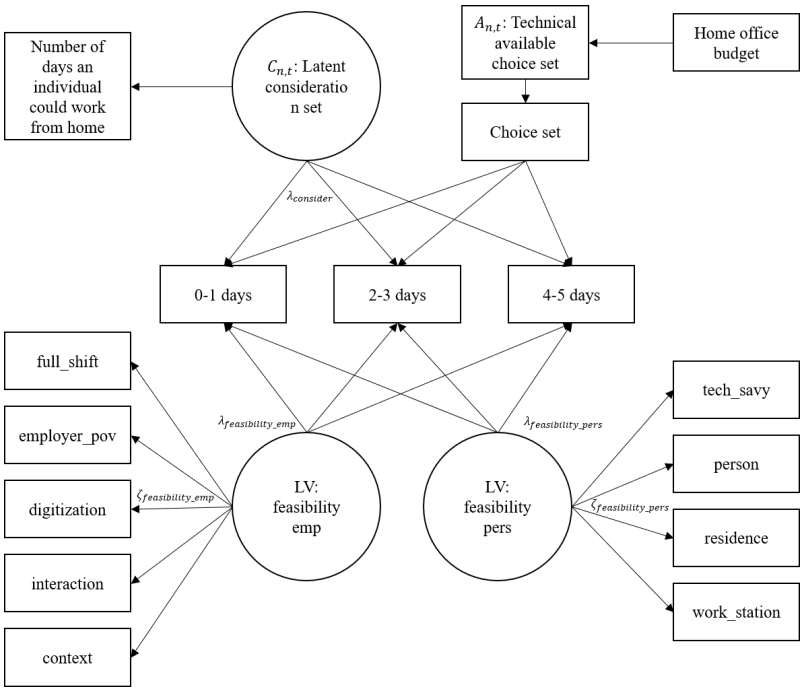
To consider or not to consider?

An endogenous choice set formation approach to model the home office frequency decision

Author(s):
Heimgartner, Daniel; Axhausen, Kay W.

Publication date:
2023-05

Permanent link:
https://doi.org/10.3929/ethz-b-000612933



To consider or not to consider? An endogenous choice set formation approach to model the home office frequency decision

Daniel Heimgartner

Kay W. Axhausen

STRC conference paper 2023 **May 15, 2023**

STRC | **23rd Swiss Transport Research Conference**
Monte Verità / Ascona, May 10-12, 2023

To consider or not to consider? An endogenous choice set formation approach to model the home office frequency decision

Daniel Heimgartner
IVT
ETH Zürich
CH-8093 Zurich
daniel.heimgartnery@ivt.baug.ethz.ch

Kay W. Axhausen
IVT
ETH Zurich
CH-8093 Zurich
axhausen@ivt.baug.ethz.ch

May 15, 2023

Abstract

Our study proposes an endogenous choice set formation (ECSF) approach to model the decision of home office frequency. We hypothesize that a significant portion of the frequency variation can be explained by an individual's home office feasibility, which accounts for two components: (1) the perceived suitability of the characteristics of the job and employment condition (employment feasibility), and (2) the individual's own perception of their suitability for working from home (personal suitability). Both dimensions are modeled as latent variables and found to explain part of the frequency decision. The employment feasibility is the dominant discriminator between the alternatives of 0-1 home office days and 2-3 days, while the personal feasibility becomes more important in explaining the choice of 4-5 home office days. Upon examining the corresponding factor loadings, it appears that the home office work station is a decisive factor in influencing the feasibility of working from home, followed by the quality of the residential environment and the perceived personal suitability for home office work. All three factors exhibit substantial loadings on the latent factor, indicating their meaningful contribution to determining the feasibility of home office work. Additionally, the technological savviness of an individual, although less influential, still has a discernible impact on the feasibility of working from home. The factor loadings of the employment feasibility are relatively less substantial and, except for the job's degree of digitization, are not statistically significant.

Keywords

Endogenous choice set formation, multinomial logit, home office frequency, home office feasibility, hybrid choice

Contents

List of Tables

List of Figures

1 Introduction

The extent to which work can be shifted from the traditional office setting to remote work from home is known to vary significantly across different industries, cities, and countries (Dingel and Neiman, 2020). While there has been extensive discussion on the topic of how many jobs can be performed remotely, it is important to note that the feasibility of working from home is not a binary concept. According to Sener and Bhat (2011), when modeling the frequency of remote work, it is essential to first estimate whether a job can be feasibly done from home or not. However, even if the characteristics of a job make it technically feasible for remote work (from the employer's perspective), it does not necessarily mean that the employee will be able to or choose to shift to remote work (from the employee's perspective) based on their own preferences.

This study aims to examine the feasibility perspective in detail and disentangle its contribution to the observed frequencies of home office usage. We propose that home office feasibility consists of two components: (1) the characteristics of the job and employment, including their perceived suitability for remote work, and (2) the individual's own perception of their suitability for working from home, including their technological proficiency, residential situation, and home office workstation characteristics.

Furthermore, we propose an endogenous choice set formation (ECSF) modeling approach to account for the fact that not all possible frequencies of home office usage are considered by the decision-maker when choosing the number of weekdays to work from home. This approach aims to provide a more comprehensive understanding of the factors that influence remote work behavior and shed light on the complexities of home office feasibility from multiple perspectives.

2 Methodology

The data was gathered during a pre-test study, where respondents were requested to provide essential details about their socioeconomic background, household structure, residence, work characteristics, work-from-home situation, and mobility behavior. Additionally, the survey instrument included Likert-scale questions that enabled the modeler to assess the feasibility of an individual's home office. These indicators are presented in Table 1.

Table 1: Home office feasibility indicators.

Latent variable	Indicator	Statement	Scale	Expected loading
feasibility_emp	full_shift	Could you shift n% of your work time to home office, without feeling pressured to return to your regular work place more often?	Definitely not - Definitely yes	+
	employer_pov	Does your employer like the idea of home office?	Unhappy face - Happy face	+
	digitization	My job can be done mostly on the computer.	Strongly disagree - Strongly agree	+
	interaction	My job requires physical/interpersonal interaction which cannot be compensated by digital channels.	Strongly disagree - Strongly agree	−
	context	My job requires a specific work environmnet (e.g., equipment, safety precautions, working outdoors, etc.).	Strongly disagree - Strongly agree	−
feasibility_pers	tech_savy	I find it easy to work with computers.	Strongly disagree - Strongly agree	+
	person	How suitable do you consider yourself as a person for home office?	Not suitable at all - Very suitable	+
	residence	How suitable do you consider your residential environment (distraction through family, noise, number of rooms, etc.) for home office?	Not suitable at all - Very suitable	+
	work_station	How suitable do you consider your home office work station for home office?	Not suitable at all - Very suitable	+

In the second part of the study, respondents took part in a stated preference (SP) experiment, in which they were presented with two work arrangement options and asked to choose one. One of the attributes of these options was the home office budget, indicating the maximum number of allowed home office days. After selecting their preferred arrangement, respondents were asked to state their intended frequency of home office use, taking into consideration the constraint imposed by the level of the home office budget attribute. The experiment comprised a total of six choice situations which were generated using a D-efficient blocked design. Each respondent was randomly assigned to one of the total 6 blocks.

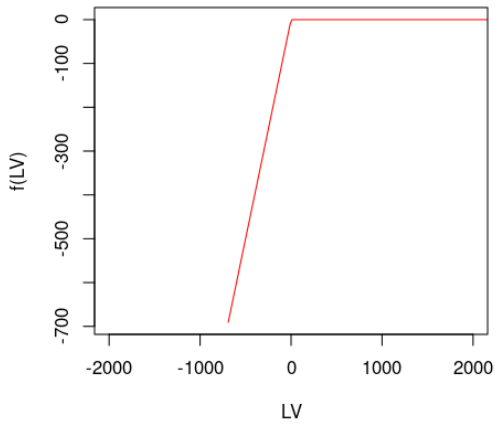
The final sample for analysis was limited to individuals who were currently working from home, resulting in a total of 65 participants.

Neglecting the composition of the choice set may introduce bias in parameter estimates, as highlighted in previous literature (e.g., Swait (1984)). It is evident that not everyone considers the option of working from home for all five days a week as feasible. Hence, it is crucial to account for individual-specific consideration sets. In our modeling approach, we draw on the work of Schmid *et al.* (2022b) where utilities are probabilistically constrained, such that only a latent subset of technical availabilities is considered in the decision-making process. Thus, the true latent choice set, denoted as $C_{n,t}$, is a subset of the technical available choice set, denoted as $A_{n,t}$, which in turn is a subset of the full choice set Γ, i.e., $C_{n,t} \subseteq A_{n,t} \subseteq \Gamma$.

$A_{n,t}$ represents the presence of a home office budget imposed by the employer, which sets a maximum limit on the number of allowed work-from-home days. The consideration set is approximated using a latent variable, which is measured with the help of an indicator where respondents are asked to disclose the percentage of work-time that can be shifted to remote work.

The utility specification of the endogenous choice set formation (ECSF) model is characterized by individuals deriving utility from different levels of remote work, specifically, working either 0-1, 2-3, or 4-5 days from home.

Figure 1: Utility discounting



$$
\begin{aligned}
U_{i,n,t} = ASC_i + \lambda_{consider,i} \log\left(\frac{1}{1+\exp(-LV_{consider,i,n})}\right) \\
+ \lambda_{feasibility\_emp,i} LV_{feasibility\_emp,n} \\
+ \lambda_{feasibility\_pers,i} LV_{feasibility\_pers,n} + \epsilon_{i,n,t}
\end{aligned} \tag{1}
$$

$$LV_{consider,i,n} = \alpha_i + \eta_{i,n}^{LV_{consider}} \tag{2}$$

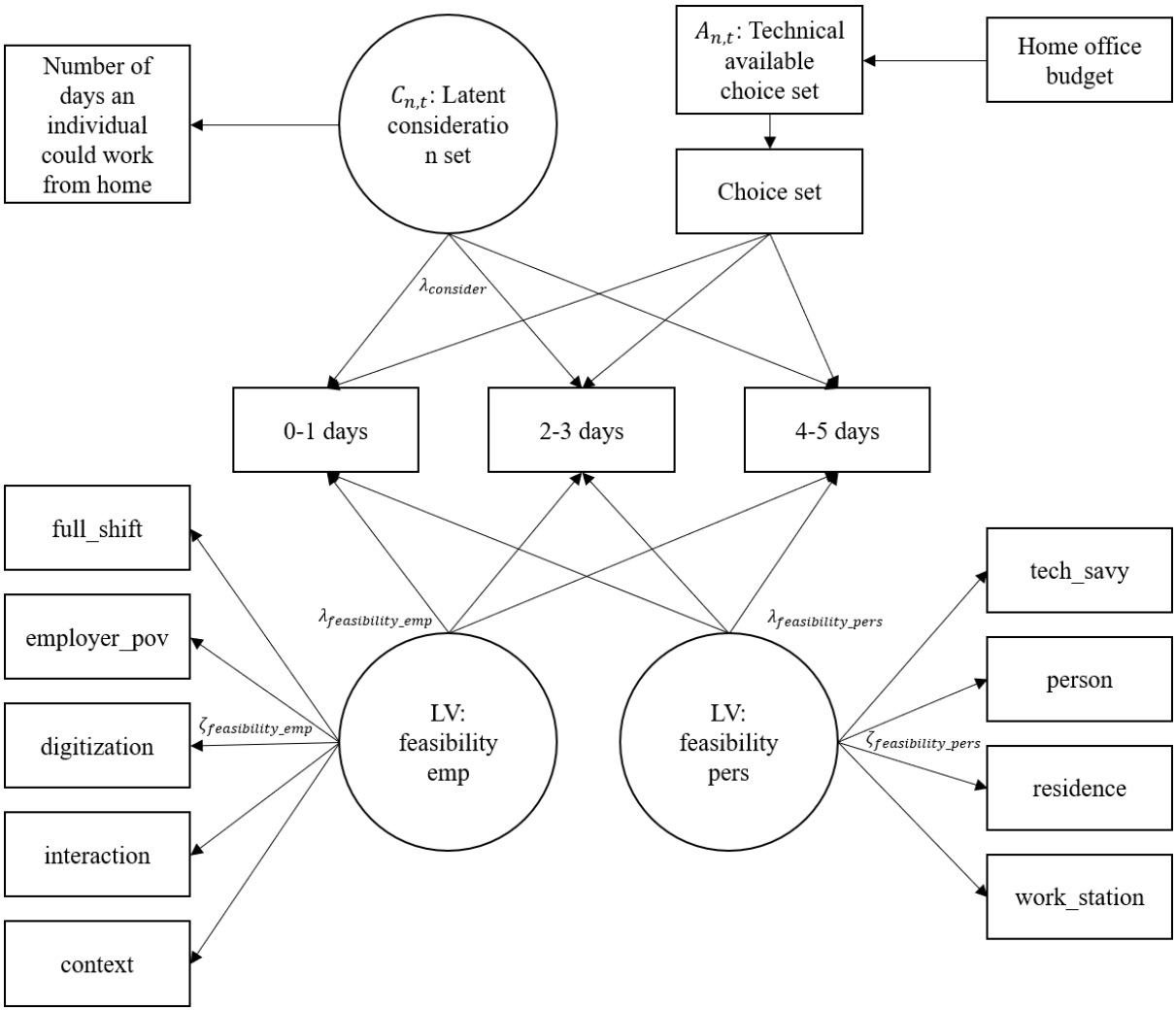
$$LV_{feasibility\_emp,n} = \eta_n^{LV_{feasibility\_emp}} \tag{3}$$

$$LV_{feasibility\_pers,n} = \eta_n^{LV_{feasibility\_pers}} \tag{4}$$

where ASC is the alternative-specific constant, LV stands for latent variables and α, γ are the parameters to be estimated. $\epsilon_{i,n,t}$ and the three η represent error terms.

The underlying mechanism of the endogenous part of the model is illustrated in Fig. 1, and operates as follows: As the latent variable $LV_{consider,i,n}$ becomes more negative, the choice probability of alternative i approaches zero, indicating utility discounting (Schmid *et al.*, 2022b). Two additional latent variables are introduced in the model, namely $LV_{feasibility\_emp}$ which captures the home office suitability work and employment characteristics, and $LV_{feasibility\_pers}$ which captures an individual's perception of his/her home office suitability as well as the suitability of their environment (e.g., residential characteristics, home office work station). A visual representation of the modeling

Figure 2: Path diagram



framework can be found in Fig. 2.

The measurement model of the implied consideration $c_{i,n}$ of alternative i is:

$$P(c_{i,n}|LV_{consider,i,n}) = \frac{\exp\left(LV_{consider,i,n}\right)}{1+\exp\left(LV_{consider,i,n}\right)}^{c_{i,n}} \frac{1}{1+\exp\left(LV_{consider,i,n}\right)}^{1-c_{i,n}} \tag{5}$$

Since the indicators for the residual measurement equations are on a Likert scale, the choice probabilities are obtained from an ordered logit model, assuming that $v_{n,i}$ represents the latent construct underlying the responses of decision-maker n to indicator i (Train,

2009).

$$\begin{aligned} P(y_{k,j,n} = 1|LV_n) &= P(\tau_{k,j+1,n} < v_{n,k} < \tau_{k,j,n}) \\ &= P(\eta_{k,n} < \tau_{k,j,n} - \zeta_k LV_n) - P(\eta_{k,n} < \tau_{k,j+1,n} - \zeta_k LV_n) \\ &= \frac{\exp(\tau_{k,j,n} - \zeta_k LV_n)}{1 + \exp(\tau_{k,j,n} - \zeta_k LV_n)} - \frac{\exp(\tau_{k,j+1,n} - \zeta_k LV_n)}{1 + \exp(\tau_{k,j+1,n} - \zeta_k LV_n)} \end{aligned} \tag{6}$$

In this equation $y_{k,j,n}$ equals one if person n selects option j for indicator k, τ_j is the jth cutoff point and ζ_k can be interpreted as factor loading in this context. The latent variables are defined as in Eq. (3) and Eq. (4).

The main choice model is a simple multinomial logit (MNL).

$$P(i_{n,t}|\mathbf{\Theta}, LV_{consider,n}, \mathbf{LV_n}) = \frac{\exp V_{i,n,t}}{\sum_j V_{j,n,t}} \tag{7}$$

where $\mathbf{LV_n}$ are the two latent feasibility variables $LV_{feasibility\_emp,n}$ and $LV_{feasibility\_pers,n}$.

The probability equations presented above yield the overall log-likelihood function for observing individual n choosing alternative i, as well as choosing all the indicators.

$$\begin{aligned} \mathcal{LL}_n(\mathbf{\Theta}) = &\prod_{t=1}^{T} P(i^*_{n,t}|\mathbf{\Theta}, LV_{consider,n}, \mathbf{LV_n}) \\ &\times \prod_{i=1}^{I} P(c_{i,n}|LV_{consider,i,n}) \\ &\times \prod_{k=1}^{K} P(y_{k,j,n}|LV_{feasibility\_emp,n}) \\ &\times \prod_{k^*=1}^{K^*} P(y_{k^*,j,n}|LV_{feasibility\_pers,n}) \end{aligned} \tag{8}$$

3 Results and discussion

As a first remark, it should be noted that no alternative-specific variables were included in the model, as shown in Eq. (1). The hypothesis was that individuals have a pre-determined fixed number of days they would like to work from home, and then choose the work arrangement that is most favorable given this pre-determined home office frequency. Therefore, our model focuses on the question of what determines this a priori frequency. Additionally, it is expected that sociodemographic variables may influence the choice of home office frequency, but at this stage of the analysis, we simplify the model and only consider the impact of a person's home office feasibility on the chosen frequency. Given an individual's answers to the indicator questions presented in Table Table 1, the main question of interest is how much of the home office variation can be explained by this feasibility measure. One could argue that this model captures the technical rather than the preference dimension of home office frequency determination.

The model parameters and fit indicators are presented in Table 2. The rho-squared, which compares the likelihood of our model to that of the observed shares model (intercepts-only), favors our model. We have also computed the prediction accuracy (PA) by accounting for the probabilistic nature, sampling the choices according to the alternative-specific choice probabilities, as described in Schmid *et al.* (2022a). This approach is expected to yield a more realistic prediction accuracy compared to assuming that the alternative with the highest probability is always chosen, which is known as first preference recovery (see Ortúzar and Willumsen (2011)). The PA calculated from a hold-out sample of 20% of the data was 62%, only slightly higher than the one from the intercept-only model. It should be noted that we compared the model to a specification without the latent consideration dimension, which resulted in minor differences in the lambda coefficients, more noteable differences in the factor loadings, as well as changes in significance levels and major differences in the ASCs (as could be expected). However, the AIC/BIC criteria favored the standard model. Further validation should be conducted with more data from the on-going main field work.

The ASCs now indicate that after controlling for consideration 4-5 days of home office has a higher base-line utility (which is clearly not the case when comparing to observed choice shares).

The analysis of the main MNL model reveals that the effect of consideration is only significant for the 4-5 days alternative, while all other coefficients are significant and have intuitive

Table 2: Model parameters.

	Alternative	Coefficient	Estimate	Robust SE
MNL model:	2-3 days	ASC_23	16.88**	6.29
		lambda_consider_23	107.82	119.03
		lambda_feasibility_emp_23	14.06**	4.84
		lambda_feasibility_pers_23	3.27*	1.56
	4-5 days	ASC_45	30.38**	12.13
		lambda_consider_45	28.46*	14.55
		lambda_feasibility_emp_45	9.26**	3.00
		lambda_feasibility_pers_45	8.51**	3.41
	Latent variable	Factor	Loading	Robust SE
Latent choice :	consideration	zeta_consider_23	0.30*	0.14
		zeta_consider_45	1.18*	0.63
	feasibility_emp	zeta_fully_shift	-0.13	0.27
		zeta_employer_pov	0.20	0.32
		zeta_digitization	0.96*	0.48
		zeta_interaction	-0.73	0.47
		zeta_context	-0.75	0.49
	feasibility_pers	zeta_tech_savy	0.59*	0.31
		zeta_person	2.49***	0.76
		zeta_residence	3.01***	0.94
		zeta_work_station	3.89**	1.66
Fit indicators:		Rho-squared vs equal shares	0.26	
		Adj. Rho-squared vs equal shares	0.22	
		Rho-squared vs observed shares	0.18	
		Adj. Rho-squared vs observed shares	0.14	

Note: $^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

signs. Interestingly, the feasibility coefficients derived from the two latent variables, namely *lambda_feasibility_emp* and *lambda_feasibility_pers*, appear somewhat counter-intuitive at first sight. Specifically, one would expect a higher *lambda_feasibility_emp* coefficient for the 4-5 days alternative compared to the 2-3 days alternative, indicating that a job with higher technical feasibility for home office would result in a higher propensity to work from home for 4-5 days. However, our findings suggest that the feasibility to work from home 4-5 days is primarily influenced by personal suitability factors, both of the person itself and his/her environment, as indicated by the *lambda_feasibility_pers* coefficient. This implies that individual preferences and personal circumstances play a stronger role in determining the feasibility of working from home for an extended period, rather than just job and employment characteristics.

Upon examining the corresponding factor loadings, it appears that the home office work station is a decisive factor in influencing the feasibility of working from home, followed

by the quality of the residential environment and the perceived personal suitability for home office work. All three factors exhibit substantial loadings on the latent factor, indicating their meaningful contribution to determining the feasibility of home office work. Additionally, the technological savviness of an individual, although less influential, still has a discernible impact on the feasibility of working from home.

The factor loadings of *feasibility_ emp* are relatively less substantial and, except for the job's degree of digitization, are not statistically significant.

4 Conclusion

The initial motivation for this analysis was the observation that SP attributes did not significantly affect home office frequency choice behavior. Instead, it appeared that respondents had a pre-determined preference for a specific home office frequency and chose the work arrangement that maximized their utility based on this preconception. As a result, the question arose as to what factors determine this a priori frequency choice. This study adopts the perspective that a latent home office feasibility factor governs the observed number of home office days.

The distinction between 0-1 home office days and 2-3 home office days is primarily influenced by what we refer to as employment feasibility, whereas personal feasibility (which encompasses perceived personal suitability, suitability of the residential environment, and the home office work station) plays a relatively more influential role in determining the alternative of 4-5 home office days.

While the constituents of employment feasibility could not be fully disentangled in our analysis (with only the degree of digitization found to be a significant factor), the personal feasibility dimension is clearer. All considered factors, including perceived personal suitability, residential environment, and quality of the home office work station, load substantially on the latent variable and play important roles in determining home office frequency choice behavior.

Our feasibility model demonstrates a prediction accuracy of approximately 62% and a reasonably high rho-squared measure. However, further efforts should be directed towards understanding the remaining variation, particularly by incorporating socioeconomic vari-

ables once more data is available. Additionally, extending the structural equations of the latent variables would allow for a more comprehensive explanation of their contributors.

We have demonstrated a meaningful application of an endogenous choice set formation approach. However, further testing with a more comprehensive data set is needed, including benchmarking against non-stochastic choice set (classical) specifications as well as latent class models. In particular, the latent class models should incorporate class-specific choice sets and the class allocation model should consider our feasibility considerations.

5 References

Dingel, J. I. and B. Neiman (2020) How many jobs can be done at home?, *Journal of Public Economics*, **189**, 104235.

Ortúzar, J. d. D. and L. G. Willumsen (2011) *Modelling Transport*, John Wiley & Sons, Hoboken.

Schmid, B., F. Becker, J. Molloy, K. W. Axhausen, J. Lüdering, J. Hagen and A. Blome (2022a) Modeling train route decisions during track works, *Journal of Rail Transport Planning & Management*, **22**, 100320.

Schmid, B., J. Swait, H. T. Kassahun and M. C. Bliemer (2022b) Endogenous choice set formation model, paper presented at the *7th International Choice Modelling Conference (ICMC 2022)*, Reykjavic.

Sener, I. N. and C. R. Bhat (2011) A copula-based sample selection model of telecommuting choice and frequency, *Environment and Planning A*, **43**, 126–145.

Swait, J. D. (1984) Probabilistic choice set generation in transportation demand models, Ph.D. Thesis, Massachusetts Institute of Technology.

Train, K. E. (2009) *Discrete Choice Methods with Simulation*, Cambridge University Press.