UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS

UCL

WORKING PAPERS SERIES

Paper 223 - June 20

Gravity Model Calibration by Rent

ISSN 1467-1298

Centre for Advanced Spatial Analysis University College London Gower St London WC1E 6BT
Tel: +44 (0)20 3108 3902 casa@ucl.ac.uk www.casa.ucl.ac.uk

Gravity Model Calibration by Rent

Robin Morphet and Zahratu Shabrina

27th May 2020

1 Introduction

The maximum entropy spatial interaction model is derived and shown to incorporate both trip cost and von Thünen (Puu 1997) rent per trip, a pure location rent. The similarity in result between calibrating against mean trip cost and calibrating against rent is then demonstrated. This is shown by generating a sequence of models using given values of β and then calibrating these by rent to give a comparison of β values. An example of the use of location rent as a measure of accessibility is outlined.

- The paper is structured as follows:
  - Section 2: The underlying gravity model is derived using maximum entropy
  - Section 3: The classic von Thünen rent/trip cost model is outlined
  - Section 4: Defining von Thünen rent within the gravity model
  - Section 5: Estimating the von Thünen rent from the balancing factors of the gravity model
  - Section 6: A description of the calibration method using J-divergence and balancing factors
  - Section 7: A demonstration of calibrating a gravity model using balancing factors and J-divergence
  - Section 8: A demonstration of the method used to derive Airbnb accessibilities to tourist destinations in London
  - Section 9: A discussion of accessibility and its equivalence to rent
  - Section 10: Conclusions

This paper is written in R Markdown and is available, together with code and data, on a GitHub repository as an R Studio project on https://github.com/robinmorphet/gravity-rent-calibration.

2 Deriving the Gravity Model

The maximum entropy model is derived by Wilson (1970). We follow that process but work in probabilities rather than trips because we regard the latter as random variables whereas the probabilities are measures. We begin by constructing the Lagrangian $\mathcal{L}$

$$\mathcal{L} = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} ln p_{ij} + \lambda_0 \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} - 1 + \sum_{i=1}^{n} \lambda_i \sum_{j=1}^{n} p_{ij} - p_{i*} + \sum_{j=1}^{n} \lambda_j \sum_{i=1}^{n} p_{ij} - p_{*j} + \beta \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} c_{ij} - \bar{c} \tag{1}$$

Differentiating $\mathcal{L}$ with respect to p_{ij} and setting the result to zero delivers the maximum entropy model thus

$$p_{ij} = e^{-\lambda_0} e^{-\lambda_i} e^{-\lambda_j} e^{-\beta c_{ij}} \tag{2}$$

Summing both sides and recognising that λ_0 is a constant we find that

$$e^{\lambda_0} = \sum_{i=1}^{n}\sum_{j=1}^{n} e^{-\lambda_i} e^{-\lambda_j} e^{-\beta c_{ij}} = Z \tag{3}$$

We call the constant Z as it corresponds to the partition function of statistical physics and write the model as

$$p_{ij} = \frac{e^{-\lambda_i} e^{-\lambda_j} e^{-\beta c_{ij}}}{Z} \tag{4}$$

We now take logarithms, multiply by $\frac{-p_{ij}}{\beta}$ and sum over i, j to get

$$\frac{1}{\beta} S = U + PV - G \tag{5}$$

where S is entropy, U is mean energy, PV is $\frac{\lambda_i + \lambda_j}{\beta}$ and G is $-\frac{1}{\beta} lnZ$. These terms are used to show the correspondence of the model to the classical gas model (Callen 1985) where G is the free energy and also the Marshallian consumer surplus, and PV the product of pressure, P and volume V. Equation 5 gives the classical definition of G as does $-\frac{1}{\beta} lnZ$. As we will see, the value of G corresponds to the Marshallian consumer surplus and PV to rent/unit area times area where area is the 2 dimensional quantity paralleled by volume V in the classical gas model and rent/unit area corresponds to pressure.

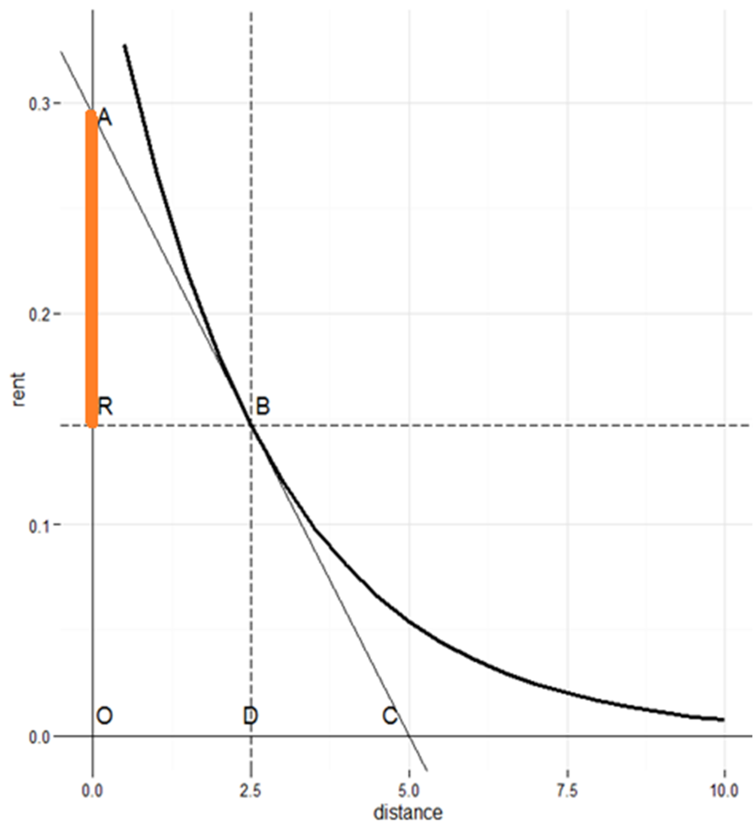

Figure 1: Von Thünen Rent v Distance

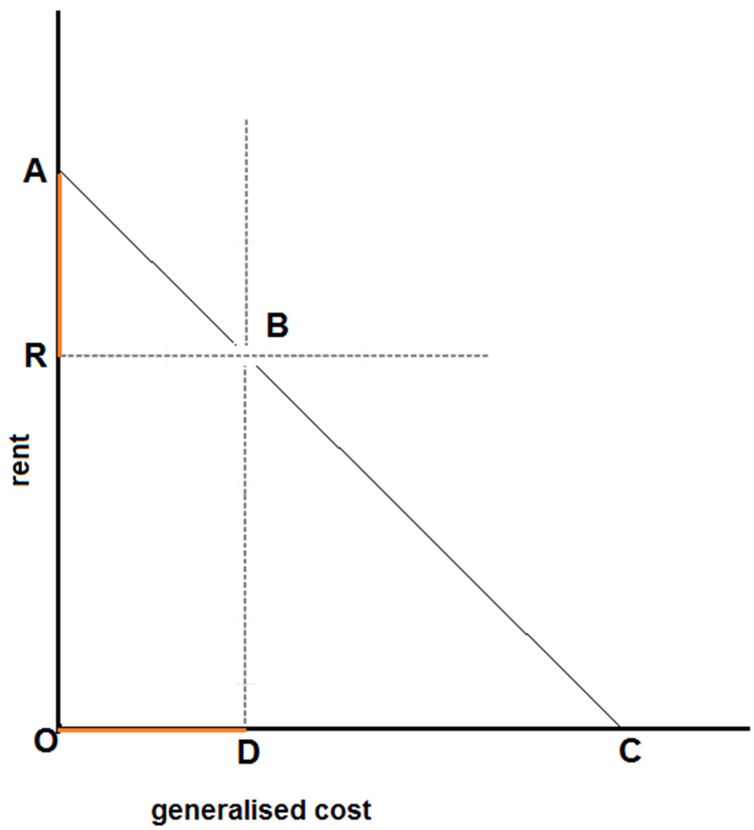

Figure 2: von Thünen Rent v Generalised Cost

3 Deriving the von Thünen Model

The von Thünen model postulates a single city surrounded by a uniform plain. Farmers sell all their product in the city. The price of the product is determined by the cost of growing it, the cost of transporting it to market and the level of profit. In the model the markets clear and the rate of profit is constant. This together with the fact that the model is deterministic and there is only a single purchaser implies that it is a model of perfect competition. A more detailed exposition can be found in Puu (1997). Figure 1 shows the rent curve against distance. The rate of cost of transport for a farmer located at D is given by the slope at B. Thus the cost of transporting produce to market fom D is AR and the level of rent paid is OR. In Figure 2 we substitute generalised cost for distance in common with standard practice in trip modelling (Dios Ortuzar and Willumsen 2011). This equalises AR and OD which gives a slope of -1. We may express this linear relationship as an equation thus:

$$rent + generalised.cost = constant \tag{6}$$

In terms of the von Thünen model the constant is the distance from the city to his "wilderness" where rent has declined to zero. In more modern terms it is that cost of transport which renders the goods in question too expensive for, or beyond the range of, the market.

4 Matching the Gravity and von Thünen Models

In the gravity model we have n destinations but in the von Thünen model we have only one. The table of trips from origin to destination in the gravity model is thus replaced by a single column showing trips from the origins to the single centre. In the von Thünen model the origin zones are annular rings with the city at their centre. We therefore set up the model adapting Equation 4 with a single destination zone, k, and without a destination constraint as all produce is sold in one place.

$$p_{ik} = \frac{e^{-\lambda_i} e^{-\beta c_{ik}}}{Z} \tag{7}$$

Of course, knowing the p_i we know the p_{ik} but for the moment we choose to ignore this and in fact all we need to know is that the p_i exist and are fixed. To make the comparison between the two models we exploit the Legendre transform (Kennerly 2011; Zia, Redish, and McKay 2009; Callen 1985) which underlies the structure of entropy maximising gravity models (Lesse 1982). Following Callen's (1985) exposition we consider a variable $Y(X)$ and form its partial differential $\frac{\partial Y}{\partial X}$. The Legendre transform, Ψ, is then given by

$$\Psi\left(Y\right) = -\frac{\partial Y}{\partial X} X + Y \tag{8}$$

Applying this equation to the straight line graph of Figure 2 and Equation 6 we get

$$\Psi\left(rent\right) = 1.rent + generalised.cost \tag{9}$$

We now determine the equivalent Legendre transform using Equation 7 but with $p_{ik} = p_i$ a constant as is Z. Rearranging and taking logarithms we get

$$\frac{\lambda_i}{\beta} = -\frac{lnZp_i}{\beta} - c_{ik} \tag{10}$$

Differentiating with respect to generalised cost we get

$$\frac{\partial\left(\frac{\lambda_i}{\beta}\right)}{\partial c_{ik}} = -1 \tag{11}$$

giving the Legendre transform

$$\Psi\left(c_{ik}\right) = \frac{\lambda_i}{\beta} + c_{ik} \tag{12}$$

Comparing Equation 12 and Equation 9 we see that

$$rent = \frac{\lambda_i}{\beta} \tag{13}$$

The rent, like the trip cost, is a cost per trip. The identification of the balancing factors with rent has a somewhat chequered history. The relationship was initially suggested by Dieter (1962) but suffered widespread rejection (e.g. Kirby 1970) although it was resuscitated to an extent in Williams and Senior (1978) who interpreted λ_i and λ_j as penalty functions in a non linear program method. Alonso (1964) appealed to the von Thünen model in his analysis of the residential housing market around a single centre (implying perfect competition) using a bid rent curve of the kind shown in Figure 3. The later identification of von Thünen rent with the balancing factors (Morphet 2012) also showed that the polycentric gravity model is a model of imperfect competition in which the measure of imperfection (reflecting the triangle of Harberger(1964)) is:

$$\frac{1}{\beta} I\left(p_{ij}, p_i p_j\right) = \frac{1}{\beta} \sum_i \sum_j p_{ij} ln \frac{p_{ij}}{p_i p_j} \tag{14}$$

the Mutual Information, I, times $\frac{1}{\beta}$ and p_i and p_j are origin and destination probabilities, their product giving a trip matrix for the case of zero cost i.e. the case of perfect competition where impediments to trade in the form of trip costs have been removed.

5 Estimating the rent from the balancing factors

The balancing factors are derived directly from the Furness iteration. We can rewrite Equation 4 as

$$p_{ij} = \frac{e^{-\lambda_i} e^{-\lambda_j} e^{-\beta c_{ij}}}{\sum_i \sum_j e^{-\lambda_i} e^{-\lambda_j} e^{-\beta c_{ij}}} \tag{15}$$

rearranging the denominator we may write

$$p_{ij} = \frac{e^{-\lambda_i} e^{-\lambda_j} e^{-\beta c_{ij}}}{\sum_i e^{-\lambda_i} \sum_j e^{-\lambda_j} e^{-\beta c_{ij}}} \tag{16}$$

and we can write for the origin balancing factors bf_i

$$bf_i = \frac{e^{-\lambda_i}}{\sum_i e^{-\lambda_i}} \tag{17}$$

and taking logarithms and the using the Z notation we have

$$ln\left(bf_i\right) = -\lambda_i - ln\left(\sum_i e^{-\lambda_i}\right) = -\lambda_i + Z_i \tag{18}$$

To identify λ_i we take the logarithm of the origin balancing factors which must then be corrected by a constant. This may be based on existing estimates of land values or by identifying a minimum value and adding a constant to the logged balancing factors to ensure that there are no negative rents. The value for λ_j is derived in a similar fashion. It should be noted that adding a constant to either λ_i or λ_j does not affect the value of p_{ij} as it is equivalent to adding a constant to $-\lambda_i$ and $-\lambda_j$ in Equation 15 which will be present in both numerator and denominator and so cancel out.

6 Calibrating the Gravity Model

The standard method for calibration, i.e. finding the value of β, is to run the model with varying values of β in a search to find that value which gives a mean trip cost sufficiently close to the observed trip cost (Hyman and Wilson 1969; Dios Ortuzar and Willumsen 2011). The method of running the model is to use a Furness iteration (Furness 1965) to balance Origins and Destinations for a given value of β. This introduces the balancing factors $e^{-\lambda_i} e^{-\lambda_j}$ of Equation 2 which relate to the rents of Equation 13. The iteration provides a result in which the balancing factors are unique (Lemma 2, Evans 1970) which ensures that the resulting trip matrix, given β, is also unique. In the method of calibration by rent, instead of matching observed and modelled mean trip costs, we match observed and modelled rents by minimising the J divergence (Rohde 2016) between them. The J divergence is an information based semimetric measure which satisfies all the conditions for a distance measure apart from the triangle inequality. In particular it is finite, symmetric and decomposable. The J divergence between two probability distributions, $p\left(x_i\right)$ and $q\left(y_i\right)$ is given by

$$J = \frac{1}{2} \sum_i \left(p\left(x_i\right) - q\left(y_i\right)\right) \ln\left(\frac{p\left(x_i\right)}{q\left(y_i\right)}\right) = \frac{1}{2}\left(I\left(p,q\right) + I\left(q,p\right)\right) \tag{19}$$

7 A Demonstration

In this demonstration we use a small model for which we know the value of β is 0.1. We then extract the origin balancing factors which will act as our target distribution. The model is then run for for several values of β which are the compared using J divergence and the value of β estimated at the point of minimum divergence. Our cost data is given by the 5x5 table:

Table: Inter Zonal Trip Costs

	1	2	3	4	5
1	10.0	14.1	14.1	14.1	14.1
2	14.1	10.0	20.0	28.3	20.0
3	14.1	20.0	10.0	20.0	28.3
4	14.1	28.3	20.0	10.0	20.0
5	14.1	20.0	28.3	20.0	10.0

the origins Oi and destinations Dj are given by

Table: Origins

1	2	3	4	5
500	500	3000	5000	1000

Table: Destinations

1	2	3	4	5
5000	3000	1000	500	500

We now set the deterrence function matrix

Table: Deterrence function

	1	2	3	4	5
1	0.3678794	0.2441433	0.2441433	0.2441433	0.2441433
2	0.2441433	0.3678794	0.1353353	0.0590129	0.1353353
3	0.2441433	0.1353353	0.3678794	0.1353353	0.0590129
4	0.2441433	0.0590129	0.1353353	0.3678794	0.1353353
5	0.2441433	0.1353353	0.0590129	0.1353353	0.3678794

We now iterate the deterrence matrix to the row an column totals which are expressed as probabilities rather than trips and we extract the origin balancing factors:

Table: Balancing Factors: Beta = 0.1

1	2	3	4	5
0.1599941	0.1609328	1.465261	3.236415	0.5107356

We now know our target values of β and of the origin balancing factors. It remains for us to construct a search for the value of β knowing only the origin balancing factors.

This gives us a list of balancing factors for each value of β. From Equation 17 we see however, that the balancing factors are standardised by their sum. For this reason it is appropriate to use the J-divergence as the measure of deviation since it too assumes probabilities standardised to sum to 1. We now compute the value of the J divergence for each set of balancing factors. We form the latter into a dataframe the head of which is shown below

Table: Values of balancing factors by Beta

beta	1	2	3	4	5
0.080	0.0323141	0.0336704	0.2724989	0.5667591	0.0947575
0.082	0.0319594	0.0332086	0.2717458	0.5685563	0.0945299
0.084	0.0316081	0.0327470	0.2709891	0.5703576	0.0942982
0.086	0.0312601	0.0322858	0.2702288	0.5721629	0.0940624
0.088	0.0309155	0.0318252	0.2694649	0.5739719	0.0938226
0.090	0.0305741	0.0313654	0.2686974	0.5757845	0.0935787
0.092	0.0302359	0.0309065	0.2679263	0.5776004	0.0933310
0.094	0.0299009	0.0304488	0.2671515	0.5794195	0.0930793
0.096	0.0295691	0.0299923	0.2663733	0.5812415	0.0928238
0.098	0.0292403	0.0295374	0.2655914	0.5830664	0.0925645
0.100	0.0289146	0.0290842	0.2648060	0.5848937	0.0923015
0.102	0.0285918	0.0286328	0.2640171	0.5867234	0.0920349
0.104	0.0282721	0.0281835	0.2632246	0.5885552	0.0917646
0.106	0.0279553	0.0277363	0.2624287	0.5903889	0.0914908
0.108	0.0276414	0.0272914	0.2616293	0.5922244	0.0912136
0.110	0.0273304	0.0268490	0.2608265	0.5940613	0.0909329
0.112	0.0270222	0.0264092	0.2600203	0.5958995	0.0906488
0.114	0.0267168	0.0259722	0.2592107	0.5977388	0.0903615
0.116	0.0264141	0.0255382	0.2583979	0.5995789	0.0900709
0.118	0.0261137	0.0251058	0.2575762	0.6014280	0.0897763
0.120	0.0258164	0.0246778	0.2567561	0.6032704	0.0894793

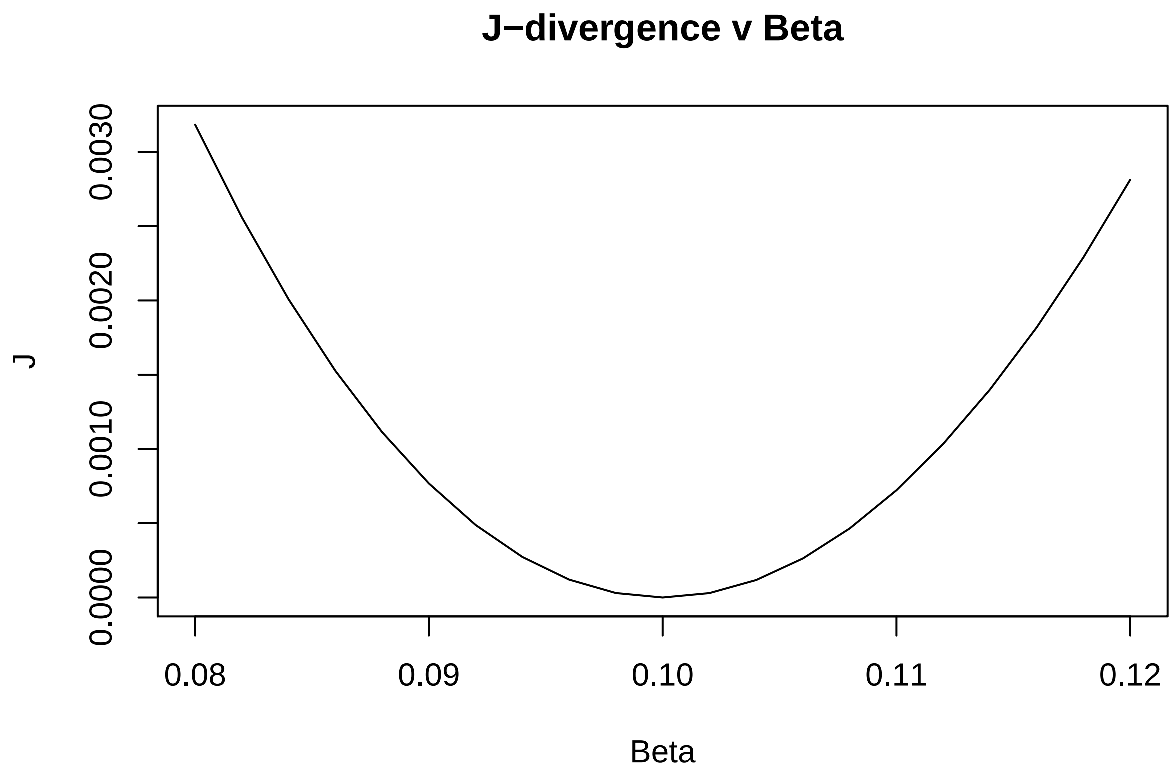

Figure 4: Calibration by balancing factors

We see from Figure 4 that we have achieved a minimum value of J at the 0.1 value of β. What we have shown in this demonstration is that we can calibrate a gravity model knowing only the balancing factors for the origins. We know however, from Equation 13 that the balancing factors are a function of rent. In the following section we explore an application in which we construct a surrogate for rent from which is then compared with the balancing factors. We then seek a minimum J fit as before.

8 An application

In this application (Shabrina 2020) we calibrate a gravity model using a surrogate for observed rent data as a target in contrast to the synthesised balancing factor data used in the previous section. The data describes the relation of Airbnb (platform-based short term holiday rentals) sites to popular tourist destination sites in London. The data is at Lower Super Output Area level which, being a relatively small area, means that some LSOAs have zero Airbnb sites and are therefore ignored in the initial calculation. So after reading in the data it was cleaned by determining an index of LSOAs with zero Airbnb sites which was then used to remove them from the data.

The cost data is journey time data from the LSOAs to the tourist attractions. This is read in as before, but with the zero LSOAs already removed. The times were then converted to minutes.

The proxy for origins was the number of Airbnb bedrooms in each LSOA and the proxy for destinations was the number of visitors to each attraction. Both origins and destinations were converted to probabilites as the model of Equation 1 is defined in terms of probabilities. This process also ensures that the origin total equals the destination total which is one of the requirements of the iteration used to estimate the model.

We now construct the target rent which we are trying to match with our model. This is, for each LSOA, the cost per bed times the number of beds. This gives us a total rent for each LSOA which we then convert to probabilities.

The search range for β is then generated. The particular range chosen is determined by trial and error but in general we would expect to find a value between 0 and 2 (Hyman and Wilson 1969). The model is then run producing balancing factors and a matrix of trips. For each value of β a set of statistics is produced. Of most concern to us here are the calculated J-divergences.

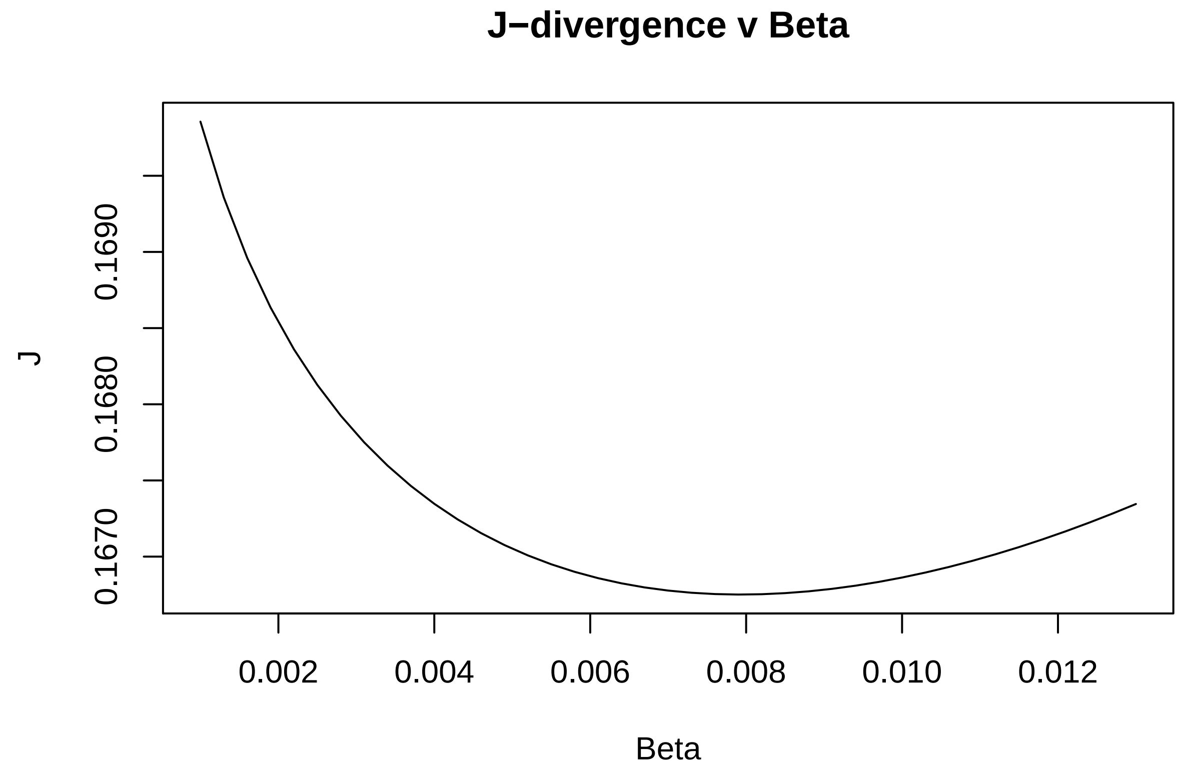

Figure 5: Calibration against zonal bed cost

Figure 5 suggests a best fit value for β of 0.008. Knowing this value we then generate the value of the balancing factors and hence of the rent proportions. These relative rents are taken as relative accessibilities and can be used to generate an accessibility surface in which accessibilities for the zones without Airbnb beds can be estimated by interpolation.

9 Accessibility

The early reference by Hansen (1959) to accessibility gave a definition which in terms of the model defined in Equation 7 can be written

$$a_i = \sum_{i}^{n} p_j e^{-\beta c_{ij}} \tag{20}$$

where a_i is accessibility.

This gravity formulation may be contrasted to Hansen's original (1959) in which c_{ij} was distance and α

equalled 2

$$a_i = \sum_j \frac{D_j}{c_{ij}^{\alpha}} \tag{21}$$

The deterrence function in Equation 20 reflects the diminishing number of trips as c_{ij} increases whereas in Equation 21 the cost function represents a diminishing appreciation with distance of the destination potentials. Hansen (1959) is arguing on intuitive grounds for a particular set of preferences whereas in the gravity model these are more akin to preferences revealed under the constraints of the model.

In practice this measure has been found unsatisfactory (Dios Ortuzar and Willumsen 2011). A review by Srour et al (2002) compared accessibility measures with empirically estimated land values confirmed the link between land value and accessibility, particularly to jobs. We are more fortunate in being able to derive the relationship theoretically from the von Thünen assumptions within the gravity model. Srour was less successful in identifying the appropriate measure of accessibility. The use of a logsum method proved problematic. Niemeyer (1997) argued for a logsum method suggesting that the logsum measure of consumer surplus equated to accessibilty. If we consider accessibility as rent then this is only true if the model is one of perfect competition i.e. of zero transport cost. The gravity model, however, is a model of imperfect competition. The logsum approach is one of utility or value maximisation with the contentions that this brings whilst the gravity model, through rent, is an approach based on cost.

It may be asked, in relation to the application above why, when we already have a proxy for rent, do we model rent at all? The answer is that the rent we are modelling is a pure location rent. The observed rents may reflect many other hedonic attributes such as the presence of local facilities, the age and construction/maintenance costs of the building, the local environment and perceptions of safety from crime. We use the gravity model to extract the pure location rent.

10 Conclusions

We have shown that the gravity model can be calibrated against the balancing factors. This should not be a surprise since the more usual method of calibration against mean trip cost uses only one parameter whereas the number of (origin) balancing factors equals the number of zones. We have shown that the J-divergence is an effective statistic to minimise in order to achieve a good fit. We have demonstrated theoretically the relation between balancing factors and location rent which we argue is a good measure of accessibility. This has been used in an application to determine accessibilities of Airbnb properties in London which seems to perform well and is shown elsewhere to perform rather better than other indices. The use of this method in transportation practice may be not so much to replace mean trip cost calibration but rather to update models continuously as new rent patterns are observed. Within the practice of Land-Use Transportation Interaction modelling it is not clear that the conventional property price iterations are consistent with the gravity rents or by implication, the underlying model itself. The recognition of the von Thünen location rent within the gravity model lays the basis for a dynamic based on trip cost reduction followed by increased rent and hence densification with the consequent increase in congestion determining the need for further trip cost reduction.

References

Alonso, W. 1964. *Location and Land Use. Toward a General Theory of Land Rent.* Cambridge, Mass: Harvard University Press.

Callen, H. B. 1985. *Thermodynamics and an Introduction to Thermostatistics.* Second. John Wiley.

Dieter, K. H. 1962. "Distribution of Work Trips in Toronto." *Journal of the City Planning Division*, Proceedings of the American Society of Civil Engineers, 1: 9–28.

Dios Ortuzar, Juan de, and Luis G Willumsen. 2011. *Modelling Transport.* John Wiley & sons.

Evans, A. W. 1970. "Some properties of trip distribution methods." *Transportation Research* 4 (1): 19–36.

Furness, K. P. 1965. "Time Function Iteration." *Traffic Engineering and Control* 7 (7): 458–60.

Hansen, Walter G. 1959. "How Accessibility Shapes Land Use." *Journal of the American Institute of Planners* 25 (2): 73–76.

Harberger, A. C. 1964. "The measurement of waste." *The American Economic Review* 54 (3): 58–76.

Hyman, G. M., and A. G. Wilson. 1969. "The effects of changes in travel costs on trip distribution and modal split." *High Speed Ground Transportation Journal*, 79–85.

Kennerly, S. 2011. "A graphical derivation of the Legendre transform." https://sites.google.com/site/samkennerly/Legendre.pdf?attredirects=0.

Kirby, H. R. 1970. "Normalizing Factors of the Gravity Model – An Interpretation." *Transportation Research* 4: 37–50.

Lesse, P. F. 1982. "A phenomenological theory of socioeconomic systems with spatial interactions." *Environment and Planning A* 14 (7): 869–88.

Morphet, R. 2012. "Von Thunen's Legendre Transform: Urban Rent and the Interaction Model." Working Paper 193. CASA Working Papers. UCL, London: Bartlett Centre for Advanced Spatial Analysis. https://www.ucl.ac.uk/bartlett/casa/publications/2013/jul/casa-working-paper-193.

Niemeier, Debbie A. 1997. "Accessibility: An Evaluation Using Consumer Welfare." *Transportation* 24 (4): 377–96.

Puu, T. 1997. "Mathematical Location and Land Use Theory."

Rohde, Nicholas. 2016. "J-Divergence Measurements of Economic Inequality." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179 (3): 847–70.

Shabrina, Zahratu. 2020. "The Impact of the Platform Economy in Cities: The Case of Airbnb." PhD thesis, UCL (University College London). https://discovery.ucl.ac.uk/id/eprint/10093740/.

Srour, Issam M, Kara M Kockelman, and Travis P Dunn. 2002. "Accessibility Indices: Connection to Residential Land Prices and Location Choices." *Transportation Research Record* 1805 (1): 25–34.

Williams, H. C. W. L., and M. L. Senior. 1978. "Accessibility, Spatial Interaction and the Evaluation of Land Use-Transportation Plans." In *Spatial Interaction Theory and Planning Models*, edited by A. Karlqvist, L. Lundqvist, F. Snickars, and J. W. Weibull, 253–87. Amsterdam: North Holland.

Wilson, A. G. 1970. *Entropy in urban and regional modelling*. London: Pion.

Zia, Royce KP, Edward F Redish, and Susan R McKay. 2009. "Making Sense of the Legendre Transform." *American Journal of Physics* 77 (7): 614–22.