A Theoretical Foundation for the Gravity Model of Factor Flows

James E. Anderson
Boston College
January 1978

Working Paper #85

A Theoretical Foundation For The Gravity Model of Factor Flows

The gravity model of commodity and factor flows is one of the most successful examples of measurement without theory to be found. It is usually specified as $X_{ij} = \alpha Y_i^{\beta} Y_j^{\gamma} N_i^{\delta} N_j^{\varepsilon} (d_{ij})^{\phi} V_{ij}$, where X_{ij} = flow of the good or factor from country or region i to country or region j, Y_i, Y_j are incomes at i and j, N_i, N_j are populations at i and j, d_{ij} is the distance between i and j, and V_{ij} is a log-normal disturbance with zero mean. In a wide variety of cross-sections, at different levels of aggregation, and sometimes with pooled cross-section and time series data, it usually gets a good fit. Unfortunately, its usefulness is severely hampered by the unidentified nature of the model: policy instruments have no identifiable role, and the stability of the equation is always suspect, especially under policy change.[1]

---

[1]The only attempt to explain the gravity equation in terms of economic theory of which I am aware is Niedercorn and Bechdolt (1969). It is not very satisfactory. They present a model explaining the geographic trip pattern of a utility maximizing individual within, say, an urban area. Trips cost money (or time) so the budget constraint binds choice. Utility is an additive function, across destinations, of population at the destination (representing total number of potential "interactions") times a function of total trips to the destination. The derived demand for trips for any O-D pair thus depends on population of the destination relative to other destinations, the cost of trips, and income at the origin. Given nice restrictions on the utility function, the gravity equation, more or less, emerges.

The problem with this model is that it simply moves the attractive force of destination population into the utility function without a fundamental economic theory. This side of the model is no better than the social physics gravity model. Essentially they have half of an economic explanation of the gravity model through the budget constraint.

A second point against the model in the present context is that while it may be a reasonable urban trip model, it cannot be at all useful in migration studies.

Recently, I have identified the commodity flow gravity equation as part of a reasonable approach to the pure cross-section estimation of expenditure systems (Anderson, 1977). Trade's share of total expenditure varies across regions or countries according to a stable unidentified function of income, population, and resource endowment variables. Traded goods are differentiated by place of origin.[2] Expenditures on individual traded goods out of total traded goods expenditures are identical functions across countries or regions. The gravity equation turns out to be an efficient way to use the information contained in the trade balance constraints.

This paper will show that a somewhat similar chain of reasoning partially identifies the cross-section gravity equation of factor flows. Growth rates of income and factors vary across regions or countries as unidentified functions of income, population, and resource endowment. Factors of production within a class like labor, or even unskilled labor, are differentiated by place of origin. Demand for factors of production is an identical function across regions or countries. The gravity equation is a reasonable approach to estimation of the demand system with use of the information contained in the restriction that new demand for factors of production must equal new supply of factors of production.

---

[2]For evidence on the reasonableness of this assumption, see P. Isard (1977).

In contrast to the commodity case, the restriction that factors of production be differentiated by place of origin is extreme. In the case of capital it may be too extreme (though the prevalence of two-way capital flows even when capital is disaggregated by type suggests some differentiation), so the remainder of this paper develops the model for labor only. The model for capital would be very similar. For the case of labor, differentiation by place of origin certainly has some plausibility (Southern migrants to the Northeast tend to be low-skilled, Northern migrants to the Southeast tend to be high-skilled, etc.), but why should labor of a given skill grade be differentiated by place of origin? Customs, language, and word-of-mouth channels of information transmission may combine to effectively differentiate by place of origin. Thus the differentiation model can perhaps be taken as a useful parable.[1] This is in part an empirical question, and in this regard note that the factor flow gravity equation almost necessarily implies differentiation by place of origin. The movement of the factor between any points i and j is fitted as a log-linear function of variables at points i and j. Any undifferentiated model makes it very difficult to: (1) explain two way flows ($i \rightarrow j$ and $j \rightarrow i$), and (2) explain why the $i \rightarrow j$ flow does not

---

[1]Uniformity of wages within a skill grade in a region is not complete, but even so, social pressures and union practices undoubtedly force more uniformity than is justified by productivity.

also depend on the strength of flows to j from all other points. The goodness-of-fit of the gravity equation suggests that the differentiation model is useful.

Part I presents the simplest case of a gravity-type factor flow model derived on the lines indicated above. While it yields a gravity-type equation, it would probably not be sensibly so estimated. In Part II, the gravity equation gains legitimacy as an efficient user of the information in the new supply = new demand restrictions. Questions of bias are immediately raised, so there is a bias-efficiency tradeoff. Part III introduces distance and policy variables creating barriers to mobility into the model. It shows that the usual treatment of distance is improper and proposes alternatives. The introduction of policy variables at least potentially opens the way to the use of gravity equations to help in estimating the long run effect of minimum wage restrictions, relocation subsidies, and similar policy variables.

I. The Simplest Model

Suppose that n regions experience uniform steady state growth. There are no barriers to factor mobility. Where Y_i is income in region i and N_i is the labor force in region i:

(1) $$\Delta Y_i/Y_i = g = \Delta N_i/N_i, \quad i=1,\ldots,n.$$

Labor is differentiated by place of origin, so at any point in time N_i consists of a stock of heterogeneous labor. Natural additions to the labor force at point i at any point in time are considered however, to take on the attributes of region i, to be i-type labor. With identical Cobb-Douglas production formation across regions, the stock demand for labor of type j in any region i is:

(2) $$F_{ji} = b_j Y_i, \quad i,\ j=1,\ldots,n.$$

The coefficient b_j is the share parameter for labor of type j in the production function. With steady state growth, and no barriers to mobility (type j labor receives the same wage everywhere), all prices may be set at unity, hence (2) is the stock demand for labor of type j in region i. Taking differences in (2) we have:

(3) $$f_{ji} = \Delta F_{ji} = b_j \Delta Y_i.$$

f_{ji} is the flow of labor from region j to region i over some interval of time such as a year. (3) assumes implicitly that migrants make once for all decisions, so that new entrants to the labor force at i are the only source of new i-type labor. For factor market equilibrium, new supply of labor to the n region system must equal new demand:

$$(4) \qquad \sum_i f_{ji} = \Delta N_j .$$

Substituting (3) into (4) and solving for b_j we obtain:

$$(5) \qquad b_j = \frac{\Delta N_j}{\sum_i \Delta Y_i} .$$

Substituting back into the factor demand equation (3) and using (1), we obtain two alternate forms of the stripped down gravity equation:

$$(6) \qquad f_{ji} = \frac{Y_i Y_j}{\sum_i Y_i} g = \frac{N_i N_j}{\sum_i N_i} g .$$

Compared to the usual gravity model, either income or population is omitted, distance is omitted, the denominator term is added (though it is constant across observations), and only the constant term g is to be estimated, income or population being constrained to unit elasticity. The right hand equality, except for the denominator and the absence of distance, is the original "social physics" form of the gravity equation applied to movement of people.

While a gravity-type equation (6) is implied by (1)-(5), it would probably not be sensible to use (6) to estimate g.

If we postpone consideration of error structure issues like simultaneous equations bias until Part II, the n+1 parameters g, $b_1, \ldots, b_n$ can be estimated from (1), with 2n observations on g; and from (3), with n observations for each parameter $b_1, \ldots, b_n$. (6) gives n^2 observations to estimate g, but is almost sure to be a "noisier" estimator than (1).

II. The Growth Rates Variable Case

Suppose that the model of Part I is altered by making growth rates of income and labor force different, and by allowing each growth rate to vary across regions or countries. Every element of the heterogeneous stock of labor in each region conforms to the local growth rate in labor rather than the growth rate at the point of origin (birth minus death rates and labor force participation take on the local features).[1] This is somewhat plausible, but an extreme assumption. As in Part I, all the entrants to the labor force at each point i have the i-type characteristic, regardless of their relatives' point of origin. And only the new entrants make migration decisions, so that the only source of new i-type labor is the addition to the labor force at point i (before migration). These are again extreme but somewhat plausible assumptions.

Turning to the unidentified growth rate functions which drive the model, they are log-linear functions of income, labor force, and endowments. Loosely speaking, for the labor force growth rate, both participation rates and natural growth rates will vary with income per capita, resources per capita, and possibly absolute size. Income growth rates depend on the resource stock, labor force growth rates, and capital growth

---

[1]This is not strictly necessary at a given time, but if we want to use the equation over time (as we certainly do for prediction), it is necessary.

rates, the latter of which depends on the savings rate, income, and the capital stock. The savings rate may depend on income per capita. Growth theory and demographic theory may well be productive in restricting the forms and independent variables of the growth rate functions. This is an important task for future work. Our present purpose is to show that the log-linear model produces the conventional gravity equation, and lends it legitimacy as an estimating device.

Instead of (1) we now have, where E_i is the resource endowment in i:

$$(1') \qquad \frac{\Delta Y_i}{Y_i} = g(Y_i, N_i, E_i)$$

$$\frac{\Delta N_i}{N_i} = h(Y_i, N_i, E_i)$$

g( ) and h( ) are log-linear in form.

(3) and (4) are reproduced as before:

$$(3) \qquad f_{ji} = b_j \Delta Y_i$$

$$(4) \qquad \sum_i f_{ji} = \Delta N_j$$

The chain of substitution which leads to the gravity equation now produces:

$$(6') \qquad f_{ji} = \frac{h(Y_j, N_j, E_j) N_j \; g(Y_i, N_i, E_i) Y_i}{\sum_i g(Y_i, N_i, E_i) Y_i}$$

(6') is the gravity equation as usually specified save for the absence of distance and the presence of the constant term in the denominator. In the stochastic forms of (1') and (6'), the gravity equation in effect offers n^2 observations to determine the parameters of n( ) and g( ), rather than the n observations in (1'). There is a great gain in efficiency of estimation. The conditional expectations (estimated values) of $g(Y_i, N_i, E_i)Y_i$, $\Delta\hat{Y}_i$, can then be used to replace ΔY_i in the demand equations (3), and the n parameters $b_1, \ldots, b_n$ estimated, each with n observations. The substitution of the estimated values $g(Y_i, N_i, E_i)Y_i = \Delta\hat{Y}_i$ for ΔY_i has the great advantage that $\Delta\hat{Y}_i$ should be independent of the error term of (3). That is, if (3) is estimated in the stochastic version of its original form, ΔY_i determines F_{ji}, but F_{ji} may also help determine ΔY_i (through its influence on demand) and we have simultaneous equations bias.

Ironically, however, (1') and (4) taken together in their stochastic form imply that the vector of Y's is not determined independently of the error terms of (1') and (3). The gravity equation (6') in its stochastic form will yield biased estimates of the parameters of g( ) and n( ). It is still logically possible to escape this implication by specifying the substitution of (1') into (4) must hold in deterministic form only ($b_j\ g(Y_i, N_i, E_i)Y_i = N_j\ h(Y_j, N_j, E_j)$, "planned" new supply equals "planned" new demand), but this is a dubious expedient. Tackling the problem directly, one good procedure might be to take as instruments for the Y's the one period lagged Y's. They should be highly correlated with current Y's and close to uncorrelated with the current error terms of (1') and (3). Efficiency of estimation is reduced, naturally, but the bias problem is reduced. One period lagged Y's may not always be good instruments, however. System procedures which take account of the simultaneity involved in (4) are possible, but messy, since (1'), (3), (4) is a non-linear system. Use of them implies further reductions in efficiency, at a further gain in cutting down asymptotic bias. For a discussion of essentially this problem in the commodity flow gravity model, see Anderson (1977). The gravity equation (6') in its stochastic form may not be a bad compromise between efficiency and bias. Empirical work is needed to deepen our understanding here.

III. Barriers to Mobility: Distance and Policy Variables

The role of distance in gravity models is to proxy direct and indirect transit costs of all sorts. The indirect costs include information costs and adjustment costs on arrival to different customs, etc. In effect all such costs act like a tax on labor, distorting the wage equilibrium imposed in the previous sections.[1] The gross return to labor of each type can now vary across regions. Differential taxation of wage income across regions acts in exactly the same way. Finally, quotas on migration in the international context create implicit taxes on migration. If these could be calculated (a difficult task) then the implicit tax would act just like the transit costs proxied by distance.[2] Denote the scale factor which distorts the gross return to labor of type j at point i from its return at point j as t_{ji}. For flow equilibrium, net returns must be equal everywhere for type i labor. If W_{jj} is the wage of type j labor at the source, then

$$W_{jj}t_{ji} = W_{ji} \qquad i,j=1,\ldots,n. \tag{7}$$

---

[1]The transit costs are of course a capital expense. To convert them to a flow tax on labor, implicitly we simply use the discounted tax stream equivalent to the capital expense.

[2]The tax on migration is again a capital expense, which must be converted to a flow equivalent.

Without loss of generality, set $W_{jj} = 1$ for all regions j. Evidently $t_{jj} = 1$. With distance proxying all costs, $t_{ji} = f(d_{ji})$.

Now consider the derivation of the gravity equation. The demand for labor equation, assuming the Cobb-Douglas form for simplicity, is now:

$$(3') \qquad f_{ji} = \frac{1}{t_{ji}} b_j \Delta Y_i$$

Using (1') and (4), the gravity equation now becomes:

$$(6'') \qquad f_{ji} = \left[\frac{\dfrac{1}{t_{ji}}}{\sum_i \dfrac{1}{t_{ji}} \dfrac{Y_i g(Y_i, N_i, E_i)}{\sum_i Y_i g(Y_i, N_i, E_i)}} \right] \frac{Y_i g(Y_i, N_i, E_i) h(Y_j, N_j, E_j) N_j}{\sum_i Y_i g(Y_i, N_i, E_i)}$$

(6'') differs from (6') by the large bracketed term. If the term $\frac{1}{t_{ji}}$ is brought outside the bracket, and is made log-linear in distance; except for the bracket and ignoring the constant denominator, (6'') is the usual gravity equation. Note that the denominator of the bracketed term cannot, however, be treated as a constant in moving across regions j. The bracket term, <u>relative economic distance</u> is a weighted average of economic distance between j and i relative to economic distances between j and all other points, where the weights are relative demands (in changes). (6'') states that the flow from j to i depends on growth in demand at i relative to growth at all points, growth in supply, and relative economic distance. Its intuitive appeal has now been thoroughly grounded in economic theory.

The usual specification of the gravity model is thus erroneous, since it omits the weighted average economic distance deflator. Unfortunately, the true relation is highly non-linear in the logs since the weights are unknown prior to estimation, as is the function $t_{ji} = f(d_{ji})$. Several considerations hold out some hope. The weights in the denominator could perhaps be approximated by weights based on last period's changes in income. In the stochastic version of the system, their error terms should be close to uncorrelated with this period's error terms and yet they should be highly correlated with this period's changes in income. The economic distance terms t_{ji} can be measured directly in so far as they reflect policy, such as differential income taxation. Thus, split t_{ji} into the policy component t^*_{ji} and the pure distance component $f^*(d_{ji})$, and $t_{ji} = t^*_{ji}\ f^*(d_{ji})$. This reduces the amount of "work" the distance term must perform, as well as being useful on its own. The resulting denominator term is now approximated as:

$$(8) \qquad \sum_i \frac{1}{t^*_{ji}} \frac{1}{f^*(d_{ji})} \frac{\Delta Y_i, t-1}{\sum_i \Delta Y_i, t-1} \simeq \sum_i \frac{1}{t_{ji}} \frac{Y_i g(Y_i, N_i, E_i)}{\sum_i Y_i g(Y_i, N_i, E_i)} .$$

Using the left hand side of (8) in (6'') produces a gravity-type equation which, while still non-linear in the logs, is not too unpleasantly non-linear to be estimated successfully, since only the exponent of $f^*(d_{ji})$ is producing non-linearity.

Note finally that even if the above approximation is not used, the usual form of the gravity equation may often not be biased too badly. The weighted average economic distance deflator will vary across origin points j only slightly with many polygonal distributions of regions. With only a small variance, the specification error from falsely treating it as a constant is likely to be small. On the other hand, if regions are distributed in highly rectangular fashion (in the limit, on a line), then the specification error is almost certain to be serious. Thus, prior knowledge about the shape of the regional distribution can guide the choice of whether to use (8) in (6''), or the usual procedure of treating the weighted average economic distance as a constant. But the more complex procedure which overcomes the specification error should ordinarily be preferred.

Note finally that while the discussion has been restricted to the Cobb-Douglas production function case for simplicity, more complex homogeneous production functions could be used. See Anderson (1977) for a discussion of the CES case.

The methods of this section and those previous can be used to identify a long run equilibrium model of factor flows in which the role of policy variables is made clear. Based on estimates from it, long run effects of policy changes can be computed.[1]

The gravity equation is shown to be a useful procedure for utilizing the information contained in market clearing relations in estimation of the system. The major problem with the approach is its assumption of differentiation by place of origin. Empirical work can perhaps illuminate the extent to which this is a useful parable. Further areas which need work are: the identification of the growth rate functions,

---

[1]For example, suppose all parameters of the system (i)-(v) have been estimated:

$$\text{(i)} \quad \Delta Y_i = g(Y_i, N_i, E_i) Y_i$$

$$\text{(ii)} \quad \Delta N_i = h(Y_i, N_i, E_i) N_i$$

$$\text{(iii)} \quad f_{ji} = b_j \, \Delta Y_i \, \frac{1}{W_{ji}}$$

$$\text{(iv)} \quad W_{ji} = W_{jj} t_{ji}$$

$$\text{(v)} \quad \sum_i f_{ji} = \Delta N_j \qquad ; \; i,j=1,\ldots,n.$$

A policy induced change in some t_{ij} is imposed, and we wish to find the vector of changes in factor prices and the implied vector of changes in migration. From (i) and (ii) we obtain estimates of ΔY_i, ΔN_i. The system (iii)-(v) contains $2n^2$ useful (non-identical, as in $W_{ii}=W_{ii}$) equations in $2n^2$ unknowns f_{ji}, W_{ji}. Thus we plug in the estimated ΔY_i, ΔN_i and the parametric change in t_{ji} and solve for the new values of f_{ji} and W_{ji}. This is a long-run quasi-general equilibrium approach.

and empirical work to deepen understanding of the tradeoff between bias and efficiency in estimation of the type of system developed in Parts II and III.

REFERENCES

Anderson, J.E., "A Theoretical Foundation for the Gravity Equation," mimeo, 1977. American Economic Review, forthcoming.

Isard, Peter, "How Far Can We Push the Law of One Price?," American Economic Review, December 1977.

Niedercorn, J., and B. Bechdolt, "An Economic Derivation of the 'Gravity Law' of Spatial Interaction," Journal of Regional Science, 1969 (Vol. 9, No. 2), pp. 273-284.