Web-based survey design for unravelling semi-compensatory choice in transport and urban planning

Sigal Kaplan[a], Shlomo Bekhor[b], Yoram Shiftan[b]

[a]*Transportation Research Institute, Technion, Haifa, Israel;* [b]*Faculty of Civil and Environmental Engineering, Technion, Haifa, Israel*

Sigal Kaplan
Faculty of Civil and Environmental Engineering
Technion - Israel Institute of Technology
Haifa 32000, Israel
Tel: +972.4.829.2381
Fax: +972.4.829.5708
Email: sigalkap@technion.ac.il

The estimation of semi-compensatory models is gaining momentum in transport planning in recent years. However, traditional survey methodologies focus on collecting solely compensatory choice data, which leads to information loss when semi-compensatory models are estimated. The present study proposes a novel web-based survey that enables collecting data about the entire semi-compensatory choice process. The web-based environment allows seamless tracking of semi-compensatory choice protocols without interfering with the natural choice process and without introducing problems related to comprehension bias, narrative inconsistency and misinterpretation of the choice protocols. The procedure is applied to rental apartment choice by students and results shed light on semi-compensatory choice by: (i) demonstrating the importance of choice set formation; (ii) unravelling the distribution of threshold selection across the population; (iii) revealing the linkage between the viable choice-set and the choice.

Keywords: two-stage choice protocols; web-based survey; semi-compensatory choice.

Introduction

The estimation of semi-compensatory models is a recent progress that has gained momentum in transport planning in the last decade. The conceptual framework of semi-compensatory models, derived from the studies of Manski (1977) and Swait and Ben-Akiva (1987), includes a probabilistic two-stage choice process, consisting of an elimination-based choice set formation upon satisfying random constraints, followed by a utility- based choice (Ben-Akiva And Boccara 1995). Although semi-compensatory models assume a two-stage choice process including choice set formation and choice, current semi-compensatory models are estimated solely on the basis of choice outcomes (Ben-Akiva and Boccara 1995, Morikawa 1995, Swait 2001,

Başar and Bhat 2004, Cantillo and Ortúzar 2005, Cantillo et al. 2006, Zheng and Guo 2008, Castro et al. 2009). As a result of the lack of information about choice set formation, current semi-compensatory models require the consideration of all the theoretically possible choice sets. This impedes their application in transportation, since transportation related choices (e.g., travel fare, route choice and destination choice) often involve a large number of alternatives and choice sets. Curiously, although information regarding the choice set formation is important for the estimation of semi-compensatory models, no attempts have been made so far to collect data regarding the entire semi-compensatory choice process. Possible reasons may be the wide agreement in the literature with the assumption that the elimination-based choice set formation process is covert (Tversky 1972), and the difficulty to impute choice sets on the basis of observational data from traditional natural experiments (Ben-Akiva and Boccara 1995).

Although data collection with respect to choice set formation is not addressed in the literature related to semi-compensatory models, threshold elicitation, which is the premise for choice set formation, has been widely discussed in the marketing literature with respect to price acceptability range. Three prevailing methods for threshold elicitation are direct surveys, conjoint-based analysis and market data.

Direct surveys consist of interviewing respondents about upper and lower limits of their acceptable range of values for a specified product. Alternative formats are the own-category procedure derived from the fundamentals of psychophysical experiments (Monroe 1971), and explicit open-ended questions (e.g., Lichtenstein et al. 1988, Wang et al. 2007). Although widely applied, direct surveys are criticized for producing incentive compatibility bias and strategic response bias (Wang et al. 2007) when utilised to elicit the acceptable price range before a decision is made. Post-

decision recall of thresholds after an actual choice should be bias free, but produces inaccurate or partial results due to memory retrieval difficulties (Gensch and Svestka 1979). Another limitation of direct surveys is their referral to a single pre-specified alternative. Encompassing the entire range of alternatives by using detailed descriptions is burdensome and expensive. Using instead general descriptions leads to inconsistent elicitation since subjects imagine different reference alternatives.

Conjoint analysis (e.g., Kohli and Mahajan 1991, Jedidi and Zhang 2002) serves to infer thresholds for multi-attribute products from preference or choice data. These methods avoid excessive attention to the price acceptability range and thus are considered as incentive neutral (Wang et al. 2007). A limitation of conjoint analysis and choice-based experiments consists in their attempt to infer an essentially non-compensatory concept (thresholds) from compensatory behaviour. In addition, these methods assume that individuals are fully informed regarding their choice sets before revealing their preference structure. However, in realistic choice situations this assumption does not hold (Shocker et al. 1991).

Market data analysis is conducted by using panel data or scanner data at designated locations. Although market data methods have the advantage of collecting revealed preference data, their disadvantages are high operating costs, dependency on historical price fluctuations and inability to encompass new products (Breidert et al. 2006). In addition, market data collection methods infer choice set information from recurrent choices under different market conditions, hence they are unsuitable for retrieving choice set information from cross-sectional data.

A major disadvantage of the three aforementioned methods is their exclusive focus on threshold data collection, rather than on the entire semi-compensatory process. A potentially powerful method for unravelling the entire semi-compensatory

choice process is the analysis of verbal decision protocols recorded during a choice task. This method, originally employed by Payne (1976), involves recording actions of individuals by asking them to "think aloud" while performing a choice task. The advantage of this method is twofold. First, the method can reveal both non-compensatory threshold selection and compensatory choice during a choice task. Second, the method is bias free when choice protocols are recorded for actual transactions. Nevertheless, traditional methods of recording choice protocols (i.e., writing transcripts and audio-video) are subject to two severe limitations that impede their application in transportation. First, recording verbal choice protocols necessitates a laboratory setting, which in turn imposes restrictions on testing behaviour in realistic choice situations. Second, verbal choice protocols involve problems of respondents' burden, narrative heterogeneity and difficulty of individuals to articulate their thoughts in a structured manner.

The present study demonstrates the application of a web-based survey to collect data about the entire semi-compensatory choice process. The survey enables to retrieve tolerated criteria thresholds and choice outcomes from both cross-sectional and panel data in complex choice situations entailing many alternatives and multiple criteria. The survey is suitable for collecting either revealed preference (RP) or stated preference (SP) data, is free of the biases that are traditionally related to threshold selection, and avoids problems of respondents' burden, narrative inconsistency and misinterpretation of the choice protocols. For reasons of completeness it should be noted that the principles of the data collection are briefly discussed by Kaplan et al. (2009), while focusing of on reservation price elicitation for marketing purposes. The current study enhances the understanding of the proposed methodology by thoroughly discussing the practical aspects of the survey application, focusing on survey issues

that are relevant to transport and urban planning, and providing a detailed description of the information regarding choice set formation and choice that can be retrieved from the survey.

The proposed web-based survey is applied to a SP choice experiment of off-campus rental apartment choices by students, as an example of a complex choice situation involving a large number of alternatives characterised by many attributes. The remainder of the paper is organized as follows. The next section presents the proposed web-based survey. The third section illustrates the application of the proposed survey within the empirical context. The fourth section focuses on the analysis of the collected data. Last, the fifth section draws conclusions and recommends further research.

Methodology

The present study demonstrates the application of an innovative web-based survey for the combined data collection of criteria thresholds, as the premise for choice set formation, and choice outcomes. The survey methodology relies on evidence from decision making studies showing that: (i) when faced with many alternatives, individuals use a sequence of an elimination-based heuristic and a compensatory evaluation (Payne et al. 1993); (ii) intrinsic constraints and attitudes can be measured similarly to the measurement of responses to physical stimuli in psychophysics (Monroe 1971); (iii) information regarding individual preferences can be inferred from choice outcomes. The following sections describe the survey, its advantages and limitations, and the prevention of possible biases, which are related to the application of the survey to collect stated preferences.

Survey overview

The web-based survey consists of a procedure for tracking the entire semi-compensatory choice process and a questionnaire aimed at collecting personal information. The procedure to record the semi-compensatory choice process consists of a two-stage choice task. At the first stage, individuals specify their tolerated criteria thresholds in order to delimit the universal realm of alternatives to their viable choice set. The threshold specification is conducted from a list of fixed criteria threshold values. At the second stage, individuals choose their preferred alternative from their retained viable choice set. In the event that the procedure does not yield any satisfactory result, individuals have the possibility to either settle for their most preferred alternative among the available ones that satisfy their tolerated criteria thresholds, update their criteria thresholds, or retreat to a "no choice" option. A questionnaire supplements this two-stage procedure by collecting individual characteristics that can be associated with the criteria threshold selection and the choice outcomes.

The two-stage choice task is based on the assumption that individuals acquire only partial information regarding the universal realm of alternatives at the criteria threshold specification stage, while they are fully informed regarding the profile of alternatives at the choice stage. Specifically, at the criteria threshold specification stage individuals are aware of the general category, for example "rental apartments", delimited by the possible range of criteria threshold values, but they do not possess detailed information regarding the attribute values of all the alternatives in the universal realm. Theoretically, such information could be retrieved by repetitive criteria threshold specification and thorough investigation of the viable choice set. Practically, such a process is unfeasible for a large number of attributes and a large

number of alternatives in the universal realm. Accordingly, the criteria threshold specification depends solely on individuals' intrinsic constraints, preferences and perceptions as well as informal knowledge of the attribute value range in the universal realm. At the choice stage, individuals are fully informed regarding the exact profiles of the alternatives that satisfy their selected criteria thresholds, and hence their choice depends on the attributes of the alternatives.

The selected individual criteria thresholds and chosen alternatives are retrieved for respondents by recording their two-stage choice protocols. As explained in the introduction, the traditional method of recording verbal choice protocols by asking people to "think aloud" necessitates a laboratory setting, is costly and time consuming, and bears problems of individuals having difficulty to articulate their thoughts in a structured manner. Instead, the current study proposes a web-based recording of choice protocols by tracking the series of individuals' typing actions during the criteria specification and choice process in a web-based environment.

The web-based survey can be used to collect either SP or RP data. In SP surveys, a hypothetical choice experiment is designed to simulate real choice conditions in which respondents choose among alternatives from a synthetic database. In RP surveys, actual transactions in digital economy are traced.

Benefits and limitations of the web-based environment

The benefit of web-based tracking of choice protocols is fourfold from the researchers' perspective. First, it helps avoiding the problems of phrasing inconsistency and misinterpretation of verbal choice protocols. Second, it prevents coding mistakes. Third, it does not interfere with the natural choice process, since individuals are unaware that their actions are recorded unless specifically informed. Last, the operational costs associated with a web-based environment in terms of time,

human resources and budget are modest when compared to other survey forms such as home interviews and phone surveys. The benefit of the web-based environment for survey respondents is twofold. First, respondents' burden is minimal as individuals are able to perform actual or hypothetical transactions as they naturally do in a digital economy environment without being asked to conduct the extra cognitive effort of "thinking aloud". Second, individuals are minimally disturbed in their daily activities due to the schedule and location flexibility provided by the web-based environment.

Although computer aided recording of choice protocols in a web-based environment is highly advantageous, a possible limitation is the necessary condition that respondents are familiar with the Internet media and have accessibility to an Internet connection. With the widespread penetration of the Internet media all over the world and among all age groups, this condition seems less and less a limitation. An additional challenge concerns sample reliability. The web-based environment offers less control over sample composition and is at higher risk of fraud than traditional surveys, since it provides a much higher degree of anonymity and location flexibility. However, in order to maintain a high level of sample reliability, the survey can be designed for controlled access by asking respondents to identify themselves through individual passwords to be verified against a list of eligible respondents.

Response reliability

Although the survey retrieves bias free data when applied as RP survey to extract information from actual transactions, it may be susceptible to strategic response bias and incentive compatibility bias when applied as SP survey to extract information from hypothetical choice situations.

Strategic response bias typically occurs in SP surveys when respondents anticipate that their responses would influence the range of the attribute values of the

alternatives. In mode choice for example, if respondents are led to believe that their responses may influence the price or the speed range of the alternative, they might specify low prices and high speed thresholds at the choice set formation stage and choose low cost and high speed alternatives at the choice stage. Strategic response bias can be mitigated by decreasing the respondents' awareness regarding their ability to influence the supply of alternatives. Three design elements of the proposed survey aim at mitigating strategic response bias: (i) the choice is made from a fixed product inventory, hence individuals are unaware that their responses can influence the range of attribute values; (ii) the design of the survey as a choice task masks the importance of choosing specific criteria thresholds at the choice set formation stage or specific attribute values at the choice stage; (iii) at the choice set formation stage, the criteria are listed in a table format that can be read both horizontally and vertically, in order to prevent respondents from associating the order of the criteria with their relative importance. Furthermore, the selection of criteria thresholds is left entirely to the respondents' discretion.

Incentive compatibility bias occurs in SP surveys when respondents do not bear the consequences of their choices. For example, respondents might choose more expensive alternatives than they could afford when no actual payments are involved. Associating the hypothetical choice with an actual probability for making a paid transaction can possibly mitigate the incentive compatibility bias (Wang et al. 2007). Specifically, after making their choice, respondents are asked to purchase the chosen alternative, depending on a lottery result. The association of the hypothetical choice with an actual transaction probability is limited to inexpensive tangible products. However, in transport and urban planning this method is difficult to apply since choices are costly (e.g., long distance travel fare, air travel fare, residential choice, car

rental choice) or less tangible (e.g., route choice or destination choice). Instead, the proposed survey mitigates incentive compatibility bias by allowing respondents to use the price threshold in order to adjust the cognitive burden associated with the choice process. A high price threshold results in many alternatives, increasing the effort needed for the choice process. A low price threshold results in a small or null set of alternatives, decreasing satisfaction from the choice process (Iyengar and Lepper 2000). Thus, respondents are encouraged to state their true price range in order to compromise between the price and the cognitive effort that is necessary in the choice process.

Empirical application of the proposed survey design

The web-based survey was applied to collect SP data about rental apartment choice by students. The survey was conducted by means of a custom designed website, which was inspired by existing real-estate portals in order to evoke the feeling of a realistic choice situation. In the survey, respondents searched a synthetic real-estate database according to a pre-specified list of criteria and threshold values. From the resulting set of alternatives, they chose their three most preferred apartments and ranked them by their order of preference. Respondents were allowed to conduct the database search multiple times by revising their selected thresholds. The final choice and the thresholds leading to it were automatically coded into the database. The next subsections describe the generation of the synthetic database, the criteria specification and the questionnaire design, present the website structure and discuss the issue of incentives.

Synthetic database generation

The synthetic apartment database was generated on the basis of a statistical analysis of four on-line real-estate portals that serve for students' rental apartment choice in Israel. The four websites are locally popular since they allow both publishing and reading free advertisements without on-line registration. The retrieved data sample contained 310 distinct apartments for rent in Haifa, which were advertised during the winter semester of 2007. The retrieved apartment inventory was analyzed by studying the distribution of apartment attribute values, calculating the correlation among attributes, and estimating regression models.

The synthetic real-estate database included 600 apartments representing six neighbourhoods in the city of Haifa. Each apartment in the database was characterised by 18 attributes including neighbourhood, walking time to campus, noise level, price, size, number of rooms and balconies, renovation status, floor, number of roommates and smoking policy, availability of view, parking, security bars, elevator, air conditioning, solar water heating and washing machine. The synthetic real-estate dataset was generated with the intention to maintain behavioural realism while providing sufficient variability in the generated values. Hence, the composition of the synthetic real-estate dataset relied on random draw of apartment attribute values from distributions with parameters based on the statistical analysis of the retrieved data inventory.

Duplicate and dominant alternatives produced by the generation process were removed from the dataset. Duplicate alternatives were defined as apartments that are not easily differentiable under the conditions of information load in terms of number of alternatives and attributes. Accordingly, alternatives were removed as duplicates when they shared the same attribute values for five out of eight criteria. Dominant alternatives were defined as apartments that were either high quality apartments with

very low monthly rent or high quality apartments located in close proximity to campus. Following the generation process and the removal of duplicate and dominant alternatives, the characteristics of the generated dataset were compared to the retrieved apartment inventory to confirm the desired similarity between the retrieved and the generated data sources.

Search criteria

The criteria for eliminating alternatives at the choice set formation stage were derived from a pilot survey among 74 students. The relevant criteria were apartment sharing, location, maximal rent price, minimum and maximum number of rooms, proximity to campus (walking time in minutes), noise level and parking availability. Rent price values were expressed in U.S. dollars, the currency commonly used in 2007 in Israel to indicate monthly rents. The rent price varied between $150 and $700 per month, with increments of $10, to reflect the actual price range observed on the market. Apartment sharing was defined as the possibility to select either a vacant apartment or an apartment with roommates. The location criterion included the six metropolitan core neighbourhoods. The number of rooms varied between 1 and 5. The walking time ranged between 5 and 30 minutes. The noise criterion included four levels differentiating between apartments on the basis of road hierarchy (i.e., local street versus arterial road) and orientation with respect to the street (i.e., street versus rear facing apartment). Parking availability differentiated between availability and unavailability of a reserved parking space.

Questionnaire design

The questionnaire aimed at collecting personal information, perceived location amenities, price perceptions, travel preferences and study preferences. Personal

information included socio-economic characteristics, transportation use, residential characteristics and smoking habits. Perceived location amenities for each neighbourhood included perceived car travel time to campus, ease of accessibility to campus by public transport, availability of job opportunities and leisure activities, and public open space availability. Information regarding attitudes perceptions and preferences was collected through indicators that were retrieved from the literature, since the development of valid and reliable scales was out of the scope of the current research. Price perceptions were based on Lichtenstein et al. (1993). Study preferences (on-campus versus home) were constructed similarly to attitudes towards telecommuting (Mokhtarian and Bagley 2000). Travel preferences were adopted from Handy et al. (2006). All the items were expressed on a seven-point Likert scale, except car travel time that was expressed in minutes.

Website design

A custom designed website was developed in order perform the survey. The website was connected to a database that contained the synthetic apartment dataset and tables for automatic coding of the respondents' questionnaire answers, selected threshold values and choice outcomes.

In order to control for sample reliability and facilitate the unique identification of the respondents, students were asked to insert their identification card number upon entering the website. This identification method allowed schedule flexibility in completing the survey and prevented problems resulting from unstable internet connectivity, since it allowed multiple entries to the website by the same student while controlling for duplicate answers. In order to tackle the problem of non-response due to privacy violation, students were also given the possibility to complete

the survey anonymously by typing a pre-specified identification number, provided that they waved their rights to the incentives.

In order to encourage respondents to provide reliable information, the students were notified that a careful and truthful completion of the questionnaire was a necessary condition for receiving the incentive. Students were also advised about the existence of a procedure for the detection of inconsistent, careless and random answers. In addition, the submission time for each task was recorded automatically in order to control for random and careless answers. In particular, extremely short completion times relatively to average completion time of the entire sample served as a proxy indicator for random or careless answers along with detection of random data patterns.

The survey was split into two parts that could be accessed separately in the website. The first part of the website consisted of the questionnaire regarding personal information followed by the two-stage choice task. The second part of the website consisted of the questionnaire regarding attitudes, perceptions and preferences. The division of the survey in parts aimed at decreasing respondent's burden by creating a sense of satisfaction upon task completion, allowing partial completion of the survey, and providing schedule flexibility in completing the entire survey. In addition, the first part of the survey was designed as self-sufficient, in order to include in the data analysis respondents who completed the first part of the survey, but dropped out before completing the second part.

The website structure is illustrated in Figure 1. The first page of the first part (part A) presented a questionnaire regarding respondents' personal information. A data verification function verified full completion of the questions in the page. Upon

successful submission of the questionnaire page, the respondents were directed to the two-stage choice task.

[Insert Figure 1 about here]

The structure of the two-stage choice task within the website consisted of the following elements: (i) a page asking the respondents to specify their tolerated criteria thresholds from a menu of criteria threshold values; (ii) a query in structured query language (SQL) for retrieving the viable choice set from the database upon submission of the specified criteria thresholds; (iii) a database containing the generated apartment dataset, from which a record-set containing the viable choice set was retrieved upon the execution of the SQL query; (iv) a page presenting the retrieved record-set of viable alternatives and inviting respondents to choose their three most preferred alternatives.

The criteria specification page did not provide any information regarding possible dependence among criteria or their order of importance. The criteria were listed in a table format to prevent inferences regarding their relative importance and their selection was entirely at the respondents' discretion. In addition, the selection of criteria thresholds did not indicate any linkage among them.

Data verification functions checked that the choice set size was manageable and that the same alternative was not ranked more than once. In the case that the SQL query yielded either an empty or an unmanageable (over 100 alternatives) choice set, the list of viable apartments was not presented and an alert message asked respondents to redefine their criteria thresholds. In the case of duplicate ranked alternatives, an alert message appeared and the choice outcomes were not coded into the database. Upon successful completion of the two-stage choice experiment, both

the specified criteria thresholds and the corresponding choice outcomes were recorded into the database.

In the second part of the website (part B), a questionnaire regarding perceptions, attitudes and preferences was presented over multiple pages due to its length. Upon automatic verification of full completion of the questions in each web page, responses were recorded into the database and respondents were directed to the next web page.

To facilitate communication with the respondents, instructions could be read and detailed maps could be viewed by opening pop-up windows at every stage of the survey. A 24-hour active phone number was provided on the first page of the website and an e-mail link was embedded on every web page in order to allow respondents to ask questions and to report in real-time interrupted sessions due to connectivity problems.

Incentives and promotion

Participation in a monetary prize raffle was offered to students as an incentive for completing the survey. Three reasons guided the choice of a prize raffle over guaranteed individual incentives: (i) large prizes from raffles are more effective than guaranteed small individual payments in web based surveys (Bosnjak and Tuten 2003); (ii) response quality, sample composition and survey outcomes of web-based surveys are not adversely affected by using a prize raffle as an incentive (Goritz 2004); (iii) prize raffles allow offering substantial incentives within budget constraints without posing limitations on sample size.

Monetary rewards were preferred over gift certificates, since money is independent from taste preferences. A total budget of 1,000 U.S. Dollars was

allocated for the raffle, and 23 prizes were defined within the $25-$250 range. The highest prize reflected the average monthly rent price for a student in a shared apartment. The prize structure was designed in accordance with rank-dependent expected utility theory (Quiggin 1991) in order to attract both risk seekers and risk averters.

The survey was conducted during the spring semester of 2007 over a period of forty days. An official e-mail message was sent to both undergraduate and graduate students. In addition, postcard-size leaflets were distributed on campus in the beginning, middle and end of the survey for a total of 3,000 postcards. Advertising the survey by postcards allowed the inclusion of all the necessary information in a small portable format that transmits a positive image as postcards are a popular advertising tool addressing young audience for leisure activities.

Empirical application: results

Data collection efficiency

The data collection efficiency of the survey is evaluated from the perspective of both researchers and the respondents. The efficiency from the researchers' perspective concerns obtaining high-quality data at low costs. The efficiency from the respondents' perspective concerns having minimal disturbance to daily activities.

Efficiency from the researchers' perspective

The efficiency of the web-based survey from the researchers' perspective is assessed by survey completion rate, drop-out rate, share of valid responses and cost-benefit per respondent. The cost-benefit per respondent and the level of correspondence between the completion rates and the advertising efforts indicate the effectiveness of data collection. The share of valid records denotes the quality of the data.

The survey completion rate reflects the efforts in advertising the survey. The highest share of completed surveys (26.2%) was obtained following the official e-mail. Each of the further advertising efforts by distributing leaflets produced an immediate 10% increase of the total completion rate.

From 1,325 distinct respondents who started the survey, 84.1% completed the survey. The database yielded records of 1,049 valid respondents, who were defined as respondents with complete and non-duplicate records, a valid student number, and non-random answers. The high share of valid survey completions (93.6% of the respondents who completed the survey) indicates the high quality of the data.

The costs of the survey totalled roughly 3,000 U.S. Dollars (2007 exchange rate) including website construction, incentives and promotion. The database yielded 1,049 valid records of respondents resulting in an extremely effective cost-benefit ratio of only 2.86 U.S. Dollars per respondent.

Efficiency from the respondents' perspective

The efficiency of the web-based environment from the respondents' perspective is assessed by looking at the hourly distribution of the completed surveys as an indicator of the usefulness of schedule flexibility. As illustrated in Figure 2, the survey completion times were spread around the clock, which indicates the usefulness and the importance of time flexibility for the respondents.

[Insert Figure 2 about here]

Sample demographics

Socio-economic characteristics of the survey respondents are summarized in Table I. The demographics of the Technion student population are not public records, but some information was retrieved from the Israeli Central Bureau of Statistics (CBS) and other sources in order to evaluate whether the sample is representative of the

student population. The percentage of female respondents is fairly close to the 35.3% female student population in the Technion (CBS 2006). The median age of 26.6 years among the survey respondents corresponds to the median age of 26.1 years of the general student population (CBS 2005a). The share of respondents living in dormitories matches the share of student population in dormitories assuming full occupancy (Technion 2007). Among the 55.1% of survey respondents currently living in their own apartments, 32.5% rent an apartment alone or with roommates, 14.0% live with a partner and 53.5% are married. These shares fairly agree with CBS records (CBS 2005b), which report that 32.8% of the students not residing with their parents rent an apartment, 9.4% live with a partner and 57.8% are married. The distribution of the place of residence is compatible to the geographic location of the Technion. The share of employed undergraduates is 43.3%, which is slightly lower than the employment rate of 52.0% among engineering students (Intel press room 2007).

[Insert Table I about here]

Choice set formation stage

The importance of elimination-based choice set formation

At the choice set formation stage, eight criteria were available to respondents for extracting their viable choice set from the apartment dataset: apartment sharing, neighbourhood, maximum monthly rent price, maximum and minimum number of rooms, noise level, reserved parking availability and proximity to campus. The average number of selected criteria by respondents was 4.8 (SD=1.2). The distribution of the number of selected criteria by respondents is presented in Figure 3.

[Insert Figure 3 about here]

The share of respondents who selected each criterion is presented in Table II. The most frequently selected criteria for delimiting the universal realm of apartments

to a viable choice set were apartment sharing, neighbourhood cluster and maximum monthly rent price. The most popular combination of thresholds selected by 17.1% of the respondents included all the criteria except for parking. The second most popular combination selected by 8.0% of the respondents included apartment sharing, neighbourhood, maximum monthly rent price, proximity to campus and noise level.

[Insert Table II about here]

Threshold distribution across the population

55.2% of the respondents retained only vacant apartments in their choice set, while 39.3% preferred to retain only shared apartments and 5.5% did not delimit their choice set according to the apartment sharing criterion.

The distribution of the selected monthly rent thresholds is presented in Figure 4. Excluding respondents who did not delimit the universal realm by selecting a monthly rent threshold, the mean of this distribution is $381.3 (SD=$134.5), which is only 14.5% higher than the mean rent price in the generated apartment dataset. Even though the price threshold was listed in $10 increments, most respondents (78.5%) selected the monthly rent thresholds in multiples of $50 (i.e., $200, $250,…,$700), which reflects the sensitivity of respondents to changes in rent prices.

[Insert Figure 4 about here]

The six neighbourhoods represent four neighbourhood types that are relevant to students' residential choice. The first (type A) is adjacent to the campus but offers little employment or leisure opportunities. The second (type B) is located in close proximity to campus, although slightly further away than type A. The third (type C) is located farther from campus in the new city centre. It is not easily accessible by public transport, but offers abundance of shopping, leisure and employment opportunities.

The fourth (type D) is located farther from campus in the old city centre. It offers very good accessibility to campus by public transport and abundance of low-price shopping opportunities, although it offers little employment or leisure opportunities. Figure 5 describes the distribution of neighbourhood type selection across respondents. Type A neighbourhood is by far the most attractive to students, while type D neighbourhood is by far the least attractive to students.

[Insert Figure 5 about here]

The option to delimit the realm of apartments by walking time to campus was available to respondents who delimited the universal realm to a neighbourhood in close proximity to campus (type A and type B). Figure 6 displays the distribution of the maximum walking time across respondents. Respondents were tolerant with respect to walking time to the campus. Interestingly, only 12.3% were willing to walk only up to 10 minutes to campus, while 44.1% of the respondents were willing to walk at least 20 minutes.

[Insert Figure 6 about here]

Table III details the two-way distribution of the range of number of rooms that was selected by respondents. Most frequently, respondents delimited their realm of alternatives to apartment size ranging from 2.0 to 4.5 rooms.

[Insert Table III about here]

Figure 7 displays the distribution of respondents according to their tolerated maximal noise level. The noise criterion was important to a large share of the respondents, as 38.4% of the respondents delimited their universal realm to apartments situated at local streets.

[Insert Figure 7 about here]

Linkage between threshold selection and individual characteristics

The criteria specification can be related only to individual characteristics since respondents had only partial information regarding the universal realm of alternatives at the choice set formation stage. Table IV presents the results of an ordered-response probit model for the monthly rent price threshold as an example for the linkage between threshold selection and individual characteristics. The model is estimated for the full sample of 1049 respondents and accounts for variables that are significant at the 0.05 level in the model based on individual characteristics.

[Insert Table IV about here]

Correlation patterns across criteria

Apartment sharing is moderately correlated with price (Spearman's Rho = 0.57) and walking distance (Spearman's Rho = 0.43), namely respondents who delimited the universal realm to shared apartment stated lower price thresholds and lower walking time thresholds than respondents who delimited the universal realm to vacant apartments. Hence, respondents were willing to share an apartment in order to save money or to live in greater proximity to campus. The price criterion is loosely correlated with the minimal tolerated number of rooms (Spearman's Rho = 0.32) and is moderately correlated with neighbourhood type C (Spearman's Rho = 0.44) indicating that respondents stated higher price threshold values for larger apartments and for apartments located in the new city center. All the correlations are significant at the 0.05 level.

Choice stage

The selected combination of tolerated criteria thresholds yielded the viable choice set for each respondent. The distribution of the viable choice set size is displayed in figure 8. The average choice set size is 17.0 alternatives (SD=17.5) and the median

choice set size is 11.0 alternatives. 48.3% of the respondents had up to 10 alternatives in their viable choice set and 86.2% of the respondents had up to 30 alternatives.

[Insert Figure 8 about here]

Respondents were given the option to sort their viable choice set according to one criterion of their liking. 72.3% chose to sort the viable choice set. 66.6% of the respondents sorted by price, while the second and third most frequently selected criteria for sorting the viable choice set were neighbourhood and proximity to campus, selected by 12.9% and 9.2% of the respondents, respectively.

After sorting the choice set, individuals chose and ranked their three most preferred apartments. The relative position of the chosen apartments on the list of viable ones after sorting influenced the choice: 60.3% of the respondents chose their first priority choice among their top five listed apartments.

Correlation across stages

Several variables, such as monthly rent price, walk time to campus and number of rooms, have a dual role both as thresholds at the choice set formation stage and as attributes at the choice stage. For the price variable, the relationship between the selected threshold and the corresponding choice outcome was investigated. Considering all three ranked choices per respondent, the monthly rent price of the chosen alternative is strongly correlated with the price threshold value (Pearson's R=0.72) and its distribution in terms of percentage difference with respect to the price threshold is presented in figure 9. For almost a third of the respondents (30.2%), the difference between the monthly rent price of the chosen apartment and the monthly rent price threshold was at most 15%, and for 51.1% the difference was at most 25%. These results establish the linkage between the consecutive choices in the two stages.

[Insert Figure 9 about here]

The monthly rent price range of the viable choice set was defined by the rent price threshold, conditional on apartment availability. Since threshold selection was conducted under conditions of partial information, that is the respondents were not aware of the "true" price range in the generated database, the actual range in the choice set could be narrower than the respondents' tolerated price range. The monthly rent price of the chosen apartments was strongly correlated with the minimum (Pearson's R = 0.78), maximum (Pearson's R = 0.85) and average price (Pearson's R = 0.90) of the apartments included in the choice set. Figure 10 presents the price of the chosen alternative in terms of percentage of the average price of the viable apartments.

[Insert Figure 10 about here]

Conclusions

The current study demonstrates the application of a novel web-based survey for collecting data about semi-compensatory choice processes. The collected data consist of criteria thresholds, which serve to delimit the universal realm of alternatives to a viable choice set, and respective choice outcomes. Data were collected by means of a custom designed website, which seamlessly tracked semi-compensatory choice protocols by recording respondents' typing actions. The procedure was applied to rental apartment choice of students as an example of a non-repetitive choice situation entailing multiple criteria and many alternatives.

As the current study breaks new ground in unravelling semi-compensatory choice, three important aspects of the choice set formation process are observed for the first time. First, elimination-based choice set formation is crucial in complex choice situations as 85.6% of the respondents selected at least four criteria thresholds for delimiting the universal realm to a viable choice set. Second, elimination-based

choice set formation is insufficient as a sole simplification strategy, since 72.3% of the respondents voluntarily chose to sort the alternatives in the viable choice set according to the criteria of their liking. The importance of sorting as an additional simplification heuristic is indicated by the fact that 60.3% of the respondents chose their first priority choice among their top five listed apartments. Third, the size of a manageable choice set is less than a dozen alternatives, as the median choice set size among survey respondents was 11.0 alternatives and 48.3% of the respondents had up to 10 alternatives in their viable choice set. 86.2% of the respondents had up to 30 alternatives, which is expected since choice sets of larger size become unmanageable (Iyengar and Lepper, 2000).

Results demonstrate the ability of the proposed survey methodology to observe the distribution of criteria thresholds across the population, to identify correlations among criteria thresholds and to understand the linkage between criteria threshold selection and individual characteristics. The results of the current study show that thresholds are correlated across criteria. Interestingly, the correlations are related solely to intrinsic perceptions of respondents since the respondents could not infer such information from the survey. In addition, results show the existence of population heterogeneity in criteria threshold selection and that a significant proportion of the variance (35.0%) in threshold selection can be explained by individual characteristics.

When applied to collect SP data, the survey may be susceptible to strategic response bias and incentive compatibility bias usually associated with direct questioning of respondents. Instead, when applied to collect RP data, the proposed survey is bias free. In the current hypothetical choice experiment, the difference between the mean of selected price threshold values and the mean rent price in the

synthetic database, which is based on actual market prices, is 14.5%. Considering that due to the complexity of the choice task respondents could not easily infer the price distribution in the synthetic database from repetitive trials, this difference reasonably indicates that the threshold values are unbiased. Probably, the design of the choice task is efficient in mitigating possible biases. This point remains to be investigated and assessed in further research.

Choice set formation and choice from the viable options are assumed to be distinct mental processes (Bovy, 2009), and hence are treated as independent in the literature of semi-compensatory models based on Manski's (1977) formulation. However, the choice set formation stage and the choice stage are not independent in that the final choice is conditional on the viable set of alternatives. Additional dependence derives from the employment of the two cognitive processes by the same individual. The current study confirms that dependence exists between the two stages by observing the price variable, which has a dual role both as a threshold at the choice set formation stage and as an attribute at the choice stage. Results show strong correlation between the two stages, as the price of the chosen product variation is strongly correlated with the price threshold as well as with the price range limits in the available choice set. These findings suggest the existence of anchoring and adjustment procedure at the choice stage. Specifically, after a-priori stating their thresholds, respondents update their preference structure by using either the thresholds or the average attribute values of the viable alternatives as external reference points. This result is of major importance since it calls for reconsideration of the currently prevailing assumption in semi-compensatory models that the error terms of the choice set formation and choice are independent.

Further development includes the application of the proposed procedure to other spatial and transportation related choices (e.g., car rental, travel fares, route choice, and recreational destination choice). The successful collection of SP data demonstrates the potential of the procedure for collecting RP data from actual on-line transactions conducted via commercial websites. In the transport sector, websites are readily available from airlines and train companies, tourist agencies and car rental agencies.

Subsequently, the proposed survey methodology can contribute to a better realisation of both the explanatory and predictive potential of semi-compensatory models and increase their ability to realistically represent behavioural decision processes. By relying solely on choice outcomes, current models have a limited ability to represent the non-compensatory choice set formation process. Frequent assumptions embedded in semi-compensatory models are independence of the thresholds from individual characteristics (Morikawa, 1995; Swait, 2001; Cantillo and Ortúzar, 2005; Cantillo et al., 2006; Castro et al., 2009), normal distribution of threshold values across the population (Swait, 2001; Cantillo and Ortúzar, 2005; Cantillo et al., 2006) and independence across thresholds of different criteria (Ben-Akiva and Boccara, 1995; Swait, 2001; Başar and Bhat, 2004; Cantillo and Ortúzar, 2005; Cantillo et al., 2006; Castro et al., 2009). The added information regarding threshold selection can contribute to the refinement of the mathematical representation of thresholds in semi-compensatory models. In addition, since analysts currently have to consider all the theoretically possible choice sets, semi-compensatory models are applicable to choice problems with a limited number of choice sets. In fact, all the aforementioned semi-compensatory models are applied to choice situations with up to nine choice sets. As demonstrated by Kaplan et al. (2009),

the collected data regarding individual choice sets can contribute to alleviating the computational complexity embedded in semi-compensatory models, as it enables the analyst to consider the actually chosen choice sets for estimation purposes, rather than to consider the entire realm of theoretically possible choice sets. Last, the current study shows that a correlation exists between the choice set formation stage and the utility maximization stage. A potential development of semi-compensatory models may include the representation of the correlation between the two stages.

References

Başar, G., Bhat, C.R. 2004. A parameterized consideration set model for airport choice: application to the San Francisco Bay area. *Transportation Research Part B: Methodological, 38 (10),* 889-904.

Ben-Akiva, M.E., Boccara, B. 1995. Discrete choice models with latent choice sets. *International Journal of Research in Marketing, 12 (1),* 9-24.

Bosnjak, M., Tuten, T.L. 2003. Prepaid and promised incentives in web surveys: an experiment. *Social Science Computer Review, 21 (2),* 208-217.

Bovy, P.H.L. 2009. On modelling route choice sets in transportation networks: a synthesis. *Transport Reviews, 29 (1),* 43-68.

Breidert, C., Hahsler, M., Reutterer, T. 2006. A review of methods for measuring willingness-to-pay. *Innovative Marketing, 4,* 8-32.

Cantillo, V., Heydecker, B., Ortúzar, J.d.D. 2006. A discrete choice model incorporating thresholds for perception in attribute values. *Transportation Research Part B: Methodological, 40 (9),* 807–825.

Cantillo, V., Ortúzar, .J.d.D. 2005. A semi-compensatory choice model with explicit attribute-thresholds of perception. *Transportation Research Part B: Methodological, 39 (7),* 641-657.

Castro, M., Martinez, F., Munizaga, M. 2009. Calibration of a logit discrete choice model with elimination of alternatives. *Proceedings of the International Choice Modelling Conference,* March 30-April 1 Harrogate.

Central Bureau of Statistics 2005a. Demographic characteristics of university candidates, students and graduates - table 22: student population by degree, age and field of studies. http://www.cbs.gov.il/www/publications/educ_demog_05/educ_demog_h.htm

Central Bureau of Statistics 2005b. Living arrangements of young adults in 2000 by demo-graphic and socio-economic characteristics - Working Paper No. 17 of the Demography Sector. http://www.cbs.gov.il/www/publications/pwmegurim17.pdf.

Central Bureau of Statistics 2006. Students in universities and other higher-education insti-tutes - table 5: percent of women in the student population by degree, field of study and institution. http://www.cbs.gov.il/www/publications/stud_04/stud_h.htm.

Gensch, D.H., Svestka, J.A. 1979. An exact hierarchical algorithm for determining aggregate statistics from individual choice data. *Management Science, 25 (10),* 939-952.

Göritz, A.S. 2004. The impact of material incentives on response quantity, response quality, sample composition, survey outcome, and cost in online access panels. *International Journal of Marketing Research, 46,* 327-345.

Handy, S., Cao, X., Mokhtarian, P.L. 2006. Self-selection in the relationship between the built environment and walking. *Journal of American Planning Association, 72,* 55-74.

Iyengar, S.S., Lepper, M.R. 2000. When choice is demotivating: can one desire too much of a good thing? *Journal of Personality and Social Psychology, 79 (6),* 995-1006.

Intel press room, 2007. Employment status among university students. http://www.intel.com/cd/corporate/pressroom/emea/heb/archive/2007/341609.htm

Jedidi, K., Zhang, J.Z. 2002. Augmenting conjoint analysis to estimate consumer reservation price. *Management Science, 48 (10),* 1350-68.

Kaplan, S., Bekhor, S., Shiftan Y. 2009. Eliciting and estimating reservation price: A semi-compensatory approach. *Journal of Business Research (in press, DOI: 10.1016/j.jbusres.2009.09.019).*

Kohli, R., Mahajan, V. 1991. A reservation-price model for optimal pricing of multi-attribute products in conjoint analysis. *Journal of Marketing Research, 28 (3),* 347-54.

Lichtenstein, D.R., Bloch, P.H., Black, W.C. 1988. Correlates of price acceptability. *Journal of Consumer Research, 15 (2),* 243-52.

Lichtenstein, D.R., Ridgway, N.M., Netemeyer, R.G. 1993. Price perceptions and consumer shopping behavior: a field study. *Journal of Marketing Research, 30 (2),* 234-245.

Manski, C. F. 1977. The structure of random utility models. *Theory and Decision, 8 (3),* 229–254.

McKelvey, R.D., Zavoina, W. 1975. A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology, 4 (1),* 103-120.

Mokhtarian, P.L., Bagley, M.N. 2000. Modeling employees' perceptions and proportional preferences of work locations: the regular workplace and telecommuting alternatives. *Transportation Research Part A: Policy and Practice, 34 (4),* 223-242.

Monroe, K.B. 1971. Measuring price thresholds by psychophysics and latitudes of acceptance. *Journal of Marketing Research, 8 (4),* 460-464.

Morikawa, T. 1995. A hybrid probabilistic choice set model with compensatory and non-compensatory choice rules. In: Hensher, D. A., King, J., Oum, T. Eds. *Proceedings of the 7th World Conference on Transport Research,* Sydney.

Payne, J.W. 1976. Task complexity and contingent processing in decision making: an information search and protocol analysis. *Organizational Behaviour and Human Performance, 16 (2),* 366-387.

Payne, J.W., Bettman, J.R., Johnson, E.J. 1993. The adaptive decision maker. Cambridge University Press, Cambridge, MA.

Quiggin, J.C. 1991. On the optimal design of lotteries. *Economica, 58 (229),* 1-16.

Shocker, A.D., Ben-Akiva, M.E, Boccara, B., Nedungadi, P. 1991. Consideration set influences on consumer decision-making and choice: issues, models and suggestions. *Marketing Letters, 2 (3),* 187-197.

Swait, J., Ben-Akiva, M.E. 1987. Incorporating random constraints in discrete choice models of choice set generation. *Transportation Research Part B: Methodological, 21 (2),* 91-102.

Swait, J. 2001. Non-compensatory choice model incorporating attribute cut-offs. *Transportation Research Part B: Methodological, 35 (10),* 903-928.

Technion, 2007. Facts about the Technion campus. http://www1.technion.ac.il/about

Tversky, A. 1972. Elimination by aspects: a theory of choice. *Psychological Review, 79,* 281-299.

Wang, T., Venkatesh, R., Chatterjee, R. 2007. Reservation price as a range: an incentive compatible measurement approach. *Journal of Marketing Research, 44 (2),* 200-13.

Zheng, J., Guo, J.Y. 2008. A destination choice model incorporating choice set formation. *Proceedings of the 87th TRB Annual Meeting, 13-17 January, Washington, D.C.*

List of tables and Figures

Tables

Table I. Sample socio-economic characteristics.
Table II. The distribution of criteria selection across respondents.
Table III. Minimum by maximum number-of-rooms criteria
Table IV. Monthly rent price threshold as a function of individual characteristics.

Figures

Figure 1. Web-based survey design.
Figure 2. Survey completion by daily period.
Figure 3. The distribution of the number of selected criteria across respondents.
Figure 4. Distribution of selected monthly rent price threshold values.
Figure 5. Selected neighbourhood types across respondents.
Figure 6. Selected maximum walking time to campus across respondents.
Figure 7. Selected noise level thresholds across respondents.
Figure 8. Size of the viable choice set across respondents.
Figure 9. Monthly rent price across choice outcomes as a function of the price threshold value.
Figure 10. Monthly rent price across choice outcomes as a function of the average price of the viable apartments.

Table I. Sample socio-economic characteristics.

Variable	Categories (%)				
Gender	Male 60.1	Female 39.9			
Marital Status	Married 29.5	Single 70.5			
Age	≤ 21 8.2	22-24 21.0	25-29 50.9	30-34 16.9	35-44 3.0
Income source	Scholarship 45.4	Full-time 11.2	Part-time 25.1	None 18.3	
Residential arrangement	Dorms 32.3	Rent alone 5.7	Rent with roommates 12.2	Co-habit with spouse 37.2	Parent's house 12.6
Place of Residence	Haifa 65.0	North Israel 23.3	Center Israel 7.8	Other 3.8	
Car availability	Every day 45.9	2-3 times weekly 9.9	Once a week 10.6	2-3 times monthly 7.8	Rarely 25.8

Table II. The distribution of criteria selection across respondents.

Criterion	Respondents who selected the criterion (%)
Apartment sharing	94.5
Maximal rent price	88.8
Neighbourhood	87.2
Maximal walking time to campus	67.5
Maximal number of rooms	59.5
Maximal noise level	58.1
Minimal number of rooms	53.7
Availability of reserved parking	8.6

Table III. Minimum by maximum number-of-rooms criteria.

min/max	1.0-1.5	2.0-2.5	3.0-3.5	4.0-4.5	> 4.5
1.0-1.5 (%)	0.19	2.86	5.24	4.10	28.31
2.0-2.5 (%)	-	1.14	11.06	17.35	5.34
3.0-3.5 (%)	-	-	0.76	10.68	8.01
4.0-4.5 (%)	-	-	-	0.29	4.67
> 4.5 (%)	-	-	-	-	0.00

Table IV. Monthly rent price threshold as a function of individual characteristics.

Variable	Category	Price threshold model	
		Coefficient	P-value
Marital status	Married = 1, Single = 0	0.540	0.000
Gender	Male = 1, Female = 0	-0.232	0.001
Age	Continuous (in years)	0.052	0.000
Monthly Expenses	< $500 [a]	-	-
	$500-$1,000	0.308	0.001
	$1,000-$1,500	0.441	0.000
	>$1,500	0.612	0.000
Employment status	Unemployed/ Scholarship [a]	-	-
	Part time job	0.155	0.058
	Full time job	0.435	0.000
Car availability	Rarely [a]		
	Weekly	0.222	0.044
	Daily	0.320	0.000
Current place of residence	Middle income neighborhoods in Haifa [a]	-	-
	High income neighborhood in Haifa	0.372	0.000
	Tel Aviv	0.303	0.029
Price Knowledge	Continuous	0.068	0.002
Threshold values	250	0.878	0.004
	300	1.297	0.000
	350	1.559	0.000
	400	1.753	0.000
	450	2.176	0.000
	500	2.526	0.000
	550	3.064	0.000
	600	3.245	0.000
	650	3.500	0.000
	700	3.598	0.000
Number of observations			1,049
Log-Likelihood at estimates			-2164.4
McKelvey-Zavoina Pseudo-R^2			0.35

[a] – base category.

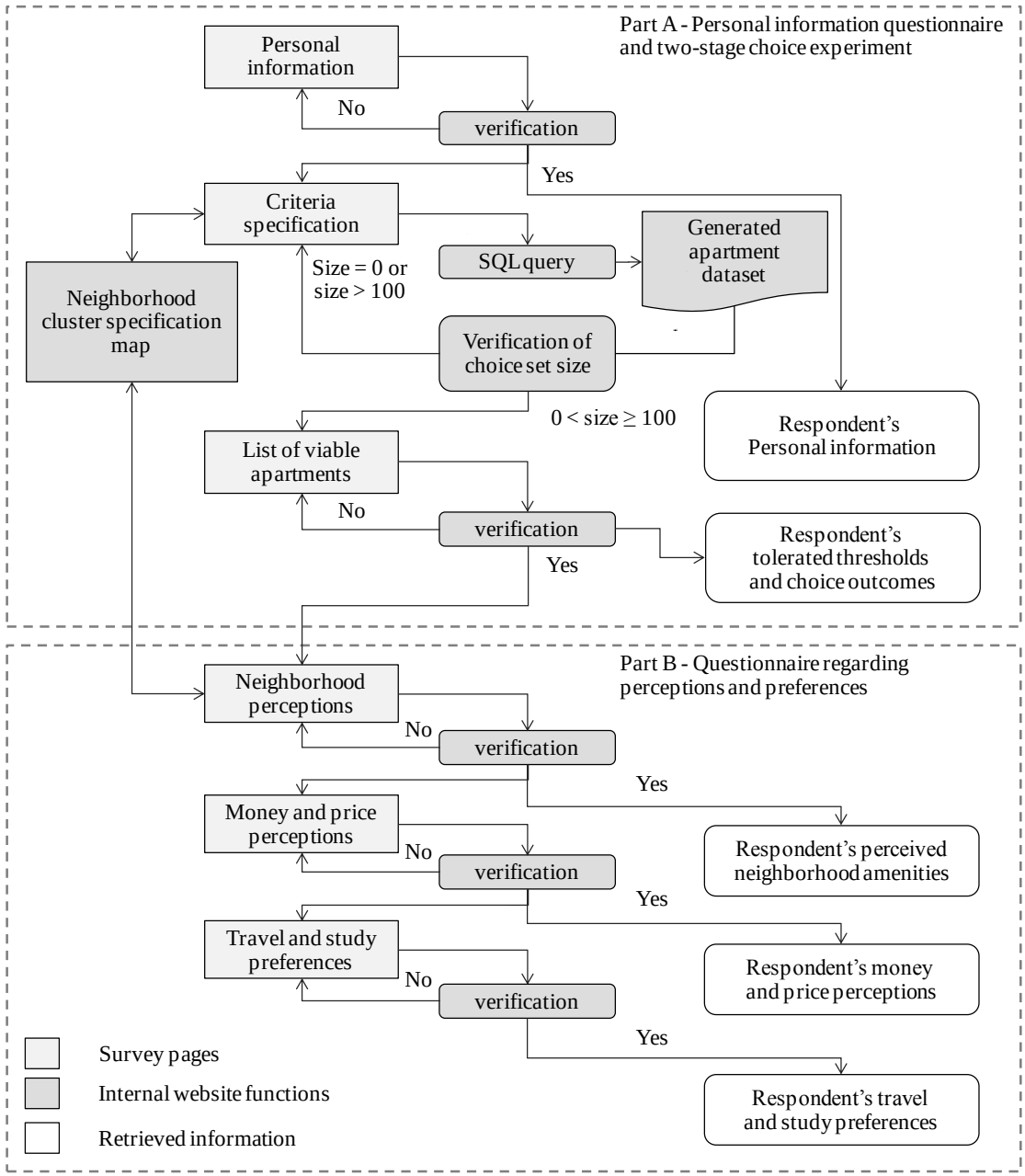

Figure 1. Web-based survey design.

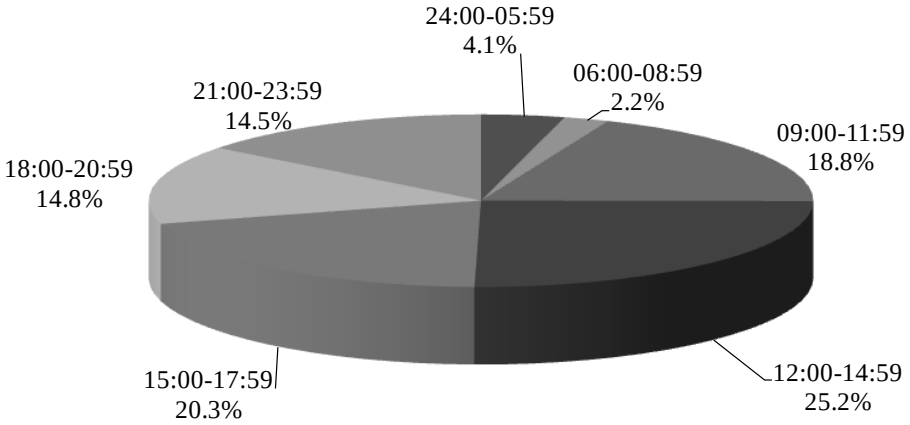

Figure 2. Survey completion by daily period.

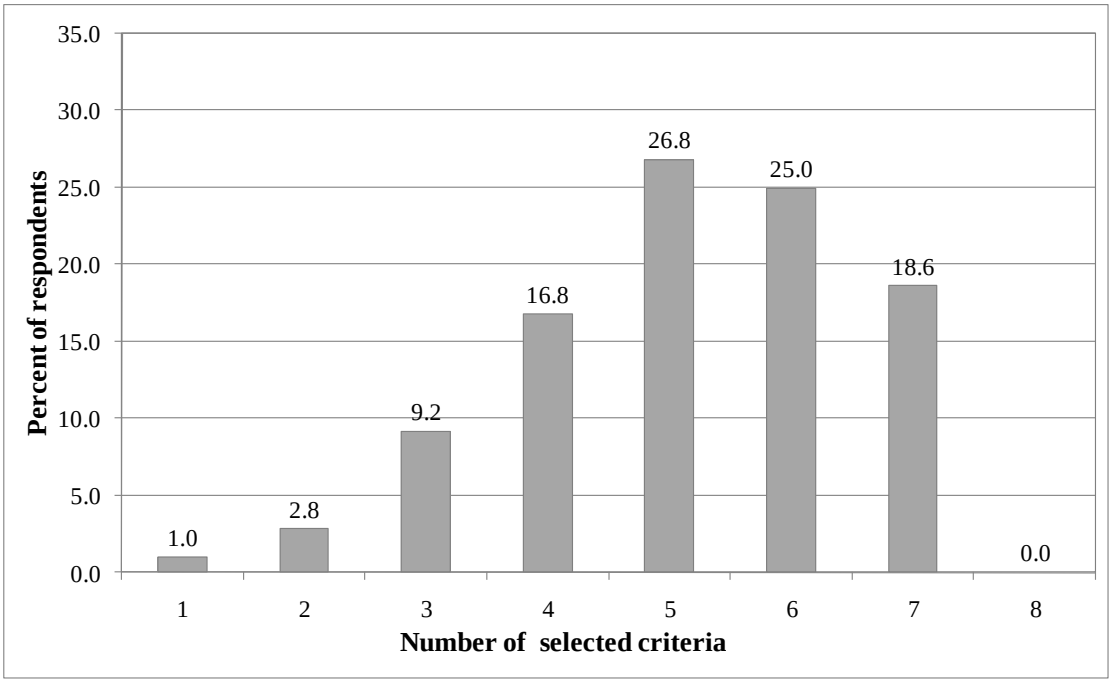

Figure 3. The distribution of the number of selected criteria across respondents.

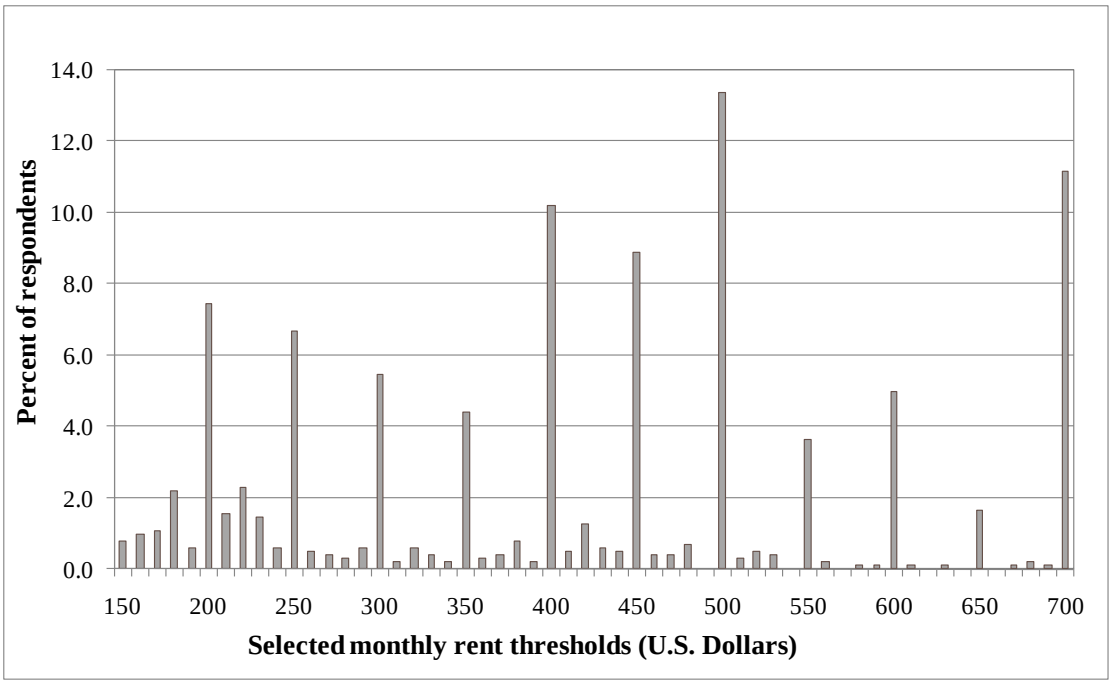

Figure 4. Distribution of selected monthly rent price threshold values.

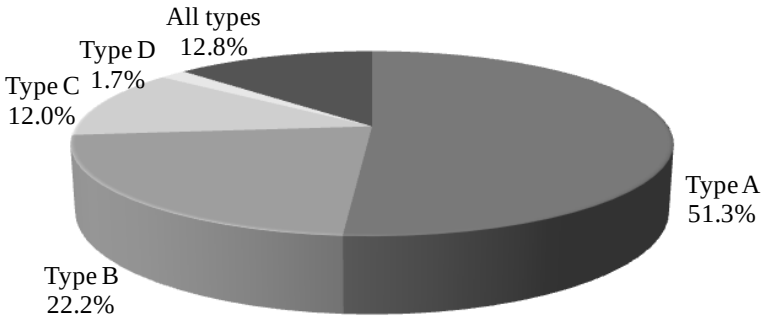

Figure 5. Selected neighbourhood types across respondents.

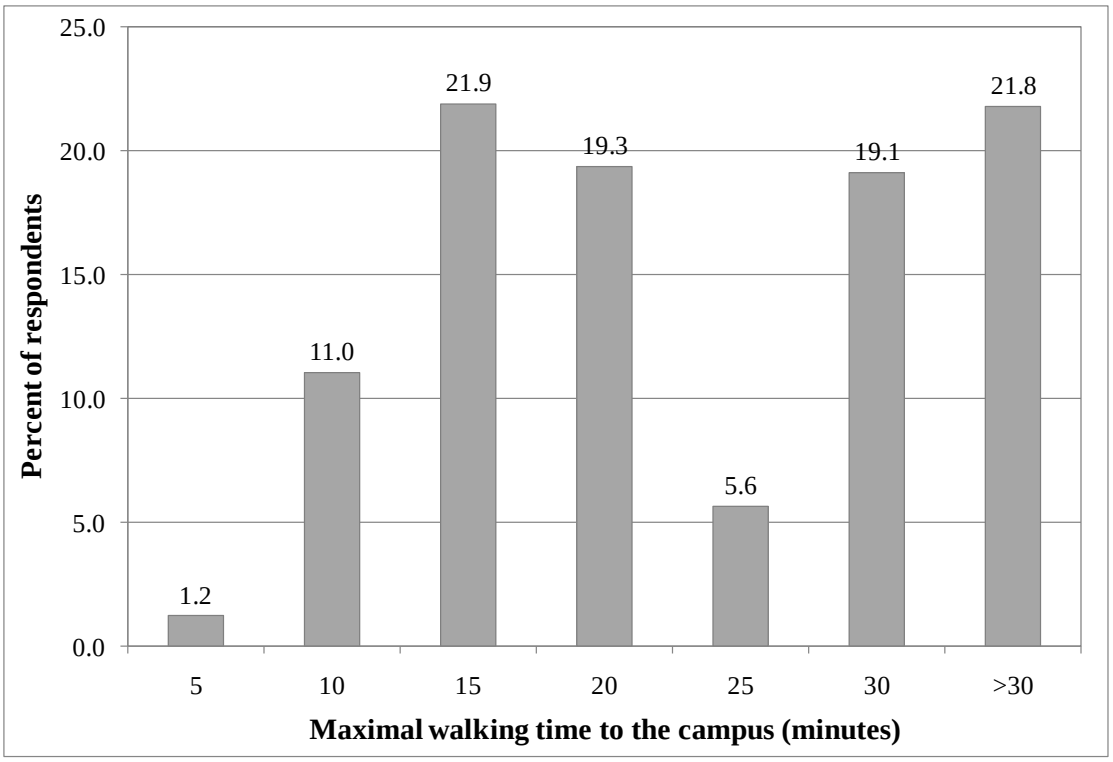

Figure 6. Selected maximum walking time to campus across respondents.

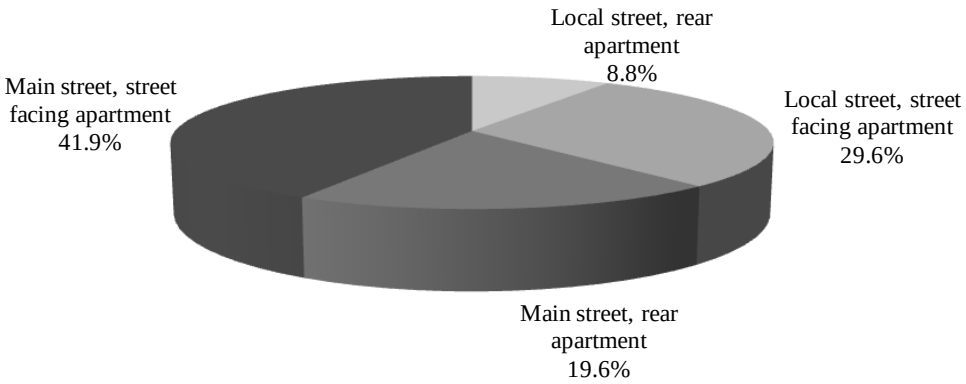

Figure 7. Selected noise level thresholds across respondents.

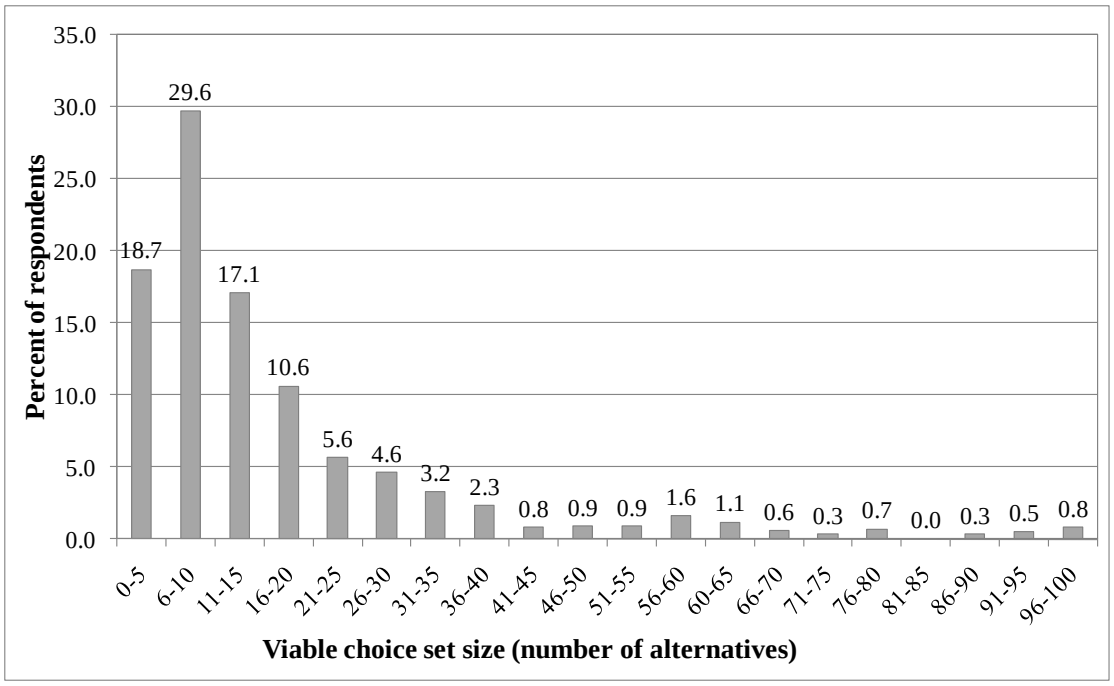

Figure 8. Size of the viable choice set across respondents.

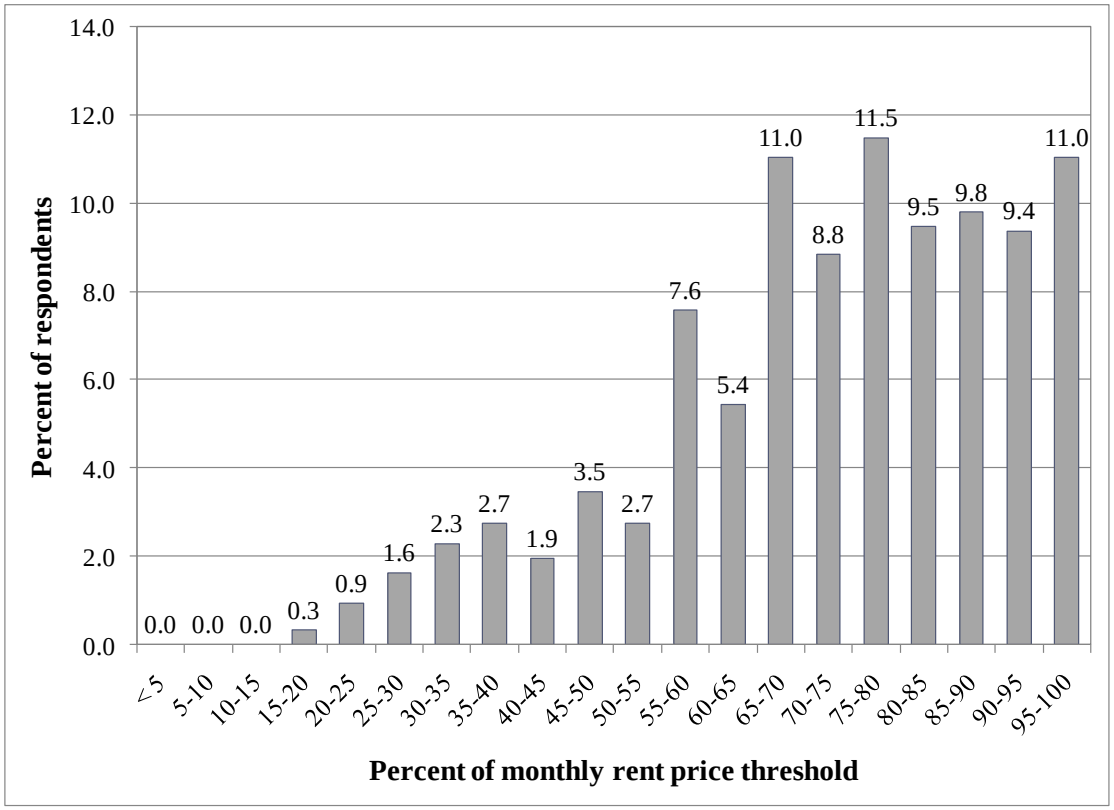

Figure 9. Monthly rent price across choice outcomes as a function of the price threshold value.

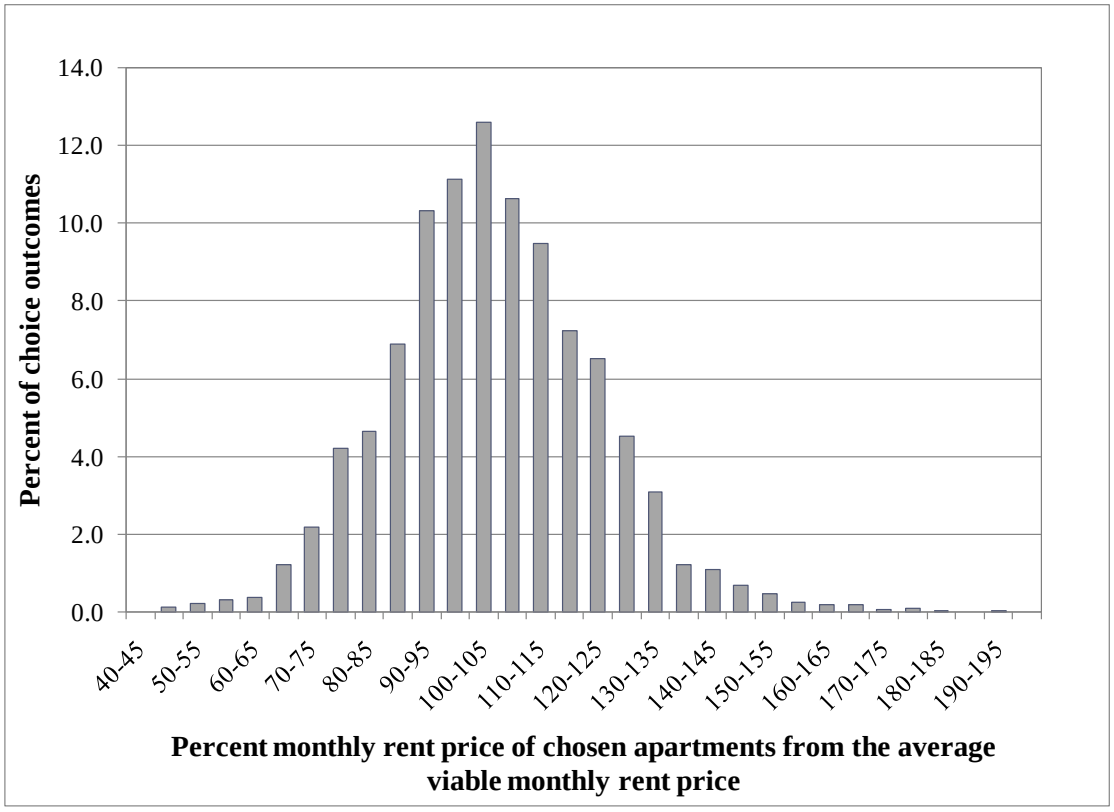

Figure 10. Monthly rent price across choice outcomes as a function of the average price of the viable apartments.