Endogenizing the rise and fall of urban subcenters via discrete programming models

P Gordon
School of Urban and Regional Planning and Department of Economics, University of Southern California, Los Angeles, CA 90089-0042, USA

J E Moore II
School of Urban and Regional Planning and Department of Civil Engineering, University of Southern California, Los Angeles, CA 90089-0042, USA
Received 19 August 1988

Abstract. Models that represent the polycentric and dispersed nature of modern cities should be able to account for the rise and fall of subcenters. Based on a review of the programming models applied to urban analysis, five properties are suggested that an adequate model should include. It should: 1 confront the simultaneity between markets for land and transportation services; 2 accommodate the reality of cities as places where externalities and common properties abound; 3 emulate the intertemporal albeit bounded nature of planning and decisionmaking; 4 fully exploit principles of economic theory; and, 5 offer computability. We develop a discrete programming model with these five properties, comparing its capabilities with those of previous approaches.

1 Introduction
The conventional monocentric model of urban economics has generated a number of useful insights (Muth, 1985). Yet, the distinctly polycentric and dispersed nature of modern urban settlements provides argument for a theory that recognizes and endogenizes the rise and fall of urban subcenters. In a recent address to the Western Regional Science Association, Richardson (1988) registered pessimism concerning the ability of conventional urban economics to solve the problem. He endorsed, instead, approaches of discrete space mathematical programming. In this paper our purpose is to formulate a programming model that simulates the spatial evolution of modern cities. The key analytical questions are: why do subcenters emerge, and what are the conditions under which subcenters are likely to be formed? The key methodological question is: can a tractable model of these phenomena be designed?

2 Literature review
Among spatial planning models, the transportation problem of linear programming has received the widest attention and elaboration from planners (Greenberg, 1978). This is reasonable because transportation linear programs are subject to efficient solution, exhibit unimodularity, and frequently serve as templates for more complicated models of travel choice.

These more complicated models have recently been surveyed by Boyce et al (1988). The intellectual advances cited in their review call attention to the fact that theorists have successfully unified models of travel, mode, and route choice. All of the models cited have as their structural core a path-flow version of the transportation problem. Boyce et al identify the work of Beckmann et al (1956) as the seminal formulation of this class.

\[
\text{Minimize } Z = \sum_{a} \int_{0}^{r_a} d_a(w)dw - \sum_{m} \sum_{a} \int_{0}^{f_{ma}} g_{ma}^{-1}(u)du ,
\]  

(1)
subject to
\[ \sum_k f_{kmn} = f_{mn}, \quad m \in M, \ n \in N, \]  
\[ \sum_k \sum_m \sum_n \delta_{akmn} f_{kmn} = f_a, \quad a \in A, \]
and
\[ f_{kmn} \geq 0, \quad k \in K, \ m \in M, \ n \in N, \] (4)

where \( d_a(.) \) is an increasing cost function of the traffic volume on link \( a; \ g_m^{-1}(.) \) is an inverse 'demand for travel' function of the equilibrium cost of travel; \( f_a \) is the endogenous flow on link \( a; \ f_{kmn} \) is the endogenous flow on path \( k \) from origin \( m \) to destination \( n; \ f_{mn} \) is the total endogenous flow from origin \( m \) to destination \( n; \) and \( \delta_{akmn} \) is a binary indicator equal to one if link \( a \) is on path \( k \) from origin \( m \) to destination \( n \), and equal to zero otherwise. The model is an endogenous path flow formulation that minimizes user equilibrium\(^{(1)}\) transportation costs given elastic demand for travel.

Though elegant, this formulation includes a number of restrictive assumptions that have been perpetuated in the literature it generated. The derivative formulations cited by Boyce et al (1988):

1 are entirely static;
2 include fixed O-D values (though not necessarily fixed interzonal flows), implying that the activity system is fixed;
3 identify congestion-dependent, steady-state network equilibria;
4 abstractly (and imprecisely) characterize imperfections in decisionmaking via entropy, information, or dispersion coefficients; and
5 treat land-uses as infinitely divisible, though this shortcoming is washed out by the more restrictive assumption of a fixed activity system.

Concurrent with the investigation of network equilibrium, an analogous stream of independent research produced correspondingly restricted models of activity location. In the interest of tractability, these land-use models either ignored transportation entirely or else exploited assumptions of fixed commodity flows and link costs (Alonso, 1964; Herbert and Stevens, 1960; Lowry, 1964). During the past decade it has become understood that both approaches to urban model building, though clearly complementary, are equally deficient. Congestion costs are determined within the process of land-use allocation and, in turn, affect this allocation. It is intellectually inconsistent to accept either transportation costs or land-uses as fixed. Neither is fixed, but modeling the redistributive effects of activity and transportation systems has generally been regarded as a difficult problem.

It is only recently that work confronting the simultaneity of land and transportation markets has been undertaken. Many of these combined models are integrated formulations, that is, separate land-use and transportation models that are iteratively linked. Typical examples of these efforts include Berechman (1981), Boyce (1978), Boyce and Southworth (1979), and Hutchinson (1975).

In a set of parallel developments, the simultaneity of land-use and transportation has also been treated via the formulation of unified, general equilibrium models. Though there are many seminal representatives of this class, the foremost example is probably Mills's (1972) linear programming model of land-use and transportation. This 'efficient resource allocation model' optimizes spatially defined production activities and transportation network investments, subject to constraints on minimum export requirements and the location of export facilities. Mills's basic formulation

\(^{(1)}\) See Wardrop, 1952.
has been extended by a number of other scholars, including Hartwick and Hartwick (1974; 1975), Kim (1978; 1979), Moore (1986), and Moore and Wiggins (1988). All of these subsequent efforts:

1 are entirely static, with the exception of one formulation by Moore and Wiggins (1989);
2 endogenize all location decisions except those pertaining to the location of export nodes;
3 either approximate steady-state network equilibria, or else ignore congestion completely;
4 assume perfect information, with the exception of recent formulations by Kim (1983; 1986); and
5 treat land-uses as infinitely divisible, with the exception of transportation facility investments in one formulation by Mills (1975).

A separate line of inquiry that exploits mathematical programming approaches can be traced to a landmark article by Koopmans and Beckmann (1957) investigating the properties of the 'assignment problem'. Their unified formulation treats indivisible, fixed-interaction activities ('plants' in their text) to be located at discrete sites.

Maximize $Z = \sum_i \sum_m a_{im} x_{im} - \sum_i \sum_m \sum_j f_{ij} d_{mn} x_{im} x_{jn}$,

subject to

$\sum_i x_{im} = 1, \quad m \in M$, \hspace{1cm} (6)

$\sum_m x_{im} = 1, \quad i \in I$, \hspace{1cm} (7)

$x_{im} \geq 0, \quad i \in I, \quad m \in M$, \hspace{1cm} (8)

where $a_{im}$ is the known profitability that accrues to activity $i$ if it locates at site $m$, $d_{mn}$ is the unit cost of interaction between an activity located at site $m$ and an activity at site $n$, and $f_{ij}$ is the annual interactivity shipment requirement. $x$ is the vector of solution variables. The constraints assure that only one indivisible activity is allocated to any site, and vice versa.

Koopmans and Beckmann showed that, if the exogenous traffic intensities are suppressed, the feasible region for this problem is a permutation matrix that constrains the variables to zero or one, as required by the logic of the model. This simplest version of the Koopmans–Beckmann problem is an assignment linear program that provides access to duality conditions. The convenient economic interpretation of the dual variables as 'plant rents' and 'site rents' is important to Koopmans and Beckmann's analysis.

In the more general case, Koopmans and Beckmann's main result can easily be summarized: if plant operators' utilities are dependent on the location (nearthess) of other activities, then the optimal solution (assignments of plants to sites) is at best unstable and at worst indeterminable. The quadratic program that is used to solve this problem was shown to yield prices (dual variables) that would not sustain any set of assignments. At least two locators were always left with an incentive to exchange positions. Koopmans and Beckmann's proof has intuitive appeal as it is likely that the full value of any site cannot really be judged without knowledge of the other activity locations. The cost of interaction with these other sites is a major determinant of the worth of any location.

Koopmans and Beckmann characterized this result by noting that their objective function only accounted for what they called 'semi-net' revenues, and that under
their assumptions the interaction costs of an individual firm cannot be known until all activities have been located. Many authors have commented on this outcome, some showing that there may be stable solutions, usually in the case of sufficiently limited activity interactions (Goldstein and Moses, 1975; Hartwick, 1974).

The question of price-sustainable location assignments aside, the Koopmans–Beckmann problem suffers from two restrictive assumptions: that traffic intensities and link costs are both exogenous. Thus, the frequency of shipments between activities is made independent of the proximity of the activities to one another. Unfortunately, these assumptions have proved notoriously difficult to shed.

Hopkins (1977) and Los (1979) used heuristic methods to solve a Koopmans–Beckmann problem that included endogenous traffic intensities. Subsequently, they extended their formulations (Hopkins and Los, 1979; Los, 1978) by endogenizing investments in the transportation network. In all cases, the resulting formulations were difficult convex or discrete convex programming problems that could not be solved optimally. The authors relied on iterative procedures reminiscent of conventional urban transportation planning efforts.

In summary, the Koopmans–Beckmann problem and its cited extensions:
1. are entirely static;
2. endogenize all location decisions;
3. either ignore congestion completely, or treat it very indirectly;
4. assume perfect information, but generally require an heuristic solution; and
5. (appropriately) treat land-uses as discrete, indivisible activities, but do so at the expense of price-sustainable results.

Gordon and Wingo (1981) showed that since simultaneous locational choice is what decimates the price-sustainability of the optimal solution to the Koopmans–Beckmann problem, it is useful to study the nature of the information problems that would arise if locators entered the market and chose locations sequentially rather than simultaneously. Gordon and Wingo suggested a simple extension that might guide planners or other third parties charged with assigning activities to optimal locations. Their formulation assumes an exogenous \(N \times N\) matrix of external (dollar) effects, \(e\). Locational proximity is assumed to be the sufficient condition for these potential externalities to become realized. These effects can be summarized by forming the \(N^2 \times N^2\) matrix, \(b\), where

\[
b_{imjn} = \begin{cases} 
  e_{ij} & \text{if } d_{mn} \leq L, \\
  0 & \text{if } d_{mn} > L.
\end{cases}
\]

The selection of \(K\) as the proximity threshold at which externalities become realized is arbitrary, and simply denotes the spatial attenuation of externalities.

In addition, Gordon and Wingo’s formulation specifies a new \(N^2 \times N^2\) set of variables, \(y_{imjn}\), that relate multiplicatively to the original choice variables via a nonlinear constraint:

\[
y_{imjn} = x_{im} x_{jn}, \quad i, j \in I; \ m, n \in M.
\]

Thus, these nontransaction interactions are accounted for in the same way as transaction-based interactions. Gordon and Wingo’s objective function points to the maximization of

\[
Z = \sum_{i} \sum_{m} a_{im} x_{im} + \sum_{i} \sum_{m} \sum_{j} \sum_{n} (b_{imjn} - f_{ij}) d_{mn} y_{imjn},
\]

subject to constraints (6) through (9). Constraint (9) could, of course, be substituted directly into the objective function.
The usefulness of this formulation, henceforth identified as GWA, is simply that it makes explicit the difference between market solutions of the linear assignment model and solutions that consider all manner of interaction effects. These differences exist in terms of contrasting outcomes, information problems, and solution properties, but only if a static, simultaneous solution is a realistic simulation of the problem to be solved.

Fortunately, the requirement for a static, simultaneous solution can easily be shed. Model GWA can be easily adapted to an intertemporal context. All that is required is (discrete) time notation and information on the costs of relocations between time periods. The maximand becomes

\[
Z = \sum_i \left[ \sum_i \sum_m a_{im}(t)x_{im}(t) - \sum_i R_i |x_{im}(t) - x_{im}(t-1)| \right],
\]

where \( R_i \) is the fixed relocation cost of activity \( i \); constraints (6) through (9) continue to hold; and the elements of \( a_{im}(t) \) are endogenous and specific to each time period. That is, initial values, \( a_{im}(0) \), are presumed to be known, and each \( a_{im}(t+1) \) is updated within the program as follows:

\[
a_{im}(t+1) = a_{im}(t) + \sum_i \sum_{m} (b_{im} - f_{ij} a_{mn}) y_{imn}(t).
\]

This means all benefits known to accrue to activity \( i \) at site \( m \) in period \( t \) are reflected in the bid of the operator of plant \( i \) for location \( m \) in the next period. Each bidder's evaluation is premised on the locational patterns that were observed in the previous period. If each time period is defined by the arrival (or departure) of exactly one new bidder, then all of the information needed for the determination of accurate bids is available. These conditions are the polar opposite of those defined for Koopmans and Beckmann's static solution.

Although the revised problem, henceforth identified as GWb, models intertemporal location, it treats all time periods simultaneously, thus implying considerable data and computational requirements. Less well informed, myopic, or more practical planners might be able to solve an alternative intertemporal problem by placing expression (12) outside of the optimization and replacing the modulus of the previous objective function with the following linear expression:

\[
\sum_i \sum_m (a_{im}(t) - R_i [1 - x_{im}(t-1)]) x_{im}(t),
\]

where \( x_{im}(t-1) \) is exogenous to time period \( t \). This linear function could be optimized in each period, subject only to constraints (5) through (8).

The planner is left with a series of tractable linear programs, henceforth identified as GWc, formulated on the plausible basis that limited information precludes simultaneous treatment of all time periods. Though this incremental perspective implies an efficiency loss, this so-called loss is nothing more than the value the planner or society would place on perfect information. Unfortunately, there is no market in which perfect information is for sale. All systems have boundaries and all optimizations are suboptimizations. We endorse a periode-specific approach because to do otherwise is to exchange tractability for the opportunity to exploit information that cannot be obtained. This intertemporal perspective has been endorsed by the work of other scholars, including that of Anas (1983) and, more recently, Werczheimer (1987).
3 Intertemporal models of urban location and travel choice

Gordon and Wingo identified a general planning model that retained the weakness of exogenous unit costs of transportation. Whereas this elaboration of Koopmans and Beckmann’s assignment model sharpens our thinking about land markets and about planning, it is deficient in the treatment of traffic flows. Exogenous traffic intensities, flows, and travels costs are unrealistic.

Extending model GWC to account for the endogeneity of travel costs, network assignment, and discrete location provides the foundation for a new class of tractable planning models, a formulation that accomplishes more than any existing alternative. As we have noted, this embellished model offers highly plausible behavioral interpretations. A suggested version of this new model, henceforth identified as GMa, follows below.

Maximize \( Z = \sum_i \sum_m a_{im} x_{im} + \sum_i \sum_m \sum_j \sum_n (b_{imjn} - f_j d_{mn}) y_{imjn} - \sum_a \int_0^{f_a} d_a(w) dw \) \tag{14}

subject to

\[ \sum_i x_{im} = 1, \quad m \in M, \] \tag{15}
\[ \sum_m x_{im} = 1, \quad i \in I, \] \tag{16}
\[ x_{im} x_{jn} = y_{imjn}, \] \tag{17}
\[ \sum_k \sum_m \sum_n y_{imjn} f_{kinjn} = f_j, \quad i, j \in I, \] \tag{18}
\[ \sum_k \sum_m \sum_j \sum_n y_{imjn} \delta_{akmn} f_{kinjn} = f_a, \quad a \in A, \] \tag{19}
\[ x_{im} = 0, 1, \] \tag{20}

and

\[ f_{kinjn} \geq 0, \] \tag{21}

where \( f_a \) is the total flow on transportation link \( a \); \( d_a(.) \) is an increasing (average) cost function of the flow on link \( a \); \( f_{kinjn} \) is the flow on path \( k \) from activity \( i \) located at site \( m \) to activity \( j \) located at site \( n \); and \( \delta_{akmn} \) is a binary indicator equal to one if link \( a \) is on path \( k \) from site \( m \) to site \( n \), and is equal to zero otherwise.

The last term of the objective function captures user equilibrium transportation costs accruing on the transportation network. Constraints (18) and (19) are the conventional path-flow constraints for static network equilibrium. Constraint (18) ensures that the traffic flows between activity \( i \) located at site \( m \), and activity \( j \) located at site \( n \), over all paths \( k \) satisfy the exogenous interaction requirement for activities \( i \) and \( j \). Constraint (19) ensures that the flows on all of the paths using link \( a \) contribute to the total flow on that link. The remainder of the formulation is precisely model GWC.

Model GMa includes endogenous costs of transportation line-haul, but retains the restrictive assumption of fixed intensities of traffic. Further, the solution to model GMa is subject to the question of price sustainability visited on the solution of the original Koopmans–Beckmann problem. It is not even clear that this formulation has a unique optimum (we have not checked for convexity), because expression (18) is cubic. Finally, the path-flow formulation requires the preenumeration of all network paths, some of which may be very similar. A formulation of link flow would
suspend the requirements for enumerated paths, but could not provide unique path flows as outputs.

Consider extension GMb corresponding to the ‘myopic’ intertemporal model GWc. Extending expression (12) provides

\[ a_{lm}(t + 1) = a_{lm}(t) + \sum_{j} \sum_{n} \left( b_{lmjn} - \sum_{k} f_{kmjn}(t) \left( \sum_{a} d_{akmn} d_{a}[f_{ja}(t)] \right) \right) y_{lmjn}(t). \]  

We obtain the final cost term in expression (22) by solving a conventional static network equilibrium problem given a known set of activity locations.

\[ \text{Minimize } Z = \sum_{a} \int_{0}^{f_{ja}(t)} d_{a}(w) dw, \]

subject to

\[ \sum_{k} \sum_{i} \sum_{m} \sum_{n} y_{lmjn}(t) f_{kmjn}(t) = f_{ij}, \]

\[ \sum_{k} \sum_{i} \sum_{m} \sum_{n} \sum_{j} y_{lmjn}(t) d_{akmn} f_{kmjn}(t) = f_{ja}(t), \]

\[ f_{kmjn}(t) \geq 0, \]

where \( y_{lmjn}(t) \) is exogenous to the optimization. This is a convex nonlinear programming problem that can be solved by a number of convenient methods, including specialized applications of the Frank–Wolfe algorithm (Eash et al, 1979). If \( d_{a}(.) \) is adequately approximated by an increasing linear function, then the objective function is merely quadratic. The outputs of this optimization include equilibrium link and path flows, but additional information on path costs is required to complete expression (22). These are obtained by summing link costs over the paths defined between each pair of sites \( m \) and \( n \). Given these path costs, bid rents in expression (22) can be completely updated.

Specification GMb provides a highly computable, incremental optimization model of discrete land-use with endogenous costs of transportation. Is it also possible to endogenize traffic intensities without sacrificing tractability?

The original version of the Koopmans–Beckmann problem was formulated for a precise match between \( N \) single-technology activities and \( N \) sites. We suggest that multiple, discrete technologies (Mills, 1972), each associated with a separate vector of traffic intensities, be defined for each activity, and that dummy sites be defined for unused production technologies. We contend that this specification extends model GMb to include endogenous transportation costs and endogenous technology choice. In this final formulation, identified as GMc, technology choice (that is, input choice) will be driven by the proximity of input sources. This will effectively endogenize traffic intensities, but in a discrete way. Further, there would be no need to resort to heuristic, iterative manipulations, such as those invoked by Hopkins and Los (1979). Instead, discrete land-use, transportation flows and costs, technology choice, and activity interactions will be simultaneously determined by solving an incremental optimization problem. Though intertemporal, this math program poses a smooth, convex optimization problem with linear constraints. The model is a treatment of the planning questions most central to the field, but with a degree of tractable simultaneity that has not been achieved by previous formulations.

4 Conclusion

Model GMc is premised on the assumption that economic actors are well informed about the present, are partially constrained by the past, and are intensely interested in the benefits accruing over short-term futures. This perspective permits planners
and actors to identify system and self-oriented alternatives, despite the operation of confounding conditions relating to the market. The key to treating limited rationality is not to constrain the information imposed on static formulations, but to differentiate between the quantity of the information available concerning the future and that concerning the present.

Model GMc lends itself nicely to the investigation of polycentric urban forms, and the simulation of polycentric development. Although little is understood about polycentric development, the broad outlines can be sketched. First, subcentering is part of a dynamic process. As most cities have grown, both residential and nonresidential activities have tended to disengage from the center. In many cases subcenters have formed. Eventually, the spread of some of these secondary centers has given rise to generally dispersed sites of employment (Richardson and Gordon, 1986). Agglomeration opportunities originally induced firms in most cities to cluster in their central business districts (CBDs). New firms, in turn, contributed new agglomeration opportunities, though at an ever decreasing rate. In addition, new growth brought on eventual crowding and congestion. As is well known, congestion costs are thought to grow at an increasing rate. We assert that further accretion at the center ceases when the agglomeration economies that would be available to a new firm locating in the CBD no longer dominate the congestion costs associated with being in the downtown, and when the possibility of greater advantage of agglomeration over congestion exists elsewhere. Thus, emergence, growth, decline, and obsolescence of individual urban subcenters is most likely the result of simple economic behaviors.

The modeling of these processes requires a dynamic approach that allows localized congestion and agglomeration effects to be endogenously determined in each period. These effects, in turn, should affect further developments. The model suggested in this paper includes all of the required properties, couched in appealing simplicity. Model GMc captures the cogency of the decisions involved. Further, we are convinced that this formulation will generate solutions consistent with empirical observations.

References
Alonso W, 1964 Location and Land Use (Harvard University Press, Cambridge, MA)
Anas A, 1983, “The effects of transportation on the tax base and development of cities”, DOT/OST/P-30/85/005, final report to the US Department of Transportation (National Technical Information Service, Springfield, VA)
Hopkins L D, 1977, "Land-use plan design—quadratic assignment and central-facility models" Environment and Planning A 9 625–642
Hopkins L D, Los M, 1979, "Location-allocation algorithms for land use plan design with fixed and substitutable interactions" Journal of Regional Science 19 345–361
Kim T J, 1983, "A combined land use—transportation model when zonal travel demand is endogenously determined" Transportation Research B 17 449–462
Los M, 1978, "Simultaneous optimization of land use and transportation" Regional Science and Urban Economics 8 21–42
Los M, 1979, "A discrete convex programming approach to the simultaneous optimization of land use and transportation" Transportation Research B 13 33–48
Mills E S, 1972, "Markets and efficient resource allocation in urban areas" Swedish Journal of Urban Economics 74 10C–113