ARTICLE

OPEN

Relationship between households' share of food expenditure and income across South African districts: a multilevel regression analysis

Kabeya Clement Mulamba [ID] [1✉]

This paper investigates the relationship between household income and the share of food expenditure in South Africa using clustered household income and socio-economic data published in the National Dynamics Income Study. To achieve this objective and because of the clustered nature of the data, a multilevel linear model, in particular, the random intercept model, was employed as a suitable method of investigation. Fixed effect methods were also used to compare results. The results confirmed that the share of household food expenditure is inversely related to household income in a district. This finding is consistent with Engel's law, such that an increase in household income is related to a decrease in the proportion allocated to food items by a typical household within a district. It was found that the relationship between household income and food expenditure does not vary across districts. The finding also pointed out that the district context matters because the district's average household income significantly explains the share of food expenditure at the household level.

[1] University of Johannesburg, Johannesburg, South Africa. ✉email: mkclement@gmail.com

Introduction

This paper aims to answer the following three questions. First, "is monthly household income related to the share of food expenditure within a district in South Africa?" Second, "does household income's effect on the share of food expenditure vary across districts?" Third, "do district (or group) contextual factors matter in explaining the relationship between monthly household income and the share of food expenditure in South Africa?" These questions were set out to empirically test Engel's law (Engel, 1895), which postulates that the share of a household's food expenditure tends to decrease as its income increases.

An extensive survey of South African households containing information on monthly income, monthly food expenditure, and other demographic characteristics from the fifth wave of the National Income Dynamics Study (NIDS) (Southern Africa Labour and Development Research Unit, 2018) was employed to gather the necessary data for empirical analysis. One advantage of the NIDS is that households are nested within districts, making this clustered (or hierarchical) data. Given this data type, applying methods that assume independence of, for instance, the dependent variable may yield spurious results. Instead, it required the adoption of an empirical approach that, on the one hand, considers the possibility of similarity in the phenomena under investigation, particularly the share of household expenditure on food items for households residing in the same districts. On the other hand, the empirical approach should acknowledge the existence of heterogeneity among households living in different districts of South Africa. Empirical analysis to answer the posed three questions was carried out in the context of Tobler's (Tobler, 1970) first law of geography, which says: "everything is related to everything else, but near things are more related than distant things." In other words, Tobler's first law of geography was pertinent to this study because households belonging to the same districts are more related because of their close contiguity. In contrast, households from different districts are less related.

Furthermore, in the spirit of Engel, and as asserted by Chai and Moneta (2010), although he used sample data from Belgian households, his conclusions point out that this law is general, meaning that it applies to any household regardless of area. Studies that have investigated the relationship between households' food expenditure and income, including those focusing on South Africa, did not necessarily consider the possibility of dependence that characterises consumption choices by households in the same geographical area in a country (Bopape and Myers, 2007; Burger et al., 2017; Gummerson and Schneider, 2013; Koch, 2022; Maitra and Ray, 2006; Posel et al., 2020; Sekhampu, 2012; Waidler and Devereux, 2019; Yatchew et al., 2003). This dependence may occur because district-wide factors similarly affect households in the same district (i.e., geographical area). For instance, district's food prices, or lifestyles may affect food expenditure of households residing in a district in a similar manner. At the same time, it is plausible to expect that the consumption choices of households located in different geographical areas would be independent of each other. These two reasons and possibly many others are one way to understand Tobler's first law of geography. For these reasons, one contribution of the present paper was to test Engel's law by accounting for and estimating the dependence mentioned above in a district while accounting for differences between districts.

Another contribution of this paper was to disentangle the effect of household income on household food expenditure into two distinct effects. First, the paper aimed to investigate the relationship between household income, taken as a deviation from the mean of households' income, and the share of food expenditure within a district. Second, the present paper sought to investigate whether the district mean of household income relates to the household share of food expenditure. According to Feaster et al. (2011), the effect of household income as the deviation of a district average on the household share of food expenditure is called the *within-district household income effect*. In contrast, a district average income effect on the household share of food expenditure is referred to as the *between-district household income effect*. The purpose of partitioning the effect of household income on household food expenditure was to determine how the within-district household income effect differs from the between-district effect of household income in South Africa. This approach was adopted to answer the third research question: determining whether the contextual effect matters in explaining the relationship between household income and food expenditure.

Previous studies, including Waidler and Devereux (2019), have shed light on the relationship between household income and food expenditure in South Africa. However, the relevance of posing the questions mentioned above in the present paper is justified because economic phenomena must also be understood in their social (locational) or temporal context. Therefore, the empirical strategy adopted in this paper was use multilevel linear models (MLM) for reasons described in section three related to methodology. Fixed effects models (FEM) and ordinary least square models (OLS) were also estimated for comparison, as also further discussed in section "Methodology and data". The structure of this paper is as follows. The section "Previous studies" discusses previous studies on the topic and how they are related to this paper. The section "Methodology and data" presents the methodology and data, whereas the section "Results" discusses the results. The conclusion is presented in the last section.

Previous studies

There is an extensive literature of empirical studies testing the relationship between household income and food expenditure globally. Most of these studies use Engel's law as a theoretical basis for their analysis. It is worth noting that these studies vary in many respects, including in methodological approaches and understating of the law as postulated by Engel. For instance, some authors discuss the shortcomings of unitary models in which households' consumption behaviour is considered a representative or an aggregate of all household members (Attanasio and Lechene, 2010; Attanasio and Valerie, 2014; de Vreyer et al., 2020). In other words, these studies argue that intra-household heterogeneity is not accounted for in the so-called unitary models. Thus, because of this, they contend that unitary models hide the influence of household members on household consumption patterns. It is worth noting that there is merit in the argument that they put forward.

However, it was not feasible to adopt such an approach in the present paper because of the limitation of the NIDS data. In effect, not all household members in a household were asked to reveal their consumption and expenditure behaviour. On the contrary, the NIDS questionnaires are purposely designed so that a mother or her representative in a household is the respondent as far as household expenditure patterns are concerned. This means that the NIDS information related to household consumption and expenditure applies to the household as a unit, not its members.

Moreover, even if the information on household members regarding consumption patterns was available, this does not preclude one from carrying out the analysis by considering the household as the unit of measurement. In this case, the issue can only be raised when there is what is referred to as an "ecological

fallacy" in the estimation or interpretation of the models. For instance, an ecological fallacy may occur when the behaviour of household members is explained as if it was that of a household to which they belong. Based on this, it is worth noting that the analysis undertaken in the present paper considers only variables at the household level, with the following justifiable exceptions. In other words, a few variables that represent the characteristics of household heads included in the analysis are the exceptions. However, their interpretation is related to the entire household, as is discussed further in the section "Methodology and data" related to methods. In this respect, conclusions drawn in the present paper only apply to the behaviour of households and never to individual members of those households.

A few strands of literature emerge concerning South Africa. One strand of the literature includes studies such as Koch (2022), Posel et al. (2020) and Yatchew et al. (2003) that employed the equivalence scales methods to deal with, among others, the issue of endogeneity in the relationship between household income and food expenditure. It is important to note that the issue of endogeneity and how it is dealt with is discussed in the next section related to the methodology.

Another strand of South African literature includes studies that have employed traditional methods, such as OLS, to investigate the relationship mentioned above. Sekhampu's (2012) work is one of those studies related to the present paper regarding the unit of analysis, households, and the selection of predictors. Nevertheless, the work by Sekhampu (2012) is different from the present study as far as these elements are concerned. The scope in Sekhampu (2012) is limited to a township called Bophelong, which is situated 70 kilometres from Johannesburg. In contrast, the analysis in the present paper deals with national data covering a total of 4900 households distributed across all districts.

Grobler and Sekhampu's (2012) work also focuses on South Africa. However, its scope is limited in terms of geographical coverage and the typology of the targeted households. In effect, this work focuses on the same area as Sekhampu (2012) while only considering households receiving government grants. As can be seen, this study is limited to one area in South Africa. It becomes difficult to generalise its conclusions as areas are different, and one must understand that area-specific factors may influence household consumption. As already discussed, the present paper fills that gap by considering a nationwide survey. Moreover, focusing only on households that receive government grants, one does not learn much about what happens with households that do not receive government grants. In other words, it is not feasible, based on Grobler and Sekhampu (2012), to determine whether the pattern of expenditure on food is the same for households that receive government grants and those that do not.

Hence, one must understand the patterns of all households to get a complete picture of what is going on regarding their consumption choices. In this respect, the approach adopted in the present paper considers three alternative predictors of interest to analyse household food expenditure share. These three predictors are each drawn from a separate dataset constructed from the NIDS, as further discussed in the section "MLM specifications". This approach is helpful because it allows one not to differentiate households based on the source of their income. Whereas in the other two cases, there is indeed differentiation.

In essence, the discussion above shows that previous studies that have examined this research question, including household food security, food expenditure, and consumption in Africa, and South Africa in particular, used different empirical strategies, as mentioned above (Babalola and Isitor 2014; Belete et al., 1999; Browne et al., 2007; Coetzee, 2013; d'Agostino et al., 2018; de Cock et al., 2013; Grobler and Sekhampu, 2012; Gummerson and

Schneider, 2013; Muzindutsi and Mjeso, 2018; Rubhara et al., 2020). Most importantly, these studies did not employ MLM and overlooked the contextual effects of geographical units (i.e., districts) in which households or individuals reside, as discussed throughout this paper. In addition, the samples used in these studies are not as nationally representative as the NIDS, and their analytical approaches do not consider the group effects issue.

Before closing this section, it is essential to note that the analysis carried out in the present paper attempted to bridge the above-identified gap, particularly in South Africa. This is because it is the first time that the relationship between households' income and food expenditure is examined using MLM and considering the district contextual effects of household income on household food expenditure and other aspects, as discussed in the following section.

Methodology and data
The MLM is the chief method adopted for analysis in this paper. However, FEM (and OLS) models were also estimated for comparison purposes. This strategy was followed because of the NIDS data's clustered nature, which creates the potential issue of non-independence of the dependent variable, as further discussed by Huang (2016). In this regard, this section discusses the specifications of both the MLM and FEM models. The parallelism between these two approaches is also discussed in this section.

Notations. For ease of interpretation of the equations and symbols used in this section, it is vital to set the scene. First, households are indexed by i ($i = 1 \dots n$), while districts are represented by j ($j = 1 \dots m$). This representation implies that two levels of units or observations were analysed. These are households that were considered as observations at Level 1 and districts as observations at Level 2. Second, the dependent variable used in the estimations is associated with households and is represented by y_{ij}, with i (1, 2 … m). The term x_{ij} represents a column vector of the independent variable of interest (i.e., household income) associated with households. Third, for MLM specification, the derived district mean of the independent variable of interest (i.e., district mean of household per capita income) is considered related to observations at Level 2 and is represented by $\bar{x}_j$. Fourth, control variables are all at Level 1, and presented by Z_{ij} is an m-by-k matrix of k independent variables. Lastly, the two continuous variables (i.e., Age and Income) were transformed into logarithms.

MLM specifications. As set out below, a three-step procedure was followed in the specification and estimation of the MLM models. In each step, the models were estimated with the restricted maximum-likelihood estimator (see Hox et al., 2017) for details regarding the mathematical formulae).

Step 1: Null model. The first step consisted of estimating a null model without independent variables, except for the intercept, as shown in Eq. (1) below.

$$y_{ij} = \gamma_{00} + u_{0j} + \varepsilon_{ij} \quad (1)$$

$$u_{0j} \sim N\left(0, \sigma_u^2 \mathrm{I}\right)$$

$$\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2 \mathrm{I}\right)$$

where γ_{00} is the grand mean of the share of monthly household food expenditure across all districts, u_{0j} is the intercept specific for district j, also known as Level 2 error term, and ε_{ij} is the disturbance terms vector or Level 1 error term, assumed to follow a normal distribution and be homoscedastic. Except for the

disturbance term, Eq. (3) is composed of two components, notably the fixed-effects (γ_{00}) and the random-effects components (u_{0j}). Instead of u_{0j} being estimated, it is its variance, σ_u^2 in practice, as the latter conveys valuable information regarding the between-districts variation. In essence, the σ_u^2 quantifies the variability of the dependent variable (i.e., the share of household expenditure on food items) between districts, whereas σ_ε^2 quantifies the variability of the dependent variable in a district.

Moreover, Eq. (1) is used to test whether the between-district variance is greater enough than the within-district variance and to explain the share of monthly household expenditure on food using the intraclass correlation coefficient (ICC) as shown in Eq. (2) below. In simple terms, the ICC measures how similar households are within a district (Huang, 2016).

$$\text{ICC} = \frac{\sigma_u^2}{\sigma_\varepsilon^2 + \sigma_u^2} \qquad (2)$$

The ICC ranges from zero to one, with zero meaning complete independence and one a completed dependence. The economic literature has not established a cut-off of ICC to consider MLM or single-level method (i.e., OLS). However, it is a convention in other social sciences, including educational research, where MLM is used chiefly to consider MLM when ICC is equal to or >0.05. While this paper borrowed from this convention to determine the application of MLM, it with worth noting that common sense requires one not to assume independence in the observations once the ICC is above zero. In other words, an ICC above zero indicates considering MLM. Hox et al. (2017) note that failing to account for dependence and using single-level methods, such as OLS, when the data at hand is clustered, may lead to misestimating coefficients' standard errors. This paper also considered OLS estimation to compare the coefficients' standard errors with those of MLE and FEM models (OLS specification is not included in the text for space reasons). The next step is to estimate a variant of a multilevel linear model, called the random intercept model, as set out below.

Step 2: Random intercept model (RI). After confirming the suitability of multilevel over single-level logistic models, a random intercept model was estimated for each dependent variable. Equation (3) below represents the specification of the random intercept linear model considered in this paper.

$$y_{ij} = \gamma_{00} + u_{0j} + \gamma_{01}\bar{x}_j + \gamma_{10}\left(x_{ij} - \bar{x}_j\right) + \sum_{k=1}^{K} \gamma_{k,20}(z_{k,ij} - \bar{z}_{k,j}) + \varepsilon_{ij} \qquad (3)$$

$$u_{0j} \sim N(0, \sigma_u^2 \mathrm{I})$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2 \mathrm{I}),$$

The meaning of terms remains the same as in Eq. (1), except for the following. First, all independent variables, including the predictor of interest, were entered in deviations of their respective district means. Second, using deviations instead of the original variables is consistent with the procedure referred to as centring within a cluster (CWC) in the literature (Enders and Tofighi, 2007; Feaster et al., 2011; Huang, 2016; McNeish and Kelley, 2019) The CWC process, as well as the inclusion of a Level 2 predictor discussed below, was adopted to account for the contextual effects of household income on food expenditure, as stated in the third objective of the present paper (Raudensbush and Bryk, 2002) The CWC of these binary (dummy) control variables was done following the procedure laid out in Sommet and Morselli (2017) In this regard, the slopes of these dummy variables, taken as deviations, would be interpreted as the effect of being in the target group on the dependent variable, in clusters, on average (Yaremych et al., 2021).

Third, by entering, for instance, the deviation of "Income", one can assess whether households with higher (or lower) income than other households in their district tend to allocate higher (lower) proportions of expenditure on food items. For example, the coefficient γ_{10} is referred to in the literature as the "within effect" of household income on the share of food expenditure (Bell et al., 2019) Concerning the district mean of the variable of interest, this was entered as a unique independent variable at Level 2. The object of this variable was to determine the effect of the district mean of household income on household food expenditure. The intuition behind the inclusion of the Level 2 variable is that households in a district are interdependent regarding food expenditure. Therefore, what influences one household's food expenditure may also impact other households directly through direct interactions with other households in a district or indirectly. Another advantage of considering the district mean of household income as a Level 2 predictor is from a policy perspective. This variable can be used to assess whether, on average, districts categorised by higher (or lower) household income exhibit higher (or lower) household expenditure share, all other things being equal. For example, the literature refers to the coefficient γ_{01} as the "between effect" of household income on the share of food expenditure. Since Eq. (3) (and subsequently Eq. (4), includes the predictor of interest centred at the district mean, as well as the district mean, the Wald test was used to assess whether their coefficients are different (Feaster et al., 2011): H_0: $\gamma_{01} = \gamma_{10} = 0$, H_A: $\gamma_{01} \neq \gamma_{10}$.

Equation (3) predicts the share of (per capita) monthly food expenditure for household i in a district j as a function of the overall intercept (γ_{00}) (which is interpreted as the expected value of the dependent variable for cluster j), the deviations of selected independent variables concerning their district averages, the district mean of the independent variable of interest, and the district's random effects (u_{0j}). In addition, this model allows the intercept to vary between districts. For instance, the intercept of the share of monthly food expenditure for household i irrespective of the district is equal to γ_{00}, whereas that of a household i in district j is equal to $\gamma_{00} + u_{0j}$. As in Eq. (3), it is the variance $\left(\sigma_u^2\right)$ that is estimated.

It is worth noting that the specification in Eq. (1) does not allow for slopes to vary between districts, although the intercept varies. It is for this reason that this specification is referred to as a random intercept model. Equation (3) is generally compared to a more unrestricted model, also referred to as a random slope model, whose specification allows for intercept and slope variation. Hence, in the third step, Eq. (4) is estimated and compared to Eq. (3) using the likelihood ratio to select a suitable model, as discussed below.

Step 3: Random slope model (RS). Equation (6) is referred to as a random slope model. For computation purposes, the analysis in this paper considers a random slope model in which only the slope of one independent variable, notably "Income", is permitted to vary between districts. Equation (4) is the random slope model considered for this paper.

$$y_{ij} = \gamma_{00} + u_{0j} + \gamma_{01}\bar{x}_j + \left[\gamma_{10} + u_{1j}\left(x_{ij} - \bar{x}_j\right)\right] + \sum_{k=1}^{K} \gamma_{k,20}(z_{k,ij} - \bar{z}_{k,j}) + \varepsilon_{ij} \quad (4)$$

$$u_{0j} \sim N(0, \sigma_{u_0}^2 \mathrm{I})$$

$$u_{1j} \sim N(0, \sigma_{u_1}^2 \mathrm{I})$$

$$\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2 \mathrm{I})$$

The meaning of terms remains the same as in Eq. (4), except that the slope of the unique independent variable of interest (i.e.,

deviation of household income or wage) was allowed to vary across districts in line with the fourth objective set in this paper. The fixed-effects component of Eq. (4) is represented by

$$\left(\gamma_{00} + \gamma_{01}\bar{x}_j + \left[\gamma_{10} + \left(x_{ij} - \bar{x}_j \right) \right] + \sum_{k=3}^{K} \gamma_{k,20}(z_{k,ij} - \bar{z}_{k,j}) + \varepsilon_{ij} \right).$$

This component includes the overall intercept and the slopes that apply to all households regardless of the districts. The random-effects part of Eq. (4) is represented by $\left(u_{0j} + u_{1j}\left(x_{ij} - \bar{x}_j \right) \right)$, which is the intercept and the slope for one independent variable for a specific district. In this respect, one can say, for instance, that the influence of the variable of interest $(x_{ij} - \bar{x}_j)$ on the share of monthly food expenditure for a household in a district j is equal to $(\gamma_{10} + u_{1j})$. As for Eqs. (1) and (3), the focus is on the estimated variances of the random effects, represented by $\sigma_{u_0}^2$ and $\sigma_{u_1}^2$. The latter quantifies the between-district variance of the slope of the independent variable of interest.

FEM specifications. Huang (2016) notes that there are other alternative methods to analyse clustered data, among which FEM is prevalent in the empirical economics literature. This issue is further discussed by Oshchepkov and Shirokanova (2022) as they compare the use of MLM with other methods for clustered data, specifically in economics. While these authors provide guidance on which approach to apply based on a set of criteria, it is essential to note that Huang (2016) argues that the research question mainly determines the choice of a suitable approach. In addition, there is a tendency in other methods to treat the issue of group clustering as a nuisance. In contrast, the idea in MLM consists of taking account of this clustering by partitioning the variance of the disturbances into two (or as many as there are levels in the data) components. Furthermore, Oshchepkov and Shirokanova (2020) assert that methods used in economics to analyse clustered data are less convenient, flexible, and efficient than MLM. Based on the brief above, the FEM specification of the relationship between household income and share of food expenditure is represented by Eq. (5) below.

$$y_{ij} = \alpha_0 + \alpha_1(x_{ij} - \bar{x}_j) + \sum_{k=1}^{K} \alpha_{k,2}(\bar{z}_{k,ij} - \bar{z}_{k,j}) + u_j + \varepsilon_{ij}$$

$$\varepsilon_j \sim N\left(0, \sigma_\varepsilon^2 I\right) \tag{5}$$

The symbols in Eq. (5) are the same as in the previous equations. However, it is crucial to note that the term u_j is considered the district-fixed (not random) effects, and no assumption is made about it, as in the MLM specifications. For comparing Eq. (5) and MLM models, two FEM specifications derived from Eq. (5) were considered convenient estimation approaches.

Homogeneous-least square dummy variable model (Ho-LSDV). The Ho-LSDV, as shown in Eq. (6) below, is the first variant of FEM considered in this paper. It is important to note that it is not different from the famous LSDV used in the literature, except that the prefix homogenous was added to indicate that the effect of the coefficient of the independent variable of interest does not vary across districts.

$$y_{ij} = \alpha_0 + \alpha_1(x_{ij} - \bar{x}_j) + \sum_{k=1}^{K} \alpha_{k,2}(z_{k,ij} - \bar{z}_{k,j}) + \sum_{j=1}^{J-1} \omega_j D_j + \varepsilon_{ij} \tag{6}$$

$$\varepsilon_{ij} \sim N\left(0, \sigma_\varepsilon^2\right)$$

where D_j is the district j, the coefficient $\omega_j\left(j = 1, \ldots, J-1\right)$ is interpreted as the difference between the effects of district j and

the reference district on the dependent variable. As can be seen, the Ho-LSDV model accounts for a clustered data structure by including the district affiliation information directly into the model. In other words, district effects are considered fixed or simply categorical predictors of the dependent variable in Ho-LSDV. In contrast, in the MLM specifications, the random effects are specified to deal with the dependence issue for households in the same districts.

Furthermore, it is also worth noting that, α_1 in Eq. (6) is interpreted in the same way as γ_{10} in Eq. (3). Bell et al. (2019) assert that the Ho-LSDV provides an estimate of "within effect" that is not biased by the "between effect". Up to this point, one can conclude that RI is like Ho-LSDV as far as the "within effect" coefficients are concerned. However, the principal difference between Ho-LSDV and RI is that the former cannot include Level 2 variables, such as the district mean household income. It also means that the Ho-LSDV does not have the "between effect". Therefore, this paper's third research question cannot be answered with the Ho-LSDV model, whereas the first question can. In other words, McNeish and Kelley (2019) argue that FEM has limitations in addressing the contextual effects of income on household food expenditure. The second variant of FEM is described in the next section.

Heterogeneous-least square dummy variable model (He-LSDV). Equation (7) represents the He-LSDV, and all symbols are as in Eq. (6), except that the term α_3 captures the interaction between the independent variable of interest, household income and district dummies. The He-LSDV specification is like RS (i.e., Eq. (4)), except that the "between effect" is not considered in the former. This is because the relationship between household income and share of food expenditure is allowed to vary across districts through the term α_3. The He-LSDV has limitations in accounting for contextual effects, as with Ho-LSDV. Therefore, only the first and second research questions can be answered with He-LSDV.

$$y_{ij} = \alpha_0 + \alpha_1\left(x_{ij} - \bar{x}_j \right) + \sum_{k=1}^{K} \alpha_{k,2}(z_{ij} - \bar{z}_j) + \alpha_3\left(x_{ij} - \bar{x}_j \right)\sum_{j=1}^{J} D_j + \sum_{j=1}^{J} \omega_j D_j + \varepsilon_{ij}$$

$$\tag{7}$$

Endogeneity challenge. Summers (1959) argued that the relationship between household income and expenditure is characterised by endogeneity. Consequently, one must apply analytical techniques such as the instrumental variable (IV) method to circumvent this difficulty so that results are not spurious and misleading. It is one of the reasons that Koch (2022) and Posel et al. (2020) proposed applying the equivalence scales method, where the relationship between household income and food expenditure is examined in the case of South Africa. While Koch (2022) adequately dealt with endogeneity, their proposed method is unsuitable to address the posed questions, particularly how to account for the richness of clustered data, as discussed throughout the present paper.

In addition to the endogeneity issue argued by Summers (1959) that is inherent to the relationship between household income and food expenditure, scholars have demonstrated the difficulty of dealing with it in MLM settings (Aiello and Bonanno, 2018; Grilli et al., 2011; Qian et al., 2020). It is clearly shown in this literature that there are various sources of endogeneity in MLM. Regardless of its origin, the consequences of endogeneity are the same concerning the reliability of results. It is also crucial to note that it is beyond the scope of this paper to discuss sources of endogeneity.

Based on the preceding, the approach proposed by Qian et al. (2020), which recommended using MLM if there are no suitable candidate methods, was followed in the present paper. According

Table 1 Description of variables used in the analysis.

Variable	Description
Dependent	
Food expenditure share	Per capita household monthly expenditure on food in South African rand as a percentage of per capita household monthly total expenditure in South African rand
Independent	
Income	Logarithm of per capita household total monthly income in South African rand
Age	Logarithm of completed years of the household head at the time of the survey
Gender	A binary variable. It is equal to one (reference) if the household head is male, whereas it is zero if female.
Marital status	A binary variable. It is equal to one (reference) if the household head is legally married or living together as husband and wife, whereas it is zero otherwise.
Education	A binary variable indicating whether a household head has completed at least a secondary education. It is equal to one (reference) if they have no secondary qualification and zero otherwise.
Ethnicity	A binary variable indicates the predominant population group to which the household belongs. It is equal to one (reference) when the predominance is black African, and it is equal to zero for non-black.
Area	A binary variable that indicates whether a household is located in an urban area or not. It is equal to one (reference) if the household is located in an urban area and zero in a non-urban area.

to Qian et al. (2020), the marginal interpretation of the coefficients of endogenous independent variables in the fixed-effects component of MLM, in this case, is no longer valid. Instead, these coefficients will have a "conditional-on-the-random-effect interpretation" (Qian et al., 2020, p. 390). As can be seen, the MLM was suitable to answer the posed research questions for this and other reasons discussed in the preceding sections.

Data. Table 1 shows the variables used to estimate the models discussed in the preceding section. These variables were sourced from the NIDS and are considered level 1 because they relate to households. Only one Level 2 variable is considered in this paper: the district means of the independent variable of interest (i.e., Income). This variable was purposely created to answer the posed research questions. Before discussing the data, it is important to say one or two words about the NIDS, a longitudinal panel survey of South African households commissioned by the National Department of Planning, Monitoring and Evaluation (Brophy et al., 2018) First, technical issues, including the sampling design adopted in the NIDS, can be found in Brophy et al. (2018).

Second, the data used for analysis in this study is mainly from the file "Household derived" of the fifth wave, which covers 13,719 respondents. However, some households were discarded because of missing information. This resulted in 4900 households being considered in the analysis after the process. Third, information from other files, notably "Household roster", "Child", and "Adults" of the fifth wave, was also gathered to construct variables representing characteristics of household heads (e.g., Age, Marital status, Gender). Last, it was noted that the number of households in a district was unevenly distributed. In effect, on average, there were 94 households in a district, with a minimum and maximum of 23 and 273 households, respectively. As can be seen, these statistics were consistent for one to carry out the analysis using the MLM method mainly.

Now turning to the variables reported in Table 1. First, the dependent variable, "Food expenditure share", was constructed from the information in the above files by ensuring that the numerator and denominator are normalised for household size. This normalisation was crucial because one could not compare households of different sizes regarding consumption choices. Many empirical studies, including Engel (1895), have pointed out that the composition of a household is a determinant of its food consumption. Thus, it must be adequately addressed in the analysis. Now, the question became if one must consider the number of household members provided by the NIDS to normalise the dependent and the independent variable of interest, as shown in Table 1 (e.g., Food expenditure share and income).

Hymans and Shapiro (1976) present a concise discussion on this matter. These authors recommend using "equivalent persons" instead of "persons" as the actual number of household members. Coincidentally, the "Child" and "Adult" files of the NIDS provide information regarding the number of household members aged between 0 and 14 years old and between 15 years old and more, respectively. As can be seen, this means that, according to the NIDS, an adult is a person aged 15 years old and more. For analysis in the present paper, the actual number of household members, based on the working definition of "equivalent person", was determined based on the information from the above-cited files from the NIDS. In effect, the number of children who are members of the household was first divided by two. This is also to say that each child member is equivalent to 0.5 full household members. After obtaining the corrected number of child members, this number was added to the number of adult members to get the total number of household members used for normalisation. Although this number may have shortcomings, it is the view of the author that it considers the issue of differences in the composition of a household from a size point of view.

While on this issue of the dependent variable, it is worth noting that there is no consensus in the empirical literature on the measure of household food expenditure or the normalisation procedure. For instance, some studies use household food expenditure (not normalised) as the dependent variable, and household size is also used as one of the predictors. Other studies, in contrast, use the share of household food expenditure. However, in the author's view, it was critical to normalise the information. The "Food expenditure share" is more suitable to test Engel's law, as discussed in the introduction. Hence, it was the only dependent variable considered in the present study.

Second, the predictor of interest, "Income", was selected based on the literature and data availability to answer the research questions. Based on Engel's law, it is expected that the relationship between "Income" and "Food expenditure share" will be negative. Third, the control predictors considered in the analysis include "Age", "Gender", "Marital Status", "Education", "Ethnicity", and "Area". Most of these control predictors depict the characteristics of household heads. Jayasinghe and Smith (2021) highlight the importance of household head characteristics in household food consumption decisions. From a practical point of view, to appreciate the relevance of household head characteristics, one must first understand the definition of a household head. Although the NIDS survey does not define a household head, it is the

Table 2 Summary statistics of continuous variables.				
Variable	Mean	Maximum	Minimum	Standard deviation
Food expenditure share	31	94	0.3	17
Age	46	92	18	14
Income	4861	868,507	51	15,596

Table 3 Summary of categorical variables.			
Variable	Category	Number of households	Percentage
Gender	Male	2701	55
	Female	2199	45
Marital status	Married	3734	76
	Not married	1166	24
Education	Education	1476	30
	No education	3424	70
Ethnicity	Black	3604	74
	Non-black	1296	26
Area	Urban	3013	61
	Non-urban	1887	39

understanding in this paper that it uses Statistics South Africa's definition.[1] A household member is considered as household head when they are the primary decision-maker, the person who owns or rents the dwelling, or the primary breadwinner.

Based on this definition, it is logical to expect that the household's consumption decisions, including food expenditure, would be significantly influenced by the household head. Thus, it is crucial to include household headship characteristics as predictors of household food expenditure, notably the household head's age, gender, marital status, and education level, as discussed in detail in Jayasinghe (2019) and Posel et al. (2020).

It is difficult to predict the expected relationship between the age of the household head and "Food expenditure Share". However, if one could associate age with income, it is possible to believe that, as the age of a household head increases, her income also increases. Thus, because of their position as a breadwinner, it is possible to observe a decrease in the share of household food expenditure, all other things being equal. The inclusion of "Gender" as a predictor in the present paper is critical, particularly in the context of South Africa, where women are believed to be disadvantaged over their male counterparts. For instance, because of historical and persistent societal and cultural norms entrenched in South Africa, one would expect that women face discrimination regarding access to means of production (e.g., land), well-paying jobs, and so forth. Though Jayasinghe and Smith (2021) do not focus on the same topic as in the present paper, they take this matter even further in justifying why the inclusion of such as predictor should be considered. These authors explain that there are subtle differences between female-headed and male-headed households regarding the composition and size of households.

Moreover, the aim of including "Gender" as a predictor in the present study is to determine whether there are differences between female-headed and male-headed households regarding food expenditure share in South Africa. It is for the same reason as for "Gender" that "Marital status" is also included as a predictor in the present study. The aim is basically to get an understanding of whether there are differences in food expenditure between households headed by married people and those that are not.

The last household head characteristic included as a predictor in the present study is "Education". As with age, it is intuitive to associate this variable with income or poverty. Therefore, it is logical to expect that households whose heads do not have an education will allocate a more significant portion of their income to food expenditure. In contrast, the opposite is possible for households whose heads are educated. Regarding "Ethnicity" and "Area", their inclusion accounts for disparities in South Africa.

Table 2 summarises the main descriptive statistics of continuous variables. A mean of 31 for "Food expenditure share" implies that, on average, 30 per cent of monthly expenditure for a typical household in South Africa is allocated to food. However, a maximum of 94 for "Food expenditure share" is problematic because it shows that some households in South Africa almost have their entire expenditure on food. It also can be seen, in Table 2, that there are disparities among households for all variables. These are evident in the large standard deviations. It is important to note that some of these disparities can be explained partly because of the use of entire cross-sections of households. However, as discussed in the following section, only the deviations from the district means of these variables were used for regression analysis (refer to the CWC process explained in the section "MLM specifications"). In this way, one does not have to worry about these disparities.

The summary of categorical variables is provided in Table 3. The first column of this table shows a variable, whereas the second depicts categories associated with the variable. The reference categories are displayed in bold font. The number of households in each category is shown in column three, and the fourth column shows the same information in percentages. For instance, one can read the second row of Table 3, which is related to the variable "Gender", as follows: 2199 out of 4900 households have females as heads, representing 45 per cent.

It can also be seen that about 70 per cent of household heads in the sample had not completed secondary education. Given that secondary education is considered the basic level of education required, this finding indicates that more effort is needed to educate masses of South Africans.

Results

Models' comparison. The purpose of this section is to compare the estimated models to determine one that is suitable for the data in line with the discussion in the section "Methodology and data". The interpretation of the suitable model's coefficients concerning the posed requestions is provided in the section "Discussion of RI estimates". In this respect, Table 4 reports, in addition to OLS and the null model, Ho-LSDV and RI, which are variants of FEM and MLM. These models assume a homogeneous or non-varying relationship between household income and expenditure within a district. Except for the intercept, the coefficients of Level 1 variables should be read as the coefficients of deviations from district means of variables of interest, as well as of the control variables in line with the process of centring within the context outlined in the section "MLM specifications". The term "District mean income" represents the coefficient of the district mean of "Income".

The first step consisted of assessing whether the dependence issue because of data clustering matters in modelling the relationship between household income and food expenditure. This process was achieved with the reported ICC statistic under the null model in Table 4. The ICC of 0.131 means that 13 per cent of the variance in households' food expenditure share can be attributed to differences between districts. This finding suggested that one cannot estimate a single-level regression because of the similarities among households in a district concerning food expenditure. LR and AIC statistics in Table 4 were also used to complement ICC for the assessment mentioned above. In the case

Table 4 Results of homogeneous models.

Parameter	OLS	Null	Ho-LSDV	RI
Fixed-effects				
Intercept	31.242*** (0.208)	32.212*** (0.879)	19.382*** (2.347)	117.976*** (9.053)
Income	−5.517*** (0.227)		−5.517*** (0.207)	−5.517*** (0.207)
District mean income				−11.361*** (1.196)
Age	−2.379*** (0.718)		−2.379*** (0.653)	−2.379*** (0.6535)
Gender	0.763* (0.451)		0 .763* (0.410)	0.763* (0.410)
Marital status	−0.949*** (0.549)		−0.949* (0.499)	−0.949* (0.499)
Education	−4.792*** (0.525)		−4.792*** (0.478)	−4.792*** (0.478)
Ethnicity	1.367** (0.682)		1.367 ** (0.620)	1.367** (0.620)
Area	−6.105*** (0.629)		−6.105*** (0.572)	−6.105*** (0.572)
Random-effects				
Residual variance (Level 1)		243.111** (4.938)		176.312** (3.584)
Intercept variance (Level 2)		37.111** (8.014)		12.350** (2.949)
ICC		0.133		
LR test against single-level		541.88*** [0.000]		171.72*** [0.000]
Log-likelihood	−20,083.693	−20,479.198	−19,527.34	−19,668.751
AIC	40,185.39	40,964.4	39,174.68	39,359.5
Wald statistic for equality				23.14*** [0.000]
Number of groups (Level 2)	52	52	52	52
Number of observations (Level 1)	4900	4900	4900	4900

Source: Author's estimations.
***, ** and * refer to significance at the 1%, 5% and 10% levels, respectively. Standard errors are in parenthesis, while *P*-values are in brackets.

of LR, the null hypothesis was rejected at a one per cent level. OLS's reported AIC is bigger than Ho-LSDV and RI's. Based on this finding, it was concluded that the data was suitable for models that consider the issue of dependence instead of OLS (i.e., single-level model), as discussed throughout this paper. It is worth noting that the above discussion and the conclusion, therefore, constitute one contribution of this paper in assessing the relationship between the two phenomena in question in the context of clustered data.

Though the assessment confirmed that OLS or any other single-level model was not relevant for the data at hand, results from OLS estimation were included in Table 4 to compare them with those from the other two models. The comparison between these models revealed the following. OLS results are like those reported for Ho-LSDV and RI (at least for coefficients in the fixed-effect component). Nevertheless, the coefficients' standard errors of OLS were overestimated. This finding was not surprising as the literature establishes that the data clustering issue is one cause of the overestimation of coefficients' standard errors in OLS. Thus, MLM (i.e., RI) and FEM (i.e., Ho-LSDV) are recommended.

Moreover, RI and Ho-LSDV coefficients and standard errors are similar. However, Ho-LSDV does not include "district mean income", a Level 2 variable. This variable captures the "between effect" or "contextual effect". RI included it to assess how income at a district level (i.e., macro level) is related to household food expenditure. Table 4 also shows that the Wald statistic is significant, indicating no equality between the coefficients

"Income" and "District mean income". It can also be noted that the inclusion of both coefficients into the RI specification was statistically sound to capture the "within effect" and "between effect" of household income on food expenditure. This finding also exposes the limitation of FEM (i.e., Ho-LSDV), as it fails to capture the contextual effect. Because of this reason, RI seemed a suitable model whose results can be used to answer the first and third research questions posed in this paper. In contrast, the second research question cannot be answered with RI. Thus, He-LSDV and RE were estimated and then compared to Ho-LSDV and RI, respectively, mainly using AIC statistics to determine whether the relationship between household income and food expenditure varies across districts.

He-LSDV and RE results are reported in Table 5. The AIC of He-LSDV is bigger than the AIC of Ho-LSDV in Table 4, which is an indication that the latter model is preferable to the former. In other words, as per the FEM model, it is statistically sound not to include the interaction between "Income" and the dummy variable district, as the relationship between household income and food expenditure is not varying. Despite this finding, Ho-LSDV is still not a suitable model, as discussed in the preceding paragraph.

Furthermore, the AIC for RE is equal to that of RI in Table 4. This finding implies that there was no significant difference between RI and RE. In other words, it cannot be concluded that the relationship between household income and food expenditure varies across districts. This conclusion also confirmed the size of the variance of the random effect of income ("Income variance

Table 5 Results of heterogeneous models.

Parameter	He-LSDV	RS
Fixed-effects		
Intercept	24.383***	117.977***
	(2.3974)	(9.054)
Income demeaned	−4.575**	−5.583***
	(1.929)	(0.226)
District mean income		−11.361***
		(1.196)
Age demeaned	−2.418***	−2.415***
	(0.660)	(0.654)
Gender/gender demeaned	0.823**	0.777*
	(0.413)	(0.410)
Marital status/marital status demeaned	−1.074**	−0.972*
	(0.502)	(0.499)
Education/education demeaned	−4.759***	−4.783***
	(0.481)	(0.478)
Ethnicity/ethnicity demeaned	1.667***	1.491**
	(0.645)	(0.625)
Area/area demeaned	−5.927***	−6.066***
	(0.599)	(0.575)
Random-effects		
Residual variance (Level 1)		175.904**
		(3.588)
Intercept variance (Level 2)		12.356**
		(2.949)
Income demeaned variance (Level 2)		0.336**
		(0.330)
LR test		173.46***
		[0.000]
Log-likelihood	−19,432.484	−19,667.882
AIC	39,086.97	39,359.76
Wald statistic		22.50***
		[0.000]
Number of groups (Level 2)	52	52
Number of observations (Level 1)	4900	4900

(Level 2)) relative to intercept variance in Table 5. In conclusion, because RI is the suitable model, the following section discusses in detail its estimates shown in Table 4 in relation to the posed research questions.

Discussion of RI estimates. First, looking at the fixed-effects component of RI, it is noted that the coefficient of "Income" is statistically significant. This finding confirms Engel's law in that when there is an increase (decrease) of per capita household monthly income, the share of food expenditure decreases (increases), conditional on random effects. Based on this result, it can be inferred that it is expected that across all districts, a typical household in a district will decrease (increase) the share of monthly food expenditure by a factor of 0.05 per cent, all other things being equal.

The findings for coefficients "Income" is different from what Grobler and Sekhampu (2012) and Sekhampu (2012) have established. Both papers found a positive relationship between household income and food expenditure. In reality, this is not contradictory whatsoever because these studies consider "Household food expenditure". In contrast, in the present study, and per Engel's law, the "share of household monthly food expenditure" is the dependent variable. It is crucial, at this stage, also to discuss the coefficient "District mean income". As already said, the reported Wald statistic is statistically significant, which means that the inclusion of "District mean income" into the model was valid. As can be seen, the reported coefficient of "District mean income" is significant, which implies that the district context explains the relationship between household food expenditure

and income in South Africa. In terms of magnitude, the coefficient of "District mean income" is greater than that of "Income" in absolute terms. This coefficient is also negative and aligned with Engel's law, as discussed in the preceding sections. Thus, one could also say that both the 'within- and between-effects' were found significant in this paper.

After discussing the results related to "Income" and "District mean income", it was essential to provide answers to the posed research questions. In this regard, it was concluded that monthly household income was related to the share of food expenditure within a district in South Africa. Similarly, district contextual factors matter in explaining this relationship. In contrast, there was no satisfactory conclusion as to whether the relationship between monthly household income and the share of food expenditure varies across districts for the period under consideration.

Regarding the coefficients of the control variables in the fixed-effects component, it is worth noting that all demographic factors determine household expenditure on food items. For instance, "Age" is significant and inversely related to "Food expenditure share". This finding suggests that an increase by one year of a household head corresponds to a decrease of 0.02 per cent in the share of monthly expenditure that a typical household in a district in South Africa allocates to food. This finding aligns with the intuition elaborated in section three that age is related to income. Hence the coefficient behaves as reported in Table 4. The reported positive estimate of "Gender" is an indication that households tend to increase the share of food expenditure when their heads are male as opposed to female-headed households in a district. On the contrary, households whose heads are married tend to decrease their share of food expenditure, all other things being equal.

The reported coefficients of "Education" in Table 4 reveal that education is inversely related to household food expenditure. This result indicates that, in a district, households whose heads do not have at least a secondary education increase (decrease) the allocation of their total monthly expenditure to food items by 4 per cent as their income decreases(increases). Furthermore, the reported ethnicity-related coefficient reveals another dimension of disparity in South African society. For instance, it can be seen from this coefficient that the relationship between household food expenditure and typical predominantly non-black households in a district is positive. In other words, predominantly non-black households are not poor compared to black households.

Another aspect of disparity in South Africa can be explored by looking at the coefficients related to the geographical areas of households. In this regard, Table 4 reports that the coefficient of "Area" is negative and statistically significant in all three models. This finding suggests that a typical urban household in a district allocates a smaller proportion of its total monthly expenditure to food items than a household in a non-urban setting. It also means a geographical divide in South Africa because urban households are more affluent than their counterparts in rural areas.

Second, the discussions were centred on the fixed-effects component up to this point. However, it can be recalled that the RI specification only allows the intercept to vary between districts. In this respect, on average, the estimated intercept for Model 1, for instance, which is "Food expenditure share" of household i in district j, is equal to $\left(12.35 + u_{0j}\right) u_j$. What is important to note here is that, for households belonging to the same districts, the intercept will not change because, in this case, u_{0j} will not vary, whereas, for households belonging to different districts, the intercept will change in the proportion of u_{0j}. In addition, Table 4 reports the "Intercept variance", and as can be seen, this statistic is significant to indicate the variability of the concerned parameter at Level 2 or between districts.

Finally, and in general, it is essential to note that the empirical findings based on the information from the NIDS household survey of the present paper demonstrate the validity of Engel's law for South Africa consistently. Together with findings for other predictors, it became pertinent to consider contextual effects in the modelling because economic agents belonging to the same districts are influenced by latent but common factors. Consequently, it means that the food expenditure behaviour for households in the same districts will contain a proportion of similarities. An MLM model was selected to estimate the relationship between household monthly income and share of food expenditure to unravel these contextual effects and address the issue of autocorrelation in the dependent variable, a, on the one hand. On the other hand, the present paper uses data covering a large sample of households across all districts in South Africa.

To the author's best knowledge, an analysis that considers contextual effects has been lacking up to this point, at least in the South African context. One aspect that requires further research on this subject matter is perhaps to understand the dynamism of this relationship over time.

Conclusion
This paper has tested the significance of the relationship between household income and food expenditure in South Africa. The topic is truly relevant in this country. It aimed to test the validity of Engel's law, mainly using a multilevel linear regression approach to account for the homogeneity or similarities in the consumption behaviour of households in the same districts. By adopting this methodological approach, the present paper was an attempt to bridge the gap in empirical studies, in particular those that focus on South Africa, in the sense that one has to account for group effects when analysing the behaviour of economic agents that are located in different geographical or social settings.

A diagnostic process found that the random intercept model was appropriate for the sample data. This implies that households in a district exhibit the same behaviour as far as food expenditure is concerned. It then was found that, for three models, the empirical results confirm the validity of Engel's law in the sense that it was established that an increase (decrease) by one South African rand in per capita household monthly income leads to a decrease (increase) of two per cent in the share of household monthly expenditure allocated to food items in a typical South African household in a district.

Data availability
This paper used quantitative data published by Southern Africa Labour and Development Research Unit. R and Stata version 17 software were also for quantitative analysis. In this regard, a Stata Do-file can be made available upon request.

Note
1 See the General Household Survey for a definition of a household head: P03182018.pdf (statssa.gov.za).

References
Aiello F, Bonanno G (2018) Multilevel empirics for small banks in local markets. Pap Reg Sci 97(4):1017–1037
Attanasio O, Lechene V (2010) Conditional cash transfers, women, and the demand for food. No. 10, 17. IFS Working Papers

Attanasio O, Valerie L (2014) Efficient responses to targeted cash transfers. J Political Econ 122(1):178–222
Babalola DA, Isitor SU (2014) Analysis of the determinants of food expenditure patterns among urban households in Nigeria: evidence from Lagos State. IOSR J Agric Vet Sci 7:71–75
Belete A, Igodan CO, M'Marete C et al. (1999) Analysis of rural household consumption expenditure in South Africa: the case of food plot holders in Tyefu irrigation scheme in the eastern cape province. Agrekon 38(2):194–203
Bell A, Fairbrother M, Jones K (2019) Fixed and random effects models: making an informed choice. Qual Quant 53(2):1051–1074
Bopape L, Myers R (2007) Analysis of household demand for food in South Africa: model selection, expenditure endogeneity, and influence of socio-demographic. Paper presented at the African Econometrics Society Annual Conference, Cape Town, South Africa
Brophy T, Branson N, Daniels RC et al. (2018) National Income Dynamics Study panel user manual. Release 2018. Version 1. Southern Africa Labour and Development Research Unit, Cape Town
Browne M, Ortmann GF, Hendriks S (2007) Expenditure elasticities for rural households in the Embo ward, Umbumbulu, Kwazulu-natal. Agrekon 46(4):566–583
Burger RP, Coetzee LC, Kreuser CF et al. (2017) Income and price elasticities of demand in South Africa: an application of the linear expenditure system. S Afr J Econ 85(4):491–514
Chai A, Moneta A (2010) Retrospectives Engels curves. J Econ Perspect 24(1):225–240
Coetzee M (2013) Finding the benefits: estimating the impacts of the south African child support Grant. S Afr J Econ 81(3):427–450
d'Agostino G, Scarlato M, Napolitano S (2018) Do cash transfers promote food security? The case of the South African Child Support Grant. J Afr Econ 27(4):430–456
de Cock N, D'Haese M, Vink N et al. (2013) Food security in rural areas of Limpopo province, South Africa. Food Secur 5(2):269–282
de Vreyer P, Lambert S, Ravallion M (2020) Unpacking household Engel curves. No. w26850. NBER Working Paper
Enders CK, Tofighi D (2007) Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. Psychol Methods 12(2):121–138
Engel E (1895) Die Lebenskosten belgischer Arbeiter-Familien früher und jetz. Int Stat Inst Bull 9(1):1–74
Feaster D, Brincks A, Robbins M et al. (2011) Multilevel models to identify contextual effects on individual group member outcomes: a family example. Fam Process 50(2):167–183
Grilli L, Rampichini C, Rampichini GL (2011) Model building issues in multilevel linear models with endogenous covariates. In: Proceedings of ISAC-INTERFACE-IFCS Workshop
Grobler WCJ, Sekhampu J (2012) Determinants of household food expenditure among household receiving government grants in South Africa. In: Proceedings of the Annual world business research conference
Gummerson E, Schneider D (2013) Eat, drink, man, woman: Gender, income share and household expenditure in South Africa. Soc Forces 91(3):813–836
Hox JJ, Moerbeek M, van de Scoot R (2017) Multilevel analysis: techniques and application (third). Routledge
Huang FL (2016) Alternatives to multilevel modeling for the analysis of clustered data. J Exp Educ 84(1):175–196
Hymans SH, Shapiro HT (1976) The allocation of household income to food consumption. J Econom 4(2):167–188
Jayasinghe M (2019) Dynamic trends in household poverty and inequality in Sri Lanka: do gender and ethnicity matter? J Asia Pac Econ 7860
Jayasinghe M, Smith C (2021) Poverty implications of household headship and food consumption economies of scales: a case study from Sri Lanka. Soc Indic Res 155(1):157–185
Koch SF (2022) Equivalence scales in a developing country with extensive inequality. S Afr J Econ 1–27
Maitra P, Ray R (2006) Household expenditure patterns and resource pooling: evidence of changes in post-apartheid South Africa. Rev Econ Househ 4(4):325–347
McNeish D, Kelley K (2019) Fixed effects models versus mixed effects models for clustered data: reviewing the approaches, disentangling the differences, and making recommendations. Psychol Methods 24(1):20–35
Muzindutsi PF, Mjeso T (2018) Analysis of South African household consumption expenditure and its determinants: application of the ARDL model. Euro Econ 3:169–179
Oshchepkov, AY, Shirokanova A (2020) Multilevel modeling for economists: why, when and how. Higher School of Economics Research (No. 233)
Oshchepkov A, Shirokanova A (2022) Bridging the gap between multilevel modeling and economic methods. Soc Sci Res 104:102689
Posel D, Casale D, Grapsa E (2020) Household variation and inequality: The implications of equivalence scales in South Africa. Afr Rev Econ Finance 12(1):102–122

Qian T, Klasnja P, Murphy SA (2020) Linear mixed models with endogenous covariates: modeling sequential treatment effects with application to a mobile health study. Stat Sci 35(3):375–390

Raudensbush SW, Bryk AS (2002) Hierarchical linear models: applications and data analysis methods (Second). SAGE Publications Ltd.

Rubhara TT, Oduniyi OS, Mudhara M et al. (2020) Analysis of household food expenditure patterns. a case of Shamva district Zimbabwe. Future Food J Food Agric Soc 8(1):20–28

Sekhampu TJ (2012) Socio-economic determinants of household food expenditure in a low-income township in South Africa. Mediterr J Soc Sci 3(3):449–453

Sommet N, Morselli D (2017) Keep calm and learn multilevel logistic modeling: a simplified three-step procedure using stata, R, Mplus, and SPSS. Int Rev Soc Psychol 30(1):203–218

Southern Africa Labour and Development Research Unit (2018) National Income Dynamics Study 2017, Wave 5. Southern Africa Labour and Development Research Unit

Summers R (1959) A note on least squares bias in household expenditure analysis. J Econom Soc 127–126

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46:234–240

Waidler J, Devereux S (2019) Social grants, remittances, and food security: does the source of income matter? Food Secur 11(3):679–702

Yaremych HE, Preacher KJ, Hedeker D (2021) Centering categorical predictors in multilevel models: best practices and interpretation. Psychol Methods

Yatchew A, Sun Y, Deri C (2003) Efficient estimation of semiparametric equivalence scales with evidence from South Africa. J Bus Econ Stat 21(2):247–257

Competing interests

The author declares no competing interests.

Ethical approval

This paper does not contain any studies with human participants performed by the author.

Informed consent

This paper does not contain any studies with human participants performed by the authors.

Additional information