Does Where You Live Affect How You Feel? Causal Evidence from an Integrated Econometric and Machine Learning Framework

Jerry Chen JC2205@CAM.AC.UK
University of Cambridge, Cambridge, United Kingdom

Li Wan LW423@CAM.AC.UK
University of Cambridge, Cambridge, United Kingdom

Abstract

We investigate whether residential relocation causally improves subjective wellbeing by leveraging household relocations in the UK Household Longitudinal Survey as natural experiments. An integrated framework combining a difference-in-differences and synthetic control ensemble with a causal forest model is applied to nearly a decade of panel data. Relocation causes an immediate and sustained improvement of 8% in subjective wellbeing; a change in the built environment type (e.g. subrurb to city) adds a further 5%. We demonstrate the complementarity and interoperability of canonical econometric and machine learning methods for causal inference on subjective panel data.

Data and Code Availability. Our analysis uses the UK Household Longitudinal Survey (Understanding Society), accessible via the UK Data Service. Geographical linkage data (LSOA-level residential location) was obtained through a Special License granted by the Office for National Statistics. ONS Area Classification, Indices of Multiple Deprivation, and Access to Health Assets and Hazards indices are publicly available. Project code is available upon request.

Institutional Review Board (IRB). IRB approval was not required as all data used are publicly available and fully deidentified.

1. Introduction

Measuring subjective wellbeing (SWB) is an essential part of assessing the progress of societies (OECD, 2013). Evaluations of inner feelings must be self-reported and pose many challenges in validity and interpretability when modelled (Krueger and Stone, 2014). Causal factors of SWB in observational studies are complex and difficult to determine (Diener et al., 2017). Domain knowledge from psychology (Byrne, 2016; Wu et al., 2022), epidemiology (Hernán and Robins, 2020), and economics (Abadie et al., 2010; Imbens and Rubin, 2015) enabled counterfactual modelling, but there is a distinct lack of studies making use of subjective data. Counterfactual analysis is at the heart of observational causal inference: canonical statistical methods can be integrated with emerging machine learning (ML) methods to deliver more accurate and policy-relevant models.

Constructing a causal model suitable for the complexity of a large-scale panel SWB dataset requires a comprehensive set of covariates, significantly increasing the data dimensionality. Due to the idiosyncratic nature of subjectivity, SWB data is noisy and nonlinear (Chen and Wan, 2024); estimating treatment effects using SWB as outcome is likely heterogeneous. Capturing the treatment heterogeneity is essential for inter-personal comparison and providing generalisable policy insights; the partitioning of population according to relevant covariates is a key strength of the forest-based ML method.

The COVID-19 pandemic accelerated residential relocation, with evidence of sustained urban-to-suburban and rural moves across the U.S., U.K., and beyond, driven by infection avoidance, environmental preferences, and a search for agency (Batty, 2020, 2022; Batty et al., 2022; Haslag and Weagley, 2022; Wang et al., 2022). A substantial literature links relocation to SWB: relocations reduce environmental stressors (Wolpert, 1965), yield enduring improvements in housing and neighbourhood satisfaction (Nowok et al., 2018; Clark, 2016), and lower suicide mortality among movers (Hagedoorn and Helbich, 2022). Though the correlation vary with hous-

ing tenure, life course events, and whether the move is voluntary or forced (Foye et al., 2018; Kang, 2019).

Built environment characteristics (such as neighbourhood aesthetics, greenspace access, noise, and pollution) are consistently associated with SWB outcomes, with non-linear effects documented using both cross-sectional (Yang et al., 2021; Huebner et al., 2022) and systematic review methods (Cooper et al., 2010; Krefis et al., 2018). However, causal identification of the relationship between built environment change and SWB remains a gap: most studies rely on correlational designs that cannot exclude reverse causality or confounding, and there is no conclusive longitudinal evidence linking relocation-induced built environment change to SWB outcomes.

We design a natural experiment using household relocation as treatment to investigate the causality between built environment and SWB. We start with statistical learning using an ensembled model of staggered difference-in-differences (DiD; Callaway and Sant'Anna 2021) and generalised synthetic control methods (SCM; Xu 2017). The ensemble model has semiparametric assumptions and relatively low data dimensionality. The statistical model directly quantifies the causal relationship between relocation and SWB, and indirectly quantifies the causal relationship between built environment change and SWB.

We build upon the statistical counterfactual model by integrating a forest-based causal learning model (Athey and Imbens, 2016; Athey et al., 2019b; Athey and Wager, 2019). The ML approach has distinct advantages to complement our statistical method: (1) integrated causal model framework; (2) direct quantification of heterogeneous treatment effects; (3) data-driven selection of covariates; (4) interpretable policy decisions. There are limited methodological options in ML to robustly handle panel data; our integrated application demonstrates interoperability between statistical and ML methods for counterfactual causal inference.

2. Data and Methods

In this section we introduce the main dataset used in our integrated application – the UK Household Longitudinal Survey and its linked geographical data. We then describe the data cleaning process and the sample inclusion criteria. The SWB outcome measurement is validated and refined to maximise model sensitivity. The statistical methods rely on an ensemble model which combines estimations from DiD and SCM. The ML model, given its empirical novelty, has significant limitations with temporal data. This is a particular challenge for tree- and forest-based ML models to address temporal dependencies and non-stationarity. The panel data is restructured for compatibility with causal forest model specifications. We propose a unified framework for integrating the ML model with the statistical ensemble model.

2.1. Data Availability

Our research is enabled by the longitudinal Understanding Society dataset (UK Household Longitudinal Survey) – the largest household panel study of its kind. With its first wave collected in 2009, our study includes responses as recently gathered as 2019. The decade-long perspective provides crucial information for investigating changes and factors affecting people's lived experiences. Not only does the dataset provide a comprehensive set of demographic and socioeconomic indicators on the respondents, it also includes well-established SWB indicators. General Health Questionnaire-12 (GHQ-12) is a 12-item questionnaire that measures short-term change in mental health and psychological functioning – it is widely and reliably used in community settings to detect a range of symptom and behaviour related to distress. It has also been extensively validated and deployed to measure broader mental health and SWB outcomes (Werneke et al., 2000; Dolan et al., 2011).

Geographical data linkage to the Understanding Society dataset allows us to identify individuals who reveal a change in residential location over the study period. Respondents report their residential location annually. Residential location at the Lower Super Output Area (LSOA) level is acquired through a Special License granted by the Office for National Statistics. LSOA-level residential location enables the following data to be linked:

1. **ONS Residential-based Area Classification (OAC)**: applying K-means clustering to 2011 UK Census variables in five domains (housing, demographic structure, household composition, socioeconomic character, and employment) to produce 8 supergroups covering all LSOAs in England; this well-established, holistic classification is used as a proxy measurement for broad built environment types, as summarised in Appendix A;

2. **Travel to Work Area (TTWA)**: geography created to approximate labour market areas (self-contained areas in which most people both live and work);

3. **English Indices of Multiple Deprivation (IMD)**: small-area composite measurement of seven main types of deprivation including income, employment, education, health, crime, access to housing and services, and living environment;

4. **Access to Health Assets and Hazards (AHAH)**: multi-dimensional index (Daras et al., 2019) concerning four domains of accessibility (retail environment, health services, physical environment, and air quality) measuring how 'healthy' neighbourhoods are.

2.2. Statistical Ensemble Model Samples

The statistical model samples must have a balanced panel of all 11 years over the survey period (2009–2019), including a complete set of responses related to the GHQ-12 questionnaire items. Counterfactual controls are constructed from individuals who never relocated over the survey period. The treatment group consists of individuals who changed residential locations at the LSOA level only once during the survey period. Those who moved before 2012 or after 2016 are further excluded. This requirement follows the treatment irreversibility assumption and allows for sufficient data points for estimating the pre- and post-treatment effects. Treated movers are further bifurcated according to change in OAC type to indirectly investigate the causality between built environment change and SWB.

Figure 1 shows the statistical model data cleaning and merging process with sample sizes at each stage.

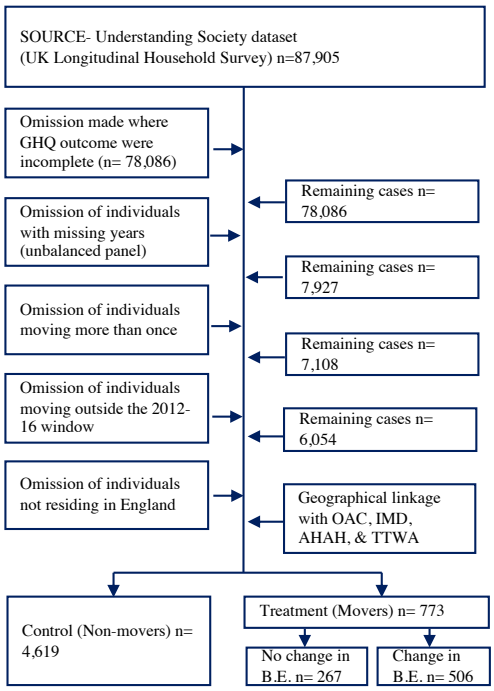


Figure 1: Statistical model sample construction

2.3. Machine Learning Model Samples

Figure 2 shows the ML model data cleaning and merging process with sample sizes at each stage.

The ML model has less stringent sample inclusion requirements. It builds on the statistical model and directly quantifies the heterogeneous treatment effects of built environment change. Rather than using change in built environment as a mediator of household relocation, both control and treatment groups in the ML model relocated – the change in built environment is treatment. As such, we compare moving

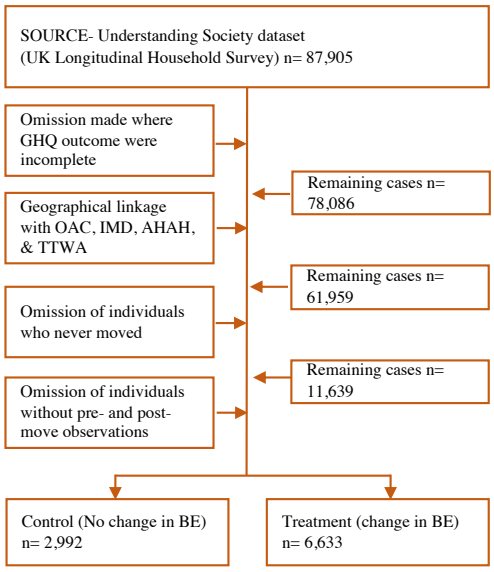


Figure 2: ML model sample construction

to a new OAC (treatment, e.g., city to countryside) to moving to the same OAC (control, e.g., suburb to suburb). As the inclusion criteria are broader, relocated sample sizes increased tenfold with regards to built environment change as treatment.

2.4. Conceptual Framework

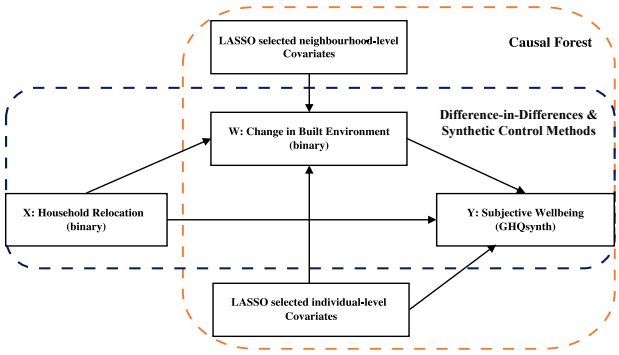


Figure 3: Conceptual framework for integration of statistical ensemble and ML models

Figure 3 shows the conceptual framework for the integrated application of statistical ensemble and ML models, taking advantage of their respective strengths. SCM and DiD methods robustly model longitudinal data, and offer interpretable visualisation of the counterfactual. Causal forest methods capture heterogeneity in the treatment effects. High-dimensional data enables better control on confounders and a data-driven selection of covariates. A unified framework allows for cross-validation of estimation results, and more importantly, predicts relocation effects and identifies subgroups most in need of built environment interventions.

2.5. Outcome Variable Refinement

We use two methods to refine our main outcome variable: first, we follow the literature (e.g., Hystad and Johnsen, 2020) and apply principal component analysis to identify a subset of the GHQ-12 questions with strong internal consistency and validity using our study samples. Second, to minimise noise and maximise variance explainable by our independent variables, we apply a random effects model and rotate each of the GHQ-12 items as the outcome variable to find the highest R^2 values. Only the treatment group samples are used in our parameter tests. The models include a binary treatment period indicator ("pre-relocation = 0"), as well as the relocation type which are vectors of "origin OAC – destination OAC" (e.g., suburban–rural).

Covariates are controlled at individual and neighbourhood levels. VIF and Hausman tests are performed to confirm model consistency, and we conclude low multicollinearity. A new variable *GHQsynth* is the composite score of the subset of GHQ-12 items most relevant to our model. GHQsynth is used as the main SWB outcome in subsequent analyses. GHQsynth partially mitigates measurement noise through data-driven item selection but cannot eliminate error inherent to subjective self-report.

2.6. Covariates Selection

For neighbourhood-level covariates: the two main covariates used are the change in Indices of Multiple Deprivation (IMD) and the change in Access to Health Assets and Hazards (AHAH). Both variables are measured in decile form where 1 is most deprived or has the least access to amenities, respectively. Both indices are composites from a diverse range of measures related to the built environment such as density, housing type, housing tenure, access to green space, and pollution levels. The constituent variables of IMD (e.g., income, education, health, crime rates, geographical barriers, living environment) and AHAH (e.g., access to GP, pharmacy, fast food, gambling, green space; as well as NO_2, PM_{10} and noise exposures) are exhaustively used in the ML model for covariates selection.

Individual-level covariates inform an unobserved latent lifestyle factor that potentially confounds the causal relationship. Demographic covariates include sex (binary), age (continuous), and residence in London (binary). Socioeconomic covariates include employment status (binary), workplace (home or office based), NSSEC classification (management, intermediate, semi-routine, manual), and degree (binary). To better capture the change in latent lifestyle, we also include binary variables representing circumstantial changes such as promotion, demotion, retirement, maternity leave, long-term sick leave, and family-care. Changes in paid work hours, overtime hours, and commuting times are also captured as continuous covariates in the ML model.

2.7. Statistical Ensemble Methods

We employ two causal models (DiD and SCM) to quantify the potential causal effects. We follow a vertical ensemble design for panel causal effects modelling (Athey et al., 2019a). Due to multiple treatment units and staggered timing, the models are separately estimated before the ensemble weightings are applied to calculate the causal effects.

In the staggered DiD model, we only observe one potential outcome for each unit. For those who do not relocate in any year, observed outcomes are untreated potential outcomes in all years. For individuals who do relocate, observed outcomes are the unit-specific potential outcomes corresponding to the particular year the individual first relocated. We define G_i as the year an individual first relocates; $D_{i,t}$ is then a binary variable equal to 1 if individual i first relocates in year t. There are T years and we denote a particular year by t, where $t \in \{2, \ldots, T\}$.

We let $Y_{i,t}(g)$ denote the potential outcome that individual i would experience at time t if they were first to relocate in year g. To incorporate the generalisation of the Average Treatment Effect on the Treated (ATT) applicable to data structures with multiple treatment groups and multiple time periods, we use the ATT for individuals who are members of a particular group g at a particular time period t. The ATT for each group g, t is:

$$ATT(g,t) = \mathbb{E}\left[Y_t(g) - Y_t(0) \mid G_g = 1\right] \tag{1}$$

In other words, the ATT calculated – at year t for a given group g of individuals relocating in the same year – is the expected GHQsynth difference between the movers and non-movers. The DGP for the average staggered DiD causal effect follows the form:

$$\theta = \sum_{g \in \mathcal{G}} \sum_{t=1}^{T} w(g,t) \cdot ATT(g,t) \tag{2}$$

where $w_{g,t}$ are weighting functions. Our model adopts the doubly-robust estimator (Sant'Anna and Zhao, 2020) that refines individual fixed-effects estimations to account for unobserved confounders and certain types of anticipatory effects.

In the generalised SCM model, Xu (2017) follows the canonical SCM as conceptualised in Abadie et al. (2010) and treats the treatment effect as given once the sample is drawn. $Y_{i,t}$ is the outcome of interest of individual i at time t. $D_{i,t}$ is a binary variable equal to 1 if individual i has relocated prior to year t, and 0 otherwise. Let $Y_{i,t}(0)$ and $Y_{i,t}(1)$ be the GHQsynth outcome for individual i at time t when $D_{i,t} = 0$ and $D_{i,t} = 1$, respectively. Let $\mathcal{T}^r$ denote the set of units in treatment. N^{tr} denotes the number of individuals in the treatment group. ATT in the post-treatment periods $t > T_0^i$ are obtained with the following DGP:

$$\text{ATT}_t = \frac{1}{N^{\text{tr}}} \sum_{i \in \mathcal{T}} \left[Y_{i,t}(1) - Y_{i,t}(0)\right] = \frac{1}{N^{\text{tr}}} \sum_{i \in \mathcal{T}} \left[\delta_{i,t}\right] \tag{3}$$

From this formula, the ATT is a simple mean of the differences between the observed treated outcomes and the synthetically generated counterfactual. $Y_{i,t}(1)$ is observed for treated units in post-treatment periods; hence the crux of this generalised method is to construct counterfactuals for each treated unit in post-treatment periods – turning the causal inference problem into one of forecasting missing data.

2.8. Machine Learning Methods

We follow the causal tree (Athey and Imbens, 2016) and extension causal forest (Athey and Wager, 2019) frameworks as our main ML methods. As with the statistical methods, causal effects are defined via a counterfactual model. $Y_i(1)$ and $Y_i(0)$ denote treated and control potential outcomes; W_i and X_i denote treatment and covariates for individual i; S_i is the partitioning covariate. The conditional average treatment effect (CATE) is:

$$\text{CATE}_x = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = x\right] \tag{4}$$

The causal forest framework estimates CATE for a target sample x by running a weighted residual-on-residual regression on samples with similar treatment effects. The model does this in two steps: (1) it first builds greedy trees which partition covariates to maximise the squared difference between subgroups; (2) it then aggregates each tree's prediction by taking a weighted average of outcomes that fall into the terminal leaf. The causal forest uses an adaptive "honest" weighting to give asymptotic guarantees:

$$\text{CATE}_x = \text{lm}\Big(Y_i - \hat{m}^{(-i)}(X_i) \sim W_i - \hat{e}^{(-i)}(X_i), \quad \text{weights} = \alpha_i(x)\Big) \tag{5}$$

where $\hat{m}^{(-i)}$ is the inversely cross-fitted conditional mean of outcome Y_i, and $\hat{e}^{(-i)}$ is the inversely cross-fitted propensity score. $\alpha_i(x)$ is the frequency with

which the i-th sample falls into the same leaf as x. Since the CATE estimation does not incorporate a temporal dimension, our SWB panel data is converted into a cross-sectional representation of the change in GHQsynth between pre- and post-relocation periods.

Data-driven selection of covariates set $\mathcal{X}$ is a crucial process and a key advantage of the forest-based model. To assess model performance and better understand heterogeneous treatment effects, we further estimate the rank-weighted average treatment effect (RATE; Yadlowsky et al. 2021). The targeting operating characteristic (TOC) is a curve comparing the overall ATE benefit of treating only a certain fraction of units. RATE is a weighted sum of the TOC curve:

$$\begin{aligned} \mathrm{TOC} = \mathbb{E}\left[Y_i(1) - Y_i(0) \mid F(S(X_i)) \geq 1 - q\right] \\ -\mathbb{E}\left[Y_i(1) - Y_i(0)\right] \end{aligned} \tag{6}$$

where S is a rule assigning scores to samples in decreasing order of treatment prioritisation, F is the distribution function of the prioritisation rule, and q represents the fraction of treated units (i.e., at $q = 0.25$ the TOC quantifies the incremental benefit of treating only the 25% largest CATEs compared with the overall ATE).

To demonstrate the empirical relevance of the causal forest method, we apply the policy tree framework of Athey and Wager (2021) to inform decisions to relocate and change built environment. Policy tree solves the treatment assignment problem:

$$\hat{\pi} = \mathrm{argmax}_{\pi \in \Pi} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left(2\pi(X_i) - 1\right) \hat{\Gamma}_i \right\} \tag{7}$$

where Π is a pre-specified policy class and Γ_i is the doubly robust score for the target estimand (Sverdrup et al., 2020). The optimal tree maximises the sum of rewards. Simple but globally optimal decision trees are easily interpretable and suitable for assessing treatment options.

3. Results

In this section we first present the results from refining the outcome variable. We sequentially report results from the statistical and ML models. The statistical ensemble model is estimated first: DiD is implemented in Stata using the `csdid` package (Rios-Avila et al., 2021); SCM is implemented in R using the `gsynth` package (Xu and Liu, 2022); the causal forest model is executed in R using the `grf` package (Athey et al., 2019b).

3.1. Refining Subjective Wellbeing Outcome

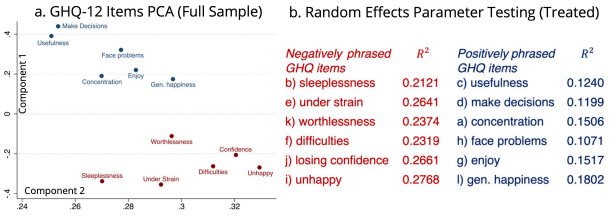


Figure 4: Refining process of model outcome variable. (a) Principal component analysis of GHQ-12 items. (b) Random effects parameter testing by R^2 values.

Figure 4a shows the results from the principal component analysis for all items contained in the GHQ-12. Component 1 explains 48.90% of the variance, and Component 2 explains 10.37% of additional variance. While the Component 1 loadings are fairly evenly distributed for the 12 items, two distinct groups emerge based on their loadings in Component 2.

The negatively phrased items identified using PCA have consistently higher R^2 values as shown in Figure 4b, indicating higher explanatory power of the dependent variables. Our composite SWB outcome measurement – *GHQsynth* – is constructed using a simple sum of the six negatively phrased GHQ items. The main outcome variable measuring SWB is ranged from 6 (lowest distress) to 24 (highest distress). Iterative latent class analysis to visualise SWB patterns across the sample is included in Appendix B.

3.2. Statistical Ensemble Model Results

We first present the results from the statistical ensemble model. The generalised SCM is particularly robust in producing counterfactuals, addressing key issues that limit the applicability of causal inference in observational studies: multiple treatment timing and multiple treatment units. We show the visualised SCM results for Models 1, 2, and 3; the equivalent DiD visualisations are included in Appendix C, D, E. The weighted average accounts for the post-treatment ATTs by computing a time-weighted mean

of individual $ATT(g,t)$ values within each model. The ensemble ATE is a simple equal-weighted mean of the DiD and SCM weighted averages: the two constituent models are not weighted by variance or pre-treatment fit, but rather treated symmetrically as lower and upper bounds of the ensemble. Bacon decomposition to assess the effects of staggered treatment adoptions is included as a robustness test in Appendix F.

Table 1 summarises the ATT results. Model 1 shows an immediate negative treatment effect in the year following relocation, representing a decrease in distress level in the treatment group as compared to the counterfactual. Moreover, the effect of decreasing distress appears sustained for the entire post-relocation period of our study, despite diminishing significance in the SCM model. Based on the ensembled ATE, our model results suggest that residential relocation causes an immediate and sustainable decrease in distress and improvements to SWB.

Model 2 shows movers to a different built environment experience decreased distress levels following relocation, and the effect is sustained for the 5-year post-relocation period. The ensemble ATE shows a larger magnitude improvement in SWB than relocation in general. Model 3 shows treatment estimates are not statistically significant, indicating a lack of causal effect from residential relocation without change in built environment. Comparing Models 2 and 3 infers that the positive causal effects from relocation shown in Model 1 are attributable to changes in built environment. We indirectly observe a causal, positive relationship between built environment changes and SWB changes.

Figure 5 shows the effects of relocation on subjective wellbeing from SCM Model 1 comparing movers to non-movers. Comparable DiD results are in Appendix C.

Figures 5a and 5b provide contextual support for our estimations. Specifically, we observe the large ratio between the control and the treatment groups, suggesting a sufficiently large pool of control donors from which the synthetic counterfactuals are constructed. Within the treatment group, we observe even distribution of treatment timing (i.e., the year of relocation). Figure 5b shows the variance of individual SWB data and the notable difference between the treated and the synthetic counterfactual (i.e., the projected distress level of relocated individuals had they not relocated).

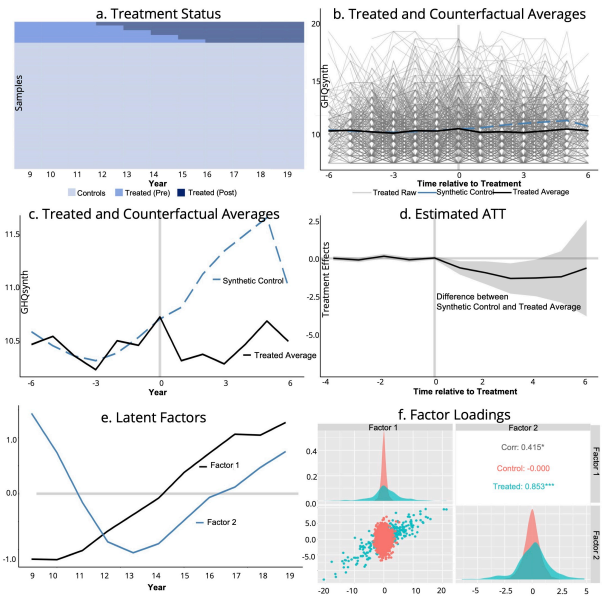


Figure 5: SCM Model 1: effects of relocation on subjective wellbeing (movers vs. non-movers). (a) Sample composition. (b) Individual SWB variance. (c) Counterfactual vs. observed treated average. (d) Average treatment effect with 95% confidence path. (e) Latent factors. (f) Latent factor loadings.

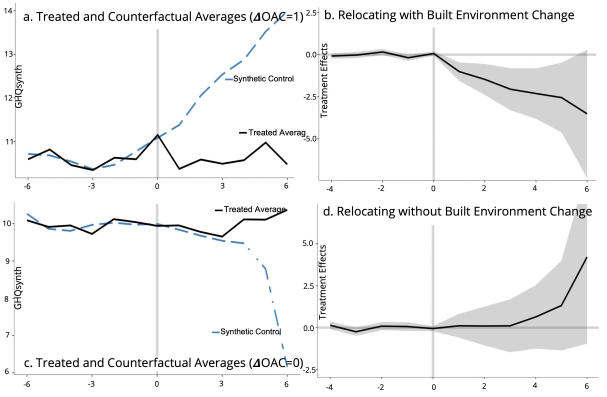


Figure 6: SCM Models 2 and 3: indirect comparative analysis of built environment change on SWB. (a) Counterfactual vs. observed, Model 2. (b) ATT with 95% CI, Model 2. (c) Counterfactual vs. observed, Model 3. (d) ATT with 95% CI, Model 3.

Table 1: Summary of the Average Treatment Effect on the Treated (ATT) for three causal models: Model 1 (Relocation); Model 2 (Relocation with built environment change); Model 3 (Relocation without built environment change). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

	Model 1		**Model 2**		**Model 3**	
T	DiD	SCM	DiD	SCM	DiD	SCM
-3	0.496^{**}	-0.101	0.336^{*}	-0.020	0.327	-0.242
-2	-0.003	0.137	-0.098	0.161	-0.120	0.094
-1	0.469^{***}	-0.089	0.463^{**}	-0.181	0.105	0.070
0	-0.635^{***}	0.023	-0.437^{**}	0.066	-0.262	-0.052
1	-0.708^{***}	-0.598^{**}	-0.501^{***}	-1.008^{***}	-0.184	0.114
2	-0.915^{***}	-0.896^{*}	-0.591^{***}	-1.470^{***}	-0.518	0.100
3	-0.762^{***}	-1.262^{**}	-0.668^{***}	-2.053^{***}	0.018	0.112
4	-0.863^{***}	-1.230^{*}	-0.395^{*}	-2.308^{**}	-0.265	0.641
5	-0.984^{***}	-1.157	-0.675^{***}	-2.549^{*}	-0.034	1.318
Weighted Avg.	-0.808^{***}	-0.991^{*}	-0.544^{***}	-1.963^{*}	-0.208	0.697
Ensemble ATE	-0.899		-1.243		0.245	

The counterfactual difference is further illustrated in Figure 5c. In the pre-treatment period, the synthetic counterfactual closely follows the trend of the observed treated average. Significant divergence occurs from the time of relocation ($T = 0$); the synthetic counterfactual continues to follow the pre-treatment trend based on latent factor regressions on the control group, whereas the average SWB of the treated samples starts to deviate from the pre-treatment trend.

Figure 5d shows the average treatment effect (i.e., the difference between treated and counterfactual) with 95% confidence path generated from 10,000 bootstrap repetitions. The first-year post-relocation coefficient is -0.598 (95% CI -1.177 to -0.021) in the SCM model. Continuing the trend of decreasing distress, from the third year after relocation the SCM estimates overtake the DiD estimates in magnitude: at $T = 3$ distress level reaches its local minimum at -1.262 (95% CI -2.213 to -0.310). Thereafter, the distress level increases marginally with diminishing statistical significance.

Figures 5e and 5f demonstrate the latent factor regression results used to generate the synthetic counterfactual. The generalised SCM model automatically identifies the optimal latent factors by minimising the mean squared prediction error, using data from the control group only. In this application there are two estimated factors depicted in Figure 5e, where the y-axis corresponds to the magnitude of latent factors (square root of eigenvalues to demonstrate relative importance).

Figure 6 shows the results from SCM Models 2 and 3 for an indirect comparative analysis on the effect of built environment changes on SWB. Comparable DiD results are in Appendix D and E.

Parameters of SCM Models 2 and 3 are identical to those used in Model 1. The treatment group in Model 1 consisted of all movers; in this step we further bifurcate the treatments based on whether or not they changed built environment upon relocation. The control groups remain consistent as all those who did not relocate during the survey period. Hence the SCM latent factor analysis returns identical results to Figures 5e and 5f and is therefore omitted.

Figures 6a and 6b show the SCM model retaining those who changed built environment as the treatment group. We observe accurate matching between the observed treated average and counterfactual, which rapidly diverges from the treated average from the first year post-relocation. Figure 6b shows the coefficient lowers from 0.066 (95% CI -0.058 to 0.192) at $T = 0$ to -1.008 (95% CI -1.562 to -0.454) at $T = 1$, continuing the downward trend until the end of our study period. ATT overall is -1.963 (95% CI -3.289 to -0.638), $p = 0.003$.

Figures 6c and 6d show the SCM model retaining those who relocated but did not change built environment. We observe close matching from the pre-treatment period continuing into the post-relocation years, with no significant divergence and the confidence path spanning both negative and positive values. ATT overall is 0.697 (95% CI −1.008 to 2.402), $p = 0.423$. We conclude no causal effect from relocation with no change in built environment.

3.3. Forest-based Machine Learning Model Results

The Average Treatment Effect estimated from the causal forest model is −0.458 (95% CI −0.594 to −0.317), in line with the levels indirectly estimated by the statistical ensemble model (the difference between Models 1 and 2). We summarise CATE based on partitioning covariates using best linear projection, and present RATE to assess CATE estimator performance. 1-depth and 2-depth policy trees are used to directly interpret the empirical findings. Overlap and fit are assessed with propensity score distribution and bias/standard deviation, respectively in Appendix G.

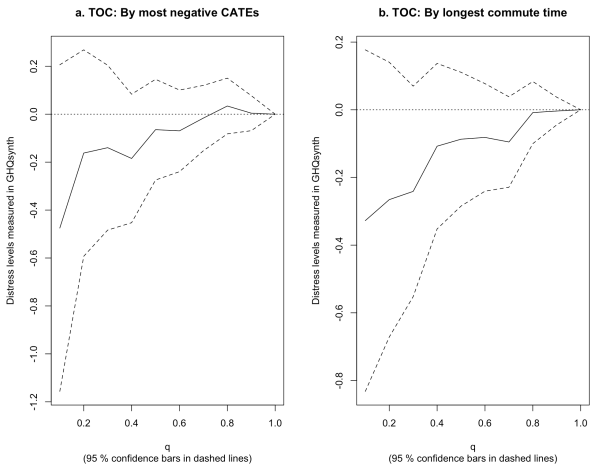


Figure 7: RATE assessment of the causal forest model using TOC. (a) by most negative CATEs. (b) by decreasing commute time.

Figure 7a visualises the heterogeneous treatment effects using RATE, where q indicates the selected treatment proportion of samples. The y-axis corresponds to the additional decrease in distress level from treating only the subset of samples. Subsets of the treated population benefit significantly more than others, demonstrating treatment heterogeneity (uniform treatment effects would result in a flat TOC).

Figure 7b contextualises treatment heterogeneity using commute time. The smaller q corresponds to samples spending the longest time commuting ($q = 0.1$ corresponds to the top 10% longest commute times). In general, moving to a different built environment delivers larger SWB improvement to longer commuters. An intriguing "stair-shaped" pattern in treatment heterogeneity is observed, where the TOC curve plateaus for certain ranges of q. Samples can be divided into three subsets – long, medium, and short commuters – for whom changing to a new built environment delivers approximately −0.3, −0.1, and ≈ 0 additional decrease in distress, respectively.

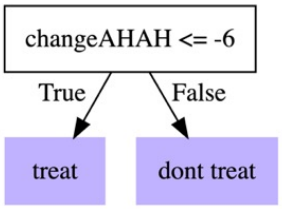


Figure 8: 1-depth policy tree (node size = 500) for built environment change as treatment.

Figure 8 demonstrates the application of the policy tree and its ability to offer simple, interpretable policy recommendations. AHAH is measured in deciles where 1 signifies least access to amenities. The partitioning criterion is whether one relocates to a significantly less accessible area (more than 60%). Decision node "treat" on the policy tree has an ATE of 0.654 (95% CI −0.139 to 1.448); "don't treat" has an ATE of −0.330 (95% CI −0.139 to 1.448). Given relocating to a neighbourhood with significantly decreased accessibility, changing built environment would increase distress. If relocating to a more accessible area, even without a change in built environment the distress level decreases.

Figure 9 demonstrates a more complex double-layered policy tree based on two partitioning criteria with four possible decision nodes. For an unemployed person outside London, relocating without change in built environment (node 1: don't treat) has ATE = 0.251 (95% CI 0.119 to 0.347); an unemployed person moving to a new built environment in

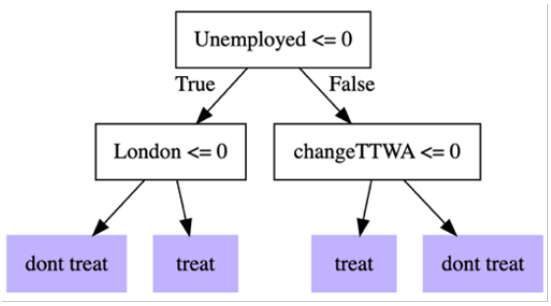


Figure 9: 2-depth policy tree (node size = 100) for built environment change as treatment.

London (node 2: treat) has ATE = 0.375 (95% CI 0.133 to 0.618). For an employed person whose move likely requires changing jobs (changing travel-to-work area), changing built environment (node 3: treat) has ATE = −0.582 (95% CI −1.297 to 0.133). For an employed person relocating long-distance without change in built environment (node 4: don't treat), ATE = −0.168 (95% CI −0.660 to 0.324).

4. Discussion

Figure 10 contextualises results from the integrated model using sample density and empirical cumulative distribution function (ECDF).

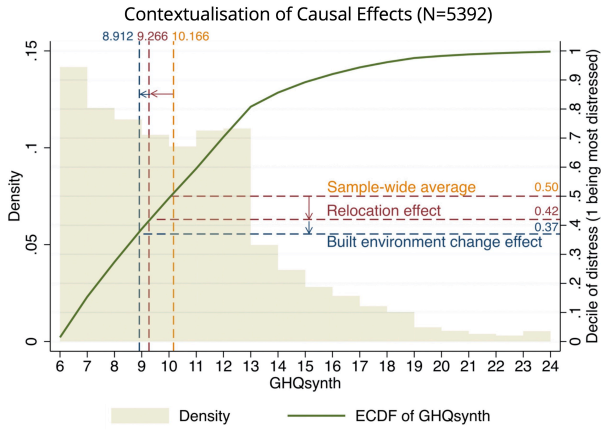


Figure 10: Contextualisation of results using sample density and empirical cumulative distribution function of GHQsynth.

The integrated model results are applied to a hypothetical individual with average SWB. The GHQsynth sample mean is 10.166. Our integrated model shows immediate and enduring positive causal effects of relocation, equivalent to an average improvement of 8% in SWB level compared to the sample population. Moving to a different built environment improves SWB by an additional 5%. Without a change in built environment type, the positive causal effects become negligible. Within the GHQsynth interval of 6–13 (>80% of samples), the slope of the ECDF is approximately linear, suggesting the contextualisation of causal effects is generalisable.

Household relocations are not randomly assigned; unobservable life events, health shocks, and measurement error in residential location cannot be eliminated from observational data of this kind. The methodology is designed to minimise confounding: covariates are controlled at individual and neighbourhood levels, two comparable model designs are applied to the same population-representative longitudinal dataset, and the doubly-robust estimator remains consistent provided either the propensity score or outcome model is correctly specified. The null result for Model 3 (Table 1) is consistent with built environment change, rather than selection into relocation per se, driving the observed SWB improvements. The treatment irreversibility assumption is enforced by excluding samples with multiple relocations; two thirds of relocated samples move only once, and multiple movers likely have distinct motivations (occupational requirements or precarious housing) that complicate comparison with single movers.

In terms of measurement, GHQ-12 captures short-term psychological distress through self-report, and GHQsynth mitigates measurement noise through data-driven item selection but does not eliminate the non-linearity and subjectivity of the scale. The OAC supergroups are derived by clustering of K-means on demographic, socioeconomic and housing census variables, so a change in OAC after relocation may reflect a shift in sociodemographic context as much as a change in the physical built environment. To address this, we include IMD and AHAH as covariates in the ML model to help distinguish physical from social environmental effects. Spillover effects of relocation on dependent household members and neighbourhood-level outcomes such as gentrification remain outside the scope of this model and require further study.

The vertical ensemble method considers characteristics of each constituent model. The DiD model co-

efficients are possibly underestimated given the likelihood of incomplete adjustment for anticipation effects (Malani and Reif, 2015; Alpert, 2016). The SCM model has limited violation of the convex hull condition leading to potential data extrapolation (Goh and Yu, 2022); based on the positive correlation between the two latent factors, this likely led to an overestimation of the coefficients. Comparing the results from the two models, SCM consistently produces larger coefficients than DiD. The ensemble design also addresses the voluntary-involuntary mover distinction: DiD assumes no anticipatory effects and is better suited to involuntary movers; SCM handles anticipatory effects and is better suited to voluntary movers. Given the limitations discussed, we contextualise the DiD estimates as lower bounds and the SCM estimates as upper bounds of the ensemble model for the relocation effect.

We highlight challenges faced by ML applications in panel data settings. Currently there are few ML methods suitable for addressing temporal dependencies and non-stationarity. We chose a causal forest model for data-driven covariates selection and for identifying heterogeneous treatment effects in distinct subpopulations. The panel data was transformed into a pseudo-cross-sectional structure where each individual has one observation representing the difference from pre-relocation to post-relocation; this sacrifices the temporal dimension visible in the SCM trajectories of Figure 6, where SWB effects fading or compounding over post-relocation years can be directly observed.

The ML model diagnostics in Appendix G confirm the overlap assumption is satisfied: propensity scores for treatment and control groups show no mass near zero or one despite the high-dimensional covariate space, and the distribution of bias scaled by outcome standard deviation is centred around zero, consistent with an unbiased ATE estimate.

For future research, Dynamic Double ML is suitable for repeated treatment for multiple treated samples (Lewis and Syrgkanis, 2021). Deep learning approaches such as RNN (Tank et al., 2021) and N-BEATS (Oreshkin et al., 2020) also have the potential to inform policy by generating intuitive and testable counterfactuals; notably the SyNBEATS method (Goldin et al., 2022).

Leveraging an integrated framework of state-of-the-art statistical and ML methods, our study quantitatively explores the evolving relationships between relocation, built environment change, and subjective wellbeing. We combine causal inference in statistics and machine learning to establish and quantify the complex causal relationships. The ability to robustly handle subjective data and spatial granularity, combined with less stringent assumptions, will enable more quantitative research to inform a wide range of empirical studies.

Acknowledgments

We acknowledge the Isaac Newton Trust (grant number: ACDF9) for the Academic Career Development Fellowship support which enabled this research.

References

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505, 2010. doi: 10.1198/jasa.2009.ap08746.

Abby Alpert. The anticipatory effects of Medicare Part D on drug utilization. *Journal of Health Economics*, 49:28–45, 2016. doi: 10.1016/j.jhealeco.2016.06.005.

Susan Athey and Guido Imbens. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016. doi: 10.1073/pnas.1510489113.

Susan Athey and Stefan Wager. Estimating treatment effects with causal forests: An application. *Observational Studies*, 5(2):37–51, 2019. doi: 10.1353/obs.2019.0001.

Susan Athey and Stefan Wager. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021. doi: 10.3982/ECTA15732.

Susan Athey, Mohsen Bayati, Guido Imbens, and Zhaonan Qu. Ensemble methods for causal effects in panel data settings. *AEA Papers and Proceedings*, 109:65–70, 2019a. doi: 10.1257/pandp.20191069.

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178, 2019b. doi: 10.1214/18-AOS1709.

Michael Batty. The coronavirus crisis: What will the post-pandemic city look like? *Environment and Planning B: Urban Analytics and City Science*, 47(4):547–552, 2020. doi: 10.1177/2399808320926912.

Michael Batty. The post-pandemic city: speculation through simulation. *Cities*, 124:103594, 2022. doi: 10.1016/j.cities.2022.103594.

Michael Batty, Julian Clifton, Peter Tyler, and Li Wan. The post-Covid city. *Cambridge Journal of Regions, Economy and Society*, 15:447–457, 2022. doi: 10.1093/cjres/rsac041.

Ruth M. J. Byrne. Counterfactual thought. *Annual Review of Psychology*, 67(1):135–157, 2016. doi: 10.1146/annurev-psych-122414-033249.

Brantly Callaway and Pedro H. C. Sant'Anna. Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021. doi: 10.1016/j.jeconom.2020.12.001.

Jerry Chen and Li Wan. Remote working and experiential wellbeing: A latent lifestyle perspective using UK time use survey before and during COVID-19. *PLOS ONE*, 19(7):e0305096, 2024. doi: 10.1371/journal.pone.0305096.

William A. V. Clark. Life events and moves under duress: disruption in the life course and mobility outcomes. *Longitudinal and Life Course Studies*, 7:218–239, 2016. doi: 10.14301/llcs.v7i3.376.

Rachel Cooper, Christopher Boyko, and Ricardo Codinhoto. The effect of the physical environment on mental wellbeing. In *Mental Capital and Wellbeing*, pages 967–1006. Wiley Blackwell, Hoboken, NJ, 2010.

Konstantinos Daras, Mark A. Green, Alec Davies, Alex Singleton, and Benjamin Barr. Access to healthy assets and hazards (AHAH) – updated version 2017. figshare, 2019.

Ed Diener, Sarah D. Pressman, Jeff Hunter, and Destiny Delgadillo-Chase. If, why, and when subjective well-being influences health, and future needed research. *Applied Psychology: Health and Well-Being*, 9(2):133–167, 2017. doi: 10.1111/aphw.12090.

Paul Dolan, Richard Layard, and Robert Metcalfe. Measuring subjective wellbeing for public policy. Technical Report Special Paper No. 23, Centre for Economic Performance, London School of Economics, 2011.

Ciara Foye, David Clapham, and Tommaso Gabrieli. Home-ownership as a social norm and positional good: Subjective wellbeing evidence from panel data. *Urban Studies*, 55:1290–1312, 2018. doi: 10.1177/0042098017695478.

Gyuhyeong Goh and Jisang Yu. Synthetic control method with convex hull restrictions: A Bayesian maximum a posteriori approach. *The Econometrics Journal*, 25(1):215–232, 2022. doi: 10.1093/ectj/utab021.

Mikhail Goldin, Alberto Abadie, and Bhramar Mukherjee. SyNBEATS: Synthetic control using neural basis expansion, 2022. Placeholder – verify full publication details with authors.

Petrus Hagedoorn and Marco Helbich. Longitudinal effects of physical and social neighbourhood change on suicide mortality: A full population cohort study among movers and non-movers in the Netherlands. *Social Science & Medicine*, 294:114690, 2022. doi: 10.1016/j.socscimed.2021.114690.

Peter H. Haslag and Daniel Weagley. From L.A. to Boise: How migration has changed during the COVID-19 pandemic, 2022.

Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, Boca Raton, 2020.

Gundula M. Huebner, Tadj Oreszczyn, Kenan Direk, and Ian Hamilton. The relationship between the built environment and subjective wellbeing — Analysis of cross-sectional data from the English Housing Survey. *Journal of Environmental Psychology*, 80:101763, 2022. doi: 10.1016/j.jenvp.2022.101763.

Sigurd W. Hystad and Bjørn Helge Johnsen. The dimensionality of the 12-item General Health Questionnaire (GHQ-12): Comparisons of factor structures and invariance across samples and time. *Frontiers in Psychology*, 11:1300, 2020. doi: 10.3389/fpsyg.2020.01300.

Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press, Cambridge, 2015. doi: 10.1017/CBO9781139025751.

Sungsoo Kang. Why low-income households become unstably housed: Evidence from the panel study of income dynamics. *Housing Policy Debate*, 29:559–587, 2019. doi: 10.1080/10511482.2018.1544161.

Anne C. Krefis, Mareike Augustin, K. Heinke Schlünzen, Jürgen Oßenbrügge, and Jobst Augustin. How does the urban environment affect health and well-being? A systematic review. *Urban Science*, 2(1):21, 2018. doi: 10.3390/urbansci2010021.

Alan B. Krueger and Arthur A. Stone. Measuring subjective wellbeing: Progress and challenges. *Science*, 346(6205):42–43, 2014. doi: 10.1126/science.1256392.

Greg Lewis and Vasilis Syrgkanis. Double/Debiased machine learning for dynamic treatment effects via g-Estimation. arXiv:2002.07285, 2021.

Anup Malani and Julian Reif. Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform. *Journal of Public Economics*, 124:1–17, 2015. doi: 10.1016/j.jpubeco.2015.01.001.

Beata Nowok, Allan Findlay, and David McCollum. Linking residential relocation desires and behaviour with life domain satisfaction. *Urban Studies*, 55:870–890, 2018. doi: 10.1177/0042098016665972.

OECD. *OECD Guidelines on Measuring Subjective Well-being*. OECD Publishing, Paris, 2013. doi: 10.1787/9789264191655-en.

Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. arXiv:1905.10437, 2020.

Fernando Rios-Avila, Brantly Callaway, and Pedro H. C. Sant'Anna. csdid: Difference-in-Differences with multiple time periods in Stata. Working paper, 2021.

Pedro H. C. Sant'Anna and Jun Zhao. Doubly robust difference-in-differences estimators. *Journal of Econometrics*, 219(1):101–122, 2020. doi: 10.1016/j.jeconom.2020.06.003.

Erik Sverdrup, Ayush Kanodia, Zhengyuan Zhou, Susan Athey, and Stefan Wager. policytree: Policy learning via doubly robust empirical welfare maximization over trees. *Journal of Open Source Software*, 5(50):2232, 2020. doi: 10.21105/joss.02232.

Alex Tank, Ian Covert, Nicholas Foti, Ali Shojaie, and Emily B. Fox. Neural Granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4267–4279, 2021. doi: 10.1109/TPAMI.2021.3065601.

Yuhao Wang, Chen Zhong, Qili Gao, and Carmen Cabrera-Arnau. Understanding internal migration in the UK before and during the COVID-19 pandemic using twitter data. *Urban Informatics*, 1:15, 2022. doi: 10.1007/s44212-022-00018-w.

Ursula Werneke, David P. Goldberg, Ilker Yalcin, and Tevfik B. Ustün. The stability of the factor structure of the General Health Questionnaire. *Psychological Medicine*, 30(4):823–829, 2000. doi: 10.1017/S0033291799002287.

Julian Wolpert. Behavioral aspects of the decision to migrate. *Papers of the Regional Science Association*, 15(1):159–169, 1965. doi: 10.1007/BF01947871.

Steven A. Wu, Shruti Sridhar, and Tobias Gerstenberg. That was close! A counterfactual simulation model of causal judgments about decisions. PsyArXiv, 2022.

Yiqing Xu. Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76, 2017. doi: 10.1017/pan.2016.2.

Yiqing Xu and Licheng Liu. gsynth: Generalized synthetic control method. R package version 1.2.1, 2022. URL https://CRAN.R-project.org/package=gsynth.

Steve Yadlowsky, Scott Fleming, Nigam Shah, Emma Brunskill, and Stefan Wager. Evaluating treatment prioritization rules via rank-weighted average treatment effects. arXiv:2111.07966, 2021.

Meng Yang, Julian Hagenauer, Martin Dijst, and Marco Helbich. Assessing the perceived changes in neighborhood physical and social environments and how they are associated with Chinese internal migrants' mental health. *BMC Public Health*, 21:1240, 2021. doi: 10.1186/s12889-021-11289-4.

Appendix A. ONS Area Classification (OAC)

The ONS Residential-based Area Classification (OAC) applies K-means clustering to 2011 UK Census variables to produce 8 supergroups covering all LSOAs in England:

1. **Supergroup 1 – Rural Residents**: Sparsely populated areas with older, owner-occupied housing, high car ownership, many employed in agriculture and skilled trades.

2. **Supergroup 2 – Cosmopolitans**: Dense urban areas with high population turnover, students and young professionals, private rented sector dominance, and ethnic diversity.

3. **Supergroup 3 – Ethnicity Central**: Areas of high ethnic minority concentration, typically in inner urban zones, with mixed tenure, high unemployment, and below-average qualifications.

4. **Supergroup 4 – Multicultural Metropolitans**: Diverse, densely populated urban areas with varied housing tenure, high population density, broad range of occupations.

5. **Supergroup 5 – Urbanites**: Mixed urban areas characterised by a balance of social and owner-occupied housing, intermediate occupations, and moderate deprivation levels.

6. **Supergroup 6 – Suburbanites**: Predominantly owner-occupied suburban areas with family households, high car ownership, and residents in managerial and professional occupations.

7. **Supergroup 7 – Constrained City Dwellers**: Urban areas with high levels of social housing, older residents, lower incomes, and limited employment opportunities.

8. **Supergroup 8 – Hard-Pressed Living**: Areas characterised by high deprivation, social rented housing, lower qualifications, routine occupations, and poor health outcomes.

Appendix B. Latent Class Analysis

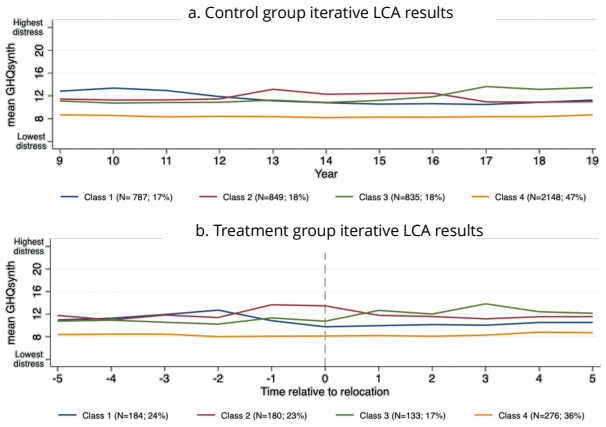


Figure A1: Iterative latent class analysis to visualise subjective wellbeing patterns.

Appendix C. Staggered DiD: Relocation Effects

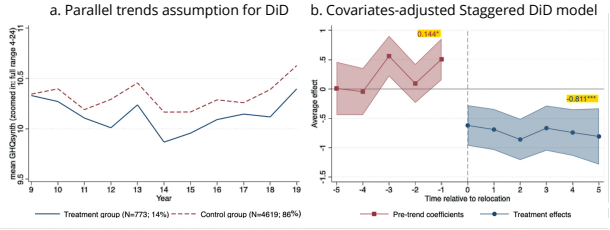


Figure A2: Staggered difference-in-differences model for relocation effects on subjective wellbeing.

Appendix D. Staggered DiD: Built Environment Change

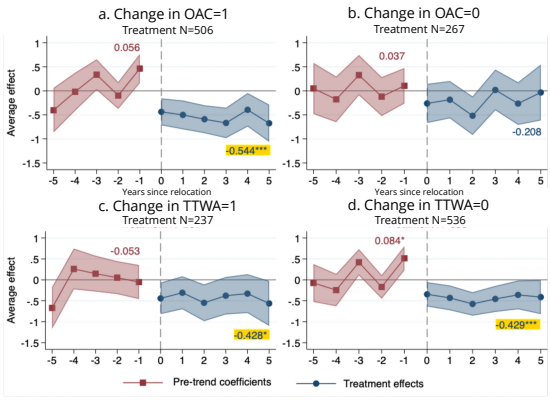


Figure A3: Staggered DiD for relocations with built environment change vs. no built environment change.

Appendix E. Staggered DiD: 2-way Interaction

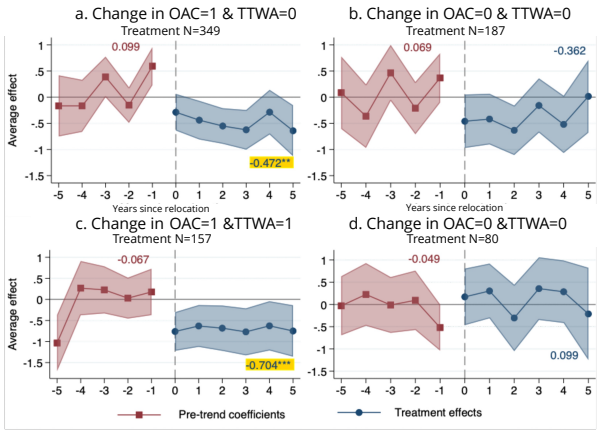


Figure A4: Staggered DiD 2-way interaction of change in built environment and change in employment area.

Appendix F. Bacon Decomposition

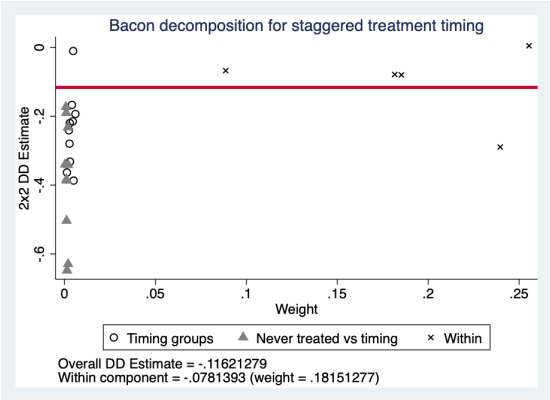
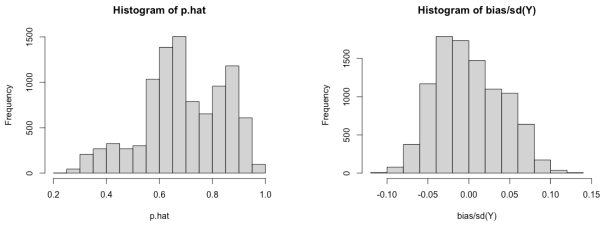


Figure A5: Bacon decomposition of staggered DiD (grouped by treatment timing).

Appendix G. ML Model Diagnostics



(*a*) Propensity score distribution

(*b*) Outcome bias and standard deviation

Figure A6: ML model diagnostics. (a) Overlap by histogram of propensity score. (b) Fit by outcome bias and standard deviation.