TI 2021-084/VIII
Tinbergen Institute Discussion Paper

Cities with Forking Paths? Agglomeration Economies in New Zealand 1976-2018

Stuart Donovan[1]
Thomas de Graaff[1]
Arthur Grimes[2]
Henri L.F. de Groot[1]
David C. Maré[2]

[1] Vrije Universiteit Amsterdam
[2] Motu Economic and Public Policy Research

Cities with forking paths? Agglomeration economies in New Zealand 1976-2018*

Stuart Donovan[1,2,3], Thomas de Graaff[1,3], Arthur Grimes[2], Henri L.F. de Groot[1,3], and David C. Maré[2]

[1]Department of Spatial Economics, Vrije Universiteit Amsterdam, The Netherlands
[2]Motu Economic and Public Policy Research, Wellington, New Zealand
[3]Tinbergen Institute, Amsterdam, The Netherlands

20th September 2021

Abstract

We consider whether external urban economic advantages (agglomeration economies) vary with time and space using a simple economic model and detailed micro-data on 134 locations in New Zealand for the period 1976–2018. We find subtle temporal variation, with estimates peaking in 1991 and then falling over the next 15-years by approximately 1%. Since 2006, however, estimates have remained broadly stable. Our results reveal more significant spatial variation: Large cities offer net benefits in production, but not in consumption, whereas small locations close to large cities ("satellites") experience agglomeration economies that are stronger than average.

Keywords: agglomeration economies, cities, productivity, consumption, New Zealand
JEL-classification: R11, R23, R30.

* The authors acknowledge financial support from "Building Better Homes Towns and Cities", National Science Challenge; Kate Preston and Shaan Badenhorst for help with data; and comments from Peter Nunns. Stuart acknowledges support from Veitch Lister Consulting. Access to the data used in this study was provided by Statistics New Zealand under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. The results presented in this study are the work of the authors, not Statistics New Zealand nor their data suppliers. Corresponding author: s.b.donovan@vu.nl.

1 Introduction

In "The Garden of Forking Paths," Borges uses a parable to illustrate the multiplicity of possible outcomes that can arise from the passage of time, observing "...time forks perpetually toward innumerable futures" (Borges et al., 1962). Analogies of this theme have been invoked to highlight barriers to the replication of scientific research and selection biases affecting publication (Gelman and Loken, 2013; Kasy, 2021). In economics, where empirical research often seeks to identify the causal effects of parameters that are subsequently used to appraise hypothetical scenarios and inform policy settings, we suggest the "forking paths" zeitgeist raises related questions about validity. That is, to what extent are economic parameters valid outside of the specific context—specifically, time and place—in which they are estimated? Researchers have begun to grapple with these gnarly epistemological questions (see, e.g., Meier and Sprenger, 2015; Rosenzweig and Udry, 2020). In this spirit, we consider the temporal stability and spatial transferability of estimates of external urban economic advantages, or "agglomeration economies". The complex mechanisms through which agglomeration can influence the choices of firms and households hint at a large number of possible "forking paths".

Why might policy-makers and researchers be interested in the stability and transferability of agglomeration economies? We posit three reasons. First, estimates of agglomeration economies are often used to appraise the effects of policies, such as transport projects, in hypothetical scenarios that extend decades into the future. In doing so, appraisal guidelines usually assume agglomeration economies are both stable and transferable—at least after controlling for differences in industrial composition (see, e.g., Table 37 in Waka Kotahi, 2020). Second, trends in agglomeration economies are central to debates on the future of cities. Whereas Cairncross (1997) and Friedman (2006) argued technology will reduce the benefits of proximity, Glaeser (2011) opined "despite the technological breakthroughs that have caused the death of distance, it turns out that the world isn't flat; it's paved." In the wake of the COVID-19 pandemic, debates about the economic advantages of cities—and whether these advantages are likely to persist in the future—have surfaced anew (see, e.g., Shenker, 2020). Third, agglomeration economies are a rubric, or "black box", used to describe several distinct microeconomic channels, such as knowledge spillovers and labour market efficiencies (Puga, 2010). Understanding how agglomeration economies vary over time and space may help illuminate aspects of these underlying channels. For these reasons, we suggest the stability and transferability of agglomeration economies is relevant to both policy and research.

Notwithstanding this relevance, few empirical studies consider whether agglomeration economies vary with time and space, at least in a systematic fashion. In terms of temporal stability, Martínez-Galarraga et al. (2008) analyse agglomeration economies in production for Spain and find estimates decline in the period 1860–1999. Similarly, the recent meta-analysis by Donovan et al. (2021) draws on estimates for the period 1960–2020 and finds evidence that, since 2000, agglomeration economies in production have fallen by 1–2%. More attention has been paid to spatial variation. Several meta-analyses—see, for example, Ahlfeldt and Pietrostefani (2019) and Donovan et al. (2021)—and primary studies—see, for example, Ahrend et al. (2017) and Chauvin et al. (2017)—allow for and report differences between countries, although the factors that cause these differences are typically left unexplored. Fewer studies, moreover, consider the potential for spatial variation *within* countries, which is likely to be more relevant to domestic policy settings. Notably for our purposes, Maré and Graham (2013) find some evidence that agglomeration economies in production vary between nine regions in New Zealand—even after controlling for the characteristics of individual firms.

Against this backdrop, we consider whether agglomeration economies vary with time and space using data on 134 locations in New Zealand for the period 1976–2018. Following Roback (1982), Gabriel and Rosenthal (2004), Y. Chen and Rosenthal (2008), and Maré and Poot (2019), we estimate a simple economic model that measures the relative advantages of locations to firms and workers in terms of the price of labour and floorspace. Although grounded in the existing literature, our study combines several useful attributes. Notably, we draw on detailed data for almost all individual full-time workers in New Zealand at approximately five-year intervals over more than forty years. Whereas Martínez-Galarraga et al. (2008) present estimates for two recent periods, namely 1965–1979 and 1985–1999, our estimates cover nine intervals in the period 1976–2018. And, in contrast to Maré and Graham (2013), who estimate agglomeration economies for nine regions in New Zealand, we estimate agglomeration economies for 134 locations. Unlike Gabriel and Rosenthal (2004) and Martínez-Galarraga et al. (2008), but like Maré and Graham (2013) and Maré and Poot (2019), we use micro-data to control for the observed characteristics of individual workers and dwellings. Together, these attributes enable us to derive estimates of agglomeration economies in production and consumption that are relatively robust and broadly comparable over time and space.

Three main insights emerge from our analysis. First, we confirm that micro-data yield smaller estimates of agglomeration economies in production, with wage elasticities that are approximately half those from aggregate data—a finding that aligns with the seminal work by Combes, Duranton and Gobillon (2008) and the recent meta-analysis by Ahlfeldt and Pietrostefani (2019). More uniquely, we find the gap between estimates derived from micro-data vis-á-vis aggregate data has not changed with time, implying that the relevance of sorting by workers is stable—at least in New Zealand. Second, we find evidence of subtle temporal variation in estimates of agglomeration economies, which arises due to the effect of agglomeration on rent. Specifically, from a nadir in 1981, estimates peak in 1991, and then decline by approximately 1%—loosely corroborating the trend in Donovan et al. (2021).[1] Since 2006, however, our estimates of agglomeration economies have remained broadly constant; contrary to claims, the world does not appear to be getting "flatter". Finally, we find evidence that agglomeration economies vary systematically between locations: Large cities in New Zealand offer net benefits in production, but not in consumption, whereas "satellite" locations that are close to large cities experience agglomeration economies that are stronger than average. Evidence of large spatial differences within a relatively small country like New Zealand—after controlling for the observed characteristics of workers and dwellings—is a notable finding.

Our findings highlight several avenues for further research. First, we see value in primary research, like Martínez-Galarraga et al. (2008) and the present study, which uses consistent data and methods to trace the evolution of agglomeration economies within countries and over time. Second, further research would ideally seek to explain temporal and spatial variation in estimates of agglomeration economies. Although we detect the fingerprints of such variation, and present informal explanations for its origins, we do not pin-point the underlying causes. Third, we suggest there is merit in developing more sophisticated measures of agglomeration. Such measures could, for example, consider multi-modal transport costs, rather than just time or distance, and measure proximity to specialised infrastructure and institutions—like major airports and universities—that may augment the microeconomic effects of agglomeration (Brueckner, 2003; X. Chen et al., 2021; Kantor and Whalley, 2014; Cermeño, 2019). Finally, we echo others by calling for more research into agglomeration economies in consumption.

The remainder of this paper reads as follows: Section 2 develops the methodology; Section 3 summarises the data; Section 4 presents the benchmark models and sensitivity tests; Section 5 discusses our findings; and Section 6 concludes.

[1] In addition to the general factors discussed in Donovan et al. (2021), such as information and communications technologies ("ICT"), we speculate that temporal variation in New Zealand in the period 1976–2001 may reflect the lingering effects of major economic policy reforms that were implemented in the 1980s.

2 Methodology

2.1 Model

Our model follows Maré and Poot (2019), which in turn adapts Gabriel and Rosenthal (2004). To start, we assume that mobile workers choose their locations given preferences for housing, H, a composite consumption good, Y, and local amenities, $f_u(A)$. We assume a representative worker's preferences, U_{it}, for city i in period t is given by:

$$U_{it} = f_u(A_{it}) H_{it}^{\alpha} Y_{it}^{1-\alpha}. \tag{1}$$

We assume all workers supply one unit of labour and earn a wage, w_{it}, that is, our model focuses on the choices of full-time workers. The price of housing, r_{it}, is locally determined, but we assume that the price of the composite good is the same everywhere—so it can be indexed in period t without loss of generality. Given the budget constraint $w_{it} = r_{it} H_{it} + Y_{it}$, we derive the following conventional Marshallian demand functions:

$$H_{it}^{*} = \alpha \frac{w_{it}}{r_{it}} \qquad \text{and} \qquad Y_{it}^{*} = (1-\alpha) w_{it}. \tag{2}$$

Substituting the Marshallian demand functions for housing, H_{it}^{*}, and the composite good, Y_{it}^{*}, from Equation (2) into Equation (1) and re-arranging yields:

$$v_{it} = \kappa_u f_u(A_{it}) \frac{w_{it}}{r_{it}^{\alpha}} = \overline{v}_t. \tag{3}$$

Where v_{it} denotes indirect utility and $\kappa_u = \alpha^{\alpha}(1-\alpha)^{1-\alpha}$ is a constant. Following Roback (1982), we impose a spatial equilibrium condition that requires that workers in all locations achieve the same fixed reservation utility, $\overline{v}_t$, in each time period.

In each location, we assume there is a representative firm that uses a commonly-available, constant returns to scale technology to produce the composite good, Y_{it}, using floorspace, H_{it}, and labour, L_{it}, as inputs. Formally, we have:

$$Y_{it} = f_y(A_{it}) H_{it}^{\gamma} L_{it}^{1-\gamma}, \tag{4}$$

where $f_y(A_{it})$ describes the effect of local amenities, A_{it}, on productivity. We assume firms pay w_{it} and r_{it} for labour and floorspace.[2] Assuming firms maximise profits, we can use the resulting first-order conditions to derive the following equilibrium condition:

$$r_{it}^{\gamma} w_{it}^{1-\gamma} = \kappa_y f_y(A_{it}), \tag{5}$$

where $\kappa_y = \gamma^{\gamma}(1-\gamma)^{1-\gamma}$. Taking logs and re-arranging the equilibrium conditions in Equations (3) and (5) yields iso-utility and iso-cost conditions for workers and firms:

$$\ln f_u(A_{it}) = \alpha \ln r_{it} - \ln w_{it} + \ln \overline{v}_t - \ln \kappa_u, \tag{6}$$

$$\ln f_y(A_{it}) = \gamma \ln r_{it} + (1-\gamma) \ln w_{it} - \ln \kappa_y. \tag{7}$$

Equations (6) and (7) measure the implicit value of local amenities to workers and firms in equilibrium as a function of rents, wages, and structural parameters. To understand the effect of agglomeration, $\ln D_{it}$, we differentiate Equations (6) and (7) as follows:

$$\frac{\partial \ln f_u(A_{it})}{\partial \ln D_{it}} = \alpha \frac{\partial \ln r_{it}}{\partial \ln D_{it}} - \frac{\partial \ln w_{it}}{\partial \ln D_{it}} = \alpha \epsilon_{it}^{r} - \epsilon_{it}^{w} = E_c, \tag{8}$$

$$\frac{\partial \ln f_y(A_{it})}{\partial \ln D_{it}} = \gamma \frac{\partial \ln r_{it}}{\partial \ln D_{it}} + (1-\gamma) \frac{\partial \ln w_{it}}{\partial \ln D_{it}} = \gamma \epsilon_{it}^{r} + (1-\gamma)\epsilon_{it}^{w} = E_p. \tag{9}$$

Where the constant terms drop out and we let $\frac{\partial \ln r_{it}}{\partial \ln D_{it}} = \epsilon_{it}^{r}$ and $\frac{\partial \ln w_{it}}{\partial \ln D_{it}} = \epsilon_{it}^{w}$, that is, elasticities of rents and wages with respect to agglomeration, which—in their most general form—vary with location, i, and time, t. Together, Equations (8) and (9) conveniently express the change in the implicit value of local amenities that follows from a change in agglomeration in terms of the latter's effects on rents and wages, as measured by their elasticities.[3] One direct implication of Equation (8) is that we can expect positive agglomeration economies in consumption, E_c, if and only if $\alpha \epsilon_{it}^{r} > \epsilon_{it}^{w}$. Bringing this model to data then involves a relatively simple two-step process: First, we estimate the elasticities of rents and wages with respect to agglomeration, ϵ_{it}^{r} and ϵ_{it}^{w}, and, second, we combine these elasticities as per Equations (8) and (9), given values for α and γ.

2 Using r_{it} as the price of floorspace for firms may cause bias if the composition of inputs into floorspace differs between workers and firms and the prices of inputs varies over time. We expect this bias is small.

3 Specifically, Equation (8) defines the change in implicit value for workers who live and work in location, i.

2.2 Estimation

To identify ϵ_{it}^{w} and ϵ_{it}^{r}, we estimate models consisting of two equations:

$$\begin{aligned}\ln w_{it} &= f^{w}(\ln D_{it}) + \tau_t^{w} + \zeta_i^{w}\\ \ln r_{it} &= f^{r}(\ln D_{it}) + \tau_t^{r} + \zeta_i^{r},\end{aligned}$$

where $f(\ln D_{it})$ denotes a function of agglomeration, $\ln D_{it}$, which can vary between equations and models; τ_t denotes individual time effects; and ζ_i denotes individual location effects. For $\ln w_{it}$ and $\ln r_{it}$, we begin by using average income and rent. In our preferred models, however, we replace these averages with estimates of the spatial income and rent premia, which are derived from a process similar to that used in Combes, Duranton and Gobillon (2008).[4] Specifically, we first regress the reported incomes and rents of individual workers and dwellings for each Census year versus a set of location fixed effects and controls for observed individual characteristics.[5] We estimate separate regressions for each year, such that the effects of observed characteristics on incomes and rents can vary with time. The location fixed effects—or, "spatial income and rent premia"—then become the dependent variables of the equations in our models.

We use Bayesian methods to estimate all models, which helps address four inter-related empirical problems.[6] First, we follow the literature on partial pooling models and treat individual time and location effects, τ_t and ζ_i, as group effects (or, "random effects") that are drawn from common population distributions with their own variances, or hyper-parameters (Gelman, Carlin et al., 2013). Compared to population effects (or, "fixed effects"), group effects pool information both within and between time periods and locations, mitigating the risk of over-fitting. With at most nine observations per location, we are especially concerned about over-fitting the location effects, ζ_i. And, unlike restricted maximum likelihood estimators, Bayesian methods directly estimate

[4] Unlike Combes, Duranton and Gobillon (2008), our data do not follow workers or dwellings over time, preventing us from including individual effects in the first-stage regressions. Donovan et al. (2021) find that including such effects reduces estimates of agglomeration economies in production by 1.1% whereas Ahlfeldt and Pietrostefani (2019) conclude controlling for "sorting" reduces estimates by 2–4%.

[5] For workers, we control for gender, age (polynomial by gender), qualification ($n = 10$, except in 1976 when $n = 4$), two-digit industry sector ($n = 54$), ethnicity ($n = 15$), religion ($n = 11$), and birthplace ($n = 12$). For dwellings, we control for the number of bedrooms ($n = 10$) and rooms ($n = 10$) as well as dwelling type ($n = 8$) and heating ($n = 8$). Results of the first-stage regressions are available on request.

[6] Specifically, all models are estimated using the statistical package `R` running in the RStudio environment with the `brms` package and default priors (R Core Team, 2021; RStudio Team, 2021; Bürkner, 2017).

the hyper-parameters that are associated with group effects. Second, our dependent variables—that is, the spatial income and rent premia—are measured with uncertainty, or error. Bayesian methods, specifically errors-in-outcomes models, enable us to directly account for uncertainty in the spatial premia—as measured by their estimated standard errors—in a theoretically consistent manner.[7] Third, inspection of our data reveals the presence of some unusual observations that are associated with small towns (c.f. Figures 1 and 2). To mitigate the influence of these observations, we allow our two response variables to follow Student's t-distributions, which—compared to Gaussian distributions—allow more mass in the tails of the probability distributions (Geweke, 1993; Gelman and Hill, 2007). Finally, Equations (8) and (9) define agglomeration economies as composite parameters that are formed from estimates of both ϵ_{it}^{w} and ϵ_{it}^{r}. By jointly estimating these equations using Bayesian methods, we can generate distributions of parameter estimates for ϵ_{it}^{w} and ϵ_{it}^{r} that directly account for correlations between the two equations.[8]

We estimate four benchmark models. Model A uses average incomes and rents and imposes common elasticities, ϵ, across all time periods and locations:

$$\begin{aligned} \ln w_{it} &\sim \mathcal{N}(\epsilon^{w} \ln D_{it} + \tau_t^{w} + \zeta_i^{w}, \sigma^{w}) \\ \ln r_{it} &\sim \mathcal{N}(\epsilon^{r} \ln D_{it} + \tau_t^{r} + \zeta_i^{r}, \sigma^{r}), \end{aligned} \tag{Model A}$$

where we follow the conventional assumption that group effects, τ and ζ, are drawn from independent normal distributions.

Model B also imposes common elasticities across all time periods and locations but uses the spatial income and rent premia within a multi-level, errors-in-outcomes setting:

$$\begin{aligned} \ln w_{it} &\sim \mathcal{N}(\ln w_{it}^{*}, s_{it}^{w}) \\ \ln r_{it} &\sim \mathcal{N}(\ln r_{it}^{*}, s_{it}^{r}) \\ \ln w_{it}^{*} &\sim \mathcal{N}(\epsilon^{w} \ln D_{it} + \tau_t^{w} + \zeta_i^{w}, \sigma^{w}) \\ \ln r_{it}^{*} &\sim \mathcal{N}(\epsilon^{r} \ln D_{it} + \tau_t^{r} + \zeta_i^{r}, \sigma^{r}), \end{aligned} \tag{Model B}$$

7 Combes, Duranton and Gobillon (2008) address errors-in-outcomes using a feasible generalized least-squares ("FGLS") estimator, which yields similar results to OLS estimators. The authors attribute this result to the large number of observations at their disposal. Subsequent researchers have tended to follow this lead, pointing to precise estimates of the spatial premia in the first-stage as reason to treat the latter deterministically (see, e.g., footnote 13 in Groot et al., 2014). In contrast, the large number of relatively small locations in our data leads to somewhat imprecisely estimated spatial premia.

8 Joint estimation, for example, allows us to model correlations between the residuals and individual effects, τ_t and ζ_i. In Section 4.2.3, we model these correlations as a sensitivity test.

where we assume the estimated spatial premia, $\ln w_{it}$ and $\ln r_{it}$, are drawn from normal distributions with true means, $\ln w^*_{it}$ and $\ln r^*_{it}$, and standard deviations, s^w_{it} and s^r_{it}, respectively, as defined by the standard errors of the premia in the first-stage regressions.

Model C then allows both of the true means, $\ln w^*_{it}$ and $\ln r^*_{it}$, to follow Student's t-distributions with additional degrees-of-freedom parameters, ν:

$$\begin{aligned}
\ln w_{it} &\sim \mathcal{N}(\ln w^*_{it}, s^w_{it}) \\
\ln r_{it} &\sim \mathcal{N}(\ln r^*_{it}, s^r_{it}) \\
\ln w^*_{it} &\sim t(\epsilon^w \ln D_{it} + \tau^w_t + \zeta^w_i, \nu^w, \sigma^w) \\
\ln r^*_{it} &\sim t(\epsilon^r \ln D_{it} + \tau^r_t + \zeta^r_i, \nu^r, \sigma^r).
\end{aligned} \tag{Model C}$$

Finally, Model D allows elasticities to vary with time and location, as per the group effects, δ_t and γ_i, respectively. In its most general form, Model D is defined as

$$\begin{aligned}
\ln w_{it} &\sim \mathcal{N}(\ln w^*_{it}, s^w_{it}) \\
\ln r_{it} &\sim \mathcal{N}(\ln r^*_{it}, s^r_{it}) \\
\ln w^*_{it} &\sim t((\epsilon^w + \delta^w_t + \gamma^w_i) \ln D_{it} + \tau^w_t + \zeta^w_i, \nu^w, \sigma^w) \\
\ln r^*_{it} &\sim t((\epsilon^r + \delta^r_t + \gamma^r_i) \ln D_{it} + \tau^r_t + \zeta^r_i, \nu^r, \sigma^r).
\end{aligned} \tag{Model D}$$

In Section 4, we estimate several variants of Model D that restrict $\delta_t = 0$ and $\gamma_i = 0$.

In short, Model A and Model B differ only in the use of aggregate data vis-á-vis spatial premia, where the latter are preferred and require that we model errors-in-outcomes. Model C then mitigates the effects of influential observations by allowing the dependent variables to follow Student's t-distributions. Despite their differences, Model A, Model B, and Model C all assume that income and rent elasticities are constant with time and space. In contrast, Model D allows the income and rent elasticities to vary with the time and location group effects, δ_t and γ_i. We are primarily interested in understanding whether there is evidence to prefer Model D, where the income and rent elasticities can vary with time and space, over Model C, where they cannot. And, in keeping with our aforementioned interest in the external validity of estimates of agglomeration economies, we focus on the out-of-sample predictive performance of these benchmark models, that is, their ability to predict observations not used in their estimation.

3 Data

Our main source of data is the New Zealand Census (the "Census"). The temporal dimension is defined by the nine waves from 1976 to 2018.[9] The cross-sectional dimension is defined by the 143 zones of the urban area classification developed by Statistics New Zealand, where we consolidate zones in metropolitan areas, namely Auckland (four zones), Wellington (four zones), Hamilton (three zones), and Napier-Hastings (two zones).[10] The resulting panel of 134 locations over nine waves forms the basis of our analyses. For each observation, we extract data on full-time workers aged 25-years or older and rented dwellings. Due to security and confidentiality provisions, we lose 26 observations associated with small towns in early Censuses—leaving us with 1,180 observations.[11] We use consumer price indices to index incomes and rents to 2018 (second quarter) levels. In 2018, the median and mean population of locations is 4,464 and 30,189, respectively.

The use of Census data confers advantages and disadvantages. In terms of advantages, the Census seeks to survey all residents and typically has average response rates in excess of 95%, mitigating problems with sample selection. And, by accessing unit records, we can control for the characteristics of individual workers and dwellings. Usefully, the Census also collects information at the place-of-residence and place-of-work, both of which are relevant to our model. Notwithstanding these advantages, Census data also has some limitations. As the Census does not track individuals over time, we cannot control for unobserved sources of heterogeneity. Census data on income and rents are also grouped into \$100 and \$10 bands, introducing some measurement error.[12] The Census also records gross income rather than labour income, where the latter is more relevant to our model.[13] Finally, the Census only collects prices for rented dwellings, which represent around one-third of all dwellings, on average, in the period we analyse.

[9] Specifically, the Census was undertaken in 1976, 1981, 1986, 1991, 1996, 2001, 2006, 2013, and 2018. Although normally undertaken every five years, the Census planned for 2011 was postponed until 2013.

[10] As urban areas, or locations, are not consistently coded across Censuses, we allocate individuals and dwellings to urban areas as defined in 2018 using the most detailed geographic coding available in each year. Where an urban area from an earlier Census is associated with more than one urban area in 2018, we allocate records to the area that contains the largest share of the 2018 population. Generally, this is a "meshblock", which contain approximately 100 people, on average. For the 1976 Census, meshblock codes were derived from undocumented administrative codes. For individuals who were away from home on the night of the Census, coding was available only at a more aggregate ("area unit") level.

[11] Specifically, we lose 13, 10, and 3 observations from the 1976, 1981, and 1986 censuses, respectively.

[12] We assign responses to the midpoint of bands and top-code responses in the highest band.

[13] Gross income appears to be an accurate measure of labour income for full-time workers aged 25-years or older. For these workers, Maré and Poot (2019) regress gross income from the 2001, 2006, and 2013 Censuses versus labour income from administrative data and find unitary coefficients and $R^2 > 0.97$.

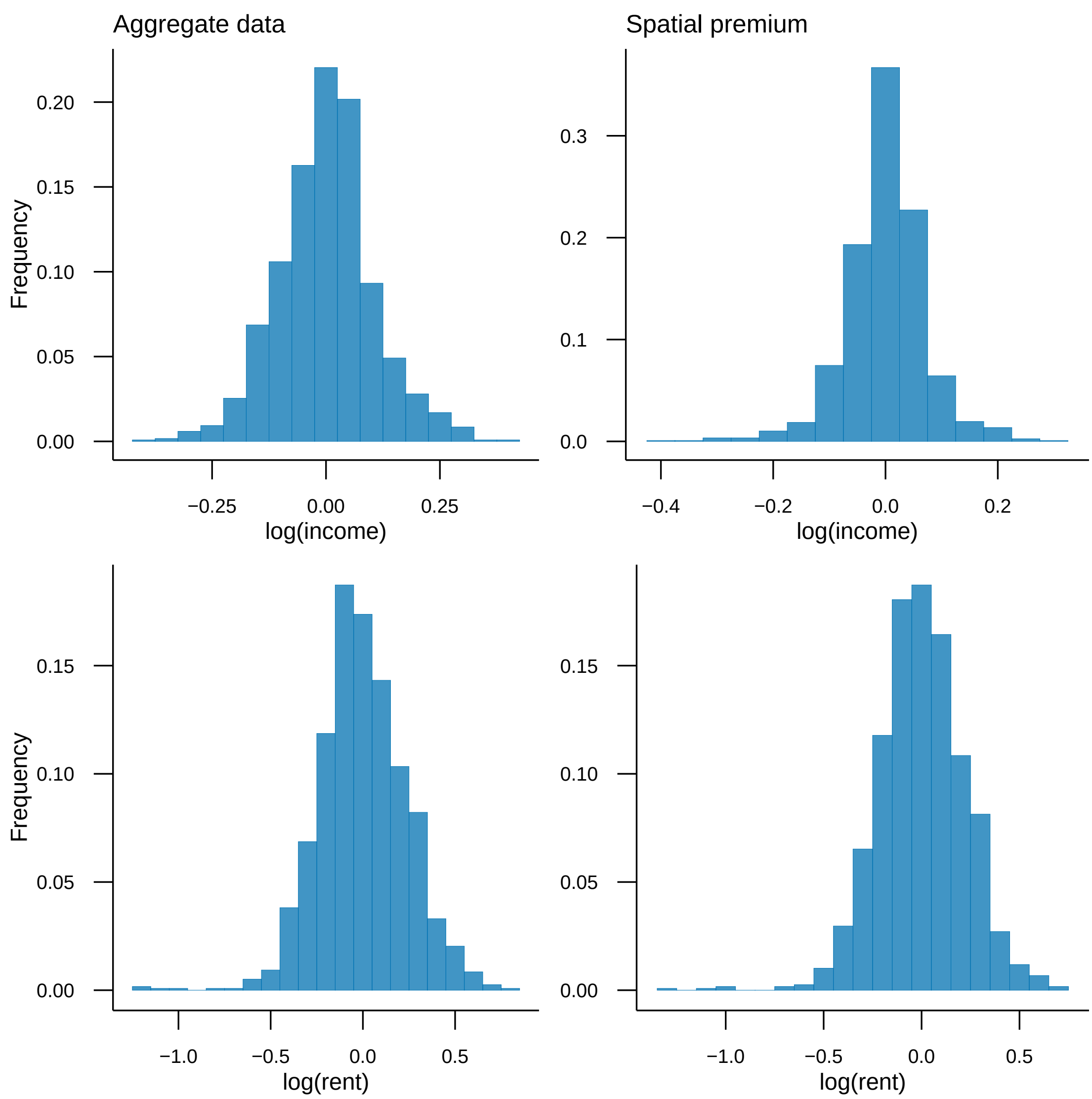

Figure 1: Histograms of dependent variables, mean-centred by year. The left and right panels show aggregate data and spatial premia; the top and bottom panels show incomes and rents, respectively.

Figure 1 presents histograms for our dependent variables, where the top and bottom panels show incomes vis-á-vis rents and the left and right panels show aggregate data vis-á-vis spatial premia, respectively. We find broadly symmetric distributions with some evidence of fat tails, especially on the low-side. Although spatial premia appear to be slightly more centred than aggregate data, the fat tails remain. Figure 2 then plots the dependent variables (vertical axes) versus agglomeration (horizontal axes), $\ln D_{it}$. We follow Holl (2012) and measure agglomeration using "effective population", $D_{it} = P_{it} + \sum_{j \neq i} \frac{P_{jt}}{d_{ij}}$, where the population, P_{it}, of location i is added to the sum of the

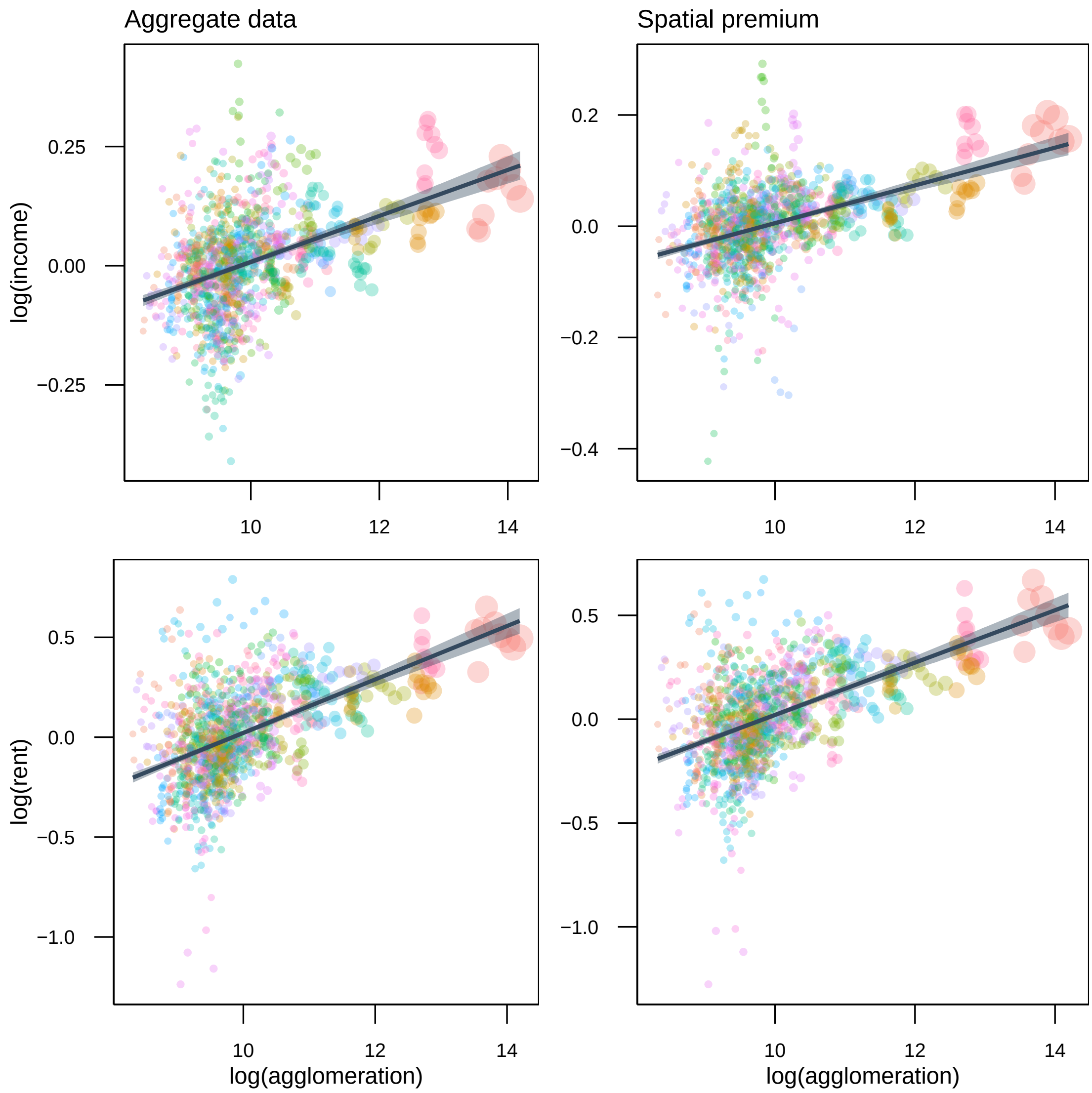

Figure 2: Scatter plots of dependent variables (vertical axes), mean-centred by year, versus agglomeration (horizontal axes), $\ln D_{it}$. The colour and size of points denotes locations and their population, P_{it}. The left and right panels show the aggregate data and spatial premia; the top and bottom panels show incomes and rents, respectively.

population, P_{jt}, of other locations j inversely-weighted by the distance in kilometres between the two, d_{ij}. We measure, d_{ij}, in 2018; to our knowledge, historical data on distances between locations does not exist. Approximately two-thirds of D_{it} stems from P_{it} with the balance from P_{jt}. In Section 5, we consider alternative specifications for agglomeration. Figure 2 reveals a positive association between prices and $\ln D_{it}$, with slightly more spread in aggregate data vis-á-vis spatial premia—suggesting some of the variation in prices is due to differences in the composition of workers and dwellings.

4 Results

4.1 Benchmark Models

Table 1 presents estimated parameters and standard errors (s.e.) for the four benchmark models specified in Section 2.2. The first panel presents the common income and rent elasticities, ϵ^w and ϵ^r, whereas the second panel presents the associated estimates of agglomeration economies, E_p and E_c. The latter are computed from the elasticities, as per Equations 8 and 9, where we follow Maré and Poot (2019) and assume the cost shares of floorspace in production and consumption are $\gamma = 0.10$ and $\alpha = 0.20$, respectively.[14] The third panel summarises the time- and location-specific elasticities that are included in each of the five variants of Model D. And, finally, the bottom panel summarises two model performance measures, R^2 and elpd, where we prefer the latter.[15]

Turning to the results, Model A—which uses aggregate data—returns income elasticities that are approximately twice as large as Model B—which uses micro-data. The elpd values indicate Model B performs worse than Model C, supporting the latter's use of Student's t-distributions to mitigate influential observations.[16] Turning to Model D, we find Model D4—which allows rent elasticities to vary with time and location—has the highest elpd of the models we test, closely followed by Model D5. Comparing Model B, Model C, and the five variants of Model D, we find stable and precise estimates for the income elasticity of 3.6–3.7%. For the rent elasticity, we find similar estimates for Model A, Model B, and Model C.[17] However, we find somewhat larger estimates in Models D2, D4, and D5, which has economically significant implications for our results: Whereas Model C predicts

[14] The assumptions for α and γ do not affect the estimated elasticities but instead shift the levels of agglomeration economies in production and consumption, as per $\frac{\partial E_p}{\partial \gamma} = \epsilon^r_{it} - \epsilon^w_{it}$ and $\frac{\partial E_c}{\partial \alpha} = \epsilon^r_{it}$. For example, if expenditure shares were 20% larger than we assume, such that $\gamma = 0.12$ and $\alpha = 0.24$, then the estimated agglomeration economies in production and consumption for Model D4 can be expected to shift by $0.02(0.301 - 0.036) = 0.005$ and $0.04(0.301) = 0.012$, respectively.

[15] The Bayesian R^2 is the median value from the posterior predictions of a model, averaged across the two equations. The expected log pointwise predictive density ("elpd") measures the out-of-sample performance of a model using efficient leave-one-out cross-validation; see Vehtari et al. (2017) for details. We note the R^2 and elpd is only comparable between Model B, Model C, and Model D, which use the same dependent variables—that is, the spatial income and rent premia that are estimated in the first-stage regressions.

[16] We find $\nu^w \approx 66$ and $\nu^r \approx 6$. The latter indicates considerable mass exists in the tails of the probability distribution for the response variable, providing further empirical support for Model C vis-á-vis Model B.

[17] This raises an interesting question: Why does the use of micro-data (Model B) result in smaller elasticities for incomes but not for rents when compared to aggregate data (Model A)? We posit that micro-data may matter more for incomes than rents because, first, workers are more mobile than dwellings and, second, we have weaker quality adjustments for dwellings. We return to the latter question in Section 5.

	Model A	Model B	Model C	Model D				
				D1	D2	D3	D4	D5
ϵ^w	0.071	0.037	0.036	0.036	0.036	0.036	0.036	0.036
	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
ϵ^r	0.182	0.165	0.182	0.191	0.384	0.181	0.301	0.303
	(0.019)	(0.017)	(0.019)	(0.024)	(0.055)	(0.018)	(0.054)	(0.052)
E_p	0.082	0.049	0.051	0.052	0.071	0.051	0.063	0.063
	(0.008)	(0.005)	(0.005)	(0.006)	(0.008)	(0.006)	(0.007)	(0.008)
E_c	-0.035	-0.004	0.000	0.002	0.041	0.000	0.024	0.024
	(0.010)	(0.007)	(0.007)	(0.008)	(0.012)	(0.007)	(0.012)	(0.012)
δ_t^w	No	No	No	Yes	No	Yes	No	Yes
γ_i^w	No	No	No	No	Yes	Yes	No	Yes
δ_t^r	No	No	No	Yes	No	No	Yes	Yes
γ_i^r	No	No	No	No	Yes	No	Yes	Yes
R^2	0.849	0.992	0.992	0.993	0.994	0.992	0.994	0.994
elpd	2,154	1,914	1,970	2,012	2,188	1,970	2,214	2,213

Table 1: Estimated elasticities for the four benchmark models with the standard errors in parentheses. All models use 1,180 observations and include individual time and location effects, τ_t and ζ_i.

agglomeration economies in consumption are approximately zero, Models D2, D4, and D5 predict they are positive, on average. Notwithstanding uncertainty in the estimates for the rent elasticity, most lie close to the 0.176–0.304 range reported in Combes, Duranton and Gobillon (2019) (c.f. Table 4, p. 1573). On balance, we suggest the results in Table 1 provide strong evidence of positive agglomeration economies in production, although more uncertainty exists for agglomeration economies in consumption. Based on the elpd, Model D4 is our preferred benchmark model to take forward for sensitivity testing. In doing so, we note the temporal and spatial variation in agglomeration economies in Model D4 originates via the rent equation, not the income equation.

4.2 Sensitivity Tests

4.2.1 Basic variants

We estimate two basic variants of Model D4 in which the income elasticity can vary with location ("S1") and time ("S2"), γ_i^w and δ_t^w. Results for these two tests are shown in the second and third columns of Table 2, respectively. Compared to Model D4, both Models S1 and S2 return similar elasticities and elpd values. That is, we find no evidence of temporal or spatial variation in the elasticity of income with respect to agglomeration.

	Model D	Sensitivity tests				
	D4	S1	S2	S3	S4	S5
ϵ^w	0.036	0.036	0.036	0.036	0.033	0.033
	(0.006)	(0.007)	(0.006)	(0.006)	(0.005)	(0.006)
ϵ^r	0.301	0.301	0.301	0.289	0.272	0.294
	(0.054)	(0.053)	(0.055)	(0.049)	(0.049)	(0.052)
η				0.149		
				(0.030)		
E_p	0.063	0.063	0.063	0.062	0.057	0.059
	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
E_c	0.024	0.024	0.024	0.022	0.021	0.026
	(0.012)	(0.013)	(0.012)	(0.011)	(0.011)	(0.012)
δ_t^w	No	No	Yes	No	No	No
γ_i^w	No	Yes	No	No	No	No
δ_t^r	Yes	Yes	Yes	Yes	Yes	Yes
γ_i^r	Yes	Yes	Yes	Yes	Yes	Yes
SUR	No	No	No	No	Yes	No
Corr.	No	No	No	No	No	Yes
R^2	0.994	0.994	0.994	0.994	0.994	0.994
elpd	$2,214$	$2,213$	$2,214$	$2,211$	$2,258$	$2,212$

Table 2: Estimated elasticities for sensitivity tests of Model D4, with standard errors in parentheses. All models use 1,180 observations and include individual time and location effects, τ_t and ζ_i.

4.2.2 Income-elasticity of rents

In another test ("S3"), we include the spatial income premium at the place-of-residence, $\ln w_{it}^{*h}$, in the rent equation, which becomes $\ln r_{it}^* \sim\ t((\epsilon^r + \delta_t^r + \gamma_i^r) \ln D_{it} + \eta \ln w_{it}^{*h} + \tau_t^r + \zeta_i^r, \nu^r, \sigma^r)$.[18] With this specification, we can test whether housing expenditure, $r_{it} H_{it}^*$, is a constant share α of income, w_{it}—that is, η =1—as predicted by the Marshallian demand for housing in Equation 2, Section 2.1. Results for Model S3 are shown in the fourth column of Table 2. We find similar elasticities to Model D4, although the estimated income elasticity of rents, η, has a median value of 0.15 and a 95% credible interval of [0.09, 0.21]. Thus, we find evidence to reject the assumption that housing costs are a constant share of income and, instead, find rents are inelastic with respect to income.[19] The elpd for Model S3 is lower than that for Model D4, which implies adding income to the rent equation brings about a deterioration in model performance.

[18] Uncertainty in $\ln w_{it}^{*h}$ is modelled using the standard errors from the first-stage regressions. We note $\ln w_{it}^{*h}$ and $\ln w_{it}^*$ denote the income premia at the place-of-residence and place-of-work, respectively.

[19] Differences in the samples used to calculate the spatial income and rent premia may cause η to be biased downwards. Specifically, whereas income premia are derived from data on all full-time workers aged 25-years or older, rent premia pertain only to the subset of dwellings that are rented. As workers who own their own home are expected to have, on average, higher incomes than those who rent, locations with higher levels of home ownership are likely to have higher $\ln w_{it}^{*h}$ but not necessarily higher $\ln r_{it}^*$—causing η to be biased downwards. We do not expect this issue to affect the estimated rent elasticities, however.

4.2.3 Additional correlations

As a final test, we allow for two additional types of correlation. First, we allow error terms to be correlated between equations ("S4"), as in a seemingly unrelated regression, or "SUR" (Zellner and Ando, 2010). Second, we allow for correlations between the individual time and location effects, τ_t and ζ_i, which are common to both equations ("S5").[20] Results for these two tests are shown in the final two columns of Table 2. In S4 we find a large, positive, and precisely-estimated correlation (0.76, s.e. 0.02) between the residuals of the two equations. In S5, we find similarly large, positive, and precisely-estimated correlations between the time (0.69; s.e. 0.20) and location (0.44, s.e. 0.12) effects. In both tests, however, the estimated income and rent elasticities—and associated agglomeration economies—are very close to those for Model D4.

5 Discussion

Which model do we prefer? In terms of performance, Model S4 has a higher elpd than Model D4. Inspection of the posterior distributions for the income and rent elasticities, however, reveals only minor differences between the two models. Although Model S4 returns slightly more precise parameter estimates, the associated estimates of agglomeration economies, E_p and E_c, are largely unchanged. That is, for all practical purposes, both Models D4 and S4 have similar economic implications. Compared to Model S4, however, Model D4 is both simpler and quicker to estimate. On this basis, we use Model D4 to inform this discussion while noting results carry over to Model S4. In terms of findings, we begin by comparing estimates derived from micro-data and aggregate data.[21] Interestingly, we find the gap between estimates derived from aggregate data and micro-data has remained broadly constant over time: That is, we have evidence that sorting matters, although its effects appear to be relatively stable—at least in New Zealand.

Figure 3 shows trends in estimates of agglomeration economies in production (left panel) and consumption (right panel) for Model D4. We find subtle temporal variation: From a nadir in 1981, estimates peak in 1991, and then gradually decline. Between 1991 and

[20] That is, we assume individual effects are drawn from bivariate normal distributions, $\tau \sim \mathcal{N}_2(0, \Sigma_\tau)$ and $\zeta \sim \mathcal{N}_2(0, \Sigma_\zeta)$, where off-diagonal elements of the covariance matrices, Σ_τ and Σ_ζ, can be non-negative.

[21] We estimate an equivalent version of Model D4 using aggregate data, although without errors-in-outcomes. Results for this model are not reported but are available on request from the authors.

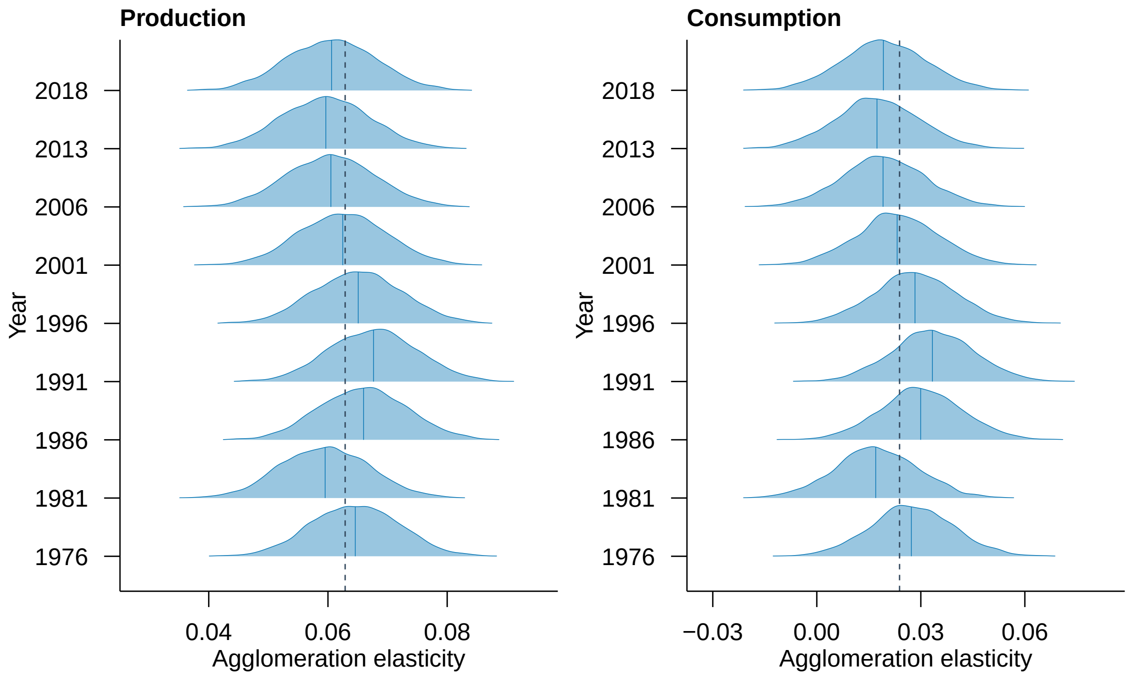

Figure 3: Trends in agglomeration economies in production (left panel) and consumption (right panel), as derived from Model D4. The dashed vertical lines denote the median for the sample.

2006, agglomeration economies in production and consumption fell by approximately 0.7% and 1.4%, respectively, but have since been relatively stable. Although the magnitude of this decline is similar to that in Donovan et al. (2021), which draws on data for 54 countries, the peak in the latter study occurs approximately one decade after that which we find here for New Zealand. Informally, we observe less temporal variation for the three most recent Censuses compared to those in the period 1976–2001. As well as the general factors discussed in Donovan et al. (2021), such as ICT technologies, we note that, in the 1980s, New Zealand implemented major economic policy reforms in response to a fiscal crisis (Evans et al., 1996). These reforms included, but were not limited to, the complete removal of agricultural subsidies, which reduced the financial returns from agriculture and led to a collapse in the price of agricultural land (Vitalis, 2007). Greater temporal variation in the period 1976–2001 may reflect the lingering effects of these shocks. This explanation has parallels to recent legal scholarship by Sitaraman et al. (2021), which argues deregulatory initiatives in the U.S. during the 1980s and 1990s—for example, in transport, communications, trade policy, and anti-trust and corporate consolidation—exacerbated urban and rural inequalities. That is, major policy reforms can have spatial dimensions. Notwithstanding variation early in the period we analyse, estimates of agglomeration economies have in recent times remained relatively stable.

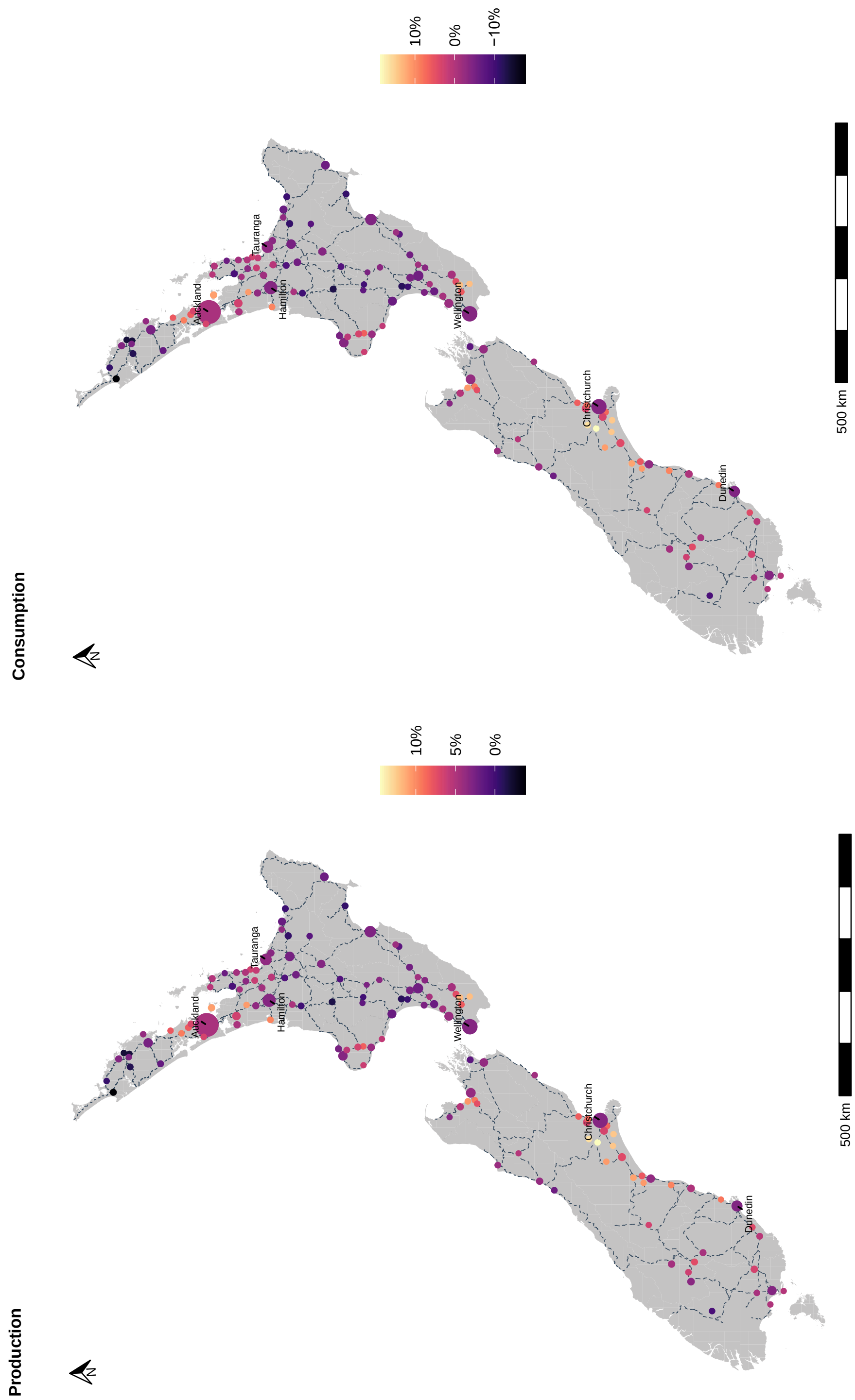

Figure 4: Agglomeration economies in production (bottom) and consumption (top). These are the median estimate for each location in 2018, as per Model D4. Labels denote the six locations with the largest populations, whereas dashed lines denote highways.

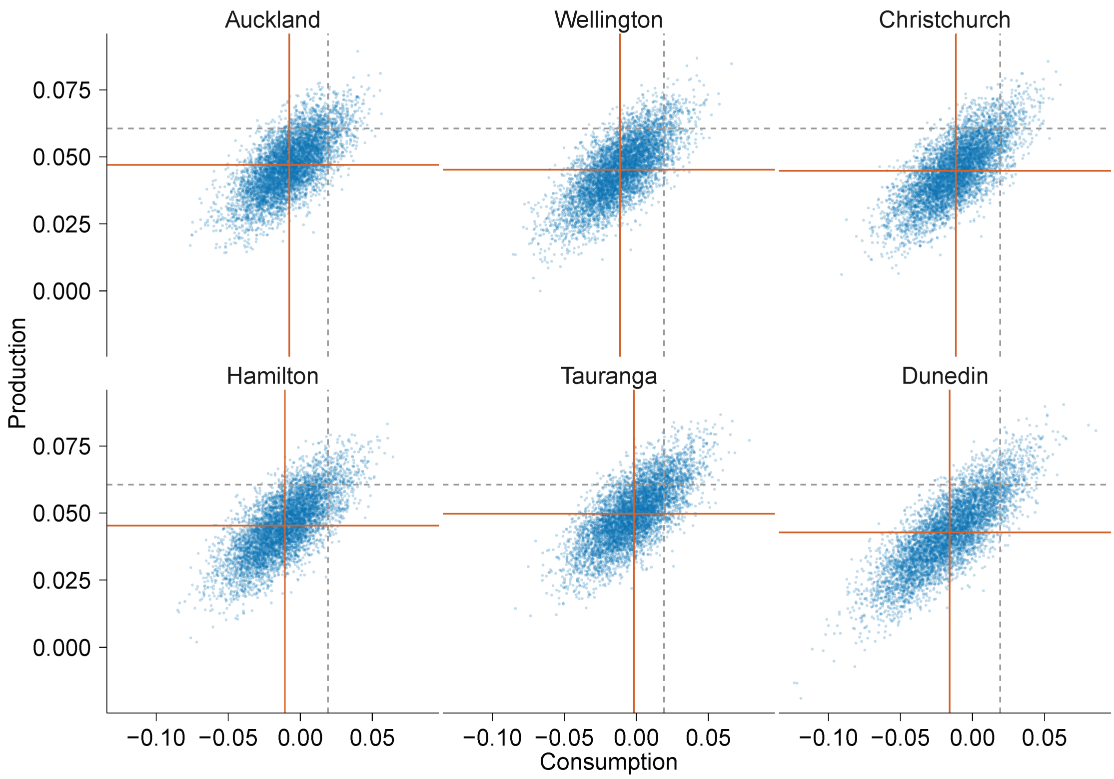

Figure 5: Agglomeration economies in production (vertical axes) and consumption (horizontal axes) for the six largest cities in New Zealand, as derived from Model D4. The dashed lines indicate the median for the sample whereas solid lines indicate the median for each location.

In contrast, we find more significant spatial variation in agglomeration economies. Figure 4 maps median estimates of agglomeration economies in production and consumption for each location in 2018. We observe some clustering in the results: Locations that are close to larger cities—such as Auckland (population 1.4 million), Wellington (population 404,000), and Christchurch (population 385,000)—tend to have stronger agglomeration economies, and vice versa for more remote locations. Figure 5 then plots agglomeration economies in production (vertical axes) versus consumption (horizontal axes) for the six largest cities. In these panels, each point represents an individual posterior estimate, whereas the solid (orange) lines indicate the median for each city and the dashed (grey) lines indicate the median for the sample. In these cities, we find estimates of agglomeration economies in production and consumption of approximately 5% and 0%, respectively, which is smaller than but close to the median for locations in the sample. That is, increased agglomeration enhances productivity but not consumption in New Zealand's largest cities. This result aligns with earlier research that finds reported well-being in New Zealand's larger urban areas is lower than average (Morrison, 2011).

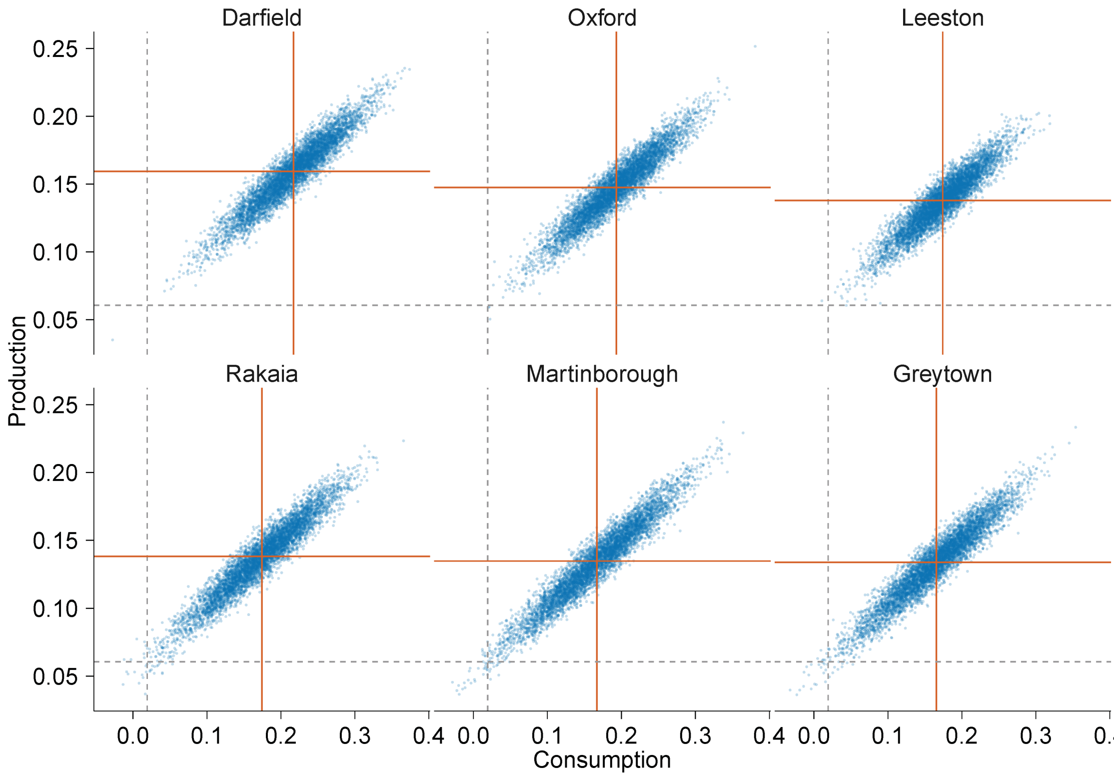

Figure 6: Agglomeration economies in production (vertical axes) and consumption (horizontal axes) for the six locations with the largest average effects, as derived from Model D4. The dashed lines indicate the median for the sample whereas solid lines indicate the median for each location.

Broad similarities in the results for larger cities in New Zealand, however, belies notable differences for smaller locations. In Figure 6, we plot the six locations with the largest agglomeration economies in production and consumption, as measured by the geometric average of the medians of the two. Here, we find agglomeration economies in production and consumption of approximately 15% and 20%, respectively. Not only are these estimates several times larger than the average but also, in these locations, agglomeration economies in consumption exceed those in production. As for why, we can only speculate. Notably, all six locations are small—with resident populations less than 3,000—and are located close to larger urban areas. Stronger agglomeration economies in "satellite" locations may indicate there are additional benefits from proximity to larger cities that are not captured in our agglomeration measure. We return to this question below.

At the other end of the spectrum, we find several locations with weaker than average agglomeration economies. Figure 7 illustrates the six locations with the lowest geometric average for estimates of agglomeration economies. For these locations, we estimate

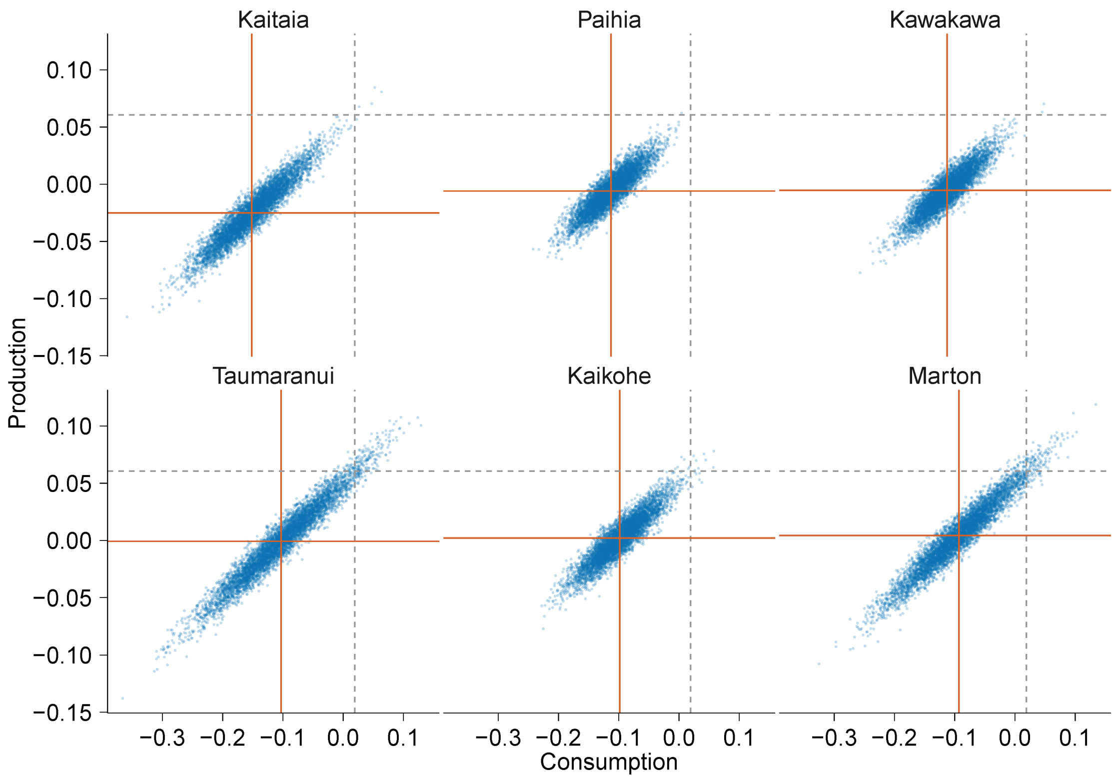

Figure 7: Agglomeration economies in production (vertical axes) and consumption (horizontal axes) for the six locations with the smaller effects, as derived from Model D4. The dashed lines indicate the median for the sample whereas solid lines indicate the median for each location.

agglomeration economies in production and consumption of approximately 0% and -10%, respectively. To explain these results, we can again only speculate. All six towns are relatively small, with populations less than 7,000 and, more interestingly, all are relatively remote. Again, we observe hints of clustering: Four of the six locations—namely, Kaitaia, Paihia, Kawakawa, and Kaikohe—are located in the same Far North District of New Zealand, whereas the other two are both located in the central North Island.

Finding that locations close to large cities experience stronger agglomeration economies, and vice versa, begs the question of whether our measure of agglomeration is misspecified? In Section 3, we adopt an ad-hoc specification for agglomeration, $D_{it} = P_{it} + \sum_{j \neq i} \frac{P_{jt}}{d_{ij}}$, in which the population, P_{it}, of location i in time t is added to the populations of other locations, P_{jt}, inversely-weighted by the travel-distance in kilometres between the two, d_{ij}. Like a "market potential" measure, this specification causes the weight that is attached to the population of other locations to decrease linearly with the distance between locations (Harris, 1954). We considered several alternative specifications for agglomeration. First, we introduced a decay factor, κ, into the measure of agglomeration, $D_{it} = P_{it} + \sum_{j \neq i} \frac{P_{jt}}{d_{ij}^{\kappa}}$, to control how the contribution of other locations, P_{jt}, attenuates with distance, d_{ij}. We estimate several models where κ ranges from 0.80 to 1.20. Although we find some evidence that $\kappa < 1$, the resulting income and rent elasticities are essentially unchanged. Second, we test a specification of agglomeration where the contribution of other locations is subject to an exponential decay that is based on time, rather than distance. Following Ahlfeldt, Redding et al. (2015), we set the coefficient of travel-time to -0.07. Compared to Model D4, we find similar elasticities and worse model performance. Third, we test a non-linear specification for agglomeration, where $\ln D_{it}$ enters with both linear and quadratic terms. Again, performance deteriorates compared to Model D4. These results suggests spatial patterns in our estimates are unlikely to arise from straightforward misspecification in the measurement of agglomeration.

To explore alternative explanations for the clustering hinted at in our results, we adopt more formal—yet still ad-hoc—definitions for types of locations. For satellites, we consider two definitions: "Satellite A" describes those locations that are within 90-minutes travel-time of cities with populations of more than 300,000 and "Satellite B" describes those locations that are within 45-minutes travel-time of cities with populations of more than 50,000, based on 2018 populations.[22] In contrast, "Remote" describes locations more than 60-minutes travel-time from cities with populations of 50,000 or more. Figure 8 relates these definitions to estimates of agglomeration economies in production. The top left panel plots estimates of agglomeration economies in production for each location (vertical axis) versus agglomeration (horizontal axis), $\ln D_i$—where the length of error bars indicates the 90% credible interval for each location. In this way, the top left panel illustrates both the heterogeneity that exists between locations and the amount of uncertainty that exists within locations. Heterogeneity in location-specific estimates of agglomeration economies exhibits a familiar "funnel" pattern, or regression to the mean, with variation tending to decrease with agglomeration, $\ln D_i$. Figure 8 also reveals that uncertainty in estimates within locations tends to reduce with agglomeration. This may simply reflect that larger cities are associated with more precise spatial premia or, perhaps, that they are less affected by idiosyncratic shocks—possibly due to greater economic diversity. We find all locations with "effective populations" $D_i \gtrapprox 50,000$ return estimates of agglomeration economies in production that are close to, albeit slightly below, the median (dashed horizontal line). Put simply, as locations increase in size, we find that our estimates of agglomeration economies converge more than they fork.

[22] In 2018, only three cities—namely, Auckland, Wellington, and Christchurch—had populations greater than 300,000 (Satellite A), whereas 17 cities had populations greater than 50,000 (Satellite B).

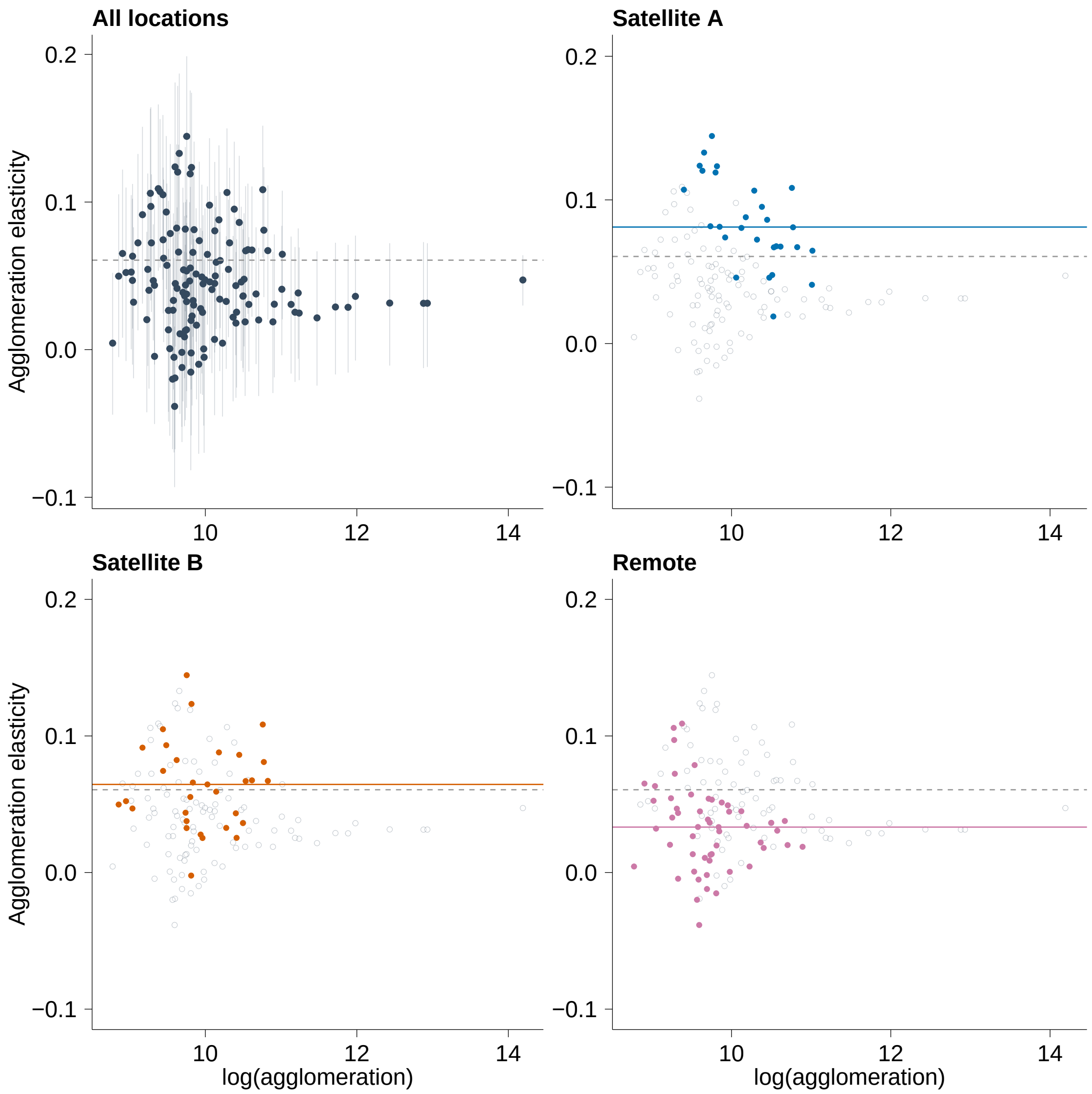

Figure 8: Top-left panel: Estimates of agglomeration economies in production for each location (vertical axis) versus agglomeration (horizontal axis), $\ln D_i$. The top-right, bottom-left, and bottom-right panels show the median estimates of agglomeration economies for "Satellite A", "Satellite B", and "Remote" locations, respectively, versus the median for the sample.

The top-right, bottom-left, and bottom-right panels in Figure 8 then show the median estimates of agglomeration economies in production for "Satellite A", "Satellite B", and "Remote" locations, respectively, as previously defined. Where locations meet the definitions for both Satellite A and B, then they are shown in both panels. Estimates for "Satellite A" and "Remote" locations tend to lie above and below the median, respectively, whereas those for "Satellite B" are more mixed. To understand whether these definitions have any bite, we adapt Model D4 to include a group-level intercept per location-type as well as an interaction between location-type and agglomeration. Notably, the inclusion of these terms improves model performance, with “Satellite A” locations returning higher than average rent elasticities. We posit “Satellite A” locations may experience stronger agglomeration economies as they offer some of the benefits of large cities but with lower overall congestion costs. Locations close to large cities may, for example, benefit from access to specialised infrastructure and institutions, like major airports and universities.[23] The latter do not feature explicitly in our measure of agglomeration but may augment its microeconomic effects, for example by expanding accessibility or enhancing knowledge spillovers (Brueckner, 2003; X. Chen et al., 2021; Kantor and Whalley, 2014; Cermeño, 2019). At the same time, these small satellite locations appear to be free of congestion.

To finish, we consider three limitations of our analysis. First, we find that rent is inelastic with respect to income (c.f. Section 4.2.2), which contravenes our model of worker preferences and might suggest the need for a more flexible approach. However, when we add income to the rent equation in Model S3 (c.f. Table 2), we find the elasticity, ϵ^r, is largely unchanged. This suggests that our results are likely to be robust to more flexible models of worker preferences in which housing is not a constant share of expenditure. Second, our estimates of the spatial rent premium do not control for some important dwelling attributes, or characteristics, such as land area and construction quality. Aspects of our methodology, however, help mitigate the potential bias introduced by unobserved dwelling attributes. Persistent differences in rent levels between locations, for example, are captured in the individual effects, ζ_i^r. This means that bias would need to originate in variation in unobserved dwelling attributes that affects locations differently over time. To the extent that temporal variation in unobserved dwelling attributes is correlated with income, then this potential bias is likely to be addressed by Model S3, which includes income in the rent equation. Third and finally, we do not address the risk of endogeneity. Our reading of the literature on wage and rent elasticities, however, leaves us somewhat sanguine about this risk. Combes, Duranton, Gobillon and Roux (2010), for example, find endogeneity in the quantity of labour supply has only modest effects on wage elasticities. Similarly, Combes, Duranton and Gobillon (2019) estimate the elasticity of urban house prices with respect to population in France and report estimates of 0.176–0.304 for pooled OLS (c.f. Table 4, p. 1573) versus 0.215–0.267 for instrumental variables (c.f. Table A1, p. 1586). As our estimated income and rent elasticities are close to those reported in other studies, including those that control for endogeneity, we suggest the latter is likely to pose only a small risk to our results and findings.

[23] The importance of access to airports may be amplified by New Zealand’s relative international remoteness, complex domestic geography, and lack of long-distance passenger rail services.

6 Conclusion

When thinking about agglomeration economies, our results suggest that "The Garden of Forking Paths" analogy may need an extra dimension: A multiplicity of outcomes can arise over both time *and* space. That said, the temporal variation we find is relatively subtle: From a nadir in 1981, estimates peak in 1991, and then gradually decline. In addition to technological shocks, like ICT, these trends may reflect the lingering effects of major economic policy reforms implemented in the 1980s. Between 1991 and 2006, agglomeration economies in production and consumption fell by approximately 0.7% and 1.4%, respectively. After 2006, however, our estimates remain stable; contrary to popular claims, technology does not appear to have made the world "flatter". In contrast, we find evidence of more notable spatial variation in agglomeration economies: Larger cities in New Zealand offer net benefits in production, but not in consumption, whereas small towns that are close to large cities, or "satellites", experience agglomeration economies that are stronger than average. Taken together, our results suggest that agglomeration economies are rather stable—at least in recent decades—but cast some doubt on their transferability—at least within New Zealand. The latter is especially relevant to smaller towns, where our results reveal a wide range of outcomes, or "forking paths."

To the extent that our findings raise questions over the stability and transferability of agglomeration economies, they also highlight the need for further research and nuanced policies. More research is needed to trace the evolution of agglomeration economies within countries and over time and, where possible, explain the resulting variation in estimates. We also repeat calls for more research into agglomeration economies in both production and consumption. Although agglomeration seems to enhance the productivity of most cities and towns in New Zealand, we find more tenuous benefits for consumption. Whether this reflects the inevitable corrosion of amenity by congestion, or a failure to adopt effective policies to mitigate congestion, remains an open question. In terms of empirical measurement, we see merit in exploring agglomeration indicators that respond to multi-modal transport costs—rather than simply time or distance—and consider proximity to specialised infrastructure and institutions, like airports and universities. Finally, we underscore the need for nuanced urban policies. Although we find evidence of stronger agglomeration economies in satellite locations, for example, these locations typically have small populations and, presumably, less congestion—both of which are endogenous attributes that may be affected by policies. The latter would, ideally, recognise and respond to these temporal and spatial forks in the road.

References

Ahlfeldt, G. M. and E. Pietrostefani (2019). 'The economic effects of density: A synthesis'. *Journal of Urban Economics* 111, pp. 93–107.

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm and N. Wolf (2015). 'The Economics of Density: Evidence from the Berlin Wall'. *Econometrica* 83.6, pp. 2127–2189.

Ahrend, R., E. Farchy, I. Kaplanis and A. C. Lembcke (2017). 'What makes cities more productive? Evidence from five OECD countries on the role of urban governance'. *Journal of Regional Science* 57.3, pp. 385–410.

Borges, J. L. et al. (1962). 'The garden of forking paths'. *Labyrinths: selected stories and other writings*, pp. 19–29.

Brueckner, J. K. (2003). 'Airline traffic and urban economic development'. *Urban Studies* 40.8, pp. 1455–1469.

Bürkner, P.-C. (2017). 'brms: An R package for Bayesian multilevel models using Stan'. *Journal of Statistical Software* 80.1, pp. 1–28.

Cairncross, F. (1997). *The death of distance: How the communications revolution will change our lives*. Vol. 302. Harvard Business School Press Boston, MA.

Cermeño, A. L. (2019). 'Do universities generate spatial spillovers? Evidence from US counties between 1930 and 2010'. *Journal of Economic Geography* 19.6, pp. 1173–1210.

Chauvin, J. P., E. Glaeser, Y. Ma and K. Tobio (2017). 'What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States'. *Journal of Urban Economics* 98, pp. 17–49.

Chen, X., C. Xuan and R. Qiu (2021). 'Understanding spatial spillover effects of airports on economic development: New evidence from China's hub airports'. *Transportation Research Part A: Policy and Practice* 143, pp. 48–60.

Chen, Y. and S. S. Rosenthal (2008). 'Local amenities and life-cycle migration: Do people move for jobs or fun?' *Journal of Urban Economics* 64.3, pp. 519–537.

Combes, P.-P., G. Duranton and L. Gobillon (2008). 'Spatial wage disparities: Sorting matters!' *Journal of Urban Economics* 63.2, pp. 723–742.

— (2019). 'The costs of agglomeration: House and land prices in French cities'. *The Review of Economic Studies* 86.4, pp. 1556–1589.

Combes, P.-P., G. Duranton, L. Gobillon and S. Roux (2010). 'Estimating Agglomeration Economies with History, Geology, and Worker Effects'. *Agglomeration Economics*. University of Chicago Press, pp. 15–66.

Donovan, S., T. de Graaff, H. L. F. de Groot and C. Koopmans (2021). *Unravelling urban advantages—A meta-analysis of agglomeration economies*. Working Paper No. 21-026/VIII. Amsterdam, The Netherlands: Tinbergen Institute.

Evans, L., A. Grimes, B. Wilkinson and D. Teece (1996). 'Economic reform in New Zealand 1984-95: The pursuit of efficiency'. *Journal of Economic Literature* 34.4, pp. 1856–1902.

Friedman, T. L. (2006). *The world is flat: The globalized world in the twenty-first century*. Penguin London.

Gabriel, S. A. and S. S. Rosenthal (2004). 'Quality of the business environment versus quality of life: Do firms and households like the same cities?' *Review of Economics and Statistics* 86.1, pp. 438–444.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari and D. B. Rubin (2013). *Bayesian data analysis*. CRC press.

Gelman, A. and J. Hill (2007). *Data analysis using regression and multilevel hierarchical models*. Vol. 1. Cambridge University Press New York, NY, USA.

Gelman, A. and E. Loken (2013). 'The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time'. *Department of Statistics, Columbia University* 348.

Geweke, J. (1993). 'Bayesian treatment of the independent Student-t linear model'. *Journal of Applied Econometrics* 8.S1, S19–S40.

Glaeser, E. L. (2011). *Triumph of the City*. Pan.

Groot, S. P. T., H. L. F. de Groot and M. J. Smit (2014). 'Regional wage differences in the Netherlands: Micro evidence on agglomeration externalities'. *Journal of Regional Science* 54.3, pp. 503–523.

Harris, C. D. (1954). 'The Market as a Factor in the Localization of Industry in the United States'. *Annals of the Association of American Geographers* 44.4, pp. 315–348.

Holl, A. (2012). 'Market potential and firm-level productivity in Spain'. *Journal of Economic Geography* 12.6, pp. 1191–1215.

Kantor, S. and A. Whalley (2014). 'Knowledge spillovers from research universities: evidence from endowment value shocks'. *Review of Economics and Statistics* 96.1, pp. 171–188.

Kasy, M. (2021). 'Of Forking Paths and Tied Hands: Selective Publication of Findings, and What Economists Should Do about It'. *Journal of Economic Perspectives* 35.3, pp. 175–92.

Maré, D. C. and D. J. Graham (2013). 'Agglomeration elasticities and firm heterogeneity'. *Journal of Urban Economics* 75, pp. 44–56.

Maré, D. C. and J. Poot (2019). *Valuing cultural diversity of cities*. Working Paper. Motu Economic and Public Policy Research.

Martínez-Galarraga, J., E. Paluzie, J. Pons and D. A. Tirado-Fabregat (2008). 'Agglomeration and labour productivity in Spain over the long term'. *Cliometrica* 2.3, pp. 195–212.

Meier, S. and C. D. Sprenger (2015). 'Temporal stability of time preferences'. *Review of Economics and Statistics* 97.2, pp. 273–286.

Morrison, P. S. (2011). 'Local expressions of subjective well-being: The New Zealand experience'. *Regional Studies* 45.8, pp. 1039–1058.

Puga, D. (2010). 'The magnitude and causes of agglomeration economies'. *Journal of Regional Science* 50.1, pp. 203–219.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Roback, J. (1982). 'Wages, rents, and the quality of life'. *Journal of Political Economy* 90.6, pp. 1257–1278.

Rosenzweig, M. R. and C. Udry (2020). 'External validity in a stochastic world: Evidence from low-income countries'. *The Review of Economic Studies* 87.1, pp. 343–381.

RStudio Team (2021). *RStudio: Integrated Development Environment for R*. Boston, MA.

Shenker, J. (2020). 'Cities after coronavirus: how Covid-19 could radically alter urban life'. *The Guardian*.

Sitaraman, G., M. Ricks and C. Serkin (2021). 'Regulation and the Geography of Inequality'. *Duke Law Journal* 70.8, pp. 1763–1836.

Vehtari, A., A. Gelman and J. Gabry (2017). 'Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC'. *Statistics and Computing* 27.5, pp. 1413–1432.

Vitalis, V. (2007). 'Agricultural subsidy reform and its implications for sustainable development: the New Zealand experience'. *Environmental Sciences* 4.1, pp. 21–40.

Waka Kotahi (2020). *Monetised Benefits and Costs Manual*. Technical Report Version 1.5. The New Zealand Transport Agency.

Zellner, A. and T. Ando (2010). 'A direct Monte Carlo approach for Bayesian analysis of the seemingly unrelated regression model'. *Journal of Econometrics* 159.1, pp. 33–45.