WORKING PAPER

ITLS-WP-15-04

The challenges and opportunities of in-depth analysis of multi-day and multi-year data

By
Claudine Moutou; Thomas Longden;
Peter Stopher and Wen Liu

February 2015

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in
Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER:	Working Paper ITLS-WP-15-04
TITLE:	**The challenges and opportunities of in-depth analysis of multi-day and multi-year data**

ABSTRACT: The paper uses a unique multi-day multi-wave panel dataset of households and their travel to conduct new in-depth analysis on the influence of life change events and travel behaviour, specifically in relation to the travel time stability of individuals that participated in five or more waves. The popularity of mobile devices offer greater low-cost opportunities for collecting detailed travel data records may increase opportunities to analyse how life change events and travel behaviour. This paper discusses issues for designing pragmatic research designs that are robust to be expanded to become longitudinal, or combined with other datasets.

KEY WORDS: ***life change events, socio-demographic data, travel-time stability, GPS measurement, panel data, multi-day data, longitudinal data***

AUTHORS: **Moutou; Longden; Stopher; Liu**

CONTACT: INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (C13)

The Australian Key Centre in Transport and Logistics Management

The University of Sydney NSW 2006 Australia

Telephone: +612 9114 1824

E-mail: business.itlsinfo@sydney.edu.au

Internet: http://sydney.edu.au/business/itls

DATE: February 2015

1. Introduction

For the past 60 or more years, the data available for analysis in transport studies has been largely limited to one-day data from a cross-sectional survey. There have been a few instances of multi-day diaries, including those undertaken in Uppsala, Sweden in 1971 (Hanson and Huff, 1986; Hanson and Huff, 1988), Karlsruhe and Halle, Germany in 2000 (Habib, Miller et al., 2008), and Reading, England in 1973 (Shapcott, 1978; Pas, 1988), among others, with each of these surveys being a one-off survey. There have also been a very limited number of panel surveys, such as the Puget Sound Panel Survey (Murakami and Watterson, 1989), the Dutch National Mobility Panel (van Wissen and Meurs, 1989), and the German Mobility Panel (Zumkeller, Madre et al., 2006), among others. However, aside from the seven day Dutch panel these panel surveys have been restricted to one or two days of data. In contrast, over the past eight years, the Institute of Transport and Logistics Studies (ITLS) of the University of Sydney has had the opportunity to collect data from panels, using Geographic Positioning Systems (GPS) as the primary mechanism for collecting travel data, enabling anything from a week to four weeks of travel data to be collected (Stopher, Zhang et al., 2009; Stopher, Moutou et al., 2013).

GPS devices were used by the ITLS team led by Stopher in 2005 as part of an evaluation study of Voluntary Travel Behaviour Change Programs (VTBCP). The GPS devices recorded the travel time for each individual for each trip for multiple days. The success of the initial pilots led to further work to evaluate the longer-term influence of VTBCP, which resulted in a six-year panel survey. While previous analyses of the multi-day and multi-year data have focused on the longer-term influence of VTBCP on amount of kilometres travelled by car, other lines of enquiry such as exploring the stability of travel time budgets and changes in the sociodemographic details of individuals and households are now possible. This is of interest from a number of perspectives: theoretical, policy decision-making, and changing expectations about research.

The notion of travel time being relatively stable over time has existed in the literature for more than 40 years. Proponents of this concept include Szalai (1972), Zahavi (1973, 1974), Zahavi and Ryan (1980), Zahavi and Talvitie (1980), Schaefer and Victor (1997), and Schaefer (2000) to name a few. Although there are disputes within the literature on the validity of the concept in applied contexts the idea of travel time constancy has had some resonance in relation to urban planning, but has been ignored in favour of economic theories of utility which place greater import on travel-time savings. It is important to note that the concept of travel time stability is not intended to mean travel-time constancy – although sometimes it is reported as such. Zahavi and Ryan (1980), for example, were clear that travel time was a function of several variables, but that where there was variation it was relatively stable. A comprehensive review of travel time budget studies, including those that have misunderstood and/or challenged the concept, can be found in Ahmed and Stopher (2014).

As identified by Ahmed and Stopher (2014) the datasets being used to support as well as dispute the concept of travel-time stability have been limited by the reliance on self-reporting (for example travel diaries, interviews) or by the number of days and years of comparable data. This is problematic as recall of travel times has been found to be an inaccurate science with many rounding up their travel time and therefore inflating errors within the dataset (Wolf 2006; Forrest and Pearson 2005; Stopher and Greaves 2009). Concern about this was a motivator for some transport researchers to experiment with the use of GPS and Global System for Mobile Communications (GSM) as a more reliable method of collecting travel time and distance travelled (Kracht 2006). Stopher and his team are using the multi-day and multi-year dataset to reinvestigate the concept of travel time stability. This paper reports on one particular line of enquiry focused on the stability of travel time using a concurrent dataset of self-reported sociodemographic changes at the individual and household level.

Analysing how travel time varies amongst types of individuals is not unusual and has been important for rationalising the willingness to pay for faster travel amongst certain strata of society. Contextual and cultural factors aside, studies regularly differentiate between travel-time benefits based on income, working status, and car ownership. However, as with other studies on travel-time stability, where there has been an analysis of the sociodemographic differences these have been limited by the quality of data about time travelled and the limited number of observations. This paper uses the annual updates to sociodemographic detail to analyse if the occurrence of life events have a statistically significant influence on an individual's travel time.

The paper is structured in the following way. Section 2 presents an overview of why changes in life events are of interest and what changes in average travel time one might expect based on the literature. Section 3 presents the dataset including how it was constructed and the categorisation of life stages. The empirical analysis in Section 4 is divided into two parts – firstly a univariate analysis to examine if a life event influences individual travel time, followed by a multivariate analysis to investigate the impact of each type of event over time. The discussion in Section 6 focuses on the challenges in the analysis of the dataset and considerations for further analysis. The paper concludes with suggestions for researchers and policy makers that would help make this type of analysis more common.

2. Incorporating life changes - a constancy of life

Life involves changes and studies into travel behaviour have long acknowledged this. Indeed the policy intent of travel behaviour research has been to find ways to understand the conditions in which individuals will be more open to changing their habitual travel behaviour to align with the desired policy goals. Transport research has, for example, focused on different stages in the life course, for example when a child gains competency to cross roads, or travel independently to school, or gains a licence to drive (for recent examples see Mitra 2013, Le Vine and Polack 2014). Life course and life span studies are established fields that follow samples of the population to study issues of health, parenting, and education using a mix of quantitative and qualitative research techniques. In transport research it has largely been associated with qualitative surveys such as mobility biographies and interviews, in part as long running travel survey panels have not been available. Mapping an individual's life course (and that of their household) in a transport context may follow a similar pattern of categorisations of life-stage events. For example key events in an individual's life course, such as having children, changing jobs, changing employment, can represent key moments to disrupt and habitualise new travel patterns, reinforce preferences for mode of travel, and invest in car ownership (Lanzendorf 2010, Scheiner and Holz-Rau 2013, Delbosc and Currie 2013). Increasingly the interaction of other factors, such as gender, aging, and changes in economic power as people move in and out of the workforce are capturing interest as concerns about transport social exclusion increase (Lucas 2012).

Collection of sociodemographic data is a secondary focus of most transport surveys. Common variables include age, sex, employment status and car ownership, which are routinely used to describe the sample. More detailed or intrusive sociodemographic detail such as income, educational level, and work status can be used to identify how sociodemographics may moderate travel behaviour, or indeed value of travel-time. In respect to travel-time research, Ahmed and Stopher (2014) identify that there are inconsistent conclusions on the influence of sociodemographics. This limits the ability to understand how individuals within the household may have an interactive influence on individual travel-time, for example the influence of changes to household size (Ahmed and Stopher 2014). The availability of a sociodemographic dataset collected alongside a multi-year panel GPS travel survey presents a unique opportunity to examine and track how circumstances at the individual and household level may have influenced changes in travel time.

The research in this paper reports on an initial investigation into the mediating influence of sociodemographic variables on changes in travel time over a multi-year period. Specifically investigating the research question: do changes in life events significantly influence an individual's travel time. It does this through a new analysis of previously collected multi-year multi-day travel surveys collected by Stopher and his team, to investigate disaggregate level change at the person level, differentiating between travel done on the weekday or weekend, or mode of travel to assess what rhythms of travel are ongoing and what are subject to fluctuations (Zumkelle, Madre et al 2006 p.364).

3. The Dataset

From 2005 to 2012, the Institute of Transport and Logistics Studies (ITLS) of the University of Sydney has collected daily travel data using GPS from households in four Australian cities (Stopher, Zhang et al., 2009; Stopher, Moutou et al., 2013). A multi-day, multi-year panel survey was conducted annually during the September-November period in Adelaide, Brisbane, Canberra and Melbourne as part of a long-term monitoring evaluations of voluntary travel behaviour change programs (VTBCP) from 2007 to 2012 (Stopher, Moutou et al., 2013). Individuals in the household aged over 14, carried the portable GPS devices for 15 consecutive days. A 7-day survey with households in Adelaide to assess short-term changes due to VTBC was conducted in 2005. In 2006, a 28-day pilot survey for the long-term evaluation was also done in Adelaide in two waves, six months apart. These three earlier data collection efforts helped to ascertain an optimal period in which respondents should carry the devices to avoid attrition, and issues with data quality was 15-days (Stopher, Clifford et al 2009). A rotating panel was used to address issues of attrition, with new households randomly selected from the target population (Stopher, Moutou et al 2013). Participants did not receive a financial incentive to participate in the GPS travel survey, but were sent newsletters over the course of the study as a strategy for maintaining levels of engagement and avoiding household attrition. In total 620 households, comprising 1778 individuals participated in the GPS travel study, with 29 households participating in 5 or more waves.

This paper focuses on these 29 households whose involvement in the GPS panel survey involved yearly updates to the sociodemographic details of the household and the individuals within it. Amongst the 29 households, ten households supplied travel data for 5 waves, six households supplied 6 waves of data, six households provided 7 waves of data, three households provided 8 waves of data and four households supplied 9 waves of data. The analysis in this paper is of a dataset constructed of two parts. One part comprises travel time data generated from the GPS travel data records of each individual's trips, each survey day, in each survey wave. The other part comprises paper forms that were used to collect detail in each wave on the individuals in the household, the cars owned, and any issues arising during the survey period (such as non-travel days, or forgetting to take the GPS device). It is these paper forms that have provided a rich amount of detail that enable the travel time data to be examined in new ways.

The paper forms were for many years assumed to be of little value. All the detail (for example, drivers' licence status, physical mobility, education status and work status) had been routinely entered into a database and already used to describe the data by sociodemographic characteristics (Zhang, Stopher et al., 2013; Stopher, Moutou et al., 2013). As new funding was secured to extend the original evaluation of VTBCP to observe longer-term effects on household travel behaviour, the Access databases were updated and improved upon. The need to review the paper forms arose from the practical difficulties of new staff setting out to combine all the sociodemographic data from all waves that had been stored in the different Access databases. Going through the paper records, however, provided an opportunity to construct a new dataset of change events for the 29 multi-wave households, and to understand more about the type of household by examining the inter-relationships amongst the individuals.

The sociodemographic detail helps to describe the data not just as a function of the number of people, but also the types of familial relationships. The 29 households included various types of relationships. Thirteen households could, for example, be described as nuclear families, with two parents and their children. Other family units included two single-parent families, one with a teenage dependent, the other with an adult dependent. There were eight couples (married or defacto), and two households with a different composition of familial relationships. While this level of description can be useful, it does not account for the changes in the household composition that change the dynamic of travel decision-making amongst members of the household, or the changing nature of travel needs as individuals in the household age and experience changes in their mobility, travel needs, and purposes for travel. The households with one person fluctuated, with only four households remaining single throughout the waves. Households with elderly dependent parents, or housemates were also present albeit in small number. Categorisation of the households by life stages provides an alternative perspective of the data that can account for the changing nature of relationships within the household, and engagement in activities outside of the household.

The types of sociodemographic changes that can be analysed are constrained by the categories that were presented on the paper forms, and the level of detail. Information about the nature of relationships between Person 1 and all other persons in the household (self, spouse/partner, father/mother, brother/sister, other relative, non-relative, live-in domestic help) were unlikely to vary, but were helpful when combined with other indicators such as age and physical limitations in defining individual level life events such as changes in the responsibility of dependents. The data collected about work and educational activities of the individuals in the household lent themselves to mapping life stages from early childhood to school to work to retirement, with supplementary data about the nature of the work (paid/unpaid, full-time/part-time), and name and location of education/work helping to add further context and distinction.

Detail collected about changes in physical limitations to travel, and information collected about vehicles and licenced drivers in the household helped to map a sequence of life changes based on mobility. The forms collected information about any physical limitations that would have an effect on types of transport each individual could use. The changes in mobility enabled mapping of different life stages from dependent child to adult to reduced mobility independence arising from aging, as well as life events that could be viewed as disruptions or enablers to mobility. The forms for example asked households “to specify for each individual what sort of driver's licence” they have. The four options (no licence, learner’s licence, provisional licence, full licence) thereby enabled the tracking of incremental change as an individual moves through the sequence of categories. Table 1 and Table 2 describe which data were used to construct 17 life events and eight household events.

Table 1: Definition of Individual Life Events (ILE) constructed from sociodemographic data

Individual Life Event (ILE)		**Data used**
Daily activity life events (Work)	Entering the workforce	Changes to worker and non-worker descriptions of individual (inc. homemaker, retired, student)
	Leaving the workforce	
	Moving work location	Change in place of work (address).
	Changing job role	Change in job title, occupation.
	Increasing work hours	Changes between full-time and part-time.
	Reducing work hours	
	Increasing voluntary hours	Change in voluntary activity
	Reducing voluntary hours	
Daily activity life events (Education)	Reducing to part-time education	Changes in education level, age, place of education.
	Starting new education	
	Leaving education	
Mobility life events	Loss of physical mobility (health)	Changes in the number of physical limitations indicated.
	Gain of physical mobility (health)	
	Gain of mobility (driver's licence)	Change of drivers' licence level (no licence, learner's licence, provisional licence, full licence)
	Loss of mobility (driver's licence)	
Responsibility life events	Gained responsibility of dependents (parenthood or carer)	Changes in number of individuals in household, age, and their relationship to each other.
	Reduced responsibility of dependents (parenthood or carer)	

Table 2: Definition of Household Life Events (HLE) constructed from sociodemographic data

Household Life Event (HLE)		**Data used**
Composition of household events	Reduction in household size	Changes in number of individuals in household, identified by name and age.
	Increase in household size	
	People changed (new)	
Mobility-type household events	Reduction in car ownership	Changes in total count of cars owned by household members
	Increase in car ownership	
	Car changed (new)	
	Reduction in bike ownership	Changes in count of bikes
	Increase in bike ownership	

Figure 1 reviews the method used to combine individual daily travel data with the occurrence of a life event at both the individual and household levels. To aggregate daily data and create a dependent variable that can be compared to individual life events that occur in a given wave the focus of the analysis concentrates on the average of daily travel within that wave for each person. The average time travelled per day is then matched to the occurrence of different life events within the same wave. In addition, a range of household life events is associated to each person in that household across each of the waves. This means that if individuals i1 and i2 had a household event occur for any person in their household then both of these individuals have that observation associated with them for that period. Aggregating the data in this manner is necessary to compare the impact of both individual and household life events on the amount of travel that an individual tends to undertake in a given day. As prior studies have found evidence that the time travelled per day tends to be similar over time and across individuals due to a travel time budget, this analysis reviews the travel time data using an average estimate for the wave with the caveat that variance is lost due to this aggregation and is not reviewed within this analysis. Future work will focus on whether redefining the dataset to focus on each daily observation makes a notable difference to the results of this paper.

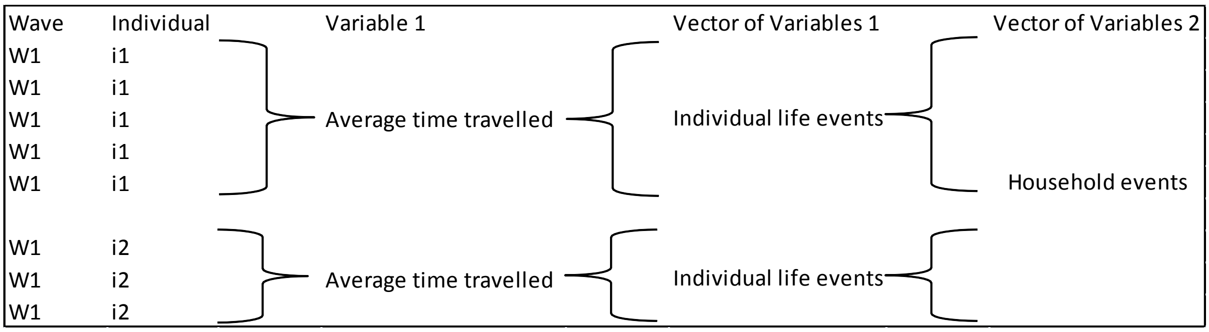

Figure 1: Combining daily data with individual and household events

3.1 Descriptive data about the dataset

This section focuses on the frequency of persons and households in the data, as well as the distribution of the average minutes travelled per day that is present within the dataset specified in the previous discussion.

Table 3 shows the frequency of household sizes and households for the waves of data collection. Changes in household composition are most apparent in waves 6, 7 and 8 when all 29 households participated. The dataset that has been compiled for the analysis has nine waves of data for the period between 2005 and 2012. Note that the second wave was collected approximately six months after the first wave and that the majority of the waves tend to include the months of October and November. Wave 1, 3 and 4 do not exactly match the October to November period, but are quite close to the periods of collection in wave 5, 6, 7, 8 and 9. Within the multivariate analysis the waves are coded as a time trend, which means that the data is coded into one indicator with the values specified as 1 for August to December 2005, 2 for March to April 2006, 2 for October to December 2006, 3 for October to December 2007, and so on. Coding wave 2 as 2 allows us to have a consistent annual time dimension and, accordingly, an annual time trend is implemented in the multivariate analysis in Section 4. Note that sensitivity testing with alternative arrangements for the time trend (with wave 2 coded as 1.5) produces minor differences in the parameter estimates.

Table 3: Frequencies across waves

		Wave 1	**Wave 2**	**Wave 3**	**Wave 4**	**Wave 5**	**Wave 6**	**Wave 7**	**Wave 8**	**Wave 9**
	Value	**2005**	**2006**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
		20 Aug. – 1 Dec.	8 Mar. – 10 Apr.	6 Oct. – 12 Dec.	12 Oct. – 14 Dec.	14 Oct. – 27 Nov.	9 Oct. – 11 Nov.	9 Oct. – 11 Nov.	6 Oct. – 8 Nov.	10 Oct. – 9 Nov.
Household size	**1**	1	1	2	3	4	5	6	6	5
	2	3	4	6	8	10	11	14	11	11
	3	0	0	0	1	2	3	2	5	3
	4	2	2	5	6	9	7	8	5	5
	5+	0	0	1	2	3	3	1	2	1
Number of households	**Total**	6	7	14	20	28	29	31	29	25

Note: these frequencies are specified at the individual level for the number of persons and the household level for the number of households. Dates of collection are inserted in the table above.

Figure 2 reviews the distribution for all days of the week, while Figure 3 and Figure 4 review the distribution for weekdays and weekend days, respectively. For the total amount of days per week, 57.4 per cent of the sample travelled between approximately 40 and 80 minutes on average per day. For weekdays, this changes to be 52.81 per cent of the sample travelling between approximately 40 and 80 minutes on average per day. In contrast the distribution of average time travelled on the weekend tended to be lower with a distribution skewed to the left and 52.90 per cent of the sample travelling between approximately 10 and 60 minutes on average per day.

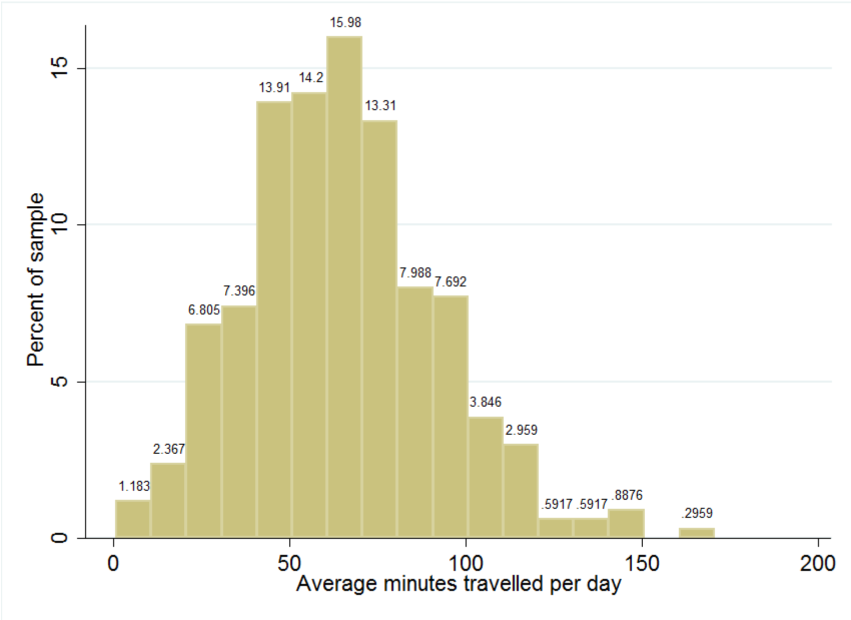

Figure 2: Distribution of average time travelled per day – All days – 372 observations

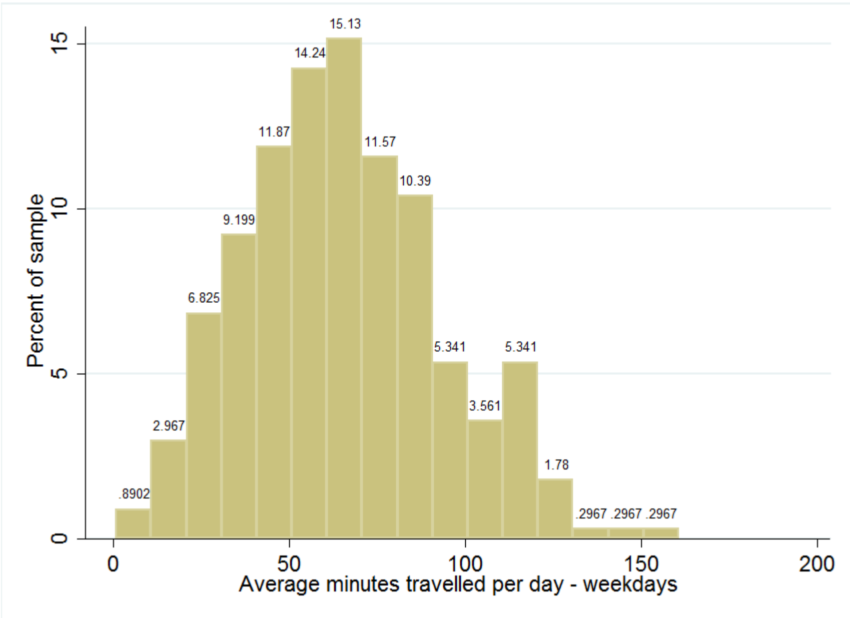

Figure 3: Distribution of average time travelled per day – Weekdays – 371 observations

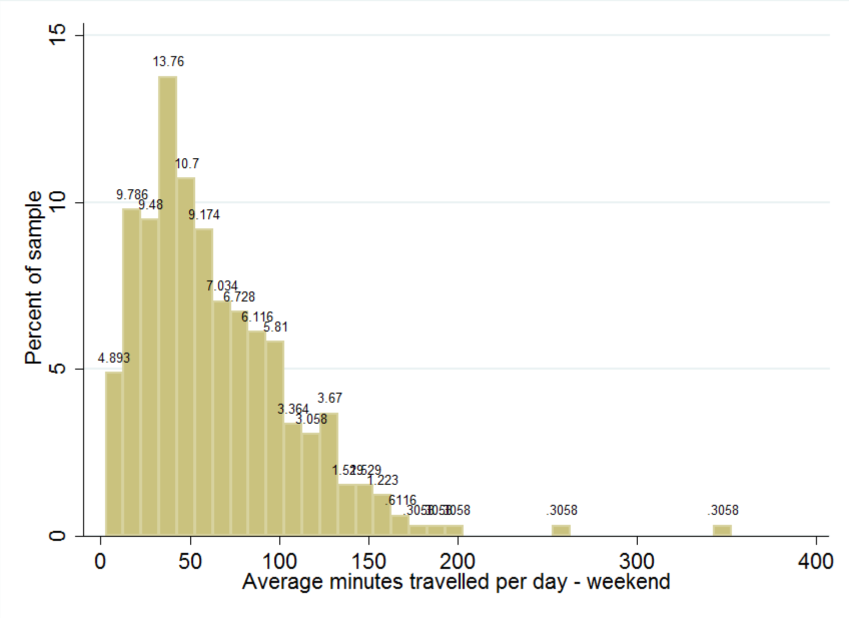

Figure 4: Distribution of average time travelled per day – Weekends – 360 observations

3.2 The categorisation of life stages

Table 4 lists the life events that are reviewed in this paper, listing the mean average travel time at the time of the event. Our analysis is based on the actual wave in which a life event occurs and does not track whether this has an impact outside of this period. The number of events across the sample differs, for example replacing a car is observed at the individual level 93 times, an increase in bike ownership being observed at the individual level 34 times and a reduction in household size is observed at the individual level 25 times. It should be noted that due to the creation of the dataset in the manner described in Figure 1, the number of occurrences at the individual level is inflated by the number of people in the household where the event occurred for every individual who is in the dataset. This is an artefact of combining the individual and household data in the manner prescribed in Figure 1. The specification of the dataset in this manner is driven by the focus of the analysis in Section 4 that aims to quantify the relation between an individual's average travel time and individual/household life events that occurred in the same period. Irrespective of this, Table 4 contains the adjusted and unadjusted frequencies. The adjusted frequency for the number of households where the events took place is provided in brackets. In the case of the household level event (HLE) where a car was replaced the adjusted frequency changes the 93 individual instances to

42 household instances – with some of the 29 households evidently changing their car more than once. Note that the multivariate analysis in Section 5 reports analysis that adjusts for clustering based on the household.

Table 4 : Mean average time travelled per day by life event and life event frequency

	Indicator	Mean Average Time Travelled	Frequency at the individual level		Indicator	Mean Average Time Travelled	Frequency at the individual level
Individual-level life events (ILE)	Entering the workforce	56.95	9	Household-level life events (HLE)	Reduction in household size	70.68	26 (8)
	Leaving the workforce	70.85	8		Increase in household size	81.76	10 (4)
	Moving work location	66.37	20		People changed (new)	76.68	14 (6)
	Changing job role	59.25	20		Reduction in car ownership	67.90	23 (11)
	Increasing work hours	50.21	13		Increase in car ownership	76.24	23 (9)
	Reducing work hours	63.06	8		Car changed (new)	68.71	90 (23)
	Increasing voluntary hours	44.47	5		Reduction in bike ownership	65.90	28 (12)
	Reducing voluntary hours	104.27	1		Increase in bike ownership	67.24	33 (11)
	Reducing to part-time education	68.07	1				
	Starting new education	61.34	8				
	Leaving education	56.86	5				
	Loss of physical mobility (health)	39.22	8				
	Gain of physical mobility (health)	75.15	2				
	Gain of mobility (driver's licence)	54.07	14				
	Loss of mobility (driver's licence)	27.44	1				
	Gained responsibility of dependents (parenthood or carer)	59.63	4				
	Reduced responsibility of dependents (parenthood or carer)	98.5	2				

Note: these frequencies are specified at the individual level as this is the form of the dataset. For the household-level life events, the frequency at the household level is presented in brackets.

4. Empirical Analysis

This section presents the empirical analysis of the dataset described in Section 3. The univariate analysis presented in Section 0 reviews the immediate impact of life event indicators on the average time travelled per day across the waves. Accordingly, the analysis focused upon whether people in the sample had a life event occur or not. Section 5 extends the analysis with a focus on multivariate analysis and does so by reviewing the impact of each life event on the average time travelled per day using regression analysis.

4.1 Univariate Analysis

Table 5 reviews the sample mean of the average time travelled for those who have and have not had an individual life event. The p values show that the sample means for all days of the week and weekend observations are statistically different from each other at a 1 per cent confidence interval. For weekday observations the sample means are different at a 5 per cent confidence interval. Table 6 reviews the sample means of average time travelled per day, which differ based on whether a household event occurred or not. In this case, the statistical tests confirm that only the weekend observations show a statistically significant difference. Table 7 reviews whether there is a statistically significant difference in the sample mean of average time travelled between the weekday and weekend observations. At a confidence interval of 5 per cent it is found that the difference between means is significant for those who had no individual life event and at a 5 per cent confidence interval the difference between means is significant for both those who have had and have not had a household life event. Note that the small sample size of the No HLE category should prompt caution when interpreting the relevant results contained in Table 6 and Table 7.

Table5 : Comparison of sample mean average time travelled per day by individual-level life event (ILE) and type of day

	All		Weekday		Weekend	
	ILE	No ILE	ILE	No ILE	ILE	No ILE
Mean	61.86	70.61	63.10	69.35	59.63	73.79
N	225	113	224	113	216	111
t-test for Equality of Means						
	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed
p value	0.005	0.005	0.053	0.050	0.005	0.010

Note: null hypothesis is that means are equal and the alternative hypothesis is that means are unequal (two sided t-test).

Table6 : Comparison of sample mean average time travelled per day by household-level life event (HLE) and type of day

	All		Weekday		Weekend	
	HLE	No HLE	HLE	No HLE	HLE	No HLE
Mean	64.79	64.72	64.71	72.82	65.80	43.45
N	318	20	317	20	307	20
t-test for Equality of Means						
	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed	Equal variances assumed	Equal variances not assumed
p value	0.990	0.991	0.209	0.274	0.025	0.002

Note: null hypothesis is that means are equal and the alternative hypothesis is that means are unequal (two sided t-test).

Table 7: Comparison of sample mean average time travelled per day by type of day and individual-level life event (ILE) or household-level life event (HLE)

	ILE		No ILE		HLE		No HLE	
	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend	Weekday	Weekend
Mean	55.65	59.63	62.69	73.79	57.46	65.80	66.20	43.45
N	254	216	125	111	357	307	22	20
p value	0.2346		0.0421		0.0056		0.0275	

Note: null hypothesis is that the difference between the means is zero and the alternative hypothesis is that the difference is not equal to zero (two sided t-test).

5. Multivariate Analysis

This section reports multivariate analysis on an unbalanced panel using fixed effects (FE), random effects (RE) and generalised least squares (GLS) regression. Regression analysis is used so as to analyse the impact of each type of life event upon the average time travelled per day across the waves of data. The regressions that were utilised are specified in equations 1, 2 and 3 with ATT_{it} representing the average time travelled per day, α_i representing the unobserved effect, δ_t as the annual time trend, and v_{it} as a composite error term specified as $\alpha_i + u_{it}$.

$$ATT_{it} = \alpha_i + \delta_t + \beta_1 X_{it} + \cdots + \beta_k X_{kt} + u_{it} \qquad (1)$$

$$ATT_{it} = \beta_0 + \delta_t + \beta_1 X_{it} + \cdots + \beta_k X_{kt} + v_{it} \qquad (2)$$

$$ATT_{it} = \beta_0 + \delta_t + \beta_1 X_{it} + \cdots + \beta_k X_{kt} + u_{it} \qquad (3)$$

As the dataset is specified as a panel with a time (t) and individual (i) dimension, there is a concern that heteroskedasticity is present. Indeed, heteroskedasticity is found when the Modified Wald test for groupwise heteroskedasticity in a FE regression model was applied. As a result, robust standard errors are applied in the initial FE and RE regressions. As the data are a combination of individual data and household data, the analysis also reviewed whether the standard errors in a FE and RE regression are sensitive to clustering based on the inclusion of household-level life event variables. As the assumption of a consistent travel time budget implies that average travel time tends to be similar over time (refer to Ahmed and Stopher 2014 for details) the analysis makes adjustments for the impact of autocorrelation upon the results of the regressions. The Wooldridge test for autocorrelation in panel data has been conducted and at a 1 per cent confidence interval the null is rejected. This means that the data have first-order autocorrelation [AR(1)]. Note that the terms autocorrelation and serial correlation are used interchangeably within this paper. As a result GLS is utilised to allow for heteroskedasticity and serial correlation simultaneously. In addition, the analysis makes adjustments for the influence of intra-household relationships with the application of cluster-robust standard errors.

With all of these issues accounted for, the following sequence of regressions was performed. Regressions 1a and 1b are FE and RE regressions with robust standard errors applied to account for heteroskedasticity. Regressions 2a and 2b are FE and RE regressions that allow for clustering based on the household groupings. Regression 3 is a GLS regression that accounts for heteroskedasticity and serial correlation simultaneously. As the Breusch and Pagan Lagrangian multiplier test is rejected at a 1 per cent confidence interval and the Hausman test[1] does not lead to a rejection of the null, the focus of this paper concentrates on the RE and GLS regressions. These are regressions 1b, 2b and 3. The importance of the GLS regression is that it allows for both heteroskedasticity and serial correlation. Even though comparability is impacted by the lack of random effects, accounting for the impact of serial correlation is important as Wooldridge (2008) notes that “much of the time serial correlation is viewed as the most important problem, because it usually has a larger impact on standard errors and the efficiency of estimators than does heteroskedasticity” (Wooldridge, 2008: 440).

Table 8 shows the regression results for the average time travelled per day for all of the days of the week, Table 9 shows the regression results for the weekday average and table 10 shows the regression results for the weekend average. This allows us to review the statistical significance of the independent variables across the types of the regressions and the periods of the week. Figure 4 provides an overview of the results by comparing the estimates for the impact of different life events on the average time travelled across periods of the week and the RE or GLS specifications. Different life events are statistically significant for different periods of the week with the average for the weekend providing notable differences to average weekday travel. Entering the workforce (+ve relationship to time travelled), a gain of physical mobility (-ve), a loss of mobility (-ve), changes in the people within the household (-ve) and a change of car ownership (+ve) are the life events that tend to be statistically different from zero for the weekend observations and not for the overall period or weekdays. Focusing upon the GLS regressions (regression 3) for the overall period, eleven variables

[1] Note that the Hausman test has been performed on regressions that are similar to 1a and 1b, except that robust standard errors were not applied.

were statistically significant. These were the constant, the person's age (-ve), leaving the workforce (+ve), increased working hours (-ve), decreased working hours (-ve), increased voluntary hours (-ve), loss of physical mobility (-ve), a gain of mobility (-ve), gaining responsibility of a dependent (-ve), having no individual life event (+ve), and a reduction in the size of the household (+ve).

It should be noted that a gain in physical mobility has only two observations within the sample and the associated parameter estimate should be interpreted with caution. In addition, high levels of correlation between the increase in household size and the addition of a new person in the household, an increase/decrease in amount of hours worked and entering the workforce, as well as the increase in car ownership and the addition of a new car, presents a situation where there is a likelihood of endogeneity bias. Refer to Table 11 in the appendix for the correlation matrix related to this analysis. Sensitivity analysis has been conducted using the GLS regressions and finds that the results remain similar irrespective of endogeneity – except for reduced working hours and an increase in household size. In the case of travel across all days, changes in the parameter estimates due to the removal of 'entering the workforce', 'people changed' and 'car changed' indicators are associated with a change in the estimate and significance associated with a reduction in working hours and an increase in household size.

Table 8 : Average time travelled per day – All days –Unbalanced panel

	Fixed effects	Random effects	Fixed effects	Random effects	GLS
	Robust st err	Robust st err	HH cluster	HH cluster	Het and AR(1)
	1a	1b	2a	2b	3
Constant	128.195*	70.540*	128.195*	70.540*	74.828*
	23.36	10.60	28.05	12.31	6.48
Female	.	1.231	.	1.231	1.126
	.	4.99	.	4.33	2.45
Age	-1.339**	-0.168	-1.339*	-0.168	-0.297*
	0.62	0.16	0.74	0.21	0.09
Entering the workforce	-3.503	0.771	-3.503	0.771	4.585
	12.59	9.11	14.47	11.68	8.47
Leaving the workforce	20.496*	22.615**	20.496*	22.615*	23.192**
	7.05	8.84	6.45	6.61	11.22
Moving work location	4.263	4.983	4.263	4.983	3.971
	4.97	4.10	4.82	3.78	3.77
Changing job role	-2.515	-1.126	-2.515	-1.126	-2.817
	6.61	5.81	4.92	4.25	4.61
Increasing work hours	-12.683	-15.617**	-12.683	-15.617**	-18.527*
	9.34	7.50	8.59	6.44	5.45
Reducing work hours	-18.439*	-18.264*	-18.439*	-18.264*	-18.147*
	5.12	6.61	5.26	4.33	10.60
Increasing voluntary hours	-20.664*	-21.285*	-20.664*	-21.285*	-14.284*
	5.57	5.92	5.88	3.96	7.59
Reducing voluntary hours	26.967*	28.479**	26.967*	28.479**	27.364
	13.82	14.06	14.90	12.06	23.03
Return to part-time education	-16.609	-16.100	-16.609	-16.100**	-46.166
	15.24	16.83	11.81	7.57	47.66
Starting new education	12.875	9.127	12.875	9.127	3.088
	12.08	7.57	13.01	8.32	7.12
Leaving education	16.600	12.219*	16.600	12.219*	6.697
	11.11	6.79	10.83	6.84	6.48
Loss of physical mobility	-4.611	-7.737	-4.611	-7.737	-14.303*
	14.59	10.02	14.86	12.72	8.50
Gain of physical mobility	14.046	15.892	14.046	15.892	11.912
	12.83	18.67	13.37	15.41	16.13
Gain of mobility (driver's licence)	-4.811	-8.414	-4.811	-8.414	-13.255**
	9.26	6.11	10.34	8.60	5.49
Loss of mobility (driver's licence)	1.152	-4.309	1.152	-4.309	-17.924
	2.61	8.47	2.89	2.80	20.50

	Fixed effects	Random effects	Fixed effects	Random effects	GLS
	Robust st err	Robust st err	HH cluster	HH cluster	Het and AR(1)
	1a	1b	2a	2b	3
Gained responsibility of dependents	-13.913*	-13.773*	-13.913**	-13.773*	-19.081**
	7.03	7.72	5.83	3.14	8.41
Reduced responsibility of dependents	-3.633	1.524	-3.633	1.524	-5.390
	8.81	19.52	10.47	12.77	17.46
No individual event	.	9.635*	.	9.635	7.178*
	.	5.684	.	6.609	2.701
Reduction in household size	-2.741	0.805	-2.741	0.805	8.511**
	5.94	4.78	7.30	6.55	3.80
Increase in household size	11.689	15.613	11.689	15.613*	15.010
	13.93	11.67	10.01	8.84	11.25
People changed (new)	2.813	-0.979	2.813	-0.979	2.707
	8.89	8.52	8.93	8.40	9.76
Reduction in car ownership	5.921	4.787	5.921	4.787	3.002
	5.16	4.82	6.45	5.63	3.78
Increase in car ownership	9.407	9.021	9.407	9.021	-0.119
	8.25	6.48	8.55	7.04	4.57
Car changed (new)	-1.575	-0.647	-1.575	-0.647	3.180
	3.65	3.19	4.68	4.06	2.25
Reduction in bike ownership	0.241	-0.498	0.241	-0.498	-0.378
	3.42	3.71	4.38	3.91	3.13
Increase in bike ownership	3.734	3.469	3.734	3.469	0.288
	5.21	4.31	5.95	5.11	3.17
Time trend	0.493	-0.742	0.493	-0.742	-0.295
	1.49	0.74	1.73	0.92	0.59
N	335	335	335	335	334
i	68	68	68	68	67
R^2 - within	0.0986	0.0930	0.0986	0.0930	
R^2 - between	0.0053	0.1209	0.0053	0.1209	
R^2 - total	0.0201	0.1209	0.0201	0.1209	
Wald $\chi^2(28)$					94.65***

Note: P Value: *** - 1% ** - 5% * - 10%. Breusch and Pagan Lagrangian multiplier test for random effects is rejected at 1% for both of the random effect specifications and hence random effects are preferred. Utilising the Hausman test results in no rejection of null and hence random effects are preferred.

Table 9 : Average time travelled per day – Weekdays –Unbalanced panel

	Fixed effects	Random effects	Fixed effects	Random effects	GLS
	Robust st err	Robust st err	HH cluster	HH cluster	Het and AR(1)
	1a	1b	2a	2b	3
Constant	39.611	65.508*	39.611	65.508*	66.451*
	23.83	11.69	26.38	12.60	6.33
Female	.	2.379	.	2.379	3.594
	.	5.52	.	4.93	2.52
Age	0.659	-0.088	0.659	-0.088	-0.221**
	0.61	0.17	0.69	0.20	0.09
Entering the workforce	-4.924	-2.396	-4.924	-2.396	-1.219
	16.03	11.92	18.30	14.88	8.12
Leaving the workforce	14.382*	16.822**	14.382	16.822**	17.924
	8.25	8.12	9.62	7.05	11.80
Moving work location	9.063*	9.131*	9.063*	9.131**	4.837
	4.94	4.90	5.30	4.59	4.25
Changing job role	-0.747	0.152	-0.747	0.152	-2.031
	6.64	5.18	4.87	3.75	4.81
Increasing work hours	-16.209	-17.695**	-16.209*	-17.695**	-15.425*
	10.44	8.67	9.46	7.82	5.32
Reducing work hours	-20.552*	-20.664*	-20.552*	-20.664*	-17.430
	4.75	6.07	3.74	2.92	10.67
Increasing voluntary hours	-18.054*	-17.927*	-18.054*	-17.927*	-14.198
	4.85	6.57	5.16	4.00	9.98
Reducing voluntary hours	32.505**	36.167**	32.505*	36.167*	40.557
	16.28	17.58	16.86	13.76	29.03
Return to part-time education	0.055	-4.013	0.055	-4.013	-30.216
	17.87	15.62	17.49	12.41	31.06
Starting new education	17.245	14.265*	17.245	14.265	13.434**
	13.38	7.35	14.93	10.12	5.84
Leaving education	17.653	15.260*	17.653	15.260	14.380**
	13.82	8.60	13.12	9.32	7.04
Loss of physical mobility	-8.214	-9.649	-8.214	-9.649	-15.611
	16.49	11.38	16.83	14.32	10.19
Gain of physical mobility	16.441*	19.088	16.441	19.088	11.497
	9.20	18.60	9.88	12.97	20.67
Gain of mobility (driver's licence)	-4.884	-8.285	-4.884	-8.285	-18.909*
	9.32	5.94	10.43	8.96	5.13
Loss of mobility (driver's licence)	3.203	-0.730	3.203	-0.730	-18.466
	3.53	8.30	4.01	3.43	21.90

	Fixed effects	Random effects	Fixed effects	Random effects	GLS
	Robust st err	Robust st err	HH cluster	HH cluster	Het and AR(1)
	1a	1b	2a	2b	3
Gained responsibility of dependents	-15.309	-13.500	-15.309	-13.500**	-12.944
	11.84	8.57	10.59	6.40	9.17
Reduced responsibility of dependents	-6.576	-2.031	-6.576	-2.031	4.682
	9.47	14.03	9.60	8.33	16.89
No individual event	.	7.907	.	7.907	4.517
	.	6.19	.	6.91	2.92
Reduction in household size	-1.025	1.732	-1.025	1.732	10.567*
	7.580	5.764	7.986	7.051	4.023
Increase in household size	4.278	7.005	4.278	7.005	13.163
	17.47	12.22	12.67	9.14	11.15
People changed (new)	9.911	7.244	9.911	7.244	3.165
	13.23	7.98	11.11	7.82	8.62
Reduction in car ownership	0.519	-0.257	0.519	-0.257	-2.628
	6.16	4.82	5.99	5.17	3.81
Increase in car ownership	7.257	6.924	7.257	6.924	3.251
	7.15	5.65	7.37	5.96	4.99
Car changed (new)	-3.528	-2.913	-3.528	-2.913	-0.419
	4.24	3.42	5.20	4.34	2.37
Reduction in bike ownership	0.727	0.261	0.727	0.261	-0.923
	3.62	3.81	4.91	4.27	3.37
Increase in bike ownership	3.247	2.971	3.247	2.971	0.010
	5.28	4.28	5.06	4.21	3.29
Time trend	-1.315	-0.636	-1.315	-0.636	0.296
	1.43	0.72	1.70	0.88	0.64
N	334	334	334	334	333
i	68	68	68	68	67
R^2 - within	0.1214	0.1171	0.1214	0.1171	
R^2 - between	0.0001	0.0603	0.0001	0.0603	
R^2 - total	0.0085	0.0937	0.0085	0.0937	
Wald $\chi^2(28)$					80.49***

Note: P Value: *** - 1% ** - 5% * - 10%. Breusch and Pagan Lagrangian multiplier test for random effects is rejected at 1% for both of the random effect specifications and hence random effects are preferred. Utilising the Hausman test results in no rejection of null and hence random effects are preferred.

Table 10: Average time travelled per day – Weekends –Unbalanced panel

	Fixed effects	Random effects	Fixed effects	Random effects	GLS
	Robust st err	Robust st err	HH cluster	HH cluster	Het and AR(1)
	1a	1b	2a	2b	3
Constant	475.991*	81.394*	475.991*	81.394*	66.175*
	48.56	13.87	59.80	15.70	10.23
Female	.	-1.989	.	-1.989	2.599
	.	6.23	.	4.80	3.87
Age	-9.084*	-0.276	-9.084*	-0.276	-0.451*
	1.25	0.20	1.52	0.25	0.14
Entering the workforce	20.702	26.232*	20.702	26.232*	24.685
	19.15	13.98	18.41	10.04	17.64
Leaving the workforce	32.641	29.979	32.641	29.979	46.014*
	23.73	21.37	25.39	23.02	16.82
Moving work location	-2.256	3.417	-2.256	3.417	3.125
	10.53	8.81	11.69	9.45	6.83
Changing job role	-3.815	2.028	-3.815	2.028	-2.707
	12.24	11.41	11.52	8.74	7.89
Increasing work hours	-25.718	-29.255*	-25.718	-29.255**	-22.649*
	20.56	11.29	21.22	13.04	13.17
Reducing work hours	-17.603	-12.080	-17.603	-12.080	-29.945*
	15.16	16.62	14.02	14.50	16.52
Increasing voluntary hours	-20.059	-25.515*	-20.059	-25.515*	-17.750
	13.68	14.42	14.25	14.66	12.85
Reducing voluntary hours	9.682	-0.035	9.682	-0.035	18.836
	24.09	19.36	22.93	16.24	26.69
Return to part-time education	9.050	30.559	9.050	30.559**	24.972
	18.59	21.39	15.33	13.92	57.37
Starting new education	-6.683	-15.641	-6.683	-15.641*	-23.125
	11.57	16.41	10.19	8.87	14.63
Leaving education	5.873	4.866	5.873	4.866	7.357
	9.85	10.44	8.42	10.98	10.43
Loss of physical mobility	15.860	-1.551	15.860	-1.551	-4.445
	10.04	8.32	10.03	8.78	10.32
Gain of physical mobility	-22.047**	-28.006*	-22.047**	-28.006*	-20.323
	8.86	7.87	9.88	8.13	17.55
Gain of mobility (driver's licence)	3.603	4.633	3.603	4.633	3.075
	16.69	10.44	17.64	13.42	8.18
Loss of mobility (driver's licence)	-3.998	-22.304**	4.633	3.603	4.633
	5.04	8.72	10.44	17.64	13.42

	Fixed effects	Random effects	Fixed effects	Random effects	GLS
	Robust st err	Robust st err	HH cluster	HH cluster	Het and AR(1)
	1a	1b	2a	2b	3
Gained responsibility of dependents	-8.293	-23.157	-8.293	-23.157**	-6.856
	18.01	15.16	11.80	10.07	13.82
Reduced responsibility of dependents	18.780	22.287	18.780	22.287	4.806
	26.52	35.21	26.59	31.18	31.81
No individual event	.	14.495*	.	14.495*	16.506*
	.	7.49	.	7.65	4.91
Reduction in household size	-4.169	5.925	-4.169	5.925	17.952**
	14.070	10.037	17.976	14.020	7.964
Increase in household size	44.993**	46.643**	44.993*	46.643*	54.361**
	22.44	23.78	25.98	24.39	24.25
People changed (new)	-31.351*	-31.334	-31.351	-31.334	-45.784**
	15.97	20.51	20.28	21.16	22.45
Reduction in car ownership	13.949	8.812	13.949	8.812	3.015
	10.44	9.22	12.77	13.01	7.99
Increase in car ownership	4.228	4.934	4.228	4.934	0.507
	11.94	9.91	11.09	10.78	7.27
Car changed (new)	2.441	5.309	2.441	5.309	7.466*
	7.65	6.06	9.72	8.73	4.26
Reduction in bike ownership	3.966	2.297	3.966	2.297	2.124
	5.91	6.23	7.57	6.81	4.66
Increase in bike ownership	6.042	6.240	6.042	6.240	7.248
	10.84	9.70	10.00	9.41	5.63
Time trend	7.352**	-1.525	7.352**	-1.525	0.239
	2.90	1.41	3.28	1.76	0.95
N	325	325	325	325	356
i	67	67	67	67	66
R^2 - within	0.0535	0.0383	0.0535	0.0383	
R^2 - between	0.0242	0.1988	0.0242	0.1988	
R^2 - total	0.0106	0.0993	0.0106	0.0993	
Wald $\chi^2_{(28)}$					77.55***

Note: P Value: *** - 1% ** - 5% * - 10%. Breusch and Pagan Lagrangian multiplier test for random effects is rejected at 1% for both of the random effect specifications and hence random effects are preferred. Utilising the Hausman test results in no rejection of null and hence random effects are preferred.

6. Discussion of challenges and opportunities

This section discusses the challenges that were faced in completing the analysis based on the dataset described in section 2. With each of the challenges discussed, this section also discusses how the analysis has attempted to overcome the issues identified. Where applicable the discussion also highlights some of the opportunities that arise from the challenges identified.

6.1 Sample Size

A small sample will limit the reliability of results for any given statistical analysis and while the dataset compiled for the analysis has a good number of observations overall, some life events had one or no observations attributed to them after the dataset was compiled. While the number of observations that had no individual life event during the whole sample period was relatively high (163 cases or approximately 32 per cent), the number of observations without a household life event attributed to that record was small (23 cases or approximately 5 per cent). As a result the univariate analysis had some small groupings when testing for differences in the sample means for household life events. In section 3.1 it was stated that the interpretation of the results associated with no household-level life event (No HLE) should be conducted with caution due to small size.

Small samples are a difficult issue to overcome once the data collection is complete. Re-sampling over time based on key characteristics will allow for a stronger dataset. In our case, the original dataset was constructed as a randomly sampled rolling panel survey with replacement of households to manage attrition within the jurisdictions. There was no controlling for household size or household composition. The sample used in this analysis was constrained by households participating in 5 or more waves. This presents an issue in relation to the construction of HLE, as the number of single-person households in the sample influences the number of No HLE. Expanding the criteria to 4 or more waves would increase the sample size, and possibly increase the number of households with No HLE. Note that section 0 also discusses re-sampling and small sample size as the discussion focuses upon attrition.

6.2 Combining Data

With a dataset that contains observations at the daily, individual and household levels the appropriate level to merge the data became an important consideration. As described in section 2, the focus of this paper is on the individual level as the basis of the empirical analysis, which is driven by the average time travelled per day and an expectation of consistency over the days surveyed due to the travel time budget hypothesis. The challenge in this case was the creation of a multi-level dataset at an aggregation that is appropriate for the analysis of time travelled per day without the introduction of bias.

With the allowance for clustering based on the household the analysis has obtained more robust results in comparison to a situation where the impact of correlation across the household grouping was not assessed. The appropriateness of aggregating the trip and daily level observations to the average time travelled per day for each wave will be the subject for future research. However, previous findings of consistency (Stopher and Zhang**,** 2011b) and the occurrence of serial-correlation implies that this aggregation is not inappropriate.

6.3 Treatment of multiple waves of data

As noted in section 2, the dataset used for the analysis has nine waves of data for the period between 2005 and 2012. The second wave was collected approximately six months after the first wave, while the majority of the other waves included the months of October and November. Within the multivariate analysis the waves are coded as an annual time trend. In accordance, these data are coded as one indicator with the values specified as 1 for 2005, 2 for March to April 2006, 2 for October to December 2006, 3 for October to December 2007, and so on. As previously noted, coding wave 2 as 2 allows us to have a consistent annual time dimension and an annual time trend implemented in the multivariate analysis. Sensitivity testing with alternative arrangements for the time trend has shown minor differences in the parameter estimates.

6.4 Regularity

The finding of serial-correlation in the analysis was mentioned in the previous sub-section, however due to the importance of the issue it is discussed a subsequent time. There is a danger of spurious results when serial-correlation is unaccounted for and accordingly GLS has been utilised so as to simultaneously account for heteroskedasticity and serial-correlation.

With respect to the analysis in section 3, establishing that serial-correlation exists implies that there is a certain regularity in the average time travelled per day across waves. In part, this points towards a consistent level of daily travel and the review of travel time budget studies by Ahmed and Stopher (2014) should be referred to for more discussion of the relevance of this finding.

6.5 Endogeneity

There is also a danger of spurious results when correlation between independent variables occurs. Upon reviewing the correlation of the indicators reviewed in the paper, unsurprisingly, significant correlation occurs in the case of indicators that were derived from another indicator. As previously noted in section 3.2, this refers to the correlation between the increase in household size and the addition of a new person in the household, an increase/decrease in amount of hours worked and entering the workforce, as well as the increase in car ownership and the addition of a new car. Sensitivity analysis has been conducted using the GLS regressions and finds that the results remain similar irrespective of endogeneity – except for changes in the reduction of working hour and an increase in household size. In the case of travel across all days, changes in the parameter estimates due to the removal of the entering the workforce, people changed and car changed indicators are associated with a change in the estimate and significance associated with an reduction in working hours and an increase in household size. As the differences are relatively minor and not unexpected, the derived indicators remain within the regression analysis so as to highlight the issue and provide an example of endogeneity bias.

6.6 Attrition

A notable loss of panel members may occur without the provision of incentives. Some level of attrition in panel surveys is expected, as circumstances change and the realities of participating in a survey become clearer to respondents (Kish, 1965). This paper presents an analysis of households that continued to participate in multiple waves despite there being no monetary incentive, but even within these households attrition existed at the individual level. Within a multi-person household there were cases of individuals choosing not to participate in the GPS travel survey component or who were absent from the household in some waves.

Re-sampling was raised in the discussion concerning sample size and this may allow for an ongoing data collection process when notable attrition may produce a small sample and/or the sample is required to remain representative of a certain population. Refer to Watson and Wooden (2012) and Watson (2006) for an interesting discussion of re-sampling within a longitudinal survey where the sampling approach focuses upon being representative of the entire Australian population. Re-sampling raises the question of how long a study will be conducted and at some point this will be related to securing long-term investment for longitudinal travel surveys and research. In this case study the original 2005-2007 short-term evaluation study was funded separately to the six wave longitudinal study, and it was only because of the compatibility of the sampled population and data collection that the datasets could be combined. Developing stronger business cases for longitudinal travel surveys is more useful than relying on happenstance. Attrition also raises the issue of basing an analysis upon an unbalanced panel, as is the case within this analysis of a sub-sample of the 620 households that participated in the GPS travel survey. Note that the construction of this dataset was based on following as many households and people over the waves that the inherent rate of attrition allowed. People were added to the dataset when household size increased, or as children became 14 years. However, there was no specific re-sampling approach for addressing individual-level attrition.

7. Conclusion

This paper examines the influence of different life change events on travel time by using an existing longitudinal GPS travel survey that had detailed travel time and sociodemographic data at the household and individual level. Although it is common to collect sociodemographic data, this paper was able to explore how regular updates to the sociodemographic detail could be used to construct life events that could have an influence on travel time, and specifically the notion of travel-time being relatively stable. Exploring the influence of life changes on travel survey datasets is normally difficult as there has not been much investment in longitudinal panels. Funding is shorter term, innovative technologies require pilots to develop the business case, yet the technologies can also become redundant quite quickly. Despite the uncertainty about the willingness of households to carry GPS devices for travel survey research, Stopher and his team have been successful in doing so, even without offering participants financial incentives.

This paper reports on the opportunities and challenges in conducting such an analysis, and presents a case for future investment in longitudinal panels. A number of issues are raised in the paper and those related to the analysis of the data are discussed in Section 6. Other issues of relevance to a community of travel survey researchers and their funders are discussed below.

The context for collecting the travel data is important. The dataset used in this paper was originally commissioned to study the long-term influence of VTBCP programs to reduce car use. However, the sample of 29 households was not coded by their exposure to VTBCP because of concerns of further reducing the sample size. Future analysis of the dataset could however be achieved if the criteria for multi-wave households is expanded to a minimum of three or four waves. Any analysis however would need to be cautious as no data were collected during the annual updates about the individual's or household's subsequent exposure to TravelSmart marketing from other sources, such as workplace or school travel behaviour programs.[2]

The secondary use of survey data, for purposes different to the original purpose is somewhat an unavoidable issue. Gaining long-term funding support is difficult, especially when using untried or expensive technology. Moreover funding for new policy concerns can become highly politicised – as has been observed in Australia for programs that were originally contributing to concerns about climate change and the need to reduce emissions from transport activities but have had funding cuts, or been shut down when there was a change of government. Moreover, the short-term funding also can make it difficult to maintain research staff and therefore presents potential problems in managing and interpreting the dataset.

The pragmatic researcher needs to design and resource a survey method and data processing process that would be appealing and cost-efficient to funders. While the constraints of the funding environment are unlikely to change anytime soon, researchers could be putting more consideration into the robustness of the methodology and data management processes if future funding were to be secured. It may be more pragmatic for the wider community of transport survey researchers to look at opportunities to pool their datasets. This would require guidance and protocols to standardize how travel data is managed, and specifically the structure of databases that would be easier to adapt for longitudinal analysis, by for example allowing the linking of updated sociodemographic data to the individual records.

Further work in quantifying the effect of such life course changes over the long-term is important to avoid over-stating the constancy of travel behaviour and mode preferences, because social norms are also subject to change. For example, obtaining a drivers' licence in the final years of secondary

[2] Although the TravelSmart program ran from 2001-2005, TravelSmart resources for employers, teachers, local government and universities continue to be available on the www.travelsmart.gov.au website.

schooling may have been a cultural norm in Australia, but there have been observations of a new trend emerging where young people are attaining their full licence later. In this paper, the analysis of travel-related life events was limited to reported changes in the availability of cars and bikes in the household. Information on public transport used were not captured in the survey paper forms. However, the larger GPS individual-level data records have inferred mode use by processing data on travel speeds and routes taken. This presents an opportunity for future work as this information could be used to construct a variable on change in use of public transport and possibly active transport (cycling and walking).

Changing expectations about the delivery of personalised travel information, through mobile devices, may place new pressure (and opportunities) on transport researchers to understand the effectiveness of travel policy interventions on different segments of the community. This paper presents an exploratory analysis of some of the opportunities that could be pursued. Life stages were difficult to analyse because nine years, although long, reduces instances of movement through life stages. While the dataset was not sufficient to provide a meaningful analysis on how life stages may affect travel time, a focus on life events was possible. Life events provided more flexibility as individuals and households were defined as having a change relative to their status in the first year of participation. With a concerted effort, travel survey researchers and their funders could work together to create more opportunities to pursue this line of research.

References

Ahmed, A. and P. Stopher (2014): 'Seventy Minutes Plus or Minus 10 — A Review of Travel Time Budget Studies', *Transport Reviews*, 34, 607-625.

Delbosc and Currie (2013): 'Exploring Attitudes of Young Adults toward Cars and Driver Licensing', paper presented to the *Australasian Transport Research Forum*, 2-4 October, Brisbane, Australia.

Hanson, S. and J.O. Huff (1986): 'Classification Issues in the Analysis of Complex Travel Behavior', *Transportation*, 13, 271-293.

Hanson, S. and J.O. Huff (1988): 'Systematic Variability in Repetitious Travel', *Transportation*, 15, 111-135.

Habib, K.M.N., E.J. Miller, and K.W. Axhausen (2008): 'Weekly Rhythm in Joint Time Expenditure for All At-Home and Out-of-Home Activities', *Transportation Research Record 2054*, 64-73.

Kracht, M. (2006): 'Using Combined GPS and GSM Tracking Information for Interactive Electronic Questionnaires', in P.R. Stopher and C.C. Stecher *(ed.) Travel Survey Methods: Quality and Future Directions*, Elsevier, Sydney, ch. 30.

Lanzendorf, M. (2010): 'Key Events and Their Effect on Mobility Biographies: The Case of Childbirth', *International Journal of Sustainable Transportation*, 4, 272-292.

Le Vine, S. and J. Polak (2014): 'Factors Associated With Young Adults Delaying and Foregoing Driving Licenses: Results from Britain', *Traffic Injury Prevention*, 15, 794-800.

Lucas, K. (2012): 'Transport and social exclusion: Where are we now?', *Transport policy*, 20, 105-113.

Mitra, R. (2013): 'Independent Mobility and Mode Choice for School Transportation: A Review and Framework for Future Research', *Transport Reviews* 33, 21-43.

Murakami, E. and W.T. Watterson (1989): 'Developing a Household Travel Panel Survey for the Puget Sound Region', *Transportation Research Record No. 1285*, 40-46.

Pas, E.I. (1988): 'Weekly Activity-Travel Behavior', *Transportation*, 15, 89-109.

Scheiner, J. and C. Holz-Rau (2013): 'A comprehensive study of life course, cohort, and period effects on changes in travel mode use', *Transportation Research Part A: Policy and Practice*, 47, 167-181.

Shapcott, M. (1978): 'Comparison of the Use of Time in Reading, England with Time Use in Other Countries', *Transactions of the Martin Centre for Architectural and Urban Studies*, 3, 231-257.

Stopher, P., E. Clifford, N. Swann and Y. Zhang (2009): 'Evaluating voluntary travel behaviour change: Suggested guidelines and case studies.' *Transport Policy*, 16, 315-324.

Stopher, P.R., Y. Zhang, J. Zhang, and B. Halling (2009): 'Results of an Evaluation of TravelSmart in South Australia', paper presented to the 32nd Australasian Transport Research Forum, Auckland, 29 September–1 October.

Stopher PR and Y. Zhang (2011a): 'Repetitiveness of Daily Travel', *Transportation Research Record*, 2230, 75-84

Stopher PR and Y. Zhang (2011b): 'Travel Time Expenditures and Travel Time Budgets - Preliminary Findings', *Proceedings of the 90th Annual Meeting of the Transportation Research Board TRB 2011*, Washington, D.C., United States, 27 January.

Stopher, P.R., C.J. Moutou and W. Liu (2013): ‘Sustainability of Voluntary Travel Behaviour Change Initiatives – A 5-Year Study’, paper presented to the *36th Annual Australasian Transport Research Forum ATRF 2013*, Brisbane, Australia, 2-4 October.

Van Wissen, L.J.G. and H.J. Meurs (1989): ‘The Dutch mobility Panel: Experiences and Evaluation’, *Transportation*, 16, 99-119.

Watson, N. (2006): ‘Options for a Top-up Sample to the HILDA Survey’, paper presented to the *ACSPRI Social Science Methodology Conference*, Sydney, 10-13 December.

Watson, N., M Wooden (2012): ‘The HILDA Survey: a case study in the design and development of a successful household panel study’, *Longitudinal and Life Course Studies*, 3, 369-381.

Wooldridge, J. (2008): *Introductory Econometrics: a modern approach*, Third Edition, Thomson South-Western, Mason.

Zhang, Y. and P.R. Stopher **(**2010): 'Exploring Travel Time Budgets with GPS Data', paper presented to the *2nd Workshop on Time Use Observatory TUO 2 2010*, Termas de Chillan, Chile, 19th March.

Zhang, Y., P.R. Stopher, and B. Halling (2013): 'Evaluation of South-Australia's TravelSmart project: Changes in community's attitudes to travel', *Transport Policy*, 26, 15-22.

Zumkeller, D., J.-L. Madre, B. Chlond, and J. Armoogum (2006): ‘Panel Surveys’, in P.R. Stopher and C.C. Stecher *(ed.) Travel Survey Methods: Quality and Future Directions*, Elsevier, Sydney, ch. 20.

Appendix

Table 1: Correlation matrix of key indicators

	Average time travelled per day	Female	Age	Entering the workforce	Leaving the workforce	Moving work location	Changing job role	Increasing work hours	Reducing work hours	Increasing voluntary hours
Average time travelled per day	1									
Female	0.055	1								
Age	-0.0825	-0.1427*	1							
Entering the workforce	-0.0477	-0.0252	-0.0396	1						
Leaving the workforce	0.0347	-0.0391	0.0333	-0.0192	1					
Moving work location	0.0146	-0.0291	0.0101	0.1734*	-0.0295	1				
Changing job role	-0.0511	-0.0379	-0.0824	0.3874*	-0.0288	0.1859*	1			
Increasing work hours	-0.1072	-0.0443	-0.0659	0.6827*	-0.0252	0.2751*	0.4442*	1		
Reducing work hours	-0.0099	-0.0391	0.0172	-0.0192	0.6606*	-0.0295	-0.0288	-0.0252	1	
Increasing voluntary hours	-0.0915	-0.0787	0.0148	0.1293*	0.1377*	0.074	0.1746*	0.0922	-0.0135	1
Reducing voluntary hours	0.0791	-0.0532	0.0571	0.3134*	-0.006	-0.0098	0.2087*	-0.0083	-0.006	-0.0045
Reducing to part-time education	0.0066	0.0529	-0.0538	-0.009	-0.0085	-0.0138	-0.0135	-0.0118	-0.0085	-0.0063
Starting new education	-0.0198	0.0527	-0.2144*	0.0493	-0.0259	0.0091	0.0112	0.0233	-0.0259	-0.0193
Leaving education	-0.0357	-0.0195	-0.0591	0.1156*	-0.0148	0.0637	0.2452*	0.2835*	-0.0148	0.3582*
Loss of physical mobility	-0.1464*	0.0252	0.2191*	0.0711	-0.0211	-0.0342	-0.0334	0.0429	-0.0211	-0.0156
Gain of physical mobility	0.0294	-0.0399	0.0562	-0.011	-0.0104	-0.0169	-0.0165	-0.0145	-0.0104	-0.0077
Gain of mobility (driver's licence)	-0.0819	-0.0781	-0.1590*	-0.0258	-0.0244	0.0146	0.0721	0.0289	-0.0244	-0.0181
Loss of mobility (driver's licence)	-0.0748	-0.0532	0.0711	-0.0063	-0.006	-0.0098	-0.0095	-0.0083	-0.006	-0.0045
Gained responsibility of dependents	-0.0208	-0.0613	0.0191	-0.0127	0.1568*	-0.0196	-0.0191	-0.0167	0.1568*	-0.009
Reduced responsibility of dependents	0.0956	0.0529	0.0567	-0.009	0.2298*	-0.0138	-0.0135	-0.0118	-0.0085	-0.0063
No individual effect	0.1517*	0.0023	0.1115	-0.0984	-0.0932	-0.1512*	-0.1477*	-0.1292*	-0.0932	-0.0692
Reduction in household size	0.0625	-0.0332	0.041	0.0925	0.0345	0.0746	0.0354	0.102	0.101	-0.0238
Increase in household size	0.109	0.0227	-0.1103	-0.0266	-0.0252	-0.0409	-0.0399	-0.0349	-0.0252	-0.0187
People changed (new)	0.0909	0.0213	-0.0798	0.1089	0.0445	0.0463	0.0019	0.0677	-0.0288	-0.0214
Reduction in car ownership	0.0309	0.0104	0.0642	0.0133	0.0182	0.0096	0.0493	0.0715	0.0182	-0.0286
Increase in car ownership	0.1138	-0.0179	-0.0058	-0.0395	0.0208	0.0501	0.0162	-0.0091	-0.0374	0.05
Car changed (new)	0.0869	-0.02	-0.1362*	0.0706	0.0167	0.0554	0.0193	0.105	0.0167	-0.0174
Reduction in bike ownership	0.0123	0.0165	-0.1335*	0.0568	0.0114	0.0334	0.0371	0.0201	0.0114	-0.031
Increase in bike ownership	0.0296	0.0607	-0.1873*	-0.0042	0.0009	-0.0155	0.0187	0.0051	0.0489	-0.035
Time trend	-0.0837	0.0093	0.0683	0.0959	0.0403	0.1239*	0.1019	0.0533	0.0705	0.0523
	Reducing voluntary hours	Reducing to part-time education	Starting new education	Leaving education	Loss of physical mobility	Gain of physical mobility	Gain of mobility (driver's licence)	Loss of mobility (driver's licence)	Gained responsibility of dep.	Reduced responsibility of dep.
Reducing voluntary hours	1									
Reducing to part-time education	-0.0028	1								
Starting new education	-0.0086	0.1579*	1							
Leaving education	-0.0049	-0.0069	-0.0211	1						
Loss of physical mobility	-0.007	-0.0099	-0.0301	-0.0171	1					
Gain of physical mobility	-0.0035	-0.0049	-0.0149	-0.0085	0.1571*	1				
Gain of mobility (driver's licence)	-0.0081	-0.0114	0.0871	0.0845	-0.0283	-0.014	1			
Loss of mobility (driver's licence)	-0.002	-0.0028	-0.0086	-0.0049	-0.007	-0.0035	-0.0081	1		
Gained responsibility of dependents	-0.004	-0.0056	-0.0172	-0.0098	-0.014	-0.0069	-0.0162	-0.004	1	
Reduced responsibility of dependents	-0.0028	-0.004	-0.0121	-0.0069	-0.0099	-0.0049	-0.0114	-0.0028	-0.0056	1
No individual effect	-0.0308	-0.0436	-0.1331*	-0.0759	-0.108	-0.0535	-0.1252*	-0.0308	-0.0618	-0.0436
Reduction in household size	-0.0106	-0.015	0.0492	0.0551	-0.0372	-0.0184	0.1580*	-0.0106	-0.0213	0.1251*
Increase in household size	-0.0083	0.1630*	0.1417*	-0.0205	-0.0292	-0.0145	0.0289	-0.0083	0.2310*	-0.0118
People changed (new)	-0.0095	0.1410*	0.1682*	-0.0235	-0.0334	-0.0165	0.0167	-0.0095	0.1998*	0.1410*
Reduction in car ownership	-0.0127	-0.018	-0.055	0.1072	0.0047	0.1733*	-0.0088	-0.0127	0.0591	0.1015
Increase in car ownership	-0.0124	0.105	-0.0534	0.0406	-0.0433	-0.0215	0.1256*	-0.0124	-0.0248	0.105
Car changed (new)	-0.0277	0.1018	0.1432*	-0.0272	-0.0097	-0.048	0.0646	-0.0277	0.0444	0.0313
Reduction in bike ownership	-0.0138	-0.0195	0.0162	-0.0339	-0.0483	-0.0239	-0.0159	-0.0138	0.1308*	-0.0195
Increase in bike ownership	-0.0156	-0.0221	0.0698	-0.0384	-0.0129	0.0556	-0.0634	-0.0156	-0.0313	-0.0221
Time trend	0.0233	0.0171	0.0197	-0.0257	0.016	0.0144	0.009	0.0233	-0.0435	-0.0148
	No individual effect	Reduction in household size	Increase in household size	People changed (new)	Reduction in car ownership	Increase in car ownership	Car changed (new)	Reduction in bike ownership	Increase in bike ownership	Time trend
No individual effect	1									
Reduction in household size	-0.0138	1								
Increase in household size	-0.0117	-0.0445	1							
People changed (new)	-0.0439	-0.0508	0.8745*	1						
Reduction in car ownership	-0.0528	0.3993*	-0.0534	-0.061	1					
Increase in car ownership	-0.0765	-0.066	0.2042*	0.2424*	-0.0792	1				
Car changed (new)	0.0352	0.1082	0.1540*	0.2144*	-0.0429	0.4128*	1			
Reduction in bike ownership	-0.0636	0.0201	0.1758*	0.1403*	-0.0085	0.0234	-0.0035	1		
Increase in bike ownership	0.0029	-0.0833	-0.0654	-0.0748	0.0447	0.0265	0.1239*	-0.1082	1	
Time trend	0.0724	0.1422*	-0.1353*	-0.0844	0.0734	0.005	0.1476*	0.0446	0.2312*	1