

**An Analysis of the Factors Influencing Differences in  
Survey-Reported and GPS-Recorded Trips**

**Stacey G. Bricka, PhD**

Texas Transportation Institute  
1106 Clayton Ln, Ste 300E, Austin, TX 78723  
Phone: 512-921-6791, Fax: 512-467-8971  
Email: [sgbricka@aim.com](mailto:sgbricka@aim.com)  
(corresponding author)

**Sudeshna Sen, PhD**

NuStats  
206 Wild Basin Rd, Ste A300, Austin, TX 78746  
Phone: 512-279-4147, Fax: 512-306-9077  
Email: [ssen@nustats.com](mailto:ssen@nustats.com)

**Rajesh Paleti**

The University of Texas at Austin  
Dept of Civil, Architectural and Environmental Engineering  
1 University Station C1761, Austin TX 78712-0278  
Phone: 512-471-4535, Fax: 512-475-8744  
Email: [rajeshp@mail.utexas.edu](mailto:rajeshp@mail.utexas.edu)

**Chandra R. Bhat, PhD**

The University of Texas at Austin  
Dept of Civil, Architectural and Environmental Engineering  
1 University Station C1761, Austin TX 78712-0278  
Phone: 512-471-4535, Fax: 512-475-8744  
Email: [bhat@mail.utexas.edu](mailto:bhat@mail.utexas.edu)

July 27, 2010

## **ABSTRACT**

Recent advances in global positioning systems (GPS) technology have resulted in a transition in household travel survey methods to test the use of GPS units to record travel details, followed by the application of an algorithm to both identify trips and impute trip purpose, typically supplemented with some level of respondent confirmation via prompted-recall surveys. As the research community evaluates this new approach to potentially replace the traditional survey-reported collection method, it is important to consider how well the GPS-recorded and algorithm-imputed details capture trip details and whether the traditional survey-reported collection method may be preferred with regards to some types of travel.

This paper considers two measures of travel intensity (survey-reported and GPS-recorded) for two trip purposes (work and non-work) as dependent variables in a joint ordered response model. The empirical analysis uses a sample from the full-study of the 2009 Indianapolis regional household travel survey. Individuals in this sample provided diary details about their travel survey day as well as carried wearable GPS units for the same 24-hour period.

The empirical results provide important insights regarding differences in measures of travel intensities related to the two different data collection modes (diary and GPS). The results suggest that more research is needed in the development of workplace identification algorithms, that GPS should continue to be used alongside rather than in lieu of the traditional diary approach, and that assignment of individuals to the GPS or diary survey approach should consider demographics and other characteristics.

## 1. INTRODUCTION

Over the past decade, the application of global positioning systems (GPS) technology within the context of household travel data collection has become commonplace. The role of GPS has primarily been for the purpose of evaluating the accuracy of trip reporting – identifying trips recorded in the GPS data streams that were not reported by the respondents in the travel diaries. Analyses of these non-reported trips have provided critical insights that have led to improvements of the survey materials and data collection methods, resulting in improved capture of trips via travel diaries.

Recent advances in GPS technology have further resulted in the introduction of lighter weight GPS units with longer battery lives, transitioning the travel survey application of GPS from in-vehicle to on-person/wearable units. This technology improvement has provided the opportunity to begin collecting GPS data for all modes of travel, particularly walking and transit. With the expansion of focus from vehicle travel to travel by all modes of travel, researchers have begun investigating in earnest the possibility of replacing paper travel diaries with GPS-recorded travel. In fact, two large-scale “proof of concept” studies are currently underway in Cincinnati and Jerusalem.

As the travel survey community evaluates the use of GPS as a primary data collection method, it is important to assess the accuracy of the GPS-recorded trips in comparison with the respondent-reported trips obtained through the traditional travel survey diary method. Efforts to date comparing survey-reported trips to GPS-recorded trips have considered the GPS data to be correct. However, GPS units themselves (and the subsequent travel extraction and activity purpose imputation algorithms) are not fail-proof – GPS traces may not capture the start of a trip if a user starts traveling before a satellite fix is obtained; the GPS signal can be lost in urban canyons or tunnels; or the GPS traces may be complete but the trip-detection algorithms may be too sensitive (over-identifying trips) or not sensitive enough (under-identifying trips).

The primary objective of this paper is to review survey-reported trips and GPS-recorded trips for the same individuals in an effort to understand how the number of trips obtained through these two methods differ. While the study does not directly shed light on which of these two sources may be closer to the “truth”, or what explicitly causes the difference between survey-reported and GPS-recorded trips, it does begin to shed light on the individual, household, travel, and contextual factors that contribute to the extent of difference between the two sources. Understanding the factors underlying these differences is a necessary prelude to fine-tuning survey collection efforts in the near term and improving GPS-data collection efforts in the medium-to-long term. In this regard, the analysis begins by considering both the data sources as providing valid measures of travel intensity, albeit potentially with some measurement error. Further, both the number of survey-reported and GPS-recorded trips are modeled jointly using a copula-based framework to account for common unobserved factors that affect these two travel intensity measures. This framework enables us to determine the type as well as magnitude of dependency or the inter-relationship between these two measures of intensities of individual travel. Since the factors determining the amount of individual travel might be different for different activity purposes, two different joint models are reported for work trips of workers (a worker is defined as a person who is either employed for pay or works voluntarily) and non-work trips of all respondents. Of course, once the joint model for the two trip-making propensities is estimated, this is translated into the effects of variables on the difference of the survey-reported and GPS-recorded trips.

This paper contributes to the travel survey literature in several important ways. It is the first to consider GPS data from a large-scale on-person subsample of a household travel survey (n=265 individuals). Prior papers on related topics have focused on in-vehicle GPS data and/or smaller scale wearable studies. The study also undertakes the analysis using recently collected GPS data (from 2009) that accommodates several technological advances unavailable in the GPS data collection efforts from the earlier part of this decade. Second, unlike earlier studies that have considered GPS units as providing the “true” measure of travel intensity, and have focused on trip under-reporting in surveys (see next section), the emphasis here is on the difference between survey-reported and GPS-recorded trips, without making any *a priori* assumptions about the accuracy of the data from the two sources. In doing so, we consider cases where the difference in trip-making between the two sources may be positive, negative, or equal (on the other hand, earlier studies have focused on the case where the difference between survey and GPS trips is negative; see, for example, Bricka and Bhat, 2006). Finally, in this transitional time from GPS as a supporting survey method to GPS as a potential (partial or full) replacement for the travel diary survey data collection method, the current study provides insights into the factors that affect the difference between the trip capture from the two sources, which can then be used in subsequent research to further invest resources on the kinds of individuals and travel environments that result in larger magnitudes of trip differences between the two data sources. One important limitation of the current study also needs to be noted: this paper accepts the survey and GPS samples as is – considerations of self-selection bias and the representativeness of the sample are not addressed in this first, but important, effort to examine differences in trip capture between the two sources. Besides, given the focus of the current research, issues of sample representativeness are of secondary consideration.

The remainder of this paper is structured in five sections. The next section provides a summary of GPS efforts to date, while Section 3 presents an overview of the Indianapolis regional household travel survey and its descriptive GPS sample characteristics. Section 4 discusses the model structure and estimation procedure. Section 5 focuses on the empirical results. The final section summarizes the important findings from the results, and recommends specific improvements for capture of travel by both diary and GPS survey methods.

## **2. GPS IN HOUSEHOLD TRAVEL SURVEYS**

Over the past decade, GPS technology has been increasingly used in travel behavior research to assess and evaluate the accuracy of travel data reported in diaries. Through such evaluations, and assuming that GPS data provides “true” measures of trip-making, it has been documented that household travel survey data, by virtue of its reliance on self-reported information by respondents, suffer from incompleteness and inaccuracies of reported trips. As a result, the under-reporting of trips is seen as a major drawback of household travel surveys.

Most prior GPS studies have explicitly focused on under-reporting of trips in household travel studies. Murakami and Wagner (1999) compared the in-vehicle GPS data collected for a six-day period with the self-reported trips for one travel day in the 1996 GPS “proof of concept” Lexington study. The study recognized the potential of GPS to identify under-reporting of trips in self-reported surveys, including inaccuracies in reporting trip start times, travel times, and trip distances. The potential of GPS to audit trip reporting levels was realized in the first GPS-enhanced household travel survey conducted in Austin, Texas in 1997 (Pearson, 2001).

To date, the levels of trip “under-reporting” estimates in the United States range from a low of 10% in Kansas City to a high of 81% in Laredo, Texas (Wolf *et al.*, 2004; Wolf, 2006).

The “under-reporting” of trips in household travel surveys can be driven by the survey methodology, respondent burden, respondent demographics and attitudes, and their travel characteristics. Wolf *et al.* (2003a; 2003b) found the main causes of trip “under-reporting” to be survey length, memory decay of respondents, failure to understand or to follow survey instructions, respondents considering the trips unimportant, unwillingness to report full details of travel, and carelessness.

Zmud and Wolf (2003) compared survey diary-reported trips (or simply “survey trips” from hereon) with GPS-recorded trips, and identified key demographic factors that may contribute to misreporting of trips, again assuming GPS to be the correct source of data (in the rest of this section, for ease in wording, we will assume the GPS data to be the “truth”, as has been the norm in earlier studies). Zmud and Wolf found that the respondents who travel less and those from smaller size households or households with fewer vehicles are more accurate reporters of trips in survey efforts. On the other hand, the likely misreporters of trips are households with three or more vehicles, households with annual incomes of less than \$50,000, respondents younger than 25 years of age, and households with three or more workers. The factors influencing survey trip under-reporting were also examined using data from the 2001 Los Angeles household travel survey and the 2002 St. Louis household travel survey, and reveal similar findings (NuStats, 2003; 2004). In these studies, the results also indicated higher under-reporting of short duration trips. Pierce *et al.* (2003) also estimated trip under-reporting by comparing the average vehicle and person trip rates, and found that trip under-reporting was more prevalent in low-income households. In addition, discretionary trips were found to be more likely to be under-reported than non-discretionary trips.

In another study, Forrest and Pearson (2005) found significant impacts of household size and household income on trip reporting accuracy. They also noted that single-member households showed the highest reporting accuracy at 63%, while households with two members had the lowest reporting accuracy at 31%. Generally, households in higher income groups showed higher reporting accuracy than those in the lower income groups. While the aforementioned studies modeled the likelihood of trip under-reporting, Bricka and Bhat (2006) jointly modeled the likelihood of trip under-reporting as well as the level of trip under-reporting (*i.e.*, the number of trips under-reported). Their findings suggested that young adults (less than 30 years of age), men, individuals with less than high school education, unemployed individuals, individuals working in clerical and manufacturing professions, workers employed at residential, industrial, and medical land-uses, and individuals in nuclear families are all more likely to under-report trips in household travel surveys than other respondents.

In addition to demographics, travel behavior characteristics were found to significantly influence trip under-reporting. Wolf (2000) claims that the “often forgotten or omitted” trips in paper diaries are trips of short duration that occur as part of a trip chain or those trips that are short round trips that do not stop at the destination long enough to trigger trip detection. Similar findings were obtained from the Kansas City GPS study by Wolf *et al.* (2004). In that effort, missed trips were found to be mostly trips made at the end of the travel day or trips associated with short stops on the way to another destination that were missed due to forgetfulness of the respondent or unimportance associated by the respondent to the stop. Forrest and Pearson (2005) found significantly high levels of trip under-reporting in Laredo, Texas for home-based non-work and non-home-based trips. Bricka and Bhat (2006) found that making a high number of trips on the survey day, traveling long distances per trip, and trip chaining affect the tendency to under-report trips in household travel surveys. Stopher *et al.* (2007) reported that individuals

with high trip making levels were more likely to underreport their travel, as were those making trips after 5 P.M. Also, trips with shorter time duration and trip distance were more likely to be under-reported, as were those associated with activities with a short duration (under 10 minutes). Their study also indicated that trips for social visits showed the highest rates of under-reporting, followed by those for picking up or dropping off a passenger.

Fewer studies have focused on the situation where the GPS unit recorded fewer trips than that reported by the respondent. Wolf *et al.* (2003a, 2003b) found GPS trips that were not identified in the in-vehicle GPS data stream but reported in the survey data. They speculated that the reasons for missed GPS trips were associated with the lack of power to the GPS units due to delayed installation at the start of the day, unplanned end-of-the-day trips after the removal of the GPS unit, or misreporting due to loss of signal or power disconnect between the GPS unit and the vehicle during midday. Wolf *et al.*'s study was one of the first to bring up possible issues associated with GPS data collection. Forrest and Pearson (2005) found some level of higher reporting from survey diaries than recorded through GPS units. They believed that the main reasons for the apparent over-reporting of home-based work trips was the effect of trip linking of survey responses (linking of trips with brief stops) or miscoding of GPS trips made to secondary work or work-related locations.

An important dimension of trip reporting assessment is the type of GPS device used for passive data collection. All the aforementioned studies in the U.S. have used in-vehicle GPS units for passive data collection. Wolf (2006) reported that the primary reasons for the "vehicle-based" approach compared to a wearable GPS-based approach are the convenient use of the power system of vehicles to charge the GPS units, better signal reception associated with mounting the device on the vehicle than being worn on the body, and reliability of data collection as compared to dependence on the respondent to remember wearing the device while traveling. However, in-vehicle GPS units are unable to collect information on trips made by other modes of transportation such as walk, bike, and transit. Due to this limitation, wearable GPS devices are now the preferred method for collecting personal travel data. The first study using wearable GPS devices was conducted in Netherlands in 1997 to test the suitability of GPS across different travel modes (Draijer *et al.*, 2000).

With the introduction of lighter-weight wearable GPS devices, there has been a steady increase in the use of these devices for recording trips. A review of the literature shows that wearable GPS travel studies have been conducted world-wide, including in the United States (Giaino *et al.*, 2009), United Kingdom (Steer Davies Gleave and Geostats, 2003), Japan (Ohmori *et al.*, 2005), Belgium (Kochan *et al.*, 2006), Netherlands (Bohte and Maat, 2008), and South Africa (Krygsman and Nel, 2009). Steer Davies Gleave and Geostats (2003) examined data from a wearable GPS subsample of London Area Transport Survey respondents for a three-day period. The study reported that 82% of all the possible days for data collection yielded useful GPS data. The remaining 18% were not as usable due to human error such as not carrying the GPS device correctly, or forgetting to carry it at all. Also, the study estimated the level of survey trip under-reporting at 16%. In the recent household travel study conducted in Western Cape, South Africa, Krygsman and Nel (2009) showed significant under-reporting of trips and tours originating from home and work. The missed trips in the self-reported diaries were primarily shorter walk or strolling trips. In addition, the study reported under-estimation of trip duration due to rounding off of trip departure and arrival times by respondents. To date, only two household travel surveys in the U.S. have used wearable GPS devices to collect personal travel data. They are the 2009 Indianapolis household travel survey (which is the focus of this paper)

and the 2009/2010 GPS household travel survey in Cincinnati. The Cincinnati household travel survey is an on-going, large scale GPS-based survey of 1,500 households that includes a prompted recall survey element (Giaino *et al.*, 2009).

As should be clear from above, most earlier studies have used GPS-recorded trips to examine the accuracy of survey-reported trips. Almost all earlier studies use the terminology “survey under-reporting” that suggests that the GPS is the “ground truth”. Unlike these earlier studies, the emphasis in the current study is to examine the factors that are responsible for more or less discord (and the direction and level of the discord) between survey-reported and GPS-recorded trips.

### **3. DATA FORMATION AND DESCRIPTIVES**

#### **3.1 Study Overview**

The empirical analysis in the current paper uses data extracted from the Central Indiana Travel Survey that was conducted in 2009, under the sponsorship of the Indianapolis Metropolitan Planning Organization. As part of the Indianapolis survey, complete demographic and travel behavior characteristics of 3,929 randomly sampled households were obtained, including details of 36,516 weekday trips for 9,337 household members. The wearable GPS component of the full-study documented travel for all persons aged 16+ across 136 households. Both CATI and GPS data are available for 272 individuals from the 136 households. However, only 265 of the 272 individuals reported data for all demographic variables of interest. Thus, all subsequent analyses in the current paper focus on these 265 individuals who answered all demographic variables of interest.

The 265 individuals considered in this analysis reported a total of 1,533 trips as part of the travel survey. The GPS units carried by these individuals recorded a total of 1,555 trips. In terms of the trips reported in the survey travel diary and/or detected in GPS:

- 57 individuals (22%) reported more trips in the travel survey than were detected in GPS.
- 131 individuals (49%) reported the same number of trips in the travel survey as were detected in GPS.
- 77 individuals (29%) reported fewer trips in the travel survey than were detected in GPS.

Prior studies evaluating GPS-detected trips, as summarized in Section 2, suggest that differences in survey-reported and GPS-recorded trips may vary based on trip purpose (work vs. non-work trips). Given the importance of the work commute in transportation planning, as well as results from earlier studies showing that non-reported survey trip rates are different between work and non-work trip purposes, this paper analyzes work trips and non-work trips separately. A work trip, in the current analysis, is defined as a trip that has one end as home and the other end as work (usually characterized as the home-based work trip in trip-based modeling). A non-work trip is one that is not a work trip (this is a combination of home-based other trips and non-home based trips, as usually characterized in trip-based modeling). These definitions were used to calculate the number of survey-reported trips in each category. For the GPS-recorded trips, the GPS contractor tagged trips as work or non-work, based on the respondent’s reported home and primary work locations. Of the 265 individuals being studied, 205 were employed, either full-time, part-time, or on a regular volunteer basis. These 205 working individuals were the focus of analysis for work trips (non-workers, by definition, did not report work trips). For the non-work trips, the GPS-recorded data showed four individuals making one non-work trip each. However, there were no corresponding non-work trips in the survey data for these same

individuals. As a result, these four individuals were removed, leaving a sample of 261 individuals for the non-work trip analysis.

The number of work trips ranges from 0 trips to 4 or more trips, while the number of non-work trips ranges from zero trips to 13 or more trips, excluding the one trip category as noted above. The data included individual demographics, individual work-related characteristics, other individual attributes, household-related variables, and day of week variables.<sup>1</sup>

### 3.2 Descriptive Analysis of the Sample

In considering work travel, most individuals reported the same number of trips in the survey as recorded by the GPS unit (63%), while 32% reported more survey trips relative to GPS trips, and 5% reported fewer survey trips relative to GPS trips. When considering non-work travel, however, there is a dramatic difference: 18% had more survey trips, 42% equally reported, and 40% had less survey trips. These trends suggest that, when it comes to work travel, the survey-reported data is perhaps more complete, but for non-work travel, the GPS data is perhaps more complete (consistent with the findings of Forrest and Pearson). However, also note that these trends may simply be an artifact of inaccurate location specificity in determining purpose from the GPS data. In particular, if the trip purpose algorithm only tags the primary workplace as “work”, but respondents report trips to secondary employment locations or report working at multiple locations such as client offices (as found in Forrest and Pearson), the result would be a higher number of work trips from the survey and a higher number of non-work trips from the GPS unit.

Several variables were considered in the descriptive analysis in this section and in the final econometric model specification discussed later in the paper. These variables include:

- (1) Individual Demographics - race (Caucasian or non-Caucasian), age, gender, education attainment (less than high school, high school, technical school/college without degree, undergraduate degree, and post graduate degree), student status (yes or no, with yes reflecting full-time and part-time students), role of the individual in the household (head of household versus not),<sup>2</sup> disability status, and driving license availability,
- (2) Individual Work-related Characteristics - employment status (full time: 30+ work hours per week, part time: less than 30 work hours per week, voluntary, retired, and unemployed), number of work days per week (less than 5 days, 5 days, and more than 5 days), work flexibility (low, medium, and high), and mode to work (car, transit, walk, bike, and other),<sup>3</sup>

---

<sup>1</sup> The Indianapolis effort focused on weekday travel only. Here, the day of week variables were used to test the hypothesis that there might be differences in travel behavior across days of the week, particularly on Mondays and Fridays.

<sup>2</sup> A person is defined as the head of the household if s/he is recruited on behalf of the household to provide responses to the recruitment survey.

<sup>3</sup> Work flexibility, mode to work, and number of work days per week are defined only for individuals who work full-time, part-time, or voluntarily. Mode to work is as reported in recruitment and represents the usual or typical mode, and not the actual mode to work reported on the travel day.



- (3) Other Individual Attributes - transit usage (whether the individual actually used transit on the travel day), non-motorized mode usage (whether the individual reported walking or biking on the travel day), and whether the person was interviewed,<sup>4</sup>
- (4) Household-related Variables - household income ( less than \$15K, \$15K to \$40K, \$40K to \$60K, \$60K to \$75K, \$75K to \$100K, and \$100K or more), location of the household (central business district residence versus non-central business district residence), length of time lived at current residence (less than an year, 1 to 2 years, 3 to 5 years, 6 to 10 years, more than 10 years), number of vehicles, number of license holders, number of workers, presence and number of children (less than 4 years, 5 to 11 years, and 12 to 15 years), and internet availability at home, and
- (5) Day of Week Variables – the weekday on which travel was recorded.

Tables 1 and 2 show the distribution of each respondent group with respect to the difference between survey-reported and GPS-recorded trips. There are three columns for each group: “survey trips > GPS trips”, “survey trips = GPS trips”, and “survey trips < GPS trips.” The “survey trips < GPS trips” column reflects the traditional comparison in earlier GPS studies (*i.e.*, the number of survey-reported trips is less than the GPS-recorded trips). The tables reveal that several variables appear to influence the direction and magnitude of the difference between survey and GPS trips. While the focus of the multivariate modeling analysis later in this paper is to rigorously identify the direction and magnitude effects of these variables, the univariate statistics from the tables reveal the following general trends (organized below by direction of difference between survey and GPS trips, and activity purpose):

- Work Trips - Survey trips > GPS trips (first column of Table 1): In terms of individual demographics, there are no substantial differences based on ethnicity, Hispanic origin, age, and gender. However, those with a high school education (relative to those with below high school education and above high school education) tend to report more work trips than detected in GPS (note the 40% entry for “high school” relative to the lower percentage entries for other education levels). Non-students also show a higher tendency to report more survey trips than GPS trips. The individual worker characteristics reveal that individuals who are full-time employees, work five or more days per week, have low work flexibility, and drive to work report more work trips than detected in GPS, relative to individuals who are not full-time employees, work less than five days per week, have moderate to high work flexibility, and do not drive to work, respectively. When considering other individual characteristics, those who use non-motorized modes for at least one work trip during the survey day tend to be less likely to report more survey trips than GPS trips relative to those who do not use non-motorized modes at all. Among household-related variables, individuals from high income (>\$75K) households, located in non-CBD areas, resident in their current location for more than 2 years, in households with no students, in household with two members, and reporting owning two vehicles are more likely to report more survey work trips than GPS trips (relative to their peers). These results are consistent with Zmud and Wolf (2003) and Forrest and Pearson (2005).

---

<sup>4</sup> The “Interviewed” variable is coded as “Yes” if the response was provided by the individual concerned and as “No” if some other household member served as a proxy when reporting travel data.

- Work Trips - Survey trips < GPS trips: For ease in presentation, we will only identify the groups who are more likely to report fewer survey trips relative to GPS trips. Of course, these results should not be viewed in the absolute, but relative to corresponding peer groups not listed. Individuals with a graduate degree are more likely to report fewer survey trips for work than detected in GPS. In addition, those who travel to work by a mode other than car (particularly non-motorized travelers) also are more likely to be in this category, as are individuals with no vehicles available to them.
- Non-Work Trips - Survey trips > GPS trips: For non-work travel, individuals of Hispanic origin are likely to report more survey trips than GPS trips, as are those with less than a high school education. Other individual attributes associated with more survey non-work trips include those who use transit or walk/bike for at least one non-work trip during the survey day. At the household level, those more likely to be in this category are those with zero or one vehicle.
- Non-Work Trips - Survey trips < GPS trips: Non-Caucasian and Hispanic individuals are likely to have fewer survey trips than GPS trips, as are workers who work more than five days per week and those who report using transit for at least one non-work trip. Friday travelers are more likely to show more GPS non-work trips than survey trips

Of course, the descriptive statistics in Tables 1 and 2 provide suggestive evidence of variations in the completeness of data obtained through the two data collection methods (survey and GPS). However, these are univariate statistics in that they do not control for the influence of other variables when examining the impact of any single variable. To obtain a comprehensive picture of the factors affecting the differences as well as relationships between measures of intensity of travel obtained through each data collection method, it is necessary to pursue a multivariate and comprehensive analysis. In the next section, we present the model structure and empirical analysis for such a methodology.

## 4. METHODOLOGY

### 4.1 Background

In our empirical case, the number of trips that an individual undertakes on a given day is available from two different sources - (1) Survey reported information and (2) GPS-recorded information. The dependent variables of interest- number of trips by activity purpose (*i.e.* work and non-work) from each of these two sources is modeled using an ordered response structure, which is based on the notion that there is an underlying latent continuous trip-making propensity measure that underlies the observed number of trips made. As discussed later, the underlying latent propensity measure is tied to the observed number of trips through threshold bounds that horizontally partition the latent measure (see McKelvey and Zavoina, 1975 for a general discussion of ordered-response models, and Bhat and Pulugurta, 1998 for a discussion in the context of travel demand models). Such an ordered-response structure explicitly accommodates the non-negative, ordinal, and discrete choice nature of number of trips. On the other hand, traditional methods that use a linear regression approach ignore the non-negative, ordinal and discrete nature of number of trips (see Bhat and Singh, 2000 for additional discussion). Further, count data models such as the Poisson or negative binomial distribution models restrict “the amount of probability mass that can be accommodated at any one point” (see Herriges *et al.*, 2008). Thus, in cases with a high fraction of ‘0’ values, as in the current empirical context, the count models are not very suitable. The alternative of adding zero-inflated approaches to

accommodate the high number of '0' values, while easy to undertake in a univariate count model, becomes difficult in the current bivariate count case where we are modeling both the number of survey-reported trips and GPS-recorded trips (for each of work trips and non-work trips) jointly. The reason for modeling the number of survey-reported and GPS-recorded trips jointly is to account for common unobserved factors that may affect both trip-making propensities of the same individual, since these are essentially measures of travel intensity of the same individual. For example, a dynamic and energetic person may make a high number of trips, which would result in a high number of survey-reported trips and GPS-recorded trips. If "being dynamic and energetic" is an individual characteristic not captured in surveys, the net effect is a positive dependence in the stochastic (unobserved) components of trip-making propensities underlying the survey and GPS trips. On the other hand, individuals who are intrinsically busy and forgetful may not have time to record all their trips, and if "being busy and forgetful" are individual characteristics not captured, the result would be negative dependence in the stochastic (unobserved) components of trip-making propensities underlying the survey and GPS trips.

In the past, unobserved dependencies of the type discussed above in two ordinal variables have been captured by using a normal marginal distribution for the unobserved error terms in the ordinal propensity equations, and coupling these error terms using a Gaussian copula. However, such a bivariate normal structure imposes the restriction that the dependence between the error terms is linear and radially symmetric (see Bhat and Eluru, 2009). In general, one does not know *a priori* what kind of dependency structure holds between the unobserved factors influencing the two propensities. Rather this is an empirical issue to be determined based on which dependency surface fits the data best. Accordingly, in this paper, we test alternative copula-based bivariate distributional assumptions to couple the error terms, and select the copula dependence form (for tying the unobserved error terms in the two propensity equations) and the univariate marginal distributions (for the unobserved error terms in each propensity equation) that provide the best fit. In doing so, we consider radially asymmetric as well as radially symmetric copula surfaces. For instance, consider the case of positive dependence in the error terms. In this case, it is possible that the positive dependence is much more at the low end of the trip propensity spectrum (because, for example, there is less opportunity for forgetting to report trips in surveys and less possibility of mis-recording of GPS trips) than at the high end of the trip propensity spectrum. That is, individuals may be clustered very tightly at the low-low levels of the two-dimensional survey-reported trip propensity and GPS-recorded trip propensity spectrum, and less so at the high-high end of the spectrum. This results in asymmetric positive dependency, which cannot be captured by the symmetric bivariate normal distribution.

Of course, the final end-objective of this paper is to model the effects of demographic, work-related and other individual attributes, household related variables, and day of week effects on the difference between the number of survey-reported and GPS-recorded trips. While it may seem that one approach to do so would be to directly use the difference between these variables in an ordered-response structure (since the difference will also be ordinal and discrete), this is not appropriate because of the potential dependence between the two trip-making propensities (as discussed earlier). The appropriate procedure would be to model the two trip-making propensities explicitly, and then obtain the effects of variables on the difference of the two dependent trip variables, as is done in the current analysis.

## 4.2 Model Structure for Joint Copula Model of Survey-Reported and GPS-Recorded Trips

The model structure is the same for the two categories of work and non-work trips, and so we present the generic structure without reference to activity purpose. For each individual  $q$  ( $q = 1, 2, \dots, Q$ ), let  $f_q$  represent the number of trips reported in the survey and let  $g_q$  represent the number of trips recorded in the GPS. Let  $m$  be an index for the number of survey-reported trips ( $m = 0, 1, 2, \dots, M$ ) and let  $n$  be the index for the number of GPS-recorded trips ( $n = 0, 1, 2, \dots, M$ ).<sup>5</sup> The equation system takes the following form:

$$\begin{aligned} f_q^* &= \alpha'x_q + v_q, \quad f_q = m \quad \text{if } \delta_{m-1} < f_q^* < \delta_m \\ g_q^* &= \beta'y_q + \eta_q, \quad g_q = n \quad \text{if } \psi_{n-1} < g_q^* < \psi_n \end{aligned} \quad (1)$$

where  $f_q^*$  and  $g_q^*$  are the latent propensity measures of travel characterizing the number of survey and GPS trips, respectively,  $x_q$  and  $y_q$  are exogenous variable vectors (with no constant terms),  $\alpha$  and  $\beta$  are corresponding coefficient vectors to be estimated,  $v_q$  and  $\eta_q$  are random error terms, and the  $\delta_m$  and  $\psi_n$  terms represent thresholds that relate the latent propensity measures  $f_q^*$  and  $g_q^*$  to their observed counterparts  $f_q$  and  $g_q$ , respectively, in the usual ordered-response fashion ( $\delta_{-1} = -\infty, \delta_M = \infty; -\infty < \delta_0 < \delta_1 < \delta_2 < \dots < \delta_{M-1} < \infty$ ) and ( $\psi_{-1} = -\infty, \psi_M = \infty; -\infty < \psi_0 < \psi_1 < \psi_2 < \dots < \psi_{M-1} < \infty$ ). The error terms  $v_q$  and  $\eta_q$  may take any parametric distribution, and the nature of the dependency between the two error terms is determined by the type of copula used. However, in the current study, we examine only logistic and normal marginal univariate distributions for the error terms, and choose the distribution that provides the best data fit. That is, the error terms  $v_q$  and  $\eta_q$  are assumed to be either standard normal or standard logistic distributed, and these error terms are assumed to be independent and identically distributed (IID) across individuals  $q$ . For presentation ease, let the marginal distribution of  $v_q$  be  $F(\cdot)$  and the marginal distribution of  $\eta_q$  be  $G(\cdot)$ . Also, for notational convenience, define  $b_{qm} = \delta_m - \alpha'x_q$  and  $d_{qn} = \psi_n - \beta'y_q$ .

With the preliminaries above, the probability that an individual  $q$  reports  $m$  trips in the survey and GPS records  $n$  trips for the same individual can be written as follows:

$$\begin{aligned} \Pr[f_q = m, g_q = n] &= \Pr[b_{q,m-1} < v_q < b_{qm}, d_{q,n-1} < \eta_q < d_{qn}] \\ &= \Pr[v_q < b_{qm}, \eta_q < d_{qn}] - \Pr[v_q < b_{qm}, \eta_q < d_{q,n-1}] \\ &\quad - \left\{ \Pr[v_q < b_{q,m-1}, \eta_q < d_{qn}] - \Pr[v_q < b_{q,m-1}, \eta_q < d_{q,n-1}] \right\} \end{aligned} \quad (2)$$

The above joint probability depends upon the dependence structure between the random variables  $v_q$  and  $\eta_q$ . In the current paper, we use a flexible copula-based approach to characterize the dependence between these error terms. As indicated earlier, the copula approach allows the testing of several types of dependence structures, so that the analyst can choose the

<sup>5</sup> We used the same categorization of the dependent variable for the GPS as well as survey-reported trips for both the work and non-work trips. So, the maximum value that the index takes in our empirical context is  $M$  in both the cases, though this need not be the case in general.

one that best fits the data rather than pre-imposing the very restrictive, but commonly used bivariate normal (BVN) distribution assumption. More generally, let the joint cumulative distribution function of  $v_q$  and  $\eta_q$  be  $H_{v,\eta}(z_{q1}, z_{q2})$ . Then,  $H_{v,\eta}(z_{q1}, z_{q2})$  can be expressed as a joint cumulative probability distribution of uniform  $[0,1]$  marginal variables  $U_1$  and  $U_2$  as below:

$$\begin{aligned} H_{v,\eta}(z_{q1}, z_{q2}) &= \Pr[v_q < z_{q1}, \eta < z_{q2}] = \Pr[F^{-1}(U_1) < z_{q1}, G^{-1}(U_2) < z_{q2}] \\ &= \Pr[U_1 < F(z_{q1}), U_2 < G(z_{q2})]. \end{aligned} \quad (3)$$

Then, by Sklar's (1973) theorem, the above joint distribution (of uniform marginal variables) can be generated by a function  $C_\theta(\dots)$  such that:

$$H_{v,\eta}(z_{q1}, z_{q2}) = C_\theta(u_{q1} = F(z_{q1}), u_{q2} = G(z_{q2})). \quad (4)$$

where  $C_\theta(\dots)$  is a copula function and  $\theta$  is a dependency parameter (assumed to be scalar), together characterizing the dependency between  $v_q$  and  $\eta_q$ .

The probability expression in Equation (2) can be re-written in terms of the copula function as:

$$\begin{aligned} \Pr[f_q = m, g_q = n] &= C_\theta[F(b_{qm}), G(d_{qn})] - C_\theta[(F(b_{qm}), G(d_{q,n-1}))] \\ &\quad - \{C_\theta[(b_{q,m-1}), G(d_{qn})] - C_\theta[F(b_{q,m-1}), G(d_{q,n-1})]\} \end{aligned} \quad (5)$$

A variety of bivariate copula functions are available, and we test several of these for appropriateness in the current empirical context. These include the independence copula (essentially equivalent to independence between  $v_q$  and  $\eta_q$ ), the traditional Gaussian copula (*i.e.*, the bivariate normal dependency structure), the Farlie-Gumbel-Morgenstern (FGM) copula, and the Archimedean class of copulas (including the Clayton, Gumbel, Frank, and Joe copulas). The reader is referred to Bhat and Eluru (2009) for a detailed discussion of these alternate copulas and the visual plots of their implied dependency. Due to space considerations, we are unable to provide additional details on the structures of the different copula types here.<sup>6</sup>

### 4.3 Model Estimation

The parameters to be estimated in the joint bivariate ordered response model include the  $\alpha$  and  $\beta$  vectors, the  $M$   $\delta_k$  parameters ( $\delta_{-1} = -\infty, \delta_M = \infty; -\infty < \delta_0 < \delta_1 < \delta_2 < \dots < \delta_{M-1} < \infty$ ), the  $M$   $\psi_n$  parameters ( $\psi_{-1} = -\infty, \psi_M = \infty; -\infty < \psi_0 < \psi_1 < \psi_2 < \dots < \psi_{M-1} < \infty$ ), and the  $\theta$  parameter characterizing the dependency between the error terms for the copula under consideration. To write the log-likelihood function, define  $I_q(m, n)$  as an indicator variable that takes the value of 1 if individual  $q$  reports  $m$  trips in survey and the GPS records  $n$  trips for the same individual and 0 otherwise. Then, the log likelihood function for the copula model takes the following form:

---

<sup>6</sup> An important note here. Many of the Archimedean copulas (including the Clayton, Gumbel, and Joe copulas) can only accommodate positive dependencies (unlike the FGM, Gaussian, and Frank copulas). Thus, these copulas cannot even handle the situation of potential negative dependence. However, to examine the appropriateness of these copulas for the potential presence of dedicated loyalty effects, one only has to re-formulate the model system in Equation (1) by introducing the  $v_q$  term in the first equation with a negative sign.

$$\log L = \sum_{q=1}^Q \sum_{m=0}^M \sum_{n=0}^M I_q(m, n) \log \Pr[f_q = m, g_q = n] \quad (6)$$

#### 4.4 Effects of Variables on the Difference between Survey and GPS Trips

Once the joint model is estimated, we can obtain the probability of all possible combination of trips  $m$  and  $n$  from the expression (5). Next, the probability of observing a difference of  $y = k$  between the survey-reported and GPS-recorded trips for an individual  $q$  can be obtained as follows:

$$h_q(y = k) = \sum_{m \in S_k} \Pr[f_q = m, g_q = m - k] \quad (7)$$

where  $S_k$  is the set of possible values of  $m$  (*i.e.*, number of survey-reported trips) for which the difference  $y$  between the survey-reported and GPS trips can be  $k$ . For example, let the ordinal trip-making categories be 0, 1, 2, and 3. Then, the probability that  $y = 1$  is the sum of the following joint probabilities:  $\Pr[f_q = 1, g_q = 1 - 1]$ ,  $\Pr[f_q = 2, g_q = 2 - 1]$ , and  $\Pr[f_q = 3, g_q = 3 - 2]$ . So, in this case  $S_k = \{1, 2, 3\}$ . Once we have the probability of each possible value of the difference  $y$ , the expected value of the difference  $y$  for an individual  $q$  can be computed as:

$$E_q(y) = \sum_k h_q(y = k) * k \quad (8)$$

A negative value for this expected value would imply more GPS-recorded trips than survey-reported trips (or under-reporting in the survey, or over-capture by the GPS unit, or some combination of the two), while a positive value would imply fewer GPS-recorded trips than survey-reported trips (or over-reporting in the survey, or under-capture by the GPS unit, or some combination of the two).

## 5. EMPIRICAL RESULTS

### 5.1 Variable Specification

Several functional forms for the variables were considered in the empirical specification. For example, age was introduced linearly, using a piecewise-linear approach, as well as dummy variables with different cut-off points. The final model specification (including the variables included, the functional form of variables, and interaction effects of variables) was based on intuitive considerations, insights from previous literature, parsimony in specification, and statistical fit/significance considerations. The final specification includes some variables that are not statistically significant at the usual 5% level of significance. In particular, we have retained some variables whose corresponding parameters have a t-statistic (for the test against zero) of close to 1 or slightly more. This is because of the relatively small sample size (from a statistical standpoint) for the analysis. Further, the effects of these variables are intuitive, and have the potential to guide future research that may have the benefit of larger sample sizes.

### 5.2 Copula Specification and Dependency Effects

The empirical analysis involved estimating models with two different univariate (*i.e.*, marginal) distribution assumptions (normal and logistic) for the error terms  $v_q$  and  $\eta_q$ , and seven different copula structures (independence, Gaussian, FGM, Clayton, Gumbel, Frank, and Joe). As

discussed in Section 4.1, in the copula approach, there is no need to assume that the marginal distributions of the  $v_q$  and  $\eta_q$  error terms are simultaneously normal (normal-normal) or logistic (logistic-logistic); instead  $v_q$  and  $\eta_q$  terms can have a normal-logistic or logistic-normal distribution. We examined all these four possible combinations for the error terms  $v_q$  and  $\eta_q$ , as well as the seven copula dependency structures, for a total of 28 copula-based models for each of the work trip and non-work trip categories. The Bayesian Information Criterion (BIC) was employed to select the best copula model from among the competing non-nested copula models with non-independent copula forms for the error terms  $v_q$  and  $\eta_q$  (see Quinn, 2007, Genius and Strazzer, 2008, Trivedi and Zimmer, 2007, page 65), since the traditional likelihood ratio test for comparing these alternative copula-based models is not applicable. The BIC for a given copula model is equal to  $-2\ln(L) + K\ln(Q)$ , where  $\ln(L)$  is the log-likelihood value at convergence,  $K$  is the number of parameters, and  $Q$  is the number of observations. The copula that results in the lowest BIC value is the preferred copula. However, since all the competing models in the current analysis have the same exogenous variables and the same number of thresholds and constants, the BIC information selection procedure measure is equivalent to selection based on the largest value of the log-likelihood function at convergence.

Among the non-independent copula models, the Clayton copula with standard normal marginal distributions for the error terms  $v_q$  and  $\eta_q$  provided the best fit for both the work as well as non-work trip categories. The log-likelihood values at convergence for these copula models are -387.27 for work trips and -1039.15 for non-work trips.<sup>7</sup> The dependence parameter in the Clayton copula is 2.713 (with a standard error of 0.620 and a t-statistic of 4.37 for the test with respect to the value of zero) for work trips and 4.129 (with a standard error of 0.506 and a t-statistic of 8.152 with respect to the value of zero) for non-work trips. Both these parameters are positive (as required in the Clayton copula) and very highly statistically significant (relative to the value of '0', which corresponds to independence), indicating the strong positive dependence among the unobserved determinants of survey-reported and GPS-recorded trips for both the work and non-work trip purposes. Another common way to quantify the dependence in the copula literature is to compute the Kendall's measure of dependence.<sup>8</sup> For the estimated association parameters, the values of the Kendall's  $\tau$  are (standard errors are in parenthesis) 0.5757 (0.056) for work trips and 0.6737 (0.027) for non-work trips. Among the final four independence copula models, the Normal-Normal Independence (NNI) model provided the best data fit, with a likelihood value of -418.77 (for work trips) and -1159.45 (for non-work trips). Since both the Clayton copula and the NNI models have the same univariate normal margins for both  $v_q$  and  $\eta_q$ , they can be compared using a likelihood ratio test. The chi-squared test statistic for

<sup>7</sup> The corresponding log-likelihood values at sample shares were -439.28 (for work trips) and -1116.58 (for non-work trips). Log-likelihood ratio tests of the final Clayton copula model with these naïve sample share models clearly reject the sample share models at any reasonable level of significance, underscoring the predictive ability of the final Clayton copula models.

<sup>8</sup> See Bhat and Eluru (2009) for a description of this dependency measure. The traditional dependence concept of correlation coefficient  $\rho$  is not informative for asymmetric distributions, and has led statisticians to use concordance measures. Basically, two random variables are labeled as being concordant (discordant) if large values of one variable are associated with large (small) values of the other, and small values of one variable are associated with small (large) values of the other. This concordance concept has led to the use of the Kendall's  $\tau$ , which is in the range between 0 and 1, assumes the value of zero under independence, and is not dependent on the margins. For the Clayton copula,  $\tau = \theta / (\theta + 2)$ .

work trips is 62.99 for work trips, and 240.59 for non-work trips, very strongly rejecting the null hypothesis of independence between the trip propensities underlying the survey-reported and GPS-recorded numbers of trips.

The dependence form of the Clayton copula implies that the dependency in unobserved components affecting the survey-reported and GPS-recorded trip propensities is strong at the left tail, but not at the right tail. Figure 1 plots the dependency scatter plot of the relationship between the unobserved components  $v_q$  and  $\eta_q$ . As can be observed, the results indicate that there is much tighter clustering (more positive dependence) at the low end of the trip propensity spectrum than at the high end of the spectrum. That is, there is much better conformity in the survey-reported and GPS-recorded trip-making at low trip intensity levels than at high levels. This is intuitive, since, as explained earlier, there is less opportunity for forgetting to report trips in surveys and less possibility of mis-recording of GPS trips at low trip intensity levels than at high trip intensity levels. The net result is an asymmetric positive dependency, which cannot be captured by the symmetric bivariate normal distribution.

### 5.3 Model Estimation Results

To conserve on space, we only present the results for the best Clayton copula model with normal marginals.<sup>9</sup> The results are provided in Tables 3 (for work trips) and 4 (for non-work trips). The coefficients provide the effects of variables on the latent trip-making propensity of an individual as obtained from the survey-reported data and the GPS-recorded data. In this section, we briefly discuss the overall effects of variables on both the trip-making propensities from the survey and GPS data (since both these propensities are measures of travel intensity of the individual). The discussion of the effects of variables on the difference between the survey-reported and GPS-recorded trips, which is the focus of this study, is presented in Section 5.4.

The thresholds at the top of the tables do not have any substantive interpretations. They simply serve the purpose of mapping the latent propensity into the observed frequency levels. Also note that, for dummy exogenous variables, the base category is explicitly identified in the tables and the text discussion below.

#### 5.3.1 Work Trips

In the category of individual demographic variables, there are no differences in work trips based on race and age for females, according to the survey-reported data. However, according to the GPS-recorded trips, individuals of Caucasian race and young women have a higher propensity to undertake work trips compared to non-Caucasians and older women, respectively. Further, the GPS data indicates that middle-aged individuals (30-50 years) have a higher work trip-making propensity than those who are older than 50 years. Both the survey and GPS data suggest higher work trip-making for young males (less than or equal to 30 years) than for older males. Individuals with education attainment of less than an undergraduate degree have a higher propensity to make work trips than those with an undergraduate degree, perhaps because those with an undergraduate degree are able to telecommute more often, given their positions in the career ladder. However, according to the GPS-reported trips (but not from the survey data), those with a graduate degree also have a higher work trip propensity than those with an undergraduate

---

<sup>9</sup> The estimates from the other copula models and the independent model were, as one would expect, different from those obtained from the Clayton copula model. Further, the standard errors of the model estimates in the Clayton model with normal marginals were, in general, smaller than those from the other models, indicating efficiency benefits as well from using the Clayton copula structure.



degree, potentially a reflection of the managerial and supervisory roles that individuals with high levels of education hold in organizations, requiring them to be in their office on a daily basis. Also, non-students make more work trips than students.

The individual work-related characteristics reflect the lower levels of work trip-making for those who are voluntarily employed compared to those who are employed for pay. Interestingly, this is the only common effect estimated from the survey and GPS data. Other results are as follows: (1) Individuals who indicate that they are part-time employed have a lower work trip propensity than those indicating full-time employment (from GPS data), even after accounting for the number of work days per week, (2) those working more than or equal to five days indicate more survey-reported work trips in the survey, but, surprisingly, the GPS data shows lower work trip-making propensity for those working more than five days relative to those working five days or less, and (3) those who drive to work report more work trips in the survey relative to those who do not drive to work, perhaps because of under-reporting of non-motorized and non-car trips or issues with identifying non-primary work locations in the GPS data.

Among other individual attributes, according to the GPS data (but not the survey data), those who reported making at least one work trip by walking or biking on the travel day have a higher work trip propensity than those who do not. The lack of effect of this variable from the survey data may be a result of individuals forgetting to report some of their very short distance trips.

The household-related variables suggest higher trip-making propensity for individuals in households with low-income and high income compared to individuals in households with middle income. An additional effect is the lower work trip propensity of individuals in households with more students, but this result manifests itself only in the survey data.

The day of week effects indicate a lower work trip-making tendency on Fridays compared to other days, potentially because of individuals occasionally taking off from work on Fridays to have a long weekend or working a compressed work week.

### 5.3.2 Non-Work Trips

Individuals of Caucasian heritage, according to the GPS data, have a higher propensity to undertake non-work trips relative to individuals of other races (this is a result that has been observed in several earlier studies; see, for example, Misra and Bhat, 2000). The race difference is discernible from the survey data only for the non-work trips of individuals from large households (*i.e.*, households with more than 4 individuals). Age effects from the survey data indicate a higher non-work trip intensity for older individuals relative to those under the age of 50. Also, individuals who are highly educated (undergraduate degree or higher) make more non-work trips than those with a lower education, while the head of the household makes fewer non-work trips than other individuals in the household.

The work-related variables indicate higher non-work trip intensities for retirees relative to other individuals. In addition, according to the survey data, those voluntarily employed and those who are not employed have higher non-work trip intensities than those working part-time or full-time. Also, the more work days per week for those who are employed leads to a lower non-work trip propensity. All these work-related effects are intuitive, and suggest time constraints based on work times and work commitments.

The survey data suggests that those who are interviewed directly have a higher non-work trip-making propensity than those whose data is provided by proxy. In the past, the fact that those whose data were obtained by proxy reported fewer non-work trips was often attributed to

the proxy reporting process (*i.e.*, the person doing the reporting did not know about non-work trips). However, as shown in Section 5.4, the GPS-recorded trips generally match the survey-reported trips for the “proxy” group, suggesting simply that this “proxy” group has lower non-work trip making propensities.

The household-related variable effects are as expected. Individuals in households with high incomes (more than \$40K) and high auto availability levels have a higher propensity to make non-work trips than those in households with low incomes and low auto availability, respectively. Higher trip-making propensities are also associated with individuals residing in the Indianapolis central business district (compared to those residing outside the central business district) and those who have lived less than two years at their current residence location (relative to “old-timers” in the region).

Finally, non-work trip-making intensity is higher on Fridays relative to other week days, which is the reverse of the effect of this variable on work trip-making.

#### **5.4 Estimates of the Effects of Variables on the Difference between Survey and GPS Trips**

The parameter estimates of variables in the joint copula model of survey and GPS trip propensities do not directly provide the magnitude of the effects of variable on the expected value of the difference between the survey-reported and GPS-recorded trips. To do so, we compute the aggregate “elasticity” effects of variables. These elasticity effects are “pseudo-elasticity” measures that determine the change in the expected value of the difference in trips between the survey-reported and GPS-recorded trips as a percentage of current expected value of trip-making level, due to a change in a dummy variable from 0 to 1 for each (and all) individuals. An issue that arises here is the current expected value of trip-making level – is it what is obtained from the survey data or what is obtained from the GPS data. In our analysis, we use the average of the expected values of trip-making levels from the survey data and the GPS data to represent current trip-making levels. Note also that all variables in Tables 3 and 4 are dummy variables, except for “number of students” and “number of vehicles per license holder”. The first of these variables is an ordinal variable, while the second is the ratio of two ordinal variables. For the purposes of estimating variable effects, we consider this second variable to also be ordinal, by changing the value of the variable through an increase in the number of vehicles variable (while leaving the number of licenses variable constant).

To compute the aggregate-level “elasticity” effect of a dummy exogenous variable, we change the value of the variable to one for the subsample of observations for which the variable takes a value of zero and to zero for the subsample of observations for which the variable takes a value of one. We then sum the shifts in the expected value of the difference in trips as predicted based on the survey and the GPS data sources in the two subsamples after reversing the sign of the shifts in the second subsample, and compute the change in the expected value of difference in survey and GPS trips across all individuals in the sample (as a percentage of current expected value of trip-making across all individuals) due to a change in the dummy variable from 0 to 1. To compute the aggregate level elasticity effect of an ordinal variable, we increase the value of the variable by 1 and compute the percentage change in expected difference due to a unit increase in the variable. Note that the expected value of the difference between the survey-reported and GPS-recorded trips can be obtained directly using Equation (8).

Table 5 provides the “elasticity” effects for work trips. The entry corresponding to the “Race” category in the table indicates that the difference in the survey-reported and GPS-recorded number of work trips is, on average, about 29.4% more in magnitude for Caucasians

than for non-Caucasians. The negative sign reveals that the survey-reported trips tend to be fewer relative to GPS-recorded trips for Caucasians compared to non-Caucasians. One way to interpret this result is as follows. Consider an “average” Caucasian and an “average” non-Caucasian, each of whom have a “true” value of trip-making of 3. Then, according to our results, if the GPS-recorded trips is “true” and the number of survey-reported trips for the non-Caucasian is equal to the number of GPS-recorded trips (*i.e.*, no difference in survey and GPS trips for non-Caucasians), then the Caucasian can be expected to report only 2.12 [=3\*(1-0.294)] work trips in the survey. That is, Caucasians under-report work trips in the survey by 29.4% compared to non-Caucasians. On the other hand, if the number of survey-reported trips is “true”, and again there is no difference in the numbers of survey and GPS trips for the non-Caucasian, then the GPS unit can be expected to record 3.88 [=3\*(1+0.294)] work trips for the Caucasian. That is, the GPS unit over-records work trips by 29.4% compared to non-Caucasians. Of course, both the survey and GPS unit may not provide the truth, in which case the conclusion would be that the difference between survey and GPS trips is 29.4% higher in magnitude (with fewer survey trips relative to GPS trips) for Caucasians than for non-Caucasians. Similarly, the first entry in the “*Age and Gender*” category in the table shows that young women (less than or equal to 30 years of age) also tend to have a higher difference by about 43% between survey and GPS trips (with fewer survey trips than GPS trips) relative to older (>50 years) women. On the other hand, the second entry in the “*Age and Gender*” category indicates that young men tend to have a higher difference by about 8.3% between survey and GPS trips (with more survey trips than GPS trips) relative to older men. Other effects may be similarly interpreted.

In reviewing the elasticities, we continue the evaluation by three groupings: (1) Individuals for whom the survey and GPS trips are most similar, (2) Individuals for whom the survey trips reported tend to be higher than GPS-recorded trips, and (3) Individuals for whom the GPS trips tend to be higher than survey trips. For the first group, the stability in recording trips does not vary that much by data collection method, which implies that we can have confidence in the data, regardless of the data collection method applied. Note that the groups that constitute the base category in Tables 5 and 6 are the ones that most belong to the first group, since the effects in these tables are shown with regard to an effective “0” difference between the survey and GPS trips (for instance, as discussed earlier, non-Caucasians have less disparity between survey and GPS trips compared to Caucasians. Further, in the case of multiple dummy variables to represent a particular variable (such as education attainment), those dummy variables with a low magnitude of elasticity in the tables also belong to the first group.

The results in Table 5 for the work trip show the following characteristics associated with each group:

- (1) Survey and GPS work trips are most similar – Non-Caucasians, older individuals (more than 50 years of age), males under the age of 30, with an undergraduate degree, non-students and those working full-time, those working fewer than five days a week, those who travel to work by non-car and non-motorized modes, and individuals in middle-income households and in households with fewer students have the most consistency in the number of work trips obtained by both data collection methods. Further, the results indicate that day of week does not play any substantial role in differences between survey trips and GPS trips. This is an interesting group in that males under the age of 30 are typically associated with “under-reporting” of work trips in household surveys (see Bricka and Bhat, 2006, for example), supporting the use of GPS technology to capture trips more accurately for young males. However, our results suggest that this group

dependably reports work trips, regardless of survey approach. Overall, for travelers with the characteristics identified above, both traditional survey diaries and GPS can be used interchangeably with confidence that the data obtained is largely the same regardless of method used.

- (2) Survey work trips greater than GPS trips – The individuals in this group include those who are employed part-time, work 5+ days per week, and travel to work typically by car. Reported household income associated with these individuals includes both those earning less than \$40,000 per year and those earning \$75,000 or more. Here, we believe that the trip purpose identification algorithms are too restrictive in identifying work locations and should be expanded to consider multiple job locations (either due to working two or more jobs, or associated with those who have multiple job locations and travel among them on a given day; note that those with low and high incomes are likely to have multiple job locations). The large differences associated with this group are particularly problematic in that trip-based and tour-based models focus on the work trip as an anchor of travel. Our results suggest that any consideration of a GPS-only study must include a stronger focus on how work locations are identified in the data streams to more accurately detect work trips.
- (3) Survey work trips less than GPS trips – This group is the most studied to date, with the traditional focus on survey “under-reporting.” As explained above, it is unclear whether the differences in work trips reported for these individuals come from under-reporting in the survey or over-detecting in the GPS, or both. The individuals comprising this group include Caucasians, females under the age of 30, individuals between the ages of 30 to 50, those with less than an undergraduate degree and those highly educated, those who volunteer, and those who reported a walk or bike trip to work on the travel day. Here, the issue may be related to trip-chaining of busy individuals who do not keep good survey records or who travel more and thus their GPS-recorded trips are subject to possible over-detection due to traffic patterns in the areas in which they travel. In addition, the fact that those who reported at least one trip by walk/bike fall into this category points to possibly an overly-sensitive trip detection algorithm being used. These results suggest that when determining which data collection method to use, volunteers and these busy mothers should be placed into the GPS data collection category.

A similar investigation into the magnitude of differences between survey-reported and GPS-recorded trips for non-work purposes is provided in Table 6. Given the difference in the number of work vs. non-work trips, the elasticity calculations have different denominators and thus cannot be directly compared. The non-work results are presented below, using the same reporting structure as for work trips:

- (1) Survey and GPS non-work trips are most similar – when considering non-work travel, the non-work trips from the survey and GPS methods are similar (or at least less dissimilar) for non-Caucasians in small households (less than 4 individuals), young individuals ( $\leq 50$  years of age), non-head individuals, those working full or part-time, those whose data is provided by proxy, individuals who do not use transit and non-motorized modes at all for non-work purposes, those with high household incomes ( $\geq \$40$  K) and resident in their current location for over two years, and for non-Friday travel (all these are relative to equivalent peers not belonging to the categories listed above). Interestingly, education attainment, work days per week, residential location, number of vehicles, and number of

licensed individuals do not have much of an impact on the dissonance (or consistency) between survey and GPS trips. Overall, these appear to be individuals with a busy and structured work-life, and with lower levels of non-work trip-making propensities (supporting the earlier literature that the lower the intensity of travel, the more accurate the reported data).

- (2) Survey non-work trips greater than GPS trips – The individuals in this group include those who perhaps can be categorized as having a more leisurely paced life: the person identified as the head of household, who volunteers, is retired, or not employed, who was interviewed in person, and who reported at least one trip by walk or bike. Here, the fact that the non-motorized travel was captured at higher rates in the survey data as compared to the GPS data speaks again to the need to review the algorithm specifications. The remaining demographic characteristics speak to a respondent group with a more leisurely life style, with potentially more time to record trip details and with very set patterns so the destinations are well known. Another possibility here is that the individuals reported their typical travel in the survey, while the GPS documented their actual travel on the travel day.
- (3) Survey non-work trips less than GPS trips – This group is again the most studied to date, with the focus on survey “under-reporting.” Individuals with more non-work GPS trips than survey trips (or at least higher negative discordance between survey and GPS trips) include Caucasians, non-Caucasians who are part of a large household (>4 individuals), individuals over the age of 50, transit users, and those who more recently moved to the region. The newcomers may drive more tentatively than those who have lived in the region for a longer time, which may be reported in the GPS as stops. Friday travel is the final characteristic of individuals in this group.

## 6. CONCLUSIONS

This paper has examined the differences in measures of intensity of travel as obtained using survey-reported and GPS-recorded data collection methods for two trip purposes (work and non-work) through the application of a joint ordered response model. The research models both the propensity to record trips of each purpose, as well as the propensity to have a discord between survey-reported and GPS-recorded numbers of trips. The empirical analysis in the paper uses data extracted from the 2009 Indianapolis regional household travel survey. Specifically, household members aged 16+ who both reported their travel patterns using a survey diary and who carried wearable GPS units for the same 24-hour diary period were selected for the analysis. The findings from this research can inform both survey-reported as well as GPS-recorded travel data collection approaches.

In terms of work trip reporting, those respondents with a higher propensity to travel for work purposes are Caucasian, younger (under the age of 30), with lower education levels. Those employed have higher work-trip making propensities than volunteers, and within workers, those who are employed full-time have higher levels than part-time workers. Those with lower propensities to travel for work include students and individuals in households with a higher number of students, those who are employed part-time or who volunteer, and those who reported travel for a Friday. These results are consistent with the literature, validating the research approach.

Individuals characterized as Caucasian, more highly educated, and not employed for pay (retired, not employed, or volunteer), with higher incomes living in the central business district

had a higher propensity to travel for non-work purposes. On the other hand, those employed for pay had lower propensities to make non-work trips, with that propensity decreasing as the number of days worked per week increased. Again, these findings are intuitive in that those with more discretionary time would make more discretionary trips.

When considering the magnitude of differences between the survey-reported and GPS-recorded trips for both work and non-work, five main conclusions can be drawn:

1. For the most part, the majority of trips identified for both work and non-work purposes are similar for the two data collection methods – 63% of all work trips and 42% of all non-work trips – were identified to have equal numbers of trips in the survey-reported and GPS-recorded data, as indicated at the top of Tables 1 and 2. However, where the magnitude of trip making identified through each survey method does not match, there are inconsistencies in terms of which survey method obtained more trips and which survey method obtained more accurate trip details. This suggests that improvements are needed in both methods to yield more stable trip rates for the trip purposes studied here. For this reason, we recommend that GPS-only studies be undertaken with caution until methodological improvements result in more consistent data being obtained from both sources.
2. GPS can and should be used to obtain travel details from the younger generation – either those still in college or those who are not 1- or 2-person worker households. The results of this study show that the levels of trip reporting for work and non-work trips among this technologically savvy respondent group are the same and given their propensity to prefer technology, GPS should be deployed to this group whenever possible. GPS should also be used for those respondents characterized as heavy travelers and those more predisposed to making non-motorized trips. This includes workers who travel as part of their work, or, as indicated in this study, busy volunteers, middle-aged individuals (30-50 years of age), and those who have characteristics associated with the tendency to trip chain, as such individuals appear not to fully record their travel using the traditional survey method.
3. Any consideration of collecting data using GPS only should be undertaken with caution, with particular care given to the specification of the trip purpose and mode detection algorithms. As indicated in this study, work trips associated with those who hold more than one job or travel to different workplaces on the travel day were not detected in the GPS data at the same levels of intensities as compared to survey data. Given that most algorithms consider the primary work locations only, respondents that work at multiple sites throughout the work day/week or work multiple-jobs at different locations report more work trips in the survey but the GPS-recorded trips reflect a portion of those as non-work trips. Other explanations include the possibility that the buffers used to compare the work location to the GPS-recorded trip location are too tight (*i.e.*, they need to be expanded to capture off-site parking) or that the work locations are within urban canyons or other locations where satellite receptivity is poor and those GPS data points are excluded from analysis. In addition, travel by transit and non-motorized modes were also associated with significant differences in trips reported, indicating that the trip detection algorithms require review.
4. The results also suggest that the traditional survey approach may be better suited for the elderly, retirees, and those who can be characterized as having more leisurely travel.

More research is needed to ascertain why these individuals report more survey trips than GPS, with a concern that perhaps these respondents are not carrying the GPS units for all trips, or that they are reporting in their surveys “typical” travel as compared to actual travel on the assigned travel day. This group also includes those new to the region, and who perhaps travel more tentatively than those who have lived in the region for more than two years. The tentative driving style of the newcomers may result in more GPS-recorded non-work trips as compared to survey-reported data due to over-detection of stops in the GPS data streams. Given the cost of deploying GPS, combined with the fact that the traditional survey method appears to capture trips sufficiently for these individuals, this may be a group to target for the traditional and less-costly diary/survey approach.

5. An interesting finding of this research, and one that requires further study, is the phenomenon of Friday travel. As reported herein, individuals who travel on Fridays tend to travel more for non-work purposes than work purposes. The work trips obtained by both the survey and GPS are consistent with each other and the literature, which indicates that both data collection methods tend to match at lower trip making intensities. For non-work travel, survey-reported trips tend to be fewer than GPS-recorded trips, suggesting that there is more trip chaining and special travel patterns on Fridays as compared to other weekdays. For this reason, we recommend that when GPS is used in a travel survey, a higher proportion of Friday travelers be equipped with the GPS units to better capture and measure travel patterns on this specific day of the week.

In summary, this paper has examined the propensities and magnitude of trip making as reported through the survey data and as recorded using GPS units for both work and non-work purposes. The findings suggest that instead of an either/or approach to data collection, that the two survey methods complement each other and should be used in tandem. Any study considering GPS-collection only should strongly consider the details used to develop the algorithms to assign trip purpose, as our results show that the use of primary workplace location only tends to under-record work trips. However, GPS should be considered as the data collection method when dealing with the younger, more technology savvy individuals as well as those that have high travel propensities or characteristics associated with trip chaining, in order to ensure that all trip details are recorded. However, for the elderly and more leisurely travelers, the traditional survey method is recommended

#### **ACKNOWLEDGEMENTS**

A special thank you to Andrew Swenson, Indianapolis MPO, for permission to utilize the data for this analysis, to Sally Amen for her assistance in preparing Tables 1 and 2, and to Lisa Macias for her assistance in typesetting and formatting this document. The contents of this paper reflect the opinion of the authors only.

**REFERENCES**

- Bhat, C.R., and N. Eluru (2009). A copula-based approach to accommodate residential self-selection effects in travel behavior modeling. *Transportation Research Part B*, 43(7), 749-765.
- Bhat, C.R., and V. Pulugurta (1998). A comparison of two alternative behavioral choice mechanisms for household auto ownership decisions. *Transportation Research Part B*, 32(1), 61-75.
- Bhat, C.R., and S.K. Singh (2000). A comprehensive daily activity-travel generation model system for workers. *Transportation Research Part A*, 34(1), 1-22.
- Bohte, W. and K. Maat (2008). Deriving and validating trip destinations and modes for multi-day GPS based travel surveys: An application in the Netherlands. Presented at the 87<sup>th</sup> Annual Meeting of the Transportation Research Board, Washington, D.C., January.
- Bricka, S., and C.R. Bhat (2006). A comparative analysis of GPS-based and travel survey based data. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1972, 9-20.
- Draijer, G., N. Kalfs and J. Perdok (2000). Global Positioning System as data collection method for travel research. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1719, 147-153.
- Forrest, T., and D. Pearson (2005). Comparison of trip determination methods in household travel surveys enhanced by a Global Positioning System. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1917, 63-71.
- Genius, M., and E. Strazzeria (2008). Applying the copula approach to sample selection modeling. *Applied Economics*, 40(11), 1443-1455.
- Giaimo, G.T., R. Anderson, and L. Wargelin (2009). Large-scale deployment of a GPS-based household travel survey in Cincinnati. Presented at the 12<sup>th</sup> TRB National Transportation Planning Applications Conference, Houston, Texas.
- Herriges, J.A., D.J. Phaneuf, and J.L. Tobias (2008). Estimating demand systems when outcomes are correlated counts. *Journal of Econometrics*, 147(2), 282-298.
- Kochan, B., T. Bellemans, D. Janssens, and G. Wets (2006). Dynamic activity-travel diary data collection, using a GPS-enabled personal digital assistant. Presented at the *Innovations in Travel Modeling Conference*, Austin.
- Krygsman, S., and J.H. Nel (2009). The use of global positioning devices in travel surveys - a developing country application. Proceedings of the 28<sup>th</sup> Southern African Transport Conference, Pretoria, South Africa.



- McKelvey, R.D., and W. Zavoina (1975). A statistical model for the analysis of ordinal level dependent variables. *Journal of Mathematical Sociology*, 4(summer), 103-120.
- Misra, R., and C.R. Bhat (2000). Activity travel patterns of non-workers in the San Francisco Bay area: Exploratory analysis. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1718, 43-51.
- Murakami, E., and D.P. Wagner (1999). Can using Global Positioning System (GPS) improve trip reporting? *Transportation Research Part C*, 7(2-3), 149-165.
- NuStats (2003). Household Travel Survey Final Report of Survey Results. East-West Coordinating Council, St. Louis.
- NuStats (2004). Year 2000 Post-Census Regional Travel Study, GPS Study Final Report. Southern California Association of Governments, Los Angeles.
- Ohmori, N., M. Nakazato and N. Harata (2005). GPS mobile phone-based activity diary survey. Proceedings of the *Eastern Asia Society for Transportation Studies*, 5, 1104-1115.
- Pearson, D.F. (2001). Global Positioning System (GPS) and travel surveys: Results from the 1997 Austin household survey. Presented at the *Eighth Conference on the Application of Transportation Planning Methods*, Corpus Christi, Texas.
- Pierce, B., J. Casas, and G. Giaimo (2003). Estimating trip rate under-reporting: Preliminary results from the Ohio household travel survey. Presented at the *Transportation Research Board 82nd Annual Meeting*, National Research Council, Washington, D.C.
- Quinn, C. (2007). The health-economic applications of copulas: Methods in applied econometric research. Health, Econometrics and Data Group (HEDG) Working Paper 07/22, Department of Economics, University of York.
- Sklar, A. (1973). Random variables, joint distribution functions, and copulas. *Kybernetika* 9, 449-460.
- Steer Davies Gleave and Geostats (2003). The use of GPS to improve travel data, study report. Prepared for the DTLR New Horizons Programme, submitted to the London Department for Transport, London.
- Stopher, P., C. FitzGerald, and M. Xu (2007). Assessing the accuracy of the Sydney household travel survey with GPS. *Transportation*, 34(6), 723-741.
- Trivedi, P.K., and D.M. Zimmer (2007). *Copula Modeling: An Introduction for Practitioners*. Foundations and Trends in Econometrics, 1(1), Now Publishers.

- Wolf, J. (2000). Using GPS data loggers to replace travel diaries in the collection of travel data. Dissertation, Georgia Institute of Technology, School of Civil and Environmental Engineering, Atlanta, Georgia.
- Wolf, J. (2006). Applications of new technologies in travel surveys. In Stopher, P.R., Stecher, C.C. (eds.) *Travel Survey Methods—Standards and Future Directions*, pp. 531–544. Elsevier, Oxford.
- Wolf, J., M. Oliviera and M. Thompson (2003a). Impact of under-reporting on mileage and travel time estimates: Results from Global Positioning System-enhanced household travel survey. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1854, pp. 189-198.
- Wolf, J., M. Loechl, M. Thompson, and C. Arce (2003b). Trip rate analysis in GPS-enhanced personal travel surveys. In Jones, P., and Stopher, P. (eds.) *Transport Survey Quality and Innovation*, Ch. 28, pp. 483-498, Elsevier, Oxford.
- Wolf, J., S. Bricka, T. Ashby, C. Gorugantua (2004). Results from the 2004 Kansas City GPS-enhanced household travel survey. Paper presented at the *TRB Conference on Data for Understanding our Nation's Travel*, November.
- Zmud, J., and J. Wolf (2003). Identifying the correlates of trip misreporting—results from the California statewide household travel survey GPS study. Presented at the *International Conference on Travel Behaviour Research*, Lucerne, Switzerland, August.

## **LIST OF TABLES**

TABLE 1 Characteristics of Respondents with Work-Related Travel

TABLE 2 Characteristics of Respondents with Non-Work Related Travel

TABLE 3 Joint Model of Travel for Work Trips

TABLE 4 Joint Model of Travel for Non-Work Trips

TABLE 5 Percentage Change in Total Expected Difference between Survey and GPS Trips for Work Travel

TABLE 6 Percentage Change in Total Expected Difference between Survey and GPS Trips for Non-Work Travel

## **LIST OF FIGURES**

FIGURE 1 Plot of Dependency Profile between Survey and GPS Trips

**TABLE 1 Characteristics of Respondents with Work-Related Travel**

Variable	Survey Trips > GPS Trips n=65	Survey Trips = GPS Trips n=130	Survey Trips < GPS Trips n=10
<b><i>Individual Demographics</i></b>			
<u>Ethnicity</u>			
Caucasian	32.4%	62.4%	5.3%
African American	33.3%	61.9%	4.8%
Other	0.0%	100.0%	0.0%
<u>Hispanic Origin</u>			
Hispanic	33.3%	66.7%	0.0%
Not Hispanic	31.6%	63.3%	5.1%
<u>Age</u>			
Age ≤ 30 years	30.8%	69.2%	0.0%
Age 31-50 years	29.1%	64.0%	7.0%
Age > 50 years	34.4%	61.3%	4.3%
<u>Gender</u>			
Male	34.0%	59.6%	6.4%
Female	29.7%	66.7%	3.6%
<u>Education Attainment</u>			
Less than high school	27.3%	72.7%	0.0%
High school	40.0%	57.1%	2.9%
College with no degree	32.4%	67.6%	0.0%
Technical school with some degree	29.4%	70.6%	0.0%
Undergraduate college with degree	32.8%	62.1%	5.2%
Graduate degree	26.0%	62.0%	12.0%
<u>Student status</u>			
Full-time student	18.8%	81.3%	0.0%
Part-time student	0.0%	100.0%	0.0%
Not a student	33.7%	60.9%	5.4%
<b><i>Individual Work Characteristics</i></b>			
<u>Employment status</u>			
Full-time employed (30 hrs +)	39.3%	53.6%	7.1%
Part-time employed (< 30 hrs)	25.0%	75.0%	0.0%
Not employed	3.4%	96.6%	0.0%
<u>Work days per week</u>			
Less than 5 days	15.8%	81.6%	2.6%
5 days	39.8%	54.0%	6.2%
More than 5 days	50.0%	43.8%	6.3%
<u>Work Flexibility</u>			
Low	39.0%	58.5%	2.4%
Medium	36.0%	56.0%	8.0%
High	20.3%	78.1%	1.6%
<u>Mode to work</u>			
Car	34.7%	61.4%	4.0%
Other	13.8%	75.9%	10.3%

**TABLE 1 (Cont.) Characteristics of Respondents with Work-Related Travel**

Variable	Survey Trips > GPS Trips n=65	Survey Trips = GPS Trips n=130	Survey Trips < GPS Trips n=10
<b><i>Other Individual Attributes</i></b>			
<u>Interviewed?</u>			
Yes	29.9%	64.6%	5.5%
No	39.0%	58.5%	2.4%
<u>Used Transit for at Least One Work Trip</u>			
Yes	NA	NA	NA
No	31.7%	63.4%	4.9%
<u>Used Walk/Bike for at Least One Work Trip</u>			
Yes	22.2%	50.0%	27.8%
No	32.6%	64.7%	2.7%
<b><i>Household Related Variables</i></b>			
<u>Household Income</u>			
Less than \$40,000	25.5%	72.3%	2.1%
\$40,000-74,999	17.4%	80.4%	2.2%
\$75,000 or more	40.2%	52.7%	7.1%
<u>Residential location</u>			
Located in CBD Region	28.3%	64.2%	7.5%
Not Located in CBD Region	35.4%	62.6%	2.0%
<u>Length lived at the current location</u>			
Less than 2 years	22.7%	72.7%	4.5%
2 years or more	32.8%	62.3%	4.9%
<u>Number of Students in the household</u>			
0	41.0%	52.4%	6.7%
1	15.8%	81.6%	2.6%
2+	25.8%	71.0%	3.2%
<u>Household Size</u>			
1	24.0%	72.0%	4.0%
2	40.8%	52.1%	7.0%
3	32.4%	64.9%	2.7%
4 or more	25.0%	70.8%	4.2%
<u>Presence of Children</u>			
None	35.5%	58.1%	6.5%
1	28.6%	71.4%	0.0%
2 or more	23.9%	71.7%	4.3%
<u>Household Vehicles</u>			
0	0.0%	83.3%	16.7%
1	19.4%	72.2%	8.3%
2	41.8%	57.1%	1.0%
3 or more	26.2%	66.2%	7.7%
<b><i>Day of Week Variables</i></b>			
<u>Day of the week</u>			
Monday – Thursday	31.8%	63.6%	4.5%
Friday	31.0%	62.1%	6.9%
<i>Overall</i>	31.7%	63.4%	4.9%

**TABLE 2 Characteristics of Respondents with Non Work-Related Travel**

Variable	Survey Trips > GPS Trips n=48	Survey Trips = GPS Trips n=109	Survey Trips < GPS Trips n=104
<b><i>Individual Demographics</i></b>			
<u>Ethnicity</u>			
Caucasian	16.9%	45.4%	37.7%
African American	18.4%	31.6%	50.0%
Other	14.3%	28.6%	57.1%
<u>Hispanic Origin</u>			
Hispanic	55.6%	11.1%	33.3%
Not Hispanic	17.1%	42.9%	40.1%
<u>Age</u>			
Age ≤ 30 years	12.2%	46.3%	41.5%
Age 31-50 years	21.5%	34.4%	44.1%
Age > 50 years	18.1%	45.7%	36.2%
<u>Gender</u>			
Male	17.6%	39.5%	42.9%
Female	19.0%	43.7%	37.3%
<u>Education Attainment</u>			
Less than high school	25.0%	37.5%	37.5%
High school	16.7%	48.3%	35.0%
College with no degree	18.8%	35.4%	45.8%
Technical school with some degree	14.3%	38.1%	47.6%
Undergraduate college with degree	22.6%	38.7%	38.7%
Graduate degree	14.8%	46.3%	38.9%
<u>Student status</u>			
Full-time student	20.8%	33.3%	45.8%
Part-time student	33.3%	50.0%	16.7%
Not a student	17.7%	42.4%	39.8%
<b><i>Individual Work Characteristics</i></b>			
<u>Employment status</u>			
Full-time employed (30 hrs +)	22.5%	30.4%	47.1%
Part-time employed (< 30 hrs)	13.5%	51.4%	35.1%
Not employed	14.0%	55.8%	30.2%
<u>Work days per week</u>			
Less than 5 days	18.2%	40.3%	41.6%
5 days	23.9%	33.6%	42.5%
More than 5 days	20.0%	26.7%	53.3%
<u>Work Flexibility</u>			
Low	20.5%	41.0%	38.5%
Medium	25.5%	26.5%	48.0%
High	15.4%	47.7%	36.9%
<u>Mode to work</u>			
Car	21.3%	35.1%	43.7%
Other	22.6%	38.7%	38.7%

**TABLE 2 (Cont. ) Characteristics of Respondents with Non Work-Related Travel**

Variable	Survey Trips > GPS Trips n=48	Survey Trips = GPS Trips n=109	Survey Trips < GPS Trips n=104
<b><i>Other Individual Attributes</i></b>			
<u>Interviewed?</u>			
Yes	21.3%	41.1%	37.7%
No	7.4%	44.4%	48.1%
<u>Used Transit for at Least One Non-Work Trips</u>			
Yes	25.0%	16.7%	58.3%
No	18.1%	43.0%	39.0%
<u>Used Walk/Bike for at Least One Non-Work Trips</u>			
Yes	40.0%	26.7%	33.3%
No	13.9%	44.9%	41.2%
<b><i>Household Related Variables</i></b>			
<u>Household Income</u>			
Less than \$40,000	20.8%	35.1%	44.2%
\$40,000-74,999	10.5%	59.6%	29.8%
\$75,000 or more	20.5%	37.8%	41.7%
<u>Residential location</u>			
Located in CBD Region	22.6%	38.7%	38.7%
Not Located in CBD Region	17.8%	42.2%	40.0%
<u>Length lived at the current location</u>			
Less than 2 years	22.6%	38.7%	38.7%
2 years or more	17.8%	42.2%	40.0%
<u>Number of Students in the household</u>			
0	17.3%	38.8%	43.9%
1	19.6%	49.0%	31.4%
2+	19.7%	42.3%	38.0%
<u>Household Size</u>			
1	11.8%	41.2%	47.1%
2	18.7%	40.7%	40.7%
3	16.7%	43.8%	39.6%
4 or more	21.6%	42.0%	36.4%
<u>Presence of Children</u>			
None	16.8%	42.5%	40.7%
1	20.5%	38.6%	40.9%
2 or more	22.0%	42.0%	36.0%
<u>Household Vehicles</u>			
0	27.3%	27.3%	45.5%
1	20.8%	37.5%	41.7%
2	16.3%	40.7%	43.1%
3 or more	19.0%	48.1%	32.9%
<b><i>Day of Week Variables</i></b>			
<u>Day of the week</u>			
Monday – Thursday	19.1%	43.6%	37.3%
Friday	13.9%	30.6%	55.6%
<i>Overall</i>	18.4%	41.8%	39.8%

TABLE 3 Joint Model of Travel for Work Trips

Variable	Survey-reported trips		GPS-recorded trips	
	Parameter	t-stat	Parameter	t-stat
<b>Threshold parameters in ordered response</b>				
Threshold 1	0.8143	3.23	0.9353	3.03
Threshold 2	1.4764	5.78	1.5402	4.74
Threshold 3	2.4004	8.96	2.3732	7.02
Threshold 4	2.5614	9.95	2.6724	7.80
<b>Individual Demographics</b>				
<i>Race (non-Caucasian is base)</i>				
Caucasian	--	--	0.3898	1.49
<i>Age and Gender (&gt; 50 years of age is base)</i>				
Less than or equal to 30 years*Female	--	--	0.4558	1.99
Between 30 and 50 years	--	--	0.2775	1.63
Less than or equal to 30 years *Male	0.4395	2.02	0.4558	1.99
<i>Education Attainment (undergraduate college with degree is base)</i>				
Less than high school	0.2305	1.45	0.4743	2.37
High school	0.2305	1.45	0.4743	2.37
College with no degree	0.2305	1.45	0.4743	2.37
Technical school with some degree	0.2305	1.45	0.4743	2.37
Undergraduate college with degree	--	--	--	--
Graduate degree	--	--	0.5764	2.73
<i>Student (non-student is base)</i>	-0.6121	-1.62	-0.7232	-1.68
<b>Individual work related characteristics</b>				
<i>Employment (full-time employment is base)</i>				
Full time	--	--	--	--
Part time	--	--	-0.9310	-3.84
Voluntary employment	-1.3493	-2.63	-0.9310	-3.84
<i>Work days per week (less than 5 days is base)</i>				
Less than 5 days	--	--	--	--
5 days	0.6138	3.18	--	--
More than 5 days	0.7416	2.10	-0.5466	-1.29
<i>Mode to work (non-car mode to work is base)</i>				
Car	0.5245	2.49	--	--
<b>Other individual attributes (no walk/bike use is base)</b>				
Uses walk/bike for at-least one work trip on travel day	--	--	0.7458	2.83
<b>Household Related variables</b>				
<i>Household income (medium income \$40K-\$75K is base)</i>				
Less than \$40K	0.3110	1.22	--	--
Between \$40K and \$75K	--	--	--	--
Greater than \$75K	0.3695	2.52	--	--
<i>Number of students</i>	-0.0972	-1.70	--	--
<b>Day of week variables (non-Friday is the base).</b>				
<i>Friday</i>	-0.3165	-1.62	-0.4819	-1.88



TABLE 4 Joint Model of Travel for Non-Work Trips

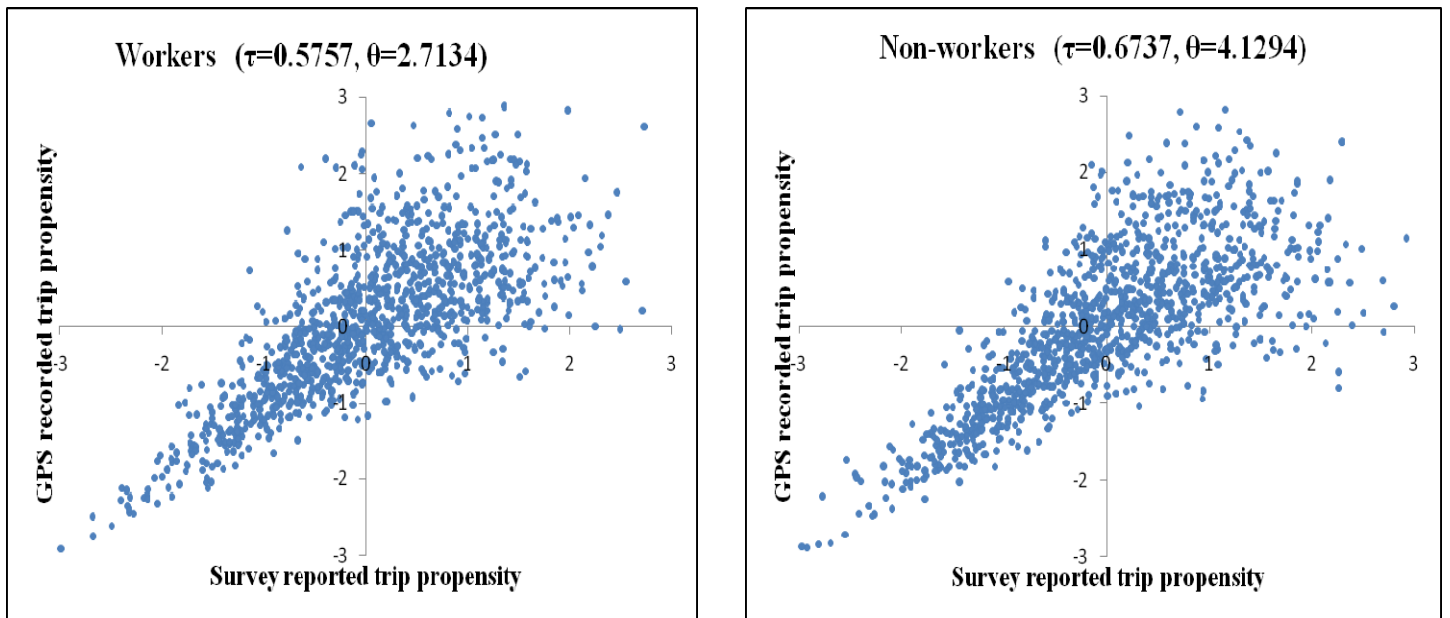
Variable	Survey-reported trips		GPS-recorded trips	
	Parameter	t-stat	Parameter	t-stat
<b>Threshold parameters in ordered response</b>				
Threshold 1	-0.1794	-0.75	-0.5903	-2.52
Threshold 2	0.2194	0.92	-0.1385	-0.60
Threshold 3	0.3987	1.67	0.0706	0.31
Threshold 4	0.7830	3.24	0.3598	1.57
Threshold 5	1.0525	4.33	0.6726	2.91
Threshold 6	1.3644	5.54	0.9930	4.25
Threshold 7	1.6459	6.57	1.2151	5.16
Threshold 8	1.9945	7.71	1.4711	6.13
Threshold 9	2.2769	8.56	1.6460	6.75
Threshold 10	2.5234	9.22	1.8700	7.53
Threshold 11	2.6522	9.51	2.0510	8.09
Threshold 12	2.9752	10.07	2.2762	8.70
<b>Individual Demographics</b>				
<i>Race (Non-Caucasian in small household is base)</i>				
Caucasian	--	--	0.1389	1.52
Non-Caucasian in Large Household	-0.3385	-1.66	--	--
<i>Age (age &lt;=50 years is base)</i>				
Less than 30 years	--	--	--	--
Between 30 and 50 years	--	--	--	--
More than 50 years	-0.1621	-2.23	--	--
<i>Education Attainment (less than undergraduate degree is base)</i>				
Less than high school	--	--	--	--
High school	--	--	--	--
College with no degree	--	--	--	--
Technical school with some degree	--	--	--	--
Undergraduate college with degree	0.2915	2.27	0.1836	1.45
Graduate degree	0.2915	2.27	0.1836	1.45
<i>Head of household (non-head is base)</i>	-0.3100	-2.35	-0.4533	-3.60
<b>Individual Work Characteristics</b>				
<i>Work Status (full or part-time employment is base)</i>				
Full time	--	--	--	--
Part time	--	--	--	--
Voluntary employment	0.3161	2.67	--	--
Retired	0.6427	2.83	0.4327	2.00
Not employed	0.1547	1.30	--	--
<i>Work days per week (&lt;5 days work/week is base)</i>				
Less than 5 days	--	--	--	--
5 days	-0.4442	-3.19	-0.3974	-3.06
More than 5 days	-0.6960	-2.56	-0.6040	-2.39
<b>Other individual attributes (self-explanatory base categories)</b>				
<i>Interviewed</i>	0.2727	3.13	--	--
<i>Uses transit for at-least one non-work trip on travel day</i>	0.6478	2.10	1.0736	3.49
<i>Uses walk/bike for at-least one non-work trip on travel day</i>	0.5105	5.12	--	--
<b>Household related variables</b>				
<i>Household income (&gt;=40K is base)</i>				
Less than \$40K	-0.3598	-2.17	-0.2813	-1.75
<i>Located in CBD region (non-CBD location is base)</i>	0.3207	2.55	0.2930	2.38
<i>Time lived at the current location (&gt;=2 years is base)</i>				
Less than two years	1.0414	5.41	1.1634	6.03
<i>Number of vehicles per license holder</i>	0.3924	2.70	0.3148	2.13
<b>Day of week variables (non-Friday is base)</b>				
<i>Friday</i>	0.2687	1.62	0.4447	2.55

**TABLE 5 Percentage Change in Total Expected Difference between Survey and GPS Trips for Work Travel**

<b>Variable</b>	<b>Percentage Change</b>
<b>Individual Demographics</b>	
<i>Race (non-Caucasian is base)</i>	
Caucasian	-29.44
<i>Age and Gender (&gt;50 years is base)</i>	
Less than or equal to 30 years*Female	-43.25
Less than or equal to 30 years *Male	8.30
Between 30 and 50 years	-23.67
<i>Education Attainment (undergraduate degree is base)</i>	
Less than undergraduate degree	-14.55
Graduate degree	-52.86
<i>Student (non-student is base)</i>	-12.05
<b>Individual work related characteristics</b>	
<i>Employment (full-time employment is base)</i>	
Part time	60.95
Voluntary employment	-50.00
<i>Work days per week (&lt;5 days per week is base)</i>	
5 days	65.95
More than 5 days	127.44
<i>Mode to work (non-car mode is base)</i>	
Car	53.53
<b>Other individual attributes</b>	
Uses walk/bike for at-least one work trip on travel day	-76.11
<b>Household Related variables</b>	
<i>Household income (medium income is base)</i>	
Less than \$40K	35.27
Greater than \$75K	40.13
Number of students	-10.44
<b>Day of week variables (non-Friday is base)</b>	
Friday	2.24

**TABLE 6 Percentage Change in Total Expected Difference between Survey and GPS Trips for Non-Work Travel**

<b>Variable</b>	<b>Percentage Change</b>
<b>Individual Demographics</b>	
<i>Race (Non-Caucasian in small household is base)</i>	
Caucasian	-8.79
Non-Caucasian in Large Household	-19.12
<i>Age (age&lt;=50 years is base)</i>	
More than 50 years	-9.45
<i>Education Attainment (less than undergraduate degree is base)</i>	
Undergraduate college or above	5.32
Head of the household (non-head is base)	10.64
<b>Individual Work Characteristics</b>	
<i>Work Status (full or part-time employment is base)</i>	
Voluntary employment	18.84
Retired	10.54
Not employed	9.12
<i>Work days per week (&lt;5 days per week is base)</i>	
5 days	-0.59
More than 5 days	-1.43
<b>Other individual attributes</b>	
Interviewed	15.73
Used transit for at-least one of the non-work trips on travel day	-28.43
Used walk/bike for at-least one of the non-work trips on travel day	20.19
<b>Household related variables</b>	
<i>Household income (&gt;=40K is base)</i>	
Less than \$40K	-2.88
Located in CBD region (non-CBD location is base)	0.01
<i>Length of stay at the current location (&gt;=2 years is base)</i>	
Less than two years	-12.88
Number of vehicles in household	1.64
Number of licensed individuals in household	-0.99
<b>Day of week variables (non-Friday is base)</b>	
Friday	-13.07



**FIGURE 1** Plot of Dependency Profile between Survey and GPS Trips