An Analysis of Gender Differences in Vehicles Miles Traveled (VMT) Using Nonparametric Methods

(Preliminary and Incomplete)

by

Brian W. Sloboda
U.S. Department of Transportation
Bureau of Transportation Statistics
Office of Advanced Studies
400 7th Street, S.W. (Room 3430)
Washington, D.C. 20590
(202) 366-5342 (Voicemail)
brian.sloboda@bts.gov

Vincent W. Yao
Institute for Economic Advancement
University of Arkansas at Little Rock
2801 South University Avenue
Little Rock, AR 72204-1099
(501) 569-8453 (Voicemail)
Email: wxyao@ualr.edu

Abstract:

In the United States as in many nations, there are often differences between the travel patterns of men and women with regards to the differences in travel. Traditionally, women make shorter work trips, make greater use of public transit, make more trips for the purpose of serving another person's travel needs, and drive far fewer miles per year than men. These differences in travel are delineated by the separate social responsibilities of men and women. However, in the past few decades, women have been participating more in the labor force. In addition, women still retain their family obligations as nurturers, shoppers, and homemakers. Given the changes in the transit patterns of women in recent decades, women's travel patterns still differed substantially from those of men. In fact, these emerging trends from transit patterns of women, their actual vehicle miles traveled (VMT) are starting to increase and may surpass the VMT of men. Additionally, it is speculated by transportation planners and policy-makers that the VMT of women will surpass the VMT of men in the future.

In the past, transportation studies have not been particularly oriented to women's' travel issues despite the presence of the data and statistical methodologies. On the other hand, surveys such as the National Household Transportation Survey (NHTS) can be used to understand trends in women's travel patterns, which are often attributed to changes in labor force participation, household structure, and attitudes. This paper will analyze the differences in the vehicle miles traveled (VMT) between men and women using nonparametric methods using the data from the NHTS data as prepared by the Bureau of Transportation Statistics (BTS) in the U.S. Department of Transportation. Moreover, the examination of the relationship of VMT and other household variables would be estimated in a flexible way which cannot be assessed using parametric modeling methods.

While women have made many advances in their participation in the labor force in the past several decades, women still continue to face the daunting task of balancing their individual, family, and career responsibilities. Women's travel continues to grow not only in the total vehicle miles of travel (VMT) but also in the number of trips, the frequency of trips, and the length of trips. Despite these significant gains, child care responsibilities are still most likely held by women, and women often must make career and job location decisions in part based on the accessibility to family from those jobs. Given their greater labor force participation in the labor force and their obligatory responsibilities, their needs for more reliable transportation continue to change over time.

Current research continues to show major differences in the travel patterns of men and women, which are often attributed to historical and economic trends.[1] However, other influential factors explaining the differences in the travel patterns are also attributed to gender roles and their obligatory responsibilities to the family and nonfamily members. At the same time, research suggests that variations in the travel patterns of women may also result from differences in the residential and labor market opportunities available to women as well as different transportation options available to them. The objective of this paper is to assess the variations of VMT by gender via nonparametric methods using the 2001 National Household Travel Survey (NHTS). Using nonparametric methods will allow for an assessment of VMT and other characteristics in more flexible ways which cannot fully assessed via the standard parametric methods.

Section II Review of the Literature

For the first half of the twentieth century, men and women had delineated economic roles. More specifically, men worked outside the home and had the responsibility for providing for the household while women had the responsibility for managing household affairs [Blau and Ferber (1986) and Becker (1991)]. However, in the latter years of the twentieth century, women began to actively participate in the labor force and must often juggle their household responsibilities with their vocational responsibilities. Consequently, their greater labor force participation has resulted in changes in their travel patterns especially for women who have children in the past twenty years (Rosenbloom and Burns, 1989). Additionally, the travel patterns of women seem to reflect their constant balance to juggle work and household responsibilities [Wachs (1987) and Hanson and Johnston (1985)].

The linking of nonwork trips to work trips has been studied by [Adler and Ben-Akiva (1979); Oster (1979);Goulias and Kitamura (1989); Golob (1986); Kitamura (1984); Oster (1978); and Nishii, Kondo, and Kitamura (1988). Strathman, Dueker, and Davis (1993) used the concept of trip chaining to examine the propensity of households to add nonwork trips to the work commute and the allocation of nonwork trips through chaining. They found that workers who commuted in peak periods had a lower propensity to form work/nonwork trips. They also found that certain household types contributed the largest amounts of peak period chaining behavior: single adults; dual-income couples; dual-income families with preschoolers; and multiworker households. These types of households were also the fastest growing type of household formations. Single-occupant commuters had a higher propensity to add nonwork trips to their commute. Trip-chaining analysis lead to the puzzling conclusion that even with the increase of congestion during peak periods, people continue to trip chain. This could indicate an inelastic demand for the activities and/or locations regardless of the time or monetary price to be paid due to congestion. Downs (1982) points out that many nonwork activities are concentrated in the peak commuting periods, as people take children to school or run errands before and after work. According to Gordon et al. (1988), between 1977 and 1983, nonwork trips grew faster than commute trips and grew during peak periods. Richardson and Gordon (1989) found that the overall growth in nonwork travel accounted for 70 to 75 percent of all weekday trips. They also determined that in every size of SMSA, nonwork travel grew three to four times faster than work trips. Davidson (1991) examined the exact nature of trip chains in a study of employees, and found that the employees were twice as likely to make stops on their way home from work as they were during the morning commute.

Wachs (1987) cites the history of differences between women and men with respect to transportation. Pas (1984) found that gender was a significant factor in daily travel-activity behavior. Using the National Personal Transportation Survyes (NPTS) surveys of 1977 and 1983-4, Gordon, Kumar, and Richardson (1987) confirmed gender differences found by earlier researchers. They found that the growth of trip-making was greater in females than males, reaching 46 percent for married nonworking wives without children and 32 percent for married nonworking mothers. From their analysis, they found that women make shorter commute trips and make more nonwork trips. Additionally, women often make shorter trips in order to economize on that particular trip, given the need to make more nonwork trips. Rosenbloom (1987) and Rosenbloom and Burns (1994), using the NPTS data and numerous local studies, have documented that elder- and child care and household responsibilities often performed by women would to need to make interconnected travel decisions involving employment and child- and elder care locations.

Pazy et al. (1996), recognizing the changing status of women's careers and its influence on commuting patterns of women. They found that in their sample, single women made shorter commutes than married women and had shorter commute times. These differences were attributed to the increased residential mobility of single women. Single women were more likely to live in the central city, while married women were domiciled in the suburbs. Hamed and Mannering (1993) and Bhat (1997) found that males were more likely to go directly home after work than females because females would often engage in such activities as shopping, personal business and recreation after work.

The review in the preceding parts of this section has focused on the increase of VMT based on gender and its antecedents. Despite these noticeable differences by gender, there is empirical evidence concerning slower projections for the VMT. Since 1977, the population of the United States has increased by 30 percent while total person VMT for daily travel has increased by 151 percent. Thus, much of the growth of the VMT can be attributed to such factors as the population age profile, auto availability, licensure rates, size of household, male and female labor participation rates, real income per capital, and land use patterns (Polzin, Chu, and Toole-Holt (2003).

Section 3: The Statistical Methodology:

Most economic analyses typically estimate models under strong parametric assumptions. In fact, the true functional form of the model is rarely known and often their misspecification of an underlying functional form to the data could lead to misleading results. Suppose there are n independent and identically distributed observations (y_i, x_i,….,x_{in}) for all i= 1,……,n from a continuous distribution with a density function of the form f(y, x,…….x_n) =f(x,y). In the latter generalized density function, y is the dependent variable and x is the vector of the regressors. Also if E|y|<∞, then the regression function B(x) =E(y|x) which is the formulation of the regression model

(1) $$Y=B(x)+e$$

where e is the error term which satisfies the condition E(e|x)=0. The latter condition states the regressors and the error term are independent of each other, and uses ordinary least squares (OLS) in the estimation process.

The estimation of (1) by OLS assumes that the data are being fitted to some specified functional form. As an alternative, (1) can also be specified by nonparametric estimation without reference to a functional form because the parametric form may be too restrictive to fit any unexpected features of the underlying the data whereas the nonparametric methods provides an alternative method for assessing relationships of the variables.[2] Nonparametric estimation was first conceptualized by Rosenblatt (1956), and a kernel estimator is used in nonparametric estimation. A kernel is nothing more than a weighted contribution of observations in a h-neighborhood of each observation, x and h controls the amount of smoothing around each datum. Put in another way, kernel estimator defines a set of weights at each observation to produce an estimate. Nadaraya (1964) and Watson (1964) or the Nadaraya-Watson kernel estimator proposed a consistent nonparametric estimator of the form

(2) $$B_n(x) = \sum_{i=1}^{n} y_i R(x)$$

From (2) R(x) is the kernel estimator which can be given as follows:

(3) $$R(x) = k\left(\frac{x_i - x}{h}\right) \Bigg/ \sum_{i=1}^{n} K\left(\frac{x_i - x}{h}\right)$$

where K is the kernel estimator which determines the shapes of the “bumps” and h determines the width. From (3), the partial derivative of the conditional mean at point x is calculated by taking the partial derivative of $B_n(x)$ with respect to x_i or

(4) $$\frac{\partial B_n(x)}{\partial x_i} = \sum_{i=1}^{n} y_i R'(x)$$

The partial derivative in (4) measures the change in the slope of B(x). In equation (3), h is the smoothing or the bandwidth value, and the selection of h requires the value to be optimal in order to fit the data properly. That is, the original plot of y_i and x_i may appear to be “bumpy.” Thus, it becomes necessary to pool the x_i’s in order to smooth out the y’s by taking a weighted average. By taking the weighted average or kernel smoothing, the greatest weight is placed on the concentration of the observations and lesser weight is placed on the lesser concentration of the observations. The latter procedure provides a smoothing of the data. For more details concerning smoothing, the reader should consult Ullah (1988), Silverman (1986), Härdle (1990), and Yatchew (1998). Racine (1998) and Stone (1974) use the leave-one-out cross validation method to smooth the data. This procedure entails the leaving out the i^{th} observation and trying different values for smoothing parameters. This procedure continues until the optimal smoothing parameter is obtained. Then, the nonparametric regression occurs via use of bootstrapping methods which generates the conditional mean, partial derivatives, and the standard errors for each observation point.

Since we are assessing the differences between the VMT distributions by gender, it would be of value to assess the equality of the two distributions. The test of equality proposed by Li and Racine (2003) provides for an empirical assessment of the differences between the two distributions. Suppose there are n_1 observations of X from an unknown distribution f(x) and there are n_2 observations of Y from an unknown distribution f(y). Then, the test statistic can be developed based on the integrated squared density difference given by

(5) $$I = \int \left[f(x) - g(x) \right]^2 dx = \int \left[f(x)dF(x) + g(x)dG(x) - 2f(x)dG(x) \right]$$

F(•) and G(•) represent the cumulative distributions for X and Y, respectively and

(6) $$\int dx = \sum_{x^d \in S^d} \int dx^c$$

Without showing the algebraic derivations, f(•) and g(•) are replaced by their kernel estimates and F(•) and G(•) are replaced by their empirical distribution function, the following statistic is derived:

(7) $$I_n = \frac{1}{n} \sum_{i=1}^{n_1} f(x_i) + \frac{1}{n_2} \sum_{i=1}^{n_2} g(y_i) - \frac{2}{n_2} \sum_{i=1}^{n_2} f(y_i)$$

Equation (7) represents the nonparmetric version of the test statistic. To facilitate a transition from the nonparametric to the parametric estimation, equation (7) in parametric form can be written as

(8) $$T_n = (n_1, n_2 h_1, \ldots .h_q)^{1/2} \frac{I_n - C_n}{\sigma} \rightarrow N(0,1)$$

Intuitively, this empirical implementation of this test statistic, namely (8) will be explained as follows. Randomly select n_1 and n_2 observations separately from the pooled sample of the two distributions. Then compute the test statistic given in equation (8) and apply the procedure N times and use the empirical distribution of the N bootstrap statistics to approximate the null distribution of T_n.

A generalized nonparametric specification of the model can be given as

(9) $$y_i = B(x_1, \ldots\ldots, x_p) + e_i = E(y_i \mid x_i, \ldots., x_p) + e_i$$

Now significance testing can also be applied to the nonparametric model. The nonparametric estimation methods yield partial derivatives which would vary over its domain. That is, the partial derivatives may only be significant over a certain range and insignificant over the rest of the range. This is in contrast to the parametric estimation where the partial derivative is typically restricted to be constant over its domain. Tests for significance are formulated to detect whether a partial derivative equals zero over the entire domain of each variable specified in (9). A test for this purpose has a null hypothesis stated in terms of the vector of partial derivatives of the conditional mean as:

$$H_O : \lambda = E_\eta \left[\frac{\partial E(Y \mid X)}{\partial X_j} \right]^2 = 0 \tag{10}$$
$$H_A : \lambda = E_\eta \left[\frac{\partial E(Y \mid X)}{\partial X_j} \right]^2 > 0$$

where η denotes a unit vector of length j and $[\partial E(Y|X)/\partial X]^2$ represents a vector of the squared derivatives. Given (10), if the null hypothesis is true, then λ is equal to zero. Otherwise, λ will exceed zero. The actual test statistic for the test is given as

$$\lambda = n^1 \sum_{i=1}^{n} \sum_{h=1}^{p} \left[\frac{\beta_h(x_i)}{SE(\beta_h(x_i))} \right] \tag{11}$$

(11) divides the partial derivative estimate, $\beta_h(x_i)$ by its asymptotic standard error in the denominator. Finally, the use of (10) entails the use of bootstrap methods to obtain its sampling distribution. Under the null hypothesis, the test statistic and the critical values are also calculated. For greater analytical details of this test statistic, the reader should consult Racine (1997).

Section 4: Data Sources and The Empirical Results

The data for this analysis is derived from the 2001 National Household Travel Survey (NHTS) which is a compilation of daily and long-distance travel, and the NHTS includes demographic characteristics of households, people, vehicles, and detailed information on daily and longer-distance travel by all modes. VMT is the movement of a privately owned vehicle for one mile regardless of the number of people in the vehicle. The other variables to be used in the analysis are as follows. The size of the household would be used since the number of people in a household may require additional trips performing household services. The number of work trips would also be incorporated into the analysis. Since the number of worktrips just contain worktrips, it may not capture the addition of trip chaining , so the variable great circle distance to work was used a proxy.

Tests of Equality Between the VMT by Gender

For the tests of equality between VMT between men and women, we apply the testing procedure as given by Li and Racine (2003). The resulting test statistic is T = 19.729 with the 90th, 95th, and 99th percentiles under the null being 0.7112, 0.0001, and .001, respectively.[3] The null of equality of male and female VMT distributions is rejected; additionally, the resampled percentiles indicate that the normal distribution provides a poor approximation to the finite-sample null distribution even for a fairly large pooled sample of the datasets for men and women.

Results of the Tests of Significance

Finally, the test of significance was applied to the data. The test of one joint test for the data or $E[Y|X] \perp X_1\ X_2\ X_3$ and the bandwidths for the testing process were chosen via leave-one-out cross validation. The results of the testing process are summarized in Table 1:

Variable	t- statistic	$t^{critical}$	p-value
Male			
X_1 X_2 X_3	.757	.4512	.01
Female			
X_1 X_2 X_3	.785	.6204	.01

The results for both male and female indicate a rejection of the joint variables (X_1 X_2 X_3) which are independent of the conditional mean Y.

Conclusions and Final Thoughts:

VMT has been widely used for various analyses related to transportation issues. A typical modeling approach has been the use of parametric regression methods which may be too restrictive since there could be underlying variations in the data which cannot be captured by the parametric methods. As an alternative, nonparametric methods were utilized to assess the differences between the VMT patterns by gender, using the 2001 NHTS data. The results have indicated that there are differences as given by the results of the kernel regression analyses between males and females. The kernel regression utilized VMT as the dependent with three regressors: the size of the household; the number of work trips; and the great circle distance to work. [4] Also the goodness of fit metrics from the kernel regression methods were improved over the parametric methods metrics of goodness of fit. Additionally, the primary limitation of these kernel regressions is that it is more difficult to appreciate in more than three dimensions or having more than two regressors in the analysis. Another issue concerning these kernel methods is the application of these methods to large datasets requires more computing time than the standard parametric methods. However, the results of such methods provide intuitive insight into the behaviors of the underlying data.

ENDNOTES

[1] In the sociology literature, this was translated into as a backlash against the feminist movement that this movement only resulted in more work. That is, women engaged more in paid jobs outside the home and return home to their "second" job of engaging in domestic labor. As this shift occurred for women, it did not result in an equivalent shift for men to engage more in domestic labor (Hochschild 1989).

[2] The simplest form of nonparametric or density estimation is a histogram as encountered in elementary statistics. The histogram is defined as $f(B_k)$ =1/nh`where i=1,…,n where B_k represents the kth bin. Think of each bar in the histogram as a bin. Also h is the bandwidth or the smoothing parameter which controls the amount of smoothing to the given data.

[3] Also the p-value was calculated for this test statistic. The p-value for this test statistic is <.001 which also indicates that the null hypothesis was rejected.

[4] The multidimensional graphs of the regression results were not provided because of space limitations. Also the results of the parametric methods were also not provided.

References

Adler, T. and M. Ben-Akiva. (1979). "A Theoretical and Empirical Model of Trip Chaining Behavior." *Transportation Research (B)*, 13B, 243-257.

Bhat, C. (1997). "Work Travel Mode Choice and Number of Nonwork Commute Stops." *Transportation Research.*

Becker, G.S. (1991). *A Treatise of the Family*. Cambridge, Massachusetts: Harvard University Press.

Blau, F.D. and M.A. Ferber (1986). *The Economics of Women, Men, and Work*. Englewood Cliffs, New Jersey: Prentice-Hall.

Davidson, D. (1991). "Impact of Suburban Employee Trip Chaining on Transportation Demand Management." *Transportation Research Record,* 1321: 82-89.

Downs, Anthony. (1992). *Stuck in Traffic: Coping with Peak-Hour Traffic Congestion*. Washington, D.C.: The Brookings Institution.

Golob, T. F. (1986). "A Nonlinear Canonical Correlation Analysis of Weekly Trip Chaining Behavior." *Transportation Research (A*), 20A, 385-399.

Gordon, P., A. Kumar and H.W. Richardson. (1988). "Beyond the Journey to Work." *Transportation Research (A),* 22A, 416-426.

Goulisa, K.G. and R. Kitamura. (1989). "Recursive Model System for Trip Generation and Trip Chaining." *Transportation Research Record*, 1236, 59-66.

Hamed, M.M. and F. Mannering. (1993). "Modeling Travelers' Postwork Activity Involvement: Toward A New Methodology." *Transportation Science*. 27(4), 381-394.

Hanson, S. and I. Johnston (1985). "Gender Differences in Work Trip Lengths: Implications and Explanations." *Urban Geography*, 6(3), 193-219.

Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge: Cambridge University Press.

Hochschild, A. (1989). *The Second Shift*. New York: Avon Books.
Kitamura, R. (1984). "A Model of Daily Time Allocation to Discretionary Out-of-Home Activities and Trips." *Transportation Research (B)*, 18B, 255-266.

Li, Q. and Racine, J. (2003), A nonparametric test for equality of distributions with mixed categorical and continuous data, Technical report, Syracuse University, Syracuse, NY USA.

Nadaraya, E. A. (1964). On estimating regression, *Theory Prob. Appl.* 10: 186-90.

Nishii, K. K. Kondo, and R. Kitamura. (1988). "Empirical Analysis of Trip Chaining Behavior." *Transportation Research Record*, 1203, 48-59.

Oster, C.V. (1978). "Household Trip-making to Multiple Destinations: The Overlooked Urban Travel Pattern." *Traffic Quarterly*, 32, 511-529.

Oster, C.V. (1979). "Second Role of the Work Trip- Visiting Nonwork Destinations." *Transportation Research Record*, 728, 79-89.

Pazy, A., I. Salomon, and T. Pintzov. (1996). "The Impacts of Women's Careers on their Commuting Behavior: A Case Study of Israeli Computer Professionals." *Transportation Research (A)*, 30(4), 269-386.

Polzin, S.E., X. Chu, and L. Toole-Holt. (2003). A Policy Paper based on the Florida Department of Transportation Trends and Conditions Research. Center for Urban Transportation Research, University of South Florida, September 2003.

Racine, J. (1997), `Consistent significance testing for nonparametric regression', *Journal of Business and Economic Statistics* 15(3), 369.379.

Racine, J. (1998), `Nonparametric maximum likelihood estimation and inference', *Working Paper* .

Richardson, H.W. and P. Gordon. (1989). "Counting Nonwork Trips." *Urban Land*. 48(9), 6-12.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density functions, *Annals of Mathematical Statistics* 27: 642-69.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Stone, C. J. (1974), `Cross-validatory choice and assessment of statistical predictions (with discussion)', *Journal of the Royal Statistical Society* 36, 111.147.

Strathman, J.G., K.J. Dueker and J.S. Davis. (1993). Congestion Management: Travel Behavior and the Use of Impact Fees, Volume I: Effects of Household Structure and Selected Travel Characteristics on Trip Chaining . Seattle, WA: Transportation Northwest (TransNow).

Ullah, A. (1988). "Nonparametric Estimation of Econometric Parameters," *Journal of Quantitative Economics*.

Wachs, M. (1987). "Men, Women, and Wheels: The Historical Basis of Sex Differences in Travel Patterns. *Transportation Research Record*, 1135, 10-16.

Watson, G. S. (1964). Smooth regression analysis, *Sankhy, Series A* 26: 359-72.

Yatchew, A. (1998). Nonparametric Regression Techniques in Economics, *Journal of Economic Literature*, 36, 669-721.