From unobserved to observed preference heterogeneity: a revealed preference methodology

Laurens Cherchye* Dieter Saelens† Reha Tuncer‡

Abstract

We present an easy-to-apply nonparametric revealed preference method to identify observed preference heterogeneity from cross-sectional data. Building on the partitioning approach that was developed by Crawford and Pendakur (2013) and Cosaert (2019), it quantifies the contribution of observable consumer characteristics to describing the identified preference heterogeneity. We demonstrate the practical usefulness of our method through an application to newly gathered experimental data on consumer choice behavior in two types of decision situations: the allocation of money (choosing between two products) and the allocation of time (choosing between leisure and work). We investigate whether the same consumer characteristics drive the observed variation in choice behavior in these two settings.

*Department of Economics, University of Leuven. E-mail: laurens.cherchye@kuleuven.be.

†Department of Economics, University of Leuven and National Bank of Belgium. E-mail: dieter.saelens@kuleuven.be.

‡Department of Economics, University of Luxembourg. E-mail: reha.tuncer@uni.lu

Keywords: revealed preference, preference heterogeneity, partitioning, observable characteristics.

JEL: C14, C38, D12

1 Introduction

The empirical analysis of demand behavior has a longstanding tradition in the microeconomics literature. An important issue relates to dealing with factors different from prices and incomes (defining consumers' budgets) that impact individual consumer behavior; this issue is commonly referred to as "preference heterogeneity". In empirical applications, the applied researcher is often bound to using cross-sectional data that consist of observed choices (prices and quantities) and consumer characteristics (age, gender, etc.). Individual preferences remain unobserved and need to be identified from the observed behavior. The typical approach then consists of pooling individuals with similar observable characteristics and assuming that they have similar preferences. Yet, empirical evidence indicates that preferences may significantly differ between even the most similar-looking individuals; "unobserved" preference heterogeneity is often prevalent.

Obviously, improperly accounting for heterogeneity in consumer preferences undermines the empirical validity of the econometric estimation and identification results. This poses a difficult question: how can we model consumer behavior in a way that accommodates the heterogeneity while preserving theoretical consistency? In the current paper, we tackle this question by making use of nonparametric revealed preference techniques in the tradition of Samuelson (1938, 1948), Houthakker (1950), Afriat (1967), Diewert (1973) and Varian (1982). We present

a novel and easy-to-implement method to identify the observable consumer characteristics that drive the preference heterogeneity underlying the observed variation in consumers' choice behavior. This effectively moves the identification analysis from "unobserved" to "observed" preference heterogeneity. In empirical demand analysis, such identification can be instrumental to stratify the sample in terms of observable characteristics prior to the demand estimation exercise and/or to define observable preference factors to be included in the demand function specification.

We adopt a partitioning approach that follows original work of Crawford and Pendakur (2013), who equally used revealed preference tools to nonparametrically deal with preference heterogeneity in empirical demand analysis. For a given cross-sectional data set, Crawford and Pendakur's approach identifies the minimum number of preference types by partitioning the sample of consumers into subsets that are each consistent with the utility maximization hypothesis for a type-specific utility function. The approach is intrinsically nonparametric in that it does not require an a priori (typically non-verifiable) functional specification of the utility functions. Operationally, it boils down to partitioning the sample into subsets so that each individual subset satisfies the Generalized Axiom of Revealed Preference (GARP); the minimum number of preference types then corresponds to the partitioning with minimum cardinality (see Crawford and Pendakur, 2013, for more details). More recently, Cosaert (2019) addressed the computational complexity of Crawford and Pendakur's original procedure.[1] He introduced a graph-theoretical approach to identify the number of types that is based on the Weak Axiom of Revealed Preferences (WARP) (instead of GARP).[2] Both Crawford and Pendakur and Cosaert focused on identifying the number of consumer types with similar preferences by only analyzing the observed choice behavior (prices and

quantities), without directly including information on the consumers' observable characteristics.[3]

We extend this earlier work by explicitly including these observed consumer characteristics in the preference identification analysis. Building on Cosaert (2019)'s partitioning procedure, we introduce a novel method that identifies the consumer characteristics that are most informative in describing the observed heterogeneity in choice behavior. In particular, our method calculates a measure that quantifies the contribution of every observable characteristic to the nonparametrically identified preference heterogeneity. By its nonparametric nature, the method has two specifically attractive features. First, it builds on a fully theory-consistent partitioning of the data that only applies the nonparametric restrictions that result from the utility maximization hypothesis. Second, it brings theory to the data in a simple and direct way: it does not require additional assumptions about anything on which economic theory is silent; empirical applications only require information on observed prices, chosen quantities and observable characteristics for a cross-section of consumers.

We illustrate the practical usefulness of our method through an application to newly gathered experimental data. In this application, we specifically focus on preference heterogeneity in a cross-section of consumers in two types of decision situations: the allocation of money (choosing between two products) and the allocation of time (choosing between leisure and work). This particular set-up allows us to analyze whether the same consumer characteristics drive the observed variation in choice behavior in the two choice settings. We also propose a subsampling procedure to prevent overfitting the data in practical applications of our preference identification method, and we suggest a permutation approach to address

statistical testing of the independence between observed heterogeneity in consumer choices and observable consumer characteristics.

Before entering our analysis, it is useful to contrast the approach that we develop in the current paper with a frequent practice in the revealed preference literature that relates the degree of consumer rationality (usually measured through Afriat's Critical Cost Efficiency Index (CCEI; see Afriat, 1973; Varian, 1990)) to observable consumer characteristics; see, for example, Choi et al. (2014). Both approaches make use of nonparametric revealed preference methods to relate observed choice behavior to consumer characteristics. However, we see at least two important differences. First, we start from cross-sectional data (with a single observed choice per individual consumer), whereas measuring the degree of individual-specific rationality requires multiple choice observations for each consumer that is evaluated. Second, the two approaches serve a conceptually different purpose: we assume rational decision-making, and our central interest is in identifying the observable characteristics that describe the revealed preference heterogeneity underlying the observed consumption behavior; while the other approach allows for irrational choice behavior, and then principally aims at associating differences in rationality with observable consumer characteristics.

The remainder of our paper unfolds as follows. Section 2 introduces our methodology. Section 3 presents the design of our experiment. Section 4 discusses our empirical results. Section 5 concludes.

2 Methodology

This section sets out our approach to quantify preference heterogeneity and explains how to relate this heterogeneity to observable characteristics defining different demographic groups. We start by briefly recapturing Cosaert (2019)'s partitioning procedure, which is based on the original work of Crawford and Pendakur (2013). Next, we show how to recover preference heterogeneity contained within various observable characteristics. We introduce a measure for the degree of preference heterogeneity that is contained in each observable characteristic, relative to the heterogeneity contained in the full sample. In essence, this requires repeatedly applying Cosaert's solution algorithm to observed subsets of the data, constructed on the basis of observable demographic information. We automated all our following procedures in Python and CPLEX; and we refer to Appendix A.1 for more details and guidelines on how to access our code.

2.1 Exploring Preference Heterogeneity

Suppose we have data on N observations of consumer behavior. For each observation $i = 1, \ldots, N$, we observe a strictly positive vector of prices $\boldsymbol{p_i} \in \mathbb{R}^{M}_{++}$ and a non-negative vector of associated quantities $\boldsymbol{q_i} \in \mathbb{R}^{M}_{+}$. Taken together, $S = \{(\boldsymbol{p_i}; \boldsymbol{q_i}), i = 1, 2, \ldots, N\}$ represents the set of all observations under study. In line with our empirical application (which considers cross-sectional data; see Sections 3 and 4), we refer to each i as a specific individual, which we assume to behave rationally:[4]

$$\boldsymbol{q_i} = \arg\max_{\boldsymbol{q}} \; U_i(\boldsymbol{q}) \quad \text{s.t.} \;\; \boldsymbol{p_i'q} \leq m_i,$$

with $U_i(.)$ representing individual i's utility function and $m_i = \boldsymbol{p}_i'\boldsymbol{q_i}$ denoting total expenditure. In other words, rationality of S ensures that there exist utility functions $U_i(.)$ that solve the utility maximization problems for all individuals, without specifying the number of utility functions required to do so. Crawford and Pendakur (2013) and Cosaert (2019) propose to bound the number of utility functions that can rationalize the data in the following sense:

Definition 1 (τ-Rationalizability)**.** The set $S = \{(\boldsymbol{p_i}; \boldsymbol{q_i}); i = 1, \ldots, N\}$ is τ-rationalizable if, for each i, there exists a utility function $U_t(.)$ (with $t \in \{1, ..., \tau\}$) such that $\boldsymbol{q_i}$ solves the following optimization problem:

$$\boldsymbol{q_i} = \arg\max_{\boldsymbol{q}} \; U_t(\boldsymbol{q}) \quad \text{s.t. } \boldsymbol{p}_i'\boldsymbol{q} \leq m_i.$$

The number of utility functions needed to rationalize S must lie between 1 and N: If there exists a single utility function that can rationalize the choice behavior of all individuals then we can set $\tau = 1$; conversely, if the choices of any two individuals cannot be rationalized by a single utility function, then $\tau = N$ (i.e. each individual is characterized by her own unique utility function). Naturally, if τ utility functions rationalize the whole sample, then any larger number of utility functions (up to N) could also rationalize the sample. Cosaert (2019)'s procedure searches for the minimum number $\hat{\tau}$ of different utility functions required to find support for the utility maximization hypothesis. It makes use of the following two concepts:

Definition 2 (Partition)**.** Given the set $S = \{(\boldsymbol{p_i}; \boldsymbol{q_i}); i = 1, \ldots, N\}$, a combination of subsets $V_1, V_2, \ldots, V_\tau$ is a partition of S if:

1. $\forall i \in \{1, ...N\}$ we have $i \in V_1 \cup V_2 \cup \cdots \cup V_\tau$,

2. $\forall s, t \in \{1, ..., \tau\} : s \neq t$ we have $V_s \cap V_t = \emptyset$.

Definition 3 (WARP-partition)**.** A partition $(V_1, \ldots, V_\tau)$ is a WARP-partition of S if and only if $\forall t \in \{1, ..., \tau\}$ and $\forall i, j \in V_t$ WARP is satisfied:

$$\boldsymbol{p}_i'\boldsymbol{q}_i \geq \boldsymbol{p}_i'\boldsymbol{q}_j \Rightarrow \boldsymbol{p}_j'\boldsymbol{q}_j < \boldsymbol{p}_j'\boldsymbol{q}_i.$$

In words, a partition assigns each individual in S to exactly one of τ non-overlapping subsets $V_1, V_2, \ldots, V_\tau$. A WARP-partition is a partition such that for each subset V_t $(t \in 1, \ldots, \tau)$, any pair of individuals is consistent with WARP, so providing an empirical condition for τ-rationalizability of S. In other words, τ equals the cardinality (i.e. number of subsets $V_{(.)}$) of a given WARP-partition. In general, WARP constitutes a necessary condition for utility maximizing behavior (see Varian, 1982). However, WARP is also a sufficient condition in a two-goods setting (see Rose, 1958). In this respect, we indicate that our empirical application in Sections 3 and 4 effectively considers two-goods settings.

To recover $\hat{\tau}$ we search for the WARP-partition with minimum cardinality, dividing S in exactly $\hat{\tau}$ non-overlapping subsets $V_1, \ldots, V_{\hat{\tau}}$.[5] Each of these $\hat{\tau}$ subsets can be interpreted as a group of rational consumers sharing the same well-behaved preferences. In other words, $\hat{\tau}$ reveals the minimum number of unique preference types in S and can be understood as a measure of the underlying preference heterogeneity within the sample. In what follows, we will maintain that $\hat{\tau}$ denotes the baseline (or unconditional) preference heterogeneity contained in S. To illustrate, $\hat{\tau} = 1$ suggests that all individuals have compatible tastes (preferences) for the

given choice settings. Conversely, $\hat{\tau} > 1$ indicates that at least two individuals have conflicting tastes, such that at least two different preference types must be present in the sample.

Cosaert's method recovers $\hat{\tau}$ by relying on insights from graph theory.[6] Suppose we construct a graph G, where each vertex represents an individual and each edge connecting a pair of vertices indicates a violation of WARP. The chromatic number of the graph $\chi(G)$ gives the smallest number of colors (discrete numbers) necessary to obtain a labeling of the graph's vertices with colors, such that no two vertices sharing the same edge have the same color. Cosaert proves that $\chi(G)$ is equivalent to $\hat{\tau}$. Summarizing, the solution to $\chi(G)$ provides the minimum cardinality for a WARP-partition of the set S.

2.2 Recovering Preference Heterogeneity within Observable Characteristics

Suppose that S contains demographic information (such as age, gender, education, etc.) in addition to price and quantity information, such that observed consumer characteristics can be matched with the observed consumption choices.[7] In this case, unobserved preference heterogeneity arises when individuals with identical observable characteristics maximize different utility functions. In terms of the discussion above, this implies that similar-looking individuals are WARP-partitioned in different subsets $V_{(.)}$.

We next propose a method to recover the preference heterogeneity as identified through Cosaert (2019)'s partitioning method that is contained within these observable characteristics. Specifically, the method divides S into non-overlapping

subsets based on observable characteristics, and then computes the minimum cardinality for a WARP-partition of each subset. This obtains the minimum number of utility functions (or alternatively, preference types) that is necessary to rationalize the subset data. This information can then be used to quantify the degree of preference heterogeneity that is contained in the observable characteristics under evaluation.

Formally, we let K denote the set of observable characteristics of interest. For a given characteristic $k \in K$ with L_k possible states, we divide S into subsets $S_{k,1}, S_{k,2}, \ldots, S_{k,L_k}$ to define a k-partition:

Definition 4 (k-partition)**.** Given the set $S = \{(\boldsymbol{p_i}; \boldsymbol{q_i}); i = 1, \ldots, N\}$ and an observable characteristic k with L_k states, a combination of subsets $S_{k,1}, S_{k,2}, \ldots, S_{k,L_k}$ is a k-partition of S if:

1. $\forall i \in \{1, ..., N\}$ we have that $i \in S_{k,1} \cup S_{k,2} \cup \cdots \cup S_{k,L_k}$,

2. $\forall l, m \in \{1, ..., L_k\} : l \neq m$ we have $S_{k,l} \cap S_{k,m} = \emptyset$.

To take a specific example, let k be the observable characteristic gender with two states ($L_k = 2$): female and male; assuming that there are no other states observed in the sample. The gender-partition then divides S in two non-overlapping subsets $S_{k,1}$ and $S_{k,2}$, where $S_{k,1}$ comprises all female observations and $S_{k,2}$ all male observations.

Using a similar logic as before, for each subset $S_{k,l}$ corresponding to state $l \in \{1, \ldots, L_k\}$ we can run Cosaert's algorithm to compute $\tau_{k,l}$ as the chromatic number of the associated graph $G_{k,l}$; this chromatic number represents the minimum number of unique preference types needed to establish a WARP-partition

of the subset $S_{k,l}$. Naturally, the number $\tau_{k,l}$ is bounded by the following two extreme cases. If no two observations belonging to any $S_{k,l}$ violate WARP, $\tau_{k,l}$ attains a lower bound of 1 for each state l. This indicates that partitioning by k fully describes the identified preference heterogeneity contained in S. By contrast, if any two observations belonging to any $S_{k,l}$ violate WARP, then, for every state l, the number $\tau_{k,l}$ attains its upper bound, which equals the number of observations in $S_{k,l}$. In this case, we can conclude that the characteristic k does not describe any preference heterogeneity in S.

To quantify the preference heterogeneity that is contained in characteristic k, we make use of the measure τ_k:

$$\tau_k = \sum_{l=1}^{L_k} \tau_{k,l}.$$

This measures the minimum number of unique preference types that is required to obtain a WARP-partition for all subsets $S_{k,l}$ associated with k; it effectively represents the preference heterogeneity contained in the data set S when using the k-partition under study. Thus, recalling that $\hat{\tau}$ represents the preference heterogeneity that is contained in the original (unpartitioned) data set S, we can quantify how much preference heterogeneity is contained in k by comparing τ_k to $\hat{\tau}$. If τ_k only marginally exceeds $\hat{\tau}$, then partitioning by k adds little preference heterogeneity to the baseline heterogeneity contained in S. This suggests that a considerable part of this baseline heterogeneity can be resolved by grouping individuals on the basis of characteristic k. By contrast, if τ_k is much larger than $\hat{\tau}$, then the baseline heterogeneity contained in S is largely unrelated to the partitioning variable.

Following this reasoning, we can use the κ_k-ratio to quantify how much ("un-

observed"/"undescribed") preference heterogeneity in S is left after conditioning on the observable characteristic k:

Definition 5 (κ_k-ratio)**.** For a given k-partition of S, the κ_k-ratio is defined as:

$$\kappa_k = \frac{\tau_k}{\hat{\tau}}.$$

By construction, the κ_k-ratio is bounded between 1 and L_k; and lower values indicate that more heterogeneity is described by the k-partition.[8] We conclude that the observable characteristic $k \in K$ with the lowest κ_k-ratio is most informative in describing the baseline preference heterogeneity. At this point, we acknowledge that there may well exist other ways to meaningfully quantify the preference heterogeneity that is described by observable characteristics. We choose to make use of κ_k-ratios to select most informative observables because we feel they have a natural interpretation (as explained above); we further illustrate the intuition of this procedure through a stylized example in Section 2.4. We leave the exploration of alternative selection procedures (e.g. in view of desirable axiomatic properties) as an interesting avenue for follow-up research.

2.3 Branching

In case multiple characteristics k obtain identical (or closely similar) κ_k-ratios, or if there still is substantial preference heterogeneity after conditioning on a first characteristic k, we may need to include an additional layer in the analysis, which we refer to as branching. Specifically, a level-1 branching boils down to finding the κ_k-ratios for all $k \in K$, using the k-partition method described above. Correspondingly, we refer to characteristic k as the level-1 branching variable. A

level-2 branching applies the same partition logic, but separately to each subset $S_{k,l}$ ($l \in 1, \ldots, L_k$) defined by the level-1 branch. More specifically, consider an additional characteristic $j \in K$ ($j \neq k$) with L_j possible states as the level-2 branching variable. We apply the k-partition method onto each subset $S_{k,l}$: we divide $S_{k,l}$ in L_j different subsets $S_{k,1|l}, S_{k,2|l}, \ldots, S_{k,L_j|l}$, where every $S_{k,h|l}$ ($h \in 1, \ldots, L_j$) is the subset of $S_{k,l}$ that contains the individuals attaining state h for the level-2 branching variable j; and we compute $\tau_{k,h|l}$ as the chromatic numbers of the corresponding graphs $G_{k,h|l}$.

Summing these chromatic numbers $\tau_{k,h|l}$ over all L_j states of the level-2 branching variable j, and then summing over all L_k states of the level-1 branching variable k, obtains the minimum number of preference types when simultaneously conditioning on the two characteristics k and j, which we denote by $\tau_{j|k}$:

$$\tau_{j|k} = \sum_{l=1}^{L_k} \sum_{h=1}^{L_j} \tau_{k,h|l}.$$

Conducting this exercise for each observable characteristic $j \in K$ ($j \neq k$) completes the level-2 branching of characteristic k. Similar to above, this then allows us to compute $\kappa_{j|k}$-ratios, which quantify the degree to which a specific level-2 branching (for level-1 branching variable k and level-2 branching variable j) describes the baseline preference heterogeneity contained in S:

Definition 6 ($\kappa_{j|k}$-ratio)**.** For a given level-2 branching of S, with level-1 branching variable k and level-2 branching variable j, the $\kappa_{j|k}$-ratio is defined as:

$$\kappa_{j|k} = \frac{\tau_{j|k}}{\hat{\tau}}.$$

Every $\kappa_{j|k}$-ratio relates the preference heterogeneity when partitioning on the basis of the observable characteristics k and j to the preference heterogeneity contained in the original (unpartitioned) data set S. Like before, we are then interested in finding the lowest $\kappa_{j|k}$-ratio among all possible pairs of observable characteristics, which corresponds to the most informative level-2 branching of S.[9] Finally, remark that for any given pair of characteristics k and j the order of the level-2 branching does not impact the $\kappa_{j|k}$-value (i.e. $\kappa_{j|k} = \kappa_{k|j}$), as changing the ordering of the partitioning variables does not affect the composition of the subsets.

2.4 Further discussion

We can further sharpen the intuition for using κ_k-ratios to select most informative observables through a stylized example. Let us consider level-1 branching for simplicity. Assume a set S with preference heterogeneity fully driven by the characteristic gender (with two states: female and male); this corresponds to $\hat{\tau} = 2$ (i.e. two gender types) reflecting the baseline heterogeneity contained in S. Of course, in practice the empirical analyst does not know that this baseline preference heterogeneity is truly driven by gender. Now suppose that the analyst considers two observable characteristics as candidates to describe the identified baseline heterogeneity: the "correct" characteristic gender and, in addition, the characteristic employment (with three states: student, employed, not employed) which does not directly drive consumer heterogeneity (but this is unknown to the analyst). Finally, assume that it happens to be the case that all females in S are students and all males are either employed or unemployed.

Let us use the notation G for the characteristic gender (with m and f for the states female and male) and E for the characteristic employment (with s, n and e for the states student, not employed and employed). Thus, we have $K = \{G, E\}$, $L_G = 2$ and $L_E = 3$. For the given set-up, our method obtains for gender:

$$\kappa_G = \frac{\tau_G}{\hat{\tau}} = \frac{\tau_{G,f} + \tau_{G,m}}{\hat{\tau}} = \frac{1+1}{2} = 1,$$

and for employment:

$$\kappa_E = \frac{\tau_E}{\hat{\tau}} = \frac{\tau_{E,s} + \tau_{E,n} + \tau_{E,e}}{\hat{\tau}} = \frac{1+1+1}{2} = 1.5.$$

As $\kappa_G < \kappa_E$, we correctly conclude that the observable characteristic gender (and not employment) is most informative in describing the baseline preference heterogeneity. In fact, we obtain that partitioning the set S on both gender and employment yields (respectively 2 and 3) subsets that all satisfy WARP. Nonetheless, our method leads us to label the characteristic gender as more informative than the characteristic employment. Intuitively, the κ_k-ratio "penalizes" the characteristic employment for adding complexity (i.e. one additional state) that does not contribute to describing the revealed preference heterogeneity in S.[10]

Two final remarks are in order. First, for any set S (and every k and j) we will have by construction:

$$\hat{\tau} \leq \tau_k \leq \tau_{j|k},$$

and thus:

$$1 \leq \kappa_k \leq \kappa_{j|k}.$$

The reasoning goes as follows. Every extra layer of branching (e.g. from baseline to level-1 and from level-1 to level-2) implies an additional partitioning of the set S. This automatically increases the number of subsets for which we need to calculate the minimum number of preference types that is required for a WARP-partition. In turn, this implies $\hat{\tau} \leq \tau_k$ and $1 \leq \kappa_k$ (when going from baseline to level-1) and $\tau_k \leq \tau_{j|k}$ and $\kappa_k \leq \kappa_{j|k}$ (when going from level-1 to level-2).

Finally, in practical applications we can graphically visualize the structure of our branching procedure by making use of dendograms. Figure 1 provides a level-2 branching example that closely mimics the structure of our own empirical application in Section 4. We first calculate the minimum cardinality of all possible WARP-partitions of the sample S. Next, we select an observable characteristic (here: gender) and k-partition the sample according to the possible states of gender (here: female, male). For each of these two subsets we then calculate the minimum number of preference types required to obtain a WARP-partition. To illustrate level-2 branching we select an additional observable characteristic (here: employment), and k-partition each of the two gender subsets by the three possible states of employment (here: student, employed, not employed). For each of the six newly created (gender, employment) subsets, we can then calculate the minimum cardinality of their WARP-partitions, which allows us to compute the associated $\kappa_{j|k}$-ratios.

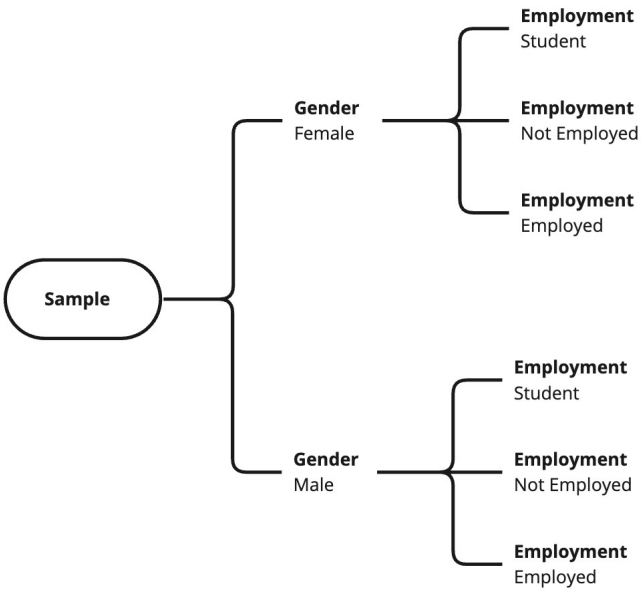

Figure 1: 1-level and 2-level branching dendogram

3 Experimental design

The experiments we devised confronted participants with two different, yet related, experimental designs.[11] First, respondents participated in a time allocation task, requiring them to allocate a given time endowment over leisure and work. The second experiment consisted of a time allocation task, which required respondents to allocate a given budget over two different products. In both experimental designs, we aimed to mimic, as closely as possible, a real-life setting known to participants. Specifically, we framed both tasks as decisions that participants might face on a daily basis, to which we added proper financial incentives. We ensured incentivization through a lottery that randomly selected a total of 10 participants and paid them their actual earnings. Incentivization constitutes an integral part of our experiment, as the use of financial rewards that are contingent on behav-

ior is a stringently enforced rule in experimental economics, which is understood to be a means to maintain strict control over incentives (Loewenstein, 1999). To ensure that incentives were compatible with our experimental design, it was announced at the beginning of the experiment that only participants who completed both tasks could take part in the lottery. The experiment was run entirely online through Qualtrics. Online experiments are increasingly popular in economics and are known to yield parallel results to laboratory experiments even in interactive settings where participants' actions impact their payoffs (Arechar et al., 2018). To prevent cheating and gerrymandering, we also implemented an IP-address based blocking, so that each participant could take part in the experiment only once. Lastly, email addresses were collected to contact participants for payment and it was announced that their data would remain anonymized once the payments were realized.

3.1 Time Allocation Task

For the time allocation task we implement a textbook consumption/labor supply model in which consumption takes the form of a Hicksian aggregate good. The experimental design resembles Kool and Botvinick (2014), as participants are given an endowment in seconds to divide between one relaxing and one mentally exhausting activity. The relaxing activity represents leisure, whereas the mentally exhausting activity represents work. For each second spent on the mentally exhausting activity, participants earn a determined wage. In our experiment, the work activity consists of typing different names that start with a specific letter of the alphabet. The leisure activity consists of watching a compilation of funny cat

videos.

After reading the instructions (which are shown in Appendix A.2.2), participants are informed about their time endowment (in seconds) together with their per second wage (expressed in eurocents). As soon as participants start the task, a timer showing the time remaining begins to count down. Participants may switch between the two activities at any time and are taken to the next task automatically once their timer reaches zero.

Participants are randomly assigned one of the 11 different budget and wage combinations shown in Table 1. Figure 2 shows the associated budget constraints, where the budgets represented by a dotted line are removed in a sensitivity analysis that we include in Appendix A.5. The choices made by participants during this task have real monetary outcomes: 5 randomly selected participants were paid for their time spent at the mentally exhausting activity, according to their randomly assigned per second wage. Appendix A.3.1 shows the maximization problem that participants faced during this task. Under rationality, participants' observed behavior allows us to reveal their time use preferences.

As a final remark, we indicate that the time allocation task considers two more budget-wage combinations than the money allocation task that we describe next (i.e. 11 budgets versus 9 budgets). By doing so, we can assess whether the empirical results produced by our method are sensitive to the number of budgets that is considered. In Appendix A.5 we compare the findings for the time allocation setting with 11 budgets (which will be our core exercise) and for a more restricted setting with only 9 budgets (where we arbitrarily exclude the budgets 6 and 8). Comfortingly, this sensitivity check confirms that our main qualitative findings are largely robust.

Budget	Endowment (in seconds)	Wage per second (in cents)	Consumption possibility
1	120	6	720
2	120	2	240
3	180	4	720
4	180	2	360
5	225	4	900
6	225	1	225
7	180	1	180
8	300	2	600
9	240	2	480
10	90	5	450
11	240	3	720

Table 1: Allocation choices in the Time Allocation Task

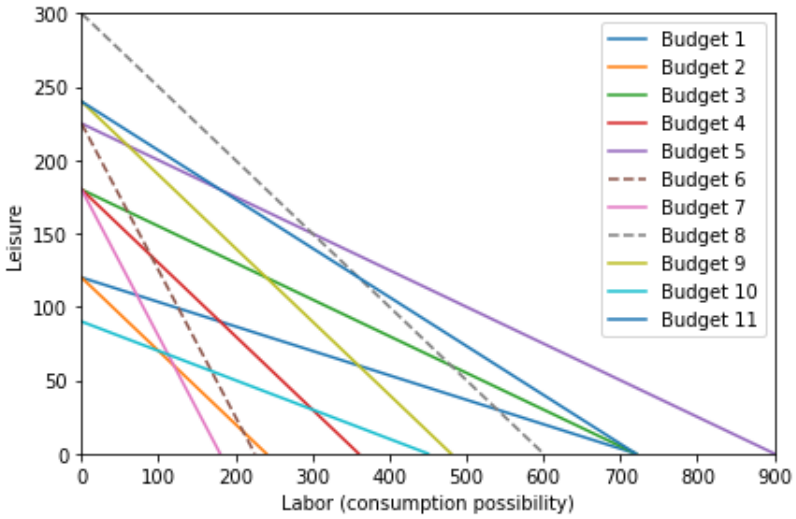

Figure 2: Budget constraints in the Time Allocation Task

3.2 Money Allocation Task

For the money allocation task, we implement a simple two-goods preference maximization model. We used vacation and shopping as goods to imitate a realistic trade-off that participants might face. The allocation task is as follows: After reading the instructions (which are shown in Appendix A.2.3), participants were informed about their budgets and the unit prices of the two goods; and they have to divide their budgets between the two goods by moving a slider. After finishing the allocation task, they are presented with the resulting quantities of each good they have chosen, and with the allocation of their budget in percentage points. Participants are then asked to confirm their allocation to finish the task; they have the possibility to change their initial allocation if wanted. Participants may adjust their preferred decision as many times as they like before finishing the task.

Similar to before, prices and endowments are randomly assigned to the participants, taking possible values as shown in Table 2. Figure 3 shows the associated budget constraints. Participants' final choices again have real monetary outcomes: 5 randomly selected participants were awarded vacation and/or shopping vouchers worth €20, according to their chosen allocations. Appendix A.3.2 shows the maximization problem that participants faced during this task. Assuming rational decision-making, we expect participants to choose their most preferred bundle from the set of affordable alternatives, which allows us to identify their consumption preferences with respect to vacation versus shopping.

Budget	**Endowment (in points)** $\bar{m}$	**Price of the Vacation good** p_1	**Price of the Shopping good** p_2
1	40	1	3
2	40	3	1
3	60	1	2
4	60	2	1
5	75	1	2
6	75	2	1
7	60	1	1
8	40	1	4
9	40	4	1

Table 2: Allocation choices in the Money Allocation Task

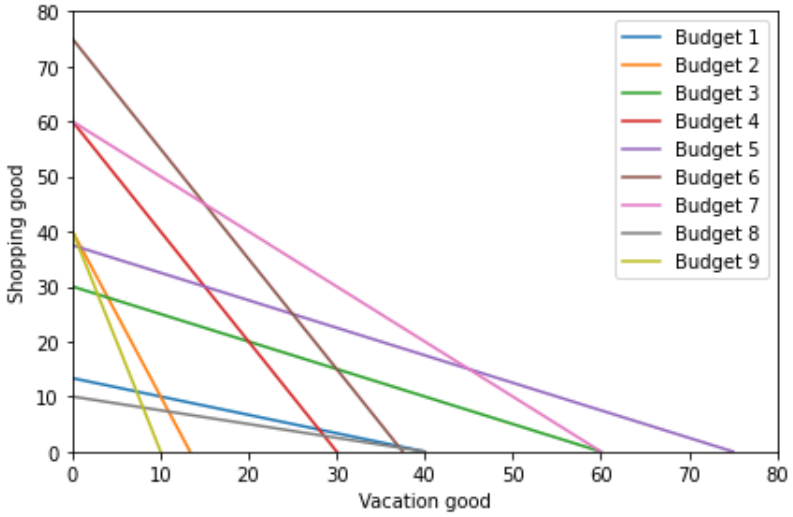

Figure 3: Budget constraints in the Money Allocation Task

3.3 Data

In total, 148 respondents correctly completed both the time allocation task and the money allocation task. Additionally, 35 respondents completed only the time allocation task, providing a maximum sample size of 183 respondents. In addition to respondents' choice behavior, we collected demographic information on age, gender, education, marital status and employment status. Table 3 describes the composition of our sample.[12]

Coincidentally, we obtain an almost even split between males and females. As we specifically targeted university students, we obtain a relatively young sample, although some older respondents also filled in the questionnaire. Next, the sample is characterized by significant heterogeneity in respondents' education levels, with a higher fraction of female respondents having obtained a bachelor's or master's degree. Further, the large majority of both males and females in our sample have never married. Finally, most respondents are from Turkey or Belgium.

Next, Table 4 documents the individuals' choice behavior, for both the time and money allocation task.[13] The first column shows the average fraction of time spent on the labor activity, and the second column shows the average budget share for the shopping good. Both tasks show considerable heterogeneity in the observed choice behavior. For example, for the money allocation task, the average budget share allocated to shopping varies from 39.9% (budget 9) to 67.5% (budget 1). Next, the reported standard deviations indicate significant dispersion of budget shares within budgets. Thus, even when different respondents face exactly the same budget and price conditions, they often choose widely different consumption allocations.[14]

Summarizing, our sample exhibits substantial heterogeneity in both the observable consumer characteristics and the selected (time and money) allocations. In the next section, we will investigate the link between these two observations. Particularly, we will study whether the observed variation in choice behavior can be described through observable characteristics driving the variation in revealed preferences.

Table 3: Description of the sample; absolute (and relative) frequencies for different categories of consumers

	Female N = 93	**Male** N = 90	**Overall** N = 183
Age			
17 to 22	31 (33.3%)	34 (37.8%)	65 (35.5%)
23 to 27	45 (48.4%)	33 (36.7%)	78 (42.6%)
28 or above	17 (18.3%)	23 (25.6%)	40 (21.9%)
Education			
Bachelor's degree (e.g. BA, BS)	37 (39.8%)	26 (28.9%)	63 (34.4%)
Lower than bachelor	26 (28.0%)	37 (41.1%)	63 (34.4%)
Master's or higher	30 (32.3%)	27 (30.0%)	57 (31.1%)
Marital Status			
Couple	14 (15.1%)	17 (18.9%)	31 (16.9%)
Single	79 (84.9%)	73 (81.1%)	152 (83.1%)
Employment			
Employed	23 (24.7%)	23 (25.6%)	46 (25.1%)
Not Employed	13 (14.0%)	13 (14.4%)	26 (14.2%)
Student	57 (61.3%)	54 (60.0%)	111 (60.7%)

Table 4: Average budget share per task (SDs in parenthesis)

Budget	Time Allocation (% of labor time)	Money Allocation (% of shopping budget)
1	76.8	67.5
	(34.6)	(22.6)
2	77.2	45.4
	(28.8)	(32.9)
3	74.2	44.2
	(35.0)	(27.4)
4	74.0	52.5
	(29.7)	(23.0)
5	64.8	55.9
	(31.8)	(31.3)
6	71.3	46.4
	(32.4)	(29.0)
7	79.5	46.2
	(24.5)	(17.9)
8	77.0	61.9
	(28.1)	(30.3)
9	66.9	39.9
	(30.5)	(23.4)
10	86.2	
	(23.0)	
11	81.1	
	(31.2)	

4 Results

We illustrate our methodology by applying it to the experimental data on time and money allocations that we described in the previous section. The exposition below reports κ_k-ratios and $\kappa_{j|k}$-ratios for various level-1 and level-2 branchings. As explained in Section 2, these ratios measure the minimum increase in unobserved preference heterogeneity associated with a given branching relative to the baseline heterogeneity in the sample. Our interest lies in recovering the (pair of) observable characteristic(s) that obtain(s) the division of the data sample that maximizes the preference heterogeneity that is described by observable characteristics. Therefore, we are interested in finding the branching that reports the lowest ratios for each task, so obtaining the characteristics that are most informative in describing preference heterogeneity. For clarity, these lowest ratios will be highlighted in bold in the tables below.

Our discussion in Section 3 highlighted the heterogeneity that is present in our data, both in terms of observable characteristics and respondents' choice behavior. This motivates our analysis below, which aims to assess whether respondents' choice behavior can be rationalized by partitioning the sample on the basis of the consumers' demographic information. Moreover, we will investigate whether the same observable characteristics can describe the observed heterogeneity in consumption decisions (i.e. money allocation task) and time decisions (i.e. time allocation task). Table 5 presents an overview of the possible states for all observable characteristics that we consider.

In what follows, we first discuss the results of our "core exercise" that uses 11 (9) budgets for the time (money) task, reporting the outcomes of both level-

1 and level-2 branchings. Subsequently, we present the results of a subsampling procedure that we can use to prevent overfitting the data. Finally, we suggest a permutation approach to address statistical testing of the independence between observed heterogeneity in consumer choices and observable consumer characteristics.

Table 5: Overview of states

Observable characteristic	Possible states
Age	17-22y, 23-27y, 28y+
Education	Lower than bachelor's, Bachelor's degree, Master's degree or higher
Employment	Student, Employed, Not Employed
Gender	Male, Female
Marital Status	Couple, Single

4.1 Full Sample Results

Our core analysis considers a total of 148 observations for both the time allocation task (11 budgets) and the money allocation task (9 budgets). The baseline heterogeneity contained in the full sample (denoted by $\hat{\tau}$ in Section 2) amounts to 19 and 21 preference types for the money and time allocation task, respectively.[15] Table 6 reports the κ_k-ratios of a level-1 branching for all observable characteristics. These ratios are computed by running Cosaert (2019)'s solution algorithm for each subset of the k-partition in question.[16] For completeness, Table 7 reports the number of unique budgets that are faced by consumers in each state that we evaluate. It appears that there is very little variation across characteristics and states: consumers always face all or all but one of the budgets under consideration. In our opinion, this makes it so that we may reasonably argue that unequal distributions of budgets across characteristics and states does not drive our empirical

results.

For both tasks, k-partitioning by marital status obtains the lowest κ_k-ratios. Specifically, for the money allocation task, sorting by marital status increases the baseline heterogeneity by around 10%, which implies 2 ($\approx 0.105 * 19$) additional preference types. Similarly, 1 ($\approx 0.048 * 21$) supplementary preference type is required to establish rationalizability for the time allocation task.[17] Moreover, for both tasks, education and age are least informative in describing choice variation. One possible explanation is that the majority of respondents in our sample are students of the same age group, which may well imply too little variation in these characteristics to describe the observed heterogeneity in choice behavior. Further, it turns out that money use preferences are more diversified than time use preferences, as indicated by the higher κ_k-values.[18]

Table 6: κ_k-ratios for level-1 branchings; full sample

	Time Allocation	Money Allocation
Marital Status	**1.048**	**1.105**
Gender	1.095	1.368
Employment	1.143	1.316
Education	1.143	1.474
Age	1.286	1.421

A priori, one may have expected the optimal level-2 branching to combine the two most informative characteristics for the level-1 branching. However, Table 8 shows that this need not be the case.[19] More specifically, the optimal level-2 branchings for the time and money allocation tasks correspond to, respectively, marital status and employment and marital status and age. Thus, both level-2 branchings include the optimal level-1 branching variable marital status. Interestingly, while age was one of the least informative characteristics when conducting a

Table 7: Number of unique budgets per state; full sample

	Time Allocation	**Money Allocation**
Gender		
Female	11	9
Male	11	9
Age		
17 to 22	11	9
23 to 27	11	9
28 or above	10	9
Education		
Bachelor's degree	11	9
Lower than bachelor	11	9
Master's or higher	10	9
Marital Status		
Couple	10	8
Single	11	9
Employment		
Employed	11	9
Not Employed	10	8
Student	11	9

level-1 branching for the money allocation task, it turns out to be most informative for level-2 branching. By contrast, for the time allocation task, age performed worst for both level-1 branching and level-2 branching. These findings seem to suggest that different characteristics drive heterogeneity in money use and time use preferences.

Table 8: $\kappa_{j|k}$-ratios for level-2 branchings; full sample

	Time Allocation	Money Allocation
Marital Status and Employment	**1.286**	1.579
Gender and Marital Status	1.381	1.632
Gender and Education	1.429	1.789
Gender and Employment	1.429	1.895
Education and Marital Status	1.381	1.579
Education and Employment	1.381	1.789
Age and Marital Status	1.429	**1.526**
Age and Employment	1.476	1.895
Age and Education	1.571	1.684
Age and Gender	1.714	2.000

4.2 Subsampling Results

In practical applications, one of the pitfalls of using WARP when checking for rationality is its binary outcome: either the data satisfy WARP or they do not.[20] Implementing such binary conditions on increasingly smaller subsets (such as in level-2 branching) makes our method vulnerable to "overfitting" the observed data.[21] To prevent this, we may resort to a commonly used machine learning method that applies repeated subsampling without replacement (Mohri et al., 2018; Alpaydin, 2020).[22] More specifically, we verify the statistical robustness of our above findings by generating 200 subsamples drawn from the original data. There is no a priori rationale for determining subsample size. For our illustrative exercise, we choose to set the size equal to 80% of the original sample size. Evidently, in more

elaborate applications, one may well experiment with alternative subsample sizes.

This subsampling procedure helps to mitigate the possible effects of outliers or measurement error. From a practical viewpoint, it may also help to differentiate between branchings yielding identical κ_k-ratios when using the full sample. Table 9 reports κ_k-ratios for the various level-1 branchings, averaged over the 200 subsamples, together with their standard deviations. Interestingly, for both tasks, the κ_k-ratios reported in Table 9 closely resemble those in Table 6, both in terms of ordering and in terms of magnitude. Next, for the time allocation task the subsampling procedure breaks the tie in κ_k-ratios for employment and education, in favor of employment.

Table 9: Mean κ_k-ratios (SDs in parenthesis) for level-1 branchings; subsampling

	Time Allocation	Money Allocation
Marital Status	**1.093**	**1.133**
	(0.067)	(0.081)
Gender	1.142	1.334
	(0.108)	(0.117)
Employment	1.177	1.315
	(0.095)	(0.105)
Education	1.192	1.470
	(0.085)	(0.107)
Age	1.329	1.437
	(0.123)	(0.090)

When applying the same subsampling procedure for the level-2 branchings, we obtain the averages and standard deviations that are reported in Table 10. In essence, our earlier findings remain unchanged: the combination marital status and employment is optimal for the time allocation task, while the combination marital status and age is optimal for the money allocation task. Moreover, the κ_k-ratios reported for the subsampling exercise again closely resemble the ratios that we computed for the full sample. However, because of variation across subsamples we

do observe some (minor) changes in the ordering of the κ_k-ratios when comparing with the full sample outcomes. Lastly, the standard deviations for the level-2 branchings are somewhat higher than those obtained for the level-1 branchings.

Table 10: Mean $\kappa_{j|k}$-ratios (SDs in parenthesis) for level-2 branchings; subsampling

	Time Allocation	Money Allocation
Marital Status and Employment	**1.365**	1.614
	(0.102)	(0.123)
Gender and Marital Status	1.403	1.604
	(0.155)	(0.135)
Gender and Education	1.483	1.808
	(0.135)	(0.143)
Gender and Employment	1.500	1.874
	(0.150)	(0.157)
Education and Marital Status	1.435	1.643
	(0.105)	(0.102)
Education and Employment	1.485	1.868
	(0.120)	(0.129)
Age and Marital Status	1.495	**1.569**
	(0.135)	(0.096)
Age and Employment	1.532	1.901
	(0.145)	(0.143)
Age and Education	1.631	1.747
	(0.133)	(0.114)
Age and Gender	1.745	1.942
	(0.155)	(0.153)

4.3 Towards statistical testing: a permutation approach

We conclude our empirical application by suggesting a method to assess the statistical significance of the dependence between the observed heterogeneity in consumer choices and observable consumer characteristics. Particularly, it may well be that a given k-partitioning of the observed consumer behavior corresponds to a low κ_k-ratio simply by coincidence, while preference heterogeneity in reality is actually independent of the characteristic k under study. Basically, this means that our partitioning methodology lacks power to effectively identify this independence

for the observed behavior under evaluation. We will argue that a permutation approach may deal with such concerns.

At this point, it is important to acknowledge that our following analysis should at best be conceived as a "proof of concept application". Most notably, formally showing the validity of the permutation procedure for statistical testing purposes requires a careful exploration of its theoretical properties. Nonetheless, we do believe that the procedure does have an attractive intuition; and, therefore, we see a formal treatment of the procedure's properties as a potentially interesting avenue for follow-up research.[23] For compactness, our following exposition will only consider marital status, which is the observable characteristic that corresponds to the optimal level-1 branching for our full sample (for both the time and money allocation task). Evidently, the proposed approach could equally well be applied to any other characteristic, any level-2 branching or any other exercise that we conducted above.

The permutation approach proceeds as follows. We randomly assign a marital state to each consumer observation by drawing without replacement from the observed marital status distribution (as shown in Table 3). Intuitively, this random assignment reflects the null hypothesis that the observed heterogeneity in consumer choices is independent of the characteristic marital status. For this newly constructed sample, we then compute the κ_k-ratio for a level-1 branching based on marital status. Repeating this procedure 200 times produces a counterfactual distribution of κ_k-ratios under the null of independence, against which we can compare the κ_k-ratios of the full sample (reported in Table 6).[24]

Columns 2-6 of Table 11 report descriptive statistics of this counterfactual distribution for the money and time allocation tasks.[25] The last column of the

table reports the percentiles of the κ_k-ratios for the level-1 branching based on marital status that we obtained for the full sample. We observe that the full sample κ_k-ratios belong to the first quartiles of the distributions for both tasks. Specifically, 19% and 16% of the permuted samples obtain a κ_k-ratio below that for the full sample for the time and money allocation tasks, respectively.

In a following step, these results may then be used for a nonparametric permutation test of the null hypothesis of independence. To illustrate this for our empirical application: given that the κ_k-ratios for marital status are in the 19-th and 16-th percentiles of the (simulated) distributions under independence (for the time and allocation money tasks, respectively), we can only reject the null for significance levels that are substantially higher than the standardly used 5% or 10%. Arguably, the fact that we cannot significantly reject independence for our data set may well be due to our rather small sample (with only 148 consumer observations). Summarizing, when concluding that our partitioning method indicates marital status as a main driver of consumer heterogeneity for the sample at hand, we must cautiously add that our permutation approach actually does not allow us to statistically reject independence at reasonable significance levels.

Table 11: κ_k-ratio distribution under independence for a level-1 branching based on marital status; permutation approach

	κ_k-ratio under independence: Simulated distribution obtained through permutation					κ_k-ratio full sample
	Minimum	25th Percentile	Median	75th Percentile	Maximum	
Time Allocation	1	1.048	1.095	1.143	1.286	19-th percentile
Money Allocation	1	1.105	1.158	1.211	1.316	16-th percentile

5 Conclusion

We have presented a revealed preference methodology to identify preference heterogeneity from cross-sectional data. Our method builds on the WARP-based partitioning procedure that was developed by Cosaert (2019). It is aimed at quantifying the contribution of observable consumer characteristics to describing the identified preference heterogeneity. This allows us to identify the observable characteristics that principally drive the observed heterogeneity in consumer choice behavior. For empirical applications, we suggest a subsampling procedure to accommodate for the risk of overfitting the observed data that follows from the binary nature of the WARP requirement for rational consumer behavior. Furthermore, we indicated that our method may provide a fruitful basis for a permutation approach to statistically test the independence between observed heterogeneity in consumer choices and observable consumer characteristics. Attractively, our method is intrinsically nonparametric, making it robust to functional specification error. Moreover, it is conceptually simple and easy-to-implement, which is convenient from a practitioner's point of view.

We have also demonstrated the empirical usefulness of our method through an application to newly gathered experimental data on consumer choice behavior. In particular, we considered two types of decision situations: the allocation of money (choosing between two products) and the allocation of time (choosing between leisure and work). By comparing the results for these two experiments, we can investigate whether the same consumer characteristics describe the heterogeneity in choice behavior in the two settings. When considering only a single observable characteristic, we single out marital status as the main driver of heterogeneity in

both settings. When combining observable characteristics, we find that marital status and employment are important drivers of heterogeneity in the time allocation task, whereas marital status and age are most relevant in the money allocation task.

Acknowledgements

We thank the Editor Daniel Gottlieb and two anonymous referees for insightful comments. We are also grateful to Sam Cosaert and Khushboo Surana for useful discussion and feedback. Laurens Cherchye thanks the Fund for Scientific Research Flanders (FWO) and the Research Fund KU Leuven for financial support. Research ethics approval was obtained by the KU Leuven Privacy and Ethical Review (reference number: G-2021-3988). The views expressed are those of the authors and do not necessarily reflect those of the National Bank of Belgium.

Notes

[1]In follow-up work, Surana (2022) proposed a revealed preference method to quantify the degree of unobserved preference heterogeneity between different individuals in a panel data setting. This method makes use of partitioning techniques that are closely similar to those used by Cosaert (2019). As such, Surana's method may fruitfully be integrated with our method to quantify the degree of preference heterogeneity between groups of observably homogeneous consumers in a cross-sectional setting.

[2]WARP is a criterion for rational (i.e. utility maximizing) consumer behavior that is somewhat weaker than GARP. The main difference between WARP and GARP is that GARP additionally imposes transitivity of preferences. Cherchye et al. (2018) characterize the conditions on the observed choice behavior (prices and quantities) under which transitivity adds bite to the

empirical analysis. We refer to Section 2.1 for additional details on WARP.

[3]To be exact, Cosaert (2019) actually also proposed a characteristics-based cluster analysis that assigns individuals to the identified partitions (i.e. preference types). While this cluster analysis is equally based on observable consumer characteristics, it serves a very different purpose (i.e. clustering consumers in the cross-section versus quantifying the contribution of observable consumer characteristics to the observed heterogeneity in choice behavior).

[4]In general, the subscript i could either index different time periods (when using time series data) or different individuals (when using cross-sectional data).

[5]Remark that such a WARP-partition need not necessarily be unique.

[6]Graph theory is the study of graphs, which are mathematical structures used to model pairwise relations between objects. A graph in this context is made up of vertices (nodes) which are connected by edges (lines). We refer to Cosaert (2019) for a detailed discussion of the graph-theoretical concepts and tools that underlie his partitioning procedure.

[7]Throughout, we assume that S does not contain individuals with missing demographic information.

[8]To see why κ_k obtains an upper bound of L_k, consider a k-partition that splits the individuals in S into L_k different non-overlapping subsets $S_{k,l}$. Under the worst possible case the total baseline heterogeneity contained in S will be contained in each of the L_k subsets $S_{k,l}$. In this case $\tau_k = L_k * \hat{\tau}$, such that $\kappa_k = L_k$. This upper bound is easily realized by considering the case when the data set S is 1-rationalizable (in terms of Definition 1), i.e. there exists a single utility function that rationalizes the choice behavior of all individuals, which yields $\hat{\tau} = 1$. In this case, any k-partition yields $\tau_k = L_k$ by construction and, thus, the ratio measure $\kappa_k = L_k$.

[9]In principle, one may further compute higher level branchings but we will not pursue this further. In this regard, we remark that higher level branchings are possible but not necessarily desirable in smaller data sets as the number of individuals in each state quickly declines after level-2 branching.

[10]Remark that, if preference heterogeneity were fully driven by the characteristic employment (instead of gender), then an analogous reasoning would yield $\kappa_G = \kappa_E = 1$ when assuming exactly the same conditions (with all students being female and all not employed and employed being male). Particularly, we would get baseline heterogeneity $\hat{\tau} = 3$ (i.e. three employment

types) and:

$$\kappa_G = \frac{\tau_G}{\hat{\tau}} = \frac{\tau_{G,f} + \tau_{G,m}}{\hat{\tau}} = \frac{1+2}{3} = 1 \text{ and } \kappa_E = \frac{\tau_E}{\hat{\tau}} = \frac{\tau_{E,s} + \tau_{E,n} + \tau_{E,e}}{\hat{\tau}} = \frac{1+1+1}{3} = 1.$$

Thus, the "correct" characteristic employment is no longer penalized for having more states than gender. In this particular (stylized) situation, not only the characteristic employment but also gender attains the lowest possible κ_k-ratio (i.e. $\kappa_G = \kappa_E = 1$) because all females and males happen to be in distinctively different employment states (i.e. females are all student (with $\tau_{G,f} = \tau_{E,s} = 1$) while males are all either not employed or employed (with $\tau_{G,m} = \tau_{E,n} + \tau_{E,e} = 2$)). Intuitively, both characteristics contribute equally to describing the baseline preference heterogeneity when controlling for their number of states (i.e. $L_G = 2$ and $L_E = 3$).

[11]Our specification of the endowments and prices in the two allocation tasks (see Tables 1 and 2) is closely similar to the one used by Andreoni and Miller (2002) in a two-goods setting. We chose to introduce some variation in the endowment-price regimes faced by the consumers in the two choice settings. For the time allocation task, we use individual wage as the value of time. As explained in Section 3.2, this corresponds to describing time allocation choices in terms of a labor supply model. See, for example, Cherchye and Vermeulen (2008) for an empirical analysis of labor supply behavior (based on observational data) that makes use of nonparametric revealed preference methods.

[12]For education, "lower than bachelor's degree" contains "high school degree or equivalent" and "some college, no degree", and "master's degree or higher" contains "master's degree" and "doctorate or professional degree". For marital status, the single state combines the "widow" option (containing only 1 observation) with "single". For employment, we merged the options "employed full time", "employed part time" and "self-employed" into the state "employed", while the state "not employed" combines the options "unable to work", "unemployed and currently (not) looking for work" and "retired".

[13]Specifically, we list budgets for the two tasks in the order they appear in Tables 1 and 2. Participants were randomly assigned to budgets and did not necessarily face the budget pairs as they are presented in Table 4.

[14]Remark that we merely consider a simple two-goods setting. Increasing the dimension of

goods would likely further increase choice dispersion.

[15]The reported difference in baseline heterogeneity may, in part, result from variation in the underlying budget constraints for both tasks. Specifically, as appears from Figures 2 and 3, the budget lines in the time allocation task intersect more frequently than those used in the money allocation task, which suggests higher discriminatory power of the revealed preference conditions for rational consumer behavior (see Bronars, 1987). As such, the time allocation task presents a more rigorous assessment of rationality, which explains the greater baseline heterogeneity required for rationalizing respondents' time choices.

[16]Remark that κ_k-ratios are independent from the number of states for a given characteristic. This is illustrated in Table 6 by the κ_k-ratios for gender (with two states: female, male) and employment (with three states: employed, not employed, student) in the money allocation task. Although employment contains more states it obtains a lower κ_k-ratio. Thus, κ_k-ratios allow comparisons across observables regardless of their number of states.

[17]The finding that preferences are most homogeneously distributed across singles and married individuals might be of particular interest to scholars in the household economics field, where individual demand functions are often estimated under the assumption that preferences remain unaffected by changes in household composition (see, for example, Browning et al., 2013). Our findings seem to suggest that preferences do depend on household composition. Of course, our reported results are mainly illustrative and pertain to rather specific choice settings. It remains to be seen whether these findings will be replicated in more general settings.

[18]Remark that this result is independent of the baseline heterogeneity $\hat{\tau}$, as the κ_k-ratios measure relative increases in preference heterogeneity.

[19]Intuitively, this relates to the fact that the level-1 branching considers the different characteristics in isolation from each other.

[20]The literature has proposed alternative procedures to relax the "sharp" binary WARP conditions. A most popular procedure is to make use of Afriat's CCEI (Afriat, 1973; Varian, 1990). As extending our method to include these relaxed rationality conditions is fairly straightforward, we will not consider this explicitly in the current paper.

[21]In statistics, overfitting describes a situation where the outcome of the analysis corresponds too closely to a particular set of data, and may therefore fail to fit to additional data. In our case,

it means that the obtained findings on observable preference heterogeneity could be specifically driven by particular small subsets of observations that are generated in the partitioning approach, which may hamper representativeness.

[22]Subsampling without replacement implies that observations from the original sample will not be sampled more than once for a given subsample, but may appear in several different subsamples. It is different from bootstrapping, which uses subsampling with replacement. Although sampling without replacement is less common due to the under-representation of outliers and reducing variance (Rao et al., 1962), subsampling with replacement is of little use for our application, as sampling multiple times an identical observation will not increase subset heterogeneity (because identical choices cannot violate WARP; see Definition 3).

[23]For example, such a treatment may develop along the lines of Cherchye et al. (forthcoming), who proposed a similar permutation approach for statistical tests of individual consumer rationality on the basis of nonparametric revealed preference conditions. We refer to Pesarin and Salmaso (2010) for a general discussion of the permutation testing approach.

[24]In principle, there are $N!$ possible permutations for a sample with N observed consumer choices. In theory this is what we should consider, but for large N it is practically infeasible to do. Therefore, we propose to restrict ourselves to considering (only) 200 randomly chosen permutations in practice. This gives an "approximation" of the wanted distribution of the κ_k-ratio under the null.

[25]We note that, because we measure the preference heterogeneity contained in an observable characteristic (i.e. τ_k using the notation of Section 2.2) as an integer, this counterfactual distribution of the κ_k-ratios (which divide τ_k by the integer-valued baseline heterogeneity) will be discrete.

References

S. N. Afriat. The construction of utility functions from expenditure data. *International economic review*, 8(1):67–77, 1967.

S. N. Afriat. On a system of inequalities in demand analysis: an extension of the classical method. *International economic review*, pages 460–472, 1973.

E. Alpaydin. *Introduction to machine learning*. MIT press, 2020.

J. Andreoni and J. Miller. Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753, 2002.

A. A. Arechar, S. Gächter, and L. Molleman. Conducting interactive experiments online. *Experimental economics*, 21:99–131, 2018.

S. G. Bronars. The power of nonparametric tests of preference maximization. *Econometrica*, 55(3):693–698, 1987.

M. Browning, P.-A. Chiappori, and A. Lewbel. Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. *Review of Economic Studies*, 80(4):1267–1303, 2013.

L. Cherchye and F. Vermeulen. Nonparametric analysis of household labor supply: goodness of fit and power of the unitary and the collective model. *The Review of Economics and Statistics*, 90(2):267–274, 2008.

L. Cherchye, T. Demuynck, and B. De Rock. Transitivity of preferences: When does it matter? *Theoretical Economics*, 13(3):1043–1076, 2018.

L. Cherchye, T. Demuynck, B. De Rock, and J. Lanier. Are consumers rational? shifting the burden of proof. *The Review of Economics and Statistics*, forthcoming.

S. Choi, S. Kariv, W. Müller, and D. Silverman. Who is (more) rational? *American Economic Review*, 104(6):1518–1550, 2014.

S. Cosaert. What types are there? *Computational Economics*, 53(2):533–554, 2019.

I. Crawford and K. Pendakur. How many types are there? *Economic Journal*, 123 (567):77–95, 2013.

W. E. Diewert. Afriat and revealed preference theory. *Review of Economic Studies*, 40(3):419–425, 1973.

H. S. Houthakker. Revealed preference and the utility function. *Economica*, 17 (66):159–174, 1950.

W. Kool and M. Botvinick. A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General*, 143(1):131, 2014.

G. Loewenstein. Experimental economics from the vantage-point of behavioural economics. *Economic Journal*, 109(453):F25–F34, 1999.

M. Mohri, A. Rostamizadeh, and A. Talwalkar. *Foundations of machine learning*. MIT press, 2018.

F. Pesarin and L. Salmaso. The permutation testing approach: a review. *Statistica*, 70(4):481–509, 2010.

J. N. Rao, H. Hartley, and W. Cochran. On a simple procedure of unequal probability sampling without replacement. *Journal of the Royal Statistical Society: Series B (Methodological)*, 24(2):482–491, 1962.

H. Rose. Consistency of preference: the two-commodity case. *Review of Economic Studies*, 25(2):124–125, 1958.

P. A. Samuelson. A note on the pure theory of consumer's behaviour. *Economica*, 5(17):61–71, 1938.

P. A. Samuelson. Consumption theory in terms of revealed preference. *Economica*, 15(60):243–253, 1948.

K. Surana. How different are we? Identifying the degree of revealed preference heterogeneity. *Mimeo*, 2022.

H. R. Varian. The nonparametric approach to demand analysis. *Econometrica*, 50 (4):945–973, 1982.

H. R. Varian. Goodness-of-fit in optimizing models. *Journal of Econometrics*, 46 (1-2):125–140, 1990.

A Appendix

A.1 Python Implementation

We automated the level-1 and level-2 branching methodology by integrating the solution algorithm from Cosaert (2019), which uses IBM ILOG CPLEX Studio (=CPLEX), with Python. The publicly available project on OSF (Open Science Framework)[26] keeps the CPLEX algorithm at the center, using it for the linear optimization, while data manipulation is run through Python, and results are stored in Excel. The working logic is as follows: Any branching exercise first conditions the data by a given observable characteristic (which is equivalent to sorting the entire data set according to the states of one observable). In a typical

data set this amounts to selecting a variable at the column level, sorting by its possible values (states), and send observations in each state separately to CPLEX to find the minimum number of unique preference types. In case of a level-2 branching we simply double sort two variables one after the other and send each overlapping value separately to CPLEX. For a complete level-1 or level-2 branching this operation is performed automatically for each variable and each of its states. Lastly, repeated subsampling requires one to run the process iteratively. The Python code automates the described process in its entirety and reports the κ-ratios for both level-1 and level-2 branchings.

A.2 Experiment Instructions

A.2.1 Introduction: Welcome to the experiment!

Welcome and thank you for participating.

Please read and follow the instructions carefully as they contain everything you need to know to participate. Participation in this study is voluntary and will take no more than 8 minutes of your time. All responses will be processed anonymously. After the details of the experiment have been explained to you, you may decline to participate if you so wish. Please note that if you choose not to fully participate, you will not be eligible to receive money.

This study has two parts. In the first part, you will choose how long you want to perform a task that requires effort and a pure leisure task. In the second part, you will be given a budget and allocate your resources between two goods. At the end of the two parts, you will be asked to complete a brief questionnaire. In this study, you may earn money and vouchers for yourself, depending on your decisions

during the tasks. After all the answers are collected, we will randomly choose 10 participants for the first task and 5 for the second task to pay their earnings. If you are chosen, your earnings will be paid once the study is concluded. The maximum amount you can earn in total is €9 for the first task and €20 worth of vacation and shopping vouchers for the second task. Distributing the survey is also rewarded: an additional €10 will be given to the participant sharing the survey link with most others. Therefore, please provide the name of the participant from which you have received the survey link on the next page.

A.2.2 Time Allocation Task

Welcome to the first part of our study (which will take at most 5 minutes to complete).

This part consists of one decision-making task that may earn you money depending partially on your decisions and partially on luck.

You will be randomly assigned 90 to 300 seconds, which you will be asked to allocate between two tasks. The first task requires you to put in effort: you will choose a letter and write as many names as you can that start with the chosen letter. For each second spent on this task you will earn a randomly determined wage, varying from 2 to 6 cents per second. This first task thus represents your willingness to work for a given wage. The second task consists of watching a popular video of funny animals we found on YouTube. You will not earn money for this task, but the video is meant to be fun and enjoyable compared to the first task. This second task represents how much you value leisure, instead of working for a given wage. Naturally, the less time you allocate for the first task, the longer you can enjoy the video.

Your goal is to allocate the assigned time endowment between work (first task) and leisure (second task). You can switch between tasks as often as wanted by simply clicking the arrow at the bottom of the page. When time is up, you will be automatically taken to the next part of the study. After all respondents have completed the experiment, we will randomly select 10 participants and pay them their earnings (based on the time spent on the first task and the per second wage). The next page reveals your time endowment together with the per second wage that may be earned by performing the first task. Good luck!

You are assigned XXX seconds. Each second you allocate to the first task earns you XXX cents. The maximum amount of money that you may earn in this part is €XXX. Attention! The next page begins the first task. A timer at the top of the page will show how many seconds are left. Note that time spent on the first task without working to write names will not earn you money.

Choose any letter of the alphabet and start writing names. Remember to capitalize the first letter and to put a space between each name.

A.2.3 Money Allocation Task

Welcome to the second part of our study.

This part consists of one decision-making task where you can earn vouchers worth €20 depending partially on your decisions and partially on luck. The task requires you to make only a single choice and will take less than 2 minutes. You will be randomly assigned a budget of 40 to 100 points to allocate between Vacation and Shopping goods. Prices of the Vacation and Shopping goods will also vary randomly. You may allocate your points however you like by using a slider on the next page showing the percentage of your budget allocated to the two goods. Your

choice will be converted to vouchers for Vacation and/or Shopping, according to the ratio of the points allocated and their relative prices. After all respondents have completed the experiment, we will randomly select 5 participants to be paid in vouchers.

For example:

You are assigned 100 points: Vacation goods each cost 1 point, while Shopping goods each cost 2 points. You decide to allocate 50% of your budget to Vacation (i.e. buying 50 units) and 50% to Shopping (i.e. buying 25 units). Because the price of the Shopping good is twice as expensive as the price of the Vacation good you are therefore 2 times more likely to win the Vacation voucher than the Shopping voucher.

The next page reveals your total number of points and the prices of both the Vacation and Shopping goods.

You are assigned XXX points. Vacation goods each cost XXX point(s). Shopping goods each cost XXX point(s).

How do you allocate your budget?

A.3 Maximization Problems for the Experimental Tasks

A.3.1 Time Allocation Task

Under the assumption of rationality, every participant $i \in N$ in the set of observations S faces the following constrained maximization problem:

$$\begin{aligned} \max_{c,a} \quad & U_i(c, \bar{A} - a) \\ \text{s.t.} \quad & pc + w(\bar{A} - a) = w\bar{A}, \end{aligned} \tag{1}$$

where $\bar{A}$ is maximum amount of time (in seconds) that the consumer can *work*, determined by her endowment in seconds. $\bar{A} - a$ is *leisure*, which is time not spent working on the mentally exhausting activity a. As consumption c represents aggregate Hicksian good, its price p is normalized at 1.

A.3.2 Money Allocation Task

Under the assumption of rationality, every participant $i \in N$ in the set of observations S faces the following constrained maximization problem:

$$\begin{aligned} \max_{q_1,q_2} \quad & U_i(q_1, q_2) \\ \text{s.t.} \quad & p_1 q_1 + p_2 q_2 \leq m, \end{aligned} \tag{2}$$

where q_1 (q_2) is the quantity of Vacation goods (Shopping goods) and p_1 (p_2) its price. The participant's endowment in points is given by m.

A.4 Pairwise Dominance Matrices (based on subsampling)

We can use our subsampling results for an additional, pairwise comparison of alternative level-1 and level-2 branching specifications. Specifically, we can compute how often one characteristic obtains a lower κ_k-ratio than a second characteristic over the 200 subsamples. We summarize our results for the level-1 branching exercise in Table 12, which reports in the form of a pairwise dominance matrix the percentage of subsamples (out of 200 draws) for which the observables in rows

obtain strictly lower ratios than those in columns. For example, for the time allocation task, we observe that marital status outperforms gender in 58.5 percent of the draws, employment in 75 percent of the draws, etc.. Further, looking at the columns, we observe that gender outperforms marital status in 18 percent of the draws, meaning that for 23.5 percent of the draws gender and marital status perform equally well.[27] Looking at the results for both tasks, we find that marital status is the overall top performer; its dominance is particularly pronounced for the money allocation task.

Table 12: Pairwise dominance matrix (rows dominate columns) for level-1 branchings (in %)

	Marital Status	Gender	Employment	Education	Age
Marital Status	0	58.5	75	83	98
Gender	18	0	54.5	60,5	94
Employment	6	22	0	42.5	94.5
Education	3	20	31.5	0	83.5
Age	0	1.5	1.5	4.5	0

(a) Time Allocation task

	Marital Status	Gender	Employment	Education	Age
Marital Status	0	98.5	98.5	100	100
Gender	0	0	28.5	79.5	73.5
Employment	0	48.5	0	87.5	87.5
Education	0	9.5	3.5	0	26.5
Age	0	12	6	52.5	0

(b) Money Allocation task

In a directly analogous manner, we may also conduct pairwise comparisons of alternative level-2 branching specifications. Table 13 summarizes these results for our application. To take a specific example, for the time allocation task we observe that marital status and employment (MS - Emp) outperform gender and marital

status (G - MS) in 49.5% of the draws, gender and education (G - Edu) in 78.5% of the draws, etc.. Looking at the columns, we find that gender and marital status outperform marital Status and employment in 27% of the draws, meaning that for 20% of the draws gender and marital status and marital status and employment perform equally well. From Table 13, we learn that the pair marital status and employment is the top performer for the time allocation task. For the money allocation task, age and marital status turns out to be the top performer, although the combination gender and marital status also performs relatively well.

A.5 Sensitivity Analysis: reducing the number of budget sets (time allocation task)

We assess the sensitivity of our empirical results with respect to the number of budgets that is considered for the time allocation task. We recall from Section 3.3 that our original sample contains 148 respondents that completed both the time allocation task and the money allocation task (giving the sample that we considered in our core exercise that is discussed in the main text) and, in addition, 35 respondents that only completed the time allocation task. This actually gives us a sample with 183 (= 148 + 35) observations completing the original time allocation task with 11 budgets.

We use this enlarged sample as a basis for our sensitivity check. Specifically, from this sample we remove the observations that were assigned budgets 6 and 8, so constructing a new sample of consumers with choices defined over (only) 9 different budgets. When dong so, we get a newly constructed sample with 148 observations (as budgets 6 and 8 contain just 35 observations), i.e. exactly the same size as

	MS - Emp	G - MS	Edu - MS	G - Edu	Edu - Emp	Age - MS	G - Emp	Age - Emp	Age - Edu	G - Age
MS - Emp	0	49.5	70	78.5	85	85.5	86	91	99.5	100
G - MS	27	0	55.5	68	69	70.5	82.5	83	96.5	99.5
Edu - MS	11.5	27.5	0	53.5	57	58.5	59	72.5	94.5	99.5
G - Edu	8	14	22.5	0	38	44.5	48	55	88	97
Edu - Emp	3.5	13	21	37	0	40.5	45.5	54	93.5	99
Age - MS	3	12.5	18	39	38.5	0	45.5	53	84.5	98
G - Emp	3	6	18.5	31.5	33.5	37.5	0	54	84.5	99.5
Age - Emp	1.5	7	13.5	26	25.5	17	25.5	0	75.5	97.5
Age - Edu	0	0.5	0	1	0.5	4	6.5	8	0	77.5
G - Age	0	0	0	1	0	0	0.5	0.5	6.5	0

(a) Time Allocation task

	Age - MS	G - MS	MS - Emp	Edu - MS	Age - Edu	G - Edu	Edu - Emp	G - Emp	Age - Emp	G - Age
Age - MS	0	51	55.5	69.5	96	95	99	98	100	99.5
G - MS	26.5	0	41.5	50	85	96.5	98	100	99.5	99.5
MS - Emp	18.5	39	0	49.5	81.5	92.5	99	100	100	100
Edu - MS	9.5	31	32	0	80.5	88.5	99.5	90.5	100	98
Age - Edu	0.5	8	7	5.5	0	63.5	83.5	78	90.5	93.5
G - Edu	0.5	0.5	3	7.5	21	0	61.5	60.5	68	78.5
Edu - Emp	0	0	0.5	0	6.5	21.5	0	42.5	50	58.5
G - Emp	0	0	0	3	10	17	39	0	50.5	63.5
Age - Emp	0	0	0	0	3.5	14.5	33	29	0	55
G - Age	0	0	0	1	2.5	10.5	24	24	29	0

(b) Money Allocation task

Table 13: Pairwise dominance matrix (rows dominate columns) for level-2 branchings (in %)

the original sample that is used in our core exercise. In this sensitivity analysis, we compare the findings on observable heterogeneity for this newly constructed sample with the results for our original sample (as described in Section 4).

In terms of baseline heterogeneity (denoted by $\hat{\tau}$), we obtain 22 and 21 preference types for the settings with 9 budgets (newly constructed sample) and 11 budgets (original sample), respectively. Given its particular construction, the sample for 9 budgets is not identical to the one for 11 budgets, even though both samples contain 148 consumer observations. This may (partly) explain this (small) difference in baseline heterogeneity.

Table 14 then reports the κ_k-ratios of a level-1 branching for all characteristics, for both the setting with 9 budgets (first column) and our core exercise with 11 budgets (second column, replicating the results shown in Table 6). For both settings, k-partitioning by marital status obtains the lowest κ_k-ratio, indicating robustness of our conclusion that marital status is a most relevant characteristic driving heterogeneity in time use preferences.

Table 14: Time allocation task: κ_k-ratios for level-1 branchings; full sample

	(Sens. analysis - 9 budgets)	(Core exercise - 11 budgets)
Marital Status	**1.136**	**1.048**
Gender	1.182	1.095
Employment	1.227	1.143
Education	1.227	1.143
Age	1.364	1.286

Next, when looking at Table 15, we find that the level-2 branching results for the sample with only 9 budgets differ slightly from those for the core exercise: the level-2 branching that is based on the characteristics marital status and employment, which came out as optimal in our core exercise, is now slightly dominated by

the branching based on gender and marital status. These results seem to indicate that our findings for the core exercise are not fully robust. At this point, however, is it important to highlight that the difference between the respective κ_k-ratios in the first column of Table 15 is rather small. In our opinion, this mainly suggests gender to be a third relevant characteristic to describe heterogeneity in time use preferences, to be considered in addition to marital status and employment.

Table 15: Time allocation task: $\kappa_{j|k}$-ratios for level-2 branchings; full sample

	(Sens. analysis - 9 budgets)	(Core exercise - 11 budgets)
Marital Status and Employment	1.409	**1.286**
Gender and Marital Status	**1.364**	1.381
Gender and Education	1.409	1.429
Gender and Employment	1.409	1.429
Education and Marital Status	1.591	1.381
Education and Employment	1.545	1.381
Age and Marital Status	1.636	1.429
Age and Employment	1.682	1.476
Age and Education	1.727	1.571
Age and Gender	1.727	1.714

Tables 16 and 17 allow us to check robustness of these findings through our subsampling procedure. In this case, lowering the number of budgets changes the optimal branchings for both the level-1 and the level-2 branchings. Similar to before, our results highlight the importance of gender (in addition to marital status and employment) to model preference heterogeneity. Specifically, for the sample with 9 budgets, k-partitioning by gender obtains the lowest κ_k-ratio, while the pair of gender and marital status characterizes the optimal level-2 branching. Again, the differences with the optimal level-1 and level-2 branchings in our core exercise (based on marital status and employment and marital status, respectively) are not very pronounced. We take these findings to indicate the potential usefulness of a level-3 branching exercise that is based on the characteristics employment, marital status and gender. Given the mainly illustrative nature of our empirical

application, however, we feel that such an exercise falls beyond the scope of the current study.

Table 16: Time allocation task: mean κ_k-ratios (SDs in parenthesis) for level-1 branchings; subsampling

	(Sens. analysis - 9 budgets)	(Core exercise - 11 budgets)
Gender	**1.118**	1.142
	(0.068)	(0.108)
Marital Status	1.167	**1.093**
	(0.066)	(0.067)
Employment	1.218	1.177
	(0.075)	(0.095)
Education	1.230	1.192
	(0.071)	(0.085)
Age	1.354	1.329
	(0.091)	(0.123)

Table 17: Time allocation task: mean $\kappa_{j|k}$-ratios (SDs in parenthesis) for level-2 branchings; subsampling

	(Sens. analysis - 9 budgets)	(Core exercise - 11 budgets)
Marital Status and Employment	1.398	**1.365**
	(0.089)	(0.102)
Gender and Marital Status	**1.353**	1.403
	(0.098)	(0.155)
Gender and Education	1.462	1.483
	(0.099)	(0.135)
Gender and Employment	1.437	1.500
	(0.103)	(0.150)
Education and Marital Status	1.550	1.435
	(0.088)	(0.105)
Education and Employment	1.560	1.485
	(0.116)	(0.120)
Age and Marital Status	1.569	1.495
	(0.106)	(0.135)
Age and Employment	1.643	1.532
	(0.121)	(0.145)
Age and Education	1.691	1.631
	(0.111)	(0.133)
Age and Gender	1.653	1.745
	(0.111)	(0.155)