Online Supplement to the Paper
"Quantifying the Relative Contribution of Factors to Household Vehicle Miles of Travel"

Abhilash C. Singh, Sebastian Astroza, Venu M. Garikapati, Ram M. Pendyala,
Chandra R. Bhat (*corresponding author*), and Patricia L. Mokhtarian

Reasons for Using Density as the Sole Built Environment (BE) variable

At the outset, we should state that the consideration of more BE variables is a key aspect that is worthy of further exploration in future research efforts. In this study, there are three key reasons for the use of density as the sole measure of built environment. We state these reasons below. For ease, we take the case of why we are, for example, not including "distance to CBD" as an additional land-use variable, as has been done in some previous studies.

First, density is a commonly used quantitative metric of the built environment. It is easily measured, quantified, calculated, and understood. Many other measures such as walkability index, pedestrian friendliness, bicycle level of service, land use diversity and variety, and proximity to transit infrastructure can also be useful measures of the built environment. However, although these attributes have been defined and used in the literature, some ambiguity remains in their definition and what exactly they represent. Besides, there is ample evidence that the many measures of land-use are highly correlated with density, which is also easy to quantify. As Brownstone (2009) stated about a decade back in a major special Transportation Research Board report "There are potentially many aspects of the built environment that could affect households' travel behavior. Naturally research has concentrated on those aspects that are easy to measure. Since most measures of the built environment are highly correlated, it may only be necessary to include a few key characteristics to capture the effects. Most national level studies only use residential and/or employment density since these are the easiest to obtain." Kim and Brownstone (2013) further state that "…..land use density is highly correlated with almost all measures of urban sprawl", referencing also Badoe and Miller (2000). Indeed, many recent studies continue to use density as the sole indicator of the built environment because of the ease of computing density and the high correlation with other built environment measures, so we are certainly not the only ones to do so (see, for example, Kim and Brownstone, 2013; Brownstone and Fang, 2014; Paleti et al., 2013; Cao and Fan, 2012; and Bhat et al., 2016). Besides, in our analysis, when we added in "distance to CBD", where the CBD was defined as the zone with the highest continuous value of density (this zone came out to be zone ID 124, bounded by West 46th street to the north, West 42nd street to the south, 7th avenue to the west, and 6th avenue to the east; Time Square is just outside this zone, less than 100 feet from the northwest tip of this CBD zone), there was only a relatively marginal improvement in data fit after including density. To confirm that this was because of the high correlation between density and "distance to CBD", we examined the correlation coefficient between these two variables in continuous form, which came out to be very close to -0.8. That is, zones that are low density are, in general, far away from the "CBD" zone, and zones with high density are, in general, close to the "CBD" zone. To further reinforce this, we examined the average distance to CBD across the records of households within the three discrete density categories of low, medium, and high density. Within those residing in the low density zones, the distance to CBD was 53 miles. Within the medium density zones, the corresponding value was 25 miles, and within the high density zones, the value was 9 miles. The strong negative correlation is readily

apparent again. As another data point, the top ten zones with the highest continuous density are all in the immediate vicinity of the CBD zone, less than a mile away.

A second reason for not including the "distance to the CBD" variable is that plenty of recent studies show that the notion of major cities being mono-centric is not valid anymore, so that the concept of "distance from CBD" becomes fuzzy. Indeed, in the specific context of New York City, Cox (2010), while acknowledging that New York City is one of the most centralized large urban areas in the world (with midtown Manhattan in and around Time Square being what may be traditionally called as the CBD point in the city), indicates that the city, in reality, is "a highly decentralized metropolitan area". Notwithstanding the clustering of high density zones in and around time square (as indicated in the earlier paragraph), approximately, 74% of the employment in the New York metro area is outside Manhattan, according to the 2010 US Census Bureau work flow tables. Indeed, the inner ring suburbs of New York City (outside the five boroughs of New York City) have more employment (28%) than Manhattan (26%). The four boroughs of New York City (Brooklyn, Queens, The Bronx, and Staten Island) hold 22% of the employment, and the outer ring suburbs hold the remaining 24% of the jobs. At the same time, a relatively small fraction of workers (35%) in the four boroughs other than Manhattan work in Manhattan, and this percentage falls rapidly down to 14% of the inner ring suburban workers who work in Manhattan and 6% of the outer ring suburban workers who work in Manhattan. Basically, there is a trend of working more closely to home, rather than the commute into Manhattan implied by assuming a monocentric city. In short, as Cox indicates, there is a "dominantly local nature of commuting in New York…..All of this is a huge change from a half-century ago….New York City has moved from virtual monocentrism to the Edge Cities polycentrism and increasingly even to an amorphous Edgeless City employment dispersion". In fact, employment in Manhattan has dropped substantially, and risen elsewhere, including in Queens and Staten Island, where both employment and population growth has been very strong in the recent past. Besides, as New York University urban scholars Solly Angel and Patrick Lamson-Hall show in a two minute animation video of how Manhattan area densities have changed over a 210 year period from 1800 to 2010 (see https://www.citylab.com/equity/2015/06/watch-210-years-of-manhattan-densification-in-2-minutes/394736/), the densities have tapered off by about 40% since 1910 and are now much more evenly spread all through Manhattan (as opposed to, during the 1800s and 1900s, the substantial density peaking patterns within the lower east side of Manhattan first and then midtown Manhattan later). Also, as stated by Cox, these researchers did find the outward movement of residences and employers to the other four boroughs of New York and to the inner and outer ring suburbs. Interestingly, the above comments about monocentric metro regions transforming into polycentric or even edgeless cities is much more universal than for New York City alone, as documented by Angel and Blei (2016), Hu et al. (2018), and Zhong et al. (2017).

Third, and as we indicated in our earlier rebuttal, there remains a methodological challenge that is worthy of additional research. In the model system considered in the paper, the residential location choice alternatives are defined by "density"; and density variables are, in turn, incorporated in the VMT equation (continuous dependent variable) to capture built environment effects. By doing so, it is possible to explicitly and easily tease out residential sorting effects (represented by error covariances) from built environment effects. If additional built environment attributes were included in the VMT equation specification (such as distance to CBD, land-use mix, network attributes, etc.), the correlation between "density" and these additional built environment measures

would render it difficult to cleanly separate residential self-selection effects from true built environment effects. Additional research and development efforts on the methodological front are needed to address this issue, namely, tease out residential self-selection and built environment effects separately in the presence of multicollinearity. In short, there is no reason to believe that there will be no self-selection in the other built environment variables just like we have considered the self-selection in density at the residential location. Certainly, it is possible that people who are loath to driving and are more walk and bicycle friendly will live closer to their work places and also have lower VMT. That is, "distance to CBD" could well be endogenous to VMT choices. Given the methodological challenges involved in such a multi-dimensional endogenous modeling set-up, and our desire to separate residential self-selection effects from true built environment effects, the authors exclusively focused on density as the measure of built environment – both to characterize residential location choice and as explanatory variables in the VMT equation.

Having said all of the above, the point made by the reviewer remains a valid one. In the absence of consideration of any other built environment attributes, is it possible that this study is under-estimating the true built environment effects? Is some of the built environment effect captured in the "unexplained" portion of the variance? These are important questions that are worthy of future research, and the paper includes a paragraph about the limitations of the study and directions for future research.

REFERENCES

Angel, S., and Blei, A.M. (2016) The spatial structure of American cities: The great majority of workplaces are no longer in CBDs, employment sub-centers, or live-work communities, *Cities*, 51, 21-35.

Badoe, D. A., and Miller, E. J. (2000) Transportation-land-use interaction: Empirical findings in North America, and the implications for modeling. *Transportation Research D*, 5(4), 235–263.

Bhat, C.R., S. Astroza, A.C. Bhat, and K. Nagel (2016) Incorporating a Multiple Discrete-Continuous Outcome in the Generalized Heterogeneous Data Model: Application to Residential Self-Selection Effects Analysis in an Activity Time-use Behavior Model, *Transportation Research Part B*, 91, 52-76

Brownstone, D. (2009) Key Relationships Between the Built Environment and VMT. Transportation Research Board Special Report 298 on Driving and the built environment: The effects of compact development on motorized travel, energy use, and CO2 emissions.

Brownstone, D., and Fang, H. (2014) A vehicle ownership and utilization choice model with endogenous residential density. *Journal of Transport and Land Use*, 7(2), 135-151.

Cao, X. and Y. Fan (2012) Exploring the Influences of Density on Travel Behavior Using Propensity Score Matching. *Environment and Planning B*, 39(3), 459-470.

Hu, L., Sun, T., and Wang, L. (2018) Evolving urban spatial structure and commuting patterns: A case study of Beijing, China. *Transportation Research Part D*, 59, 11-22.

Kim, J., and Brownstone, D. (2013) The impact of residential density on vehicle usage and fuel consumption: Evidence from national samples. *Energy Economics*, 40, 196-206.

Paleti, R., C.R. Bhat, R.M. Pendyala, and K.G. Goulias (2013) Modeling of Household Vehicle Type Choice Accommodating Spatial Dependence Effects. *Transportation Research Record: Journal of the Transportation Research Board*, 2343, 86-94.

Zhong, C., Schläpfer, M., Müller Arisona, S., Batty, M., Ratti, C., and Schmitt, G. (2017) Revealing centrality in the spatial structure of cities from human activity patterns. *Urban Studies*, 54(2), 437-455.