Some adaptations of Multiple Discrete-Continuous Extreme Value (MDCEV) models for a computationally tractable treatment of complementarity and substitution effects, and reduced influence of budget assumptions

David Palma*, Stephane Hess†
Institute for Transport Studies and Choice Modelling Centre
University of Leeds

21st April 2020

Abstract

Many decisions can be represented as interrelated discrete and continuous choices, i.e. what and how much to choose from a set of finite alternatives (incidence and quantity of consumption). In the last twenty years, several models of Karush-Kuhn-Tucker demand systems have been developed and used to study these kinds of decisions. While strongly grounded in economic theory, most of these models have two limitations: they require specifying a budget, and usually omit any complementarity effects. In this paper, we propose an extension to the Multiple Discrete Continuous Extreme Value (MDCEV) model that requires no budget definition and incorporates complementarity and substitution effects. The extension is based on the hypothesis that total expenditure on the alternatives under consideration is small compared to the overall budget. This allows us to use a linear utility function for the numeraire good, leading to a likelihood function without the budget or numeraire good in it. The lack of a budget is specially useful when forecasting, as it avoids cascading errors due to an inaccurate budget specifications. The inclusion of complementarity and substitution effects enriches the interpretability of the model, while its functional form avoids theoretical issues present in previous formulations. Alongside the model derivation, we discuss its main properties, and propose a forecasting algorithm for it. We also report four applications of the model to datasets about time use, household expenditure, supermarket scanner data, and trip generation. A computational implementation of the model is available online.

Keywords: multiple discrete continuous; MDCEV; budgetless; Kuhn-Tucker demand; complementarity; substitution

*D.Palma@leeds.ac.uk
†S.Hess@leeds.ac.uk

1 Introduction

Many choices can be represented as multiple discrete continuous decisions. In these, a decision maker faces a finite set of alternatives, and must choose how much to "consume" of each one, potentially consuming none, one or multiple alternatives. Examples of these situation include activities performed during a day, grocery shopping, investment allocation, etc. Traditional choice models are not well suited for these situations, as they only allow the choice of a single alternative. Continuous models, on the other hand, often underestimate the probability of zero consumption for individual alternatives, also known as the "corner solution". Joint models, where the continuous choice is conditional on the discrete one, usually lack a strong grounding in economic theory, though there are exceptions (Hausman et al., 1995).

The Karush-Kuhn-Tucker multiple discrete continuous (MDC) consumer demand models (Bhat, 2008, 2018; Chintagunta, 1993; Hanemann, 1978; Kim et al., 2002; Mehta and Ma, 2012; Phaneuf and Herriges, 1999; Song and Chintagunta, 2007; Wales and Woodland, 1983) attend to the issues mentioned in the previous paragraph. These models begin by explicitly formulating the consumer utility maximisation problem, assuming either a direct or indirect utility function with associated randomness. Then the optimal solution is derived through the use of Karush-Kuhn-Tucker conditions. Finally, the likelihood function of these conditions is written given the distributional assumptions on the utility function. Nowadays, one of the most popular models of this category is the Multiple Discrete Continuous Extreme Value (MDCEV) model (Bhat, 2008). It has been applied in different areas, such as transport (Jäggi et al., 2012), time use (Enam et al., 2018), social interactions (Calastri et al., 2017), alcohol purchase (Lu et al., 2017), energy consumption (Jeong et al., 2011), investment decisions (Lim and Kim, 2015), household expenditure data (Ferdous et al., 2010), price promotions (Richards et al., 2012), and tourism (Pellegrini et al., 2017).

However, MDC models are not without challenges of their own. The most immediate one is that they require an explicit budget to be estimated. While determining the budget can be easy in time use applications, it can be challenging in other scenarios. For example, in purchase decisions, the budget will rarely be an individual's full income, as there is likely mental accounting and recurring expenses to account for, all of which are not observable. Investment decisions face a similar problem, as the total budget may expand or shrink as a function of expected performance of the investment alternatives. There are other scenarios where even the simple definition of a budget is problematic, for example when modelling the number of recreational trips during a year, or the number of activities performed by an individual during a week. The problem becomes more acute in forecasting. Any predictions from a model require a budget, and predicting the budget, e.g. the income of individuals in the future, is another problem in itself, and introduces cascading errors in the forecast values. A common solution in past work has been to use the total expenditure as the budget, but this is also problematic as it prevents total expenditure from growing or decreasing given changes in the attributes of alternatives.

The budget issue was approached by Pinjari et al. (2016) and Dumont et al. (2013). Pinjari et al. (2016) use a two-stage approach. In the first stage, they use either a stochastic frontier or a log-linear regression to estimate the expected budget, and in the second stage they use the expected budget in an MDCEV model. They compare the performance of both approaches against arbitrarily determined budgets. When using the stochastic frontier method, they assume the budget to be an unobservable characteristic of decision makers, defined as the maximum amount they are willing to spend. This implies that the expected budget under this approach tends to be bigger than the total expenditure. The log-linear regression, on the other hand, attempts to predict total expenditure, so it leads to expected budgets that are of the same magnitude as the total expenditure. While both approaches offer similar performance, and both outperform the arbitrarily determined budget, the stochastic frontier approach leads to bigger expected budgets, therefore allowing for more variability in the forecast, as the total expenditure has room to grow if the attributes of the alternatives improve.

Dumont et al. (2013) propose a different two-step approach to estimate the budget. In the first step, they propose estimating a Structural Equation Model (SEM) where the budget is a latent variable, whose structural equation has socio-demographics as explanatory variables. The budget can have several indicators, such as average expenditure in the category during the last three months, expected expenditure in the future, and ownership of goods from the same category. Income is also considered a latent variable, with at least stated income as indicator. More formally, the latent budget B_n and latent income I_n relate as follows :

$$B_n = Z_n \zeta_z + \zeta_I I_n + \eta_n \tag{1}$$

$$I_n = \xi_n \tag{2}$$

$$y_{nj} = \lambda_j B_n + \sigma_j \varepsilon_{nj} \tag{3}$$

$$S_n = \lambda_s I_n + \sigma_s \varepsilon_{ns} \tag{4}$$

where Z_n are socio-demographics of individual n, y_{nj} is indicator j of the budget, S_n is the stated income, $\eta_n, \xi_n, \varepsilon_{nj}$ and ε_{ns} are standard normal error terms, and $\zeta_z, \zeta_I, \lambda_j, \sigma_j, \lambda_s$ and σ_s are parameters to be estimated. As expected, authors report lower log-likelihoods when using the SEM approximation to the budget than when using maximum expenditure, but they do note an improvement in the MDC parameters significance levels. They do not report changes in forecast performance, making it difficult to evaluate the performance of the proposed approach.

Bhat (2018) uses a model with a linear utility function for the outside good, just as we do in this paper, effectively removing the requirement of defining a budget. In this paper, we explore the consequences of this functional form, and offer a theoretical justification for it. Furthermore, we explicitly account for complementarity and substitution effects, allowing for cross-price elasticities of demand to arise naturally from the formulation.

A second limitation of many Kuhn-Tucker demand model formulations is that they omit direct complementarity and substitution effects (Bhat, 2008, 2018; Chintagunta, 1993; Hanemann, 1978; Kim et al., 2002; Phaneuf and Herriges, 1999; Wales and Woodland, 1983). Substitution and complementarity define relationships between the demand for pairs of products. If the demand for one of them increases, then the demand for the other is reduced in the case of substitution and increased in the case of complementarity (Hicks and Allen, 1934). While the budget constraint naturally induces substitution between products, this is only an indirect effect. In fact, substitution due to changes of the budget are more commonly associated with what Manchanda et al. (1999) call a *coincidence* effect, i.e. a shock that is common to all available alternatives. Complementarity effects, on the other hand, are completely excluded from many formulations such as the MDCEV model. Other models, however, have considered these effects in multiple ways, usually leading to different shortcomings.

Bhat et al. (2015) present an MDC model with non-additive utility functions, which allow for complementarity and substitution effects. This is essentially an MDCEV model with an additional term in the utility function interacting consumptions of pairs of alternatives. This interaction is weighted by a θ_{ij} parameter whose value determines the nature of the interaction. A positive value of θ_{ij} implies complementarity between alternatives i and j, a negative one substitution, and $\theta_{ij} = 0$ implies no complementarity or substitution. The formulation by Bhat et al. (2015) has three main drawbacks. The first is that the utility function is valid only for some values of θ_{ij}, as other values would lead to a negative marginal utility of consumption. While this is not necessarily problematic, as it is easy to bound the maximum values of a parameter, the issue lies in that condition involving the level of consumption. In particular, the condition is as follows:

$$\frac{\partial U}{\partial x_k} = \psi_k + \sum_{l \neq k} \theta_{kl} \gamma_l \log \left(\frac{x_k}{\gamma_k} + 1 \right) > 0 \; \forall k, l, \tag{5}$$

where l and k enumerate alternatives, U is the direct utility function, x is the level of consumption, and $\psi_k, \theta_{kl}, \gamma_l$ and γ_k are parameters to be estimated. As the logarithm is not a bounded function, whether or not this condition is satisfied will depend on the level of consumption x of each individual, making it impossible to assess the correctness of a model without associating it to a particular dataset. This hinders model transferability from one dataset to another, and jeopardises forecasting, as only scenarios that fulfil the condition above should be permissible forecasts.

The second issue with the solution proposed by Bhat et al. (2015) is that the stochasticity is introduced midway through the derivation of the model in the Karush-Kuhn-Tacker conditions, and not in the initial formulation of the model. While this is merely a formal issue, it does imply that the origin of the randomness is not clear, and it is not possible to easily associate it with unobserved variables or measurement errors, as would be the case in more traditional econometric models. The third issue is that γ parameters have a role both in satiation and in the interaction term (i.e. complementarity and substitution) of the utility, making their interpretation difficult.

Pellegrini et al. (2019) refine the model proposed in Bhat et al. (2015) by proposing a different interaction term in the utility function. While this new formulation leads to an improved fit and provides a clear interpretation of γ parameters, it retains at least the first issue associated to the formulation of Bhat et al. (2015). Song and Chintagunta (2007) propose a different model considering complementarity and substitution effects, but their formulation restricts these to be symmetrical, i.e. the model should contain as much complementarity as it does substitution. More formally, the parameters measuring complementarity and substitution must add up to zero. There are no theoretical reasons for this to necessarily be the case in any given application. Finally, Mehta and Ma (2012) propose a model with a similar formulation of complementarity and substitution effects to the one proposed by Song and Chintagunta (2007), but without the symmetry constraint. Nonetheless the applicability of their model, just as in Bhat et al. (2015), is linked to the levels of consumption in each particular data set. Therefore, transferability of the model to new forecasting scenarios cannot be ensured.

In this paper, we propose an extension to the MDCEV model that does not require the definition of a budget, and includes explicit complementarity and substitution effects. Our approach is a suitable approximation of a full MDC model for situations where the expenditure on all alternatives that are included in the model is small compared to the overall budget, which allows us to drop the budget from the model likelihood. Also, the proposed formulation for the complementarity and substitution part of the utility leads to a utility function whose validity does not depend on the particular dataset under analysis, incorporates stochasticity in a natural way, and is associated with a clear interpretation of parameters.

The remainder of this document is structured as follows. The next section introduces the model formulation, derivation and likelihood function, and closes with an efficient forecasting algorithm for it. Section 3 discusses the identification of the model parameters, some constraints that theory and estimation imposes on them, and compares the model forecast performance to that of a similar model with a budget constraint in a situation where we have an explicit budget. Section 4 presents applications of the proposed model formulation to four different datasets, dealing with time use, household expenditure, supermarket scanner data, and number of trips, respectively. The paper closes with a brief summary of the proposed model formulation capabilities and limitations.

2 Incorporating complementarity, substitution and an implicit budget into the MDCEV model

2.1 Model formulation

Consider the classical (consumer) utility maximisation problem, where an individual n must decide what products k to consume from a set of alternatives, by maximising his or her utility subject to a budget constraint (Eqn. 6).

$$
\begin{aligned}
Max_{x_n} \qquad & u_0(x_{n0}) + \sum_{k=1}^{K} u_k(x_{nk}) + \sum_{k=1}^{K-1} \sum_{l=k+1}^{K} u_{kl}(x_{nk}, x_{nl}) \\
s.t. \qquad & x_{n0}p_{n0} + \sum_{k=1}^{K} x_{nk}p_{nk} = B_n
\end{aligned}
\tag{6}
$$

where $n = 1...N$ indexes individuals and $k = 1...K$ alternatives, $x_n = [x_{n0}, x_{n1}, ..., x_{nK}]$ is a vector grouping the consumed amount of each alternative (product), p_{nk} is the price of alternative k faced by individual n, and B_n is the total budget available to individual n. x_{n0} is an *outside* or *numeraire* good, i.e. a good that aggregates all consumption outside of the category of interest. For example, if the researcher is interested in modelling demand for food, $x_{n1}, ..., x_{nK}$ would represent expenditure in different food categories (the *inside* goods), while x_{n0} would represent the aggregate expenditure in housing, transport, leisure, etc. It is usually assumed that $p_{n0} = 1$, so that x_{n0} becomes the total expenditure on categories other than the one of interest, where $x_{n0} > 0$.

We assume the following functional forms for the different parts of the utility function.

$$
u_0(x_{n0}) = \psi_{n0} x_{n0} \tag{7}
$$

$$
u_k(x_{nk}) = \psi_{nk}\gamma_k \log\left(\frac{x_{nk}}{\gamma_k} + 1\right) \tag{8}
$$

$$
u_{kl}(x_{nk}, x_{nl}) = \delta_{kl}\left(1 - e^{-\delta_k x_{nk}}\right)\left(1 - e^{-\delta_l x_{nl}}\right) \tag{9}
$$

We assume a linear utility function for the outside good (Eqn. 7), as this will later on allow us to drop both the outside good consumption x_{n0} and the budget B from the final model formulation. A utility function that is linear in the consumed amount of goods contradicts classical economic theory, as it does not lead to positive and diminishing marginal utility of consumption. However, this should be considered as an appropriate approximation for the case where most of the budget is spent on the outside good, and only a relatively small amount is spent on the inside goods. In such a case, changes in the total expenditure of inside goods would lead to a relatively small change in the consumed amount for the outside good, and therefore a negligible change in the marginal utility of it. This means that our model is appropriate for cases where the total expenditure in the inside goods is small compared to the total budget, something we corroborate using simulation in Section 3.3. Furthermore, the model we propose allows for parametrisation of the (constant) marginal utility of the outside good ψ_{n0}, permitting differences across individuals in the sample, based for example on observed socio-demographic characteristics.

We take the definition of u_k from Bhat (2008). In this formulation, ψ_{nk} represents alternative k's *base utility*, i.e. its marginal utility at zero consumption. This parameter could be interpreted as the scale of the utility of product k. The γ_k parameters, on the other hand, relate mainly to consumption satiation, by altering the curvature of alternative k's utility function. In general, a higher γ_k indicates higher consumption of alternative k, when consumed. While a common interpretation is that ψ_{nk} and γ_k determine what and how much of alternative k to consume, respectively, this is not completely true. There is a level of interaction between these parameters, and in some circumstances a low value of ψ_{nk} can be compensated by a high value of γ_k (Bhat, 2008, 2018).

Parameters ψ_{nk} must always be positive, as they represent the marginal utility of alternatives at the point of zero consumption. We ensure this using the following definition.

$$\begin{aligned} \psi_{n0} &= e^{\alpha z_{n0}} \\ \psi_{nk} &= e^{\beta_k z_{nk} + \varepsilon_{nk}} \end{aligned} \tag{10}$$

where z_{n0} is a row vector of characteristics of the decision maker that are expected to correlate with that individual's marginal utility of the outside good (e.g. socio-demographics); α is a column vector of parameters representing the weights of those characteristics on the marginal utility of the outside good; z_{nk} are attributes of alternative k; β_k are vectors of parameters representing weights of those attributes on the alternative's base utility; and ε_{nk} is a random disturbance term. We only include random disturbances in the base utility of the inside goods, as this leads to a closed-form likelihood function. We discuss the inclusion of a random disturbance in the marginal utility of the outside good in Section 3.4.

The final component of the utility function, $u_{kl}(x_{nk}, x_{nl})$, captures the complementarity and substitution effects between inside goods. Figure 1 presents the behaviour of this component for a set of δ_{kl}, δ_k, and $delta_l$ parameters, and different values of x_{nk} and x_{nl}, which are assumed to be equal. If $\delta_{kl} > 0$, there is complementarity between alternatives k and l, as this component will increase the overall utility. If $\delta_{kl} < 0$, there is a substitution effect between alternatives k and l, as u_{kl} becomes more negative as x_{nk} and x_{nl} increase. If $\delta_{kl} = 0$, the consumption of both alternatives is independent of each other. While the δ_{kl} parameter controls the sign and magnitude of the complementarity or substitution effect, parameters δ_k and δ_l determine the curvature of the effect. However, as we discuss in Section 3.1, these parameters are difficult to identify, and we recommend setting them to fixed values in model estimation. The value of u_{kl} is bounded to the interval $[0, \delta_{kl})$, ensuring transferability of estimated models to other datasets, a point we discuss in Section 3.2.

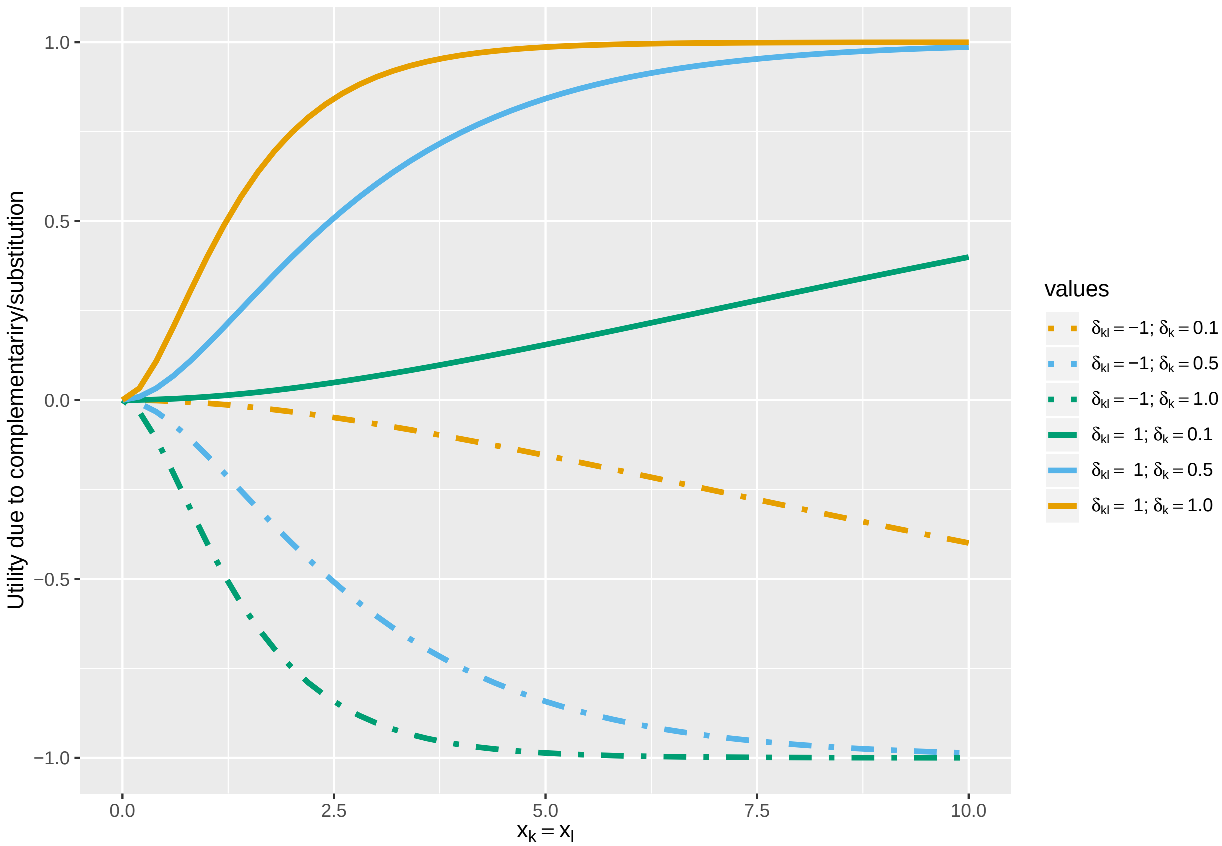

Figure 1: Complementarity/substitution component of the utility.

2.2 Model derivation

To solve the optimisation problem, we begin by writing its Lagrangian (Eqn. 11) and Karush-Kuhn-Tacker conditions of optimality (eqns. 12 and 13).

$$Lagr(x_n) = u_0(x_{n0}) + \sum_{k=1}^{K} u_k(x_{nk}) + \sum_{k=1}^{K-1} \sum_{l=k+1}^{K} u_{kl}(x_{nk}, x_{nl}) - \lambda \left(x_{n0}p_{n0} + \sum_{k=1}^{K} x_{nk}p_{nk} - B_n \right) \quad (11)$$

$$\frac{\partial Lagr}{\partial x_{n0}} = 0 \quad : \quad \psi_{n0} = \lambda p_0 \tag{12}$$

$$\frac{\partial Lagr}{\partial x_{nk}} = 0 \quad : \quad \frac{\psi_{nk}}{\frac{x_{nk}}{\gamma_k}+1} + \delta_k e^{-\delta_k x_{nk}} \sum_{l \neq k} \delta_{kl} \left(1 - e^{-\delta_l x_{nl}}\right) \leq \lambda p_{nk} \tag{13}$$

Eqn. 13 will be an equality when alternative k is consumed (i.e. $x^*_{nk} > 0$, with x^*_{nk} the consumption at the optimum, i.e. the observed consumption). Eqn. 13 will be an inequality when $x^*_{nk} = 0$. In other words, the marginal utility of any consumed product k at the optimum level of consumption will be λ scaled by the alternative's price p_{nk}. Instead, if the product is not consumed, its marginal utility will be lower. By combining eqns. 12 and 13, we obtain:

$$\frac{\psi_{nk}}{\frac{x_{nk}}{\gamma_k}+1} + \delta_k e^{-\delta_k x_{nk}} \sum_{l \neq k} \delta_{kl} \left(1 - e^{-\delta_l x_{nl}}\right) \leq \psi_{n0} \frac{p_{nk}}{p_{n0}} \tag{14}$$

Replacing ψ_{n0} and ψ_{nk} by their definitions (Eqn. 10), and isolating the random component ε_{nk}, we obtain

$$\begin{aligned} \varepsilon_{nk} &\leq -W_{nk} \qquad (15) \\ W_{nk} &= z_{nk}\beta_k - \log\left(\frac{x_{nk}}{\gamma_k}+1\right) - \log\left(\psi_{n0}\frac{p_{nk}}{p_{n0}} - \delta_k e^{-\delta_k x_{nk}} \sum_{l \neq k} \delta_{kl}\left(1 - e^{-\delta_l x_{nl}}\right)\right) \end{aligned}$$

Now we only need a distributional assumption on ε_{nk}, and to apply the Change of Variable Theorem from ε_{nk} to x_{nk} (only over the consumed alternatives) to obtain the likelihood function of the model. We assume ε_{nk} to be independent of each other, and to be distributed identically following a Gumbel distribution with mean zero and a scale $sigma$ to be estimated. Then, if f and F are the density and cumulative distribution functions of ε_{nk}, respectively, we can write the likelihood function as follows:

$$Like(x_{nk}) = |J| \prod_{k=1}^{M_n} f(-W_{nk}) \prod_{k=M+1}^{K} F(-W_{nk}) \tag{16}$$

$$
\begin{aligned}
J_{ii} &= \frac{1}{x_i+\gamma_i} + \frac{\delta_i E_i}{\psi_0 \frac{p_k}{p_0} - E_i} \\
J_{ij} &= \frac{-\delta_{ij}\delta_i\delta_j e^{-\delta_i x_i} e^{-\delta_j x_j}}{\psi_0 \frac{p_k}{p_0} - E_i} \\
E_i &= \delta_i e^{-\delta_i x_i} \sum_{l\neq i} \delta_{il}(1-e^{-\delta_l x_l})
\end{aligned} \tag{17}
$$

In this set of equations, $|J|$ is the value of the determinant of the Jacobian J of W_{nk}, whose elements are defined in Eqn. 17 (index n has been dropped for clarity, i indexes rows, and j columns). No obvious compact form exists for this determinant. M_n is the number of products consumed by individual n, and in Eqn. 16 we have re-ordered the alternatives so that the consumed alternatives hold the indexes $k = 1...M_m$, and the non-consumed alternatives hold indexes $k = (M_n+1)...K$. If no alternative is consumed, both the Jacobian and the first product of Eqn. 16 drop out, and only the second product remains (this time for $k = 1...K$).

Replacing the Gumbel density and cumulative distribution functions $f(x) = e^{-e^{-\sigma x}}$ and $F(x) = \sigma e^{-x+e^{-x}}$ into equation 16, we obtain Eqn. 18.

$$
Like(x_{nk}) = |J| \frac{1}{\sigma^M} \frac{\prod_{k=1}^{M_n} e^{\frac{W_{nk}}{\sigma}}}{\prod_{k=1}^{K} e^{-e^{\frac{W_{nk}}{\sigma}}}} \tag{18}
$$

2.3 Forecasting

Once the model has been estimated, forecasting requires solving the original maximisation problem proposed in Eqn. 6 several times, each time using different draws of ε_{nk} from a Gumbel(0,σ) distribution, and then averaging the result across these draws.

To solve the optimisation problem we once again use the Lagrangian in Eqn. 11 and the KKT conditions in eqns. 12 and 13, leading us to Eqn. 14. Assuming an equality and isolating x_{nk}, we obtain

$$
x_{nk} = h(x_{nk}) = \gamma_k \left(\frac{\psi_{nk}}{\psi_{n0}\frac{p_{nk}}{p_{n0}} - E_{nk}} - 1 \right) \tag{19}
$$

where the definition of E_{nk} can be found in Eqn. 17, and where it depends on the value of all x_n. Eqn. 19 is a fixed point problem, i.e. a problem of the form $x = h(x)$. According to the Existence and Uniqueness theorem, as the right part of Eqn. 19 is continuous in x_n over the closed interval $[0, \frac{B_n}{p_{nk}}]$, at least one solution to the problem exists. However, we cannot ensure that the solution is unique. We solve Eqn. 19 through the following iterative approach:

1. Set $r = 0$ and $x_n^{(r)} = [x_{n1}^{(r)}, ..., x_{nK}^{(r)}]$ to zero.
2. For each $k \in \{1, 2, ..., K\}$
   2.1. Set $s = 0$ and calculate $E_{nk}^{(r)}$.
   2.2. Set $x_{nk}^{(r)(s)}$ to a random starting value.
   2.3. Make $x_{nk}^{(r)(s+1)} = h(x_{nk}^{(r)(s)})$.
   2.4. If $|x_{nk}^{(r)(s+1)} - x_{nk}^{(r)(s)}| > \tau$ and $s < S$, go to step 2.3.
   2.5. If $x_{nk}^{(r)} < 0$ or $|\frac{\partial U_n}{\partial x_{nk}} - \frac{\partial U_n}{\partial x_{n0}}| > \tau$, or $|x_{nk}^{(r)(s+1)} - x_{nk}^{(r)(s)}| > \tau$ make $x_{nk}^{(r)} = 0$, otherwise make $x_{nk}^{(r)} = x_{nk}^{(r)(s+1)}$
3. If $|x_n^{(r)} - x_n^{(r)}| > \tau$ and $r < S$ go to 2.

where S is the maximum number of iterations allowed, and τ indicates the convergence tolerance parameter, which can be set to the desired precision. This procedure must be performed multiple times for each individual n, each time with a different set of draws for the ε_{nk} disturbances. Then results for each set of draws must be averaged. As x_{nk} is not bounded, big value of ε_{nk} would lead to a unreasonably big value of x_{nk}. Therefore, we recommend truncating the draws of ε_{nk} to the 99.5% quantile.

3 Model properties

In this section, we discuss some of the most relevant properties of the model, namely the identifiability and some theoretical constraints on its parameters, its performance when compared to an equivalent model with an explicit budget, and a way to introduce additional unobserved variability (mixing).

3.1 Identification of parameters

When estimating the proposed model, the modeller should consider the following three points regarding identifiability of parameters.

First, individuals who do not consume any inside good should **not** be excluded from the sample. Even though these observations do not provide any information on the value of ψ_{nk}, they do provide information of the value of ψ_{n0} in relation to the inside goods.

Second, there should be no constant (intercept) in the definition of ψ_{n0}, i.e. z_{n0} should not contain an element equal to 1 for every individual. As utility does not have any meaningful units, we require setting a base against which all other utilities are measured. To do this, we recommend setting the intercept of the outside good to zero. Any variable that changes across individuals can be included in z_{n0}, even if they are not centred around zero. We recommend populating z_{n0} with characteristics of individual n, such as socio-demographics, and in particular, individual n's income. Including income in this way does not imply that the budget is equal to the income, but only that the marginal utility of the outside good depends on it. We would expect a negative coefficient for income if included in ψ_0, as an increase of income usually leads to increased overall consumption, and therefore a smaller marginal utility of the outside good. In general, a negative coefficient in α indicates that an increase in the corresponding explanatory variable leads to increased consumption. The opposite is true for a positive coefficient.

Third, while it is theoretically possible to identify parameters δ_{kl}, δ_k and δ_l, in practice this is very difficult. We recommend two simplifications: set $\delta_k = \delta_0$, $\forall k \in \{1, ..., K\}$, and fix $\delta_0 = -\frac{\log\left(1-\sqrt{p}\right)}{q}$, where p is a percentage (we recommend $p = 0.95$), and q is the quantile associated with p in the consumption vector of all inside goods in the whole dataset, excluding zeros. The first simplification is mostly a theoretical one. Different values of δ_k are difficult to interpret, as they imply that complementarity and substitution effects are dominated by attributes of individual alternatives, as opposed to the relation between pairs of alternatives (already captured by the $\delta_k l$ parameters).

The recommended normalisation of δ_0 ensures that when $x_{nk} = x_{nl} = q$, we have that $u_{kl} = p\delta_{kl}$ (see Figure 2). In other words, we fix the curvature of u_{kl} to make sure it discriminates between at least $100p\%$ of the sample. For the top $100(1-p)\%$ of observed consumption, the complementarity or substitution effect will be very close to δ_{kl}, regardless of how big the consumption is. The modeller can adjust the value of p depending on the desired level of discrimination for high consumptions. Values of p closer to 1 lead to increased discrimination for high consumption levels, but in the presence of outliers, could lead to decreased discrimination for low levels of consumption and generate numerical instability during model estimation.

The value of δ_0, and more generally δ_k, are difficult to estimate because as they get smaller, they become harder to identify. As discussed in the previous paragraph, smaller values of δ_k will lead to better differentiation of the complementarity and substitution effects. Therefore, the estimation process of parameters will naturally tend towards small values of δ_k. However, u_{kl} behaves as a linear function of the product of consumptions $x_{nk}x_{nl}$ for small values of δ_k, as replacing $e^{-\delta_k x_{nk}}$ by its Taylor series expansion shows in Eqn. 20.

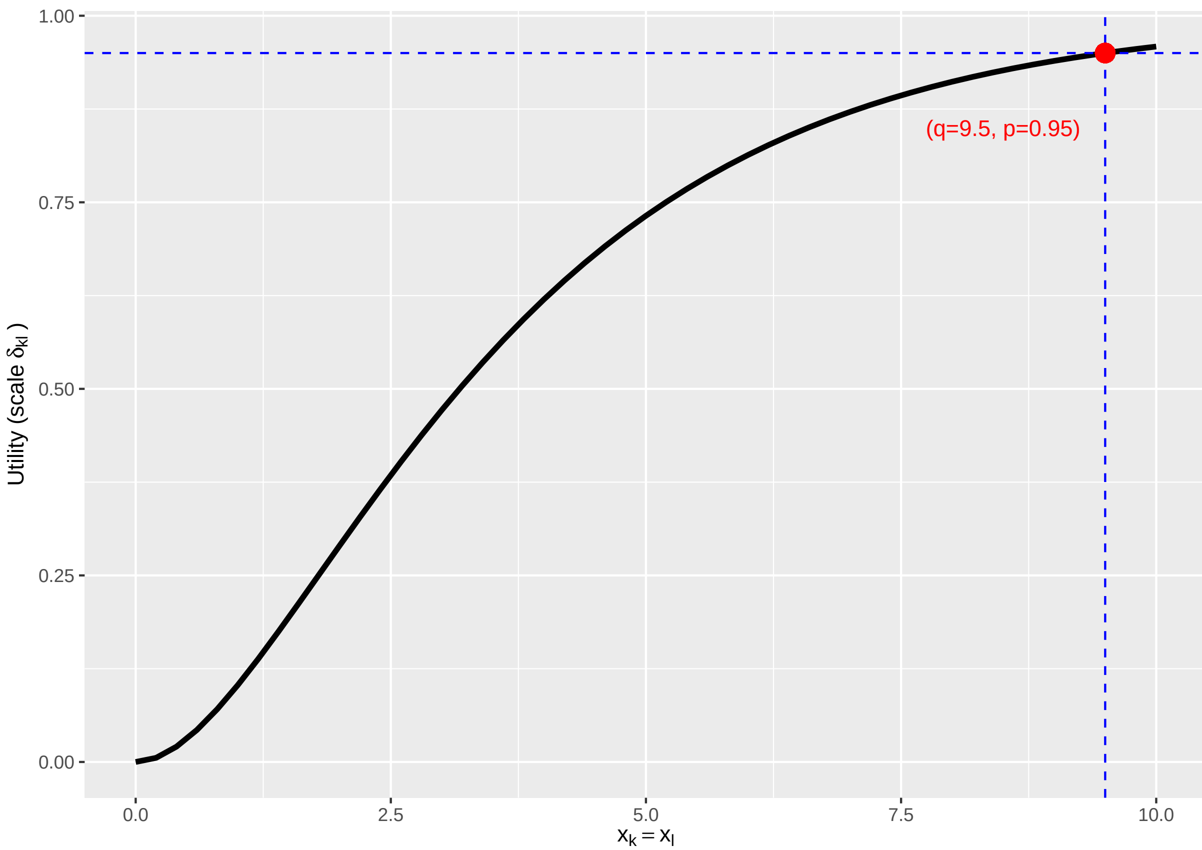

Figure 2: Complementarity/substitution component of the utility.

$$
\begin{aligned}
u_{kl}(x_{nk}) &= \delta_{kl}\left(1 - e^{-\delta_k x_{nk}}\right)\left(1 - e^{-\delta_l x_{nl}}\right) \qquad (20)\\
&= \delta_{kl}\left(1 - \sum_{r=0}^{\infty}\frac{(\delta_k x_{nk})^r}{r!}\right)\left(1 - \sum_{r=0}^{\infty}\frac{(\delta_l x_{nl})^r}{r!}\right)\\
&= \delta_{kl}\left(1 - \left(1 - \delta_k x_{nk} + \frac{(-\delta_k x_{nk})^2}{2} + ...\right)\right)\left(1 - \left(1 - \delta_l x_{nl} + \frac{(-\delta_l x_{nl})^2}{2} + ...\right)\right)\\
&= \delta_{kl}\left(\delta_k x_{nk}\right)\left(\delta_l x_{nl}\right)\\
&= \delta_{kl}\delta_k\delta_l\left(x_{nk}x_{nl}\right)
\end{aligned}
$$

As δ_k and δ_l are small, δ_k^2 and δ_l^2 are very close to zero, allowing us to truncate the Taylor series at the second term. As the final expression in Eqn. 20 shows, u_{kl} behaves as a linear function with a slope equal to $\delta_{kl}\delta_k\delta_l$. As the slope is formed by the product of multiple parameters, these cannot be independently identified.

3.2 Constraints on estimated parameters

Both economic theory and the solution of the optimisation problem in Eqn. 6 require all marginal utilities to be positive and decreasing. In the case of the proposed model, this implies the inequality in Eqn. 21:

$$\frac{\psi_{nk}}{\frac{x_{nk}}{\gamma_k}+1} + \delta_k e^{-\delta_k x_{nk}} \sum_{l\neq k} \delta_{kl}\left(1-e^{-\delta_l x_{nl}}\right) > 0 \qquad (21)$$

Additionally, the second logarithm in Eqn. 15 requires that

$$\psi_{n0}\frac{p_{nk}}{p_{n0}} - \delta_k e^{-\delta_k x_{nk}} \sum_{l\neq k} \delta_{kl}\left(1-e^{-\delta_l x_{nl}}\right) > 0 \qquad (22)$$

These conditions are functions of x_{nk}, making their fulfillment dependent on the particular dataset at hand. We would like to instead derive dataset-independent conditions. This is possible by noting that the impact of x_{nk} in both conditions is bounded by its exponential transformation to the interval $0 \leq e^{-\delta_k x_{nk}} \leq 1$ (because $\delta_k \geq 0$ and $x_{nk} \geq 0$). This allows us to derive more general conditions than Eqns. 21 and 22 by simply analysing the extreme cases $x_{nk} = 0$ and $x_{nk} = \infty$, as the value of the conditions for all other x_{nk} values will fall between these. These extreme cases have the benefit of removing x_{nk} from the conditions. Table 1 display the values of the constraints in Eqns. 21 and 22 for different extreme values of x_{nk}.

Table 1: Constraints on proposed model parameters for extreme levels of consumption

x_k	$x_{l:\delta_{kl}>0}$	$x_{l:\delta_{kl}<0}$	Eqn. 21	Eqn. 22
0	0	0	$\psi_k + \delta_k > 0$	$\psi_0 \frac{p_k}{p_0} > 0$
0	0	∞	$\psi_k - \delta_k \Delta^- > 0$	$\psi_0 \frac{p_k}{p_0} + \delta_k \Delta^- > 0$
0	∞	0	$\psi_k + \delta_k \Delta^+ > 0$	$\psi_0 \frac{p_k}{p_0} - \delta_k \Delta^+ > 0$
0	∞	∞	$\psi_k - \delta_k \Delta^- + \delta_k \Delta^+ > 0$	$\psi_0 \frac{p_k}{p_0} + \delta_k \Delta^- - \delta_k \Delta^+ > 0$
∞	any	any	$0^+ > 0$	$\psi_0 \frac{p_k}{p_0} > 0$
Where:			$\Delta^- = \sum_{l:\delta_{kl}<0} \lvert\delta_{kl}\rvert$	$\Delta^+ = \sum_{l:\delta_{kl}>0} \lvert\delta_{kl}\rvert$

Most of the conditions in Table 1 are always fulfilled, because ϕ_k, δ_k, γ_k, p_k, p_0, Δ^- and Δ^+ are all equal or bigger than zero. Eqn. 21 for $x_k = \infty$ will also always be true as zero is approached from the right (i.e. from positive values). The critical constraints from Table 1 are summarised in Eqn. 23

$$-\frac{\psi_{nk}}{\delta_k} < \sum_{l:\,\delta_{kl}<0} \delta_{kl} < \sum_{l:\,\delta_{kl}>0} \delta_{kl} < \frac{\psi_{n0} p_{nk}}{\delta_k p_{n0}} \ \forall k \tag{23}$$

Conditions in Eqn. 23 are based on extreme cases, so they represent sufficient but not necessary conditions for the validity of the parameters. In other words, estimated parameters need only to comply with Eqns. 21 and 22, but satisfying Eqn. 23 guarantees that those conditions are met, as Eqn. 23 is more stringent than Eqns. 21 and 22.

If individuals in the dataset behave rationally and in accordance with economic theory, then the estimated parameters should naturally comply with Eqn. 23. At the time of writing, we have not experienced any issues of running into inconsistent parameters, nor have we had to impose parameter constraints during estimation to enforce compliance with Eqn. 23.

3.3 Suitability of the linear utility of the outside good

We propose the linear utility for the outside good as an approximation of the case where expenditure on the inside goods (i.e. considered alternatives) is small compared to that on the outside good. In these cases, we expect only very small changes to the marginal utility of the outside good due to changes in the consumption of the inside goods. For example, consider consumption of the yoghurt product category. The expenditure on yoghurt will be small compared to the total expenditure on food, and even smaller compared to the entire disposable income of the household. By using the proposed model, the modeller does not need to determine what the correct budget is, but only needs to know that total expenditure in the category of interest is small compared to the budget, whatever that may be.

If our interpretation of the proposed model is appropriate, then its forecast performance should be similar to that of an equivalent model with explicit budget when the expenditure on the outside good is large compared to that on the inside goods. We tested this assumption through simulation. We first created 80 different datasets of 500 observations each, assuming an MDCEV data generation process (Bhat, 2008). The MDCEV model is very similar to the one proposed in this paper, but with an explicit (and enforced) budget, and without the explicit complementarity and substitution component of the utility. Each dataset had three inside goods (always available) plus an outside good. We generated them assuming β_k (the base utility intercepts) and γ_k (the satiation parameters) to follow uniform distributions across samples between -4.5 and 1.5, and between 0 and 10, respectively. We defined the budget as $B_n = 50 + \eta_n z_n$, where η_n followed a standard normal distribution across samples, and z_n followed a standard uniform within each dataset. We estimated an MDCEV and a proposed model on each dataset, using the true budget

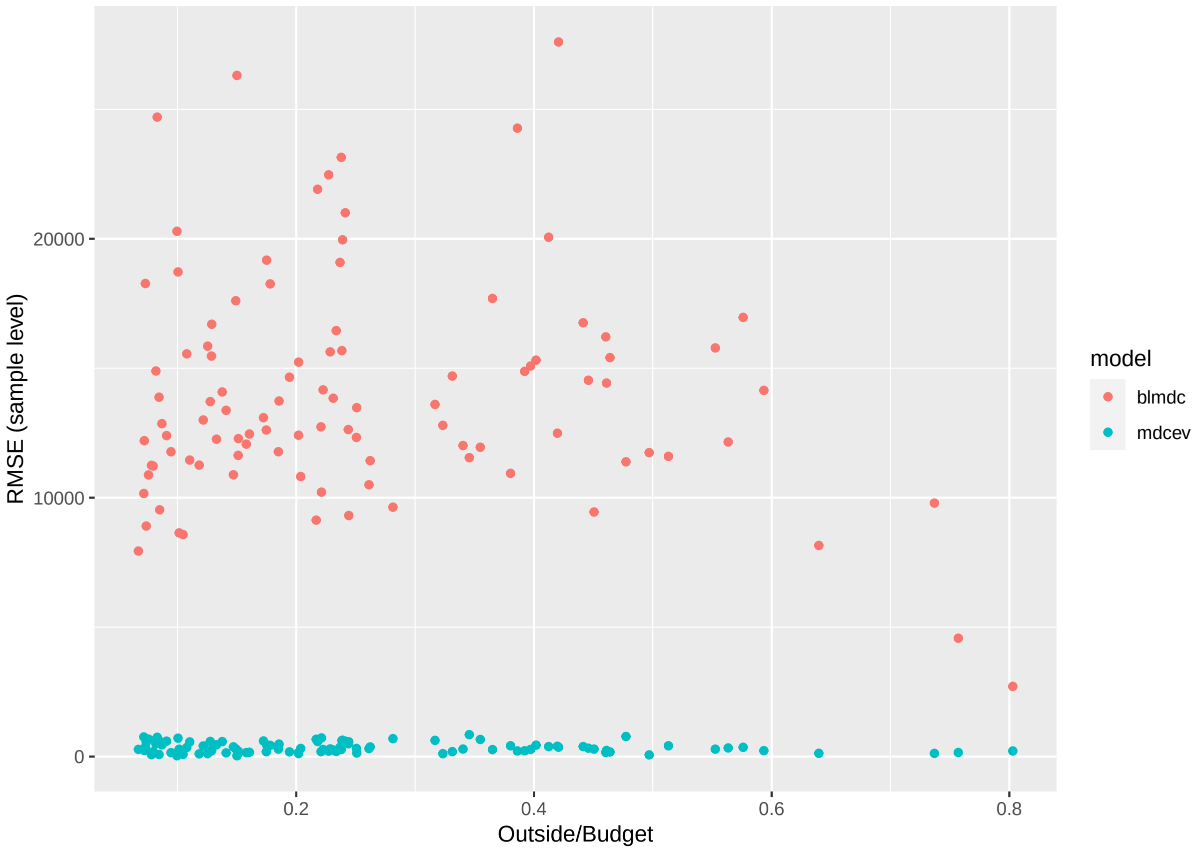

Figure 3: Compared forecast error of the proposed model to an MDCEV model (Bhat, 2008) with explicit budget, on data generated by MDCEV

for the MDCEV. Then we forecast consumption with each of the models for the same data as used in estimation, and calculated the Root Mean Squared Error (RMSE) of the forecast aggregate demand in the whole sample.

As Figure 3 shows, the forecast performance of the proposed model approaches that of an MDCEV with explicit budget as the expenditure on the outside good increases. We observe that the proposed model reaches similar errors to those from the MDCEV model when the outside good captures 40% of the total budget or more. This point of equivalence, however, will depend on the dataset, so no fixed threshold can be proposed for when the linear approximation of the marginal utility of the outside good underlying the proposed model is appropriate.

3.4 Unobserved shocks to consumption

Manchanda et al. (1999) propose that demand models should allow for *coincidence*, complementarity and substitution effects to achieve a realistic representation of demand. These effects should be represented both at a systematic and stochastic level, i.e. due to observed and unobserved factors. In this context, *coincidence* effects, are shocks to demand unrelated to complementarity and substitution effects, and they can influence the demand for alternatives independently or jointly. There can be many causes for coincidence effects. For example, in a retail environment, purchase cycles of two unrelated products could coincide (e.g. beer and baby food), or preferences for a single product could vary, or a common shock could influence demand for all products at the same time (e.g. an increase or reduction in income).

The proposed model captures systematic variations in the complementarity, substitution and coincidence effects, both independently for each alternative, as well as common to all of them. The substitution and complementarity effects are captured by the δ_{nkl} parameters. Independent coincidence effects are captured by systematic variations in the definition of each alternative's base utilities (ψ_{nk}). Common coincidence effects are captured by the definition of the outside good marginal utility (ψ_{n0}). All of these parameters can be defined as functions of an individual's characteristics, allowing for great flexibility.

As presented in Section 2, stochastic (random) effects are only considered in the proposed model for independent coincidence effects. As independent random error terms (ε_{nk}) are only included in the base utility of inside goods, these are alternative-specific. Nevertheless, further variability due to unobserved factors can be included through mixing. This, however, leads to a likelihood function without a closed form.

Including unobserved variability in common coincidence effects requires introducing stochasticity in the definition of ψ_{n0}. This can be achieved by defining $\psi_{n0} = e^{z_{n0}\alpha+\varepsilon_{n0}}$, with ε_{n0} following a cumulative density function $f(\varepsilon_{n0})$ with mean zero and standard deviation σ_0. This change implies that the likelihood function of the model now becomes the one in Eqn. 24.

$$Like(x_{nk}) = \int_{\varepsilon_0=-\infty}^{\infty} |J| \frac{1}{\sigma^M} \frac{\prod_{k=1}^{M_n} e^{\frac{W_{nk}}{\sigma}}}{\prod_{k=1}^{K} e^{-e^{\frac{W_{nk}}{\sigma}}}} f(\varepsilon_0) d\varepsilon_0 \tag{24}$$

This corresponds to a mixed model, with the integral being necessary to obtain the expected value of $\psi_n 0$, as it is now a random component. This integral is unlikely to have a closed form solution, but its value can be approximated using Monte Carlo methods (Train, 2009).

Similarly, stochasticity can be included in the complementarity and substitution effects by defining $\delta_{kl} = \mu_{kl} + \sigma_{kl} * \eta_{kl}$, where μ_{kl} and σ_{kl} are parameters to be estimated representing the mean and standard deviation of the effect, and η_{kl} is a random error component with mean zero (e.g. following a standard normal distribution). Once again, calculation of the likelihood using this set-up would imply the use of simulation.

4 Model application and comparison

In this section we apply the proposed model to four different datasets. The first dataset records time use, where all participants face the same budget (24 hours a day), and all alternatives - in this case, activities - have the same price (one unit of time). As the budget is known, this is not a dataset where the proposed model would traditionally be applied, but it is useful to measure the loss of forecast performance as compared to a model with an explicit budget constraint. The second dataset deals with household expenditure, where budgets vary between different households, but consumption is aggregated to categories, so prices are still unitary (one unit of money). This dataset helps us illustrate how the forecast performance of the proposed model compares to a model with explicit budget when the budget is misspecified, a case particularly relevant in forecasting. The third dataset contains scanner data from a supermarket, where both budgets and prices vary from one observation to the next. This dataset allows us to compare the sensitivity to price of the proposed model against a model with explicit budget. The last dataset summarises the number of trips for different purposes. This dataset is a case where the very definition of a budget is problematic, as there is no evident limit on the number of trips during a day, given that this relates to the length of each trip.

In every case, we use the MDCEV (Bhat, 2008) as the model with explicit budget. This model is equivalent to the proposed model, but with an explicit budget and without the explicit complementarity and substitution component in the utility function. More formally, the MDCEV is the model obtained from Eqn. 6 when $u_0(x_{n0}) = e^{\varepsilon_0} \frac{x_{n0}^{\alpha}}{\alpha}$, $u_{kl} = 0$, and ε_0, ε_{nk} follow independent Gumbel distributions with scale parameter σ.

4.1 Fixed budget and fixed prices: time use dataset

The first dataset records time use of 447 individuals across 2,826 days. Details about the data collection can be found in Calastri et al. (2020), and an application to time use analysis using this data can be found in Calastri et al. (2019). Only out-of-home activities are registered in the dataset, which we aggregate to six plus the outside good, as described in Table 2.

We estimated four different models using the Time Use data. The first, *Traditional MDCEV*, is the only model with explicit budget. *Extended MDCEV 1* is the proposed model with the marginal utility of the outside good fixed to zero. *Extended MDCEV 2* is the proposed model with socio-demographics explaining the marginal utility of the outside good. *Extended MDCEV 3* is also the proposed model, but this time with the consumed amount of the outside good as

Table 2: Main descriptive statistics of the time use database

	Engagement	Total consumption (H)	Average consumption when engaged (H)
Home (outside)	100.00%	51467	18.21
Work	40.30%	8170	7.17
School	3.01%	299	3.52
Shopping	27.71%	1408	1.80
Private	18.93%	1253	2.34
Leisure	41.54%	5227	4.45

explanatory variable for its own marginal utility. In practice, if we observed the consumed amount of the outside good, we would simply use a model with explicit budget, such as the MDCEV. However, *Extended MDCEV 3* is useful as a benchmark of the maximum forecast performance of the proposed model, if an ideal explanatory variable z_0 were available. While *Extended MDCEV 1* represents a baseline for the proposed model, *Extended MDCEV 2* represents what would be a classical use of this model, with socio-demographics in z_0. On the other hand, *Extended MDCEV 3* represents an ideal case where a variable that is perfectly correlated with the outside good is used to explain its marginal utility.

In the case of time use, we observe the budget (24 hours a day for everyone), giving a clear advantage to the *Traditional MDCEV* model and making it the model of choice for any modeller. Nevertheless, we are interested in exploring the consistency of results across the MDCEV and proposed models, as well as the loss of forecast performance due to not observing the (correct) budget. Table 3 presents the estimated parameters, likelihood and root mean squared error (RMSE) of forecasted consumption at the aggregate sample level for each model.

The parameter estimates are consistent across models. The β parameters across *Traditional MDCEV*, *Extended MDCEV 1* and *Extended MDCEV 2*, have the same sign. For *Extended MDCEV 3*, this tendency only applies for β interaction parameters, while the main effects ($\beta_{\texttt{work}}, \beta_{\texttt{school}}$, etc.) seem to have a constant (approx. 1.7) added to them. This is reasonable as unlike the other applications of the proposed model, in *Extended MDCEV 3*, the average value of ψ_{n0} is not $e^0 = 1$, but $\frac{1}{N}\sum_n e^{\alpha_{outside} x_{n0}} = 5.061 = e^{1.622}$. The γ parameters, on the other hand, share the same order across models (i.e. work is higher, followed by leisure, school, private business and shopping), but their magnitudes is only similar across *Extended MDCEV 1* and *Extended MDCEV 2*. The same is true for the complementarity and substitution parameters across the proposed models.

Parameter interpretation is equivalent across models, except for the α parameters in the *Traditional MDCEV* and the applications of the proposed model. While α measures satiation in

Table 3: Comparison of the proposed (extended) and traditional MDCEV models on a time use dataset

	Traditional MDCEV		Extended MDCEV 1		Extended MDCEV 2		Extended MDCEV 3	
	Estimate	t-ratio*	Estimate	t-ratio*	Estimate	t-ratio*	Estimate	t-ratio*
α Constant	0.000	7.61						
α Female					-0.032	-2.00		
α Weekend					0.031	1.59		
α Outside							0.084	21.46
γ Work	4.899	17.44	16.586	9.21	16.428	9.29	3.975	10.95
γ School	3.099	5.66	9.166	6.19	9.026	6.27	3.430	8.12
γ Shopping	0.426	13.82	3.086	6.32	3.049	6.38	1.414	11.10
γ Private	0.620	9.18	4.533	5.46	4.476	5.51	1.811	13.98
γ Leisure	2.096	16.61	10.522	7.78	10.456	7.85	3.162	11.66
β Work	-3.717	-32.24	-0.201	-5.51	-0.220	-5.70	1.575	16.44
x Full time	1.325	9.84	0.294	6.98	0.298	7.04	0.379	7.76
x weekend	-2.861	-14.66	-0.740	-8.78	-0.713	-8.39	-1.002	-13.95
β School	-7.418	-21.48	-1.211	-8.32	-1.222	-8.36	0.090	0.53
x 30 or younger	2.344	5.84	0.629	5.26	0.630	5.25	0.867	5.40
β Shopping	-3.804	-73.08	-0.311	-11.20	-0.318	-10.58	1.417	17.89
β Private	-4.279	-69.23	-0.416	-11.07	-0.424	-10.78	1.296	28.93
β Leisure	-3.400	-60.29	-0.189	-8.42	-0.208	-8.32	1.542	15.89
x weekend	0.295	3.99	0.082	4.36	0.114	4.20	0.242	8.47
σ	1.000	(fixed)	0.274	10.46	0.276	10.57	0.375	26.88
δ Work - School			-0.780	-5.64	-0.782	-5.77	-7.158	-9.45
δ Work - Shopping			-0.151	-1.86	-0.184	-2.22	-7.549	-7.15
δ Work - Private business			-0.336	-3.06	-0.361	-3.24	-10.720	-20.28
δ Work - Leisure			-0.141	-2.27	-0.138	-2.25	-7.549	-5.22
δ School - Shopping			-0.383	-1.43	-0.414	-1.58	-12.301	-17.49
δ School - Private business			-0.047	-0.20	-0.082	-0.34	-7.605	-2.09
δ School - Leisure			0.009	0.05	0.014	0.07	-7.118	-2.39
δ Shopping - Private business			0.217	2.17	0.217	2.15	-7.853	-9.03
δ Shopping - Leisure			0.133	2.04	0.134	2.06	-7.848	-7.18
δ Private business - Leisure			0.128	1.84	0.132	1.89	-6.982	-14.70
δ_0			0.298	(fixed)	0.298	(fixed)	0.298	(fixed)
Parameters		15		25		27		26
Loglikelihood		-15007		-15188.8		-15181.3		-13915.3
RMSE		407		958		960		667

* *Robust t-ratio*

the traditional MDCEV model, in the proposed model it represents the impact of the associated explanatory variable (z_0) on the marginal utility of the outside good (ψ_0). In the proposed model,

$\alpha > 0$ implies a positive effect of z_0 on ψ_0, and a decreasing consumption of the inside goods when z_0 grows. On the other hand, $\alpha < 0$ implies a negative effect of z_0 on ψ_0, and an increasing consumption of the inside goods when z_0 grows. In this particular application, the negative sign of $\alpha_{\texttt{female}}$ in *Extended MDCEV 2* indicates that, after controlling for other variables, women on average perform more out-of-home activities than men. The effect, however, is too small to be reflected at the aggregate level. On the other hand, the positive sign of $\alpha_{\texttt{weekend}}$ and $\alpha_{\texttt{outside}}$ implies that less time is spent on out-of-home activities during the weekend, and more time is spent at home (as expected). This is visible from the data, where on average, 19.9 hours are spent at home during weekend days, while only 17.4 are spent at home during week days.

Concerning the γ parameters, work has the highest values because it is most commonly consumed for longer periods than other activities, excluding the outside good, as Table 2 shows. As expected, the δ_{kl} parameters are dominated by substitution effects (i.e. $\delta_{kl} < 0$), due to the strongly constrained nature of time allocation. The δ_{kl} parameters show strong substitution between Work and School, as most individuals in the sample either work or study, but not both at the same time. However, models *Extended MDCEV 1* and *Extended MDCEV 2* do capture significant complementarity between Shopping and Private business, and between Shopping and Leisure. This is reasonable, as all these activities are usually carried out in the city centre, and are therefore easier to combine in a single chain of trips. *Extended MDCEV 3* does not capture this complementarity, probably because the inclusion of the outside good in ψ_0 strengthens the substitution effects in the model.

Table 4 compares the correlation between time use in the dataset to the complementarity/substitution parameters (δ_{kl}) in *Extended MDCEV 2*. We focus our comparison on *Extended MDCEV 2* as this is the most realistic use case of the proposed model in this dataset. The sign of correlations and δ_{kl} parameters agree in most cases, except for non-significant parameters and in the case of *shopping*. This last case is probably due to *shopping-private business*, and *shopping-leisure* having very small negative correlations. Indeed, only correlations smaller than -0.05 seem to lead to significant substitution parameters (i.e. $\delta_{kl} < 0$). In Section 4.3, we again compare correlations and complementarity/substitution parameters, but in a dataset where the budget constraint is less strenuous, finding a much stronger connection between them.

As expected, the *Traditional MDCEV* model achieves better likelihood and a lower RMSE than all applications of the proposed model. This is due to the *Traditional MDCEV* model using the exact true budget, giving it an advantage over the proposed models. This way, the *Traditional MDCEV* model achieves an RMSE 60% the size of the best proposed model. Therefore, we recommend using models with explicit budget, such as the traditional MDCEV model, when the budget is known with certainty in the estimation (training) dataset, and will be known with certainty in all forecast scenarios.

Table 4: Correlation and complementarity/substitution parameters in time use dataset.

	Work	School	Shopping	Private B.	Leisure
Work		-0.78	-0.18	-0.36	-0.14
School	-0.06		(-0.41)	(-0.08)	(0.01)
Shopping	-0.08	-0.03		0.22	0.13
Private B.	-0.09	0.00	-0.01		(0.13)
Leisure	-0.17	-0.01	-0.01	-0.04	

Correlations in the lower triangular matrix.
δ_{kl} parameters from Extended MDCEV 2 in upper triangular matrix.
Non significant parameters are in parenthesis.

4.2 Variable budget and fixed prices: expenditure dataset

The second dataset records expenditure during a fortnight for 10,460 Chilean households, aggregated to a dozen categories: *food*, *alcoholic beverages*, *clothing*, *bills* (rent and utilities), *homeware*, *health*, *transport*, *communications* (IT), *leisure*, *education*, *restaurants*, and *other*. This data comes from the 7^{th} Chilean Expenditure Survey (Bilbao, 2013). We use the expenditure in *bills* as the outside good, as all households in the sample pay rent or utilities, and as this on average the biggest expenditure of most households. Table 5 presents a summary of the data. All data is in thousands of Chilean pesos (kCLP).

We estimated five different models with the available data. *MDCEV 100* is a an MDCEV model (Bhat, 2008) with explicit budget equal to each household total expenditure, i.e. with the true (correct) budget. We also estimated two additional MDCEV models, one assuming only 75% and another 125% of the true budget, which we call *MDCEV 75* and *MDCEV 125*, respectively. We also estimated two proposed models. *Extended MDCEV 1* is the proposed mode, as described in 2.1, including socio-demographics in the definition of the outside good marginal utility (ψ_0), and only the most relevant complementarity/substitution parameters (δ). We identified the most relevant interactions between consumption of different activities through a Principal Component Analysis, and estimated a subset of δ parameters that allowed us to capture these major interactions. *Extended MDCEV 2* is similar to the previous model, but it fixes all δ parameters to zero and adds a random component $\varepsilon_n 0$ following a normal distribution with mean zero and standard deviation σ_0 to the definition of ψ_0 (see section 3.4). Parameter estimates, as well as maximum log-likelihood values for *MDCEV 100*, *Extended MDCEV 1* and *Extended MDCEV 2* are presented in Table 6. Complementarity and substitution parameters of *Extended MDCEV 1* are presented in 7, as in *Extended MDCEV 2* all complementarity and substitution parameters are set to zero.

Our objective was to analyse how errors in the definition of the budget lead to different forecast

Table 5: Main descriptive statistics of the expenditure data

	Fraction of the sample who bought	Total consumption (kCLP)	Average consumption when bought (kCLP)	Average consumption when bought (fraction of budget)
Food	99.6%	1532154	147.04	19.8%
Alcohol	53.8%	132523	23.55	2.8%
Clothing	53.5%	378139	67.57	5.8%
Bills	100.0%	2703618	258.47	32.7%
Homeware	88.1%	585591	63.55	5.3%
Health	72.7%	543183	71.39	5.9%
Transport	92.2%	1421741	147.50	11.7%
Communications	80.8%	418142	49.51	5.3%
Leisure	84.0%	580010	65.99	5.7%
Education	56.9%	641883	107.95	8.7%
Restaurants	69.8%	364162	49.91	4.2%
Others	93.9%	742852	75.62	6.7%

errors in models with explicit budget. To do this, after estimating each model, we forecast demand with them on the same training set. We did this multiple times, each time assuming a different value of the budget. Different budgets lead to different forecasts in the MDCEV models, but not in the proposed model. Figure 4 presents the results of this exercise. We used the root mean squared error (RMSE) of the aggregate predictions in the sample as an indicator of error in the forecast.

As Figure 4 shows, the forecast performance of the proposed model does not change as a function of the budget. Instead, the MDCEV models achieve a better forecast performance than the proposed models when the forecast budget is close to the estimation budget, but their error grows in a quadratic way with the budget misspecification. It does not seem to be very important how the estimation budget is defined in MDCEV models. For example, the estimation budget could be defined as the total income of the household or just the total expenditure on the inside goods plus one. In fact, the MDCEV models seem to perform better when the expenditure on the outside good is small. However, once a budget has been used during estimation, it is very important to accurately and consistently predict the budget for any forecasting scenario. In our application, a misspecification of the forecast budget by more than 13% leads to bigger errors than the *Extended MDCEV 1* model.

These results reveal that in contexts where the forecasting of the budget implies even mild uncertainty, the proposed model can ensure a bounded level of error in the forecast.

Table 6: Comparison of MDCEV and proposed model on expenditure dataset

	Extended MDCEV 1		Extended MDCEV 2		MDCEV 100	
	Estimate	t-ratio*	Estimate	t-ratio*	Estimate	t-ratio*
α Household (hh) size (people)	-0.043	-16.12	-0.059	-17.87		
α hh head works	-0.085	-19.75	-0.112	-21.41		
α hh head's years of education	-0.049	-39.51	-0.061	-48.43		
α hh head is female	0.052	7.76	0.064	7.56		
α hh head's age	-0.025	-4.17	-0.056	-7.55		
α Constant					0.000	76.57
γ Food	39.846	30.08	23.975	24.59	3.043	8.70
γ Alcoholic beverages	20.948	44.81	18.558	51.65	7.900	57.22
γ Clothing	46.653	42.20	42.709	44.15	16.678	46.32
γ Homeware	23.570	37.12	19.464	41.31	5.165	43.06
γ Health	32.998	39.25	28.673	41.42	9.389	45.47
γ Transportation	43.630	40.09	34.560	42.25	9.147	39.75
γ Communications	32.027	56.58	26.563	59.16	10.230	59.85
γ Leisure	29.626	43.85	24.044	47.66	7.238	49.37
γ Education	57.075	31.05	59.901	33.77	19.179	29.84
γ Restaurants	27.900	43.98	23.862	47.24	8.364	51.70
γ Others	24.912	37.83	18.958	40.69	4.335	37.96
β Food	0.250	8.01	0.320	7.69	-1.720	-15.13
β Alcoholic beverages	-1.217	-37.49	-1.429	-36.41	-5.451	-300.71
β Clothing	-1.050	-30.88	-1.386	-36	-5.363	-304.20
β Homeware	-0.564	-18.92	-0.685	-19.19	-3.922	-200.90
β Health	-0.776	-24.30	-1.048	-28.59	-4.684	-286.90
β Transport	-0.285	-9.47	-0.480	-13.57	-3.553	-156.02
β Communications	-0.558	-17.33	-0.769	-21.09	-4.283	-243.63
β Leisure	-0.546	-17.63	-0.742	-20.71	-4.098	-230.39
β Education	-0.944	-28.88	-1.373	-36.17	-5.322	-295.63
β Restaurants	-0.799	-24.71	-1.050	-28.64	-4.723	-279.73
β Others	-0.287	-9.65	-0.409	-11.71	-3.331	-138.10
σ	0.564	102.57	0.533	127.57	1.000	(fixed)
σ_0			0.3655	74.86		
Number of parameters		36		29		23.00
Likelihood		-505873.8		-498488.2		-492514.7

Table 7: Complementarity and substitution parameters of Extended MDCEV 1 on expenditure dataset.

	Estimate	t-ratio*
δ Food, Alcohol	7.980	10.71
δ Homeware, Health	4.497	7.56
δ Homeware, Transportation	-3.719	-5.97
δ Homeware, Communications	9.308	12.29
δ Homeware, Leisure	5.413	8.15
δ Homeware, Education	-7.901	-12.92
δ Homeware, Restaurants	4.077	6.42
δ Homeware, Others	5.554	8.35
δ_0	0.010	(fixed)

* *Robust t-ratio*

Table 8: Main descriptive statistics of the scanner data

	Fraction of the sample who bought	Total consumption (USD)	Average consumption when bought (USD)	Average consumption when bought (fraction of budget)
Oranges	24.0%	758	0.79	0.99%
Peaches	28.0%	988	0.88	1.22%
Pears	20.7%	645	0.78	1.01%
Pineapple	43.4%	1406	0.81	1.15%

4.3 Variable budget and variable prices: supermarket scanner dataset

The third application deals with scanner data from a chain of supermarkets (Venkatesan, 2014). After dropping all records of transactions from households with missing socio-demographic characteristics, and limiting the analysis to only four product categories, the dataset contains 4,002 purchase baskets from 656 households. All the considered product categories are fresh fruits: oranges, peaches, pears, and pineapples. Each fruit can be purchased in packs of different weights, but to simplify the analysis, we calculated the average price per Kg of each product, and expressed the amount purchased in Kg. We assumed the budget to be 9.7% of the household income, i.e. the average expenditure on food by a household in the US (USDA Economic Research Service, 2019), and the consumption of the outside good equal to the difference between the budget and the expenditure in the fruit. Table 8 summarises consumption in the dataset.

Our objective with this dataset was to compare the forecast sensitivity to changes in price

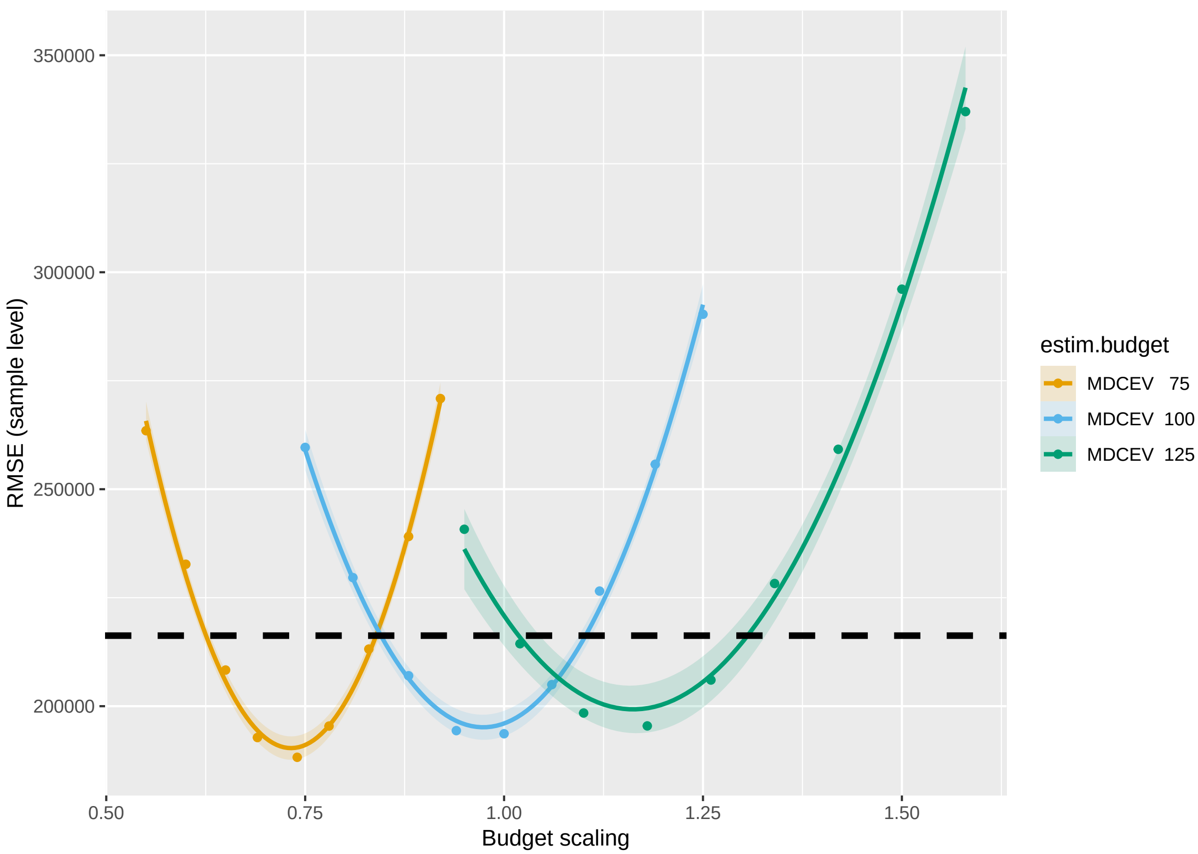

Figure 4: Compared forecast error of the Extended MDCEV 1 model (horizontal line) to MDCEV models (Bhat, 2008) when the budget is misspecified

of the proposed model, to that of a model with explicit budget. We estimated two models on the supermarket dataset. One traditional MDCEV model with the explicit budget as defined in the previous paragraph, and one proposed model (*Extended MDCEV*). The parameter estimates and optimal log-likelihood of these models are shown in Table 9. We then repeatedly forecast consumption on the training dataset, changing the price of oranges each time from 75% to 150% of their original price. In each case, we recorded the aggregated forecasted demand at the sample level. Figure 5 plots the changes in forecast for both the MDCEV and proposed models.

As can be seen in Figure 5, both the traditional and extended MDCEV models show stronger variation for the demand for oranges (the product whose price is changing) than for any other product. However, the scale of the change is very different across models. The *Traditional MDCEV* forecasts a much higher and more elastic demand, while the *Extended MDCEV* predicts a lower

Table 9: Parameters estimates of traditional and proposed (extened) MDCEV models on the supermarket scanner dataset

	Extended MDCEV		Traditional MDCEV	
	Estimate	t-ratio*	Estimate	t-ratio*
α Household size	-0.023	-1.86		
α Age	-0.022	-1.97		
α Constant			-12.014	-113.49
γ Oran	0.637	7.18	0.224	12.21
γ Peach	0.729	10.03	0.353	18.97
γ Pear	0.952	10.70	0.346	20.22
γ Pineapple	0.660	12.46	0.313	22.54
β Orange	0.249	1.34	-5.559	-41.18
β Peach	0.288	2.43	-5.079	-64.71
β Pear	0.134	1.01	-5.385	-63.34
β Pineapple	0.588	6.18	-4.655	-63.28
β Discount	0.479	6.18	2.366	43.81
δ Oranges - Peaches	-0.223	-2.47		
δ Oranges - Pears	-0.077	-1.29		
δ Oranges - Pineapple	-0.386	-4.87		
δ Peaches - Pears	0.273	4.44		
δ Peaches - Pineapple	-0.526	-8.69		
δ Pears - Pineapple	-0.313	-7.26		
σ	0.498	11.88		
δ_0	2.161	(fixed)	1.000	(fixed)
Number of parameters		18		10
Log-likelihood		-10264.57		-12287.81

* *Robust t-ratio*

and less elastic one. While it is not possible to determine what model forecast is more accurate for hypothetical prices, at the observed prices, the *Extended MDCEV* forecast is more accurate at the original prices, with an RMSE of the aggregate demand equal to 429 against 2,974 of the *Traditional MDCEV*.

Table 10 compares the correlation between products in the dataset to the complementarity/substitution parameters (δ_{kl}) in the proposed model estimated in the supermarket scanner dataset. All signs match between the two group of values, with substitution effects dominating, except between pears and peaches. Among significant parameters, a higher correlation leads to a

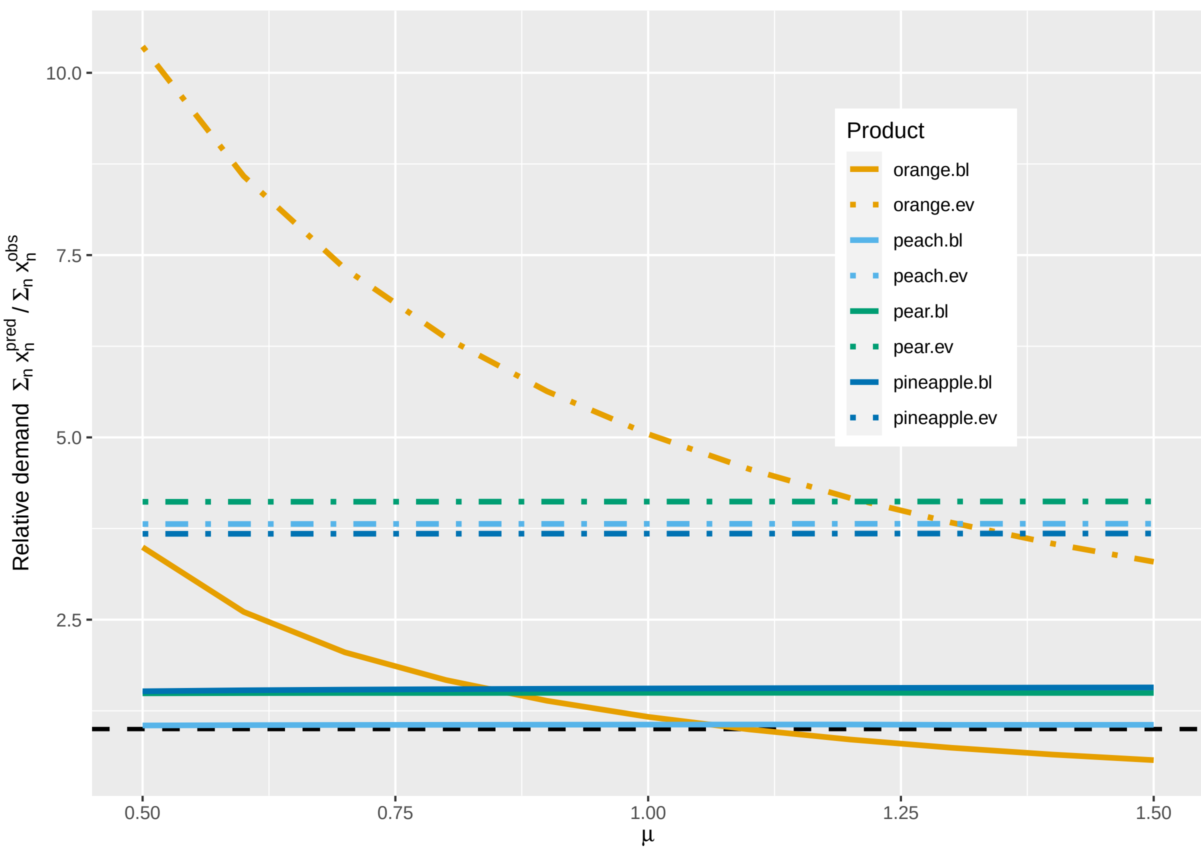

Figure 5: Relative aggregated sample demand forecasted by the traditional and extended MDCEV models for variations in the price of oranges. The black line indicates unity (i.e. original demand).

higher δ_{kl} parameter, but -as expected due to the form of u_{kl} this relationship is not linear.

In summary, the δ_{kl} parameters in the proposed model seem to efficiently capture the complementarity/substitution effects in datasets where the budget is large as compared to the expenditure in inside goods. These parameters are capable of inducing cross elasticities between alternatives in the model, despite not having the budget constraint to force substitution between them. And even though a direct interpretation of these parameters values is difficult (even in relative terms), their sign, significance and magnitude hint at whether substitution or complementarity dominates the relationship between a pair of alternatives, and in which pairs these effects are higher.

Table 10: Correlation and complementarity/substitution parameters in supermarket scanner dataset.

	Oranges	Peaches	Pears	Pineapples
Oranges		-0.22	-0.08*	-0.39
Peaches	-0.04		0.27	-0.53
Pears	-0.05	0.13		-0.31
Pineapples	-0.08	-0.16	-0.09	

Correlations in the lower triangular matrix.
δ_{kl} *parameters in the upper triangular matrix.*
** Non significant parameter*

4.4 Unknown budget: Number of trips by purpose dataset

The last application deals with number of trips generated by a household, split across different purposes: *work, study, personal business, leisure* and *return home*. Data comes from the 2012 Origin-Destination survey of Santiago, Chile (Social, 2014). The database contains observations for a single day from 10,927 households. Table 11 summarises the average number of trips per purpose by socio-demographic characteristics.

Our objective with this dataset is to compare forecast performance between a model with explicit budget (*Traditional MDCEV*) and the proposed model (*Extended MDCEV*), when the definition of the budget is arbitrary. In theory, the budget in our dataset should be the maximum amount of trips a household could generate during a day, but this value is very difficult to determine. Defining the budget as any lower (but more reasonable) value would be an arbitrary decision. A common approach in situations without an evident budget is to use the observed total consumption as the budget (Bhat and Sen, 2006). We follow this approach when estimating the MDCEV model, assuming the budget to be equal to the observed total number of trips. When forecasting with the MDCEV model, we predict the budget using a linear regression on the observed data (see Table 12). In the case of the proposed model, we have no need to make assumptions on the budget, as it is not needed during estimation nor forecasting.

We used the same socio-demographics described in Table 11 as explanatory variables in both the traditional and extended MDCEV models, removing the non-significant factors. The utility functions were linear in both cases. The marginal utility of the outside good was set to zero for the proposed model.

Forecasting performance was measured through cross-validation, by randomly selecting 75% of the data for model estimation, and forecasting for the remaining 25%. This process was repeated 30 times, allowing us to obtain a mean RMSE at the observation and sample level, and a measure

Table 11: Main descriptive statistics of the number of trips database

		Number of trips							Homes
		Work	Study	Per. B.	Shopping	Leisure	Ret. home	All	
Household size	1	0.57	0.05	0.7	0.39	0.13	1.31	3.15	1205
	2	0.87	0.16	0.83	0.57	0.13	2.08	4.64	2285
	3	1.32	0.61	0.93	0.51	0.16	2.94	6.47	2610
	4	1.56	1.16	1.16	0.55	0.15	3.95	8.53	2497
	5	1.78	1.68	1.26	0.59	0.22	4.84	10.37	1414
	≥ 6	2.28	2.15	1.43	0.68	0.22	5.89	12.65	916
Number of vehicles	0	1.16	0.78	0.82	0.54	0.12	3.05	6.46	6475
	1	1.46	0.91	1.17	0.52	0.17	3.5	7.73	3508
	≥ 2	2.11	1.08	1.81	0.67	0.41	4.35	10.44	944
Bicycles	0	1.17	0.71	0.99	0.53	0.13	3.01	6.55	5991
	≥ 1	1.54	1.01	1.05	0.56	0.19	3.66	8.02	4936
Number of workers	0	0.03	0.41	1.1	0.72	0.13	2.08	4.46	1659
	1	0.96	0.87	1.04	0.55	0.16	3.04	6.61	4580
	2	1.84	0.98	0.97	0.46	0.16	3.67	8.08	3346
	3	2.74	0.97	0.98	0.52	0.21	4.56	9.97	1017
	≥ 4	3.9	1.04	0.97	0.54	0.22	5.6	12.26	325
Household income	Low	0.61	0.74	0.97	0.6	0.12	2.7	5.75	3691
	Mid	1.34	0.92	0.93	0.51	0.12	3.34	7.16	3605
	High	2.04	0.89	1.16	0.52	0.23	3.86	8.7	3631
Total		1.34	0.85	1.02	0.54	0.16	3.3	7.21	10927

of their variability. Table 13 presents the coefficients of each model when estimated with the whole dataset, and the estimated forecast performance obtained through the cross-validation procedure.

Establishing parallels between the parameters of both models is difficult, as they have very different scale and intercepts. Interpretation of the *Traditional MDCEV* model is difficult, as to determine the overall marginal effect of an explanatory variable, the analysts should consider the effect of that variable both in the *Traditional MDCEV* model and in the linear regression predicting the budget of individuals. For example, even though the number of vehicles has a negative effect on the base utility of work trips in the *Traditional MDCEV* model (see Table 13), the same variable has a positive effect on the budget (see Table 12), making it difficult to predict

Table 12: Linear regression for budget prediction

	Estimate	t-value
Intercept	0.746	10.13
HH. size	1.644	72.39
N. vehicles	1.009	21.29
HH. income	0.251	5.37
N. workers	0.094	2.43
Bicycle	0.234	3.81
R^2		0.47

what the overall effect will be. On the other hand, interpretation of the proposed model is more straightforward, with number of vehicles having a positive effect on the number of work trips.

The *Extended MDCEV* model captures substitution among most trip purposes, as shown in Table 14. As travel demand is conditional on the activities an individual performs during a day, and as performing one activity implies giving up on another, we expected a large amount of substitution between trip purposes. The only purpose that evades this reasoning is *return home*, as it is clearly complementary with other kind of trip purposes. The only exception is *leisure*, showing substitution with return home. This is probably because *leisure* trips are most commonly chained with other trips, such as *work* or *study*, therefore correlating less with *return home*.

In term of forecast performance, the *Extended MDCEV* model is significantly better than the *Traditional MDCEV* model at the sample level (t-ratio=4.5, see bottom of Table 13). At the individual level, both models perform equivalently, though these kinds of models are rarely estimated to forecast at the individual level. On top of a better forecast performance, the proposed model does not require the additional extra step of predicting the budget, making analysis simpler and requiring fewer assumptions.

5 Conclusions

Many decisions can be represented by interrelated discrete and continuous choices, i.e. select what (incidence) and how much (quantity) to choose from a set of finite alternatives. A few examples include purchase decisions at a retail store (what to buy and how much of it), time use (what activities to perform and for how long), investment decisions (what instruments to buy or projects to execute and how much to invest in each), energy matrix choice (what energy sources to use and how much of each), etc. Among other approaches, this kind of decisions have been modelled using Kuhn-Tucker demand systems, which derive econometric models directly from the consumer utility maximising problem. This provides a strong grounding in economic theory, but

Table 13: Parameter estimates and forecast performance for models on number of trips dataset

		Extended MDCEV		Traditional MDCEV	
		Estimate	t-ratio*	Estimate	t-ratio*
	α			0.0000	89.18
Work	β Intercept	-0.0044	-3.88	2.0246	29.98
	β HH. size	0.0016	6.84		
	β N. vehicles	0.0017	3.93	-0.3614	-9.02
	β Bicycle	0.0019	3.48	-0.0951	-3.21
	β HH. income			-0.1158	-3.81
	β N. workers	0.0165	27.58	0.8574	22.45
	γ	41.2179	33.46	0.4666	60.39
Study	β Intercept	-0.0254	-14.40	0.4104	5.37
	β HH. size	0.0110	25.31	0.6415	45.29
	β N. vehicles	0.0035	8.66	-0.2753	-6.31
	β Bicycle	0.0033	5.60		
	β HH. income			-0.2373	-5.91
	β N. workers	-0.0060	-13.19	-0.2355	-5.64
	γ	51.9430	38.62	0.7342	67.13
Personal	β Intercept	-0.0028	-3.09	2.0569	27.03
Business	β HH. size	0.0043	20.37	0.1633	11.55
	β N. vehicles	0.0052	12.99	-0.1211	-3.04
	β Bicycle			-0.2657	-7.20
	β HH. income	0.0010	3.24	-0.0684	-2.78
	β N. workers	-0.0038	-10.87	-0.3255	-8.01
	γ	63.9743	36.50	1.3254	70.34
Shopping	β Intercept	-0.0084	-7.75	1.9081	23.71
	β HH. size	0.0040	17.27	0.0698	4.42
	β N. vehicles	0.0024	5.79	-0.3417	-7.59
	β Bicycle	0.0014	2.53	-0.1647	-3.91
	β HH. income			-0.1384	-4.25
	β N. workers	-0.0026	-6.27	-0.2554	-5.87
	γ	51.2864	32.66	1.2787	85.52
Leisure	β Intercept	-0.0292	-16.10	0.0000	(fixed)
	β HH. size	0.0026	7.18		
	β N. vehicles	0.0072	11.74		
	β Bicycle	0.0051	5.27		
	β HH. income	0.0018	4.74		
	β N. workers				
	γ	48.3174	26.18	1.4826	60.89
Return	β Intercept	0.0035	5.65	5.8443	45.94
home	β HH. size	0.0076	26.43	0.2345	28.41
	β N. vehicles	0.0027	8.17	-0.3297	-8.93
	β Bicycle	0.0015	3.37	-0.1563	-6.88
	β HH. income			-0.1490	-6.03
	β N. workers	0.0020	6.77	0.0933	2.54
	γ	74.2110	39.93	0.0423	9.43
σ		0.0162	39.83	1.0000	(fixed)
α				0.0000	89.18
δ_0		0.7352	(fixed)		
Loglikelihood			-77164		65409
Number of parameters			51		35
RMSE obs. level		1.00	0.02	1.03	0.02
RMSE samp. level		73.10	23.77	228.93	25.06

Table 14: Correlation and complementarity/substitution parameters in trips dataset.

	Work	Study	P. Business	Shopping	Leisure	Return H.
Work		-0.0142	-0.0292	-0.0337	-0.0110	0.0090
Study	0.12		-0.0168	-0.0243	-0.0139	0.0066
P. Business	-0.02	0.14		-0.0219	-0.0047	0.0076
Shopping	-0.09	-0.05	0.04			0.0147
Leisure	0.12	0.02	0.11	0.06		-0.0234
Return H.	0.39	0.61	0.47	0.32	0.20	

Correlations and δ_k parameters in the lower and upper triangular matrix, respectively. All parameters are significant at 95% confidence.

also implies the necessity to define a budget, and imposes limitations on the definition of the utility function, leading to the omission of relevant effects (most notably complementarity and substitution) in most implementations.

In this paper, we proposed an extension to the Multiple Discrete Continuous Extreme Value (MDCEV) model, a type of Kuhn-Tucker demand model, that does not require the analyst to define a budget, and which incorporates complementarity and substitution effects. The lack of need to define a budget is particularly useful when forecasting, as it avoids cascading errors due to inaccuracies in the budget forecast (see section 4.2). The inclusion of explicit complementarity and substitution effects, on the other hand, enriches the interpretability and realism of the model (Manchanda et al., 1999). In addition, the functional form selected to represent complementarity and substitution effects in the utility function avoids multiple issues present in previous formulations proposed in the literature (see section 1).

The proposed model is based on the hypothesis that total expenditure on the alternatives under consideration is small compared to the overall budget. This hypothesis allows us to approximate the utility of the numeraire good by a linear function, hence removing the necessity to define a budget. Indeed, simulations show that the proposed model forecast approaches that of a model with explicit budget when the hypothesis above is fulfilled. Such an assumption is realistic in most daily consumption decisions, but should always be justified when using the model. As the proposed model requires less information than an equivalent model with explicit budget, its forecast performance is also generally lower. Therefore, if the budget can be determined with a great degree of confidence, then we recommend using a model with explicit budget, such as the MDCEV (Bhat, 2008). However, if there is even mild uncertainty in the budget, the proposed model might be a useful alternative, as it makes the prediction error independent from the budget estimation.

An R implementation of the proposed model is available for R, as an extension of the *Apollo* package (Hess and Palma, 2019). To download this extension and see examples, visit ApolloChoiceModelling.com.

The model proposed in this paper contributes to the literature on Kuhn-Tucker system demand models to study multiple-discrete choices. There are still several avenues for improvement and further investigation. New functional forms for the complementarity and substitution term in the direct utility function could be explored, with special emphasis on those leading to a compact form of the Jacobian in the likelihood function. More generally, including a random component in the marginal utility of the outside good would be a useful development, specifically if it leads to a closed-form likelihood function. Finally, alternative formulations based on indirect utility functions could be less restrictive, as they avoid assumptions on the shape of decision makers' direct utility functions.

6 Acknowledgements

The authors thank Abdul Pinjari for his valuable feedback on an earlier version of the proposed model, presented at the 6^{th} International Choice Modelling Conference. The authors also acknowledge the financial support by the European Research Council through the consolidator Grant 615596-DECISIONS.

References

Bhat, C.R., 2008. The multiple discrete-continuous extreme value (mdcev) model: role of utility function parameters, identification considerations, and model extensions. Transportation Research Part B: Methodological 42, 274–303.

Bhat, C.R., 2018. A new flexible multiple discrete–continuous extreme value (mdcev) choice model. Transportation Research Part B: Methodological 110, 261–279.

Bhat, C.R., Castro, M., Pinjari, A.R., 2015. Allowing for complementarity and rich substitution patterns in multiple discrete–continuous models. Transportation Research Part B: Methodological 81, 59–77.

Bhat, C.R., Sen, S., 2006. Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (mdcev) model. Transportation Research Part B: Methodological 40, 35–53.

Bilbao, F., 2013. VII Encuesta de presupuestos familiares. techreport. Instituto Nacional de Estadísticas Chile.

Calastri, C., Hess, S., Daly, A., Carrasco, J.A., 2017. Does the social context help with understanding and predicting the choice of activity type and duration? an application of the multiple discrete-continuous nested extreme value model to activity diary data. Transportation Research Part A: Policy and Practice 104, 1–20.

Calastri, C., Hess, S., Palma, D., Crastes dit Sourd, R., 2019. Capturing relationship strength: a choice model for leisure time, frequency of interaction and ranking in name generators. Working paper.

Calastri, C., dit Sourd, R.C., Hess, S., 2020. We want it all: experiences from a survey seeking to capture social network structures, lifetime events and short-term travel and activity planning. Transportation 47, 175–201. doi:10.1007/s11116-018-9858-7.

Chintagunta, P.K., 1993. Investigating purchase incidence, brand choice and purchase quantity decisions of households. Marketing Science 12, 184–208.

Dumont, J., Hess, S., Daly, A., Ferdous, N., 2013. The use of the multiple discrete continuous extreme value modeling framework when the budget is latent: two consumer package good examples, in: 3rd International Choice Modelling Conference, Sydney.

Enam, A., Konduri, K.C., Eluru, N., Ravulaparthy, S., 2018. Relationship between well-being and daily time use of elderly: evidence from the disabilities and use of time survey. Transportation 45, 1783–1810.

Ferdous, N., Pinjari, A.R., Bhat, C.R., Pendyala, R.M., 2010. A comprehensive analysis of household transportation expenditures relative to other goods and services: an application to united states consumer expenditure data. Transportation 37, 363–390.

Hanemann, W.M., 1978. A methodological and empirical study of the recreation benefits from water quality improvement. Department of Agricultural and Resource Economics, University of California.

Hausman, J.A., Leonard, G.K., McFadden, D., 1995. A utility-consistent, combined discrete choice and count data model assessing recreational use losses due to natural resource damage. Journal of Public Economics 56, 1–30.

Hess, S., Palma, D., 2019. Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. Journal of Choice Modelling , 100170.

Hicks, J.R., Allen, R.G., 1934. A reconsideration of the theory of value. part i. Economica 1, 52–76.

Jäggi, B., Erath, A., Dobler, C., Axhausen, K.W., 2012. Modeling household fleet choice as function of fuel price by using a multiple discrete–continuous choice model. Transportation Research Record 2302, 174–183.

Jeong, J., Kim, C.S., Lee, J., 2011. Household electricity and gas consumption for heating homes. Energy Policy 39, 2679–2687.

Kim, J., Allenby, G.M., Rossi, P.E., 2002. Modeling consumer demand for variety. Marketing Science 21, 229–250.

Lim, S., Kim, Y., 2015. How to design public venture capital funds: Empirical evidence from s outh k orea. Journal of Small Business Management 53, 843–867.

Lu, H., Hess, S., Daly, A., Rohr, C., 2017. Measuring the impact of alcohol multi-buy promotions on consumers' purchase behaviour. Journal of choice modelling 24, 75–95.

Manchanda, P., Ansari, A., Gupta, S., 1999. The "shopping basket": A model for multicategory purchase incidence decisions. Marketing science 18, 95–114.

Mehta, N., Ma, Y., 2012. A multicategory model of consumers' purchase incidence, quantity, and brand choice decisions: Methodological issues and implications on promotional decisions. Journal of Marketing Research 49, 435–451.

Pellegrini, A., Sarman, I., Scagnolari, S., 2017. Stochastic frontier estimation of holiday budgets for multiple discrete-continuous extreme value model: An application to tourist expenditure analysis, in: 5yh International Choice Modelling Conference.

Pellegrini, A., Saxena, S., Pinjari, A., Dekker, T., 2019. Alternative non-additively separable utility functions for random utility maximization based multiple discrete continuous models, in: 6th International Choice Modelling Conference.

Phaneuf, D.J., Herriges, J.A., 1999. Choice set definition issues in a kuhn-tucker model of recreation demand. Marine Resource Economics 14, 343–355.

Pinjari, A.R., Augustin, B., Sivaraman, V., Imani, A.F., Eluru, N., Pendyala, R.M., 2016. Stochastic frontier estimation of budgets for kuhn–tucker demand systems: Application to activity time-use analysis. Transportation Research Part A: Policy and Practice 88, 117–133.

Richards, T.J., Gómez, M.I., Pofahl, G., 2012. A multiple-discrete/continuous model of price promotion. Journal of Retailing 88, 206–225.

Social, O., 2014. Actualización y recolección de información del sistema de transporte urbano, IX Etapa: Encuesta Origen Destino Santiago 2012. Encuesta origen destino de viajes 2012. Technical Report. Universidad Alberto Hurtado.

Song, I., Chintagunta, P.K., 2007. A discrete–continuous model for multicategory purchase behavior of households. Journal of Marketing Research 44, 595–612.

Train, K.E., 2009. Discrete choice methods with simulation. Cambridge university press.

USDA Economic Research Service, 2019. Food prices and spending. https://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/food-prices-and-spending/. Accessed: 2019-11-27.

Venkatesan, R., 2014. The complete journey. techreport. Dunnhumby Source Files.

Wales, T.J., Woodland, A.D., 1983. Estimation of consumer demand systems with binding non-negativity constraints. Journal of Econometrics 21, 263–285.