Chapter 13
Finite Automata

In Paelinck (2002), attention was drawn to a special algebra—called a min-algebra—that might rule quite a few spatial econometrics specifications; hereafter, applications of this idea are be presented in the form of finite automata.

A finite automaton specification (for a formal definition, see Linz, 1996, p. 2) can be viewed as an "if"-specification; in symbolic terms

$$y: \text{if}(\alpha x_i + \beta < \gamma z_i + \delta; \alpha x_i + \beta; \gamma zi + \delta), \tag{13.1}$$

which reads as follows: if $\alpha x_i + \beta < \gamma z_i + \delta$, then $\alpha x_i + \beta$, else $\gamma z_i + \delta$.

Hereafter, a two-region example of a dynamic finite automaton is outlined.

13.1 A Finite Automaton Bi-regional Dynamic Model

Consider the following numerical case, for which the variables are defined as follows

x_{it}: joint location factors (i = 1, 2);

y_{it}: production levels (i = 1, 2).

The model is specified as follows

$$x_{1,t+1}: \text{if}(y_{1t}/x_{1t} < y_{2t}/x_{2t}; 3; 1), \tag{13.2}$$

$$y_{1,t+1}: \text{if}(y_{1t}/x_{1t} < y_{2t}/x_{2t}; 1.05\ y_{1t}; 1.02\ y_{1t}), \tag{13.3}$$

$$x_{2,t+1}: \text{if}(y_{2t}/x_{2t} < y_{1t}/x_{1t}; 1.5; 1.2) \tag{13.4}$$

$$y_{2,t+1}: \text{if}(y_{2t}/x_{2t} < y_{1t}/x_{1t}; 1.05\ y_{2t}; 1.02\ y_{2t}) \tag{13.5}$$

The logic of the model is as follows: as long as one region has its joint location factors "undercharged" compared to those of the other one—in the sense that they are still attractive for further activity locations—, their joint value stays at a higher level, the growth rate there also being higher, and vice versa.

D.A. Griffith, J.H.P. Paelinck, *Non-standard Spatial Statistics and Spatial Econometrics*, Advances in Geographic Information Science 1,
DOI 10.1007/978-3-642-16043-1_13, © Springer-Verlag Berlin Heidelberg 2011

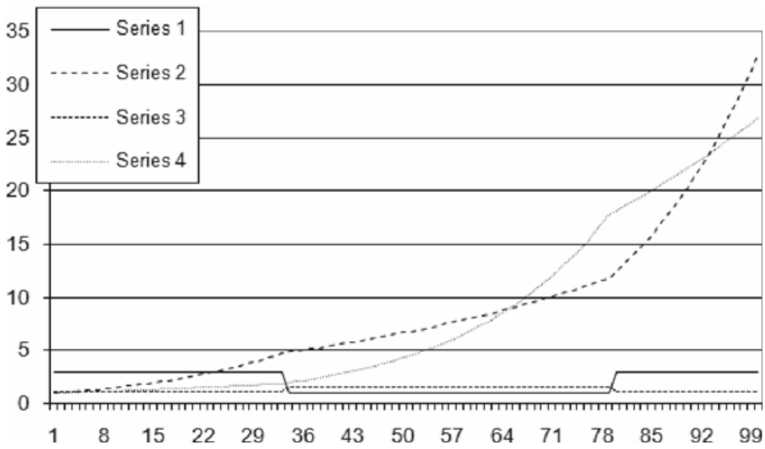

Fig. 13.1 Simulation of a dynamic finite automaton

With the initializing vector

$$(x_{10}, y_{10}, x_{20}, y_{20}) = (3, 1, 1.2, 1) \tag{13.6}$$

the resulting simulation is that presented by Fig. 13.1.

This figure portrays a dynamics in which regions at times lose their competitive edge in attracting activities.

The following problem is that of estimating such a model, possibly using an *if*-condition. Eight parameters have to be computed, to wit for each region the two levels of the location factors, x_i and $x_i{}^*$, and the growth rates, ρ_i and $\rho_i{}^*$, the lower levels being denoted by asterisks.

An example of an *if*-constraint is the following

$$\text{if}(y_{1t}/\xi_{1t} < y_{2t}/\xi_{2t}; 1; 0) \tag{13.7}$$

where ξ_{it} is one of the values of x_{it}, $x_{it}{}^*$ determined according to conditions (13.2) and (13.4). The specification of conditions (13.7) implies the introduction of a norm for the x_{it}, $x_{it}{}^*$ values. Expression (13.7) was combined with the following term appearing in the objective function φ—sum of some function of those terms, as will be seen subsequently—for minimization purposes

$$r_{1t} - \lambda_{1t}\rho_1 - (1 - \lambda_{1t})\rho_1{}^*, \tag{13.8}$$

where, for region *1* and time *t*, r_{1t} is the observed growth rate, and λ_{1t} is a binary variable, the value of which is determined by condition (13.7) for the previous period, which reproduces condition (13.3) above.

Table 13.1 presents the data used; mark the split of the r_{it}-values in four opposite groups (e.g., for $t=1,\ldots 4$, $r_{1t}<r_{2t}$, then the reverse occurs).

Table 13.1 Data for estimating finite automaton parameters

t	r_{1t}	y_{1t}	r_{2t}	y_{2t}
1	0.05	1.00	0.02	1.00
2	0.06	1.05	0.01	1.02
3	0.04	1.11	0.03	1.03
4	0.01	1.14	0.06	1.06
5	0.03	1.15	0.05	1.12
6	0.02	1.18	0.04	1.18
7	0.06	1.20	0.01	1.22
8	0.05	1.27	0.03	1.23
9	0.04	1.33	0.02	1.27
10	0.02	1.38	0.04	1.30
11	0.03	1.41	0.06	1.35
12	–	1.45	–	1.43

Table 13.2 Parameter values of the estimated finite automaton

ρ_1	ρ_1^*	x_1	x_1^*	ρ_2	ρ_2^*	x_2	x_2^*	φ
0.05	0.022	0.8911	1.0022	0.05	0.02	1.0526	1.0541	0.00148

Table 13.2 presents the estimated parameters, φ being the obtained minimum of the objective function, in which expression (13.8) was squared; use was made of a reduced gradient method (Fylstrom et al., 1998)

The split into four periods mentioned earlier is correctly pictured by the binary variables λ_{it}; to improve the estimates of the parameters, the exercise was repeated keeping the binary variables stable and endogenizing the y_{it}-values, as is explained in the ensuing discussion.

Figure 13.2 portrays results of a simulation of the estimated model over the first twelve time periods. The four-period pattern just mentioned is fully reproduced, but the values, especially in region 1, deviate from the observed ones.

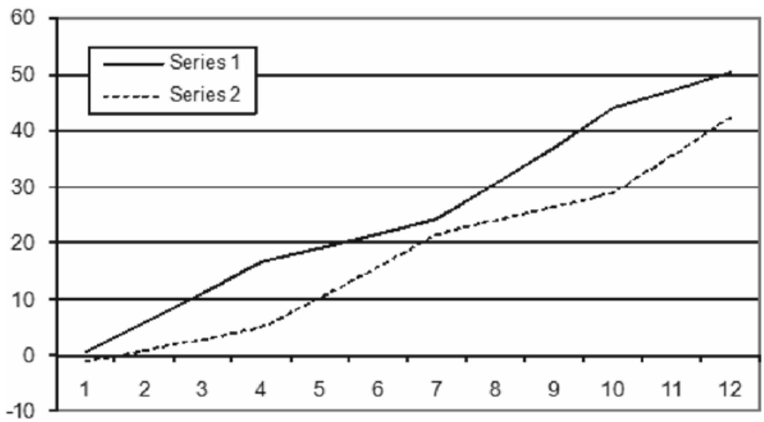

Fig. 13.2 Simulation of the estimated finite automaton

A possible alternative to objective function (13.8) is to reformulate the specification in terms of the production levels y_{it}, and use SDLS (see Sect. 11.1.3) with the computation of an optimal starting point for an endogenous simulation (Paelinck, 1990b). This method minimizes some difference between the observed values and the *endogenously simulated* values of the y_{it}s, those simulated values being at the same time generated within the estimating procedure. This approach is described next.

Of note is that condition (13.7) is expressed in terms of strict inequalities, which implies that if an equality is present, the higher x_i and ρ_i levels automatically are assigned to the other region. An alternative is furnished by

$$(\lambda_{1t} - \omega)(y_{1t}/\xi_{1t} - y_{2t}/\xi_{2t}) \leq 0, \tag{13.9}$$

with $0 < \omega < 1$, which adds another degree of freedom (to be taken up by the estimation procedure) when the second factor in expression (13.9) is zero. This procedure has been applied to the y_{it} data of Table 13.1, again with a quadratic objective function, and subsequently keeping binary variables—which correctly split the overall period into four components—stable.

Table 13.3 presents the obtained parameters; φ_ψ is the value of the objective function, φ_ρ that of φ in Table 13.2.

Figure 13.3 presents the simulated values over the first 12 time periods, again showing that the four characteristic groups mentioned earlier are still correctly

Table 13.3 Parameter values introducing conditions (13.9)

ρ_1	$\rho_1{}^*$	x_1	$x_1{}^*$	ρ_2	$\rho_2{}^*$	x_2	$x_2{}^*$	φ_ψ	φ_ρ
0.0483	0.0215	0.9704	1.0134	0.0442	0.0196	.1.0022	1.0139	15.3791	0.0015

Fig. 13.3 Second simulation of the estimated finite automaton, parameters from Table 13.3

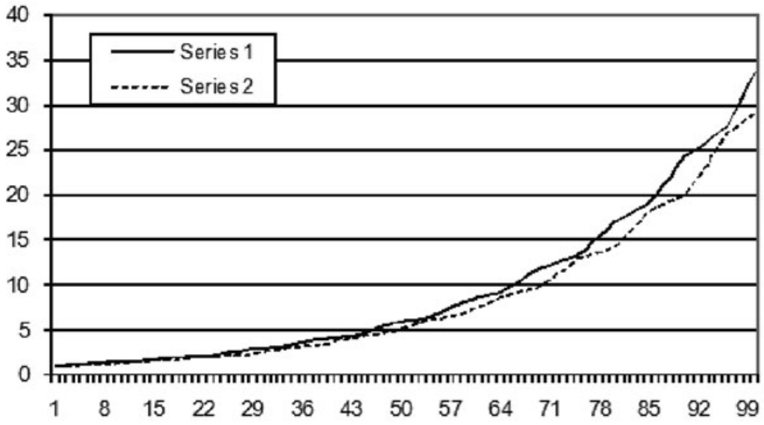

Fig. 13.4 Figure 13.3 results extended over 100 periods, parameters from Table 13.3

represented, as well as that the fit is considerably improved; Fig. 13.4 extends the projection.

To conclude, in the estimation procedure, dependence on initial conditions and multiple solutions do seem to be crucial. The latter especially holds for the x_{it} and $x_{it}{}^*$ values, which is due to the multiplicities that can satisfy conditions like (13.7) or (13.9). These points are left for further investigation.

13.2 An Empirical Application

To subject the model developed above to an empirical test in a well-documented case, gross regional product numbers for the Netherlands have been investigated. They were divided in two macro-regional sets, one for the western provinces (Noord-Holland, Zuid-Holland, Utrecht, the so-called "Rimcity"), the other one comprising the data for the remaining provinces (source: CBS, 2003; 10^9 EUROS). Given the low inflation rate (in the order of 1% annually) no price correction was applied. Table 13.4 presents the data.

The data were analyzed with the methodology discussed in Sect. 13.1, using a quadratic objective function for the product levels, and, alternatively, a binary and a fuzzy version. Table 13.5 summarizes the results. The λ_t parameters are generated by constraints (13.9), and are relative to the Rimcity, their values for the other provinces being the complements to *1*.

The curious finding, at first sight, is the respective values of the growth rates for the non-Rimcity provinces: whatever the state of their location factors' attractiveness, they follow the ups and downs of the Rimcity growth rates. This result is completely in line with the Rimcity indeed being the "motor" of the Dutch economy (Paelinck, 1973, pp. 25–40, especially pp. 37–40), imposing its evolutionary rhythm on the other regions, which corresponds to a sort of *Fick* diffusion in thermodynamics (Braun, 1975, pp, 645 a.f.; Philibert, 2005, pp. 2–3).

Table 13.4 Gross regional product data for two macro-regions in the Netherlands

Years	Rimcity	Other provinces
1998	111.0	103.7
1990	116.5	110.0
1991	122.4	116.6
1992	127.3	121.6
1993	131.3	125.5
1994	136.9	130.3
1995	142.8	136.1
1996	147.3	141.3
1997	156.5	147.3
1998	166.8	156.5
1999	176.5	164.7
2000	189.8	177.1

Table 13.5 Parameter values of the model defined by Eqs. (13.2), (13.3), (13.4), and (13.5) with data from Table 13.4

Parameters	Values, binary case	Values, fuzzy case
ρ_1	0.0638	0.0794
$\rho_1{}^*$	0.0385	0.0386
x_1	1,8797	1.1099
$x_1{}^*$	00.9001	0.9675
ρ_2	0.0409	0.0394
$\rho_2{}^*$	0.0538	0.0705
x_2	1.4010	0.9278
$x_2{}^*$	0.6084	0.9564
λ_1	1	0.5976
λ_2	0	0.1
λ_3	0	0.1
λ_4	0	0
λ_5	0	0
λ_6	0	0
λ_7	0	0
λ_8	1	0.4112
λ_9	1	0.6291
λ_{10}	1	0.5021
λ_{11}	1	1
φ_y	16.3326	9.4769

From a technical point of view, the constraint parameter estimates are consistent with one another, which hints at the adequacy of the binary estimation; the lower φ_y-value is in line with the *Le Châtelier*-principle (Samuelson, 1955, pp. 36 a.f.).

Figures 13.5 and 13.6 further down portray the binary and fuzzy cases, respectively.

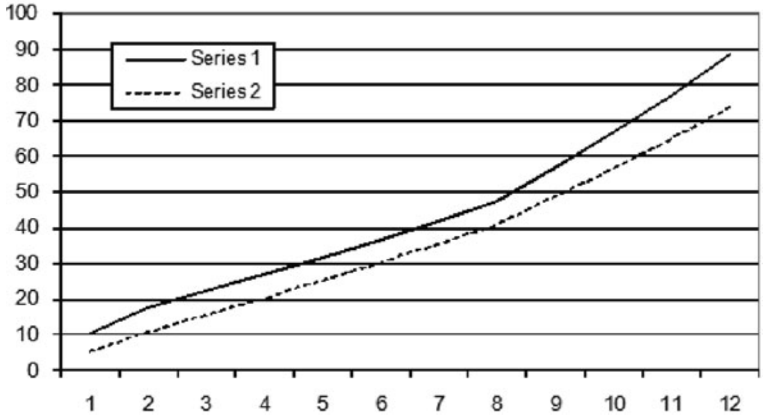

Fig. 13.5 Dutch model, binary case

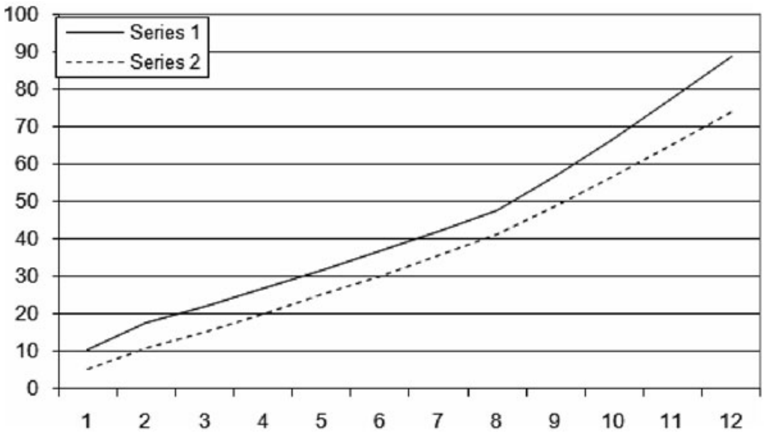

Fig. 13.6 Dutch model, fuzzy case

13.3 Conclusion

It has been shown that finite automata models can be an appropriate specification for multiregional models; the reason is that the development logic of a multiregional system needs for its modeling a special algebra, and a corresponding set-up of the corresponding estimation procedure.

Three more points should still be made.

First, starting an exercise in spatial econometric modeling with a complexity analysis of the data is advisable; obvious candidates for simple exogenous variables are their space-time coordinates. An example can be found in Getis and Paelinck (2004), where regional product data for the Netherlands were analyzed. A model specification implies the choice of exogenous variables, and possibly endogenous ones—in interdependent models—or lagged endogenous variables—in dynamic models—, therefore, they too should be implied in a complexity approach.

Second, the specifications presented can readily be generalized to three or more alternatives (regions, test specifications). For the finite automaton version, for example, the following specification shows how *and* and *or* statements can be added

$$y_i : \text{if}\,((cz_i + d < ax_i + b) \ \text{and} \ (eu_i + f < ax_i + b)\,;\ (cz_i + d) \ \text{or} \ (eu_i + f)\,;\ ax_i + b) \tag{13.10}$$

Finally, one cannot escape from the fact that hypothesis testing is fundamentally theory-laden (Aznar Grasa, 1989, p. 10). Accordingly, theoretical spatial economics will remain an indispensable guide to spatial econometric modeling.