The Interplay between Teleworking Choice and Commute Distance

Katherine E. Asmussen

The University of Texas at Austin Department of Civil, Architectural and Environmental Engineering 301 E. Dean Keeton St. Stop C1761, Austin TX 78712, USA Email: kasmussen29@utexas.edu

Aupal Mondal

The University of Texas at Austin Department of Civil, Architectural and Environmental Engineering 301 E. Dean Keeton St. Stop C1761, Austin TX 78712, USA Email: <u>aupal.mondal@utexas.edu</u>

Chandra R. Bhat (corresponding author)

The University of Texas at Austin Department of Civil, Architectural and Environmental Engineering 301 E. Dean Keeton St. Stop C1761, Austin TX 78712, USA Tel: 1-512-471-4535; Email: <u>bhat@mail.utexas.edu</u>

ABSTRACT

In this study, we model the telework-commute distance decision process as a package choice, and use a latent segmentation approach to recognize that some individuals may choose their telework arrangement first and then choose their commute distance (their residential location), while others may adopt a reverse causal behavioral process. The latent segment modeling methodology combines a multiple discrete-continuous probit model for telework adoption/intensity as the proportion of monthly workdays worked from home, from the office, and from a variable third workplace, with a log-linear regression model for commute distance (the home-to-office distance). The data for our analysis is drawn from a 2022 survey of Texas residents. Results suggest that about 20% of the population are likely "movers", whose telework choices led to a residential move that changed commute distance, while the other 80% are "stayers", whose commute distance at their predetermined (to telework) residential location changed telework habits. We further quantify the impact of teleworking on commute distances by computing the expected commute distance for several telework distribution preferences. A shift from 100% in-person work to 100% home-based remote work increases average commute distance by 64.8%, while working from a third workplace for any portion of a month reduces commute lengths relative to only home-based telework. In terms of monthly commute vehicle miles traveled (VMT), working from home for less than 18.6% of the month (about once a week) per worker actually increases overall monthly commute VMT. Monthly commute VMT reductions become tangible only at about 30-40% telework for each worker (about two days of telework per week for full-time workers with 22 workdays per month). In general, between 0 to 40% home-based telework per worker (with the remaining percentage at the office), there is a 10% commute VMT reduction for each 10% increase in home-based telework. In the range of 40% to 80% home-based telework, there is about a 13% commute VMT reduction for every 10% increase in home-based telework. And in the range of 80%-100% homebased telework reduction, there is about a 15% commute VMT reduction for every 10% increase in home-based telework.

Keywords: Telework, commute distance, hybrid work, residential location, causal relationship, latent segments

1. INTRODUCTION

During the peak of the COVID-19 pandemic, and prior to the widespread availability of vaccinations, a significant proportion of the workforce began working remotely (primarily from home). This represented a dramatic shift from before the pandemic. To be sure, in the United States, only six percent of workers primarily worked from a remote location (home or other) before the pandemic, while about three-quarters had never worked from a remote location (Coate, 2021).¹ During the peak of the pandemic in 2020, these percentages rose to 30-40 percent of all employees primarily working remote, and fell to less than 40 percent never having worked remotely (Coate, 2021; Saad and Wigert, 2021; Flynn, 2023). Indeed, many workers had no option but to work remotely at the height of the pandemic due to strict lockdown mandates from local governments. However, since mid-2021, with the availability of vaccines and other means to control COVID, employers have reintroduced the option, and in some cases the requirement, to return to the inperson designated workplace (in the rest of this paper, for convenience, the in-person designated workplace will be referred in short as the "office"). But only about 10% of the US working population has returned full-time to working from the office, while rates of never working remotely remain under 40 percent (Flynn, 2023), leaving a sizeable percentage of the workforce adopting a hybrid work location approach of working remotely on some work days and working from the office on other work days (see Asmussen et al., 2024). Such a hybrid approach combines the benefits of working remotely as well from the office in terms of social, personal and professional considerations. For example, working remotely may offer a better work-family life balance for many individuals with time and cost savings, while working from the office can offer socialization and better career promotion opportunities (see Belostecinic et al., 2021 and Tahlyan et al., 2022). Overall, there is now a widespread consensus that remote work will be more prevalent than in the pre-COVID era, based not only on employee experiences during the pandemic, and altered work location preferences caused by the pandemic (see, for example, Stefaniec et al., 2022 and Asmussen et al., 2023), but also because of generally lower resistance from employers to remote work. Besides, remote work has also been spurred on by the relative ubiquitousness of communication broadband technology within homes and other non-office work places, facilitating seamless communication, access to cloud-based applications, and video-conferencing (Bamieh and Ziegler, 2022; Agarwal, 2023). Also, emerging artificial intelligence, robotics, and augmented and virtual reality (AR and VR) technologies are enabling individuals in many professions (such as in the construction, production, and health sectors) to pursue their traditionally in-person jobs remotely, at least in a hybrid setting of working remotely on some days and from the office on other days (Higgins, 2017; Anderson et al., 2022).

The increasing embrace of a hybrid work location arrangement, while facilitated by emerging technology and providing many employee-level benefits, may also produce "rebound" effects on residential and daily mobility practices (the term "rebound" here is associated with the potential negative effects of teleworking, including residential relocation to live farther away from the work place because of the lower intensity of commuting). In fact, research studies from even before the pandemic (when hybrid work arrangements were more of an exception than the norm) have already alluded to such rebound effects. In the context of residential choices specifically, the focus on such rebound effects, in large part, has been because of the observed positive association between working remotely at least on some workdays and commute distances (we use the terms

¹ The verbiage "primarily worked from a remote location" used in this sentence indicates that an employee never commuted to an in-person out-of-home designated workplace during the week prior to completion of the American Community Survey used by Coate (2021) in the analysis.

remote work and telework interchangeably in this paper). That is, teleworkers are almost always associated with longer commute distances. But this pre-COVID literature is rather equivocal in terms of the directional causal effect between teleworking on the one hand and commute distances on the other that leads up to this positive association. In particular, some studies find (or implicitly assume) that the ability to telework drives (or at least substantially facilitates) the consideration of more far-flung (from the office location) residential locations that otherwise may not have been considered. If so, this would imply that teleworking is leading to a higher propensity to move residences that are then associated with longer commute distances, and therefore potentially sprawl (see, for example, Tayyaran and Khan, 2007; Zhu and Mason, 2014; Alonso et al., 2017; Shabanpour et al., 2018; DeVos et al., 2019; Lennox, 2020). On the other hand, another set of studies find (or implicitly assume) that the positive association between teleworking and commute distances is because those who are already located far away from their workplace (that is, already happen to have a long commute distance, due to reasons not particularly related to teleworking ability) are more likely to embrace telework as a way to reduce the time, financial cost, and stress of long commutes (Ory and Mokhtarian, 2006; Helminen and Ristimäki, 2007; de Abreu e Silva and Melo, 2018; Mitra and Saphores, 2019; Gubins et al., 2019). Additionally, it is also possible that there is no direct causality between teleworking and commute distance, and that the association between the two is purely due to unobserved individual personality and lifestyle factors. Especially when trying to extract out causal effects from cross-sectional data (and almost all earlier studies of the telework-commute distance relationship have been based on crosssectional data), ignoring the jointness of the telework and commute distance decisions can lead to biased estimates of the causal direction of effects and their magnitudes (see Bhat 2015 and Tran et al., 2016). For example, consider the case when a person is extroverted in personality and that this personality trait is not available in any form from a survey. Such a person may like to maximize social time within and beyond the office working space, thus deciding to work from the office more (telework less) and also stay close to the workplace (short commute). If so, ignoring this unobserved negative association between working from the office and commute distance would inappropriately elevate the positive causal effect of telework frequency on commute distance (if that is the causal direction considered) and also incorrectly magnify the positive causal effect of commute distance on telework frequency (if this is the causal direction considered).

In addition to the not-completely-understood causal direction/jointness issue from earlier pre-COVID studies, of course, the telework and commute (or residential choice) landscape has also changed quite considerably from before the pandemic. This is not simply in terms of the higher telework adoption (whether a worker teleworks at all or not), but also in terms of the (a) intensity of telework, (b) locations of telework, and (c) consequent potential alteration of the teleworkcommute distance relationship. In terms of the intensity of telework, telework has evolved from a predominantly occasional work-from-home practice to a routine hybrid work arrangement. This calls for additional emphasis on telework frequency rather than a simple binary distinction of whether an individual adopts teleworking at all or not (as was the emphasis of most pre-COVID studies prior to the pandemic; see Asmussen et al., 2023 for a detailed review). In terms of telework locations, while home was the principal telework location in pre-COVID days, the emergence of a third workplace (such as a café, hotel room, or even a beachfront) as another popular remote work location adds a new dimension to teleworking arrangements and the association with commute distances (Stiles and Smart, 2021; Frick and Marx, 2021; Asmussen et al., 2024). For instance, individuals who might be more likely to work from a third workplace (rather than working from home or office) may also be associated with residences close to dense urban areas

with more third workplace options. Given the general land-use structure of most cities wherein workplaces are more likely to be located in dense areas (see Glaeser et al., 2001, Tran et al., 2016, and Lorenzo, 2021), this may cause an overall negative association between third workplace teleworking and commute distance that is different from the generally observed positive association between the home telework-commute distance relationship (note that the term "commute distance" in this paper refers strictly to the home-office distance, even if an individual works from a third workplace on some days). Thus, it is useful to consider the third workplace as a possible telework location distinct from home to better accommodate the effects of the increasing adoption of a combination of both home and third workplace telework arrangements on the net effect on the telework-commute distance association. Finally, related to the above, in terms of the alteration of the telework-commute distance relationship, the pandemic has arguably altered the nature of the interplay between telework and commute distance. As telework is more accessible today, individuals may make conscious decisions to reside in locations that are farther away from their work office, or closer to it, depending on observed and unobserved individual/household demographic, lifecycle, lifestyle and personality characteristics (de Abreu e Silva, 2022). In particular, while earlier studies suggest that residential location choice determines telework decisions much more so than telework decisions determine residential location choice (see, for example, Ory and Mokhtarian, 2006 and Mitra and Saphores, 2019), the pandemic may have potentially shifted the balance between these two causal directions of effect. Overall, modeling telework and commute distance as a package choice becomes more important today to (1) accommodate both unobserved factors influencing the two choice decisions, (2) capture the "true" causal effect of one decision on another, and (3) allow the possibility that each of the two causal structures prevails within different specific population groups.

The three factors just discussed motivate the current study. In this paper, we model the workplace location (WPL)-commute distance decision-making process as a package choice to account for unobserved factors as well as use a latent segmentation approach to recognize that some individuals may choose their telework arrangement first and then choose their commute distance (i.e., their residential location), while others may adopt a reverse causal behavioral process. The data for the analysis is drawn from a 2022 survey of Texas residents who are workers, have a designated office site (even if they never go in to the office), and have the option of working away from the office. As in most earlier studies, we investigate this inter-relationship between telework arrangements and commute distance using a cross-sectional data set. While a panel (longitudinal) dataset may appear to be more suited to such an analysis because of a temporal component associated with the telework-commute relationship, the disruption caused by the pandemic can obfuscate the extraction of relationships from the use of before-after COVID panels, especially given the short time frame since the onset of the pandemic. At the same time, we believe that studying the telework-commute distance relationship in the new altered state (without reference to the pre-COVID state) using cross-sectional data has its own merits, particularly so because the worst of the pandemic was in the rearview mirror in February-March 2022 (when our survey was conducted), with high rates of vaccination, lifting of masking/social distancing mandates, reopening of all establishments and workplaces, and a substantially diminished threat of death (or even severe complications) due to the virus.² Importantly, our survey was deployed after the Omicron variant had passed its peak in Texas, and there were no mandatory pandemicrelated safety measures in place in Texas. Actually, all COVID-related mandates/restrictions had been lifted in the state as early as March 2021 (Office of the Texas Governor, 2021). Also, when asked about what they believe will be their teleworking arrangement in the "not-so-distant future", respondents indicated a work modality distribution that was largely similar to the current work modality distribution, suggesting that, at least in Texas, the effects of COVID on work patterns had reached a reasonably stable "equilibrium state" from which to tease out the telework-residential location joint decision process.

The modeling methodology combines a multiple discrete-continuous probit (MDCP) model for telework adoption/intensity in terms of the proportion of monthly workdays worked (a) from home, (b) from the office, and (c) from a variable third workplace, with a log-linear regression model for commute distance (that is, the home-office distance). To our knowledge, the resulting sample selection-type model, based on a multiple discrete-continuous sample selection mechanism, is a first in the literature. Further, to determine the causality of the relationship between telework and commute distance, we employ a latent segmentation model, characterizing the population into two segments: (1) those for whom telework choices impact commute distance, and (2) those for whom commute distance influences telework choices. When investigating such causal effects, we accommodate (control for) unobserved individual-specific factors that would otherwise manifest themselves as spurious causal effects. Of course, causality is admittedly an elusive term in all modeling contexts, including when interpreting the effects of exogenous variables such as demographics (leave alone telework choices and commute distance interactions). But, relative to other efforts that have pre-imposed a specific pre-analysis causal direction of effect over the entire set of individuals, we not only allow heterogeneity across individuals in the causal direction pathway effect (through the latent segmentation), but also control for unobserved associative effects between telework choice and commute distance. Even so, and as with any other model, we cannot claim that our estimated effects are the "true" causal effects at play, but we can at least be more confident that, after considering heterogeneity and associative effects, our estimates can only be closer to the true "causal" effect of one endogenous outcome on the other. In the rest of this presentation, we will use the term "causal" without apostrophes, with the recognition that we will never know what the "true" causal directionality of effect and the true causal effect magnitudes are.

The rest of this paper is organized as follows. Section 2 provides a brief overview of the relevant literature, and positions the current research within the larger literature landscape. Section 3 presents the survey administration process, data preparation steps, and the analytic framework.

² Also, without underplaying the value of panel data for telework-commute distance analysis, it should be noted that panel data are not a panacea to study joint decision processes and identify causality. After all, the use of the term causality in this paper, as it necessarily is in all interpretations of the term in the scientific literature, is in epistemic terms (see Williamson, 2006). Thus, given causal processes are inherently latent (we can never actually observe the continuous process of cause-effect thoughts happening in a human brain), all that we really observe is a multivariate distribution of the outcomes, whether from panel data or cross-sectional data. Regardless of the type of data available, the best we can do as analysts is to use a structured methodological framework to tease out causal inferences in an epistemic way through the decomposition of this multivariate distribution into a recursive sequence of cause-effect relationships. In this regard, cross-sectional (CS) data, even if handicapped by the lack of observations over time, and if supplemented by a structured modeling process, can be viewed as incorporating the time element of causality implicitly during the decomposition process (see Wunsch et al., 2010). In the current study, we adopt such a structured decomposition. As indicated by Bhat (2022), none of the many data collection approaches can be unequivocally considered as providing the "ideal data" or even the "best data" for causal analysis.

Section 4 presents model estimation results and goodness of fit measures. Section 5 discusses the approach to estimate the repercussions of teleworking on commute-related vehicle miles of travel, and discusses policy implications. Finally, Section 6 concludes the paper with a summary discussion, along with an identification of future research directions and study limitations.

2. LITERATURE OVERVIEW

The global shift to remote work during the "lockdown" period of the pandemic has generated substantial interest among researchers regarding teleworking. Even prior to the outbreak, telework received a good bit of attention, though much of this literature focused on telework adoption (whether or not an individual teleworked once or more over a specific period, usually a week) and not telework frequency. Also, pre-pandemic investigations confined themselves to home-based work as the alternative to working from the office. Since the pandemic, there have some additional studies focusing on telework adoption/frequency as well as that have considered a third work place (see, for example, Asmussen et al., 2023, Asmussen et al., 2024, Stiles and Smart, 2021, and Cheng et al., 2022). In this paper, we do not provide an exhaustive overview of this pre- and postpandemic literature on telework adoption/frequency, except those that have expressly considered commute distance as an exogenous determinant of telework. The reader is referred to Asmussen et al., 2024 for a recent review of the broader telework adoption/frequency literature. Similarly, there has been substantial research on the short-term telework impacts on overall travel (work plus non-work) over the entire day or over the peak period of the day (see, for example, Pendyala et al., 1991, Walls and Safirova, 2004, Choo et al., 2005, Perch-Nielsen et al., 2014, Kim, 2017, Lachapelle et al., 2018, de Abreu e Silva and Melo, 2018, Hook et al., 2020, Wöhner, 2022, and Caldarola and Sorrell, 2022). Some studies have also focused on the short-term mobile-source emissions impacts of teleworking (see, for example, Giovanis, 2018, Shabanpour et al., 2018, Cerqueira et al., 2020, Tenailleau et al., 2021, and Krasilnikova and Levin-Keitel, 2022). Recent reviews of such short-term telework impacts on travel are available in O'Brien and Aliabadi (2020) and Hostettler Macias et al. (2022). Some other studies have examined the longer-term telework impacts on significant life decisions such as vehicle purchasing, dwelling unit choice, occupation, and family planning (see, for example, Andrey et al., 2004, White et al., 2007, Blumenberg and King, 2019, O'Brien and Aliabadi, 2020, Cuerdo-Vilches et al., 2021, and Torbacki, 2021). But, of most relevance to the current study is the interplay between telework choices and residential location choices, with the latter being proxied by commute distance. Accordingly, in the next two sections, we focus on these telework-commute distance studies. However, as already discussed, the land-use structure in most cities is such that workplaces continue to be clustered in dense urban areas and commute distances are longer for those residing in suburb/rural areas, with the net result that there is a strong positive association between commute distance and suburb/rural area living. Thus, even if not explicitly considering commute distance, we will also present an overview of studies that consider residential built-environment associations with teleworking through the characterization of residential neighborhoods as urban, suburban, or rural.

The next section (Section 2.1) provides an overview of the majority of telework-commute distance studies that assume a specific unidirectional causal structure by treating one of the dimensions (telework adoption/frequency or commute distance) as the endogenous outcome and the other as an exogenous variable. Subsequently, Section 2.2 points out the distinct gap in the literature on considering the potential package and bi-directional nature of the telework-commute distance linkage. Section 2.3 identifies the salient aspects of the current study.

2.1. Unidirectional Causal Linkage

Almost all previous studies either consider teleworking as being predetermined before residential choice (that is, uses teleworking adoption/frequency as an exogenous variable in predicting commute distance), or consider residential location as being made rather independently of telework arrangements and then commute distance influencing telework choices (that is, uses commute distance as an exogenous variable in predicting telework decisions). In the rest of this paper, for brevity, we will refer to the first causal direction of effect as TC (teleworking first, commute distance next) and the second as CT (commute distance first, teleworking next).³

Within the group of TC studies, the general consensus is that higher rates of teleworking adoption/frequency lead to longer commute distances and higher propensities to live in suburbs/rural areas (Nilles, 1991; Mokhtarian, 1991; Melo and de Abreu e Silva, 2017; Lennox, 2020; DeVos et al., 2018; DeVos et al., 2019; Chakrabarti, 2018). Nilles (1991) and Mokhtarian (1991), in two of the earliest empirical studies on the topic, used data from California in the US to study the teleworking effects on commute distance. Both the studies observed that about 6% of teleworkers (that is, who had at least one day of remote work per month (for the Nilles (1991) study or per week (for the Mokhtarian (1991) study) had moved or had plans to move farther away from work after beginning to telework, though the sample size of teleworkers in both studies was small (less than 75 teleworkers). In more recent studies, using data from the Netherlands, DeVos et al. (2019) employed an ordinary least squares (OLS) regression to explore the impact of telework adoption (telework at least once a week) on the commute length (though in time and not distance) and noted that, on average, the commute time of teleworkers is 12% longer than that of nonteleworkers. They also observe that the effect of telework on commute distance varies by occupation sector and residential neighborhood. In the context of the latter, they conclude that the effects of teleworking on commuting time is most pronounced in moderately dense urban areas, and decreases as one moves toward more dense urban areas. DeVos et al., 2019 also observe that employees who transition to teleworking from a completely non-teleworking arrangement experience an increase in commute time of up to 24%. Furthermore, those who telework once a week see an increase of 16%, while those who telework more frequently may experience an even longer commute of up to 17% (DeVos et al., 2019). In another relatively recent study from data in the US, Chakrabarti (2018) reports that the one-way commute distance of individuals who adopt teleworking (telework at least one day a month) is, on average, four miles longer than a traditional non-telework commuter. Also, in a study in Switzerland, Ravalet and Rérat (2019) examined data over a five year period (2010-2015), and again observed that, in 2010, those who worked remotely at least one day during the week had a one-way commute distance that is, on average, 4.3 kilometers longer than a non-teleworker. This difference increased to 8.5 kilometers in 2015.

Within the larger second set of CT studies, the general consensus is that individuals residing farther from their office are more likely to adopt teleworking and with higher frequency, with the view to appropriate long commute times for other uses, as well as to save on high

³ To be complete, we should note that de Abreu e Silva and Melo (2018) use several statistical goodness-of-fit measures from a path analysis modeling approach to compare three different relationships between telework frequency and commute distance, including: a) teleworking frequency as a function of commuting distance, b) commuting distance as a function of telework frequency, and c) a simultaneous bidirectional relationship between commuting distance and teleworking frequency. But no information (other than data fit measures) is provided on the simultaneous relationship model. Besides, this analysis was conducted before the pandemic in Spain, with very low telework frequencies (about 82% of individuals never telework frequency as a function of commute distance (the CT direction); was the preferred specification. And this direction is pretty much the entirety of de Abreu e Silva and Melo's analysis.

commuting costs (see for example, Helminen and Ristimäki, 2007, Asmussen et al., 2023, Asmussen et al., 2024, Mergener and Mansfeld, 2021, Adobati and Debernardi, 2022). Along similar lines, some studies have investigated the impact of the residential built-environment. The findings from such studies has not been always consistent. Most such studies conclude that those residing in suburbs/outskirts of cities (that is, generally those with long commute distances) are more likely to be high-frequency teleworkers (see, for example, Ettema, 2010, Fu et al., 2012, Bhuiyan et al., 2020, Kim and Long, 2022, de Abreu e Silva, 2022, Tahlyan et al., 2022), in part because the typically larger homes in suburbs makes things more conducive to home-teleworking (Isono and Nara, 2022). Mariotti et al. (2022) found this to be particularly the case with "knowledge" workers (as opposed to "essential workers" whose jobs require physical presence at the work site), though they also note that "knowledge workers" residing in very rural areas are not particularly inclined to telework because of less accessibility to reliable and fast broadband networks in remote areas. Contrary to the findings discussed above, a handful of studies have reported that it is those who live in urban and densely populated areas (that is, those with typically short commute distances) who are more likely to adopt telework and do so frequently (see Beck et al., 2020, Haider and Anwar, 2023, López Soler et al., 2021). These studies argue that the urbanization of telework is influenced by social factors rather than purely spatial ones, as is commonly discussed in relation to commute distance. For instance, urban employees may feel they have sufficient social interactions outside of work due to the lively and dynamic nature of city life; therefore, these employees may not need to derive additional utility from social interactions in their in-person work office, leading them to have a higher preference for remote work (Castrillon et al., 2020). Researchers also highlight the digital divide between urban and rural residents, with remote work being more feasible in urban areas where jobs tend to offer greater flexibility, in contrast to rural areas where more hands-on positions, such as manufacturing, require in-person work (Davies, 2021).

2.2. Package/Bidirectional Causal Linkage

There is an ongoing debate on the direction of causality between telework choices and commuting distance, as should be obvious from the discussion in the previous section. Even as they adopt a specific causal direction of effect, several studies in the literature have acknowledged the lack of literature exploring potential bidirectional causality (see, for example, Mokhtarian et al., 2004; Muhammad et al., 2007; DeVos et al., 2019; Ravalet and Rérat, 2019; Elldér, 2020; Hostettler Macias et al., 2022). To exemplify this, O'Brien and Aliabadi (2020) state that existing studies cannot easily distinguish between the following two possibilities: (1) an employee in a large suburban home being more likely to telework due to the long commute distance office, and, (2) an employee buying a large suburban home because they have the opportunity to telework and the need for larger space (e.g., a home office).

However, to our knowledge, only one study in the literature has attempted to address this telework-commute distance causal relationship direction head-on. Ory and Mokhtarian (2006) (OM for short) used data from a 10-year retrospective survey of California workers that included current, former, and non-teleworkers (at the time of the survey). They defined a teleworker as any individual who worked remotely at least two days a month on average, for at least three consecutive months. Through a set of detailed temporal ordering-based descriptive analyses, they attempt to understand how telework impacted one-way commute lengths, and vice-versa. The suggestion from their results is that teleworkers who moved residences (that is, these individuals may be considered to fall in the TC causal direction of linkage) moved closer to their office by an

average commute distance of about 4 miles, while those non-teleworkers who moved residences and then started telework (may be considered to fall in the CT causal direction of linkage) moved farther from their office by an average commute distance of close to 15 miles (for non-teleworkers before and after a move, the average commute distance following a move increased by about 1.2 miles). These results clearly show the kind of obfuscation in the results that can happen if causal directions of effect are ignored. Specifically, the positive association between teleworking and commute distances, based on the OM study, appears to be driven primarily by the CT causal direction of effect rather than the TC causal direction of effect. If these two causal directions are not disentangled, and a strictly TC only causal direction is imposed (as in the TC studies identified in the earlier section), one would naturally get the biased result that teleworking is substantially responsible for longer commute distance. Overall, OM's results also suggest that telework itself does not play any substantial role in residential relocation, with a vast majority of relocation instances (about 84%) not preceded and not succeeded by teleworking, and only 6% of relocation instances in which teleworking ability was identified as an important reason for the relocation.

Of course, the OM study was completed well before the pandemic at a time when teleworking adoption/frequency rates were low. In fact, this is reflected in the OM study in which a person is designated as a teleworker based on a simple binary classification. The total sample size (of teleworkers and non-teleworkers) was also low (218 individuals), which restricted their study to a strictly descriptive analysis of the data without considering heterogeneity across individuals in telework and residential choices. Besides, even with the elegant temporal ordering approach of the analysis, additional investigations into whether the inferred "causal" directions are actually associations based on unobserved factors or "true" causal effects remains an open question. In the current paper, we revisit this package/causality nature of the telework-commute distance linkage in a work arrangement landscape that has been altered by the pandemic, using a multivariate statistical framework that considers the cardinal frequency of telework choices over the period of a month from three different work place locations: home, office, and a third workplace.

2.3. The Current Study

There are several novel aspects of the current study. First, we employ a latent segmentation model to acknowledge that one section of the population may have one causal structure, while the other may have the other. Our investigation also is able to reveal the relative magnitudes of the two groups. While earlier pre-COVID literature suggests that (or even assumes a priori that) distant (from work) residential choices are made without much consideration of telework ability/choice (that is, that the TC causal direction of effect is rather small to non-existent, and that it is the CT causal direction primarily at play; see, for example, OM, 2006, Muhammad et al., 2007; Kim, 2017; Mitra and Saphores, 2019), there is some evidence recently that the increased prevalence/acceptance of teleworking brought on by the pandemic may indeed have generated, for a sizeable set of individuals, new housing desires and a higher tolerance for longer commute distances (see Lei and Liu, 2022). In this context, to our knowledge, this is the first study to propose a formal latent segmentation statistical methodology to capture the market size of each of the TC and CT causal effect directions, and apply it in an altered environment after the onset of the pandemic. Second, the telework choice outcome takes the form of a multiple discrete-continuous (MDC) variable in our analysis that expressly allows the possibility of hybrid work arrangements over the period of an entire month, with a cardinal representation of telework frequency associated

with the number of days of work from a remote work place location as well as the office. In contrast, few earlier studies have considered telework frequency at all, and those that have consider it primarily in an ordinal form (such as "less than once per month", "several times a month", "once a week", "2-4 times a week, and "always", or similar ordinal categorizations) or use a strict time frame of one week that may not fully capture teleworking arrangements over longer periods (see, for example, Popuri and Bhat, 2003, Sener and Bhat, 2011; de Abreu e Silva and Melo, 2018; Shabanpour et al., 2018; Beck et al., 2020; Zhang et al., 2020). Our cardinal representation for frequency enables us to more precisely estimate the effect of teleworking on commute VMT. Third, we consider "home" and "third workplace" (such as a coffee shop or a hotel room) as two distinct telework locations, and estimate the adoption/frequency of telework from each of these distinct locations. As discussed earlier, this enables us to estimate a different kind of teleworkcommute distance relationship for the third workplace (for the so-called digital nomads) relative to home.⁴ Fourth, even within each latent segment, we consider the two outcomes of interest; telework adoption/frequency and commute distance; as resulting from a joint package decision that is impacted by observed as well as unobserved individual personality/lifestyle characteristics (that is, we account for error covariances between the two outcome variables in each latent segment). After accommodating such observed and unobserved characteristics, the direct endogenous effect (i.e., the effect of one dependent outcome on the other) may be considered a "true" causal effect within the context of a joint package decision. Fifth, and related to the above point, from an econometric standpoint, we formulate, for the first time to our knowledge in the econometric literature, a switching/sample selection model of commute distance based on an MDC selection choice. Further, we propose a methodology to compute treatment effects within this new model context, and implement it to estimate the net effect of telework on commute distance.

3. DATA

3.1. The Survey

The primary source of data for the study originated from a workplace location (WPL) choice survey undertaken by the authors as part of a project funded by the Texas Department of Transportation. The survey administration process included a suite of recruitment strategies: e-mails to city chambers of commerce across the state of Texas, professional organizations, media outlets, and a database of roughly 55,000 Texas residents' email addresses. This multi-pronged recruitment effort enabled us to reach employees in a range of industries and sectors. Survey access was restricted to individuals who were employed Texas residents, had a designated physical office in the local area at the time of the survey (even if they never commuted to that location), and had no specific mandatory requirement to be at the office all days of the month (except during periods of absence from work due to sick leave or other sanctioned leaves). Of the 1,400 respondents from the target population (workers in Texas), 463 individuals did not adequately respond to commute and current workplace-related questions. We also removed 113 individuals who reported to have switched employers during COVID, just to maintain stability of the office location so that the final sample for analysis included only those who worked in the same firm and with the same office location. This left us with a final sample of 824 respondents.

The survey elicited information about travel behavior and mobility choices, mode usage, and activity engagement. Pertinent to this study, the survey elicited information on work

⁴ The term "digital nomad" has existed even before the pandemic. But the label started receiving renewed attention after the COVID pandemic, referring broadly to anyone who works remote (potentially at different places on different workdays) and uses a digital device to connect to the office (see Hannonen, 2020 and Nash et al., 2021).

arrangements at the time of the survey, including the number of workdays in the past month, and the number of workdays worked at each of three workplace locations: home, office, and a third workplace. Additionally, individual and household socio-demographics, and information on the current office and home location, were obtained. Finally, respondents were asked to report both the distance and time between their residence and in-person work office.

3.2. Data Description

3.2.1 Sample Description

Table 1 provides information on the individual/household characteristics of the sample of 824 individuals used in our analysis. Our sample consists of only Texas employed individuals, which we are able to compare with the gender, age, and income from the Census Bureau's five-year American Community Survey (ACS) data. The ACS also does provide Texas worker-specific descriptive data related to the formal education level, presence of children (by age) and occupation split in Texas, but groups them differently than in our survey, so we are unable to compare these descriptive statistics from those from our sample. Additionally, the household structure and residential attributes in Table 1 are not available specific to Texas workers in the ACS. So, Table 1 only shows the comparable ACS-based Texas worker percentages for gender, age, and income. Clearly, the comparison with the ACS data shows that our sample exhibits a high proportion of women, those aged 50-64 years, and those with a household annual income in the range of \$100-150K, relative to the Texas working population (U.S. Census Bureau, 2022).

In terms of residential attributes, about one-fifth of sample respondents report residing in an urban area and a similar percentage report residing in a rural area, while a vast majority (57.2%) report residing in suburban areas (respondents were asked to characterize their neighborhood of living in one of these three categories). Most sample respondents also report owning their home, and indicate that they have a private study in their home. Finally, in terms of job characteristics, the office location variables indicate an expected skew toward high density locations.⁵ Not surprisingly, a high proportion of respondents are full-time (working 30 or more hours per week) and not self-employed. A substantial percentage (80.7%) of respondents work 20-24 days of the month, with a non-insignificant 13.0% working 25 days or more in the month. Also, respondents are largely employed in occupations that are remote-work friendly (e.g., education/social services, managerial/technical, and professional services). Very small percentages are employed in occupations that may be viewed as more of "essential" in-person in nature (such as retail sales/food services and healthcare).

As indicated earlier, we are unable to compare many demographic/residential attributes with the population characteristics of Texas workers. However, we are able to examine the representativeness of our sample on work characteristics by comparing with the 2020 Texas Census (Texas Demographic Center, 2022).⁶ In fact, we also considered an alternative weighted approach for presenting the descriptive statistics in Table 1, by developing weights based on the ACS-based Texas worker population splits of age, gender, and income. But, the resulting weighted

⁵ Respondents identified the zip code of their work location, which was used to link built environment variables and compute density as the ratio of number of jobs to unprotected acreage. Based on Ramsey and Bell (2014), zip codes with an employment density less than 0.3 jobs per unprotected acre of land are classified as low density, while those with 0.7 or more jobs per unprotected acre of land are classified as high density. All other zip codes are classified as medium density.

⁶ Note, however, that the 2020 Texas Census does not provide comparable demographic/residential attribute data to our sample, because it only provides the demographic characteristics of the Texas population as a whole without partitioning based on being employed or not.

sample descriptive statistics for work characteristics are generally worse (when compared to the 2020 Texas Census) than those from our unweighted sample. In particular, while there is an underrepresentation of part-time employees (work hours per week < 30 hours) in our sample (4.4% of the sample relative to 11.4% of the Texas population), the weighted sample reflects an overrepresentation by about an equal amount. Both the weighted and unweighted samples are fairly representative of the professional services and information/finance industry sectors that we are able to compare with the 2020 Texas Census, being off by less than 3.5% from the population values (11.9% of the Texas population works in the professional services industry, while 8.5% of the Texas population works in the information/finance sectors). With regard to commute time, in Texas, the average commute time from the Census data is 26.4 minutes, while that from our unweighted sample is 24.0 minutes, compared to a substantial underestimation from the weighted sample of 14.7 minutes. Additionally, the average number of days an employee works in a month from the Census data is 22 days; the unweighted sample mean is very close at 21.8 days relative to 20.8 days from the weighted sample). Also, relative to the 75.4% of employees with a commute time longer than 15 minutes, our unweighted sample statistic of 74.3% is far better than the corresponding weighted sample statistic of only 44%.

Overall, while we could have presented the weighted statistics in Table 1 rather than the unweighted statistics, we choose the latter because weighting just based on age, gender, and income actually makes the descriptive statistics for the work characteristics generally worse off. This is because the combination of the variables in our unweighted sample is such that it reflects the employment characteristics of the Texas population better than the restrictive weighting scheme dominated by just three demographic variables. Thus, we stick with the unweighted sample statistics in Table 1 as well as present the descriptive statistics of the main work-related outcome variables in the next section too based on the unweighted sample (but we provide the weighted statistics too for comparison purposes).

3.2.2 Main Outcome Variables Description

There are two main outcome variables in this study. The first outcome of interest relates to teleworking choice and the second is the commute distance (from the individual's home to the office). There is no good information to compare the teleworking adoption/frequency outcome from our survey with what was on the ground in 2022, and the Texas Census does not provide information on commute distance among Texas workers (only provides information on commute time).

The first telework choice dimension takes the form of a multiple discrete-continuous (MDC) variable (see Bhat 2008 for further details about such variables). Essentially, an MDC type variable is characterized by the discrete choice of multiple alternatives (i.e., whether or not an alternative is chosen), along with a continuous quantity allocated to each chosen alternative. In the context of the survey used in this study, individuals were asked to provide the distribution of their "days of work over the past month" across three work place location (WPL) alternatives – "Work from home", "Work from the office", and "Work from a third workplace". Here, the discrete dimension refers to whether or not each of these WPL alternatives is chosen at least once during the month, and the continuous dimension refers to the (cardinal) number of workdays in the month allocated to each WPL that has a non-zero allocation. Strictly speaking, the alternatives in the MDC outcome represent WPL alternatives, though we will refer to the MDC outcome itself as the telework choice, because, given a certain number of workdays in a month (considered exogenous), the number of office location days can be obtained immediately from the number of home and

third WPL days (that is, we will refer to the MDC outcome as telework choice, though the alternatives of this telework choice will include all the three WPL alternatives of home, office, and third workplace). In this context, in our MDC representation of the telework decision, we use the (cardinal) <u>fraction</u> of days of the month worked at each WPL as the actual dependent outcome (rather than the actual number of days of month worked at each WPL), with the fractions having to necessarily sum to one across the three WPLs for each respondent. This fractional-split MDC outcome is modeled using the approach proposed by Bhat (2008), which expressly acknowledges that each WPL may satisfy specific functional, social, productivity, emotional, privacy, visibility, networking, financial and other personal/professional objectives to different extents. As a consequence, individuals will choose, over a certain period of time and within the context of their chosen/current career, a combination of the three different WPLs to satisfy their respective personal and professional desires. As discussed at length by Asmussen et al. (2023), telework decisions today may be better characterized as a utility-maximizing multiple discrete "horizontal" choice situation in which individuals balance the many pros and cons of each WPL to determine an ideal mix of WPLs over the period of a month (or other specific time periods).

Figure 1 provides the descriptive (unweighted) statistics of the MDC telework dimension across the 824 individuals, using a Venn diagram for ease in understanding. The results show that, over the course of a month, 47.5% of respondents have at least one working day from home, while 93.8% have at least one working day from the office and only 5.2% with at least one workplace instance. These percentages are shown at the outer edges of each of the WPL circles (the corresponding weighted statistics are 41.2% with at least one working day from home, 95.3% with at least one working day from the office, and 7.3% with at least one third workplace instance). Further, and also shown at the outer edges, the mean number of days of participation from home across those who have at least one day of home-based work in a month is 9.8 days (weighted value is 10.7 days). The corresponding mean for those with at least one day of work from office is 17.9 days (weighted value is 16.9 days) and for work from the third workplace is 4.4 days (weighted value if 4.3 days). Figure 1 also shows that fully remote work, as was the work arrangement during the height of the pandemic for a significant percentage of workers, is no longer the case today, with only 51 (=47+3+1) of the 824 workers (constituting 6.2% of the unweighted sample) working fully remote (the corresponding weighted sample percentage of fully remote workers is 6.5%). But, also important to note is that only 427 individuals (51.8% of individuals) work all days of the month from the work office with no remote work whatsoever (that is, 48.2% work remotely at least once a month; the corresponding weighted statistic shows a lower amount of "at-least-oncea-month remote work" at 41.6%).

Other numbers in the Venn diagram provide the number (percentage) of individuals in other combinations of WPL locations. The significant overlap in the circles (that is, individuals working from more than one WPL) is obvious and reinforces the fact that telework choices over a period of a month is a multiple discrete-continuous situation. Specifically, close to 42.4% of the occasions involve the choice of a hybrid of WPLs rather than a single WPL to work from (the corresponding weighted statistic is lower at 37.1%). The most popular WPL combination is only working from the work office (at 51.8% of the entire sample), while the second most chosen option is working from both home and the office (at 37.3% of the entire sample). The WPL combinations that include working from the third workplace have the lowest share. Not surprisingly, the third workplace alternative, if teleworked from, is typically part of a hybrid work arrangement of participating in all three WPLs.

The second outcome of interest, the commute distance, is used as a continuous variable in our analysis. The median one-way commute distance reported in our sample is 10.0 miles (much lower at 6.0 miles based for the weighted sample), while the average one-way commute distance is 13.7 miles (lower at 11.3 miles in the weighted sample). Note that we use the natural logarithm of the commute distance in our modeling analysis as the endogenous outcome to accommodate the rightward skew, and positivity, of the outcome.

Figure 2 presents the overall relationship between telework (regardless of the location of telework, whether home or the third workplace) and commute distance (one-way). From the figure, we observe a gradual increase in the mean commute distance with an increase in telework, but only up to 75% telework (in terms of number of days in the month teleworked). This is consistent with the general literature finding that teleworkers (and frequent teleworkers) are associated with longer commutes. However, those with the highest levels of teleworking (>75% telework) are the ones with the shortest mean commute distances, with those working fully remote associated with the shortest distance between their home and office. In fact, those working fully remote are associated with a mean commute distance that is close to 1.25 miles shorter than those working fully from the office. Figure 3 presents the relationship in reverse, with teleworking as a function of commute distance. From this figure, we notice that the mean telework percentage (remote work days as a percentage of total work days) increases from 19.7% for those with a commute distance in the 0-8 mile range to 28.2% for those with a commute distance in the 8.01-15 miles range. However, there is a drop in mean telework percentage beyond the commute distance of 15 miles. Of course, these figures do not account for heterogeneity across individuals, nor do they provide any sense of whether the relationships observed are pure associations (through unobserved individual-related factors), or causal TC or CT effects. But it does immediately raise the importance of considering telework frequency in cardinal form and over longer time periods than a week, rather than treating telework in binary form or in an ordinal form within a short time frame such as a week.

4. MODELING FRAMEWORK

The joint modeling framework consists of an MDC variable for the workplace location dimension and a continuous outcome variable (in logarithmic form) for the commute distance dimension. The model system consists of several components. A latent segmentation model component is employed to probabilistically assign individuals into one of two causal segments, considering their individual and household-level attributes. Subsequently, within these segments, both the workplace location and commute distance endogenous outcomes of interest are jointly modeled (accommodating unobserved factors), incorporating socio-demographic, household, employment, residential, workplace and other observed variables. Furthermore, within each segment, the unidirectional effects of one endogenous outcome on the other are modeled. Figure 4 provides a simplified conceptual visualization of the model framework. In this section, we first provide the mathematical formulation for this joint model system within each of the two latent segments representing the TC and CT causal effect directions, and subsequently discuss the latent segmentation modeling approach.

4.1. Joint Model Framework Within Each Segment

Consider an individual q (q=1, 2, 3, ..., Q) facing a multi-dimensional MDC and continuous outcome choice system, and assume that the individual belongs to a specific segment h. However, for the time being, we will drop the subscript h for presentation ease and re-introduce this segment-

based index later on in Section 4.2. We will first discuss the MDC component followed by the continuous outcome component. The MDC component discussion is based on Bhat (2008), Bhat et al. (2013), Bhat et al. (2014), Bhat (2015), and Bhat et al. (2016)

For the MDC component, let $x_q = (x_{q1}, x_{q2}, ..., x_{qK})$ be the vector containing the proportion of number of days of work from each WPL k (k=1, 2, 3, ..., K; K=3 in our empirical application). Then, following Bhat (2008), consumer q maximizes utility subject to a proportion binding constraint:

$$\max U_{q}(\boldsymbol{x}_{q}) = \sum_{k=1}^{K} \gamma_{qk} \psi_{qk} \ln\left(\frac{x_{qk}}{\gamma_{qk}} + 1\right)$$
(1)

s.t. $\sum_{k=1}^{K} x_{qk} = 1$,

 $U_q(\mathbf{x}_q)$ refers to the utility accrued by the individual in the allocation across WPLs as contained in vector \mathbf{x}_q . ψ_{qk} is the baseline preference of WPL k, determining participation and in the alternative k for individual q, while γ_{qk} ($\gamma_{qk} > 0$) is a translation parameter that serves the dual role of enabling zero allocations (that is, corner solutions) as well as accommodating satiation effects. Lower values of γ_{qk} imply higher satiation for WPL k (or lower proportion of number of days of work from WPL k; see Bhat, 2008 for details). To develop an econometric model, we parameterize ψ_{qk} and add stochasticity as follows:

$$\psi_{qk} = \exp(\boldsymbol{\beta}' \boldsymbol{z}_{qk} + \boldsymbol{\xi}_{qk}), \tag{2}$$

where z_{qk} is an *E*-dimensional exogenous variable vector corresponding to individual q and alternative k (including a constant), as well as, for the CT segment (but not for the TC segment), the commute distance endogenous outcome. $\boldsymbol{\beta}$ is a corresponding coefficient vector (of dimension $E \times 1$), and ξ_{qk} captures idiosyncratic (unobserved) characteristics. For each consumer, across the error terms ξ_{qk} , we assume a multivariate normal distribution: $\xi_q = (\xi_{q1}, \xi_{q2}, ..., \xi_{qK})' \sim MVN_K(\boldsymbol{\theta}_K, \boldsymbol{\Lambda})$, where $MVN_K(\boldsymbol{\theta}_K, \boldsymbol{\Lambda})$ indicates a *K*-variate normal distribution with a mean vector of zeros denoted by $\boldsymbol{\theta}_K$ and a covariance matrix $\boldsymbol{\Lambda}$.

The optimal consumption allocations may be derived from Equation (1) by forming the Lagrangian and applying the Karush-Kuhn-Tucker (KKT) conditions as follows (after substituting Equation (2) in Equation (1)):

$$\mathcal{I}_{q} = \sum_{k=1}^{K} \gamma_{qk} \exp(\beta' z_{qk} + \xi_{qk}) \ln\left(\frac{x_{qk}}{\gamma_{qk}} + 1\right) - \lambda_{q} \left(\sum_{k=1}^{K} x_{qk} - 1\right),$$
(3)

where λ_q is the Lagrangian multiplier associated with the binding proportionality constraint. Then, the first-order conditions (FOCs) for the optimal fractional allocations (the x_{qk}^* values) are given by:

$$\exp(\beta' z_{qk} + \xi_{qk}) \left(\frac{x_{qk}^*}{\gamma_{qk}} + 1\right)^{-1} - \lambda_q = 0, \text{ if } x_{qk}^* > 0, \ k = 1, 2, ..., K$$
(4)

$$\exp(\beta' z_{qk} + \xi_{qk}) \left(\frac{x_{qk}^*}{\gamma_{qk}} + 1\right)^{-1} - \lambda_q < 0, \text{ if } x_{qk}^* = 0, \ k = 1, 2, ..., K$$

Of course, in addition to satisfying the FOCs above, the optimal fractional allocations should also meet the constraint $\sum_{k=1}^{K} x_{qk}^* = 1$. That is, only *K*-1 of the x_{qk}^* values need to be estimated. To accommodate this constraint, let m_q be the consumed good with the lowest value of *k* for the q^{th} consumer (note that at least one WPL must be selected by the worker during the month; that is, $x_{qk}^* > 0$ for at least one *k*). For the good m_q , since it is, by construction, consumed, the Lagrangian multiplier may be derived from Equation (4) as:

$$\lambda_q = \exp(\boldsymbol{\beta}' \boldsymbol{z}_{qm_q} + \boldsymbol{\xi}_{qm_q}) \left(\frac{\boldsymbol{x}_{qm_q}^*}{\gamma_{qm_q}} + 1 \right)^{-1}.$$
(5)

Substituting for λ_q from above into Equation (4) for the other goods k (k = 1, 2, ..., K; $k \neq m_q$), and taking logarithms, we can rewrite the KKT conditions as:

$$V_{qk} + \xi_{qk} = V_{qm_q} + \xi_{qm_q}, \text{ if } x_{qk}^* > 0, \ k = 1, 2, ..., K, \ k \neq m_q$$

$$V_{qk} + \xi_{qk} < V_{qm_q} + \xi_{qm_q}, \text{ if } x_{qk}^* = 0, \ k = 1, 2, ..., K, \ k \neq m_q ,$$
(6)

where $V_{qk} = \beta' z_{qk} - \ln\left(\frac{x_{qk}^*}{\gamma_{qk}} + 1\right)$. Letting $y_{qk} = V_{qk} + \xi_{qk}$, and $y_{qkm_q}^* = y_{qk} - y_{qm_q}$, the KKT conditions

in Equation (6) are equivalent to:

$$y_{qkm_q}^* = 0, \text{ if } x_{qk}^* > 0, \ k = 1, 2, ..., K, \ k \neq m_q$$

$$y_{qkm_q}^* < 0, \text{ if } x_{qk}^* = 0, \ k = 1, 2, ..., K, \ k \neq m_q.$$
(7)

Three significant identification issues warrant attention in this context, as the KKT conditions outlined above are based on differences, as reflected in the $y_{qkm_q}^*$ terms. These identification considerations are discussed in detail in Bhat et al. (2013)

<u>First</u>, the constant and coefficients on consumer-specific exogenous variables cannot be identified within the term $\beta' z_{qk}$ for one of the *K* WPLs, and so one of the WPLs is chosen as the base. Second, because only error differences matter, only the elements of the covariance matrix Λ_1 of $\varepsilon_{qk1} = \xi_{qk} - \xi_{q1}$, $k \neq 1$ are estimable (taking error differences with respect to the first WPL without any loss in generality). But the KKT conditions use differences relative to the first consumed good m_q for consumer q. Correspondingly, the analyst needs the covariance matrix Λ_{m_q} when translating the KKT conditions to the qth individual's consumption probability. Since m_q is not the same across consumers q, the covariance matrix Λ_{m_q} is also not the same across consumers. But all the Λ_{m_q} matrices must conform to the same covariance matrix Λ of the original error term vector ξ_q . A way to ensure this consistency is to construct Λ from Λ_1 by expanding Λ_1 with an additional conformable column of zero elements to the left and an additional conformable row of zero elements on top. Third, for scale normalization purposes, we restrict,

again without any loss in generality, the second row-second column element of Λ to the value of one.

For efficiency in bookkeeping and ease of presentation, we provide the vectorized forms of the key variables discussed so far. To do so, stack y_{qk} , V_{qk} , ξ_{qk} , and γ_{qk} into $K \times 1$ vectors: $\mathbf{y}_q = (y_{q1}, y_{q2}, ..., y_{qK})'$, $V_q = (V_{q1}, V_{q2}, ..., V_{qK})'$, $\xi_q = (\xi_{q1}, \xi_{q2}, ..., \xi_{qK})'$, and $\mathbf{\gamma} = (\gamma_{q1}, \gamma_{q2}, ..., \gamma_{qK})'$, and let $\mathbf{z}_q = (\mathbf{z}_{q1}, \mathbf{z}_{q2}, ..., \mathbf{z}_{qK})'$ be a $K \times E$ matrix of variable attributes. Then, we may write, in vector notation,

$$y_q = V_q + \xi_q \,. \tag{8}$$

Equation (8) above defines the underlying system of equations governing the MDC component (WPL) of the joint framework. We will now discuss the continuous component of the framework, which is related to the commute distance dimension. To proceed, let \tilde{y}_q be the continuous variable denoting the logarithm of the commute distance for individual q. Then, in the usual linear regression fashion, we may write:

$$\tilde{y}_q = \boldsymbol{\delta}' \boldsymbol{w}_q + \tilde{\boldsymbol{\varepsilon}}_q, \tag{9}$$

where w_q is an $(A \times 1)$ vector of exogenous variables (including a constant) as well as, for the TC segment (but not the CT segment), the observed values of WPL proportions of the multiplediscrete choice outcome variable. δ is a corresponding compatible coefficient vector and $\tilde{\varepsilon}_q$ is the random error term, assumed to be normally distributed with mean zero and scale σ .

To configure the joint framework, we may combine Equations (8) and (9) in the following

way, forming a unified system of equations. To do so, let $\vec{y}_q = \begin{pmatrix} y_q \\ \overline{y}_q \end{pmatrix}$. Also, define the following

vector \boldsymbol{B}_a and the covariance matrix $\boldsymbol{\Upsilon}$:

$$\boldsymbol{B}_{q} = \begin{pmatrix} \boldsymbol{V}_{q} \\ \boldsymbol{\delta}'\boldsymbol{w}_{q} \end{pmatrix} \quad [(K+1)\times1] \text{ vector } \qquad \boldsymbol{\Upsilon} = \begin{pmatrix} \boldsymbol{\Lambda} & \boldsymbol{\Omega} \\ \boldsymbol{\Omega}' & \sigma \end{pmatrix} \quad [(K+1)\times(K+1)] \text{ matrix}$$

where, Ω is a $(K \times 1)$ vector capturing the correlations between the continuous commute distance variable and the MDC alternatives (i.e., the workplace locations), with the first element of this vector being zero (to be consistent with the identification issues discussed earlier in the context of the Λ matrix). Then, we can write $\ddot{y}_q \sim MVN_{K+1}(B_q, \Upsilon)$.

For model estimation, if we knew the latent segment to which an individual belonged, we can derive the probability of the observed vector $\mathbf{x}_q^* = (x_{q1}^*, x_{q2}^*, ..., x_{qK}^*)$ containing the proportion of number of days of work from each WPL *k* and the observed (log of) commute distance \bar{y}_q . To obtain this probability expression, we will first define a $[K \times (K+1)]$ contrast matrix \mathbf{M}_q . The contrast matrix provides an efficient matrix mechanism to take the appropriate differences of the underlying functions described in Equation (7) with respect to the reference good m_q and to construct the differenced covariance matrix, and is akin to the 'mask' matrix used in multinomial probit models (see for example, Chapter 5 of Train, 2009). To construct this contrast matrix, first initialize all the elements of the matrix to zeros. Then, fill the first (*K*-1) rows and *K* columns with an identity matrix of size *K*-1 with an extra column of "-1" values added at the m_q^{th} column.

Finally, insert a value of one at the K^{th} row and $(K+1)^{th}$ column position. With \mathbf{M}_q as defined, we may write $\mathbf{\tilde{y}}_q^* = \mathbf{M}_q \mathbf{\tilde{y}}_q \sim MVN_K(\mathbf{H}_q, \mathbf{\Psi}_q)$, where $\mathbf{H}_q = \mathbf{M}_q \mathbf{B}_q$ and $\mathbf{\Psi}_q = \mathbf{M}_q \mathbf{\Upsilon} \mathbf{M}_q'$. Next, partition the vector $\mathbf{\tilde{y}}_q^*$ into a sub-vector $\mathbf{\tilde{y}}_{q,NC}^*$ of length $D_{q,NC} \times 1$ ($0 \le D_{q,NC} \le K - 1$) for the non-consumed WPLs, and another sub-vector $\mathbf{\tilde{y}}_{q,CC}^*$ of length $D_{q,CC} \times 1$ ($1 \le D_{q,CC} \le K$) for the consumed WPLs (except the base WPL m_q) and the continuous outcome (note that $D_{q,NC} + D_{q,CC} = K$). Let

 $\tilde{\mathbf{y}}_{q}^{*} = \left(\left(\tilde{\mathbf{y}}_{q,NC}^{*}\right)', \left(\tilde{\mathbf{y}}_{q,CC}^{*}\right)'\right)'$, which may be obtained from $\bar{\mathbf{y}}_{q}^{*}$ as $\tilde{\mathbf{y}}_{q}^{*} = \mathbf{R}_{q}\bar{\mathbf{y}}_{q}^{*}$, where \mathbf{R}_{q} is a rearrangement matrix of dimension $K \times K$ with zeros and ones. The first $(K-1) \times (K-1)$ submatrix of \mathbf{R}_{q} corresponds to the MDC component. \mathbf{R}_{q} is designed as follows. Consider a consumer q who selects WPLs 1 and 2 for consumption. Thus, $m_{q} = 1$, the number of non-consumed goods equals one (corresponding to WPL 3), and the number of non-base consumed WPLs also equals one (corresponding to WPL 2, with good 1 serving as the base needed to take utility differentials). Then, the first $(K-1) \times (K-1)$ sub-matrix (2×2 sub-matrix in our empirical case) of the rearrangement matrix \mathbf{R}_{q} can be defined with respect to goods 2 and 3 in the following way, with the element at the K^{th} row and K^{th} column of \mathbf{R}_{q} being one (and the rest of the elements of the K^{th} row and K^{th} column being zero):

$$\mathbf{R}_{q} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{q,NC} \\ \mathbf{R}_{q,CC} \end{bmatrix},\tag{10}$$

where the upper sub-matrix $\mathbf{R}_{q,NC}$ corresponds to the non-consumed goods (of dimension $D_{q,NC} \times K$) and the lower sub-matrix $\mathbf{R}_{q,CC}$ corresponds to the consumed goods and the continuous outcome (of dimension $D_{q,CC} \times K$). Consistent with the above re-arrangement, define $\tilde{H}_q = \mathbf{R}_q H_q$ and $\tilde{\Psi}_q = \mathbf{R}_q \Psi_q \mathbf{R}'_q$.

In the rest of this section, we will use the following key notation: $f_G(.; \mu, \Sigma)$ for the multivariate normal density function of dimension *G* with mean vector μ and covariance matrix Σ , ω_{Σ} for the diagonal matrix of standard deviations of Σ (with its r^{th} element being $\omega_{\Sigma,r}$), $\phi_G(.;\Sigma^*)$ for the multivariate standard normal density function of dimension *G* and correlation matrix Σ^* , such that $\Sigma^* = \omega_{\Sigma}^{-1} \Sigma \omega_{\Sigma}^{-1}$, $F_G(.;\mu,\Sigma)$ for the multivariate normal cumulative distribution function of dimension *G* with mean vector μ and covariance matrix Σ , and $\Phi_G(.;\Sigma^*)$ for the multivariate standard normal cumulative distribution function of dimension *G* with mean vector μ and covariance matrix Σ , and $\Phi_G(.;\Sigma^*)$ for the multivariate standard normal cumulative distribution function of dimension *G* and correlation matrix Σ^* . Let τ be the collection of parameters to be estimated within each latent segment: $\tau = (\beta', \delta', \gamma', [\text{Vechup}(\Upsilon)]')'$, where the operator Vechup(.) row-vectorizes the non-zero upper diagonal elements of a matrix. Then, if we knew the latent segment to which an individual belonged, the joint probability of the observed WPL fractional vector $\mathbf{x}_a^* = (\mathbf{x}_{a_1}^*, \mathbf{x}_{a_2}^*, ..., \mathbf{x}_{a_K}^*)$ and

the observed log(commute distance) \breve{y}_q for individual q (which from an estimation standpoint may also be considered as the likelihood function for individual q given the parameter vector τ) may be written as:

$$\operatorname{Prob}_{q}(\boldsymbol{x}_{q}^{*}, \boldsymbol{y}_{q}) = L_{q}(\boldsymbol{\tau}) = \operatorname{det}(\boldsymbol{J}_{q}) \int_{\boldsymbol{t}_{q,NC}=-\infty}^{0} f_{K}\left(\boldsymbol{t}_{q,NC}, \boldsymbol{\theta}_{L_{q,CC}}; \boldsymbol{\tilde{H}}_{q}, \boldsymbol{\tilde{\Psi}}_{q}\right) d\boldsymbol{t}_{q,NC}, \qquad (11)$$

where det(J_q) is the determinant of the Jacobian of the transformation from y_q to the consumption quantities x_a^* (see Bhat, 2008):

$$\det(\boldsymbol{J}_{q}) = \left\{ \prod_{k \in \mathcal{C}_{q}} \frac{1}{\boldsymbol{x}_{qk}^{*} + \boldsymbol{\gamma}_{qk}} \right\} \left\{ \sum_{k \in \mathcal{C}_{q}} \left(\boldsymbol{x}_{qk}^{*} + \boldsymbol{\gamma}_{qk} \right) \right\},\tag{12}$$

where \mathcal{T}_{q} is the set of WPLs with non-zero work days for consumer q (including WPL m_{q}).

Using the marginal and conditional distribution properties of the multivariate normal distribution, the probability in Equation (11) can be written as:

$$\operatorname{Prob}_{q}(\boldsymbol{x}_{q}^{*}, \boldsymbol{\breve{y}}_{q}) = L_{q}(\boldsymbol{\tau}) = \operatorname{det}(\boldsymbol{J}_{q}) \times f_{D_{q,C}}(\boldsymbol{\theta}_{D_{q,CC}}; \boldsymbol{\dot{H}}_{q,CC}, \boldsymbol{\Psi}_{q,CC}) \times F_{D_{q,NC}}(\boldsymbol{\theta}_{D_{q,NC}}; \boldsymbol{H}_{q,NC}, \boldsymbol{\Psi}_{q,NC})$$
(13)
$$= \operatorname{det}(\boldsymbol{J}_{q}) \times \left(\prod_{g=1}^{D_{q,CC}} \boldsymbol{\omega}_{\boldsymbol{\Psi}_{q,CC}, g}\right)^{-1} \left(\phi_{D_{q,CC}}(\boldsymbol{\omega}_{\boldsymbol{\Psi}_{q,CC}}^{-1}(-\boldsymbol{\tilde{H}}_{q,CC}), \boldsymbol{\tilde{\Psi}}_{q,CC}^{*})\right) \times \Phi_{D_{q,NC}}(\boldsymbol{\omega}_{\boldsymbol{\Psi}_{q,NC}}^{-1}(-\boldsymbol{H}_{q,NC}), \boldsymbol{\tilde{\Psi}}_{q,NC}^{*}),$$
(13)
where $\boldsymbol{\tilde{H}}_{q,NC} = \mathbf{R}_{q,NC}\boldsymbol{H}_{q}, \quad \boldsymbol{\tilde{H}}_{q,CC} = \mathbf{R}_{q,CC}\boldsymbol{H}_{q}, \quad \boldsymbol{\tilde{\Psi}}_{q,NC} = \mathbf{R}_{q,NC}\boldsymbol{\Psi}_{q}\mathbf{R}_{q,NC}', \quad \boldsymbol{\tilde{\Psi}}_{q,CC} = \mathbf{R}_{q,CC}\boldsymbol{\Psi}_{q}\mathbf{R}_{q,CC}',$
$$\boldsymbol{\tilde{\Psi}}_{q,NC,CC} = \mathbf{R}_{q,NC}\boldsymbol{\Psi}_{q}\mathbf{R}_{q,CC}', \quad \boldsymbol{\tilde{H}}_{q,NC} = \boldsymbol{\tilde{H}}_{q,NC} + \boldsymbol{\tilde{\Psi}}_{q,NC,CC}(\boldsymbol{\tilde{\Psi}}_{q,CC})^{-1}(-\boldsymbol{\tilde{H}}_{q,CC}),$$

$$\boldsymbol{\tilde{\Psi}}_{q,NC} = \boldsymbol{\tilde{\Psi}}_{q,NC} - \boldsymbol{\tilde{\Psi}}_{q,NC,CC}(\boldsymbol{\tilde{\Psi}}_{q,CC})^{-1}\boldsymbol{\tilde{\Psi}}_{q,NC,CC}', \quad \boldsymbol{\tilde{\Psi}}_{q,CC}^{*} = \boldsymbol{\omega}_{\boldsymbol{\tilde{\Psi}}_{q,CC}}^{-1} \boldsymbol{\tilde{\Psi}}_{q,CC}', \quad \boldsymbol{\omega}_{\boldsymbol{\tilde{\Psi}}_{q,CC}}^{-1}, \text{ and }$$

$$\breve{\Psi}_{q,NC}^* = \boldsymbol{\omega}_{\breve{\Psi}_{q,NC}}^{-1} \breve{\Psi}_{q,NC} \boldsymbol{\omega}_{\breve{\Psi}_{q,NC}}^{-1}.$$

The multivariate normal cumulative distribution (MVNCD) function in Equation (13) is of dimension $D_{q,NC}$, which can have a dimensionality of up to (*K*-1). The positive-definiteness of Υ can be ensured by using a Cholesky-decomposition method and estimating these Cholesky-decomposed parameters.

4.2. Latent Segmentation Model

The derivation thus far discusses the joint modeling framework assuming we know the causal directionality (whether the TC or CT causal direction) at play for the individual. That is, the probability in Equation (13) is for the (TC or CT) segment to which the individual belongs. However, we do not observe this causal directionality.⁷ Thus, we use a latent segmentation

⁷ The survey did not include a question in the survey instrument asking respondents to indicate which causal direction applied to their decision process. In any case, while such self-reported causality may be another data point for triangulation purposes, cognitive psychology literature indicates that self-reports are much better at recalling individual events than at tracing the causality sequence between events. This is because of a number of hard-wired cognitive biases (see, for example, Sloman and Lagnado, 2015; Hastie, 2015; Le Pelley et al., 2017, and Watkins, 2019) that tend to creep up when trying to ascribe causality, including (but not limited to) putting forward the first explanation that comes to mind (the "take-the-first heuristic"), simply not responding or arbitrarily making up a relationship to avoid cognitive effort ("law of least mental effort"), and attempting to rationalize a decision process as

approach in our analysis, in which we write the probability of the individual being in segment h (whether in the TC or CT segment) itself as a function of individual exogenous variables.⁸ In this latent segmentation approach, the probability of the observed WPL fractions and commute distance in Equation (13) is specific to a particular segment, h, which may be represented as

Prob_{*qh*} $(\boldsymbol{x}_{q}^{*}, \boldsymbol{y}_{q}) = L_{qh}(\boldsymbol{\tau}_{h})$, with parameters $\boldsymbol{\tau}_{h} = \left(\boldsymbol{\beta}_{h}^{\prime}, \boldsymbol{\delta}_{h}^{\prime}, \boldsymbol{\alpha}_{h} \boldsymbol{\gamma}_{h}, [\text{Vechup}(\boldsymbol{\Upsilon}_{h})]^{\prime}\right)^{\prime}$ specific to latent segment *h*. Although the actual assignment of individual *q* to a specific segment is not observed, it is possible to attribute a probability π_{qh} (h = 1, 2, ..., H; H = 2 in our empirical analysis) to individual *q* belonging to segment *h*. The conditions that $0 \le \pi_{qh} \le 1$ and $\sum_{h=1}^{H} \pi_{qh} = 1$ must be met.

To enforce these restrictions, following Bhat (1997), the following logit link function is used:

$$\pi_{qh} = \frac{\exp(\boldsymbol{v}_h'\boldsymbol{a}_q)}{\sum_{j=1}^{H} \exp(\boldsymbol{v}_j'\boldsymbol{a}_q)},$$
(14)

where, \boldsymbol{v}_h is the corresponding $(J \times 1)$ vector of parameters specific to segment h, \boldsymbol{a}_q is the corresponding $(J \times 1)$ vector of individual characteristics of individual q, and $\boldsymbol{v}_H = 0$ serves as a vector identification condition. Defining $\boldsymbol{\varpi} = [\boldsymbol{\tau}'_1, ..., \boldsymbol{\tau}'_H; \boldsymbol{v}'_h, ..., \boldsymbol{v}'_H]'$, then the overall probability of the observed WPL fractions and commute distance (that is, the likelihood function for individual q) is:

$$L_{q}(\boldsymbol{\varpi}) = \sum_{h=1}^{H} \pi_{qh} \times \operatorname{Prob}_{qh}(\boldsymbol{x}_{q}^{*}, \boldsymbol{y}_{q}) = \sum_{h=1}^{H} \pi_{qh} \times L_{qh}(\boldsymbol{\tau}_{h})$$
(15)

and the overall likelihood function is then given as:

$$L(\boldsymbol{\varpi}) = \prod_{q} L_{q}(\boldsymbol{\varpi})$$

Bhat's (2018) matrix-based approximation method for evaluating the multivariate normal cumulative distribution (MVNCD) function may employed to evaluate the integral embedded in the likelihood function for any value of K. In the empirical case at hand, this MVNCD function evaluation is only of dimensionality two, which can be evaluated rapidly.

a carefully orchestrated cause-effect pathway even if such clarity in thought may not have actually existed (labeled sometimes as the "causality heuristic").

⁸ The latent segmentation approach is essentially an application of finite-mixture models. Finite mixture models may be used to allow for greater flexibility in modeling unobserved heterogeneity in the population. As the number of mixture components increases, this becomes tantamount to a non-parametric unobserved heterogeneity distribution (see Bhat and Lavieri, 2018 and McLachlan et al., 2019) for a detailed discussion. But another (and most widely used) application of finite mixture models, as in the current paper, is to consider situations where each mixture component is viewed as representing a distinct population group (or cluster of individuals) with unique sensitivities to exogenous variables or unique decision mechanisms (see McLachlan and Peel, 2000). This one-to-one correspondence between clusters and mixture components has a long history, connecting mixture models with cluster models (also sometimes referred to as classification models), beginning with the work of Tiedeman (1955) (see McNicholas, 2016 and McLachlan et al., 2019 for reviews of finite-mixture based cluster/classification models). In the transportation field, latent segmentation has been used to represent the causal direction of effect at play in many research papers, including Waddell et al., 2007, Angueira et al., 2019, Astroza et al., 2019, Keya et al., 2021, and Batur et al., 2024.

5. MODEL ESTIMATION RESULTS

The final model specification was developed through a systematic method of testing several functional forms and combinations of explanatory variables, while removing statistically insignificant ones. Estimation was undertaken using codes and routines written by the research group in the GAUSS matrix programming language (Aptech, 2022). Exogenous variables collected in categorical form as opposed to a continuous form were considered as dummy variables in the most disaggregate form available, and progressively combined based on statistical tests to yield parsimonious specifications. Further, we explored the effect of work days from each WPL alternative on commute distance (for the TC causal latent segment) in direct fractional form (as a continuous endogenous explanatory variable) as well as ranges of the fraction (as a series of dummy endogenous explanatory variable). Similarly, we considered the effect of (ln) commute distance on telework choice (for the CT causal latent segment) directly as a continuous endogenous explanatory variable as well as a series of commute distances. In all our specification explorations, the dummy variable representations for telework frequency (for the TC latent segment) and commute distance (for the CT segment) outperformed the corresponding continuous specifications.

A few other notes about our estimation procedure and specification. First, the objective of this paper is to investigate individual-level causal relationships between exogenous variables and the outcomes of interest, as well as the causal relationships among the outcomes themselves. In undertaking such a causal analysis, there is no reason that the exogenous variables should strictly represent the population of interest. For example, telework and commute distances may vary across individuals of different age groups, but such demographic heterogeneity is considered through the exogenous "age" category variables. Thus, as long as there is adequate variation in the age variable in the sample to test a variety of functional forms across different age ranges, whether our sample is representative of the Texas population age distribution or not is really immaterial for accurately identifying the age. Additionally, because our sampling strategy itself is not based on whether an individual teleworks or not and what an individual's commute distance is (that is, our sample corresponds to the case of exogenous sampling where the sample collection process itself is not predicated on the endogenous outcome values), an unweighted estimation is the appropriate approach, because it provides consistent estimates as well as yields more efficient estimates relative to a weighted estimation procedure (see Wooldridge, 1995 and Solon et al., 2015 for an extensive discussion). Second, in addition to considering heterogeneity across individuals in the effect of observed variables on the baseline preference function (the ψ_{ak} function of Equation (2)),

we also accommodate heterogeneity in the satiation parameters (the γ_{qk} parameters) through a

 $\gamma_{qk} = \exp(\kappa'_k \upsilon_q)$ parametrization, where υ_q is a vector of independent variable characteristics and

 κ_k is a vector to be estimated. Third, we used a t-statistic threshold of 1.00 to retain variables (corresponding to a 0.32 level of significance or 68% confidence level), because of the moderate-sized sample used in the analysis, the small share of individuals who use the "home only" WPL and all combinations involving the third workplace location, and the potential for such included variables to guide future WPL investigations with larger sample sizes.⁹ Fourth, we restricted the

⁹ The reader will also note that choosing a level of significance is closely related to Type II error. Especially, one needs to worry not only about including variables incorrectly (Type I error), but also rejecting variables incorrectly (Type II error). In selecting a level of significance, we face a trade-off between making the two different types of error. In multidimensional models, we would rather make a slightly larger Type I error if it means identifying variables

effects of exogenous variables to be the same across the two latent segments for each of the telework and commute distance model components, because there is no theoretical or conceptual reason to expect that exogenous variables will affect each of the telework and commute distance dimensions differently across segments simply based on the direction of causality. For example, there is no reason to believe that the tendency of essential workers to telework less than nonessential workers should differ based on whether the TC or CT causal direction is at play. Similarly, there is no reason to believe that the gender or age of a person should differentially affect commute distance based solely on the causal direction at play. For the same reason, we maintain the same covariance matrix elements for the telework component of the model system across the two segments, and the same variance for (ln) commute distance across the two segments. But, to allow for the possibility that the unobserved association between the telework and commute distance dimensions may differ across the segments, we allow the covariance elements across the telework and commute distance dimensions to vary by segment. Also, because we are having different causal direction of effects (that introduce the telework outcomes in the commute distance component for the TC segment, and the commute distance outcome in the telework outcome for the CT segment), we leave the constants to be freely estimated across the segments. Fifth, we attempted a host of different interaction effects of variables, especially those related to gendered lifecycle (interaction of gender with presence of children and household structure), which did turn up to be important. In addition, we also explored the interaction of telework decisions and the residential neighborhood density dummy variables (urban, suburban, and rural) and the employment office density dummy variables (low, medium, and high density) for the TC causal segment (to capture any differences in telework effects on commute distances based on residential and/or office density). Similar interaction effects of commute distance with the residential/office density dummy variables were also considered for the CT causal segment (to capture any differences in commute distance effects on telework decisions based on residential/office density locations). None of these turned out to be statistically significant at our designated threshold level. The model results are presented by variable within each broad group going down the rows of Table 2. The first three columns of the table present the telework model component results and the last column presents the results for the (ln)commute distance model component. In the telework model component, the effect of each variable on the logarithm of the baseline preference (that is, the effects of variables on $\ln(\psi_{ak})$, or the elements of the β vector) is listed, followed by any effect that the variable may have on the logarithm of the γ_{qk} satiation parameters (that is, the elements of the κ_k vector). The former baseline preference effect appears without explicit labeling (for compactness in presentation), while the latter effect is identified specifically in Table 2 through the labeling as a "satiation". A positive element in the κ_k vector has the effect of increasing the

that are suggestive and that may help inform future specifications using larger sample sizes and/or more balanced telework location splits. Besides, we estimated models by removing exogenous variable with coefficients that had less than a 90% confidence level (t-statistic of 1.645) and also a 95% confidence level (t-statistic of 1.96) to check on the robustness of the results (Lu and White, 2014, Srivastava et al., 2018, and Brennan et al. 2021 suggest this kind of robustness testing). However, using such more stringent confidence levels did not materially change (in terms of magnitudes and signs) the estimates reported in Table 2 or the predicted WPL work arrangement-commute distance relationship reported in Table 6. Important also here is that, despite having low t-statistics, the variables included in our specification had little movement in their estimated effect sizes across a whole suite of different specifications we tested, further bolstering their consideration and inclusion. Overall, though, we also admit that, in cases where the t-statistics are quite low, some caution should be exercised because these variables have a good probability (in some cases as much as 23% probability) that they are actually not of any consequence in the population.

 γ_{qk} parameter, implying that an increase in the corresponding variable diminishes satiation effects and, correspondingly, leads to an increase in days of work for WPL alternative *k*, conditional on participation in the WPL alternative *k*. Conversely, a negative coefficient indicates a decrease in the γ_{qk} parameter, implying that an increase in the corresponding variable intensifies satiation effect and, correspondingly, leads to a decrease in days of work for WPL alternative *k*, conditional on participation in the WPL alternative *k*. However, these satiation effects start taking effect from the baseline parameter, which serves as the initial reference point of marginal utility of an alternative (corresponding to the situation when no days have been invested in that WPL alternative; see Bhat, 2008). Consequently, as also discussed in Asmussen et al. (2023), the proportion of days of work allocated to each WPL alternative depends on a "combination of the baseline parameters and the satiation parameters across all WPL alternatives for that individual.".

In instances in Table 2 where some variables have no effect on specific WPL alternatives in the telework choice model (denoted by a "—"), the corresponding alternatives constitute the base category in introducing the variable effect. At the bottom of the table, we provide the covariance matrix (in differenced form for the error terms of the telework choice model, differenced with respect to the "home" work location). The results in Table 2 are effects conditional on belonging to each of the TC and CT segments. The membership model results themselves (for belonging to the TC and CT segments) are presented in Table 3. For quick identification, we label the two segments (from the standpoint of whether telework led to a residential move or not) as the "mover" segment for the TC segment (that is, telework choice led to the residential move that changed commute distance) and the "stayer" segment for the CT segment (that is, people in their respective "current" home locations with predetermined commute distances changed their telework habits). This is purely for intuitiveness as we present the results, and does not mean that those in the CT segment cannot move residences at all; just that those moves would have little to do with telework choices.

5.1. Variable Effects

Individual and Household Demographics

Among the category of individual and household demographic effects, the results in Table 2 do suggest effects of the gendered lifestyle on both telework choice and commute distance, though many of these effects are only statistically significant at the 75-90% confidence level and so are suggestive and warrant additional investigations in future studies. With that said, and contrary to many earlier studies before the pandemic, women (except those who are single parents) appear to be more predisposed to work from the office (less inclined to telework) compared to men. Nonetheless, this higher office WPL preference among non-single women does align with some recent literature analyzing telework patterns after the onset of the pandemic (see Asmussen et al., 2023; Sweet and Scott (2022)), suggesting that women, relative to men, may prefer the office environment to separate their professional and personal/family responsibilities. Or this could also be a reflection of a gendering of telework productivity as perceived by employers, resulting in employer-related pressure felt by women to show up more regularly at the work office. At the same time, non-single women tend to have shorter commute distances compared to men, consistent with earlier studies before the onset of the pandemic (Axisa et al., 2012; Blumberg and King, 2019; de Abreu e Silva and Melo, 2018; Hu, 2021). This points to women, in most households, continuing to serve as secondary breadwinners, and finding jobs close to predetermined residential locations. However, the gender effects change in the presence of a child (below the age of 18 years) in the household, with single mothers having a higher preference for remote work relative to (single and non-single) men. For single mothers who are generally the most time-poor and income-poor (see Beckman, 2022), working from home offers some level of time and cost-efficiency (in terms of savings in commute travel time and cost investments in work apparel), while also allowing more time for interaction with children (Schieman et al., 2021). Our results also point to single parents (both mothers and fathers) bearing the burden of longer commute distances compared to non-single workers (parents and not parents). This may be attributable to single parents choosing their place of residence based on proximity to their children's school, extracurricular activities, and a reliable support system, rather than proximity to their workplace, given their unique circumstances (Baker et al., 2016). Interestingly, beyond the interaction of presence of children and being a single mother, our results do not show any baseline preference effects of the presence of children. However, there are satiation effects associated with the presence of children. Parents (single or otherwise, and mother or father), with children under 12 years old, spend fewer days working from the office relative to households without children, conditional on working from the office at all. This may be attributable to familial responsibilities associated with child care and other childrenrelated activities. In contrast, parents of children between the ages of 13 and 17 may actually spend more days working from the office, again conditional on partaking in office-based work. Older children are more socially and physically independent than younger children (Schieman et al., 2021; Nguyen, 2021), facilitating parents to work from their office for both professional and social gain.

Age is another factor influencing telework and commute distance preference. Older workers (50 years or older) have a lower baseline preference to telework relative to their younger coworkers. This is presumably because older workers enjoy the concept of "workplace anchoring" i.e., working in one established location, and enjoy the structure, socialization and professional visibility of their familiar in-person office (Tahlyan et al., 2022; Robbennolt et al., 2023). Additionally, the control they have over both their workspace and over employees they may manage provides added perceived benefits of the office to older workers. Overall, teleworking may not have the same allure for older workers relative to younger workers. Besides, younger workers generally relish the flexibility, control, and freedom offered by teleworking for both their work and play hours (Asdecker, 2022; Raišienė et al., 2021; Nguyen, 2021). Further, the youngest employees (younger than 30 years old) are more likely (than their older peers) to lead a "wanderlust digital nomad" lifestyle (see Mouratidis, 2018; Stickel, 2020; and Kersting et al., 2021), explaining the higher preference of this youngest worker group for the third WPL. The satiation effects of age reinforce the disinclination of older individuals to telework, with the "not-so-young" individuals (age 40 years of older) having a lower frequency of working from home if they choose to work from home at all. With regard to age effects on commute distance, which we found statistically significant at only the 80% confidence level, the oldest cohort of workers (over 50 years of age) generally have a higher commute distance compared to their younger peers, attributable to middleaged and older individuals having distinct housing preferences that lead them to seek larger homes or quieter neighborhoods located farther away from urban centers and job locations.

The "level of formal education" effect reveals a higher inclination for home-based telework among those with a bachelor's degree or higher. Higher levels of education often lead to increased job opportunities that offer greater flexibility in work arrangements, including the option to work remotely (see López-Igual and Rodríguez-Modroño, 2020; Gaduena et al., 2022). Furthermore, employees holding a bachelor's degree or higher typically have shorter commute distances compared to those without a college degree; individuals with high levels of formal education tend to reside in neighborhoods with higher mean home values that are closer to job-dense city centers (see Robbennolt et al., 2023). Finally, within the group of individual and household demographics, employees with an annual household income of \$100,000 or more typically have a higher baseline preference for telework (both from home and a third workplace) compared to employees from lower income households. High income-earning individuals are better able to negotiating in their ability to work remotely (see for example, Astroza et al., 2020; Asmussen et al., 2023), while those in the lower income category may want to show up more often at their workplace to promote themselves (and their careers) to upper management through face-to-face interactions (Jaff and Hamsa, 2021).

Residential Attributes

Relative to an urban or suburban household, those who live in a rural area have a higher predisposition to work from the office. It has been well established that those living in low density areas are typically less likely to work from home, as the regular work office environment serves as a particularly important socialization outlet in such communities (see Castrillon et al., 2020). An additional housing-related attribute that influences work location preferences is the presence of a private study within an individual's home (Patton, 2019). Unsurprisingly, individuals who have a private study in their homes exhibit a stronger inclination to work from home compared to those without a private study offers an environment that promotes productivity and privacy, and allows individuals to work with fewer distractions and interruptions (Wöhrmann and Ebner, 2021). Proceeding to the satiation effects, residing in suburban areas is associated with a higher frequency of home-based teleworking for those who have non-zero days of teleworking from home.

Employment Characteristics

From Table 2, we observe a reduced preference for working from the office or a third workplace, in comparison to working from home, among individuals whose office is located in areas with high employment density, as also reported by Asmussen et al. (2024). Several factors may be at play here. Issues such as paid parking, local congestion, and unreliable traffic conditions, which are often prevalent in high density job areas, disincentivize individuals from commuting to their office (Foucault and Galasso, 2020). Additionally, commute distances tend to be shorter for individuals working in high employment density areas.

Employment status effects are captured both through number of work hours per week and number of work days per month. Part-time workers (those working less than 30 hours a week) are more likely to have shorter commute distances than those employed full-time (30 hours or more per week), which is a commonly found trend across the literature (Axisa et al., 2012; Sen et al., 2021; Maheshwari et al., 2022). Also, employees who work only a few days per month (1 to 5) have a higher disinclination to work from the office (higher baseline telework preference), presumably because of the flexible nature of their work. At the other extreme of work days per month, employees engaged in overtime work (working 25 days or more per month) appear to have a higher preference for the third workplace, though this finding is at only the 85% confidence level. This result may be a result of such individuals, with their demanding schedules, seeking location variety throughout the month, opting for alternative workplaces beyond their home or in-person office.

Individuals employed in the healthcare industry, relative to those employed in more nonessential sectors, have a higher baseline preference to work from the office (see also Astroza et al., 2020 and Chong et al., 2020 for similar findings). On the other hand, workers in professional

services or public administration have the lowest preference for the work office, as the nature of their jobs allows greater opportunity for telework.

Endogenous Effects

Based on the structure of our latent segmentation model, WPL choice is specified to impact commute distance only in the first latent segment (pertaining to the TC or "mover" segment). While testing the impact of WPL choice on commute distance, we employed the WPL choice variable in several functional forms including a continuous form, different combinations of grouped forms of WPL frequency, as well as interactions with built environment variables and sociodemographic characteristics. Eventually, two groupings of non-interacted home-based telework frequency percentage turned out to consistently provide the best data fit. These groupings, pertaining to the percentage of days teleworked in a month were (i) 0%-50% and (ii) 50.1%-100%. Furthermore, we employed a single variable to capture third WPL-based telework frequency effects on commute distance, corresponding to positive participation (at least one day per month) in the third WPL. This specification was because of the low number of individuals who partake in work from the third WPL. The results indicate that employees who telework from home for more than half of their work days per month tend to have longer commute distances compared to those who work primarily from the office, consistent with the majority of existing research (as discussed earlier in the literature overview). That is, for employees belonging to the TC ("mover") segment, our analysis indicates that it is not whether a person adopts home-based teleworking or not, but the frequency of telework days that turns out to be important in the effect on commute distance. In particular, our analysis suggests little difference in the causal effect of home-based telework on commute distances (if that direction of causality is the one that holds) among non-teleworkers from home and those teleworking even up to 50% of their work days from home. It is only beyond the threshold of 50% of workdays teleworked that the telework effect on commute distance becomes palpable. Further, our analysis underscores the importance of differentiating between home-based and third WPL teleworking in the impacts on commute distance. When an employee in the TC ("mover" segment) works any of their work days per month from a third WPL, the commute distance to the office is actually likely to be shorter. This may be a reflection of the supplemental usage of a third WPL for the office, rather than as a telework alternative to the office. That is, employees may work from a coffee shop or other shared spaces near their office, perhaps as an occasional getaway from the office.

Next, based on the structure of our latent segmentation model, commute distance is specified to impact WPL choice in the second latent segment (that is, the CT or "stayer" segment). Again, while testing the impact of commute distance on WPL choice, we employed the commute distance variable in several functional forms including the continuous form as well as incrementally grouped dummy variable forms. After a series of rigorous tests and based on data fit measures, the final model was specified using three commute distance groupings – 8 miles or less, 8.01 to 15 miles (which we used as the base in our final model), and longer than 15 miles (note that all individuals had a non-zero commute distance, because the survey was administered only to those with a designated physical office). As expected, our results from Table 2, consistent with the earlier literature, indicate that short commute distances (8 miles or less one-way) are associated with a higher baseline preference for office-based work. However, this commute distance effect on WPL is not monotonic. As commute distance surpasses the 15-mile threshold, employees appear to show an increased likelihood again of working from the office, though there is about a 25% possibility that this effect is a chance occurrence. While this finding needs

additional scrutiny in future studies, the observation is consistent with the right side of the histogram in Figure 3, where higher average commute distances correspond to lower levels of remote work participation. Note that this effect is after accommodating observed factors (such as older individuals having longer commute distances as well as lower telework preferences) as well as unobserved factors through covariance effects in the error terms. One plausible explanation for this causal effect of commute distance on WPL in the CT ("stayer") segment is that individuals in this segment truly enjoy their commutes to the point that longer commute distances spur them to go into the office frequently.

The Constants

The baseline preference and satiation constants in the telework MDC model component, and the constants in the commute distance regression, for both the latent segments do not have any substantive behavioral interpretations. They provide the best data fit, after accommodating for the effects of exogenous variables. In this regard, it may be tempting to draw conclusions about telework and commute distance disparities among the two latent segments based on the relative magnitudes of the constants across the two segments, especially because the effects of all exogenous variables are fixed across the two segments (for reasons discussed earlier) as well as the variances of the error terms are the same across the two segments. But the composition of the two segments itself will be different, and so the ranges of the exogenous variables in the two segments will vary. Thus, it is best not to make any conclusions based on a relative comparison of the constant values across segments.

5.1.1 Covariances

The covariance matrices for each of the latent segments are presented at the bottom of Table 2. Several elements of the covariance matrix turned out to be significant at the 85% confidence level or above, suggesting the presence of unobserved factors impacting both the commute distance and the WPL choices and reinforcing the "joint" nature of decision-making. However, the elements of this differenced covariance matrix are not interpretable without imposing a structural assumption on the matrix (as several non-differenced matrix may be constructed from the same differenced matrix). In this regard, if we assume that the error term for the "work from home" alternative has a low variance and is uncorrelated with all other error terms, the implication would be that unobserved factors that increase the baseline preference for office work also increase the baseline preference for the third WPL (consistent with the notion of the supplemental usage of a third WPL for office-based work, as discussed earlier). Also, for the TC ("mover") segment, unobserved factors that increase the baseline preference for the third WPL also increase commute distance (though this covariance is very mild), while, for the CT ("stayer") segment, unobserved factors that increase the preference for the work office lowers commute distance. This last association may be because, as discussed earlier in Section 1, an extroverted person may like to maximize social time within and beyond the office working space, thus deciding to work from the office more (telework less) and also stay close to the workplace (short commute).

5.1.2 Characteristics of Latent Segments

Table 4 provides the effects of individual and household-level demographics and residential attributes on segment membership, with the CT segment (the "stayer" segment) as the base segment (these estimates correspond to the μ_1 vector for the first TC segment with $\mu_2 = 0$ for the

CT segment for identification, as discussed in Section 4.2). Thus, a positive (negative) sign for a variable in the TC segment in Table 3 indicates that individuals with the variable characteristic are more (less) likely to be assigned to the TC segment relative to the CT segment.

Two lifecycle variables have an impact on the membership of latent segments. First, according to our results, though only at an 80% confidence level, employees in households with children aged 12 years or less appear to be less likely to belong to the TC ("mover") segment, a plausible reflection of a well-established support network for pre-adolescent childcare. In contemporary societies, decisions regarding residential location and the working arrangements of employed parents have become joint decisions, dependent on the age of their children. In addition to the presence of children, the presence of a partner also has a membership effect, with those living with their partner less likely to be in the TC ("mover") segment, presumably due to the additional considerations and compromises that arise from relocation-related shared decision-making.

In terms of age, younger employees (under 30 years) are more likely to be in the TC ("mover") segment relative to their older peers. While we recognize that this finding may be a result of a 16% chance occurrence, this seems rather intuitive because younger adults often have fewer commitments, such as families or mortgages, and may be more willing to transplant themselves to another locality based on their teleworking opportunity and frequency. Employees from lower-income households earning less than \$50,000 per year may also belong more to the TC ("mover") segment. These households often have tighter budgets, and may choose their residential location based on affordability or the ability to commute to work economically (see Robbennolt et al., 2023).

Finally, workers residing in suburban neighborhoods are more represented in the TC ("mover") segment relative to those residing in urban and rural neighborhood. This may be attributed to the COVID-19 triggered phenomenon known as the "suburban flight" (urban and rural dwellers descending onto suburbia), which continues to be observed even today due to the widespread shift to telework. While urban dwellers look for better housing alternatives in and around the suburban neighborhoods, rural dwellers seek better opportunities for recreational and social interaction in the more accessible suburbs.

The constant reported toward the end of Table 3 has no substantive interpretable meaning, but provides an adjustment factor. The negative value, however, suggests a smaller segment size of the TC ("mover") segment compared to the CT ("stayer") segment, as we discuss below. Another important point to note here is that the causal direction at play (that is, segment to which an individual belongs) is not impacted by any employment characteristic (such as number of hours of work, number of workdays per week, occupation sector, or company size). This is evidenced by the lack of any employment attribute in Table 3. The indication then is that whether an individual is a "mover" or a "stayer" is more determined by basic lifecycle and related demographic attributes, not by employment attributes. This is consistent with recent housing literature (see, for example, Lei and Liu, 2022 and Robbennolt et al., 2023) that suggests that housing choices in a post-COVID era are increasingly lifecycle- and lifestyle-driven rather than tied to work-related characteristics.

5.2. Latent Segment Sizes and Characteristics

Using the latent class assignment for each individual, we are able to determine the size of each segment through the procedure developed by Bhat (1997). The last row of Table 3 provides the size of the latent segments in terms of percentages; 20.3% of the sample is estimated to belong to

the TC ("mover") segment, while 79.7% is estimated to belong to the CT ("stayer") segment. This indicates that, even in the COVID-era, the causal direction is primarily in the form of residential locations being determined without regard to teleworking arrangement, and then individuals teleworking based on commute distance. However, there is a sizeable fraction of the sample for whom teleworking opportunity and frequency impacts residential choice. That is, unlike earlier studies that consider a single causal directionality, our study indicates heterogeneity in the population in the causal direction of effect.

Table 4 provides a detailed overview of the variations of the two segments by demographic attributes in terms of actual percentages (using the procedure in Bhat, 1997). The first broad numeric column "Percent within segment" provides the split of a variable within each segment; Thus, within the first segment where WPL preference impacts commute distance (CD), 60.7% percent are women and 39.3% percent are men. Within the second segment where CD impacts WPL preference, the corresponding split between women and men is 59.5% and 40.5%, respectively. Another way to present the segment sizes is the entries corresponding to the broad column entitled "Percent within attribute". These columns provide the binary split between the two segments for each of the individual variables (or attributes). Here we observe that only 20.7% of women belong to segment one (compared to 19.8% of men), while 79.3% of women belong to segment two (compared to 80.2% of men), which nearly align with the segment size splits toward the bottom of Table 3 (20.3% of the sample is part of segment one, while 79.7% is likely a member of segment two), displaying that there is little difference within the gender attribute across the two segments. Other entries in the table pertaining to other exogenous variables can be interpreted similarly. While the latent segmentation results from Table 3 only suggest a directional influence of the sociodemographic attributes' impact on segment membership, the segment size values from Table 4 help us better understand the composition of each segment. The percentages of sociodemographic split "within segment" and "within attribute" is conformable with the results obtained in Table 3; as one would expect, the maximum shifts and changes are observed for the sociodemographic variables that turned out to be significant in the membership model (see, for example, the rows pertaining to the presence of children, age, household income and population density variables), while other sociodemographic attributes have little variation across the segments.¹⁰

Overall, our results do suggest that home-based telework will inevitably lead to urban sprawl (even if the monthly commute VMT may reduce at moderate to high telework levels, as discussed later in Section 5.4). Our results reveal that about a fifth of employees (a predicted 20.3% of our sample) are expected "movers", and are likely to base their residential location choice directly on telework opportunities. Thus, telework may result in the abandonment of urban residences near employment-dense areas, with a corresponding increase in the demand for new, rapidly constructed, and often low-quality tract housing developments in suburban or rural areas. While private residential developers may benefit from this trend, increased residential sprawl leads to several challenging issues for many other populations, due, but not limited to, (a) increased infrastructure costs to preserve, modernize and expand highways to the outskirts of cities and the growing suburbs, (b) increased air and other environmental pollution from increased vehicle emissions and the destruction of natural habitats, (c) lack of affordable housing and decrease in

¹⁰ Table 4 shows variations in segment memberships even for variables that do not explicitly appear in the latent membership model of Table 3. This is because the exogenous variables appearing in Table 3 are differentially associated with the exogenous variables not appearing in Table 3. This causes segment membership differences for all variables in Table 4.

neighborhood quality both within increasingly abandoned cities as well as the growing suburbs, (d) social fragmentation and segregation, (e) increased energy consumption, (f) underserved schools and other public services, and (g) misdirected expenditures of tax money. For instance, rather than investing in improving existing communities, there has been a recent shift in U.S. tax funds towards subsidizing new sprawling suburban communities. This shift has resulted in cities and counties spending millions of dollars on new schools, infrastructure for roadways, water and sewer lines, and fire and police protection, while placing higher tax burdens on remaining urban residents. It is crucial for governing organizations to recognize the implications of residential sprawl, so that appropriate budgets can be allocated accurately and fairly across all types of local and state regions as the demographic composition of each area begins and continues to change.

Based on individual and household demographics, "movers" inducing residential sprawl are likely families with children (families with younger children in particular), employees younger than 50 years (with the highest likelihood for employees less than 30 years old), and those with an annual income lower than \$100,000. Given that these households are the primary contributors to residential sprawl, governing and planning bodies may implement effective measures to discourage residential relocation. To retain those with limited budgets, such as single parents and low-income households, targeted initiatives such as rent or home insurance subsidies can be implemented in urban centers, incentivizing them to remain in their current urban dwellings. In contrast, younger employees can be influenced by city-wide marketing campaigns that emphasize the advantages and aesthetic appeal of residing within the city limits. However, it is crucial to acknowledge that these movers often come from communities already grappling with equity challenges. While their departure from urban centers can have repercussions for the entire community, their settlement in newly developed suburban areas can give rise to specific issues for the movers themselves. The expansion of homes and communities further apart amplifies the cost of delivering essential services, potentially leading to less targeted and responsive provisions for the residents of these suburban areas, who may bear the burden of higher local, state, and federal taxes. Consequently, households already lacking financial resources find themselves faced with increased tax obligations, further straining their finances. Moreover, availability of essential amenities such as quality public schools and community services need to be ensured by the government for marginalized sections of the society. Therefore, not only must the government devise strategies to discourage suburban immigration, they must also strive to meet the basic needs of these "movers" (in case, they decide to migrate to the suburban regions) from an equity standpoint.

5.3. Data Fit Measures and Robustness Discussion

The goodness of fit for our proposed joint latent segmentation model may be compared with other restrictive models to evaluate the performance gain in data fit. Specifically, we compare our proposed model with three other restrictive versions: (i) a model that considers the jointness in the telework and commute distance outcomes, but ignores the latent segmentation (Joint Unsegmented Model), (ii) a model that ignores the jointness in the telework and commute distance outcomes, but considers the latent segmentation (Independent Segmented Model), and (iii) a model that ignores both the jointness and the latent segmentation (Independent Unsegmented Model). Since these models are nested within our proposed "Joint Segmented Model", their performances can be compared using the likelihood ratio test.

The data fit measures are provided in Table 5. The log-likelihood at convergence for our proposed model is lower than the three restrictive versions. The likelihood ratio tests (when the

proposed model is compared to the three restrictive versions) yield values that are much higher than the critical chi-squared table values at any reasonable level of significance (at the respective degrees of freedom). The adjusted rho-squared value also favors our proposed model relative to the other three models, indicating a superior fit.¹¹

To further examine data fit issues beyond the rather difficult-to-interpret likelihood fit measures, we use the estimated parameters to develop non-likelihood based data fit metrics by predicting ln(commute distances) as well as telework choices and compare with the observed values in the sample. To do so, first, within each of the causal direction segments and for each individual, we use a minor modification of the Pinjari and Bhat (2021) procedure for MDC model forecasting (using draws from the multivariate normal distribution characterizing the errors in the telework split MDC and ln(commute distance) model components) to predict telework splits as well as ln(commute distances). Then, employing the estimated latent segmentation parameters too, we next predict, for each individual, the telework splits and ln(commute distances) across both the causality segments. Using these individual forecasted values, we compute aggregate data fit metrics and compare with the observed data. We do so at three different outcome levels. At the first level, we compare the predicted number of individuals with positive participation in each of the seven possible discrete combinations of work place locations (home only, office only, third WPL only, home and office, home and third WPL, office and third WPL, and all three locations) versus the actual number of individuals with positive participation in the corresponding discrete combination. We achieve this by computing the percentage prediction error in each of the seven possible combinations, and then compute a weighted average percentage error (WAPE) across all the seven possible discrete WPL combinations (we will refer to this metric as the discrete *participation WAPE*). We also compute, at this discrete level, the probability of correctly predicting the right discrete category combination (among the seven possible WPL locations) for each individual, and then compute a weighted average probability of correct prediction or WAPCP (referred to henceforth as the WAPCP of discrete participation). At the second level, we compare the predicted versus observed aggregate splits for fraction of days at each of the three distinct work place locations and compute the WAPE at this continuous WPL split level (we will refer to this metric as the average fractional split of days worked from each WPL WAPE). Finally, we compute the average of the predicted ln(commute distances) among individuals predicted to fall within each of the seven discrete WPL combinations, and compare with the average of the actual observed In(commute distances) within the actually chosen WPL combination, followed by the computation of a WAPE measure across all seven combinations (the average ln(commute distance) WAPE).

The resulting non-likelihood based data fit metrics are presented toward the bottom of Table 5. These metrics indicate reasonable WAPE and WAPCP values across the board of the different models. To provide a benchmark of the performance of the models, a simple sample share

¹¹ Note also that the "seemingly low" rho-squared value of our proposed model should not be interpreted as poor fit. As McFadden (1977; page 35) states in the context of simple multinomial/binary choice models, "Those unfamiliar with the rho-squared measure should be forewarned that its values tend to be considerably lower than those of the R-squared index... For example, values of 0.2 to 0.4 for rho-squared represent *excellent* fit." In a more recent paper, Giselmar et al. (2018) indicate that, in simple binary logistic models, rho-bar squared values in the 0.11-0.32 represent very good model fit with sample sizes of more than 200. In this context, rho-bar squared values of the order of 0.05 in way more advanced joint model systems such as the latent segmentation joint MDC-continuous model systems must not be dismissed as "poor fit". As further indicated by Ben-Akiva and Lerman (1985; page 91), "Values of rho-bar squared will depend on the type of model being estimated. The measure is most useful in comparing two specifications developed on the exact same data".

assignment to the seven discrete combinations would result in a WAPCP of 0.412, while a naïve models that assigns the average of commute distance to all individuals results in a WAPE of 31.33%. The corresponding Table 5 values are much better. But, again, the superior fit of the proposed model relative to the other three models is also clear (and consistently so) across all the fit metrics. In particular, even the closest competitor to the proposed model (that is, the independent segmented model) has a WAPE that is higher than the proposed model by about 20% on the dimensions of discrete participation and average fractional split of days worked from each WPL, and by about 11% for the commute distance dimension. Interesting also is that, at least in our empirical context, the independent segmented model performs better than the joint unsegmented model, suggesting that, between considering the population heterogeneity in the causal direction of effect in the telework-commute distance relationship on the one hand and the "spurious" association between telework choice and commute distance on the other, the former effect is more important. However, the results also emphasize that both these issues will generally be at play and it is best to expressly accommodate both of these when investigating the interplay between telework and commute distance.

Another important note here beyond data fit relates to considerations of robustness of our estimation and the model results. There are two main components to the model, one being the MDC component (of telework choice and the continuous commute distance dimension) and the second being the latent segmentation assignment component that is overlaid on the MDC component. The MDC component employed here is the same as the framework of Bhat's (2015) Generalized Heterogeneous Data Model (GHDM) and its subsequent extensions (Bhat et al., 2016). In prior studies, we have conducted extensive simulation analyses to demonstrate the replicability and robustness of this MDC modeling approach. The issue that remains then is the robustness and stability after overlaying the latent segmentation component over the MDC component. As discussed at length in Bhat (1997), latent segmentation models of the type used here can lead to convergence instabilities/difficulties, and it is not uncommon to use expectationmaximization (EM) type approaches in such situation (as proposed by Bhat, 1997 in the context of discrete choice outcomes). While some authors have found EM approaches to be helpful, some others (see, for example, Greene, 2016) have not found much value in these EM approaches. In our estimation, we did not have any need to resort to EM approaches because there was stability and rather smooth convergence directly with the maximum likelihood inference procedure. Part of the reason is perhaps because we provided good start values by (a) first estimating purely exogenous variable models without any latent segmentation, (b) then feeding these values along with zero coefficients on the endogenous outcome effects on each other in the latent segmentation model, and (c) starting the iterations with close to zero values (but not zero values) for the coefficients in the latent segmentation assignment model component (see Table 3 result; for the iterations to unfold, the latent segmentation assignment component start values must be slightly away from zero). Once the entire model was estimated, we ran a number of different estimations with quite substantial perturbations from the original start values provided (though one needs to be careful not to get too far from the original non-latent class estimation coefficients; the critical issue here is that the probability values for each individual need to be able to be computed at the provided start values at the first iteration). In our many runs, the convergence was to the same point (while this does not preclude the possibility of a local optima, as is the case with literally all models except the simple multinomial logit model, this kind of checking provides some confidence regarding the robustness and stability of our modeling approach). Finally, as discussed in footnote

9, the overall estimated effect of each variables changed little across a range of different variable specifications, providing additional confidence regarding the robustness of the model results.

5.4. Effects of WPL Arrangement on Commute Distance and Monthly Commute VMT

The estimation results in Sections 5.1, 5.2, and 5.3 offer insights into the effects of exogenous variables on (a) the baseline preferences and satiation parameters for the WPL preference, (b) the (ln) commute distance outcome, and (c) the latent segmentation assignment. But they do not establish an intuitive picture of the true magnitude effect of telework arrangements on commute distance. On the other hand, transportation planners and professionals are interested in understanding the tangible travel demand impacts of telework behavior on commute vehicle miles traveled (VMT). Toward this end, we use the findings from our model to estimate the expected commute VMT of travel for different workplace location arrangements across an entire month.

As earlier, let x_{q1} , x_{q2} , and x_{q3} be the WPL frequency proportions (over a period of one month) for the work from home, work from the office, and work from a third WPL, respectively, for individual q. Let \tilde{y}_q be the continuous variable denoting the logarithm of the commute distance for individual q. Then using our joint model, one can compute the density function for \tilde{y}_q at a specific continuous value of t for an individual belonging to segment h and with WPL frequency proportions x_{q1} , x_{q2} , and x_{q3} as follows:

$$f_{\bar{y}_{q},q,h}(t \mid (x_{q1}, x_{q2}, x_{q3})) = \frac{p_{q,h}(\bar{y}_{q} = t, x_{q1}, x_{q2}, x_{q3})}{p_{q,h}(x_{q1}, x_{q2}, x_{q3})},$$
(16)

where the numerator is the probability density function at a specific value t of ln(commute distance) and specific proportions of work from the three WPLs, and the denominator is the probability density function at the specific proportions of work from the three WPLs. The corresponding mixture density function for the individual at the specific continuous value t, given our latent segmentation scheme, as:

$$f_{\bar{y}_{q},q}(t \mid (x_{q1}, x_{q2}, x_{q3})) = \sum_{h=1}^{2} \pi_{qh} f_{\bar{y}_{q},q,h}(t \mid (x_{q1}, x_{q2}, x_{q3})),$$

where π_{qh} is the probability of individual *q* belonging to segment *h*. Next, the expected value of the logarithm of the commute distance for individual *q*, given the WPL distribution, may be written as:

$$\varpi_{q} = E(\breve{y}_{q} | (x_{q1}, x_{q2}, x_{q3}))$$

$$= \int_{-\infty}^{\infty} t \times \Big[f_{\breve{y}_{q}, q}(t | (x_{q1}, x_{q2}, x_{q3})) \Big] dt$$

$$(17)$$

And, the variance of the logarithm of the commute distance, given the WPL distribution, may be written as:

$$\sigma_{q}^{2} = Var(\tilde{y}_{q} | (x_{q1}, x_{q2}, x_{q3}))$$

$$= \left(\int_{-\infty}^{\infty} t^{2} \times \left[f_{\tilde{y}_{q}, q}(t | (x_{q1}, x_{q2}, x_{q3})) \right] dt \right) - \left(\int_{-\infty}^{\infty} t \times \left[f_{\tilde{y}_{q}, q}(t | (x_{q1}, x_{q2}, x_{q3})) \right] dt \right)^{2}$$
(18)

Then, the expected commute distance for individual q for the given WPL proportions is:

$$\delta_q \mid (x_{q1}, x_{q2}, x_{q3}) = \exp\left(\varpi_q + \frac{\sigma_q^2}{2}\right) \tag{19}$$

From the above, we are able to estimate the overall commute related VMT per month, given the number of workdays of the individual and the proportions. Let the total number of work days for individual q be d_q (this is observed exogenously). The total expected monthly two-way commute VMT, given the WPL proportions, is:

$$VMT_{q} | (x_{q1}, x_{q2}, x_{q3}) = 2 \times d_{q} \times x_{q2} \times \left[\delta_{q} | (x_{q1}, x_{q2}, x_{q3}) \right]$$
(20)

This value can be computed for each individual, after predicting the optimal allocations for x_{q1}, x_{q2} , and x_{q3} from the MDC model (which can be undertaken using the forecasting procedure of Pinjari and Bhat, 2021) and providing the number of days of work per month d_q as input.

In the current paper, to keep our analysis and presentation tractable, while also getting a sense of the impact of teleworking on commute VMT, we focus on ten most commonly occurring proportions (percentages) of telework arrangements in our sample (rounding off to reasonable values) and undertake the analysis by applying these arrangements to <u>every worker</u>. In this sense, all the scenarios presented here are counterfactuals, representing what would happen if each individual in the sample had a specific scenario of telework arrangement (rather than the individual's current actual telework arrangement), and comparing those results with the alternative counterfactual of each individual working full-time from the office. Also, since a majority of the sample did not work from the third workplace (only 5.2% of the sample ever worked from a third workplace), we only select two cases in which there is any third workplace participation. Further, in our subsequent analysis, we took the integration in Equations (17) and (18) from the minimum to the maximum (in the natural logarithm of miles), based on the distribution of one-way commute distribution in the sample (this precludes long one-way commute distances, such as 500 miles, that would be infeasible within the context of a working day; note that our sample was restricted to individuals who had a designated physical office in the local area at the time of the survey).

The selected combinations along with their expected commute distances (based on Equation (19) above for each individual, and then averaged across all individuals) are presented in Table 6. The first broad set of columns present the workplace location combination, followed by the expected commute distance for that combination and the expected commute distance when 100% of the month is worked from the office. The next column presents the percent increase in expected commute distance as the WPL arrangement shifts from 100% at the office to a different WPL combination, followed by an analysis of overall changes in monthly two-way commute VMT in the last three columns of the table. Focusing first on the expected one-way commute distance is for a shift to 100% home-based telework (see the eighth combination in the table). Specifically, relative to 100% office work, shifting to 100% home-based remote work increases average commute distances by nearly 65% (from 12.5 miles to 20.6 miles; see the fifth and sixth numeric columns of the table for combination #8).¹² Also, to be noted is that home-based telework of any amount is

¹² The value of our multivariate latent framework becomes clear here. As opposed to Figure 2, where the descriptive statistics indicate a reduction of the one-way commute distance from 13.25 miles to 12.06 miles when going from 100% office-based work to 100% telework, our results in Table 6 indicate the clear increase in commute distance with telework after controlling for exogenous effects, unobserved factors, and latent segmentation. This is because of the improved ability to uncover the "true" relationship between telework frequency and commute distance.

associated with an increase in commute distance. This may feel at odds with the finding in Section 5.1 that the causal effect of telework on commute distance kicks in only at \geq 50% home-based telework, but note that the direct causal effect applies only to the small-sized TC segment. In addition to this effect, there is also the increased work-from-office for those with low commute distances (<8 miles) in the large-sized CT segment.¹³ Thus, the direction of causal effects in each segment, along with the negative unobserved covariance effect between the office WPL and commute distance (as discussed in Section 5.1.1), all reinforce and lead up to the increased commute distance associated with any non-zero home-based telework arrangement. In this context, consider the impact of teleworking (primarily from home) on commute distance when employees work remotely once a week (4.4 days a month), which appears to be a rather popular hybrid work arrangement. Indeed, a recent report shows that, today, the average employee works 5.8 remote workdays per month (Flynn, 2023). In our analysis, a once-a-week teleworking arrangement would fall within the range of 10% to 20% remote work. Our findings reveal a shift of 17.6% (for 10% home-based and 90% office) and 21.6% (for 20% home-based and 80% office) in terms of increased commute distance. Overall, considering only home-based telework and office-based work, there is an increase of about 0.5 miles (about 4%) in expected one-way commute distance for every 10% increase in home-based telework from 0% telework to 80% telework. But, for an increase from 80% to 90% home-based telework, the expected one-way commute distance increases by 0.8 miles (6.4% increase), while for an increase from 90% home-based telework to 100% home-based telework, the expected one-way commute distance jumps by 1.7 miles (corresponding to a 13.6% increase).

For the cases with a positive participation percentage for the third workplace that takes away from home-based telework (the last two row combinations in the table), the results show that third WPL telework leads to a modest reduction in the commute distance in comparison to homebased telework (compare combinations 1 and 9, and 2 and 10). This decrease in commute distance with an increase in third WPL percentage that draws away from home-based telework is a combination of two reinforcing effects: (1) the negative effect of any work from the third WPL on commute distance from segment 1 (which overshadows the mild increase in commute distance associated with increased third WPL telework arising from the positive correlation in this first TC segment between third WPL telework and commute distance, as discussed in Section 5.1.1), and (2) the decrease in commute distance associated with increased third WPL telework arising from the relatively strong correlation in the second (and dominant) CT segment between third WPL telework and commute distance, also as discussed in Section 5.1.1).

Of course, the above changes in one-way commute distance based on teleworking provides only half the picture. The more critical issue is the contribution of the commute distance to monthly commute VMT. Does a frequent office goer with low commute distance produce an overall higher monthly VMT, or a rare office goer with high commute distance generate higher VMT? To address this issue, one needs to consider the total commute VMT effect based on teleworking, as provided in Equation (20). For this analysis, and given that the third WPL tends to be close to the residence of individuals, we assume that any commute VMT to the third WPL is zero (we need to do so because we do not have information on WPL locations, and these could vary from day to day). The results are provided in the last three columns of Table 6. The individual-level expected monthly two-way commute VMT, as obtained from Equation (20), is averaged across all sample

¹³ There is also the increased work-from-office for those with high commute distances (>15 miles) in the large-sized CT segment, but this effect is much weaker than the increased work-from-office for those with low commute distances (≤ 8 miles).

individuals to obtain the "Average (Across Workers) of Expected Monthly 2-way Commute VMT per worker for Given Combination". Low levels of telework (to be precise up to 18.6% from home) are associated with added VMT generation comparable to that of no telework at all, due to the combination of a lengthened average commute distance and high participation at the office WPL. In particular, the 10% telework from home scenario is associated with a higher monthly commute VMT (5.8% higher) relative to the no telework-from-home scenario. This could be a cause for concern for transportation professionals, especially as many companies continue to encourage a hybrid structure for WPL arrangements. Our results suggest that limited remote work arrangements cannot be a tool for VMT reduction, and may even exacerbate the traffic congestion situation. This is particularly so because individuals with limited remote work flexibility (say one day a week) are likely to choose Monday or Friday as their remote work day, which, when combined with the longer commute distances, would imply higher VMT in this group on midweek days. Commute VMT reductions become tangible only at about 30-40% telework (about two days of telework per week for full-time workers with 22 workdays per month). In general, between 0-40% home-based telework and the remaining at the office, there is a 10% commute VMT reduction for each 10% increase in home-based telework. Beyond 40% telework, a 60% work from home scenario results in an estimated 45.3% commute VMT reduction, while 80% or greater teleworking arrangement leads to a reduction of 71.0% in monthly commute VMT. That is, in the range of 40% to 80% home-based telework, there is about a 13% commute VMT reduction for every 10% increase in home-based telework. And in the range of 80%-100% home-based telework reduction, there is about a 15% commute VMT reduction for every 10% increase in home-based telework. Based on these findings, moderate to high levels of home-based teleworking indeed can serve to reduce commute VMT, and it would behoove transportation planning agencies to incentivize such levels of telework frequency whenever feasible and possible. At the same time, through persuasion or even regulation, cities and state/federal labor agencies may want to discourage employers from providing a telework option involving only very limited days of a week or a month.

Another observation from the one-way commute VMT statistics in the last two rows of Table 6 is that channeling some level of telework away from home to a third WPL reduces commute VMT (albeit rather modestly). Specifically, when comparing combinations 1 and 9, and also 2 and 10, we notice a drop of about 1.3-1.4% in VMT for the combinations with the third WPL relative to the comparative telework intensities solely from home. One approach for developers and land-use planners to foster third WPL participation may be to initially focus marketing and new construction of facilities (such as WeWork centers or spacious coffee shops and cafes) in areas where "stayers" predominantly reside - in suburban regions. Suburban dwellers may be less likely to relocate as a result of their newfound remote work opportunities, and may appreciate a third WPL when it is introduced within their established community. A second approach for fostering growth in third WPL telework participation is to target all income brackets. Currently, higher-income households exhibit a greater propensity for remote work from third workplaces compared to lower-income groups (as observed in Table 2). Therefore, management of third WPLs must offer affordable monthly membership options and provide amenities equivalent to those found in traditional office spaces. By catering to the needs of lower-income employees, who typically rely on such amenities in their regular workplaces, broader participation across all income groups can be achieved. Overall, these findings reinforce the importance of considering differential adoption and locations of teleworking preferences in transportation demand forecasting models at a disaggregate level to accurately represent the evolving commute VMT.

6. CONCLUSION

In the current landscape, there is a widespread consensus that telework adoption will be at levels higher than observed prior to the COVID-19 pandemic. This shift is primarily driven by reduced resistance from employers and the firsthand experiences of employees with remote work arrangements. Consequently, understanding the evolving relationship between telework and commute distance is of significant interest not only to transportation professionals but also to urban and land-use planners, geographers, labor market researchers, as well as experts in air quality and environmental studies. In this paper, we model the workplace location (WPL)-commute distance decision-making process as a package choice to account for unobserved factors as well as use a latent segmentation approach to recognize that some individuals may choose their telework arrangement first and then choose their commute distance (this forms the TC segment in our analysis), while others may adopt a reverse causal behavioral process (the CT segment). The data for the analysis is drawn from a 2022 survey of Texas residents who are workers, have a designated office site (even if they never go in to the office), and have the option of working away from the office. The latent segment modeling methodology combines a multiple discrete-continuous probit (MDCP) model for telework adoption/intensity in terms of the proportion of monthly workdays worked (a) from home, (b) from the office, and (c) from a variable third workplace, with a loglinear regression model for commute distance (that is, home-to-office distance). To our knowledge, the resulting sample selection-type model, based on a multiple discrete-continuous sample selection mechanism, is a first in the literature.

This model structure is broadly applicable to a variety of contexts with a joint MDCcontinuous model (to capture unobserved associations through error correlations) combined with a latent segmentation scheme (to capture different causal directions of effect between the MDC outcome and the continuous outcome). Examples of such applications include linking telework MDC choice with other continuous travel measures (such as local non-work trip mileage and long distance trip mileage), vehicle body/fuel type MDC choice (types of vehicles owned by a household and motorized mileage on each vehicle) with non-motorized mode mileage, non-sleep activity participation MDC (types of non-sleep activities participated in during some other time unit and the corresponding time investments) and sleep duration, physical activity participation MDC (types of physical activities participated in during some time unit and the corresponding time investments) and health biomarkers (such as body mass index). In each of these cases, the latent segmentation approach recognizes that, after accounting for the jointness in the outcomes, the causal direction between the outcomes may vary across individuals. For example, in the last case, high physical activity participation may lead to good health biomarker metrics for many individuals, but it is also possible that poor health biomarker metrics lead to lower physical activity participation in the first place for other individuals. Of course, our model system can also be easily be extended to cases where the continuous outcome is replaced with a count outcome or an ordinal outcome or even a nominal outcome (leading to latent segment MDC-count models or latent segment MDC-ordinal models or latent segment MDC-nominal outcome models). Moreover, this methodology could be extended to different empirical contexts within the emerging technology field. Teleworking, while prevalent even before COVID-19, has assumed a distinctive role in the post-pandemic era, due to the widespread availability of broadband communication technology in households and alternative work settings, all of which enable smooth communication, access to cloud-based applications, and video conferencing. Similarly, online shopping has grown as an additional (to in-person shopping) channel option for daily activities and appointments. Such

changes in time use or rearrangement of activity participation due to emerging technology's digitizing influence impact travel patterns and can be modeled using a framework similar to ours. Furthermore, advancements in vehicle technology, such as electric and autonomous features, are also reshaping residential location choices, commute distances, and activity behaviors. In this setting, our latent segmentation framework can be employed to investigate whether the adoption of autonomous vehicle features leads to an increase in annual VMT, or whether individuals who drive more are predisposed to be the early adopters of advanced autonomous vehicles. Such an application would mirror a study by Asmussen et al. (2022) on partially automated vehicles, where only one direction of effect was considered—specifically, whether Partially Automated Driving (PAF) technology affects annual VMT. Exploring the potential heterogeneity of causal pathway effects, as conducted in our current study, could offer valuable insights into the complex interplay between technology adoption and travel behavior.

The results from our proposed model indicate that men, younger employees, and those with higher formal education and income are more likely to be home-based teleworkers, while women (who are not single parents), older employees, rural dwellers, those with offices in medium-to-low job density areas, and healthcare workers prefer working from their office. Employees likely working from third WPLs include younger adults and those from higher income households. Finally, as expected, relative to having a commute distance between 8.01 and 15 miles, those with shorter commutes are high frequency office workers, though unexpectedly, those with commutes longer than 15 miles are also high frequency office workers. With respect to the commute distance outcome, our results indicate that men, single parents, older employees, those with offices in medium-to-low density areas, and full-time workers, are likely to have longer commutes than their fellow counterparts. Additionally, high frequency home-based teleworkers (working over 50% of their month from their homes) are expected to have longer commute distances, while any participation in the third WPL leads to a shorter commute. Results from the latent segmentation membership model and segment profile analysis demonstrate that about a fifth of the sample belong to the TC segment, while a larger 4/5th portion falls within the CT segment. Such a split highlights the bidirectional causal linkage between telework choices and commute distance. Individuals in the TC segment are likely unmarried, from households with older children, are younger than 30 years old, have an annual household income lower than \$50,000, and reside in an urban or rural region.

We further quantify the impact of teleworking on commute distances by computing the expected commute distance for several commonly observed WPL distribution preferences. Using such a metric, we note that a shift from 100% office-based work to 100% home-based telework increases average commute distance by 64.8%, while working from a third WPL for any portion of a month has little impact on commute lengths. In terms of monthly commute VMT, working from home for less than 20% of the month (about once a week) has negligible impact on VMT reduction, while working from home for more than 30-40% of the days (about twice a week) starts to have a reasonable impact. Additionally, a shift from office-based work to third WPL telework significantly decreases average monthly commute VMT.

As with any research investigation, there are several directions for future research. While this study explores multiple dimensions of commute distance, it was limited to only the commute distance to the office. Commute distance to a third WPL was not considered, and therefore limited information is available on overall VMT changes due to third WPL participation. Further, because the third workplace location may change at each occasion, there may even be multiple third WPL commute distances. These issues warrant additional attention in future studies. Additionally, while

the sample size for this research is reasonable (824 respondents), a larger sample sizes in future work would be better able to more definitively characterize (or at least confirm) the effects of variables found in the current study, and the relative sizes of the two causal pathway segments. It would also help estimate additional interactions among variables that could not be identified in the current study. Despite these limitations, our study provides a methodical approach to explore the nuanced and intricate interplay between telework arrangements and residential location decisions. The findings provide important insights into the commute VMT effects of teleworking, which can serve as valuable input to land-use and traffic management policies.

ACKNOWLEDGEMENTS

This research was partially supported by the U.S. Department of Transportation through the Center for Understanding Future Travel Behavior and Demand (TBD) (Grant No. 69A3552344815 and No. 69A3552348320). The survey data used in this paper was collected as part of a Texas Department of Transportation (TxDOT) funded project. The authors are grateful to Lisa Macias and Sarah McGavick for help in survey instrument design, and Lisa Macias for additional help in formatting this document. Three anonymous reviewers provided useful comments on an earlier version of this paper. Finally, the authors thank Michiel Bliemer for feedback that helped in preparing the revised version.

REFERENCES

- Adobati, F., and Debernardi, A. (2022). The breath of the Metropolis: Smart working and new urban geographies. *Sustainability*, *14*(2), 1028.
- Agarwal, M. (2023, September). *The Transformational Impact of Technology on Remote Work* [Post]. LinkedIn. <u>https://www.linkedin.com/pulse/transformational-impact-technology-remote-work-mayank-agarwal/</u>.
- Alonso, A., Monzon, A., and Wang, Y. (2017). Modelling land use and transport policies to measure their contribution to urban challenges: The case of Madrid. *Sustainability*, 9(3), 378.
- Anderson, M., Guido-Sanz, F., Talbert, S., Blackwell, C. W., Dial, M., McMahan, R. P., and Diaz, D. A. (2022). Augmented Reality (AR) as a prebrief for acute care simulation. *Clinical Simulation in Nursing*, 69, 40-48.
- Andrey, J. C., Burns, K. R., and Doherty, S. T. (2004). Toward sustainable transportation: Exploring transportation decision making in teleworking households in a mid-sized Canadian city. *Canadian Journal of Urban Research*, 13(2), 257-277.
- Angueira, J., Konduri, K. C., Chakour, V., and Eluru, N. (2019). Exploring the relationship between vehicle type choice and distance traveled: a latent segmentation approach. *Transportation Letters*, *11*(3), 146-157.

Aptech (2022). Gauss (21) [Computer Software]. Aptech Systems.

- Asdecker, B. (2022). Travel-related influencer content on Instagram: How social media fuels wanderlust and how to mitigate the effect. *Sustainability*, 14(2), 855.
- Asmussen, K. E., Mondal, A., and Bhat, C. R. (2022). Adoption of partially automated vehicle technology features and impacts on vehicle miles of travel (VMT). *Transportation Research Part A*, *158*, 156-179.
- Asmussen, K. E., Mondal, A., Bhat, C. R., and Pendyala, R. M. (2023). On modeling future workplace location decisions: An analysis of Texas employees. *Transportation Research Part A*, 172, 103671.

- Asmussen, K. E., Mondal, A., Batur, I., Dirks, A., Pendyala, R. M., and Bhat, C. R. (2024). An investigation of individual-level telework arrangements in the COVID-era. *Transportation Research Part A*, 179, 103888.
- Astroza, S., Garikapati, V. M., Pendyala, R. M., Bhat, C. R., and Mokhtarian, P. L. (2019). Representing heterogeneity in structural relationships among multiple choice variables using a latent segmentation approach. *Transportation*, 46, 1755-1784.
- Astroza, S., Tirachini, A., Hurtubia, R., Carrasco, J. A., Guevara, A., Munizaga, M., Figueroa, M., and Torres, V. (2020). Mobility changes, teleworking, and remote communication during the COVID-19 pandemic in Chile. *Findings*, 13489.
- Axisa, J. J., Scott, D. M., and Newbold, K. B. (2012). Factors influencing commute distance: a case study of Toronto's commuter shed. *Journal of Transport Geography*, 24, 123-129.
- Baker, T. L., Wise, J., Kelley, G., and Skiba, R. J. (2016). Identifying barriers: Creating solutions to improve family engagement. *School Community Journal*, *26*(2), 161-184.
- Bamieh, O., and Ziegler, L. (2022). Are remote work options the new standard? Evidence from vacancy postings during the COVID-19 crisis. *Labour Economics*, *76*, 102179.
- Batur, I., Dirks, A. C., Mondal, A., Bhat, C. R., and Pendyala, R. M. (2024). Does ridehailing use affect vehicle ownership or vice versa? An exploratory investigation of the relationship using a latent market segmentation approach. *Transportation Research Record*, *2678*(3), 252-265.
- Beck, M. J., Hensher, D. A., and Wei, E. (2020). Slowly coming out of COVID-19 restrictions in Australia: Implications for working from home and commuting trips by car and public transport. *Journal of Transport Geography*, 88, 102846.
- Beckman, M. (2022). Single Fathers Have 8 Times More Wealth Than Single Mothers, Data Shows. NASDAQ. Available at: <u>https://www.nasdaq.com/articles/single-fathers-have-8-times-more-wealth-than-single-mothers-data-shows</u>. Accessed on May 30, 2023.
- Belostecinic, G., Mogoş, R. I., Popescu, M. L., Burlacu, S., Rădulescu, C. V., Bodislav, D. A., Bran, F., and Oancea-Negescu, M. D. (2021). Teleworking-An economic and social impact during COVID-19 pandemic: A data mining analysis. *International Journal of Environmental Research and Public Health*, 19(1), 298.
- Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*, 9. MIT Press.
- Bhat, C. R. (1997). An endogenous segmentation mode choice model with an application to intercity travel. *Transportation Science*, *31*(1), 34-48.
- Bhat, C. R. (2008). The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B*, 42(3), 274-303.
- Bhat, C. R. (2015). A new generalized heterogeneous data model (GHDM) to jointly model mixed types of dependent variables. *Transportation Research Part B*, 79, 50-77.
- Bhat, C. R. (2018). New matrix-based methods for the analytic evaluation of the multivariate cumulative normal distribution function. *Transportation Research Part B*, 109, 238-256.
- Bhat, C. R. (2022). Data collection methods and activity-travel behavior analysis in a fast evolving technological future. *ITE Journal*, 92(10), 29-33.
- Bhat, C. R., Castro, M., and Khan, M. (2013). A new estimation approach for the multiple discrete– continuous probit (MDCP) choice model. *Transportation Research Part B*, 55, 1-22.
- Bhat, C. R., Astroza, S., Sidharthan, R., and Bhat, P. C. (2014). A multivariate hurdle count data model with an endogenous multiple discrete–continuous selection system. *Transportation Research Part B*, *63*, 77-97.

- Bhat, C. R., Astroza, S., Bhat, A. C., and Nagel, K. (2016). Incorporating a multiple discretecontinuous 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.
- Bhat, C. R. and Lavieri, P. S. (2018). A new mixed MNP model accommodating a variety of dependent non-normal coefficient distributions. *Theory and Decision*, 84(2), 239-275.
- Bhuiyan, M. A. A., Rifaat, S. M., Tay, R., and De Barros, A. (2020). Influence of community design and sociodemographic characteristics on teleworking. *Sustainability*, *12*(14), 5781.
- Blumenberg, E., and King, H. (2019). Low-income workers, residential location, and the changing commute in the United States. *Built Environment*, *45*(4), 563-581.
- Brennan, A., Angus, C., Pryce, R., Buykx, P., Henney, M., Gillespie, D., Holmes, J., and Meier, P. S. (2021). Potential effects of minimum unit pricing at local authority level on alcohol-attributed harms in North West and North East England: A modelling study. *Public Health Research*, 9(4).
- Caldarola, B., and Sorrell, S. (2022). Do teleworkers travel less? Evidence from the English national travel survey. *Transportation Research Part A*, 159, 282-303.
- Castrillon, C. I. M., Beckenkamp, P. R., Ferreira, M. L., Michell, J. A., de Aguiar Mendes, V. A., Luscombe, G. M., Stamatakis, E., and Ferreira, P. H. (2020). Are people in the bush really physically active? A systematic review and meta-analysis of physical activity and sedentary behaviour in rural Australians populations. *Journal of Global Health*, 10(1), 010410.
- Cerqueira, E. D. V., Motte-Baumvol, B., Chevallier, L. B., and Bonin, O. (2020). Does working from home reduce CO2 emissions? An analysis of travel patterns as dictated by workplaces. *Transportation Research Part D*, *83*, 102338.
- Chakrabarti, S. (2018). Does telecommuting promote sustainable travel and physical activity?. *Journal of Transport and Health*, 9, 19-33.
- Cheng, Y. T., Lavieri, P. S., and Astroza, S. (2022). How did the COVID-19 pandemic impact the location and duration of work activities? A latent class time-use study. *Findings*, 35621.
- Chong, S., Huang, Y., and Chang, C. H. D. (2020). Supporting interdependent telework employees: A moderated-mediation model linking daily COVID-19 task setbacks to nextday work withdrawal. *Journal of Applied Psychology*, 105(12), 1408.
- Choo, S., Mokhtarian, P. L., and Salomon, I. (2005). Does telecommuting reduce vehicle-miles traveled? An aggregate time series analysis for the US. *Transportation*, *32*, 37-64.
- Coate, P. (2021). Remote Work Before, During and After the Pandemic. NCCI. Available at: <u>https://www.ncci.com/SecureDocuments/QEB/QEB_Q4_2020_RemoteWork.html</u>. Accessed on March 1, 2023.
- Cuerdo-Vilches, T., Navas-Martín, M. Á., March, S., and Oteiza, I. (2021). Adequacy of telework spaces in homes during the lockdown in Madrid, according to socioeconomic factors and home features. *Sustainable Cities and Society*, *75*, 103262.
- Davies, A. (2021). COVID-19 and ICT-supported remote working: Opportunities for rural economies. *World*, 2(1), 139-152.
- de Abreu e Silva, J. (2022). Residential preferences, telework perceptions, and the intention to telework: insights from the Lisbon Metropolitan Area during the COVID-19 pandemic. *Regional Science Policy and Practice*, *14*, 142-161.
- de Abreu e Silva, J., and Melo, P. C. (2018). Does home-based telework reduce household total travel? A path analysis using single and two worker British households. *Journal of Transport Geography*, 73, 148-162.

- DeVos, D., Meijers, E., and van Ham, M. (2018). Working from home and the willingness to accept a longer commute. *The Annals of Regional Science*, *61*(2), 375-398.
- DeVos, D., van Ham, M., and Meijers, E. J. (2019). Working from home and commuting: Heterogeneity over time, space, and occupations. IZA Discussion Paper No. 12578, IZA Institute of Labor Economics, Bonn, Germany
- Elldér, E. (2020). Telework and daily travel: New evidence from Sweden. *Journal of Transport Geography*, 86, 102777.
- Ettema, D. (2010). The impact of telecommuting on residential relocation and residential preferences: a latent class modeling approach. *Journal of Transport and Land Use*, *3*(1), 7-24.
- Flynn, J. (2023). 25 Trending Remote Work Statistics [2023]: Facts, Trends, and Projections. Zippia. Available at: <u>https://www.zippia.com/advice/remote-work-statistics/#:~:text=66%25%20of%20U.S.%20employees%20work,5.8%20remote%20work</u> <u>days%20per%20month</u>. Accessed on: June 18, 2023.
- Foucault, M., and Galasso, V. (2020). Working after COVID-19: Cross-country evidence from real-time survey data. Science Po CEVIPOF Research Note, 9.
- Frick, N., and Marx, J. (2021). Integrating digital nomads in corporate structures: Managerial contemplations. In Proceedings of the 54th Hawaii International Conference on System Sciences (p. 2111).
- Fu, M., Kelly, J. A., Clinch, J. P., and King, F. (2012). Environmental policy implications of working from home: Modelling the impacts of land-use, infrastructure and sociodemographics. *Energy Policy*, 47, 416-423.
- Gaduena, A., Caboverde, C. E., and Flaminiano, J. P. (2022). Telework potential in the Philippines. *The Economic and Labour Relations Review*, 33(2), 434-454.
- Giovanis, E. (2018). The relationship between teleworking, traffic and air pollution. *Atmospheric Pollution Research*, 9(1), 1-14.
- Giselmar A. J., Hemmert, G. A. J., Schons, L. M., Wieseke, J., and Schimmelpfennig, H. (2018). Log-likelihood-based Pseudo-*R*² in Logistic Regression: Deriving Sample-sensitive Benchmarks. *Sociological Methods & Research*, 47(3), 507-531
- Glaeser, E. L., Kahn, M. E., and Chu, C. (2001). Job sprawl: Employment location in US metropolitan areas. Washington, DC: Brookings Institution, Center on Urban and Metropolitan Policy.
- Greene, W. H. (2016). *Nlogit Version 6.0 Reference Guide*. Econometric Software Inc., Plainview, NY.
- Gubins, S., van Ommeren, J., and de Graaff, T. (2019). Does new information technology change commuting behavior?. *The Annals of Regional Science*, *62*(1), 187-210.
- Haider, M., and Anwar, A. I. (2023). The prevalence of telework under Covid-19 in Canada. *Information Technology and People*, 36(1), 196-223.
- Hannonen, O. (2020). In search of a digital nomad: defining the phenomenon. *Information Technology and Tourism*, 22(3), 335-353.
- Hastie, R. (2015). Causal thinking in judgments. In: Keren G, Wu G, eds. *The Wiley Blackwell Handbook of Judgment and Decision Making*, 21, 590-628. John Wiley & Sons, Ltd., Chichester, UK.
- Helminen, V., and Ristimäki, M. (2007). Relationships between commuting distance, frequency and telework in Finland. *Journal of Transport Geography*, 15(5), 331-342.

- Higgins, V. (2017). Augmented and Virtual Reality: The future of work, not just play. *Professional Safety*, 62(6), 86.
- Hook, A., Sovacool, B. K., and Sorrell, S. (2020). A systematic review of the energy and climate impacts of teleworking. *Environmental Research Letters*, *15*(9), 093003.
- Hostettler Macias, L., Ravalet, E., and Rérat, P. (2022). Potential rebound effects of teleworking on residential and daily mobility. *Geography Compass*, 16(9), e12657.
- Hu, L. (2021). Gender differences in commuting travel in the US: Interactive effects of race/ethnicity and household structure. *Transportation*, 48(2), 909-929.
- Isono, I., and Nara, K. (2022). COVID-19, Telework Patterns Within a City, and Changes in Urban Structure: Preliminary Findings. Economic Research Institute for ASEAN and East Asia.
- Jaff, M. M., and Hamsa, A. A. K. (2021). Exploratory factor analysis on advantages and disadvantages of potential female telecommuters in Kuala Lumpur. *Asian Transport Studies*, *7*, 100034.
- Kersting, M., Matthies, E., Lahner, J., and Schlüter, J. (2021). A socioeconomic analysis of commuting professionals. *Transportation*, 48, 2127-2158.
- Keya, N., Anowar, S., Bhowmik, T., and Eluru, N. (2021). A joint framework for modeling freight mode and destination choice: Application to the US commodity flow survey data. *Transportation Research Part E*, *146*, 102208.
- Kim, S. N. (2017). Is telecommuting sustainable? An alternative approach to estimating the impact of home-based telecommuting on household travel. *International Journal of Sustainable Transportation*, 11(2), 72-85.
- Kim, J., and Long, D. (2022). What flattened the house-price gradient? The role of work-fromhome and decreased commuting cost. SSRN Electronic Journal. 10.2139/ssrn.4193300.
- Krasilnikova, N., and Levin-Keitel, M. (2022). Telework as a Game-Changer for Sustainability? Transitions in Work, Workplace and Socio-Spatial Arrangements. *Sustainability*, 14(11), 6765.
- Lachapelle, U., Tanguay, G. A., and Neumark-Gaudet, L. (2018). Telecommuting and sustainable travel: Reduction of overall travel time, increases in non-motorised travel and congestion relief?. Urban Studies, 55(10), 2226-2244.
- Le Pelley M. E., Griffiths, O., and Beesley, T. (2017). Associative accounts of causal cognition. In: Waldmann MR, ed. *The Oxford Handbook of Causal Reasoning*. Oxford University Press, New York, NY.
- Lei, L., and Liu, X. (2022). The COVID-19 pandemic and residential mobility intentions in the United States: Evidence from Google Trends data. *Population, Space and Place*, 28(6), e2581.
- Lennox, J. (2020). More working from home will change the shape and size of cities. Centre of Policy Studies, Victoria University.
- López Soler, J. R., Christidis, P., and Vassallo, J. M. (2021). Teleworking and online shopping: Socio-economic factors affecting their impact on transport demand. *Sustainability*, 13(13), 7211.
- López-Igual, P., and Rodríguez-Modroño, P. (2020). Who is teleworking and where from? Exploring the main determinants of telework in Europe. *Sustainability*, *12*(21), 8797.
- Lorenzo, G. (2021). Where the Jobs Are: Urban, Suburban, and Rural Migrations and the Hollowing Out of Middle-Skill Jobs. *Workforce Monitor*. Available at: <u>https://wfmonitor.com/2020/06/12/where-the-jobs-are-urban-suburban-rural-migrations-and-the-hollowing-out-of-middle-skill-jobs/</u>. Accessed on: May 22, 2023.

- Lu, X., and White, H. (2014). Robustness checks and robustness tests in applied economics. *Journal of Econometrics*, 178, 194-206.
- Maheshwari, R., De Vos, J., Van Acker, V., and Witlox, F. (2022). Analyzing the association between satisfaction with commuting time and satisfaction with life domains. *Journal of Transport and Land Use*, 15(1), 231-248.
- Mariotti, I., Di Matteo, D., and Rossi, F. (2022). Who were the losers and winners during the Covid-19 pandemic? The rise of remote working in suburban areas. *Regional Studies, Regional Science*, 9(1), 685-708.
- McFadden, D. (1977). Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments, Cowles Foundation Discussion Papers. 707. <u>https://elischolar.library.yale.edu/cowles-discussion-paper-series/707</u>, accessed December 10, 2023.
- McLachlan, G. J., and Peel, D. (2000). Finite Mixture Models. Wiley, New York.
- McLachlan, G. J., Lee, S. X., and Rathnayake, S. I. (2019). Finite mixture models. *Annual Review* of Statistics and Its Applications, 6, 355-378.
- McNicholas, P. D. (2016). *Mixture Model-Based Classification* (1st ed.). Chapman and Hall/CRC.
- Melo, P. C., and de Abreu e Silva, J. (2017). Home telework and household commuting patterns in Great Britain. *Transportation Research Part A*, *103*, 1-24.
- Mergener, A., and Mansfeld, L. (2021). Being spatially mobile without daily commuting? How working from home patterns relate to company-home distances. BIBB-Preprint. Federal Institute for Vocational Education and Training, Bonn.
- Mitra, S. K., and Saphores, J. D. M. (2019). Why do they live so far from work? Determinants of long-distance commuting in California. *Journal of Transport Geography*, 80, 102489.
- Mokhtarian, P. L. (1991). An empirical analysis of the travel impacts of telecommuting. In *Proceedings, 6th International Conference on Travel Behavior, 1*, 147-160.
- Mokhtarian, P. L., Collantes, G. O., and Gertz, C. (2004). Telecommuting, residential location, and commute-distance traveled: evidence from State of California employees. *Environment and Planning A*, *36*(10), 1877-1897.
- Mouratidis, G. (2018). Digital nomads: Travel, remote work and alternative lifestyles. *MS Thesis*, Department of Applied Cultural Analysis, Lund University.
- Muhammad, S., Ottens, H. F., Ettema, D., and de Jong, T. (2007). Telecommuting and residential locational preferences: A case study of the Netherlands. *Journal of Housing and the Built Environment*, *22*, 339-358.
- Nash, C., Jarrahi, M. H., and Sutherland, W. (2021). Nomadic work and location independence: The role of space in shaping the work of digital nomads. *Human Behavior and Emerging Technologies*, 3(2), 271-282.
- Nguyen, M. H. (2021). Factors influencing home-based telework in Hanoi (Vietnam) during and after the COVID-19 era. *Transportation*, 48(6), 3207-3238.
- Nilles, J. M. (1991). Telecommuting and urban sprawl: Mitigator or inciter? *Transportation*, *18*(4), 411-432.
- O'Brien, W., and Aliabadi, F.Y. (2020). Does telecommuting save energy? A critical review of quantitative studies and their research methods. *Energy and Buildings*, 225, 110298.
- Office of the Texas Governor (2021, March 2). *Relating to the opening of Texas in response to the COVID-19 disaster* [Press release]. <u>https://gov.texas.gov/news/post/governor-abbott-lifts-mask-mandate-opens-texas-100-percent</u>.

- Ory, D. T., and Mokhtarian, P. L. (2006). Which came first, the telecommuting or the residential relocation? An empirical analysis of causality. *Urban Geography*, 27(7), 590-609.
- Patton, E. (2019). Where does work belong? Home-based work and communication technology within the American middle-class postwar home. *Technology and Culture*, 60(2), 523-552.
- Pendyala, R. M., Goulias, K. G., and Kitamura, R. (1991). Impact of telecommuting on spatial and temporal patterns of household travel. *Transportation*, *18*, 383-409.
- Perch-Nielsen, S., von Felten, N., Henzen, C., Meissner, J., Sprenger, M., Bieri, M., Arvantis, S., Seliger, F. and Ley, M. (2014) Auswirkungen neuer Arbeitsformen auf den Energieverbrauch und das Mobilitätsverhalten von Arbeitnehmenden, 290955. Bern: Bundesamt für Energie NFE.
- Pinjari, A. R., and Bhat C. R. (2021). Computationally efficient forecasting procedures for Kuhn-Tucket consumer demand model systems: Applications to residential energy consumption analysis. *Journal of Choice Modeling*, 39, 100283.
- Popuri, Y. D., and Bhat, C. R. (2003). On modeling choice and frequency of home-based telecommuting. *Transportation Research Record*, 1858(1), 55-60.
- Raišienė, A. G., Rapuano, V., and Varkulevičiūtė, K. (2021). Sensitive men and hardy women: How do millennials, xennials and gen x manage to work from home?. *Journal of Open Innovation: Technology, market, and complexity*, 7(2), 106.
- Ramsey, K., and Bell, A. (2014). Smart Location Database, Version 2.0 User Guide. Available at: <u>https://www.epa.gov/sites/production/files/2014-03/documents/sld_userguide.pdf</u>. Accessed on May 15, 2022.
- Ravalet, E., and Rérat, P. (2019). Teleworking: Decreasing mobility or increasing tolerance of commuting distances?. *Built Environment*, 45(4), 582-602.
- Robbennolt, D., Haddad, A., and Bhat, C. R. (2023). Housing choice in an evolving remote work landscape. Technical paper, Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin.
- Saad, L, and Wigert, B. (2021). Remote Work Persisting and Trending Permanent. Gallup. Available at: <u>https://news.gallup.com/poll/355907/remote-work-persisting-trending-permanent.aspx</u>. Accessed on March 1, 2023.
- Schieman, S., Badawy, P. J., A. Milkie, M., and Bierman, A. (2021). Work-life conflict during the COVID-19 pandemic. *Socius*, 7. <u>https://doi.org/10.1177/2378023120982856</u>.
- Sen, S., Charles, M. B., and Harrison, J. L. (2021). Determinants of commute distance in south east Queensland, Australia: Implications for usage-based pricing in lower-density urban settings. *Urban Policy and Research*, 39(3), 292-307.
- Sener, I. N., and Bhat, C. R. (2011). A copula-based sample selection model of telecommuting choice and frequency. *Environment and Planning A*, 43(1), 126-145.
- Shabanpour, R., Golshani, N., Tayarani, M., Auld, J., and Mohammadian, A. K. (2018). Analysis of telecommuting behavior and impacts on travel demand and the environment. *Transportation Research Part D*, *62*, 563-576.
- Sloman S. A., and Lagnado D. (2015). Causality in thought. *Annual Review of Psychology*, 66(1), 223-247. doi:10.1146/annurev-psych-010814-015135.
- Solon, G., Haider, S., and Wooldridge, J. (2015). What are we weighting for? *Journal of Human Resources*, 50(2), 301-316. <u>https://doi.org/10.3386/w18859.</u>
- Srivastava, A., Van Passel, S., and Laes, E. (2018). Assessing the success of electricity demand response programs: A meta-analysis. *Energy Research & Social Science*, 40, 110-117.

- Stefaniec, A., Brazil, W., Whitney, W., and Caulfield, B. (2022). Desire to work from home: Results of an Irish study. *Journal of Transport Geography*, *104*, 103416.
- Stickel, M. D. (2020). Challenges and opportunities of digital nomadism and its implications for tomorrow's workforce. Master's Work Project, NOVA School of Business and Economics, Portugal.
- Stiles, J., and Smart, M. J. (2021). Working at home and elsewhere: daily work location, telework, and travel among United States knowledge workers. *Transportation*, *48*(5), 2461-2491.
- Sweet, M., and Scott, D. M. (2022). Insights into the future of telework in Canada: Modeling the trajectory of telework across a pandemic. *Sustainable Cities and Society*, 87, 104175.
- Tahlyan, D., Said, M., Mahmassani, H., Stathopoulos, A., Walker, J., and Shaheen, S. (2022). For whom did telework not work during the pandemic? Understanding the factors impacting telework satisfaction in the US using a multiple indicator multiple cause (MIMIC) model. *Transportation Research Part A*, 155, 387-402.
- Tayyaran, M. R., and Khan, A. M. (2007). Telecommuting and residential location decisions: Combined stated and revealed preferences model. *Canadian Journal of Civil Engineering*, 34(10), 1324-1333.
- Tenailleau, Q. M., Tannier, C., Vuidel, G., Tissandier, P., and Bernard, N. (2021). Assessing the impact of telework enhancing policies for reducing car emissions: Exploring calculation methods for data-missing urban areas–Example of a medium-sized European city (Besançon, France). Urban Climate, 38, 100876.
- Texas Demographic Center (2022). 2020 Census. Available at: <u>https://demographics.texas.gov/</u>. Accessed December 12, 2022.
- Tiedeman, D. V. (1955). On the study of types. *Symposium on Pattern Analysis*. 1-14. Randolph Field, TX: Air University, USF School of Aviation Medicine.
- Torbacki, W. (2021). Achieving sustainable mobility in the Szczecin metropolitan area in the post-COVID-19 era: The DEMATEL and PROMETHEE II approach. *Sustainability*, *13*(22), 12672.
- Train, K. E. (2009). Discrete Choice Methods with Simulation. Cambridge University Press.
- Tran, M. T., Zhang, J., Chikaraishi, M., and Fujiwara, A. (2016). A joint analysis of residential location, work location and commuting mode choices in Hanoi, Vietnam. *Journal of Transport Geography*, 54, 181-193.
- U.S. Census Bureau (2022). Employment Status in Texas. <u>https://data.census.gov/table/ACSST1Y2022.S2301?q=Texas%20Employment</u>. Accessed November 12, 2023.
- Waddell, P. C., Bhat, C. B., Eluru, N., Wang, L., and Pendyala, R. M. (2007). Modeling interdependence in household residence and workplace choices. *Transportation Research Record*, 2003, 84-92.
- Walls, M., and Safirova, E. (2004). A review of the literature on telecommuting and its implications for vehicle travel and emissions. Discussion Papers 10492, Resources for the Future.
- Watkins, T. R. (2019) Understanding uncertainty and bias to improve causal inference in health intervention research. PhD dissertation, The University of Sydney.
- White, P., Christodoulou, G., Mackett, R., Titheridge, H., Thoreau, R., and Polak, J. (2007). The role of teleworking in Britain: its implications for the transport system and economic evaluation. In *European Transport Conference*. Association for European Transport.

- Williamson, J. (2006). Dispositional versus epistemic causality. *Minds and Machines*, 16, 259-276.
- Wöhner, F. (2022). Work flexibly, travel less? The impact of telework and flextime on mobility behavior in Switzerland. *Journal of Transport Geography*, *102*, 103390.
- Wöhrmann, A. M., and Ebner, C. (2021). Understanding the bright side and the dark side of telework: An empirical analysis of working conditions and psychosomatic health complaints. *New Technology, Work and Employment*, 36(3), 348-370.
- Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions. *Journal of Econometrics*, 68(1), 115-132.
- Wunsch, G., Russo, F., and Mouchart, M. (2010). Do we necessarily need longitudinal data to infer causal relations?. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, 106(1), 5-18.
- Zhang, S., Moeckel, R., Moreno, A. T., Shuai, B., and Gao, J. (2020). A work-life conflict perspective on telework. *Transportation Research Part A*, 141, 51-68.
- Zhu, P., and Mason, S. G. (2014). The impact of telecommuting on personal vehicle usage and environmental sustainability. *International Journal of Environmental Science and Technology*, 11, 2185-2200.



Figure 1: Sample Distribution of Individuals among Seven WPL Combinations



Figure 2: Relationship between Remote Work Frequency and Average Commute Distance



Figure 3: Relationship between Commute Distance and Average Remote Work Participation



Figure 4: Latent Segmentation Model Framework

Variable	Count	%	TX%	Variable	Count	%	TX%
Individual Demographics				Residential Attributes			
Gender				Neighborhood Density Type			
Female	492	59.7	43.6	Rural	187	22.7	
Male	332	40.3	56.4	Suburban	471	57.2	
Age				Urban	166	20.1	
18 to 29	38	4.6	24.0	Home Tenure			
30 to 39	93	11.4	23.2	Owns home	699	84.8	
40 to 49	175	21.2	21.1	Rents home	125	15.2	
50 to 64	386	46.8	24.2	Private Study			
65 or older	132	16.0	5.3	Yes	600	72.8	
Formal Education Level				No	224	27.2	
Less than a Bachelor's degree	133	16.1		Employment Characteristics			
Bachelor's degree	290	35.2		Office Density Type			
Graduate degree	401	48.7		Low	193	23.4	
Household Demographics				Medium	45	5.5	
Presence of Children (including ages)				High	586	71.1	
No children	598	72.6		Employment Status and Structure			
Child(ren) aged 0 to 4	46*	5.6		Full-time (work hrs per week≥30), self-employed	106	12.9	
Child(ren) aged 5 to 12	128*	15.5		Part-time (work hrs per week<30), self-employed	10	1.2	
Child(ren) aged 13 to 17	118*	14.3		Full-time, not self-employed	682	82.8	
Household structure				Part-time, not self-employed	26	3.2	
Lives alone	106	12.9		Number of Work Days per Month			
Does not live with partner or alone and has no kids	106	12.9		1-5 days	19	2.3	
Does not live with partner and has kids (single parent)	46	3.9		6 to 10 days	17	2.1	
Lives with partner and has no kids	128	46.8		11 to 19 days	16	1.9	
Lives with partner and has kids	118	23.5		20 to 24 days	665	80.7	
Household Annual Income				25 days or more	107	13.0	
Less than \$25,000	6	0.6	18.3	Occupation			
\$25,000 to \$49,999	40	4.7	21.2	Healthcare	43	5.2	
\$50,000 to \$74,999	108	13.0	17.6	Retail Sales/Food Services	23	2.8	
\$75,000 to \$99,999	114	13.7	12.5	Education/Social Services	375	45.5	
\$100,000 to \$149,999	217	33.6	15.6	Information/Finance	46	5.6	
\$150,000 to \$249,999	191	23.5	7.0	Managerial/Technical	188	21.8	
\$250,000 or more	148	10.9	7.8	Professional Services	108	13.1	
				Public Administration	41	5.0	

Table 1: Sample Distribution of Exogenous Variables (N=824)

*226 households have children in them. The children categories do not add up to this number because some households have more than one child, who are in multiple age groups. "—" indicates that census comparisons are not available on the employed Texas population.

Table	2: Estimation	Results for	Endogenous	Variables	Within	Each S	Segment

	Telework Choice							In (Commuto	
Explanatory Variables	Work fro	m Home	Work f	rom the	Work from a		Distance)		
(base category)	Cooff	t stat	Work	Office	Third	WPL	Cooff	t stat	
Individual and Household Demographics	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat	
Gendered lifecycle variables									
Female			0.095	1.21			-0.077	-1 25	
Single Mothers			-0.377	-1 75			0.077	1.25	
Single Parents (father and mothers)				1.75			0.223	1.28	
Presence of children 0-12 years (satiation)			-0.330	-1.54			NA	0	
Presence of children 13-17 years (satiation)			0.301	1.19			NA		
Age (29 years or younger)			010 0 1	,					
30 to 49 years old					-0.365	-1.77			
50 years or older	-0.173	-1.85			-0.365	-1.77	0.083	1.26	
30 to 39 years (satiation)							NA		
40 years or older (satiation)	-0.642	-2.32					NA		
Formal Education Level (less than a									
Bachelor's degree)									
Bachelor's degree or higher	0.216	1.82					-0.283	-3.49	
Income (less than \$100,000)									
\$100,000 or more	0.281	3.01			0.303	1.63			
Residential Attributes									
Population Density (Urban)									
Suburban									
Rural			0.788	6.54					
Suburban (satiation)	0.342	1.69					NA		
Private Study (no)									
Yes	0.310	3.11							
Employment Characteristics									
Employment density (medium-to-low)									
High			-0.188	-1.80	-0.262	-1.68	-0.113	-1.69	
Employment Status (employed full-time -									
\geq 30 hours per wk)									
Employed part-time (<30 hours/wk)							-0.504	-2.70	
# Days Worked per Month (6 to 24 days									
1 to 5 days			-0.528	-2.28					
25 days or more (over time)					0.198	1.44			
Occupation (managerial, technical,									
service job)			0.000	1.60					
Professional Samiaas			0.299	1.60					
Professional Services			-0.480	-4.06					
Fudic Administration			-0.66/	-3.3/					
<u>Enaogenous Effects</u>									
Home-based remote work $70 (\leq 50\%)$	NT A		NTA		NT A		0 (71	2 44	
Third WPL based remote work (none)	INA		NA		NA		0.071	5.44	
At least once a month	NIA		NIA		NIA		0.601	2.26	
At least once a month One Way Commute Distance (between	INA		NA		INA		-0.001	-5.20	
8 01 and 15 miles)									
<8 miles			0 496	2 65			NA		
>15 miles			0.211	1 47			NA		
Constants			5.211	1.7/			1111		
TC Segment ("Movers")			0 742	2 19	-0.261	-0.71	1 1 9 9	8 09	
CT Segment ("Stavers")			1 237	4 66	-0 503	_1 10	2 809	26.29	
Constants (Satiation)			1.257	1.00	0.505	1.17	2.007	20.27	
TC Segment ("Movers")	-0.137	-0 33	0.394	0.68	-0.511	-0.76	NA		
CT Segment ("Stayers")	-0.268	-0.83	-0.232	-0.83	-0.852	-1.44	NA		

	Telework Choice							I m (Communito			
Explanatory Variables (base category)	Work from Home		Work from the Work Office		Work from a Third WPL		a L	Distance)			
	Coeff.	t-stat	Coeff	: t-	stat	Coeff.	t-s	tat	Coeff.	t-stat	
TC Segment ("The Movers")											
	Work f	from the V	Vork	Wo	ork fro	m a Thir	d	(ln)Commute			
				WPL#			Distance		ce		
Work from the Work Office#	1.00				0.526*				0.011		
Work from a Third WPL#						0.856*			0.529		
(ln)Commute Distance										0.568*	
CT Segment ("The Stayers")											
	Work f	from the V	Vork	Work from a Thi		m a Thir	d		(ln)Commute		
		Office#		WPL#				Distance			
Work from the Work Office#		.000&		0.526*			-0.153*				
Work from a Third WPL#			. 0.856*		56*		-0.027				
(ln)Commute Distance										0.568*	

[#] Differenced with respect to "Work from Home" alternative * Significant at 85% confidence level & Fixed value

Explanatory Variables	TC Seg	gment	CT Segment The Stavers	
(base category)	Coeff.	t-stat	(base)	
Individual and Household-level Characteristics				
Gendered lifecycle variables				
Presence of child(ren) aged 12 years or less	-0.507	-1.28		
Lives with partner	-0.812	-2.85		
Age (65 years or older)				
18 to 29 years old	0.795	1.41		
Income (\$50,000 or more)				
Less than \$50,000	0.798	1.61		
<u>Residential Characteristic</u>				
Population Density (rural or urban)				
Suburban	-1.461	-4.77		
Constant	-0.212	-0.79		
Segment Size	20.3 16	% 8	79.7% 656	

Table 3: Latent Segmentation Assignment Model

		Percent (%) within	Percent (
		segr	nent	attri	attribute		
		The	The	The	The	Overall	
	Attributes	Movers	Stavers	Movers	Stavers	Sample	
		Segment 1	Segment 2	Segment 1	Segment 2	(%)	
		WPL→CD	CD→WPL	WPL→CD	CD→WPL		
Individual and	Household-level Characteristics	S S S	CD 7 MIL	WIE / OD	007111		
~ 1	Female	60.7	59.5	20.7	79.3	59.7	
Gender	Male	39.3	40.5	19.8	80.2	40.3	
	No children present						
	(regardless of age)	70.2	73.2	19.7	80.3	72.6	
	Children present					05.4	
Household	(regardless of age)	29.8	26.8	22.0	78.0	27.4	
children	Child(ren) 0 to 12 years old					10.1	
	present	22.6	16.9	25.5	74.5	18.1	
	Child(ren) 13 to 17 years old					14.2	
	present	13.1	14.8	18.3	81.7	14.3	
	Single Mother	7.7	2.3	46.5	53.5	3.3	
Household	Single Father	1.8	0.5	4.6	95.4	1.0	
structure	Parent who lives with partner	20.7	24.2	18.0	82.0	23.5	
	Others*	69.8	73.0	19.7	80.3	72.2	
	18 to 29	10.7	3.2	46.1	53.9	4.7	
Age	30 to 39	11.9	10.5	22.6	77.4	10.8	
	40 to 49	20.8	21.5	20.1	79.9	21.4	
	50 or older	56.6	64.8	18.1	81.9	63.1	
Education	Some college or less	16.1	16.2	20.5	79.5	16.1	
level	Bachelor's degree or higher	83.9	83.8	20.3	79.7	83.9	
TT	Up to \$50,000	7.1	5.2	25.4	74.6	5.6	
Household	\$50,000 to \$99,999	34.5	25.0	26.1	73.9	26.9	
income	\$100,000 or more	58.3	69.8	17.6	82.4	67.5	
Residential Ch	aracteristics						
Dopulation	Rural	27.3	21.5	24.5	75.5	22.7	
density	Suburban	42.9	60.8	15.2	84.8	57.2	
defisity	Urban	29.8	17.7	30.1	69.9	20.1	
Drivoto study	Yes	66.7	74.5	18.6	81.4	72.8	
T IIvate study	No	33.3	25.5	25.0	75.0	27.2	
In-Person Wor	rkplace Characteristic						
Б 1 (Low	26.7	22.7	22.7	77.3	23.4	
Employment	Medium	5.4	5.5	20.6	79.4	5.5	
density	High	67.9	71.8	19.5	80.5	71.1	
Job-level Char	racteristics		•		•		
Employment	Employed part-time	4.2	2.7	26.5	73.5	4.3	
status	Employed full-time	95.8	97.3	20.1	79.9	95.7	
	Managerial or technical job	77.9	76.4	20.6	79.4	75.6	
Occurrentier	Healthcare	4.8	5.3	19.6	80.4	5.2	
Occupation	Professional Services	12.5	13.3	19.2	80.8	13.2	
	Public Administration	4.8	5.0	19.9	80.1	5.0	
		Perce	nt (%)	20.3	79.7	100	
1	Segment size	1	N	168	656	824	

Table 4: Profiles of the Two Latent Market Segments

WPL: Workplace Location; CD: Commute Distance

*We did not include the individual segment profiles for all household structures, as the ones excluded were all fairly equivalent to the overall sample and segment size across all metrics (as the "lives with a partner" structure is). Household structures under the term "other" include: live alone, lives with partner with no children, live with parents, and live with unrelated roommates.

Measure Type	Metric	Proposed Joint Segmented Model	Independent Segmented Model	Joint Unsegmented Model	Independent Unsegmented Model			
	Log-likelihood at convergence	-1793.74	-1824.29	-1839.61	-1847.44			
	Number of non-constant parameters	43	38	38	31			
	Log-likelihood at constants-only	-1960.87	-1960.87	-1960.87	-1960.87			
Likelihood Based Maasumas	Adjusted Rho-squared value	0.0623	0.0503	0.0425	0.0420			
Wieasures	Likelihood ratio (LR) test: Proposed vs Independent Segmented	$LR = 64.54 > \chi^2_{(7,0.05)} = 14.07$						
	Likelihood ratio (LR) test: Proposed vs Joint Unsegmented	based vs $LR = 91.76 > \chi^2_{(7,0.05)} = 14.07$						
	Likelihood ratio (LR) test: Proposed vs Independent Unsegmented	$LR = 107.40 > \chi^2_{(14,0.05)} = 23.68$						
	Discrete Participation WAPE	9.83%	11.65%	14.32%	16.26%			
Non- likelihood	Weighted Average Probability of Correct Prediction (WAPCP) for Discrete Combinations	0.504	0.494	0.464	0.458			
Based Measures	Average Fractional Split of Days Worked from Each Workplace Location WAPE	1.80%	2.20%	4.60%	5.40%			
	Average ln(Commute Distance) WAPE	18.75%	20.81%	26.4%	28.34%			

Workplace Location Arrangement (across an entire month)		Expected Commute	Expected Commute	Percentage Increase in	Average (across workers) of Expected Monthly	Average (across workers) of Expected Monthly 2-way	Percentage		
Num	Home	Office	Third	Distance for Given Combination (A)	Distance for 100% Work from Office (B)	Expected Commute Distance 100*[A-B]/B	2-way Commute VMT per Worker for Given Combination	Commute VMT per Worker for 100% Work from Office	Increase in 2-way Commute VMT
1	10%	90%	0%	14.7 miles		17.6%	572.6 miles		5.8%
2	20%	80%	0%	15.2 miles		21.6%	526.3 miles	541.0 miles	-2.7%
3	30%	70%	0%	15.7 miles		25.6%	475.6 miles		-12.1%
4	40%	60%	0%	16.1 miles		28.8%	418.1 miles		-22.7%
5	60%	40%	0%	17.1 miles	12.5 miles	36.8%	296.0 miles		-45.3%
6	80%	20%	0%	18.1 miles	12.5 miles	44.8%	156.7 miles		-71.0%
7	90%	10%	0%	18.9 miles		51.2%	81.8 miles		-84.9%
8	100%	0%	0%	20.6 miles		64.8%	0 miles		-100.0%
9	5%	90%	5%	14.5 miles		16.0%	564.8 miles		4.4%
10	10%	80%	10%	15.0 miles		20.0%	519.4 miles		-4.0%

Table 6: Effect of WPL Arrangement on Commute Distance