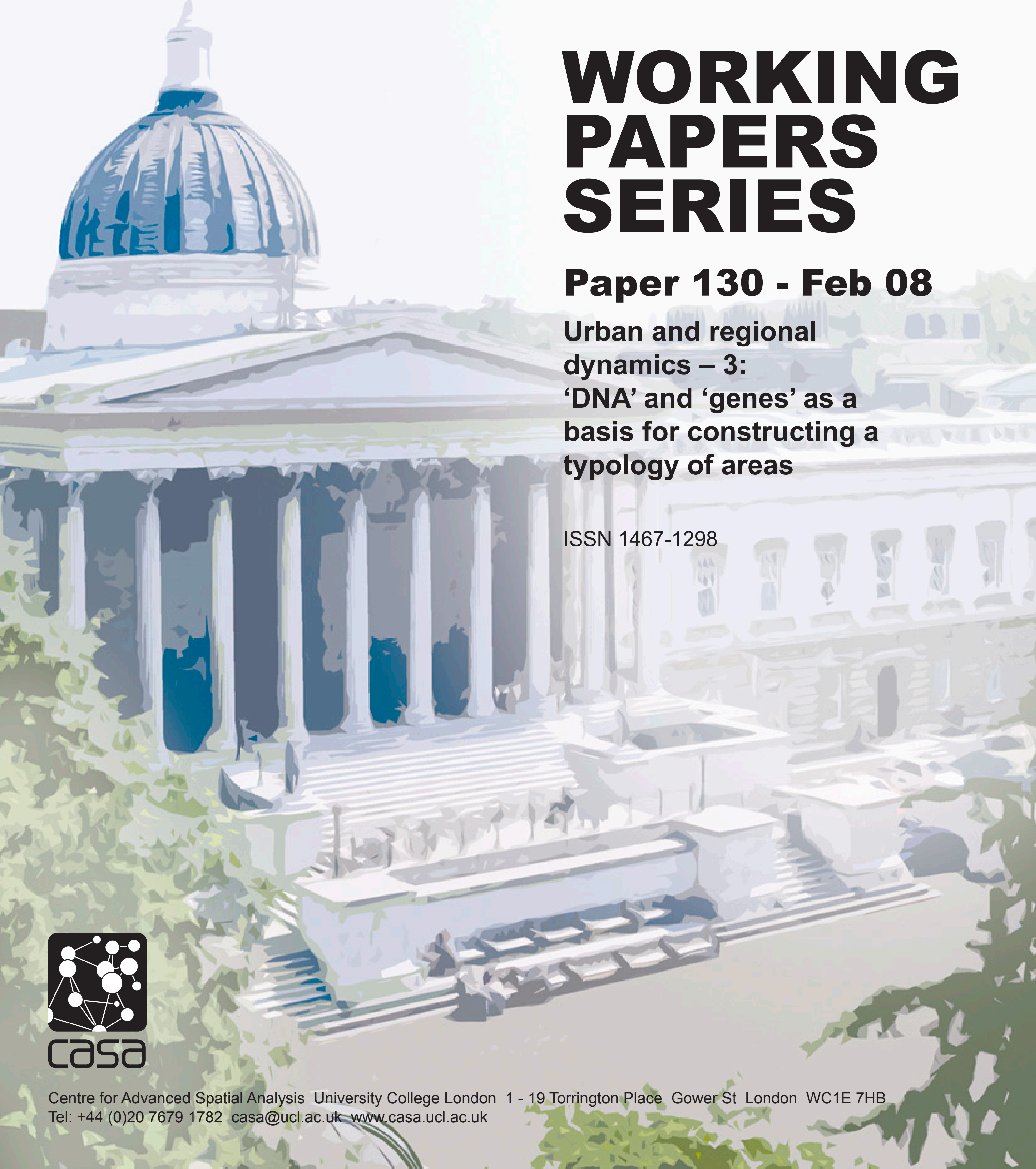
UCL CENTRE FOR ADVANCED SPATIAL ANALYSIS



WORKING PAPERS SERIES

Paper 130 - Feb 08

Urban and regional dynamics – 3: ‘DNA’ and ‘genes’ as a basis for constructing a typology of areas

ISSN 1467-1298

Centre for Advanced Spatial Analysis University College London 1 - 19 Torrington Place Gower St London WC1E 7HB
Tel: +44 (0)20 7679 1782 casa@ucl.ac.uk www.casa.ucl.ac.uk

Urban and regional dynamics – 3: ‘DNA’ and ‘genes’ as a basis for constructing a typology of areas

Alan Wilson
Centre for Advanced Spatial Analysis, University College London, WC1E 7HB

22 February 2008

1. Introduction.

This is the third paper in a series on urban and regional dynamics. They are intended to lay the basis for a future research programme. In the first paper, Wilson (2008-A), a core model was presented which can be used to demonstrate the consequences of interdependencies in the evolution of urban and regional systems. In the second, Wilson (2008-B), an hierarchical model was developed. In this paper, the argument is taken a stage further and the mechanisms or urban system evolution are further exposed. The exposition in this paper is based on the model presented in the first paper and to avoid repetition, the detail of that is not repeated here.

2. System description

We deploy the system description developed in Wilson (2008-A). We make fundamental distinctions between *structure* at and between locations, which is relatively slowly changing, *activities* at a location, and *interactions* between locations. We take the following as the principal sectors:

- Population
- Housing
- Public services[1] such as schools and hospitals
- Retail services
- Capacity in the economy – buildings, offices, equipment etc
- Government[2]

Labour is generated by converting the population at a location into an active workforce. We then make a distinction between the capacity of a sector at a location and the corresponding level of activity. (The population/workforce distinction is one example.)

---

[1] ‘Public services’ is an approximate description, used for convenience for sectors such as education and health in which the bulk of the provision is government funded

[2] We will assume an aspatial ‘pool’ into which taxes flow providing the funds for government spending

The principal spatial interactions are:

- The journey to work
- Population to housing
- Population to public services
- Population to retail centres
- Business to public services
- Business to retail services
- Business to business
- Government (spending) to population, public services, retail services and the economy

To achieve subdivisions are introduced – for example, population by income, retail facilities by type and so on. This leads to the following definitions of key variables:

- P_i^m, the number of type m people in zone i
- H_i^k, the number of type k houses in zone i
- V_j^n, the capacity of type-n consumer services in j; F_j^{nm}, the take-up (on a suitable measure – school places e.g.) by m-type people
- W_j^n, the capacity of type-n retail facilities in j; D_j^{nm}, the take-up of these facilities by m-type people
- Q_j^n, the capacity of the nth sector in zone j; X_j^n, the product of goods in sector n in j
- G_j^n, government spend in n in j

We distinguish the public services sector (such as education and health) and the retail sector from the general set of economic sectors to facilitate model construction and policy interpretation later. That is, we are distinguishing the V and W sectors from the other X-sectors (and the population sector, P, in so far as it is being treated as a producer of labour). Transport infrastructure is handled through interaction costs and total budgets available to support transport for each of the interaction arrays.

The main interaction variables are[3]:

- Y_{ij}^{mVn}, Y_{ij}^{mWn}, Y_{ij}^{mQn}, the flows to work in sectors V, W and Q from population group m in zone i to sector n in zone j for V, W and Q respectively;
- N_{ij}^{mnk}, the allocation of type m people who work in sector n in j to type k houses in i;
- U_{ij}^{mn}, the flow of type m people in zone i to consumer services of type n in j;
- S_{ij}^{mn}, the flow to retail facilities of type n;
- J_{ij}^{mn}, the flow of goods sector m in i to the consumer services sectors n in j;

---

[3] We will note later the possibility of indices such as m and n themselves being lists

- K_{ij}^{mn}, the flow of goods from sector m in i to the retail services sectors n in j;
- M_{ij}^{mn}, the flow of goods from sector m in i to sector n in j.

We will also need to define variables that represent aggregates:

- E_j^m, total employment of a type (taken as a superscript m) at location j[4];
- I_i^m, total income of type m people, resident in i;
- T_j^n, ..., tax take from various sectors in j;
- G_j^m, government spending in m in j.

3. Evolution and development – 'genetics' and 'physiology'.

We showed in Wilson (2008-A) that a comprehensive model could be developed that in principle captured the main interdependencies in the system of interest. The flows are all estimated through spatial interaction modelling linked in a more complex way than usual through input-output accounts. A mechanism was given for the dynamics of the structural variables and we indicated that development and evolution could be complex and could involve a series of phase changes. In this paper, we wish to explore the possibility of characterising cities and investigating their potential for further development by considering the underlying structural variables to be the urban analogues of DNA.

The key question is: how do cities evolve and develop? [5] We can argue as follows. Take an arbitrary starting point in time, t=0. The structural variables can be taken as {P, H, V, W, Q, L, p, c, G}[6]. The interaction models represent the workings of the city in a steady state over some time period, say one year (but again, this is an arbitrary choice for illustration). These 'workings' represent the 'physiology' of the model. The structural variables can then be considered to represent the 'DNA' – the 'genetics' - and we can explore whether the elements of these can be combined into 'genes' as explanatory variables for different forms of evolution.

The development of the system is driven through changes in the structural variables - the slow dynamics. Consider now the following possibilities:

---

[4] This will be taken as a sub-index within the n-list

[5] Can we usefully distinguish these notions? In Biology, an organism develops on the basis of the template provided by its genes – built from its DNA. Evolution involves the mutation of these genes. There is some interesting recent research linking evolution and development: that there may be elements of the DNA which is redundant in one situation but which can be activated in response to environmental stimuli to accelerate the development programme more rapidly than could be achieved through mutation – the so-called evo-devo hypothesis. Once the organism is fully developed, its 'workings' are described by its physiology.

[6] Note that, for completeness, we have included the vector of prices, p, here. We have also added 'c', to denote the matrix $\{c_{ij}\}$ to represent transport connectivity. This is an indirect representation of the transport infrastructure and we will explore how to deepen this later.

- Taking {P, H, V, W, Q, L, p, c, G} to be the city's DNA fingerprint, we may then be able to find a way of developing such an array to characterise different kinds of city and their potentials for growth or decline. This may be done by building variables from the DNA that play the analogue of genes.
- In many instances, running the core, supposedly steady-state, model will show the city to be dysfunctional: the accounting equations cannot be satisfied, for example. We can then ask questions like: how do we modify {P, H, V, W, X, L, p, c, G} through investment to change this? This would be a form of genetic medicine.

In other words, we may be able to characterise a variety of 'successful' cities and others that are less so. We consider these questions in turn in the next two sections.

4. The genetic structure of cities.

4.1. Retail structure as an example.

Consider the retail model as a special case – building on the argument in Harris and Wilson (1978). Define, as usual, S_{ij} as the flow of spending power from residents of i to shops in j; let e_i be spending per head and P_i the population of i. W_j is a measure of the attractiveness of shops in j. α and β are parameters. The vector $\{W_j\}$ can then be taken as a representation of urban structure – the configuration of W_js. If many W_js are non-zero, then this represents a dispersed system. At the other extreme, if only one is non-zero, then that is a very centralised system. A spatial interaction model can be built for the flows on the same basis as the transport model. Then:

$$S_{ij} = A_i e_i P_i W_j \exp(-\beta c_{ij}) \qquad (1)$$

where

$$A_i = 1/ \Sigma_k W_k \exp(-\beta c_{ik}) \qquad (2)$$

to ensure that

$$\Sigma_j S_{ij} = e_i P_i \qquad (3)$$

with

$$D_j = \Sigma_i S_{ij} = \Sigma_i e_i P_i W_j \exp(-\beta c_{ij})/ \Sigma_k W_k \exp(-\beta c_{ik}) \qquad (4)$$

If Wj is taken, for simplicity, as a size measure, then we might seek to measure attractiveness by raising that to a power, α, and so replacing W_j by W_j^{α} We can replace W_j in these equations by W_j^{α} throughout. Note, then, that W_j^{α} can be written

$$W_j^{\alpha} = \exp(\alpha \log W_j) \qquad (5)$$

and the core equations can be written

$$S_{ij} = A_i e_i P_i \exp(\alpha \log W_j - \beta c_{ij}) \qquad (6)$$

where

$$A_i = 1/ \Sigma_k \exp(\alpha \log W_k - \beta c_{ik}) \qquad (7)$$

This shows explicitly that $\log W_j$ can be taken as a measure of the utility of an individual going to a shopping centre of size Wj but at a transport cost, or disutility, represented by c_{ik}.

We now add a suitable hypothesis for representing the dynamics:

$$dW_j/dt = \varepsilon\,(D_j - KW_j) \qquad (8)$$

where K is a constant such that KW_j can be taken as the (notional) cost of running the shopping centre in j. This equation then says that if the centre if profitable, it grows; if not, it declines. The parameter ε determines the speed of response to these signals.

The equilibrium position is given by

$$D_j = KW_j \qquad (9)$$

which can be written out in full as

$$\Sigma i\{e_iP_iW_j\exp(-\beta c_{ij})/\Sigma_k W_k \exp(-\beta c_{ik})\} = KW_j \qquad (10)$$

and these are nonlinear equations in the $\{W_j\}$.

It is possible to characterise the kinds of configurations that can arise for different regions of α and β space: for larger α and lower β, there are a smaller number of larger centres; and vice versa[7]. This can be interpreted to an extent for a particular zone, say j, by fixing all the W_k, k not equal to j. A key challenge is to solve this problem with all the W_js varying simultaneously. There are many procedures for solving the equations (10) iteratively but we constantly need to bear in mind the sensitivity to the initial conditions – the path dependence – and the possibility of multiple solutions.

In this case, the structural variables are [{ei}, {Pi}, {Wj}, {cij}, α, β]. The β parameter can be considered to arise from the total resource available for making retail trips, C say (and this could be disaggregated by zone 1) and so C would be an alternative representation of the transport structural variable. α is about preferences for the use of larger centres relative to smaller ones.

In the terms of our analysis, the $\{S_{ij}\}$ array represents the physiology and the structural array, $\{W_j\}$ is modified slowly over time in a manner determined by equation (8). We know from previous work (Clarke and Wilson, 1985) that if parameters such as α and β change, there will be structural changes, and these can include sudden phase transitions. But what we can also see is that any changes in the other structural variables can also bring about changes in the $\{W_j\}$, and, again, possibly phase transitions. This shows the importance of interdependence in the more comprehensive version of the model. For example, the $\{e_i\}$ terms will be dependent on income flows from work and the proportion of income that can

---

[7] Clarke and Wilson (1985)

be spent on retail goods (relative, say, to housing). (We should be able to determine C from these income flows and associated models and hence calculate a parameter like β endogenously.)

We also know from previous work that while we can show through simulation that there are characteristic structures associated with particular values of the α and β parameters, the nearest we can get to an analytical account involves one W_j assuming that $\{W_k, k \neq j\}$ are fixed.

Thus, in the general case, we have to deal with a very large number of structural variables in a context where we have understood analytical intractability for a simpler case. Can we use a genetic analysis to make progress with this challenge?

4.2. Structure in the comprehensive model.

To take the argument further, we need to identify the key drivers. We do this in the first instance through Figure 1 below. There are connections to the external world through imports and exports. We can now note that additional key drivers are migration, investment by organisations and investment by Government agencies (local, regional or national) – the last already anticipated through the introduction of the G-arrays. These driving flows are shown on the figure. The genes then come into play by considering the drivers of migration, economic investment and Government investment. Migration might be driven by housing and work availability and the quality of consumer and retail services and the quality of the social environment; economic investment by profitability which in turn will depend on the availability of skilled labour, other inputs and markets. Both will be influenced by transport connectivities. The various submodels will have to incorporate some preferences that are not economic – for example, for (relative) social differentiation in residential areas (and correspondingly in schools). Government agencies will be obliged to supply what are agreed to be public goods – such as schools and hospitals – and to make these sufficiently attractive to lever migration and economic investment in appropriate directions. In particular, in terms of the model-based foundations, government agencies need to find the bifurcation points at which sudden changes for the better can be achieved.

An analysis of the physiology then reveals a high degree of interdependence. (An interesting question is whether this can be measured in some way?) Much of the interdependence is rooted in the input-output model. What we have seen already is that the set of drivers are spread around that model: it is no longer something simple like final demand that is determining. It is this insight, and the known path dependence of dynamical systems, that makes the notion of a path through a sequence of ‘initial conditions’ an important idea.

At the starting point, households will be located, workers will be employed, deriving income which is channelled back to households. Organisations will be profitable, or not; or funded by the Government, or not. What the different interaction models represent is an intersection of supply and demand but which, in each case, is mediated by other models. Ideally, the interdependence needs to be made explicit through linking submodels. For example, consider household income derived from employment. Ideally we need a model which shows how that is allocated between different areas of expenditure and what ends up in, say, the retailing 'pot' is then part of a demand function for retail goods. But that cannot be separated from earned income and the way it is divided. Typical, even comprehensive, urban models do not tackle this issue.

There is a further issue: the analytical challenge so far has been stated mainly in terms of money flows. There are also time budgets to be handled.[8]

Assume then that the population and organisations (again internal and external in each case) and government agencies each respond to signals – indicators – that give them measures of the current situation. Such indicators can be calculated from the model outputs and, in effect, they are the kinds of things that we have already recorded in the core model equations (in Wilson, 2008-A) for the slow dynamics: retailers responding to profits and losses for example. The particularly interesting case is that of government agencies as noted above. These are then the mechanisms for representing the evolution of the structural variables.

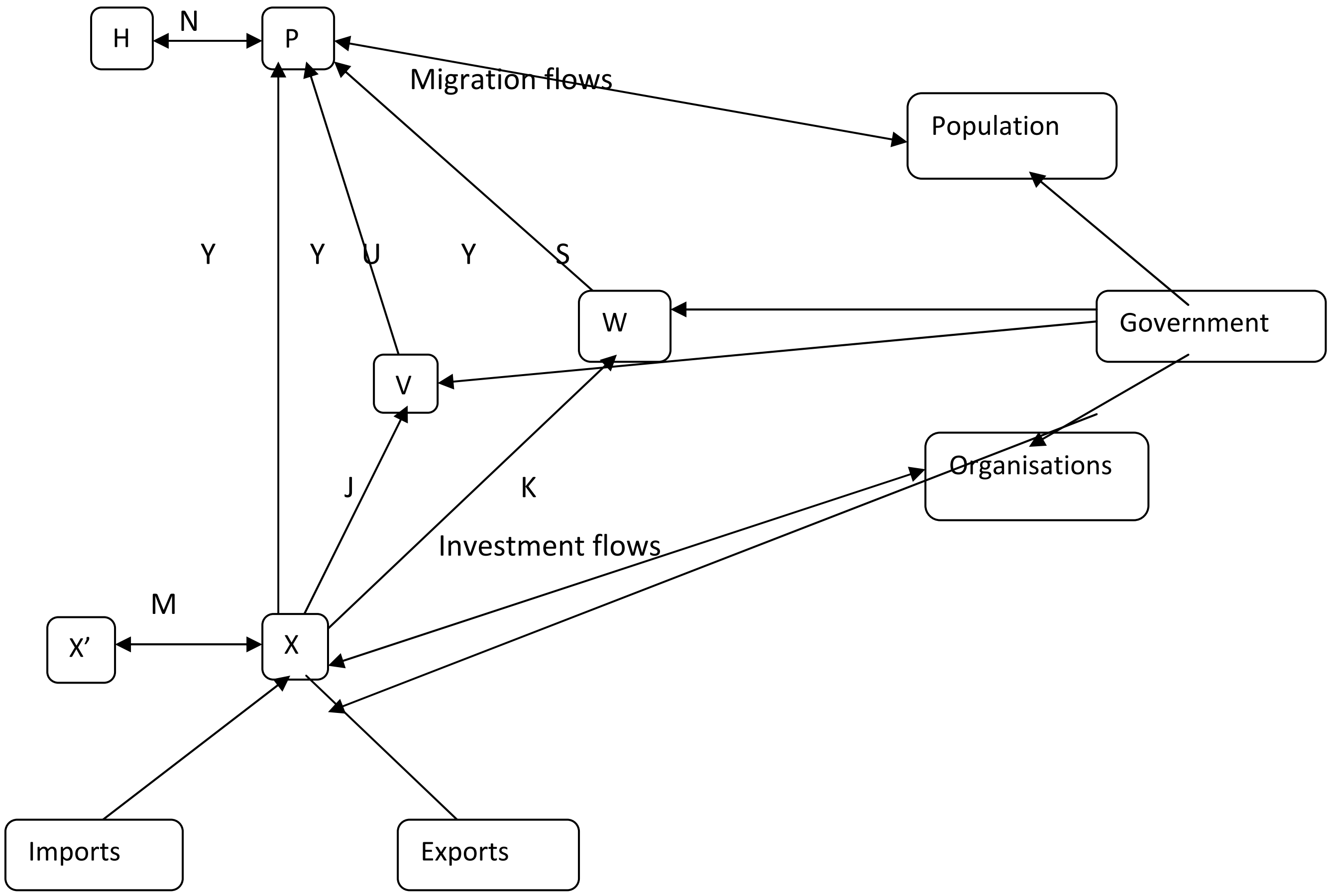


Figure 2.

---

[8] Cf. Becker (1965)

Can we summarise this complexity by identifying ‘genes’? Candidate genes might include:

- Proportions of the population by age group, skill levels, occupations and income
- Housing, by price and quality[9]
- Access to consumer services
- Access to retail
- Economic indicators:

    - ‘Variety’
    - Clustering
    - Proportions in different levels: resources-agriculture-manufacturing-service-post service
    - Imports and exports; balance of payments
    - Average income from employment
    - Income from other sources

- Transport connectivity – measured with ‘generalised costs’
- Land use densities

These could be presented at an aggregate level as implied above, or, more ambitiously, they could be presented at a zonal level – and then the gene structure would be the distribution of these indicators across zones. Can we then address the following kind of question? How does {P, H, V, W, X, L. p, c, G} limit the future development of the city? Technically, this is a question of the regions of phase space which are accessible on the basis of the structure at time zero and any exogenous adjustments that areas made.

We also have an opportunity to analyse at the zonal *or* sectoral levels. Indicators can be derived from the gene stock or from the physiology – the interaction arrays – and some are combinations of these. It is interesting to see these perspectives by arranging the input-output table in appropriate ways. It is possible to make *either* the outer structure of the table zonal – as (i, j) – *or* sectoral- as (m, n). Figures 5 and 6 show these perspectives.

---

[9] ‘Quality of environment’ to include crime levels etc

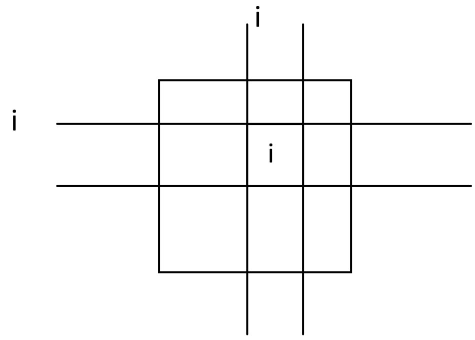


Figure 2. The (i, i) element with zones as outer structure

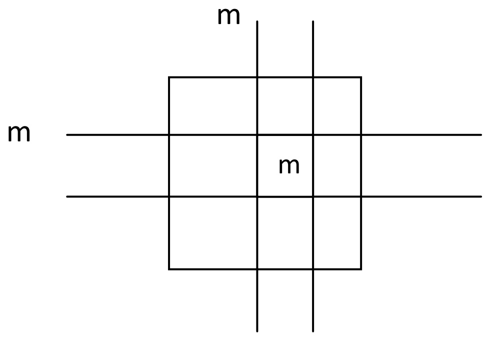


Figure 3. The (m, m) element, with sectors as he outer structure

The rows and columns that go through the (i, i) segment of Figure 2 give a full account of the workings of the economy of zone i. Those that go through the (m, m) segment of Figure 3 give the workings of a particular sector. Through these formulations, we can calculate the balance of payments for a zone or for a sector. This kind of analysis could be used to characterise types of zone and to identify problem areas. We could write down the gene signatures, for example, for

- A high-class residential area
- A low-income residential area
- An industrial area

and so on.

5. Next steps.

There are two main potential areas of application. We have already indicated that this analysis should be a means of building typologies, both of towns and cities and of areas within them; and also of urban economic structures. This leads to another interesting idea. Since geodemographics is also a basis for classification, perhaps the two notions could be linked: geodemographic analysis could be seen as the X-ray crystallography of urban gene identification!

The second area is in planning and policy development. Can we identify desired future states and (a) explore their accessibility and (b) chart paths towards them by identifying the genes that can be modified! Urban genetic medicine!

References

Becker, G. S. (1965) A theory of the allocation of time, *Economic Journal, 754*, pp. 493-517.

Clarke, M. and Wilson, A. G. (1985) The dynamics of urban spatial structure: the progress of a research programme, *Transactions, Institute of British Geographers*, *10*, pp. 427-451.

Harris, B. and Wilson A. G. (1978) Equilibrium values and dynamics of attractiveness terms in production-constrained spatial-interaction models, *Environment and Planning, A, 10*, pp 371-88, 1978.

Wilson, A. G. (2008-A) A core model for the analysis of urban dynamics, Working Paper 128, Centre for Advanced Spatial Analysis, University College London.

Wilson, A. G. (2008-B) Urban and regional dynamics – 2: an hierarchical model of interacting regions, Working Paper 129, Centre for Advanced Spatial Analysis, University College London.

AGWArts/DNAModel170208