potentially meaningful ethnic, gender and location differences in the estimated scales; and (d) a square-root scale is a reasonable approximation to what is observed in the data, which suggests the need to re-examine estimates of inequality (Posel, Casale, and Grapsa 2020).[4]

however, he does not report equivalence scales.

[2]Even though demands can be uniquely recovered from cost functions, cost functions (or expenditure functions) cannot be uniquely recovered from demand curves. Further, demands tell us about preferences related to household structure in consumption space, while equivalence requires information about preferences in household structure - consumption space. Because these spaces are not the same, identification is not generally unique. However, cost of living indexes are estimable, as the ratios of cost of living indexes for different types of households.

[3]Recent research has offered theoretical extensions that allow for some dependence between scales and income (Donaldson and Pendakur 2006), or has made use of satisfaction data, rather than income and expenditure data (Abanokova, Dang, and Lokshin 2020). We leave those considerations for future research.

[4]Undertaking separate models for the estimation of equivalence scales, as is done by Posel, Casale, and Grapsa (2020), for

2 Methods

The estimation of equivalence scales is indirect, driven by the underlying assumption that the share of food in the budget is a reasonable indication of household welfare (Engel 1857). Figure 1, which illustrates the relationship between the household budget share and (the log of) total household expenditure, supports this assumption, because the share falls as total expenditure rises. The share, itself, is, of course, more than a function of total expenditure. Of particular interest is the association between household structure characteristics, including the number of children and the number of adults, and the food share. We allow for characteristics – such as population group, provincial and urban/rural locales, and the gender of the household head – to control for preference heterogeneity and other potential determinants of the budget share.

Defining the food budget share as the ratio of food expenditure to total expenditure ($w_i = x_{fi}/x_i$) yields the following relationship, where the share is a function of (the log of) total expenditure, household structure characteristics ($\mathbf{D}_i$), such as the number of children and adults, and other characteristics ($\mathbf{Z}_i$).

$$w_i = w(\ln x_i, \mathbf{D}_i, \mathbf{Z}_i) \tag{1}$$

2.1 Linear budget share

For the analysis, we consider two specifications. The first of which is linear in the paramaters. We alter the typical approach summarised by Deaton (1987), which underpins recent research (Posel, Casale, and Grapsa 2020). Our first modification is the use of a series of binary variables capturing household structure. Thus, we have binary indicators for the number of children and the number of adults in the household.[5] Our second modification, noted above, is to use additional controls likely to influence preferences and expenditure behaviour. These modifications lead to the linear share regression in (2), where ε_i is the error; potential endogeneity is discussed below.

$$w_i = \alpha_0 + \alpha_1 \ln x_i + \sum_j \gamma_j Z_{ij} + \sum_j \rho_j D_{ij} + \varepsilon_i, \tag{2}$$

example, implies some form of selection. In the case of South Africa, where different ethnic groups were treated differently (and likely still are, despite legislative changes), selection is plausible. However, since the underlying reference groups in such an analysis are not the same, and there is no selection mechnanism estimated, whether or not those groups can be aggregated for the estimation of population level inequality is not addressed. Thus, it is not entirely clear what their approach identifies. We leave that as an open question for future research.

[5]For a household with two adults and three children, the separate binary indicators “two adults” and “three children” will be turned on, while all the other indicators, such as “three adults”, “four adults”, … , and “one child”, “two children”, “four children, … are switched off. Please, see the empirical results in Table A.1 for all of the binary indicators included in the model.

The assumption that the food share is a reasonable welfare proxy underpins the analytic approach to the indirect estimation of scales from equation (2), which is straightforward to estimate. Intuitively, the goal is to determine the addition or reduction of funds that would be needed by the household to reach the same level of welfare (as measured by the expenditure share) as some reference household. In order to calculate the scale, we set the reference share equal to a generic "other" household, as in (3), and solve for the expenditure ratio. For this exercise, the reference household has one adult and no children; therefore, the fitted reference estimate does not include any D_{ij}.

$$\alpha_1 + \beta_1 \ln x_i^a + \sum_j \gamma_j Z_{ij}^a + \sum_j \rho_j D_{ij}^a = \alpha_1 + \beta_1 \ln x^r + \sum_j \gamma_j Z_j^r. \tag{3}$$

Rearranging (3) determines the relative income ratio, and, thus, the equivalence scale, which we denote by Λ^E for this linear Engel-type scale.

$$\Lambda_i^E = \frac{x_i^a}{x^r} = \exp\left[\frac{1}{\beta_1}\left\{-\sum_j \rho_j D_{ij}^a - \sum_j \gamma_j \left(Z_{ij}^a - Z_j^r\right)\right\}\right] \tag{4}$$

Thus, the equivalence scale is determined by the adult/child make-up of the non-reference household and the relative differences in other characteristics across the households. In practice, we define the reference household so that all characteristics ($\mathbf{Z}_i$) are "off", as well, which further simplifies the calculation. Since the scales in (4) are calculated from estimates, in order to determine the variability of these scales, they must be bootstrapped. We undertake 399 parametric bootstrap replications to determine the variability of the scales.

2.2 Semi-parametric budget share

The second specification modifies Yatchew, Sun, and Deri (2003), who offer a relatively simpler way to impose base-independence, see below, than suggested by Pendakur (1999). Yatchew, Sun, and Deri (2003) include just a few variables and parameters, and their approach allows for estimation across a range of households; Pendakur (1999) estimates pairwise by household type. We offer a more detailed derivation of base-independence in Appendix C. Our approach is also similar to Ye and Koch (2021), although their focus is required energy expenditure and energy poverty. Similar to Ye and Koch (2021), we extend Yatchew, Sun, and Deri (2003) to include binary indicators for household structure and other characteristics. Specifically, we consider a semi-parametric model of the food share that includes a linear index function, where f is unknown, and the remaining variables are as above; ν is the error and potential endogeneity is addressed below.[6]

$$
\begin{aligned}
w_i &= f\left(\ln x_i - \Delta_i\right) + \nu_i \\
&= f\left(\ln x_i - \sum_j \xi_j Z_{ij} - \sum_j \delta_j D_{ij}\right) + \nu_i
\end{aligned}
\tag{5}
$$

Identification in the semi-parametric model requires the coefficient on $\ln x$ in equation (5) to be one, and helps clarify base independence. By fixing the parameter on $\ln x$, the addition or reduction to $\ln x$ needed to equate one household's share to another is based solely on the estimated average shifts in the relationship between log expenditure and the budget share; these are driven by the differences in the characteristics ($\mathbf{D}_i$) and ($\mathbf{Z}_i$) in equation (5).[7] Although, as in the linear setting, the scales are determined from setting shares equal across different household types, the semi-parametric approach offers a more direct route to the scales. The estimates from the model need only be exponentiated, while the minus sign in equation (5) must also be accommodated. Allowing for a separate parameter on ln expenditure in the estimation routine would result in two dimensions of variability (unknown f and unknown parameter) with only one control variable ln expenditure); thus, estimates would not be available. Instead of allowing for freedom in $\ln x$, differences in the characteristics trace out differences in the underlying expenditure-share relationship; subsequently, these are the expenditure adjustment parameters and underpin the scales.

Since the function f is not known, it will be estimated nonparametrically. The combination of the linear index and the unknown function yields a semi-parametric model that can be estimated following Ichimura (1993). Rather than using difference procedures for estimation, as suggested by Yatchew, Sun, and Deri (2003), we make use of bandwidth and leave-one-out cross-validation procedures, which are implemented via

[6] Base independence also allows for vertical translations, such that $\nu_i = \nu(p_i) + \epsilon_i$, where p represents prices; however, price data is problematic in this setting for a variety of reasons. For example, food purchases are recorded as expenditure, rather than items, while prices are separately available for a subset of items that are sold in the country, and not necessarily for the items that make up expenditure. In future research, we will work to match food expenditures and prices across the country to the purchases made, in an effort to also capture these vertical translations. We also note, however, that it will require us to examine another dimension of potential endogeneity that relates to the effect of measurement error in the price data. Instead, we note that expenditure captures both prices and quantities, such that the horizontal translations we estimate are averaging over the price effects.

[7] Equation (5) clarifies the idea that we are estimating a function over the linear index $\ln x_i - \Delta_i$, such that Δ_i represents horizontal translations of the functional relation f. In other words, Δ represents the change to log expenditure required to equate food shares to a baseline household type (where $\Delta = 0$). Thus, setting shares equal, assuming $\Delta = 0$ for the reference household, yields

$$
w_0 = f(\ln x_0) = f(\ln x_1 - \Delta_1) = w_1 \tag{6}
$$

Since the functions are the same, straightforward manipulation that assumes f can be inverted shows that Δ is the log of the relevant equivalence scale.

$$
\begin{aligned}
\ln x_0 &= \ln x_1 - \Delta \\
\Delta &= \ln x_1 - \ln x_0 \\
\exp \Delta &= \frac{x_1}{x_0}
\end{aligned}
\tag{7}
$$

When all of the characteristics are "off", f is the average reference household nonlinear relationship between the log of expenditure and the budget share. If one of those characteristics were to be turned back "on", the estimate of that characteristic would represent the average horizontal shift from the reference household average nonlinear relationship.

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 binary nature of all of the control variables, which we describe below, yields equivalence scales that can be easily calculated. The scale, which we label Λ^S to connote the underlying semi-parametric model, will be the exponentiated value of the sum of the estimates of characteristics that differ from the reference household's characteristics (recall the reference value $\Delta = 0$, as are $\mathbf{D}$ and $\mathbf{Z}$, as described in footnote 7). Standard errors are calculated via the Delta method.

$$\Lambda_i^S = \exp\left(-\sum_j \delta_j D_{ij} - \sum_j \xi_j Z_{ij}\right). \tag{8}$$

2.3 $(A + \kappa K)^\psi$

One popular scale estimator is $(A + \kappa K)^\psi$, where A represents the number of adults in the household, K represents the number of children, κ represents the child cost (which is generally assumed to be less then one, implying that children require fewer resources than adults) and ψ represents household economies of scale, which are also assumed to be less than one. This assumption implies that even though food, itself, is exclusive, there are public good components in food consumption, such as the ability to purchase in bulk (Posel, Casale, and Grapsa 2020). Nicholson (1976) notes that child cost estimates can be overstated, due to the fact that child costs are disproportionately food- and other necessity-related.

Early literature in South Africa assumed $\kappa = 0.5$ and $\psi = 0.9$, based on May, Carter, and Posel (1995). On the other hand, the World Health Organization placed young child calorie requirements at 64% of an adult (Woolard and Leibbrandt 1999). Posel, Casale, and Grapsa (2016) applied a range of values, including $\kappa = \{0.5, 0.9\}$, separately for younger and older children, as well as $\psi = \{0.7, 0.9, 1\}$, while Posel, Casale, and Grapsa (2020) estimate κ to be close to 0.9 and ψ to be close to 0.8. Our results, below, suggest that economies of scale parameter is smaller, and that the child cost is likely higher.

We estimate κ and ψ via non-linear least squares, based on each of the preceding models for comparative purposes. Specifically, we predict scales for all family types $\mathbf{D}$, ignoring the $\mathbf{Z}$ effects (i.e., setting each $Z_{ij} = 0$ at the equivalence scale calculation phase for all households, although it should be noted that we do

[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. Code for all of the analysis is available from the authors upon request.

estimate basic models that do not include **Z** variables). Thus, we predict:

$$\begin{aligned}\hat{\Lambda}_i^S &= \exp\left(-\sum_j \delta_j D_{ij}\right)\\ \hat{\Lambda}_i^E &= \exp\left[\frac{1}{\beta_1}\left\{-\sum_j \rho_j D_{ij}^a\right\}\right]\end{aligned} \tag{9}$$

Given $\hat{\Lambda}_i^\ell$, where $\ell \in \{E, S\}$, we estimate, via non-linear least squares,

$$\hat{\Lambda}_i^\ell = (A + \kappa^\ell K)^{\psi^\ell} + u_i. \tag{10}$$

Because the scales are predicted (and, in this case, only from household size and composition information), the estimation data will incorporate numerous replicated observations. Thus, the estimation process will automatically place more weight on household types that are more common, and, thus, the reported estimates should be understood in that context. Furthermore, estimation with categorical controls requires at least one to be the refernce. In our models, the reference household is a single female-headed, African household located in an urban area of the Western Cape. Thus, the reported scales for the "Additional Control" models relate to female-headed, African households located in urban areas of the Western Cape.

2.4 Endogeneity

The final concern to be addressed in the analysis 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 if it were simultaneous to expenditure choices (Summers 1959). One might also worry that household size is endogenous, especially if household size is reacting to income opportunities (Edmonds, Mammen, and Miller 2005; Klasen and Woolard 2008). In the case of the former, and assuming two-stage budgeting, household income could be used as an instrument. In the case of the latter, income is not expected to meet the exclusion criteria, and, therefore, alternative exclusion restrictions or other identifyind assumptions are necessary.

Within the linearized version of the model, the heteroskedasticity instrumental variable procedure (Lewbel 2012) is used to control for the endogeneity of expenditure. The model is estimated using the `REndo` package (Gui et al. 2021), which also includes instrumental variable performance tests from the `AER` package (Kleiber and Zeileis 2008). In the nonlinear setting, one could follow Imbens and Newey (2009), which is 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. However, since an instrument may not be readily available, 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 control that has a large support - we use income; importantly, income is not assumed to be purely exogenous. The control function in the second stage is the residual from a nonparametric regression of expenditure against income, **D** and **Z** from the model.[9] The binary response model in Dong (2010) is estimated semi-parametrically, as in Klein and Spady (1993), which is a binary version of the continuous model we estimate. Thus, the procedure is directly applicable in our setting.

3 Data

We apply the above methods to the South African Living Conditions Survey (LCS) 2014/2015 (Stats SA 2017). As its name implies, it is meant to contribute to the understanding of living conditions and poverty in South Africa. For our purposes, it captures household expenditure, household size and structure and some information on gender, ethnicity and household location, which we use in our analysis. The LCS uses classification of individual consumption by purpose (COICOP) categories, and food expenditures lie in Category 01, which we use for the analysis; every subcategory of food expenditure is summed within a household; we do not include food purchased away from home, which is a separate category.

The food budget share is the proportion of the household budget devoted to food consumption. We use consumption expenditure instead of income for all the estimates in the analysis, because it can be more practical to use consumption rather than income. Deaton and Grosh (2000), for example, argue that measured consumption is smoother than measured income, because, in developing countries, fewer employees are under long-term formal employment contracts and home production can be extensive. However, as outlined above, we do use income as a continuous control with wide support, when dealing with endogeneity.

In South Africa, or at least in this data, there are few differences between total consumption and total consumption in-kind. Despite that, for the analysis, we use household consumption expenditure capturing both monetary and in-kind payment for all goods and services, and the money value of the consumption of home-made products. The data was captured over the course of a year, with different samples in different provinces used. Due to the timing, all reported expenditures are inflated/deflated to April 2015, the midpoint of the survey year, using the consumer price index.[10]

[9] For the nonparmatric estimation, we follow Li and Racine (2004), which is implemented by the `np` package (Hayfield and Racine 2008) in `R` (R Core Team 2021). We applied an epanechnikov kernel for the continuous variables and the Li and Racine (2007) kernel for categorical/discrete variables. That kernel is an extension of the kernel proposed by Wang and van Ryzin (1981) that works well for both ordered and unordered discrete variables.

[10] The data from the survey is collated in a number of files, including a person file, a household file and an expenditure and

4 Results

4.1 Descriptives

We begin by describing the data used in the analysis, which we present in Table 1. The initial data included 23380 households. However, we lose 312 observations, due to missing data on included variables, as well as zero recorded income or expenditure. We find that the average household is just under 4, and that it is made up of approximately 2.5 adults and just more than one child. Male-headed households are slightly more common than female-headed households, while households are most likely to be headed by Africans (85%) or Coloureds (11%); the former is the majority in the country, and in official statistics are referred to as Black-Afrcans, while the latter is the second largest population group.

Table 1: Descriptive statistics of analysis data, N = 23068

Statistic	Mean	St. Dev.	Min	Max
Household Size	3.73	2.31	1	14
Kids in HH	1.16	1.36	0	6
Adults in HH	2.57	1.43	1	8
Male HH Head	0.55	0.50	0	1
African HH Head	0.81	0.39	0	1
Coloured HH Head	0.11	0.31	0	1
Asian HH Head	0.02	0.14	0	1
White HH Head	0.06	0.24	0	1
Western Cape	0.12	0.32	0	1
Eastern Cape	0.13	0.33	0	1
Northern Cape	0.06	0.23	0	1
Free State	0.09	0.29	0	1
KwaZulu-Natal	0.16	0.36	0	1
Northwest Province	0.09	0.28	0	1
Gauteng Province	0.14	0.35	0	1
Mpumalanga Province	0.10	0.30	0	1
Limpopo Province	0.12	0.33	0	1
Urban Formal	0.54	0.50	0	1
Urban Informal	0.07	0.25	0	1
Traditional Area	0.35	0.48	0	1
Rural Formal	0.04	0.19	0	1
Monthly Log Expenditure	8.32	1.02	4.79	12.30
Monthly Log Income	8.42	1.24	2.30	13.40
Food Share	0.25	0.17	0.0002	0.93

In terms of location, households are spread across all of the provinces, but are most likely to be found in traditional areas (which are most easily described as rural villages) and urban formal areas (primarily towns

income file. For the analysis, we use `haven` (Wickham and Miller 2021), the `tidyverse` (Wickham et al. 2019), `stargazer` (Hlavac 2018), `knitr` (Xie 2014), `kableExtra` (Zhu 2021) and `rmarkdown` (Xie, Dervieux, and Riederer 2020), which are packages for `R`, to organize the data for the analysis, prepare the data in tables and write the paper in a completely repeatable manner (Racine 2019). Code for the preparation of the data, figures, tables and all empirical modelling is available from the authors, upon request.

and cities). We do see, however, that informal locations are non-negligible, as nearly 7% of households are found in undeveloped areas, broadly labeled this way. The remainder of households live in primarily agricultural areas, referred to as formal rural areas.

Finally, as suggested by Deaton and Grosh (2000), household income is more varied than household expenditure, which offers us support in its use in the analysis. Relatedly, we see that household income has wider support than household expenditure, which validates its use in the endogeneity procedure outlined by Dong (2010) that we apply here.

4.2 Food shares

As we see in Table 1, the average proportion of food in the budget is 0.25, but the average does not tell the complete story. Figure 1 illustrates bivariate nonparamtric regressions of the food share against the log of household expenditure for four different family types. Overall, the relationship between the food share and total household expenditure is negative, implying that the share of food in the budget is a reasonable measure of welfare. We also see that food shares at nearly every level of total expenditure are larger for larger households, while the increase appears to be smaller as household size increases, implying that there are household economies of scale. However, we explore both of those features, and others, more carefully in the analysis, since bivariate regressions do not control for other factors that might also influence the budget share - total household expenditure relationship.

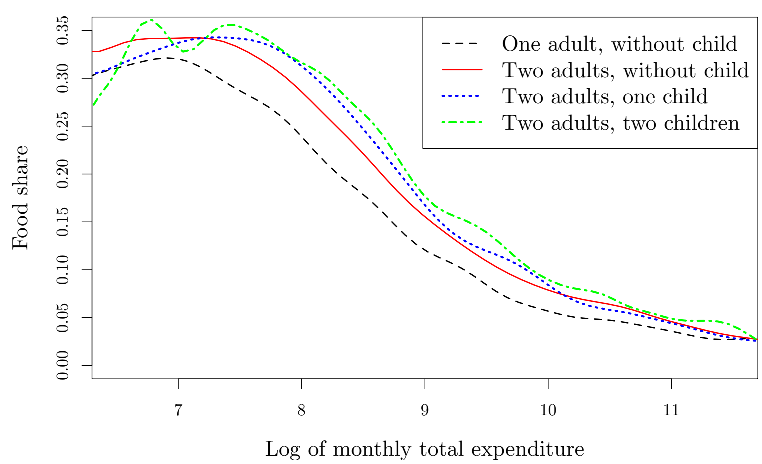

Figure 1: Fitted nonparametric regressions of household food shares against total household log expenditure: selected households sizes

4.3 Share Estimates

We begin by indirectly estimating equivalence scales following our Working-Leser and semi-paremetric methods. Estimates from those models are then plugged into the appropriate scale calculation; we present scales in the following subsection.

4.3.1 Working-Leser model

For the analysis, we estimated linear regressions, as well as semi-parametric regressions (discussed below), the parameters from which are reported in the appendix, since our main interest is the indirecly estimated scales. Our first set of models contains only log expenditure, household size and household composition controls, denoted as the "Basic Model" in Table A.1, while the other included additional controls for location, ethnic group and gender of the household head. The other models are labeled as "Additional Controls" in the same table. Under the basic model, as well as the additional controls model, we separately estimated an exogenous model and an endogenous model, where we allow for log expenditure to be endogenous, and we apply the heteroskedasticity instrumental variables procedure from Lewbel (2012); we also used income as a separately specified instrument in this procedure.

Regardless of controls or endogeneity, we find that an increase in total expenditure is associated with a reduction in the household's food budget share. We also find that additional children and additional adults are associated with an increase in the average food share. Although we do not statistically verify the following generalisations, the results suggest that each additional child and each additional adult increase the average food budget share by less than the previous additional child or adult. Recall that a binary variable like "Three Adults" denotes a household that has three adults, while "Two Adults" indicates that there are only two adults. Importantly, these indicators are defined to be mutually exclusive, rather than cumulative.

In terms of gender and ethnicity, we see that the budget share in coloured households is higher, on average, while the budget share is lower, on average, for white and Asian households, when compared to African-headed households. The average share is also lower for male-headed households. Compared to urban formal households, the food budget share is higher for all other locations: urban informal, traditional and rural formal, while provincial differences are all negative, compared to the Western Cape province (the reference province), which is one of the wealthier provinces. Thus, the additional controls capture some resource differences that affect the underlying relationship between the food budget share and expenditure.

Once endogeneity is addressed, we find that household size and composition effects are smaller, in the basic model, but larger in the additional controls model (although we do not test whether they are statistically significantly smaller or larger), as are the expenditure effects (in absolute value). In the additional controls model, there are also urban/rural changes – they all get smaller – while the provincial fixed effects are fairly similar to their exogenous counterparts. The real difference, in our view, is the change that we observe in gender and ethnic fixed effect estimates. The male-headed household difference in the budget share gets smaller, while the Asian difference becomes statistically insignificant, and the white household difference changes sign, while remaining statistically significant. In conclusion, the endogeneity results suggest that total household expenditure is correlated with both unobserved factors and most of the variables that we include in our models; therefore, it is appropriate to correct for endogeneity.

4.3.2 Semi-parametric model

The semi-parametric model is described by equation (5) and, as was the case with the Working-Leser model, the parameter estimates are presented as a "Basic Model" and "Additional Controls" model; each of those allow for exogeneity and endogeneity (with respect to household expenditure), where endogeneity corrections follow Dong (2010). The parameter estimates are located in Table A.2.

Recall the two main difference between equations (5) and (2), which is that (i) the ln expenditure estimate is set to unity for identification, such that (ii) Δ defines the semi-parametric model adjustment to ln expenditure for different households. The negative sign in the equation reminds us that the adjustment is in the denominator, i.e., it reduces ln expenditure. For that reason, the estimated effects in the semi-parametric model are of the opposite sign to the linear model and they are direct adjustments to ln expenditure, such that the model parameters should be interpreted differently.

Although not formally tested, we see that, as there are more adults and children, the estimated parameter is larger (in absolute value terms), while it appears that the incremental adjustment is smaller, as the size of the household increases. Given the negative sign in equation (5), these results imply that as households get larger, they need more total expenditure (which lowers the household share) to reach the same level of welfare as a smaller household. In the semi-parametric model, we also find that endogeneity matters, especially for the ethnicity and location parameters; however, endogeneity does not appear to have much of an impact on the household size and composition parameters, until we reach large numbers of children or adults. Even though the estimated signs in the Working-Leser and semi-parametric models are different, the underlying implications on equivalence scales are qualitatively similar, to which we now turn.

4.4 Scales

In the following sub-subsections, we describe the estimated scales, which are presented in tables in the appendix. Since those tables are quite long, we also illustrate the scales in Figures D.1 to D.8.

4.4.1 Working-Leser model

The calculated scales, with bootstrapped standard errors, resulting from the Working-Leser share models are presented in Table B.1. This table presents scales from all four estimated Working-Leser models (with and without additional controls, as well as correcting and not for endogeneity) and household sizes up to eight adults and six children. The results suggest that there are extensive economies of scale. For example, the estimated equivalence scale for a two-adult, five-child household ranges from 2.8 to 3.8, half or less the household size. The results also suggest that child costs are not constant; the parameter estimates in Table A.1 differ for each binary indicator for the number of children in the household. In particular, the third child and the fifth child are associated with a jump in the equivalence scale that is in excess of the increase seen for the second and fourth children, while the sixth child is associated with a negligible change in the scale. On the other hand, an additional adult does not appear to differentially influence the scale, regardless of how many adults are already in the household.

4.4.2 Semi-parametric model

The Working-Leser scales are based on an approach that is fairly common in the literature, although it does not assume base-independence (Blackorby and Donaldson 1993; Pendakur 1999), and, therefore, it is not clear what is being estimated (Blundell and Lewbel 1991). For that reason, we estimate semi-parametric regression models that, at least in our context, incoporate base-independence. This allows us to gain insight into the scale difference arising from the independence of base assumption.

The calculated scales, with Delta-method standard errors, derived from the semi-parametric model are presented in Table B.2. The table includes scales derived from four separate semi-parametric models (ones with and without additioinal controls, as well as ones that do and do not correct for endogeneity). Furthermore, scales up to an eight adults and six children are presented.

Broadly, as with the Working-Leser model, there are extensive economies of scale, although they are estimated to be larger via the semi-parametric approach. For example, the calculated scale for the two-adult and five-child household ranges from 2.4 to 2.9, rather than 2.8 to 3.8, as was the case in the Working-Leser founded shares. We also find that the fifth child has an economically meaningful impact on the equivalence scale, as before, although the third child impact is more in line with the other children. Adults continue to have a fairly consistent impact on the scales, as the number of adults increases.

4.4.3 Typical child cost and household economies of scale: $(A + \kappa K)^{\psi}$

In Tables B.1 and B.2, we present a series of equivalence scales underpinned by differing sets of assumptions. Our results suggest that endogeneity is present (results available upon request), and that location, gender and ethnic differences exist in food consumption decisions. The literature also suggests that base-independence is more appropriate. Thus, the scales in the last column of Table B.2 are plausibly better. Despite that, the table is not easily summarised, and, from a policy perspective, a simpler approach would likely be appreciated and better understood. For that reason, we have used our calculated scales to estimate a common child cost and household economies of scale model as in: $(A + \kappa K)^{\psi}$. These results, for all of the models, are reported in Table 2. Estimates were conducted with non-linear least squares, and we report 95% confidence intervals for the estimated parameters.

The results point to child costs that are quite varied, which agrees with the underlying empirical results finding that the third and fifth children affected the budget shares by more than for other children. The results also suggest that the child costs could be relatively larger than generally assumed in the literature. Although Nicholson (1976) suggests that child costs tend to be over-estimated, due to the fact that children mostly consume food and other basic necessities, other reasons may also be plausible. For example, in South Africa, those looking after young children are able to access a child support grant, and, even though it is not large, Aguëro, Carter, and Woolard (2006) find that it improves nutrition, especially amongst children. Access to the grant is conditional on having children, and the grant is received for each child supported. Although beyond the scope of our current analysis, it is possible that our scale estimates are capturing the relaxation of a household budget constraint, at least amongst some households with children.

The results also point to extensive household economies of scale, which are of a magnitude different than previously estimated in the literature, although Koch (2018), using 2010-11 data finds similar economies of scale. The estimates range from 0.31 to 0.58 across all models, which is well below the 0.9 assumed by May, Carter, and Posel (1995), the 0.7, which was the lower end assumed by Posel, Casale, and Grapsa (2016), the 0.77-0.85 estimated by Posel, Casale, and Grapsa (2020), or unity as implied by the per capita scale that dominates much of the inequality literature (Leibbrandt et al. 2010).

Thus, to summarise, a child cost near unity cannot be rejected in the analysis, while a household scale parameter near unity can not be accepted. Furthermore, the confidence interval for the scale parameter is nearer 0.5 than anything much larger. When taken together, the relatively large child costs and the relatively small scale economies parameter suggest that the square root scale, as applied by Aaberge and Melby (1998) and relatively recent OECD-based studies (Economic Co-operation and Development 2011), is a reasonably approximation in the South African context.

Table 2: Estimate of child cost and household economies of scale

	Basic Model		Additional Controls	
	Exogenous	Endogenous	Exogenous	Endogenous
Child Cost	1.9310	1.2459	2.0148	1.2314
	(1.04 – 2.82)	(0.83 – 1.66)	(1.03 – 2.99)	(0.82 – 1.64)
Scale Economy	0.5257	0.4988	0.4815	0.4674
	(0.47 – 0.58)	(0.46 – 0.54)	(0.42 – 0.54)	(0.43 – 0.50)
Child Cost	1.2950	1.2459	1.0621	1.6995
	(0.85 – 1.74)	(0.83 – 1.66)	(0.68 – 1.45)	(0.91 – 2.49)
Scale Economy	0.4745	0.4988	0.4283	0.3511
	(0.44 – 0.51)	(0.46 – 0.54)	(0.39 – 0.46)	(0.31 – 0.39)

Child cost β and economies θ estimated nonlinearly, along with 95% confidence intervals in (), assuming $\hat{\Lambda}^\ell = (A + \beta K)^\theta$ as in equation (10); estimation data based on calculated scales from all of the Working-Leser and semi-parametric models. In the top panel, scales are based on the Working-Leser budget shares. In the bottom panel, the scales are underpinned by the semi-parametric model.

5 Conclusion

In this paper, we presented estimates of equivalence scales from a developing country with extensive inequality – South Africa. Our estimates are underpinned by the 2014/15 Living Conditions Survey (Stats SA 2017), which is the most recent survey capturing the necessary information. Our estimates are also underscored by base-independence, which we incorporate through a linear index semi-parametric model of each household's food budget share. However, we offer comparator estimates that arise from a simpler to apply modification of the Working-Leser form outlined by (Deaton 1987). We also incorporate a broad range of location and household-specific information, and we account for endogeneity via both heteroscedasticity-based exclusion restrictions for the Working-Leser form (Lewbel 2012) and non-parametrically estimated residuals arising from the first-stage inclusion of a control variable (that may not satisfy the usual instrumental variable exclusion restriction) with continuous support (Dong 2010).

Our results point to relatively large child costs, which might arise either from the possibility that food-based equivalence scales could be overstated (Nicholson 1976) or from the child-support grant that relaxes poorer household budget constraints and has been found to improve child nutrition (Aguëro, Carter, and Woolard 2006). Our results also point to economies of scale that are more extensive than those previously used or estimated. Instead, our results suggest that the square-root scale is more appropriate than any scale used previously for South African inequality and poverty research.

Although our approach allows for the calculation of scales that vary across all regions of the country and account for many other household preference differences not previously addressed in the literature, we limit our presentation of scales in order to save space. In future research, it will be useful to present a complete set of scales across a variety of regions, and compare the effect of those and other scales on inequality across and within regions and the country, as a whole. It will also be beneficial to extend the preceding analysis to account for price differences, make use of additional food sufficiency information that is available in surveys like ours, and investigate, in more depth, the relatively large child costs that we uncover.

References

Aaberge, Rolf, and Ingrid Melby. 1998. "THE SENSITIVITY OF INCOME INEQUALITY TO CHOICE OF EQUIVALENCE SCALES," no. 4, 44: 565–69. https://doi.org/https://doi.org/10.1111/j.1475-4991.1998.tb00299.x.

Abanokova, Kseniya, Hai-Anh H. Dang, and Michael M. Lokshin. 2020. *The Important Role of Equivalence Scales : Household Size, Composition, and Poverty Dynamics in the Russian Federation*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/33872.

Aguëro, Jorge M., Michael R. Carter, and Ingrid Woolard. 2006. *The Impact of Unconditional Cash Transfers on Nutrition: The South African Child Support Grant*. Cape Town: University of Cape Town. http://hdl.handle.net/11090/46.

Ataguba, J O E. 2012. "Reassessing Catastrophic Health-Care Payments with a Nigerian Case Study." *Health Economics, Policy and Law* 7: 309–26.

Banks, James, Richard Blundell, and Arthur Lewbel. 1997. "Quadratic Engel Curves and Consumer Demand." *The Review of Economics and Statistics* 79 (4): 527–39. https://doi.org/10.1162/003465397557015.

Blackorby, Charles, and David Donaldson. 1993. "Adult-Equivalent Scales and the Economic Implementation of Interpersonal Comparisons of Well-Being." *Social Choice and Welfare* 10: 335–61.

Blundell, R, and A Lewbel. 1991. "The Information Content of Equivalence Scales." *Journal of Econometrics* 50: 49–68.

Daley, Angela, Thesia Garner, Shelley Phipps, and Eva Sierminska. 2020. "Differences Across Countries and Time in Household Expenditure Patterns: Implications for the Estimation of Equivalence Scales." *International Review of Applied Economics* 34 (6): 734–57. https://doi.org/10.1080/02692171.2020.1781798.

Deaton, Angus. 1987. *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. Washington, DC: The World Bank.

Deaton, Angus, and Margaret Grosh. 2000. "Consumption." In *Designing Household Survey Questionnaires for Developing Countries*, edited by Margaret Grosh and Paul Glewwe, 91–134. Washington, DC: World Bank.

Donaldson, David, and Krishna Pendakur. 2006. "The Identification of Fixed Costs from Consumer Behavior." *Journal of Business & Economic Statistics* 24 (3): 255–65. https://doi.org/10.1198/073500106000000035.

Dong, Yingyang. 2010. "Endogenous Regressor Binary Choice Models Without Instruments, with an Application to Migration." *Economics Letters* 107 (1): 33–35. https://doi.org/10.1016/j.econlet.2009.12.017.

Economic Co-operation, Organisation for, and Development. 2011. *Divided We Stand: Why Inequality Keeps Rising*. OECD Publishing Paris.

Edmonds, Eric V., Kristin Mammen, and Douglas L. Miller. 2005. "Rearranging the Family? Income Support and Elderly Living Arrangements in a Low-Income Country." *The Journal of Human Resources* 40 (1): 186–207. http://www.jstor.org/stable/4129570.

Engel, E. 1857. "Die Productions- Und Consumtionsverhaltnisse Des Konigreichs Sachsen." In *Dielebenkostenbelgischerarbeiter-Familien*, edited by E Engel. Dresden: C Heinrich.

Gui, Raluca, Markus Meierer, Rene Algesheimer, and Patrik Schilter. 2021. *REndo: Fitting Linear Models with Endogenous Regressors Using Latent Instrumental Variables*. https://CRAN.R-project.org/package=REndo.

Hasan, Syed Abul. 2016. "Engel Curves and Equivalence Scales for Bangladesh." *Journal of the Asia Pacific Economy* 21 (2): 301–15. https://doi.org/10.1080/13547860.2015.1137471.

Hayfield, Tristen, and Jeffrey S. Racine. 2008. "Nonparametric Econometrics: The Np Package." *Journal of Statistical Software* 27 (5). http://www.jstatsoft.org/v27/i05/.

Hlavac, Marek. 2018. *Stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI). https://CRAN.R-project.org/package=stargazer.

Ichimura, Hidehiko. 1993. "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models." *Journal of Econometrics* 58 (1-2): 71–120. https://doi.org/https://doi.org/10.1016/0304-4076(93)90114-K.

Imbens, Guido W, and Whitney K Newey. 2009. "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity." *Econometrica* 77 (5): 1481–1512. https://doi.org/10.3982/ECTA7108.

Klasen, Stephan, and Ingrid Woolard. 2008. "Surviving Unemployment Without State Support: Unemployment and Household Formation in South Africa." *Journal of African Economies* 18 (1): 1–51. https://doi.org/10.1093/jae/ejn007.

Kleiber, Christian, and Achim Zeileis. 2008. *Applied Econometrics with R*. New York: Springer-Verlag. https://CRAN.R-project.org/package=AER.

Klein, Roger W., and Richard H. Spady. 1993. "An Efficient Semiparametric Estimator for Binary Response Models." *Econometrica* 61 (2): 387–421. http://www.jstor.org/stable/2951556.

Koch, Steven F. 2018. "Catastrophic Health Payments: Does the Equivalence Scale Matter?" *Health Policy and Planning* 33 (8): 966–73.

Koch, Steven F, and Naomi Setshegetso. 2020. "Catastrophic Health Expenditures Arising from Out-of-Pocket Payments: Evidence from South African Income and Expenditure Surveys." *PLOS ONE* 15 (8): 1–14. https://doi.org/10.1371/journal.pone.0237217.

———. 2021. "Progressivity of Out-of-Pocket Payments and Its Determinants Decomposed over Time." *Development Southern Africa* 38 (5): 731–49. https://doi.org/10.1080/0376835X.2021.1880373.

Koch, Steven F, and Gauthier Tshiswaka-Kashalala. 2018. "The Demand for Reproductive Health Care." *Journal of African Economies* 27 (4): 405–29. https://doi.org/10.1093/jae/ejx042.

Leibbrandt, Murray, Ingrid Woolard, Arden Finn, and Jonathan Argent. 2010. "Trends in South African Income Distribution and Poverty Since the Fall of Apartheid," no. 101. https://doi.org/https://doi.org/https://doi.org/10.1787/5kmms0t7p1ms-en.

Lewbel, Arthur. 2012. "Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models." *Journal of Business & Economic Statistics* 30 (1): 67–80. https://doi.org/10.1080/07350015.2012.643126.

Li, Qi, and Jeffrey S Racine. 2007. *Nonparametric Econometrics: Theory and Practice*. Princeton, NJ, US: Princeton University Press.

Li, Qi, and Jeffrey S. Racine. 2004. "Cross-Validated Local Linear Nonparametric Regression." *Statistica Sinica* 14 (2): 485–512.

May, Julian, D Budlender, R Mokate, C Rogerson, and A Stavrou. 1995. "Poverty and Inequality in South Africa." Pretoria: Report prepared for the Office of the Deputy President, Republic of South Africa.

May, Julian, Michael Carter, and Dorrit Posel. 1995. "The Composition and Persistence of Poverty in Rural South Africa: An Entitlements Approach." Land and Agriculture Policy Centre Policy Paper No. 15.

Munoz Boudet, Ana Maria, Antra Bhatt, Ginette Azcona, Jayne Yoo, and Kathleen Beegle. 2021. *A Global View of Poverty, Gender, and Household Composition*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/35183.

Munoz Boudet, Ana Maria, Paola Buitrago, Benedicte Leroy De La Briere, David Newhouse, Eliana Rubiano Matulevich, Kinnon Scott, and Pablo Suarez-Becerra. 2018. *Gender Differences in Poverty and Household Composition Through the Life-Cycle : A Global Perspective*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/35183.

Nicholson, J L. 1976. "Appraisal of Different Methods of Estimating Equivalence Scales and Their Results." *Review of Income and Wealth* 22: 1–11.

Pendakur, Krishna. 1999. "Semiparametric Estimates and Tests of Base-Independent Equivalence Scales." *Journal of Econometrics* 80 (1): 1–40.

Pollak, R A, and T J Wales. 1979. "Welfare Comparisons and Equivalence Scales." *American Economic Review* 69: 216–21.

Posel, Dorrit, Daniela Casale, and Erofili Grapsa. 2016. "Re-Estimating Gender Differences in Income in South Africa: The Implications of Equivalence Scales." *Development Southern Africa* 33 (4): 425–41.

———. 2020. "Household Variation and Inequality : The Implications of Equivalence Scales in South Africa." *African Review of Economics and Finance* 12 (1): 102–22. https://doi.org/10.10520/EJC-1d056a3683.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Racine, Jeffrey S. 2019. "Energy, Economics, Replication & Reproduction." *Energy Economics* 82: 264–75. https://doi.org/https://doi.org/10.1016/j.eneco.2017.06.027.

Stats SA. 2017. *Living Conditions Survey 2014-2015 [dataset]. Version 1.1.* Statistics South Africa (Stats SA).

Summers, Robert. 1959. "A Note on Least Squares Bias in Household Expenditure Analysis." *Econometrica* 27: 121–26.

Wang, Min-Chiang, and John van Ryzin. 1981. "A Class of Smooth Estimators for Discrete Distributions." *Biometrika* 68 (1): 301–9. https://doi.org/10.2307/2335831.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, and Evan Miller. 2021. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files.* https://CRAN.R-project.org/package=haven.

Woolard, Ingrid. 2002. "Income Inequality and Poverty: Methods of Estimation and Some Policy Applications for South Africa." PhD Thesis, University of Cape Town.

Woolard, Ingrid, and Murray Leibbrandt. 1999. "Measuring Poverty in South Africa." DPRU Working Papers No. 99/33.

———. 2001. "Measuring Poverty in South Africa." In *Fighting Poverty: Labour Markets and Inequality in South Africa*, edited by Haroon Bhorat, Murray Leibbrandt, M Maziya, Servaas van der Berg, and Ingrid Woolard, 41–73. Cape Town: University of Cape Town Press.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook.* Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown-cookbook.

Xu, K. 2005. "Distribution of Health Payments and Catastrophic Expenditures: Methodology." Geneva, Switzerland: World Health Organization Discussion Paper 2-2005.

Yatchew, Adonis, Yiguo Sun, and Catherine Deri. 2003. "Efficient Estimation of Semiparametric Equivalence Scales with Evidence from South Africa." *Journal of Business & Economic Statistics* 21 (2): 247–57.

Ye, Yuxiang, and Steven F Koch. 2021. "Measuring Energy Poverty in South Africa Based on Household Required Energy Consumption." *Energy Economics* 103: 105553. https://doi.org/https://doi.org/10.1016/j.eneco.2021.105553.

Ye, Yuxiang, Steven F Koch, and Jiangfeng Zhang. forthcoming. "Modelling Required Energy Consumption with Equivalence Scales." *The Energy Journal*, forthcoming.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.

A Parameter estimates

Table A.1: Estimates from Linear Food Budget Share Equation

	Basic Model		Additional Controls	
Variables	Exogenous	Endogenous	Exogenous	Endogenous
Intercept	1.1683[a]	1.1132[a]	0.8580[a]	1.1588[a]
	(0.015)	(0.011)	(0.011)	(0.015)
Log Expenditure	-0.1153[a]	-0.1105[a]	-0.0764[a]	-0.1140[a]
	(0.002)	(0.001)	(0.001)	(0.002)
One child	0.0198[a]	0.0145[a]	0.0106[a]	0.0163[a]
	(0.003)	(0.003)	(0.003)	(0.003)
Two Children	0.0397[a]	0.0323[a]	0.0248[a]	0.0329[a]
	(0.003)	(0.003)	(0.003)	(0.003)
Three Children	0.0500[a]	0.0461[a]	0.0393[a]	0.0449[a]
	(0.004)	(0.004)	(0.004)	(0.004)
Four Children	0.0530[a]	0.0565[a]	0.0490[a]	0.0533[a]
	(0.005)	(0.005)	(0.005)	(0.005)
Five Children	0.0569[a]	0.0778[a]	0.0695[a]	0.0755[a]
	(0.008)	(0.008)	(0.008)	(0.008)
Six Children	0.0540[a]	0.0788[a]	0.0677[a]	0.0741[a]
	(0.012)	(0.011)	(0.011)	(0.011)
Two Adults	0.0386[a]	0.0447[a]	0.0310[a]	0.0436[a]
	(0.003)	(0.003)	(0.003)	(0.003)
Three Adults	0.0482[a]	0.0556[a]	0.0358[a]	0.0550[a]
	(0.003)	(0.003)	(0.003)	(0.003)
Four Adults	0.0600[a]	0.0619[a]	0.0391[a]	0.0610[a]
	(0.004)	(0.004)	(0.004)	(0.004)
Five Adults	0.0744[a]	0.0746[a]	0.0478[a]	0.0733[a]
	(0.005)	(0.005)	(0.005)	(0.005)
Six Adults	0.0681[a]	0.0733[a]	0.0464[a]	0.0724[a]
	(0.007)	(0.007)	(0.006)	(0.007)
Seven Adults	0.0884[a]	0.0900[a]	0.0598[a]	0.0875[a]

	(0.010)	(0.010)	(0.010)	(0.010)
Eight Adults	0.0666^{a}	0.0808^{a}	0.0465^{a}	0.0753^{a}
	(0.014)	(0.014)	(0.014)	(0.014)
Male-Headed			-0.0103^{a}	-0.0041^{d}
			(0.002)	(0.002)
Coloured			0.0122^{a}	0.0191^{a}
			(0.004)	(0.004)
Asian			-0.0391^{a}	-0.0082
			(0.007)	(0.008)
White			-0.0146^{a}	0.0439^{a}
			(0.005)	(0.005)
Eastern Cape			-0.0209^{a}	-0.0273^{a}
			(0.005)	(0.005)
Northern Cape			-0.0313^{a}	-0.0409^{a}
			(0.005)	(0.005)
Free State			-0.0451^{a}	-0.0513^{a}
			(0.005)	(0.005)
KwaZulu-Natal			-0.0147^{a}	-0.0203^{a}
			(0.004)	(0.005)
Northwest			-0.0441^{a}	-0.0488^{a}
			(0.005)	(0.005)
Gauteng			-0.0319^{a}	-0.0289^{a}
			(0.004)	(0.004)
Mpumalanga			-0.0217^{a}	-0.0230^{a}
			(0.005)	(0.005)
Limpopo			-0.0465^{a}	-0.0491^{a}
			(0.005)	(0.005)
Urban Informal			0.0237^{a}	0.0011
			(0.004)	(0.004)
Traditional Area			0.0507^{a}	0.0296^{a}
			(0.003)	(0.003)
Rural Formal			0.0597^{a}	0.0413^{a}

			(0.005)	(0.005)

Estimates based on both linear models and heteroskedasticity IV models (with an additional instrument: log income). The following notation and significance levels are listed: ^{a} - 0.005, ^{b} - 0.01, ^{c} - 0.05, ^{d} - 0.1.

Table A.2: Estimates from semi-parametric food budget share equation

	Basic Model		Additional Controls	
Variables	Exogenous	Endogenous	Exogenous	Endogenous
Log Expenditure	1.0000^a	1.0000^a	1.0000^a	1.0000^a
	(0.000)	(0.000)	(0.000)	(0.000)
One child	-0.2026^a	-0.1735^a	-0.1747^a	-0.1822^a
	(0.026)	(0.002)	(0.001)	(0.001)
Two Children	-0.3351^a	-0.2917^a	-0.3082^a	-0.2705^a
	(0.028)	(0.002)	(0.001)	(0.001)
Three Children	-0.4501^a	-0.3769^a	-0.3374^a	-0.3452^a
	(0.031)	(0.002)	(0.002)	(0.001)
Four Children	-0.4864^a	-0.3759^a	-0.4611^a	-0.4275^a
	(0.052)	(0.003)	(0.003)	(0.002)
Five Children	-0.7506^a	-0.5879^a	-0.6117^a	-0.5641^a
	(0.083)	(0.005)	(0.005)	(0.003)
Six Children	-0.7637^a	-0.5121^a	-0.6611^a	-0.5442^a
	(0.098)	(0.006)	(0.007)	(0.004)
Two Adults	-0.3149^a	-0.2866^a	-0.2579^a	-0.2966^a
	(0.026)	(0.001)	(0.001)	(0.001)
Three Adults	-0.4093^a	-0.3911^a	-0.3682^a	-0.3701^a
	(0.030)	(0.002)	(0.001)	(0.001)
Four Adults	-0.4931^a	-0.4560^a	-0.4460^a	-0.4908^a
	(0.035)	(0.002)	(0.002)	(0.001)
Five Adults	-0.5645^a	-0.5105^a	-0.4186^a	-0.4891^a
	(0.046)	(0.002)	(0.003)	(0.001)

Six Adults	-0.5773^a	-0.6221^a	-0.4757^a	-0.5794^a
	(0.060)	(0.004)	(0.004)	(0.002)
Seven Adults	-0.6915^a	-0.6027^a	-0.5818^a	-0.5700^a
	(0.102)	(0.005)	(0.006)	(0.003)
Eight Adults	-0.7068^a	-0.4428^a	-0.7397^a	-0.3339^a
	(0.156)	(0.009)	(0.007)	(0.005)
Male-Headed			0.0842^a	0.0414^a
			(0.001)	(0.001)
Coloured			-0.1639^a	-0.1923^a
			(0.002)	(0.001)
Asian			0.4086^a	0.1989^a
			(0.004)	(0.003)
White			0.1621^a	-0.0555^a
			(0.003)	(0.002)
Eastern Cape			0.3521^a	0.1160^a
			(0.002)	(0.001)
Northern Cape			0.2868^a	0.2261^a
			(0.003)	(0.001)
Free State			0.5731^a	0.3513^a
			(0.002)	(0.001)
KwaZulu-Natal			0.3083^a	0.1006^a
			(0.002)	(0.001)
Northwest			0.4218^a	0.2926^a
			(0.002)	(0.001)
Gauteng			0.4324^a	0.1426^a
			(0.002)	(0.001)
Mpumalanga			0.3705^a	0.1407^a
			(0.002)	(0.001)
Limpopo			0.6848^a	0.3155^a
			(0.002)	(0.001)
Urban Informal			-0.2384^a	-0.0310^a
			(0.002)	(0.001)

Traditional Area		-0.5382^a	-0.2613^a
		(0.001)	(0.001)
Rural Formal		-0.7451^a	-0.4963^a
		(0.003)	(0.002)

Estimates based on both semi-parametric models and semiparamtric control function models (with an additional instrument: log income). The following notation and significance levels are listed: ^{a} - 0.005, ^{b} - 0.01, ^{c} - 0.05, ^{d} - 0.1. Also note that I am not reporting the control function estimate here... Probably should...

B Equivalence scales

Table B.1: Estimate of equivalence scales for household type, underpinned by linear regression model

		Basic Model		Added Controls	
Adults	Kids	Exogenous	Endogenous	Exogenous	Endogenous
1	0	1.0000	1.0000	1.0000	1.0000
1	0	(0.000)	(0.000)	(0.000)	(0.000)
1	1	1.1967	1.1400	1.1493	1.1535
1	1	(0.038)	(0.027)	(0.037)	(0.027)
1	2	1.4601	1.3396	1.3843	1.3348
1	2	(0.050)	(0.033)	(0.047)	(0.034)
1	3	1.8009	1.5172	1.6734	1.4823
1	3	(0.090)	(0.054)	(0.084)	(0.052)
1	4	2.0956	1.6677	1.9000	1.5957
1	4	(0.136)	(0.093)	(0.142)	(0.087)
1	5	2.6765	2.0224	2.4841	1.9396
1	5	(0.281)	(0.157)	(0.286)	(0.148)
1	6	2.7460	2.0397	2.4269	1.9144
1	6	(0.394)	(0.265)	(0.411)	(0.232)
2	0	1.4325	1.4988	1.5001	1.4662

2	0	(0.046)	(0.038)	(0.046)	(0.037)
2	1	1.7143	1.7087	1.7241	1.6913
2	1	(0.068)	(0.054)	(0.067)	(0.052)
2	2	2.0915	2.0078	2.0766	1.9571
2	2	(0.091)	(0.064)	(0.089)	(0.062)
2	3	2.5798	2.2739	2.5103	2.1733
2	3	(0.137)	(0.093)	(0.131)	(0.085)
2	4	3.0019	2.4996	2.8502	2.3396
2	4	(0.202)	(0.143)	(0.223)	(0.130)
2	5	3.8340	3.0311	3.7265	2.8438
2	5	(0.405)	(0.243)	(0.424)	(0.223)
2	6	3.9336	3.0572	3.6407	2.8069
2	6	(0.577)	(0.400)	(0.596)	(0.348)
3	0	1.5724	1.6542	1.5972	1.6193
3	0	(0.060)	(0.049)	(0.058)	(0.046)
3	1	1.8818	1.8859	1.8357	1.8680
3	1	(0.080)	(0.062)	(0.081)	(0.058)
3	2	2.2959	2.2161	2.2110	2.1615
3	2	(0.100)	(0.073)	(0.101)	(0.072)
3	3	2.8318	2.5097	2.6727	2.4003
3	3	(0.146)	(0.107)	(0.148)	(0.093)
3	4	3.2952	2.7588	3.0347	2.5839
3	4	(0.228)	(0.157)	(0.238)	(0.146)
3	5	4.2086	3.3454	3.9677	3.1409
3	5	(0.442)	(0.265)	(0.462)	(0.245)
3	6	4.3179	3.3742	3.8763	3.1001
3	6	(0.641)	(0.445)	(0.652)	(0.387)
4	0	1.6555	1.7502	1.6681	1.7075
4	0	(0.072)	(0.061)	(0.066)	(0.057)
4	1	1.9811	1.9953	1.9172	1.9697
4	1	(0.091)	(0.074)	(0.085)	(0.071)
4	2	2.4171	2.3446	2.3091	2.2793

4	2	(0.120)	(0.089)	(0.112)	(0.083)
4	3	2.9814	2.6553	2.7913	2.5311
4	3	(0.170)	(0.121)	(0.165)	(0.110)
4	4	3.4692	2.9188	3.1693	2.7247
4	4	(0.241)	(0.175)	(0.261)	(0.153)
4	5	4.4308	3.5395	4.1437	3.3120
4	5	(0.466)	(0.287)	(0.492)	(0.261)
4	6	4.5460	3.5699	4.0483	3.2690
4	6	(0.681)	(0.467)	(0.698)	(0.415)
5	0	1.8838	1.9631	1.8703	1.9014
5	0	(0.100)	(0.089)	(0.103)	(0.081)
5	1	2.2544	2.2380	2.1496	2.1934
5	1	(0.124)	(0.104)	(0.127)	(0.098)
5	2	2.7505	2.6298	2.5890	2.5381
5	2	(0.153)	(0.127)	(0.155)	(0.116)
5	3	3.3926	2.9783	3.1296	2.8185
5	3	(0.207)	(0.153)	(0.229)	(0.144)
5	4	3.9478	3.2739	3.5535	3.0341
5	4	(0.297)	(0.215)	(0.312)	(0.197)
5	5	5.0421	3.9700	4.6460	3.6881
5	5	(0.531)	(0.340)	(0.568)	(0.307)
5	6	5.1731	4.0041	4.5390	3.6402
5	6	(0.766)	(0.543)	(0.837)	(0.465)
6	0	1.8546	1.9412	1.8345	1.8877
6	0	(0.137)	(0.117)	(0.145)	(0.111)
6	1	2.2195	2.2130	2.1084	2.1775
6	1	(0.168)	(0.137)	(0.177)	(0.129)
6	2	2.7079	2.6005	2.5395	2.5198
6	2	(0.206)	(0.162)	(0.224)	(0.145)
6	3	3.3400	2.9451	3.0698	2.7982
6	3	(0.267)	(0.200)	(0.292)	(0.178)
6	4	3.8866	3.2374	3.4855	3.0122

6	4	(0.348)	(0.233)	(0.366)	(0.218)
6	5	4.9639	3.9258	4.5571	3.6614
6	5	(0.581)	(0.366)	(0.637)	(0.320)
6	6	5.0929	3.9595	4.4522	3.6139
6	6	(0.776)	(0.554)	(0.857)	(0.466)
7	0	2.2444	2.2570	2.1888	2.1538
7	0	(0.280)	(0.213)	(0.273)	(0.194)
7	1	2.6859	2.5730	2.5157	2.4845
7	1	(0.334)	(0.248)	(0.333)	(0.221)
7	2	3.2769	3.0235	3.0300	2.8749
7	2	(0.419)	(0.291)	(0.401)	(0.260)
7	3	4.0419	3.4242	3.6627	3.1926
7	3	(0.518)	(0.319)	(0.510)	(0.293)
7	4	4.7033	3.7640	4.1587	3.4367
7	4	(0.637)	(0.395)	(0.627)	(0.339)
7	5	6.0070	4.5644	5.4373	4.1775
7	5	(0.944)	(0.523)	(0.970)	(0.460)
7	6	6.1630	4.6036	5.3122	4.1233
7	6	(1.183)	(0.739)	(1.173)	(0.596)
8	0	1.9448	2.0765	1.8377	1.9362
8	0	(0.352)	(0.286)	(0.369)	(0.244)
8	1	2.3274	2.3673	2.1121	2.2335
8	1	(0.421)	(0.330)	(0.442)	(0.282)
8	2	2.8395	2.7817	2.5439	2.5846
8	2	(0.513)	(0.383)	(0.546)	(0.328)
8	3	3.5024	3.1503	3.0751	2.8701
8	3	(0.648)	(0.447)	(0.681)	(0.375)
8	4	4.0755	3.4630	3.4916	3.0896
8	4	(0.768)	(0.479)	(0.791)	(0.401)
8	5	5.2052	4.1994	4.5651	3.7555
8	5	(1.030)	(0.633)	(1.081)	(0.521)
8	6	5.3405	4.2355	4.4600	3.7068

8	6	(1.271)	(0.772)	(1.296)	(0.642)

Estimated equivalence scale by household type, and bootstrapped standard error (399 replications). Estimates underpinned by exogenous and endogenous linear model including only basic controls. The following notation and significance levels are presented: ^{a} - 0.005, ^{b} - 0.01, ^{c} - 0.05, ^{d} - 0.1.

Table B.2: Estimate of equivalence scales for household type, underpinned by semi-parametric regression model

		Basic Model		Added Controls	
Adults	Kids	Exogenous	Endogenous	Exogenous	Endogenous
1	0	1.0000	1.0000	1.0000	1.0000
1	0	(0.000)	(0.000)	(0.000)	(0.000)
1	1	1.2246	1.1895	1.1909	1.1998
1	1	(0.021)	(0.001)	(0.001)	(0.001)
1	2	1.3981	1.3387	1.3610	1.3107
1	2	(0.020)	(0.001)	(0.001)	(0.001)
1	3	1.5684	1.4578	1.4013	1.4123
1	3	(0.020)	(0.001)	(0.001)	(0.001)
1	4	1.6264	1.4562	1.5859	1.5335
1	4	(0.032)	(0.002)	(0.002)	(0.001)
1	5	2.1182	1.8003	1.8435	1.7579
1	5	(0.039)	(0.003)	(0.002)	(0.002)
1	6	2.1463	1.6687	1.9368	1.7233
1	6	(0.046)	(0.003)	(0.004)	(0.003)
2	0	1.3701	1.3319	1.2943	1.3453
2	0	(0.019)	(0.001)	(0.001)	(0.001)
2	1	1.6778	1.5843	1.5413	1.6142
2	1	(0.040)	(0.002)	(0.002)	(0.001)
2	2	1.9155	1.7830	1.7614	1.7633
2	2	(0.038)	(0.002)	(0.002)	(0.001)
2	3	2.1489	1.9416	1.8136	1.9000
2	3	(0.038)	(0.002)	(0.002)	(0.001)
2	4	2.2283	1.9396	2.0525	2.0630
2	4	(0.051)	(0.003)	(0.003)	(0.002)
2	5	2.9022	2.3978	2.3860	2.3649
2	5	(0.058)	(0.004)	(0.003)	(0.002)
2	6	2.9407	2.2227	2.5068	2.3183

2	6	(0.065)	(0.004)	(0.005)	(0.003)
3	0	1.5057	1.4786	1.4452	1.4479
3	0	(0.020)	(0.001)	(0.001)	(0.001)
3	1	1.8439	1.7588	1.7210	1.7372
3	1	(0.041)	(0.002)	(0.002)	(0.001)
3	2	2.1051	1.9793	1.9668	1.8977
3	2	(0.039)	(0.002)	(0.002)	(0.001)
3	3	2.3617	2.1554	2.0251	2.0449
3	3	(0.039)	(0.003)	(0.002)	(0.001)
3	4	2.4489	2.1532	2.2919	2.2203
3	4	(0.052)	(0.003)	(0.003)	(0.002)
3	5	3.1895	2.6619	2.6642	2.5452
3	5	(0.059)	(0.004)	(0.003)	(0.002)
3	6	3.2318	2.4674	2.7991	2.4951
3	6	(0.065)	(0.004)	(0.005)	(0.003)
4	0	1.6374	1.5778	1.5621	1.6336
4	0	(0.021)	(0.001)	(0.001)	(0.001)
4	1	2.0052	1.8768	1.8602	1.9601
4	1	(0.043)	(0.003)	(0.002)	(0.001)
4	2	2.2893	2.1122	2.1259	2.1411
4	2	(0.041)	(0.003)	(0.002)	(0.001)
4	3	2.5682	2.3000	2.1890	2.3072
4	3	(0.041)	(0.003)	(0.002)	(0.001)
4	4	2.6631	2.2977	2.4773	2.5051
4	4	(0.053)	(0.003)	(0.003)	(0.002)
4	5	3.4684	2.8405	2.8797	2.8716
4	5	(0.061)	(0.004)	(0.004)	(0.002)
4	6	3.5144	2.6329	3.0255	2.8151
4	6	(0.067)	(0.005)	(0.005)	(0.003)
5	0	1.7585	1.6662	1.5199	1.6309
5	0	(0.026)	(0.001)	(0.002)	(0.001)
5	1	2.1534	1.9819	1.8100	1.9568

5	1	(0.047)	(0.003)	(0.003)	(0.002)
5	2	2.4585	2.2305	2.0685	2.1375
5	2	(0.046)	(0.003)	(0.003)	(0.002)
5	3	2.7581	2.4289	2.1298	2.3033
5	3	(0.046)	(0.003)	(0.003)	(0.002)
5	4	2.8600	2.4263	2.4104	2.5009
5	4	(0.058)	(0.003)	(0.003)	(0.002)
5	5	3.7248	2.9996	2.8020	2.8668
5	5	(0.065)	(0.004)	(0.004)	(0.002)
5	6	3.7742	2.7804	2.9438	2.8104
5	6	(0.072)	(0.005)	(0.005)	(0.003)
6	0	1.7813	1.8628	1.6091	1.7849
6	0	(0.034)	(0.002)	(0.003)	(0.001)
6	1	2.1813	2.2158	1.9162	2.1416
6	1	(0.055)	(0.004)	(0.004)	(0.002)
6	2	2.4903	2.4937	2.1899	2.3394
6	2	(0.053)	(0.003)	(0.004)	(0.002)
6	3	2.7938	2.7155	2.2548	2.5208
6	3	(0.053)	(0.004)	(0.004)	(0.002)
6	4	2.8970	2.7127	2.5518	2.7370
6	4	(0.066)	(0.004)	(0.004)	(0.002)
6	5	3.7731	3.3536	2.9664	3.1376
6	5	(0.073)	(0.005)	(0.005)	(0.003)
6	6	3.8231	3.1086	3.1165	3.0758
6	6	(0.080)	(0.006)	(0.006)	(0.004)
7	0	1.9967	1.8271	1.7892	1.7683
7	0	(0.051)	(0.003)	(0.003)	(0.002)
7	1	2.4451	2.1733	2.1307	2.1217
7	1	(0.072)	(0.004)	(0.005)	(0.002)
7	2	2.7916	2.4459	2.4351	2.3177
7	2	(0.071)	(0.004)	(0.005)	(0.002)
7	3	3.1317	2.6635	2.5072	2.4975

7	3	(0.070)	(0.004)	(0.005)	(0.002)
7	4	3.2475	2.6607	2.8375	2.7117
7	4	(0.083)	(0.005)	(0.005)	(0.003)
7	5	4.2294	3.2893	3.2985	3.1085
7	5	(0.090)	(0.006)	(0.006)	(0.003)
7	6	4.2855	3.0490	3.4654	3.0473
7	6	(0.097)	(0.006)	(0.007)	(0.004)
8	0	2.0276	1.5571	2.0953	1.3964
8	0	(0.077)	(0.006)	(0.003)	(0.003)
8	1	2.4829	1.8522	2.4952	1.6754
8	1	(0.098)	(0.007)	(0.005)	(0.004)
8	2	2.8347	2.0845	2.8516	1.8302
8	2	(0.097)	(0.007)	(0.005)	(0.004)
8	3	3.1801	2.2699	2.9362	1.9721
8	3	(0.097)	(0.007)	(0.005)	(0.004)
8	4	3.2977	2.2676	3.3229	2.1412
8	4	(0.109)	(0.008)	(0.005)	(0.004)
8	5	4.2948	2.8033	3.8628	2.4546
8	5	(0.116)	(0.009)	(0.006)	(0.005)
8	6	4.3518	2.5985	4.0583	2.4063
8	6	(0.123)	(0.009)	(0.007)	(0.006)

Estimated equivalence scale by household type, and bootstrapped standard error (399 replications). Estimates underpinned by exogenous and endogenous linear model including only basic controls.

C Mathematical development of base independence

In what follows, I describe base independence, relating that to the empirical model in the text. Let us begin with a notion of equal utility for different types of households, where the scale might be a function of total expenditure.

$$V\left(p,x,z\right)=V\left(p,\frac{x}{\Delta(p,x,z)},z^{r}\right) \tag{C.1}$$

From indirect utility, we can derive expenditure shares from Roy's Identity in semi-log form.

$$w_j(p,x,z)=-\frac{\partial V/\partial\ln p_j}{\partial V/\partial\ln x}=-\frac{\partial V/\partial p_j}{\partial V/\partial x}\times\frac{p_j}{x} \tag{C.2}$$

For the reference household, the share equation is based on simple substitution of terms, yielding the share, keeping in mind the assumption that $x^r\equiv x/\Delta$; thus, $\Delta(p,x^r,z^r)=1$.

$$w_j(p,x^r,z^r)=-\frac{\partial V/\partial\ln p_j}{\partial V/\partial\ln x^r} \tag{C.3}$$

For any non-reference household, we split the numerator and denominator into separate components. First, the numerator, although we include the term following $\times$ in equation (C.2); we drop the specific share notation j for convenience, such that subscripts represent the term over which the derivative is taken.

$$\begin{aligned}-\frac{\partial V}{\partial p_j}\times\frac{p_j}{x}&=\left[-V_p-V_x\frac{x}{\Delta^2}\left(-\frac{\partial\Delta}{\partial p}\right)\right]\frac{p}{x}\\&=-V_p\times\frac{p}{x}+\frac{V_x}{\Delta}\left(\frac{\partial\Delta}{\partial p}\times\frac{p}{\Delta}\right)\\&=-\frac{V_pp}{x}+\frac{V_x}{\Delta}\eta_{\Delta p}\end{aligned} \tag{C.4}$$

Next, we turn our attention to the denominator.

$$\begin{aligned}\frac{\partial V}{\partial x}&=\frac{V_x}{\Delta}+V_x\frac{x}{\Delta^2}\left(-\frac{\partial\Delta}{\partial x}\right)\\&=\frac{V_x}{\Delta}\left(1-\frac{\partial\Delta}{\partial x}\frac{x}{\Delta}\right)\\&=\frac{V_x}{\Delta}\left(1-\eta_{\Delta x}\right)\end{aligned} \tag{C.5}$$

Placing equations (C.4) and (C.5) into (C.2) results in our comparison with the reference share.

$$\begin{aligned}
w_j(p,x,z) &= -\left(\frac{V_p p}{x} + \frac{V_x}{\Delta}\eta_{\Delta p}\right)\left(\frac{\Delta}{V_x(1-\eta_{\Delta x})}\right) \\
&= \left(-\frac{V_p p}{x}\frac{\Delta}{V_x(1-\eta_{\Delta x})}\right) + \left(\frac{V_x \eta_{\Delta p}}{\Delta}\frac{\Delta}{V_x(1-\eta_{\Delta x})}\right) \\
&= \left[\frac{1}{1-\eta_{\Delta x}}\right]\left(-\frac{V_p}{V_x}\frac{p\Delta}{x}\right) + \frac{\eta_{\Delta p}}{1-\eta_{\Delta x}} \\
&= \frac{1}{1-\eta_{\Delta x}}\left[\left(-\frac{V_p}{V_x}\frac{p}{x/\Delta}\right) + \eta_{\Delta p}\right]
\end{aligned} \tag{C.6}$$

Writing more clearly,

$$w_j(p,x,z) = \frac{w_j(p,x^r,z^r) + \eta_{\Delta p}}{1-\eta_{\Delta x}} \tag{C.7}$$

It is fairly clear from equation (C.7) that if w_j, $\eta_{\Delta p}$ and $\eta_{\Delta x}$ are all allowed to vary, a solution will not be unique. Thus, we impose equivalence scale exactness, such that the scale cannot depend on utility, and, thus, cannot depend on expenditure.

$$w_j(p,x,z) = w_j(p,x^r,z^r) + \eta_{\Delta p} \tag{C.8}$$

Under base independence, then, $\eta_{\Delta p}$ is a vertical translation of the budget share. Unfortunately, are not able to incorporate prices in the analysis, thus, we are left to estimate

$$w_j(x,z) = w_j(x^r,z^r) \tag{C.9}$$

In effect, this is the premise in Yatchew, Sun, and Deri (2003), which we describe in footnote 7.

D Plots

D.1 Linear scales - exogenous

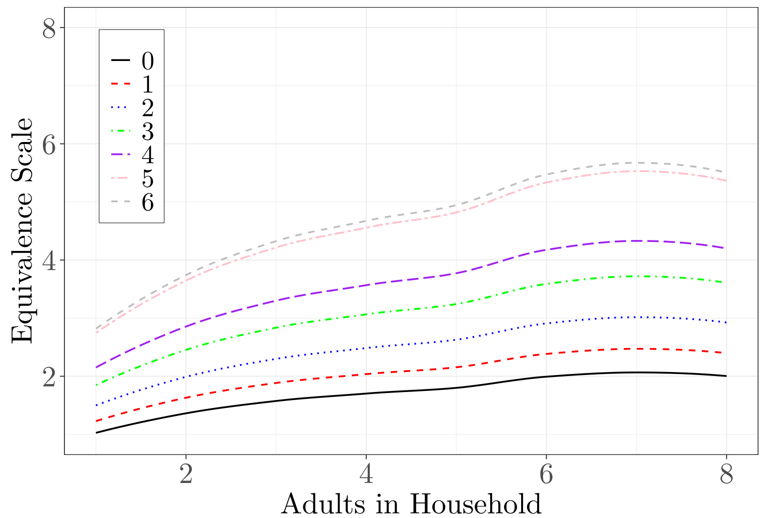

Figure D.1: Estimated equivalence scales assuming exogeneity and linear shares: Curves are separated for the number of children in the household.

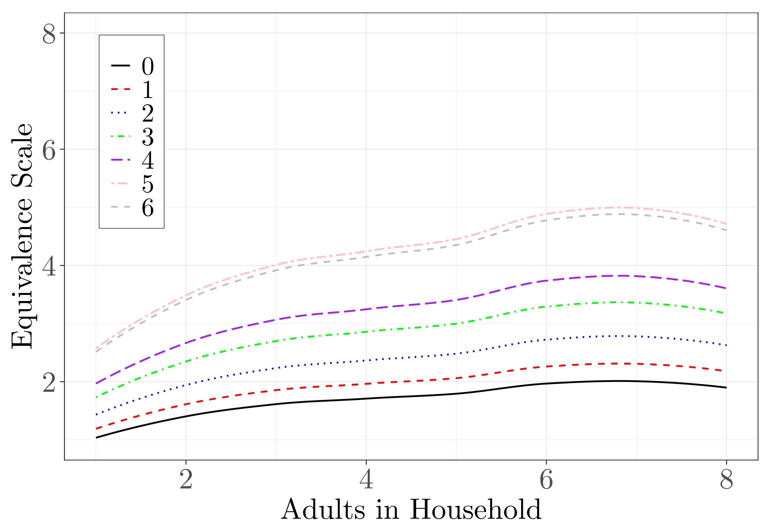

Figure D.2: Estimated equivalence scales assuming exogeneity and linear shares, but controlling for a variety of location and other demographic informaiton: Curves are separated for the number of children in the household.

D.2 Linear scales - endogenous

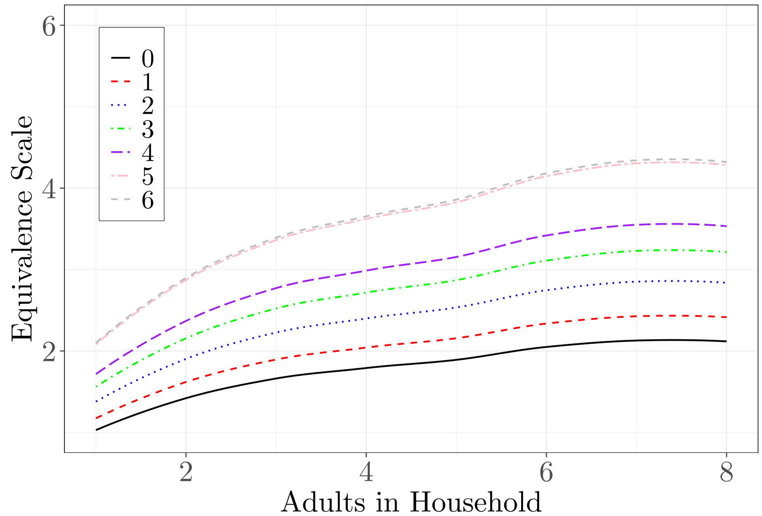

Figure D.3: Estimated equivalence scales assuming endogeneity and linear shares: Curves are separated for the number of children in the household.

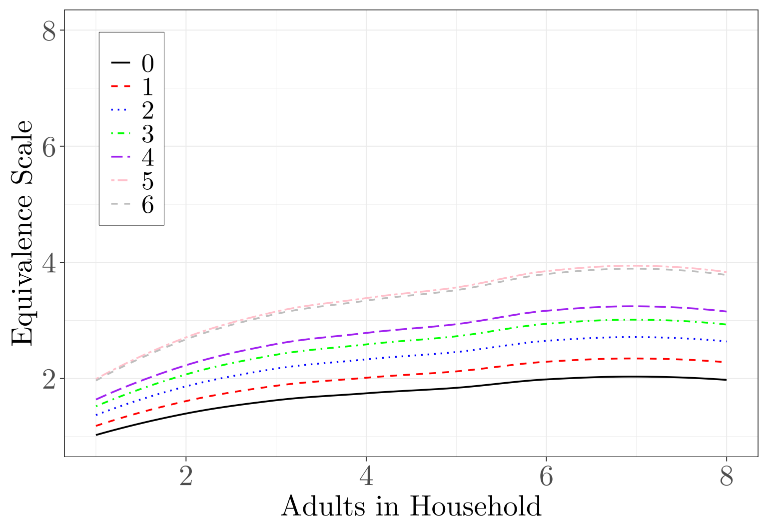

Figure D.4: Estimated equivalence scales assuming endogeneity and linear shares, but controlling for a variety of location and other demographic informaiton: Curves are separated for the number of children in the household.

D.3 Semiparmetric scales - exogenous

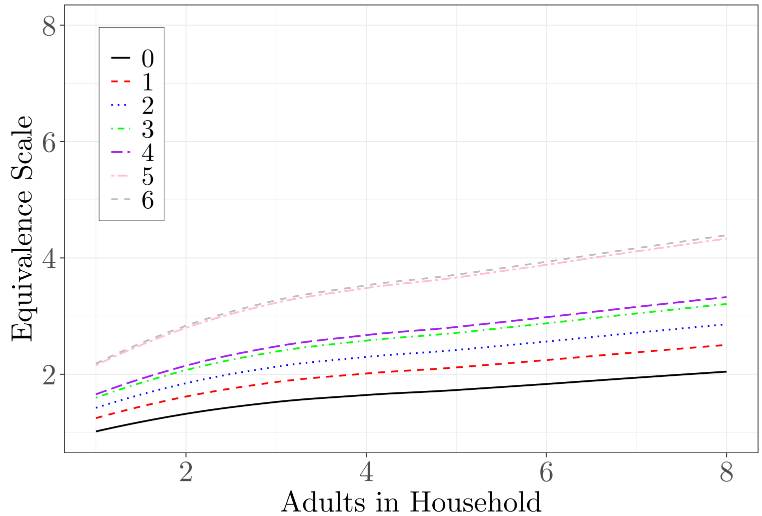

Figure D.5: Estimated equivalence scales assuming exogeneity and semiparamtric share estimation: Curves are separated for the number of children in the household.

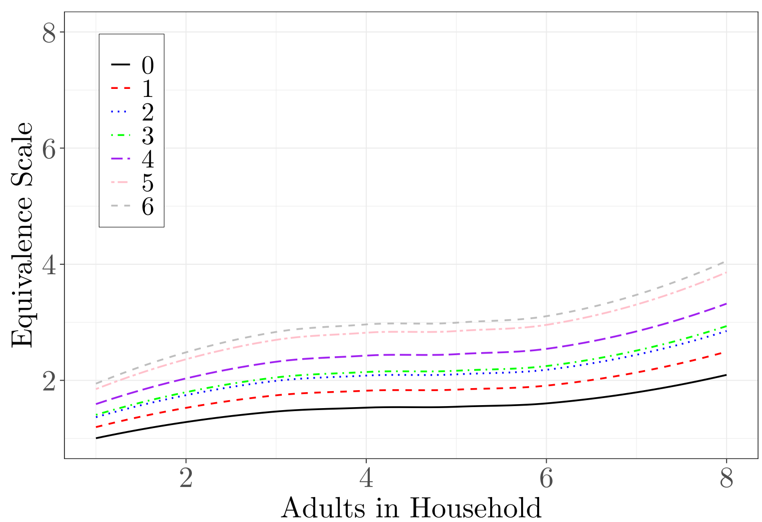

Figure D.6: Estimated equivalence scales assuming exogeneity and semiparamtric share estimation, but controlling for a variety of location and other demographic informaiton:: Curves are separated for the number of children in the household.

D.4 Semi-parametric scales: endogenous

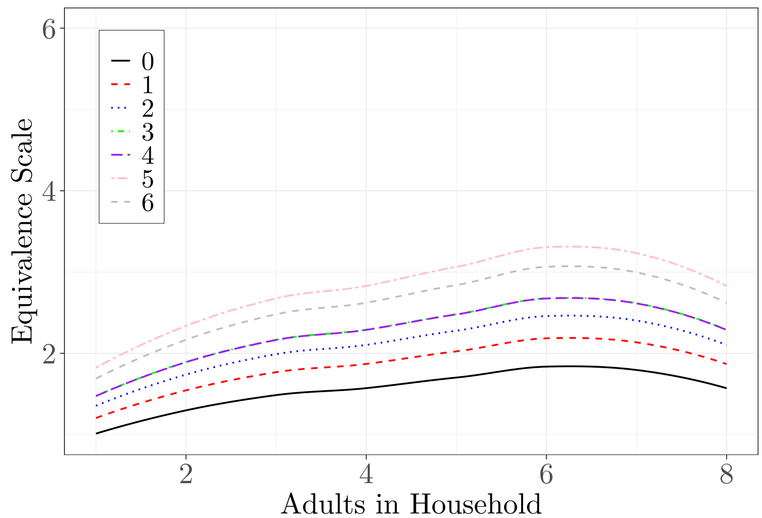

Figure D.7: Estimated equivalence scales assuming endogeneity and semiparametric share estimation: Curves are separated for the number of children in the household.

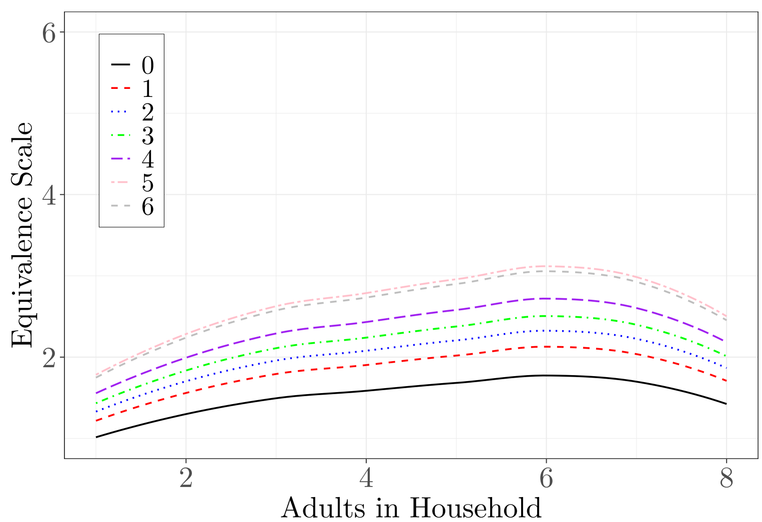

Figure D.8: Estimated equivalence scales assuming endogeneity and semiparametric share estimation, but controlling for a variety of location and other demographic informaiton: Curves are separated for the number of children in the household.