Predicting the Fourth Industrial Revolution and Subsequent Changes in the Urban Environment and Human Behavior: A Review on the Validity of Research about the Built Environment–Travel Behavior Relationship*

Tae-Hyoung Tommy Gim**

Abstract: The Fourth Industrial Revolution is expected to greatly influence green space, leisure, land use, transportation use and service, and related residential choice. In the face of the Fourth Industrial Revolution that calls on the construction of a new environmental policy and management system and the prediction of changes in the urban environment and human behavior, the purpose of this study is to examine how to improve research validity for a better construction/prediction and to provide considerations for an appropriate use of big data. Using the case of Google Flu Trends, this study argues that the first limitation of big data is that they draw only correlation, not causality, which increases the chance of the misestimation. As the second limitation of big data, their lack of scientific sampling is discussed, using the cases of a Twitter survey about Hurricane Sandy and Boston’s StreetBump (pothole response) program. Then, this study examines which conditions are required to improve research validity, through an analysis of previous studies on the built environment–travel behavior relationship. As for studies on the relationship, the hottest topic in the current literature is residential self-selection: Individual features work as a confounder that makes the relationship spurious. The spurious relationship accordingly causes the built environment effect to be misestimated. In this sense, considering the four conditions for constructing internal validity or causality -- causal mechanism, covariation, nonspuriousness, and time precedence -- this study critically evaluates the methods of previous studies, including regression–based approaches such as OLS and 2SLS regression, longitudinal design including panel analysis, quasi–longitudinal analysis and recursive and nonrecursive structural equation modeling, and consonant–dissonant matching. Also investigated are two recent models that actively measure the level of residential self–selection, propensity score matching and

* This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT [grant number 2016R1C1B1015545].
** Assistant Professor, Graduate School of Environmental Studies, Interdisciplinary Program in Landscape Architecture, and Environmental Planning Institute, Seoul National University
sample selection model. As a result, previous studies turn out to have delivered mixed results regardless of methodology. Specifically, they are not in agreement regarding the magnitude of the confounding effect of residential self-selection. This study ends by highlighting the fact that as a validity threat, selection bias lowers not only internal validity, but also external validity through its interaction with an explanatory variable, suggesting further research on the generalizability of analytical results.

**Key Words:** Fourth Industrial Revolution, Big Data, Validity, Built Environment, Travel Behavior

### 1. Introduction

As one of the hottest topics in academia nowadays, the Fourth Industrial Revolution has been proposed in January 2016 at the World Economic Forum (also called Davos Forum) by its president, Klaus Schwab. The Fourth Revolution is featured by the rapid development in all fields of information and communication technology (ICT) and especially by their convergence and integration. Among others, the revolution is led by artificial intelligence (AI) and robot, autonomous (also connected and electric) vehicle, big data, clouding and internet of things (IoT), 3D printing, and nano- and bio-technology. It is expected to influence people more rapidly and more widely than the previous three revolutions.

At this juncture, the Korean Ministry of the Environment (MOE) launched “Smart Environmental Management Task Force” this March in order to explore environmental policies that are supported by cutting-edge intelligent technologies, as would be made possible owing to the emergence of the Fourth Revolution. In July, the MOE held the “Fourth Industrial Revolution and the Environment” conference to explore promising projects and collect expert and public opinions. At the conference, the task force argued on the potential for precautionary smart environmental management and presented policy approaches and
key projects through a review of domestic and international policy cases and technical trends. A total of four policy approaches were adopted: development of a platform for creating new data value; revolution of environmental surveillance, monitoring, prediction, and management; provision of smart environmental service; and support of the revolution of environmental technology and industry ecosystem. Also, among the key projects, the task force identified the construction of the “smart air pollutants management and forecasting system” as the first project to be supported by AI and IoT. Other projects with potential include (1) advancing the “environmental pollution monitoring system” through drone, IoT and AI technologies, (2) constructing the “integrated platform for environmental big data” for the purpose of storing environmental information in the form of cloud-based big data, and (3) establishing the basis of the “individual-level real-time system for the environmental pollution exposure assessment” through the combination of the air- and water-quality environmental statistics and individuals’ health and activity information.

According to Klaus Schwab, the four Industrial Revolutions are all distinct: The first is featured by the mechanization of the production process based on water and steam power, the second by the construction of mass production systems through electric power, and the third by the automation of the production process as supported by electronic engineering and IT as well as by the expansion of the Internet. Finally, the Fourth Revolution will be led by the convergence of cutting-edge technologies that cross the boarders of AI and other ICT fields.

The features of the four revolutions have exerted substantial influences on cities in relation to their functions, which were subsequently reflected in their spatial structures. In agricultural
societies before the First Revolution, cities functioned as trading centers and they were traditionally surrounded by walls and formed centered on spatially well-defined markets. Then, in a well-suited form of expediting the movement of goods, harbor cities began to emerge and they became a hub of transportation networks. In line with the First Industrial Revolution, mass production companies were built on urban fringes, and this formed manufacturing cities that were supported by railroads for delivering mass-produced commodities. Then, in the period of the Second Revolution, industrial restructuring facilitated the third (service) industry and the expansion of automobiles and highways brought about rapid suburbanization. The Third Revolution reinforced the scale of the urbanization and generated IT cities and knowledge-centered edge cities. Finally, the Fourth Revolution is concerned with AI-supported road diets and subsequent green space expansion, better care for the quality of life and accessibility to leisure activities. Therefore, the Fourth Revolution is expected to result in innovative changes of the current commute/business/commerce-centered urban structures.

Particularly, urban residents in the era of the Fourth Industrial Revolution would not drive automobiles; instead, they may use public transit or autonomous vehicles. Also, in terms of automobile ownership, they will rely on car sharing service according to the advancement of IT and Internet technologies. This means that people will return from the suburbia to the central city in which the technologies are well provided. (This concentration to the city might lead to the “counter”-donut effect, which implies a populated center with an empty suburban ring.) Also, these people may highly value leisure activities and deeply care for the ecological environment. Therefore, along with autonomous vehicles and
connected cars, the improvement of environmental concerns and expansion of renewable energy will make electric vehicles become an important axis of the future transportation systems.

Meanwhile, in light of the report of the Davos Forum in January 2016, “The Future of Jobs”, the built environment and travel behavior are expected to change due to “changing nature of work, flexible work”, which was identified as the first of various socioeconomic drivers of the Fourth Revolution. (When the Revolution is discussed in urban studies, the main focus is on the emergence of AI- and big data-supported smart cities and how it changes the built environment. This study intentionally excludes smart city-related descriptions that have duly been provided in other studies.)

The promotion of flexible working and telecommuting systems will increase the number of those who are working at home or at public transit hubs near home, which in fact indicates the advancement of TOD’s (transit-oriented developments). Meanwhile, in addition to automobile alternatives (public transit and nonmotorized modes), the full delivery of autonomous and connected vehicles will change the meaning of travel behavior. That is, in the vehicles, people do not have to be solely involved in driving but instead can benefit from a variety of smart tool- and Internet-connected services. As such, travel will be no longer considered sheer disutility, so transportation theories/models that assume about rational humans who minimize the disutility of travel (e.g., four-step travel demand model) may lose its validity.

In line with the Fourth Revolution, the expansion of car sharing services will reduce registered automobiles and parking spaces. Also, the optimized algorithm of autonomous vehicles will lessen traffic volumes. Then, governments will consider road diets. This may in turn lead to a
virtuous circle of reduced automobile ownership and lessened travel distances/traffic volumes. (Once the issue of less optimal automobile use is addressed -- automobiles are used for a short period of time at similar hours and parked for the most time of the day -- only 10% of the current number of registered vehicles is predicted to fully satisfy all travel demands.) Recovered land spaces that are generated by road diets would turn to green spaces owing to the public’s higher concerns/needs for leisure facilities and natural settings.

Interestingly, the urban built environment led by the technologies of the Fourth Revolution is very similar to that of pre–industrial European cities. In a small-size, well-defined areas, a large population will efficiently live and easily walk or use smart modes of transportation. They all facilitate the formation of the community and social capital through better communication among residents. These outcomes are consistent with the goals of the neo–traditional development (NTD), compact city, and smart growth.

However, it is not easy at all to predict how the compact built environment will change human behavior. It is particularly so if we consider the mixed outcomes of empirical studies on the built environment–travel behavior relationship (Crane, 2000; Hall, 2001) as have long been analyzed since the 1950s (e.g., Adams, 1959; Hamburg, 1958; Levinson and Wynn, 1963; Mitchell and Rapkin, 1954) in which the Second Revolution accompanied by the expansion of highways and suburban developments aggravated automobile dependence. In this sense, the first purpose of this study is to identify reasons for the inconsistent results of previous studies in regard to their research validity (particularly, internal and external validity) and accordingly, to propose valid methodology as a way of predicting changes in the future built
environment and travel patterns.

While the Fourth Revolution is centered on data innovations (e.g., AI, IoT, and big data), the second purpose is to deal with the issue of data misuse. In the “Fourth Industrial Revolution and the Environment” conference in July, the MOE chose the construction of the “smart air pollutants management and forecasting system” as the first among other key projects. They subsequently have showcased the speed and accuracy of the Particulate Matter Forecasting and Alert System. By contrast, this study will review a similar case and explain that without methodological (research) validity, the system would be at the even greater risk of incorrect forecasting (i.e., greater estimation error).

II. Issues with Big Data

Although the term big data has no standardized definition, it is commonly summarized with 3V’s: Volume, Velocity, and Variety (Chen, Mao and Liu, 2014) (also, the so-called wanna-V’s may or may not be added: Variability, Veracity, Validity, and Value). Volume indicates that the size of big data often amounts to some terabytes or petabytes, and they cannot be easily handled by traditional database tools to the degree to which a new method is required for the search, collection, storage, processing, analysis, and presentation of big data. Velocity suggests that data are captured at a very high speed and also should be processed very quickly. The last, Variety, means that data are collected from various sources and in different formats.

The value of big data is in the possibility of detecting a relationship that can be ignored in an analysis of small-size data. However, as
criticized in 2014 in a New York Times article, “Eight (No, Nine!) Problems with Big Data”, only correlations are detected with big data: The data cannot be used to present which relationships are meaningful (i.e., a causal inference is impossible with big data). Another limitation is that as stated above as their strength, big data often produce too many significant correlations (i.e., the bigger the sample size, the smaller the standard error, leading to a deflated p-value). That is, statistically, big data carries the risk of false positive error (i.e., Type I error).

The second problem of big data is that they have inherent bias due to a lack of scientific sampling. An example of the unsuccessful use of big data is that when people’s responses to Hurricane Sandy were surveyed via Twitter, data were collected mainly in Manhattan (Martín, Li and Cutter, 2017). It is because this area has a high density of affluent residents who can afford to tweet and produce data. By contrast, those areas that have been the most severely affected (Breezy Point, Coney Island, and Rockaway) produced virtually no data. First, it is owing to a lower smartphone penetration rate in the areas. Second, although the fewer smartphone users attempted to tweet, the areas were the first in which electricity and cellular networks were shut down and thus, data production itself was impossible.

Another typical case regarding the sampling issue of big data is Boston’s StreetBump program (O’Leary, 2013). Boston has to deal with about 20,000 potholes per annum, and to efficiently allocate its resources, the City of Boston introduced a smartphone app called StreetBump. The app uses accelerometer and GPS data to passively detect a pothole and immediately report the case to the city. However,
this smart move made the resources to be biased to areas with more smartphone owners (e.g., areas that the youth frequently visit). In contrast, those who are in even greater need of the city’s resources, that is, low-income people and the low-income elderly in particular, often do without smartphone; hence, they were rather excluded from the city service.

Probably the most well-known example of the unsuccessful use of big data -- it also has a meaningful implication for the MOE’s Particulate Matter Forecasting and Alert System -- is Google Flu Trends (GFT). This Google search-based program detects correlations between terms searched for by people online and whether they have influenza symptoms in order for better outbreak tracking/monitoring. Google researchers introduced the functionality of GFT through the journal Nature (Ginsberg et al., 2009). According to the researchers, GFT was capable of tracking the diffusion of influenza across the U.S. without a single medical checkup. The tracking result was delivered much earlier than that of the U.S. Centers for Disease Control and Prevention (CDC): CDC spent more than a week for the flu diffusion tracking while GFT’s delay was merely one day. The major difference was that the CDC tracking system should receive medical records from doctors unlike the Google search-based GFT.

After a successful launch over the first several years, this theory-free and data-rich GFT model lost its validity. According to a study published in Science (Lazer et al., 2009), which argued against the Nature study, GFT excessively overestimated the maximum flu level, except the very first year. By observing 108-week cases starting from August 2011, the study found that for 100 weeks, the flu occurrences were overestimated and in some cases, the estimate was about double
the CDC’s real data. The misestimation was mainly attributed to the fact that GFT detected correlation with its algorithm, but not causality.

At this juncture, this study aims to examine what components are required in big data studies to build causality to secure research validity. As such, a contribution of this study is to provide a direction for studies that predict changes in the built environment and human behavior as led by the Fourth Industrial Revolution.

III. The Relationship between the Built Environment and Behavior: Urban Sprawl and Compact City

How the built environment would affect behavior has been studied in relation to travel patterns in the field of urban and transportation planning/engineering in which the built environment was studied in relation to the compact city/urban sprawl concept. Here, the compact city is a synonym of the above-mentioned traditional or neo-traditional development (NTD) and smart growth and also, it has quite a few components in common with TOD and new urbanism (Kim and Moon, 2011; Won, 2014).

As opposed to the sprawling city, the compact city is equipped with higher densities, land use mix, better connectivity, and pedestrian-friendly street designs (Boarnet and Crane, 2001) and among others, primarily evaluated centered on 3D’s (Cervero and Kockelman, 1997): Density refers to the concentration of the development (e.g., population and housing), Diversity means land use mix and balance, and Design indicates road network and urban design characteristics. These 3D’s increase accessibility, which is the core component of the built
environment (Gim, 2013). Ewing (2008) similarly argued: “Ultimately, what distinguishes sprawl from alternative development patterns is poor accessibility of related land uses to one another” (p.521).

Discussions on the built environment-travel behavior relationship have begun in the mid-1950s (e.g., Adams, 1959; Levinson and Wynn, 1963; Mitchell and Rapkin, 1954), but they have been articulated in a study by Newman and Kenworthy (1989). However, the study has continuously been criticized for its low precision and accuracy -- mainly because the unit of analysis (UOA) was not the individual, but the city on a larger scale, and control variables were not enough -- (Handy, 2005a, 2005b) so the researchers (Kenworthy and Laube, 1999) repeated the study in 1999 by expanding the sample and including more variables. Over again, in a very recent study (Newman and Kenworthy, 2015), they argued that a strong relationship exists between urban density and automobile use. Specifically, with a sample of 58 high income cities, they analyzed the relationship between density and per capita automobile travel and found that statistically, the relationship is strongly negative ($R^2 = 0.8392$). Especially, a modest density increase in the most sprawling areas was found to result in the strongest reduction in automobile travel.

In later empirical studies conducted at the individual level, not the city level, however, the relationship between the built environment and automobile travel reduction was reported to be mixed (Crane, 2000: Hall, 2001). It is particularly because the empirically found magnitude of the built environment effect was not consistent, rather than its significance. For instance, Ewing et al. (2007) delivered the following result in their book, Growing Cooler: “The vast majority of these studies together show significant relationships between development patterns
and travel behavior. Today, only the direction of causality and strength of the effects seems to be seriously debated” (p.65). As a criticism of the book, Handy and Mokhtarian (2008) argued: “Unfortunately, those are exactly the points on which policymakers need answers” (p.95). Especially, the current literature focuses not on the directionality of the causality, which is confirmed by checking temporal precedence of a variable, but rather on its magnitude (Cao, Mokhtarian, and Handy, 2009).

Indeed, discussions on the magnitude deserve to be highlighted because it allows planners who consider land use interventions to better predict the efficacy of a land use plan and policy. From the same perspective, Ewing and Cervero (2010) provided a list of the built environment effects as estimated for varying study areas and similarly, the U.S. Environmental Protection Agency (2013) conducted a comprehensive literature review to quantify the change in travel behavior according to the unit change in the built environment (e.g., addition of one subway station).

As for the estimation of the magnitude of the built environment effect, the hottest topic in the current literature is residential self-selection. If a compact area is selected by those who inherently have attitudinal or sociodemographic features that make people choose to live in the compact environment, the fact that they reduce automobile travel and increase alternative walk, bike, transit travel is not the product of the built environment, but that of the inherent features. If so, the inherent features work as a confounder by which the relationship between the built environment and travel behavior becomes spurious. Due to this spurious relationship, the built environment effect is misestimated. Such a misestimation issue has already been indicated with the above-discussed
case of GFT and this means that the issue may also apply to the case of the MOE’s Particulate Matter Forecasting and Alert System. Therefore, this study investigates residential self-selection (validity and as a validity threat, self-selection) and by reviewing updated discussions on the self-selection, it provides recommendations for more valid research.

IV. Validity

1. Types of Validity

Why studies on the same topic deliver mixed results is attributed to differing levels of validity. Here, validity is the degree to which a proposition, inference, or conclusion is close to the fact, that is, the best available approximation to the truth. Validity is built step-wise: in the order of conclusion, internal, construct, and external validity (see Figure 1). Construct validity is secured if research variables have an actual relationship while internal validity is concerned with whether the relationship is causal. Construct validity refers to the degree to which the causal relationship can be extended to those concepts that the variables were meant to measure (e.g., if population density and land use mix were used to evaluate the compact city level, construct validity confirms that the two indices well represent the compact city concept). Lastly, external validity presents whether the relationship can be transferred more widely to other subjects, places, and times.
2. Internal Validity

1) Selection Bias and Self-Selection

Among the four types of validity, residential self-selection has been considered a major threat to internal validity. Internal validity is defined as the accuracy of the inference on the causality that exists between the supposed explanatory and outcome variables. As a threat to internal validity, the selection (or selectivity) bias occurs when the experimental and control groups are heterogeneous. In social sciences in which air-tight laboratory control is not plausible, the random assignment of the subjects to the two groups are virtually impossible and the assignment is somewhat done at the researcher’s discretion, which makes social science studies often exposed to the bias. Meanwhile, if the selection of the experimental group subjects relies on volunteerism, the selection is particularly called self-selection. As a well-known example, if researchers evaluate the efficacy of a new
diabetes medicine and to this aim, construct an experimental group by recruiting volunteers, they cannot tell whether a reduced blood sugar level is owing to the medicine or the volunteers’ commitment to taking care of their symptom. Thus, the efficacy could be overestimated.

The misestimation by self-selection is because the (unconsidered) characteristics of the UOA affect the explanatory variable and then, the accumulated effect works on the outcome variable (UOA characteristics $\rightarrow$ explanatory variable $\rightarrow$ outcome variable). As such, the explanatory variable becomes non-extraneous, leading to endogeneity bias, which accordingly produces a spurious relationship between the explanatory and outcome variables (explained by the unconsidered variable).

How to control for self-selection includes randomization (random assignment) and matching. The most appropriate way is to randomly assign sampled subjects into the experimental and control groups. (Here, random assignment differs from random sampling, which refers to the random pick-out of a sample -- a group of subjects -- from the entire population.) If randomization is not possible, an alternative is matching. In order to comprise the experimental and control groups, a researcher pairs two subjects based on hypothesis-related variables (e.g., for the gender variable, a pair has two males) and assigns one into the experimental group and the other into the control group. Individual matching examines all characteristics at the same time to detect a pair (e.g., in the case of marriage and gender, four possible categories: unmarried male, unmarried female, married male, and married female), whereas matching by frequency only considers the overall composition (e.g., in the case of gender, only the gender ratios of the experimental and control groups are checked) for convenience purposes.
2) Residential Self-Selection

Among different types of self-selection, people’s self-selection into residential areas is particularly called residential self-selection. This concept originated in the mid-1950s in political economics. Tiebout (1956) examined the optimal level of the revenue and expenditure of the local government and argued that “[i]f consumer-voters are fully mobile, the appropriate local governments … are adopted by the consumer-voters” (p.424), that is, ultimately, people self-select into a residential area according to their internal characteristics. Gim (2013) detailed residential self-selection regarding the built environment–travel behavior relationship by presenting that “[p]eople tend to choose a residence according to their sociodemographic and attitudinal characteristics, and land use around their residence may subsequently affect their travel behavior” (p.413). People’s internal characteristics that bring about the self-selection are “travel abilities, needs and preferences” (Mokhtarian and Cao, 2008, p.205; Næss, 2009, p.297) and among them, the importance of preferences/attitudes has been highlighted: in this sense, Bohte et al. (2009) defined residential self-selection as “how households choose a residential location that conforms to their travel-related attitudes” (p.326).

The issue of residential self-selection, which causes endogeneity bias and accordingly threatens internal validity, has been suspected since the mid-1990s (e.g., Kitamura, Mokhtarian and Laidet, 1997), but in reality, it is not easy to control for the self-selection. That is, it is virtually impossible to randomly assign individuals into residential areas or to deliberately match them according to their internal characteristics; on the contrary, it is rather more likely that people with similar attitudinal and sociodemographic characteristics live
together in the same neighborhood. Considering such a residential segregation, land use and environmental planners have discussed methodological solutions for testing the causality of the built environment-travel behavior relationship (Mokhtarian and Cao, 2008).

3. Four Conditions for Establishing Causality

Internal validity, which is jeopardized by residential self-selection, means the validity of a causal inference and thus, one should examine which conditions are required to establish causality. The four conditions include: (1) causal mechanism, (2) covariation (association), (3) non-spuriousness (control for third variables), and (4) time precedence (Babbie, 2004).

First, the causal mechanism means a theoretical base. A relationship without the base is no more than a coincidence. A theory that supports the built environment-travel relationship is the derived demand theory. For instance, the following econometric trip-making model is based on the theory.

\[ T = f(c, S, B) \quad \text{〈Equation 1〉} \]

Where,
\[ T = \text{total travel distance/time or total trip frequency} \]
\[ c = \text{travel price/costs (e.g., trip time/duration, gas price)} \]
\[ S = \text{internal characteristics (sociodemographic, attitudinal, etc.)} \]
\[ B = \text{built environment characteristics} \]

\[ c = f(B) \text{ and } T = f(B, S) \quad \text{〈Equation 2〉} \]

In 〈Equation 1〉, travel behavior is deemed a function of travel
costs, the individual’s inherent characteristics, and the built environment characteristics. In <Equation 2>, the costs are assumed to be a function of the built environment, so by using the reduced form on the right, one can predict travel behavior by considering the individual and built environment characteristics.

The second condition for establishing the causality is covariation. This means that if a supposed cause changes, the effect should also change. As closely related to the second condition, the third condition is non-spuriousness, which denotes that a third extraneous variable should be controlled for. As a typical issue in relation to residential self-selection, sociodemographic and attitudinal variables determine built environment variations through residential choice, so these antecedents should be controlled for in order to accurately estimate the built environment-travel behavior relationship.

<Figure 2> presents a well-known nonspuriousness example and its application to the built environment-travel behavior relationship. Let’s assume that a researcher conducts a study for the prediction of the financial loss from a fire and found a positive relationship between the number of firefighters dispatched to the fire site and the financial loss. Such a result, which is contrary to the initial expectation, is actually attributed to the fire size, which determines both of the variables. That is, the larger the fire is, the more the number of dispatched firefighters as well as the financial loss is. Then, the firefighter size-financial loss relationship would actually be negative, insignificant, or albeit positive, different from the observed one. This is why extraneous variables should be controlled for. [How to establish the causality by controlling for extraneous variables has been summarized in a literature review by Mokhtarian and van Herick (2016).]
Lastly, time precedence, which is also a step forward from the covariation condition, suggests that changes in the supposed cause should be present before those in the outcome. Unlike nonspuriousness, this temporal precedence became highlighted only very recently in the residential self-selection literature: The relationship between the individual’s internal characteristics and the built environment can be two-way (i.e., the directionality should be tested).

Although the choice of the built environment is made according to individual characteristics, the built environment can also affect an individual’s life situations (especially, automobile ownership) and attitudes (i.e., built environment → sociodemographic and attitudinal characteristics). Such a reverse relationship can be firstly explained by the learning process. Næss (2009) argued that by altering attitudes towards automobile travel, compact land use reduces automobile ownership and use, so if this effect is considered, previous studies without consideration of the effect are likely to have rather under-estimated -- not over-estimated -- the built environment effect. Bohle et al. (2009) also suspected that if people move to a new area, they begin to appreciate transit/nonmotorized travel options in the area and thus, they may break old habit/attitudes. Secondly, the built environment-internal characteristics relationship can be explained from...
the perspective of cognitive dissonance. If attitudes and behavior are not in agreement (moving to an area with an unpreferred environment), rationalization takes place, that is, because of a difficulty in changing behavior (moving again to another area), people change their attitudes and justify their residential choice.

In the relationship of internal characteristics-built environment-behavior, if a two-way causal relationship is theoretically supported between the preceding two concepts, one should test which way of the directions is significant or if both directions are significant. Thus, in addition to non-spuriousness, confirming time precedence is a major condition for building causality in the built environment-travel behavior relationship.

V. Major Research Methods: Controlling for Endogeneity Bias

Among analytical methods that control for residential self-selection, the most frequent are regression-based approaches (e.g., multiple linear and two-stage least squares). However, the approaches cannot check the time precedence in the relationship of antecedent-cause-effect. That is, if the individual’s characteristics are considered the antecedent, the effect of the built environment is overestimated (if the characteristics are uncontrolled for); this is what most previous studies have suspected (Bohte, Maat and van Wee, 2009). However, as argued by Naess (2009), if the antecedent is the built environment, not the individual’s characteristics, the built environment effect is under-estimated (if the characteristics are included in an analytical
model as extraneous control variables).

The best way of checking the time precedence is longitudinal approaches such as panel analysis. Especially, Mokhtarian and Cao (2008) recommended longitudinal SEM (structural equation modeling), which satisfies all four conditions for building causality. Except a panel study based on the Puget Sound Transportation Panel (PSTP) (Krizek, 2003), however, such longitudinal studies have been rarely conducted despite a continuous suggestion by the U.S. Transportation Research Board (TRB) (1995, 2009). The major obstacle to a longitudinal/panel study is time, financial, and administrative costs for a long-time data collection (Gim, 2016). Another issue over the data collection period is attrition, that is, dropouts of panel participants.

Even though panel analysis is possible, it still has an accuracy issue (Gim, 2016). The rationale of panel analysis is that if an individual is exposed to different built environment settings due to his/her move to another area or the urban development of the current area, then it works as an intervention in the independent variable, similar to an experiment in the laboratory. However, the intervention is less than a true experiment (Shadish, Cook and Campbell, 2002): The move is often made according to life-cycle changes (e.g., marriage and employment) and urban development occurs in areas whose poor settings require such a development (Mokhtarian and Cao, 2008). Also, the panel measurement has an issue of attenuation: Because of a time lag, travel behavior may not be measured when the effect of the built environment is the largest (Gim, 2016).

In response to the weakness of the panel study, especially as a way of lessening data collection burdens, a quasi-longitudinal study has been conducted (e.g., Cao, Mokhtarian and Handy, 2007a; Handy, Cao
and Mokhtarian, 2005). The researchers directly asked about previous travel patterns, that is, how their current patterns differ from those of the past: The level of the changes was measured on a Likert-type rating scale. However, this quasi-longitudinal approach raises accuracy issues and as acknowledged by the researchers, changes in individual variables other than travel patterns (e.g., attitudes) were not measured, that is, implicitly assumed to be fixed.

As identified by Mokhtarian and Cao (2008), the strengths of longitudinal SEM can be similarly expected through conventional SEM based on cross-sectional data. Figure 3 presents recursive (above) and nonrecursive (below) cross-sectional SEM. First, according to recursive SEM, two theory-supported competitive models (each has the A-B and B-A paths) are specified and their model fits are compared to test the directionality/causality (Kahng and Kwon, 2008). However, almost all previous SEM studies (e.g., Bagley and Mokhtarian, 2002; Cao, Mokhtarian and Handy, 2007b; Gim, 2011; Ory, 2007; Van Acker and Witlox, 2010; Van Acker, Witlox and van Wee, 2007) used the respecification/modification approach instead of the competition approach. According to the respecification approach, only one conceptual model is specified by path removal and addition (insignificant paths are removed and based on the modification index, paths that contribute to a higher model fit are added), and it becomes gradually closer to the optimal model. The major pitfall of this respecification approach is that even though the data support the B-A path instead of the initial A-B path, it cannot be detected. Also, even if the competition approach is adopted, the issue is that unless theoretical models are very different, their model fits often become so similar that researchers cannot choose one over the other (Martens
and Haase, 2006); refer to Scheiner and Holz-Rau (2007) for an actual example regarding the built environment–travel behavior relationship.

(Figure 3) Recursive (above) and nonrecursive (below) structural equation modeling

In contrast to recursive SEM, nonrecursive SEM specifies, as shown in the bottom of <Figure 3>, a two-way causal path (note: correlation ↔, one-way causal →, two-way causal/reciprocal ⇔), and it can test whether data support one, both, or no paths. Thus, while this type of SEM is based on cross-sectional data, it works similar with longitudinal analysis.

Probably the biggest issue with nonrecursive SEM is that it requires valid instruments. The instrumental variables should be supported not only by data, but also in theory. As shown in <Figure 3>, in the path of A (instrumental), B (explanatory), and C (outcome), A should have an effect only on B, not on C (strictly, on the error of C) (paths that should be unsupported for the validity of the instrument are colored in gray). A popular example is a study on whether naturalized citizens in
the U.S. earn higher incomes if they take English names. An appropriate instrument is the linguistic complexity of their original names. The more complex their original names are, the more likely they change to English names, but the complexity does not directly affect the incomes (linguistic complexity of the original name \(\rightarrow\) chances of taking English names, but not the complexity \(\rightarrow\) higher incomes). Excluding such an exceptional case, it is quite difficult to identify valid instruments that are supported both theoretically and empirically. With low instrument validity, the estimation error becomes even larger than that of a model without the instrument [e.g., OLS regression].

Regarding residential self-selection, the selection bias can be directly controlled for through a matching method (rather than from the approach of the causality establishment). A representative method is consonant-dissonant matching. This method separates people into four types: people who prefer a compact area and live in such an area (consonant), those who prefer a sprawling area, but live in a compact area (dissonant), etc. Then, the four types of people are compared by their travel patterns. Details are as follows (Mokhtarian and Cao, 2008, p.211):

> Altitudes toward residential location type are used to classify survey respondents as consonant (well-matched) or dissonant (poorly matched) with respect to their current residential location. The travel behavior of dissonant residents is then compared to that of consonant residents in the type of neighborhood in which they would rather live, and in their current neighborhood. If the travel behavior of dissonant residents is more similar to that of the consonant residents in their desired type of neighborhood, it suggests that their predis-positions dominate their travel behavior. If their travel behavior is more similar to that fo the consonant
residents in their *current* neighborhood, it suggests that the built environment exerts a separate influence that outweighs a contrary predisposition.

However, consonant–dissonant matching simplifies the level of urban sprawl/compactness (compact vs. non-compact or high density vs. low density) and the selection of study areas is also somewhat arbitrary. In order to address this issue, recent studies employed mathematical matching, called propensity score matching (Cao, 2010; Cao, Xu and Fan, 2010; Mokhtarian and van Herick, 2016). A benefit of these studies is that they are capable not only of controlling for the self-selection effect, but also of comparing it with the true effect of the built environment (i.e., among the total observed effect, the proportion of the self-selection effect can be estimated).

### VI. Discussion and Conclusions

#### 1. (Interim) Conclusions of Current Research

This study discussed a variety of ways that previous studies employed to control for residential self-selection. Then, does the self-selection make the built environment–travel relationship over-estimated or under-estimated, that is, what would be the direction of the misestimation? Firstly, as a follow-up of the 2008 study on methodology for controlling for residential self-selection (Mokhtarian and Cao, 2008), Cao et al. (2009) did a qualitative review of the literature and found that more accurate methodology leads to a smaller effect of the built environment, that is, as has previously been suspected, the self-selection
makes the built environment-travel behavior relationship overestimated.

In contrast to the qualitative review (Cao, Mokhtarian and Handy, 2009), quantitative reviews or meta-analytical studies (Ewing and Cervero, 2010; Gim, 2012) found that residential self-selection makes the built environment effect underestimated, so if the self-selection is controlled for, the built environment-travel behavior relationship turns out to be stronger. Regarding the different finding from that by Cao et al. (2009), Ewing and Cervero (2010) explained as follows (p.276):

Still, we are left with a question. Most of the literature reviewde by [Cao et al. (2009)] shows that the effect of the built environment on travel is attenuated by controlling for self-selection, whereas we find no effect(or enhanced effects) after controlling for self-selection. The difference may lie in the different samples included in our study and that of [Cao et al. (2009)], or in the crude way we operationalized self-selection, lumping all studies that conttol for self-selection together regardless of methodology.

Particularly in relation to the second explanation -- different accuracy levels in evaluating residential self-selection -- Gim (2012) conducted a further analysis by separating methodological characteristics into technical rigor and research-design rigor. Technical rigor was considered higher in ascending order of correlation, simple linear regression, and multiple linear regression and research-design rigor was concerned with validity threats such as temporal precedence, sample representativeness, and measurement accuracy. As a result, he found that as argued by Cao et al. (2009), higher technical rigor (i.e., more advanced analytical techniques) results in a smaller effect of the built environment, but stricter research-design rigor causes a larger built environment effect. He further argued that between the
two types of rigor, research-design rigor makes a larger contribution to the analytical result, so the built environment effect would be larger in an ideal study with the perfect technical and research-design rigor.

Unlike earlier studies that focused just on the direction of the misestimation according to residential self-selection, the most recent studies attempted to actively quantify the level of the misestimation in order to compare it with the true built environment effect. However, their findings/arguments are not in agreement.

First, propensity score matching understands the causal effect as the difference between the effect that an individual receives by living in a certain area (factual) and the effect by not living in the area (counterfactual) (Cao, 2010; Cao, Xu and Fan, 2010). Based on propensity score matching, Cao (2010) argued that “the causal influences of neighborhood type are likely to be overstated by 64% for utilitarian walking frequency and 16% for recreational walking frequency” (p.487). One limitation of this method is that it cannot take into account interactions between built environmental variables and other covariates. A second alternative is Heckman’s sample selection model by which mathematical results are produced as if random assignment is used. Based on the model, Cao (2009) reported that more than 3/4 of the observed effect of the built environment on the driving distance originates actually from the built environment (accounting for 16% of the individual’s total driving distance). As its limitation, however, the sample selection model indispensably removes part of the sample: For a better composition of the experimental and control groups, a larger proportion of the sample should be removed and in this vein, this method tries to increase internal validity at the cost of external validity. Third, nonrecursive
SEM can also estimate the size of the self-selection effect. By comparing two competing models (with and without control for the self-selection), Gim (2016) found that among the total observed effect of the built environment on travel behavior, the proportion of the self-selection effect is 73.24% in the case of commuting, 73.24% for shopping travel, and 1.06% in the leisure travel model.

Notably, several studies systematically reviewed empirical studies that employed propensity score matching, sample selection model, and others that actively estimate the level of the residential self-selection effect. The literature reviews agreed on the significance of the confounding effect per se (Cao, 2014), but not on the magnitude. Næss (2014) regarded it as “a tempest in a teapot” (p.76) by arguing that while residential self-selection has a negligible effect on the built environment-travel behavior relationship, academia has an excessive concern for the self-selection effect. In contrast, Mokhtarian and van Herick (2016) reported that the self-selection effect would be considerable, amounting to 2-66% of the observed built environment effect.

While the current literature focuses primarily on the confounding effect of residential self-selection, that is, its threat to internal validity, this study attempts to take a step further by arguing for research on its effect on external validity. That is, whether a study area has more or fewer self-selected residents possibly affects the generalizibility of the findings of the very study.

2. Suggestions for Securing External Validity

As a validity threat, selection bias jeopardizes internal validity by itself, in the sense that it constructs a control group to be not
comparable with the experimental group. This is what studies have highlighted so far. Meanwhile, what is missing is that through its interaction with an explanatory variable, selection bias also threatens external validity (Connaway and Powell, 2010; Sylvia and Sylvia, 2012). Here, external validity is defined as “the extent to which one can generalize the results of the research to the populations and settings of interest in the hypothesis” (Hoyle, Harris and Judd, 2002, p.126).

Thus, to correct the selection bias, it is not enough to employ randomization (random assignment) as discussed in “4.2.1 Selection Bias and Self-Selection”. Also required is random sampling (sample representativeness). (Mechanical or statistical) randomization improves internal validity, but not external validity: Analytical results are applicable only to the randomly assigned sample, which may be biased and not highly representative of the population. In comparison, a sample based on random sampling is neither less nor more sensitive to the built environment than the population.

As argued by Ory (2007), theoretically, sample variation is more important than sample representativeness for inferential statistics. Also, internal validity should always be secured before external validity. Regarding residential self-selection, however, the composition of the sample itself is suspected to bring about differences in its effect. He and Zhang (2014) found that denser, better mixed areas have more self-selected people than sprawled areas and in this vein, Wang and Lin (2014) argued that the conclusion of Western studies may not be transferrable to Asian compact areas.

One possibility is that the magnitude of the land use–travel relationship differs by the size of self-selected residents in the study area. As in (Table 1), the self-selection effect appears to be larger in
more compact areas [as similarly suspected by He and Zhang (2014)]. Thus, a plausible hypothesis about residential self-selection on external validity is that the effect of the compact built environment is smaller in more compact areas that tend to have a higher proportion of self-selected (e.g., in the case of a compact area, less automobile-inclined) people, and so does the self-selection effect.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Measures</th>
<th>Study areas</th>
<th>Proportion of residential self-selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhou and Kockelman</td>
<td>Built environment effect on VMT (vehicle miles of travel)</td>
<td>Austin, Texas</td>
<td>1/10</td>
</tr>
<tr>
<td>Cao (2009)</td>
<td>Built environment effect on VMT</td>
<td>North Carolina</td>
<td>1/4</td>
</tr>
</tbody>
</table>

All in all, empirical studies on residential self-selection continued to examine how strongly selection bias affects internal validity. In comparison, future studies need to empirically examine individuals’ sociodemographic and attitudinal features not only as a confounder or a threat to internal validity, but also as an (interaction) moderator, that is, as a threat to external validity.

References


Ory, D. T., 2007, *Structural equation modeling of relative desired travel amounts*, *Civil and environmental engineering*, University of California, Davis, Davis, CA.


Houghton Mifflin.
Transportation Research Board, National Academies Press.

Tae-Hyoung Tommy Gim: Dr. Gim received his Ph.D. degree in city and regional planning from the Georgia Institute of Technology in Atlanta, Georgia, U.S.A. He is currently an assistant professor of environmental planning at Seoul National University and the founder and director of its Integrated Planning Lab. His fields of expertise include land use-transportation-environment interactions, spatial analysis, and survey and quantitative methods (taehyoung.gim@snu.ac.kr).

Received: 16 September 2017
Revised: 29 September 2017
Accepted: 23 October 2017