A Theoretical Foundation for the Gravity Equation

By JAMES E. ANDERSON*

Probably the most successful empirical trade device of the last twenty-five years is the gravity equation. Applied to a wide variety of goods and factors moving over regional and national borders under differing circumstances, it usually produces a good fit. Unfortunately, as is widely recognized, its use for policy is severely hampered by its "unidentified" properties. Insertion into the equation of policy instruments such as border taxes has no theoretical justification; and inference about the effect of taxes from examining changes in the equation over times when taxes have changed carries no guarantee of validity.

The gravity equation ordinarily is specified as

$$(1) \qquad M_{ijk} = \alpha_k Y_i^{\beta_k} Y_j^{\gamma_k} N_j^{\xi_k} N_j^{\epsilon_k} d_{ij}^{\mu_k} U_{ijk}$$

where M_{ijk} is the dollar flow of good or factor k from country or region i to country or region j, Y_i and Y_j are incomes in i and j, N_i and N_j are population in i and j, and d_{ij} is the distance between countries (regions) i and j. The U_{ijk} is a lognormally distributed error term with $E(ln\ U_{ijk}) = 0$. Frequently the flows are aggregated across goods. Ordinarily the equation is run on cross-section data and sometimes on pooled data. Typical estimates find income elasticities not significantly different from one and significantly different from zero, and population elasticities around $-.4$ usually significantly different from zero.[1]

The intent of this paper is to provide a theoretical explanation for the gravity equation applied to commodities. It uses the properties of expenditure systems with a maintained hypothesis of identical homothetic preferences across regions. Products are differentiated by place of origin (for a justification, see Peter Isard). The gravity model constrains the pure expenditure system by specifying that the share of national expenditure accounted for by spending on tradeables (openness to trade) is a stable unidentified reduced-form function of income and population. The share of total tradeable goods expenditure accounted for by each tradeable good category across regions is an identified (through preferences) function of transit cost variables. Partial identification is achieved. While other interpretations are possible (see for example Edward Leamer and Robert Stern),[2] the one advanced here has

*Professor of economics, Boston College. I am indebted to Marvin Kraus and Edward Leamer for helpful comments.

[1]See for example Norman D. Aitken.

[2]They offer three explanations. The first, based on physics, has little interest. The second identifies the equation loosely as a reduced form with exogenous demand-side variables (importer income and population) and supply-side variables (exporter income and population). Alternatively, the importer and exporter characteristics identify the size of the foreign sector in each, with any flow a function of size at either end. The third interpretation is based on a probability model. Let Z_i be country i's total imports, an unidentified reduced-form function of income, population, and other possibly unobservable variables. The set $\{Z_i/T\}$, where $T = \Sigma_i Z_i$, world trade, has the form of a probability distribution. Alternatively Z_i/T is a trade potential. The probability of the occurrence of flow between i and j is taken to be $Z_i Z_j/T^2$. Alternatively, potential between i and j is the product of the i and j potentials. The expected size of the flow given T is then $M_{ij} = Z_i Z_j/T$. The term T is constant in a cross-section study and can be neglected. Resistance to trade, proxied by distance, can be inserted and with the *log*-linear form for all functions, we have the gravity equation. This interpretation has the advantage of explaining the multiplicative functional form, and has a useful flexibility. Leamer subsequently developed a hybrid version of it to explain aggregate imports of good k by country i. In the hybrid model, Z_j becomes $Z_{j(i)} = \Sigma_{j \neq i} Z_j$. Also, the parameters of the Z_i and $Z_{j(i)}$ functions are permitted to vary by commodity group k. Leamer's hybrid is thus

$$M_{ik} = Z_i^k Z_{j(i)}^k \Psi(\mathbf{d}_i, \mathbf{t}_i, \mathbf{t}_{j(i)})$$

where $\mathbf{d}_i$ is a vector of distance from i to all other countries, $\mathbf{t}_i$ is a vector of i's tariffs, and $\mathbf{t}_{j(i)}$ is a vector of all other countries' tariffs. The problem with either the Leamer-Stern gravity interpretation or the Leamer hybrid is that while the potential or probabilfour distinct advantages. First, it explains the multiplicative form of the equation. Second, it permits an interpretation of distance in the equation, identifying the estimated coefficient, and can be used as part of an attack on estimating the effect of instrument changes. Third, the vague underlying assumption of identical "structure" across regions or countries is straightforwardly interpreted as identical expenditure functions. This suggests appropriate disaggregation. Finally, following the logic of the present interpretation implies that the usual estimator of the gravity equation may be biased, requiring change in the method of estimation.

The present interpretation of the gravity model makes it part of an alternative method of doing cross-section budget studies. The bias problems now uncovered may be quite severe, especially with transit costs varying considerably, but there are efficiency gains to trade off against them. The background of difficulty in modelling trade flows requires respect for any potentially promising method. This paper shows that the gravity model may merit continued development and use.

Section I develops the simplest linear expenditure model, which produces an equation like (1) but with the last three variables omitted and with Y_i and Y_j constrained to have unit elasticity. The major portion of the explanatory power of the gravity model is thus encompassed. While yielding a gravity equation, the models would never sensibly be so estimated.

In the next two sections, the gravity approach gains legitimacy as a device offering large gains in efficiency of estimation at a possible cost of bias. An important fact of life is large interregional and international variations in shares of total expenditure accounted for by traded goods, even across regions or countries where spending patterns are reasonably similar (for example, the set of developed Western countries). These are assumed to vary as a function of national income and population. Total trade expenditure is distributed across individual categories by share functions which are identical across countries. With this structure, Section II produces a gravity equation with potentially attractive properties. Section III discusses estimation of the model, shows that the usual technique may produce biased results, and suggests alternatives. Section IV integrates in distance (as a proxy for transport costs) and border taxes producing a full model suggesting the possibility of identifying long-run tariff elasticities. A constant elasticity of substitution (*CES*) case developed in the Appendix provides further details.

Two areas for theoretical development may be noted. The major remaining unidentified part of the equation is the function stipulating that trade's share of budgets is dependent on income and population. While this is a well-established empirical relation, it would be nice to have an explanation. None is offered here.

Use of pooled cross-section and time-series data requires a further development of the model also not attempted here. Two requirements should be noted. First, a theory of how short-run responses to price changes (revealed over time) are related to long-run responses (revealed over the cross section) is needed. Second, a theory of how short-run responses to income changes (revealed over time in the Keynesian type of trade model) are related to long-run responses must be constructed. Part of the second story must be the relation of trade balance to asset accumulation stressed in the recent monetary approach to the balance of payments.

I. The Pure Expenditure System Model

The simplest possible gravity-type model stems from a rearrangement of a Cobb-Douglas expenditure system. Assume that each country is completely specialized in the production of its own good (as in a Keynesian-type trade model), so there is one good for each country. No tariffs or transport costs exist. The fraction of income spent on the product of country i is

ity story is plausible, it lacks a compelling economic justification.

denoted b_i and is the same in all countries (i.e., there are identical Cobb-Douglas preferences everywhere). With cross-section analysis, prices are constant at equilibrium values and units are chosen such that they are all unity. Consumption in value and quantity terms of good i in country j (= imports of good i by country j) is thus

$$M_{ij} = b_i Y_j \tag{2}$$

where Y_j is income in country j.

The requirement that income must equal sales implies that

$$Y_i = b_i(\Sigma_j Y_j) \tag{3}$$

Solving (3) for b_i and substituting into (2), we obtain

$$M_{ij} = Y_i Y_j / \Sigma Y_j \tag{4}$$

This is the simplest form of "gravity" model. If we disregard error structure,[3] a generalization of equation (4) can be estimated by ordinary least squares, with exponents on Y_i, Y_j unrestricted. In a pure cross section, the denominator is an irrelevant scale term. The income elasticities produced (disregarding bias) should not differ significantly from unity. The functional form of the gravity equation and a major portion of the explanatory power is encompassed by the expenditure system model.

II. The Trade-Share-Expenditure System Model

The gravity equation of Section I is based on identical Cobb-Douglas preferences, implying identical expenditure shares and gravity equation income elasticities of unity. It could be fancied up by allowing policy induced price differences to produce different expenditure shares in a less restrictive preference form such as the *CES*, but there is little point in the exercise. While a gravity equation is produced by such a framework, the real variables of interest are the non-income-dependent expenditure shares. The gravity equation is a silly specification from an econometric standpoint since it substitutes out the share (which in the Cobb-Douglas case is the only parameter). This section appends to the Cobb-Douglas expenditure system for traded goods a differing traded-nontraded goods split and produces an unrestricted (non-unit income elasticity) gravity equation. The next section shows that the gravity equation becomes far more sensible.

Traded-goods shares of total expenditure vary widely across regions and countries. Hollis Chenery and others subsequently have found that in cross-section data such shares are "explained" rather well by income and population. Moreover, the linear or *log*-linear regression line of traded goods shares on income and population tends to be stable over time. No identification of this relationship is attempted here, but loosely, income per capita is an exogenous demand-side factor, and population (country size) a supply-side factor. Trade shares "should" increase with income per capita and decrease with size. Leamer and Stern have also suggested including an endowment measure as an explanatory variable which would act somewhat like size.[4] Accepting the stability of the trade-share function, the expenditure system model combines with it to produce the gravity equation.

Assume that all countries produce a traded and a nontraded good. The overall preference function assumed in this formu-

[3]Note that the manipulation which justified the presence of Y_i in (4) means that there may be simultaneous equation bias, with the Y's not being independent of error terms postulated for the equations (2) and (3). This problem is discussed in Section III.

[4]It is easy to construct examples where the trade share is a simple closed form function of factor endowment variables. Consider the small-country case in which all traded goods may be treated as a composite and there is one nontraded good. Stipulate a simple non-Cobb-Douglas utility function in traded goods and the nontraded good and assume the linear version of the two-sector production model. When a model of this sort is solved for an interior equilibrium-traded-goods share assuming trade balance, it produces a relatively simple function of the endowments. Results analogous to the Johnson pro- and antitrade bias analysis are embodied in the function. I have been unable, however, to discover a specification which produces a traded-goods share function which is *log*-linear in income and factor endowments.

lation is weakly separable with respect to the partition between traded and nontraded goods: $u = u\,(g(\text{traded goods}), \text{nontraded goods})$. Then given the level of expenditure on traded goods, individual traded-goods demands are determined as if a homothetic utility function in traded goods alone $g(\ )$ were maximized subject to a budget constraint involving the level of expenditure on traded goods. The individual traded-goods shares of total trade expenditure with homotheticity are functions of traded-goods prices only.[5] For simplicity, it is assumed $g(\ )$ has the Cobb-Douglas form in the rest of the text. Within the class of traded goods, since preferences are identical, expenditure shares for any good are identical across countries. Thus, for any consuming country j, Θ_i is the expenditure on country i's tradeable good divided by total expenditure in j on tradeables; i.e., Θ_i is an exponent of $g(\ )$. Let ϕ_j be the share of expenditure on all traded goods in total expenditure of country j and $\phi_j = F(Y_j, N_j)$.

Demand for i's tradeable good in country j (j's imports of i's good) is

$$M_{ij} = \Theta_i \phi_j Y_j \tag{5}$$

The balance-of-trade relation for country i implies

$$Y_i \phi_i = \Big(\sum_j Y_j \phi_j\Big) \Theta_i \tag{6}$$

value of imports of i plus domestic spending on domestic tradeables	=	value of exports of i plus domestic spending on domestic tradeables

Solving (6) for Θ_i and substituting into (5), we have

$$M_{ij} = \frac{\phi_i Y_i \phi_j Y_j}{\sum_j \phi_j Y_j} = \frac{\phi_i Y_i \phi_j Y_j}{\sum_i \sum_j M_{ij}} \tag{7}$$

With $F(Y_i, N_i)$ taking on a *log*-linear form, (7) is the deterministic form of the gravity equation (1) with the distance term suppressed and a scale term appended. More realistically, if trade imbalance due to long-term capital account transactions is a function of (Y_i, N_i), we may write the "basic" balance $Y_i \phi_i m_i = (\Sigma_j Y_j \phi_j)\Theta_i$, with $m_i = m(Y_i, N_i)$, and substitute into (6) and (7). This yields[6]

$$M_{ij} = \frac{m_i \phi_i Y_i \phi_j Y_j}{\sum_i \sum_i M_{ij}} \tag{8}$$

With *log*-linear forms for m and F, (8) is again essentially the deterministic gravity equation.

III. Estimation Efficiency

The model of linear expenditures of Section I, while implying a gravity equation, would never sensibly be so estimated. Homothetic preferences identical across countries imply identical expenditure share functions, and these can be estimated directly, treating distance and trade taxes appropriately in a manner set out in Section IV. Simply divide the demand equations of Section I or their analogue by own income and find the mean over j of M_{ij}/Y_j = the estimator of Θ_i. The stochastic budget constraint information can also be utilized to (in effect) add one observation since[7] $\Theta_i = Y_i/\Sigma_j Y_j$.

The trade-share model of Section II on the other hand lends some legitimacy to the gravity model. Eventually we will allow

[5]Homotheticity of $g(\ )$ is imposed because the presence of traded-goods expenditure as an argument in the Θ_i function will greatly complicate estimation. Separability is imposed to permit the two-stage decision process which removes nontraded-goods prices from the Θ_i function. Some justification for separability may be found in the observation that traded goods are far more similar to each other than they are to nontraded goods. Homotheticity would imply that, *ceteris paribus*, larger nations' trade expenditures are scalar expansions of smaller nations' trade expenditures. This may not do great violence to reality. A gravity-type model which imposes neither restriction is possible, but it would be far more complex and difficult to estimate. As with most empirical work on preferences, the restrictions' main appeal is convenience.

[6]Balance-of-payments disequilibrium could be treated as part of the error terms in estimation. Use of the model is best restricted to equilibrium years, however, since error terms due to disequilibrium may be correlated with Y_i, N_i, and therefore cause errors in estimation.

[7]Again, we disregard the problem that the Y's depend on the error term through the budget constraint.

many tradeables for each country, with tariffs and transport costs present, but initially, as before, assume only one tradeable in each and no barriers to trade. The system to be estimated is

$$(5') \qquad M_{ij} = \Theta_i \phi_j Y_j U_{ij}$$

$$(6') \qquad m_i \phi_i Y_i = \Theta_i \sum_j \phi_j Y_j$$

where U_{ij} is a *log*-normal disturbance with $E(ln\ U_{ij}) = 0$. Note that (6′) states that planned expenditures (reduced or increased by the capital account factor) = planned sales, and has no error term. Efficient estimation requires that the information in (6′) be utilized. The most convenient way to do this, since the constraint is highly nonlinear in the Y's, is to substitute out Θ_i and estimate the gravity equation:

$$(8) \quad M_{ij} = \frac{m(Y_i, N_i) F(Y_i, N_i) Y_i F(Y_j, N_j) Y_j}{\sum_j F(Y_j, N_j) Y_j} U_{ij}$$

With the *log*-linear form for $m(\ )$ and $F(\ )$,

$$m(Y_i, N_i) = k_m Y_i^{m_y} N_i^{m_N}$$

and

$$F(Y_j, N_n) = k_\phi H_j^{\phi_y} N_j^{\phi_N}$$

and the denominator made a constant term we have

$$(8') \quad M_{ij} = (k_m Y_i^{m_y} N_i^{m_N})(k_\phi Y_i^{\phi_y} N_i^{\phi_N}) Y_i \cdot (k_\phi Y_j^{\phi_y} N_j^{\phi_N}) Y_j U_{ij} \div k'$$
$$= (k_m k_\phi^2) Y_i^{m_y + \phi_y + 1} N_i^{m_N + \phi_N} \cdot Y_j^{\phi_y + 1} N_j^{\phi_N} U_{ij} \div k'$$

This is the aggregate form of (1) with the distance term omitted. Ordinarily it would be fitted on a subset of countries in the world. Exports to the rest of the world are exogenous and imports from it are excluded from the fitting. When this is done, the denominator is still the sum of world trade expenditures, and (6′) implies that (8) and (8′) assume that Θ_i is the same in the excluded countries as in the included countries. Alternatively, (6′) can be interpreted as a payments union multilateral balance constraint (which includes as a special case the rest of the world account being always zero). The denominator of (8) and (8′) then has only the included group's trade expenditure.[8] Under either interpretation, the identifying restrictions immediately allow recovery of all structural exponents from the estimator of (8′). Either interpretation will also permit the complex constant term $(k_m k_\phi^2 / k')$ to be unravelled, though the estimators $\hat{k}_m, \hat{k}_\phi, \hat{k}'$ have only large sample unbiasedness.[9] Finally, form the set of estimated values for traded-goods expenditures:

$$(9) \qquad \hat{\phi}_j Y_j = \hat{k}_\phi Y_j^{\hat{\phi}_y + 1} N_j^{\hat{\phi}_N}$$

The individual traded-goods shares Θ_i can be estimated using the instruments $\hat{\phi}_j Y_j$ (which are asymptotically uncorrelated with U_{ij}):

[8] If neither of these alternatives is palatable, exports to the rest of the world can be fixed at $M_{i,n+1}$. Trade balance is now

$$(a) \qquad m_i \phi_i Y_i = \Theta_i \sum_j \phi_j Y_j + M_{i,n+1}$$

When the trade balance is solved for Θ_i and substituted, the gravity equation is

$$(b) \qquad M_{ij} = \frac{(m_i \phi_i Y_i - M_{i,n+1}) \phi_j Y_j}{\sum_j \phi_j Y_j}$$

Non-linear methods must be used to estimate (b).

[9] Consider the worldwide identity of preferences case. Using conditional expectations in (6′), the trade balance requirement implies that

$$k_m k_\phi Y^{1 + \phi_y + m_y} N_i^{\phi_N + m_N} = \sum_{j=1}^{n} E(M_{ij}) + M_{i,n+1}$$

where $M_{i,n+1}$ is the exogenous nonrandom export of i to the rest of the world. Replacing $E(M_{ij})$ with its solved values $\hat{M}_{ij}$, and the exponents with their estimated values, we have

(a)

$$\hat{k}_m \hat{k}_\phi = [\sum_{j=1}^{n} \hat{M}_{ij} + M_{i,n+1}]\ Y_i^{-(1 + \hat{\phi}_y + \hat{m}_y)} N_i^{-(\hat{\phi}_N + \hat{m}_N)}$$

Using the definition of k', we have

$$(b) \qquad \hat{k}' = \hat{k}_\phi (\sum_{j=1}^{n} Y_j^{1 + \hat{\phi}_y} N_j^{\hat{\phi}_N}) + \sum_{i=1}^{n} M_{i,n+1}$$

Finally the estimated constant term k, is theoretically related to the three constants $\hat{k}_m, \hat{k}_\phi, \hat{k}'$ by

$$(c) \qquad \hat{k} = \hat{k}_m \hat{k}^2 / \hat{k}'$$

(a)–(c) can be solved explicitly for the three constants $\hat{k}_m, \hat{k}_\phi, \hat{k}'$.

$$(10) \qquad M_{ij} = \hat{\Theta}_i \phi_j Y_j U_{ij}$$

which is estimated across countries for country i's exports (including the rest of the world's exports to included countries), subject to the restriction that $\Sigma \Theta_i = 1$. The alternative without the gravity model is to estimate the ϕ's by regressing $\Sigma_i M_{ij}/Y_j$ on $F(Y_j, N_j)$ and then repeating the second stage. The gravity equation in effect squares the number of observations used in estimating the parameters of $F(Y, N)$. For the limited number of cross-section observations available, the gain in efficiency should be large.

It is ironic, however, that the very simultaneity which allows substitution for Θ_i may imply that the Y's are mutually determined with the error terms of the expenditure system. The model (5')-(6') postulates no random term in the trade balance constraint and thus allows treatment of the Y's as predetermined. Suppose alternatively that the trade balance constraint is a stochastic form:

$$(6'') \qquad m_i(\sum_j M_{ji}) = \sum_j M_{ij}$$

$$\text{or} \qquad m_i \phi_i Y_i (\sum_j \Theta_j U_{ji}) = \Theta_i (\sum_j \phi_j Y_j U_{ij})$$

Then (8) becomes

$$(8'') \qquad M_{ij} = \frac{m_i \phi_i Y_i \phi_j Y_j}{\sum_j \phi_j Y_j} \epsilon_{ij}$$

$$\text{where} \quad \epsilon_{ij} = U_{ij} \sum_j \Theta_j U_{ji} / \sum_j \frac{\phi_j Y_j}{\sum_j \phi_j Y_j} U_{ij}$$

A regression based on (8'') with $m(\ )$ and $F(\ )$ assigned the multiplicative form will produce biased results (with unknown direction) due to the dependency of the Y's on the error terms. The relative stability of the equation over time in some applications may suggest that the bias is not serious, but this is conjectural. Several alternatives are possible, two of which will be discussed here. Something like the gravity equation would be desirable, replacing the Y's in the equation with the instruments highly correlated with the Y's but independent of the demand equation error terms. One such instrument might be lagged income, particularly for years when last year's income seems unlikely to be correlated with this year's error term. Bias remains, but may be reduced. The other alternative is to attempt dealing with simultaneity directly. Suppose a subset of countries is considered and preferences for traded goods are everywhere the same. Rest of the world demand is considered exogenous. Run the gravity equation using ordinary least squares on the *log* of the equation (or non-linear estimation of the equation with an additive disturbance term) and obtain estimates of ϕ_i, $\hat{\phi}_i$, and m_i, $\hat{m}_i$; $i = 1, \ldots, n$.[10] The trade balance equations in matrix notation are

$$(11) \qquad (diag\, \hat{m})(diag\, \hat{\phi})\, Y = (\hat{\Theta}\iota')(diag\, \hat{\phi}) \cdot Y + M_{.n+1}$$

where $M_{.n+1}$ = rest of the world demand, an $n \times 1$ vector

Y = $n \times 1$ vector of incomes

$(diag\, \hat{\phi})$ = $n \times n$ diagonal matrix with $\hat{\phi}_i$, $i = 1, \ldots, n$ on the diagonal

$\hat{\Theta}$ = $n \times 1$ vector of $\hat{\Theta}_i$, $i = 1, \ldots, n$

ι' = $1 \times n$ row vector of ones

$(diag\, \hat{m})$ = $n \times n$ diagonal matrix with $\hat{m}_i$, $i = 1, \ldots, n$ on the diagonal

or[11] $Y = (diag\, \hat{\phi})^{-1}[\mathrm{I} - \hat{\Theta}\iota']^{-1} M_{.n+1}$

The left-hand side contains instruments for the Y's which attempt to deal with the simultaneity problem, and which can then be inserted into the gravity equation and used to reestimate the m's, ϕ's, and Θ's. Continued iteration would have no necessarily desirable property.

Either instrument would probably be preferable if the simultaneity problem were severe, as in the European Economic Community (*EEC*). For groups of countries with

[10]Since $E(ln\ \epsilon_{ij}) \neq 0$, previous methods of identifying k_m and k_ϕ cannot be used. For simplicity, this detail can be evaded by assuming $k_m = k_\phi = 1$.

[11]Provided $\Sigma_i \Theta_i < 1$, $I - \Theta\iota'$ has an inverse. We deal with a subset of goods and countries, so the condition is fulfilled. $M_{.n+1} = \Theta \phi_{n+1} Y_{n+1}$ with the assumption of identical preferences for the rest of the world.

relatively small interdependence, the gravity equation with the Y's directly used might be preferable, the greater efficiency of direct use of Y's dominating the bias. These are, of course, only rules of thumb.[12]

IV. Many Goods, Tariffs, and Distance

Now consider the gravity equation under the complication of many commodity classes of goods flowing between each country i and j, with a full set of national tariffs in each country, and with transport costs proxied by distance. Preferences for traded goods are identical across countries and are homothetic, with the traded-goods share, as before, a function of income and population. Within each commodity class, goods are considered to be differentiated by place of origin.[13] The gravity equation still has use in the estimation of trade-flow equations of this system. As in Section III, it is a device for increasing the efficiency of estimation of the trade-share function parameters. Unfortunately, tariffs and transport costs create added sources of bias in estimation of both stages. The gain may still outweigh the loss.

The landed value at country j of commodity class k goods produced in country i is $M_{ijk}\tau_{ijk}$, where M_{ijk} is the foreign port value and τ_{ijk} is the transit cost factor (including all border adjustments and transport costs). With identical homothetic preferences for traded goods, the traded-goods expenditure shares are identical functions $\Theta_{ik}(\tau_j)$, where τ_j is the vector of the τ_{ijk}'s for country j. Demand for import ik (with foreign port prices of unity as before) is

$$M_{ijk} = \frac{1}{\tau_{ijk}} \Theta_{ik}(\tau_j)\phi_j Y_j \tag{12}$$

Aggregate trade flows between i and j are thus

$$M_{ij} = \sum_k M_{ijk} = \phi_j Y_j \sum_k \frac{1}{\tau_{ijk}} \Theta_{ik}(\tau_j) \tag{13}$$

The trade balance relation is

$$m_i\phi_i Y_i = \sum_j M_{ij} = \sum_j \phi_j Y_j \sum_k \frac{1}{\tau_{ijk}} \Theta_{ik}(\tau_j) \tag{14}$$

Previously we set all the $\tau_{ijk} = 1$ and could divide both sides of (14) by $\Sigma_j \phi_j Y_j$ to obtain the aggregate share parameter for country i goods on the right: $\Sigma_k \Theta_{ik}$. The left-hand side was then substituted into (13) to obtain the gravity equation

$$M_{ij} = \frac{m_i\phi_i Y_i \phi_j Y_j}{\sum_j \phi_j Y_j} \tag{8}$$

Note that with many goods, only the aggregate version of the gravity equation is valid under the present interpretation.[14]

[12]We should note that in principle minimum-distance or full-information maximum-likelihood techniques can be used to estimate the parameters of system (5′)-(6′) and its generalization. Any model which can be solved to isolate its disturbance terms can be so treated (see A. R. Gallant). The costliness and convergence difficulties with such non-linear techniques makes compromises like those in the text attractive.

[13]Isard offers an empirical justification for this assumption. On a theoretical level, note that the gravity model almost necessarily implies differentiation by place of origin. How else can (i) two-way flows be explained, and (ii) the flow of good k between points i and j be modelled as a function of variables at i and j alone?

[14]Leamer has extended the gravity-type cross-section model to aggregation across partner countries j, estimating aggregate outward flows M_{ik}. His model shares with the present interpretation unidentified reduced-form trade potential functions similar to the trade-share functions above. It is further unidentified because Leamer gives no precise economic reason for the appearance and form of the appearance in the equation of trade potential at both ends of the trade flow. Nevertheless, it has a certain plausibility and allows extension to the estimation of tariff elasticities. In correspondence concerning an earlier version of this paper, Leamer suggested a method of obtaining commodity-specific gravity equations not involving the trade potential interpretation. As before, the demand function is

$$M_{ijk} = \Theta_{ik}\phi_j Y_j \tag{a}$$

Country i's income from sales of the ik good is

$$Y_{ik} = \Theta_{ik}\sum_j \phi_j Y_j \tag{b}$$

Suppose that the commodity classes k are defined so that income from their production across countries

With the τ_{ijk} departing from unity the division of both sides of (14) by $\Sigma_j \phi_j Y_j$ produces

$$(15)\quad \frac{m_i \phi_i Y_i}{\sum_j \phi_j Y_j} = \sum_j \frac{\phi_j Y_j}{\sum_j \phi_j Y_j} \cdot \sum_k \frac{1}{\tau_{ijk}} \Theta_{ik}(\tau_j)$$

The gravity equation substitutes for the share in (13), $\Sigma_k (1/\tau_{ijk}) \Theta_{ik}(\tau_j)$, a weighted average of such shares across all countries j. This will cause bias of unknown sign in the gravity equation parameter estimator based on the stochastic version of (12)–(14), and subsequently in the parameter estimator of the demand equation (12).[15] Other factors being equal, the bias will be less the more closely the transit costs resemble one another. In the limit, we return to the model of Section III. Evidently, similarity of transit costs should be a criterion for selecting countries in the cross-section sample. This is too stringent a criterion to permit the viability of the gravity model and flies in the face of the role it assigns to distance. To breathe life back into it, we can argue that with dissimilarity of restricted types it may still be possible to escape with small bias.

If transit costs of all sorts are an increasing function of distance and the same across commodities ($\tau_{ijk} = f(d_{ij})$ with $f(0) = 1$ and $f' > 0$), then with Cobb-Douglas preferences the demand equation and trade balance equations are

$$(13')\quad M_{ij} = (\sum_k \Theta_{ik}) \phi_j Y_j \frac{1}{f(d_{ij})} U_{ij}$$

$$(15')\quad m_i \phi_i Y_i = (\sum_k \Theta_{ik}) \sum_j \phi_j Y_j \frac{1}{f(d_{ij})}$$

Equation (13′) states that the foreign port value of country j's demand for all of i's goods equals country j's total expenditure on traded goods (in home prices), $\phi_j Y_j$, times the common aggregate traded-goods expenditure share for i's goods $\Sigma_k \Theta_{ik}$ deflated by the transit cost factor. Equation (15′) states that country i's expenditure on all traded goods at i's prices $\phi_i Y_i$ times the capital account scale factor m_i must equal the value at country i of i's exports to all countries. The gravity equation can now be derived as

$$(16)\quad M_{ij} = \frac{m_i \phi_i Y_i \phi_j Y_j}{\sum_j \phi_j Y_j} \cdot \frac{1}{f(d_{ij})} \cdot \left[\sum_j \frac{\phi_j Y_j}{\sum_j \phi_j Y_j} \cdot \frac{1}{f(d_{ij})} \right]^{-1} U_{ij}$$

With m and the ϕ's made *log*-linear functions of income and population, (16) resembles (1), with three differences. First, (16) is an aggregate equation rather than commodity specific. Second, $1/f(d_{ij})$ is not a *log*-linear function.[16] Finally, the square bracket term is missing in (1). It can be interpreted as saying that the flow from i to j depends on economic distance from i to j relative to a trade-weighted average of economic distance from i to all points in the system. The model leading to (16) is probably the best case one can make for the aggregate gravity equation as it is usually fitted in practice. The square bracket term might have little variation across origin points i for a group of countries distributed geographically in a polygon (for example, the *EEC*). Changing origin point i will lengthen some distances and shorten others, with the potential for little change in the weighted average. With small enough bias,

is a stable function of *GNP*, population and resource endowments E_i:

$$(c)\qquad Y_{ik} = \gamma^k(Y_i, N_i, E_i)\, Y_i$$

Substituting (b) and (c) into (a) we obtain a gravity form:

$$(d)\qquad M_{ijk} = \frac{\gamma^k(Y_i, N_i, E_i)\, Y_i \phi_j(Y_j, N_j, E_j)\, Y_j}{\sum_j \phi_j Y_j}$$

This may be a promising approach, although the stability of the γ^k functions is probably more controversial than the stability of the ϕ_j functions.

[15]The demand equation as before would be estimated using $\hat{\phi}_j Y_j$ as an instrument. Note that the parameters to be estimated would in principle include substitution parameters.

[16]Practitioners of the gravity model use it only because it is so convenient, and some have adopted more theoretically appealing forms.

the greater efficiency of the gravity equation (1) in aggregate form in arriving at estimates of $\phi_j Y_j$ dominates,[17] and the Θ's can be estimated from

$$(12') \quad M_{ijk} = [(\widehat{\frac{1}{f(d_{ij})}}) (\hat{\phi}_j Y_j)] \, \Theta_{ik} U_{ijk}$$

If the bias from omitting the bracketed term is likely to be substantial, (16) can be estimated with constant weights in the bracket term equal to observed trade total expenditure shares. Non-linear least squares is required, implying some loss in efficiency. Which procedure is preferable depends on the tradeoff.

Practitioners may be able to get away with restrictions less extreme than either the identical transit costs or Cobb-Douglas assumptions. Consider the *CES* preference case where trade taxes are the same across all countries j for any good k of country i (often an assumption which comes close to reality, as in the *EEC*). Assume transport cost factors depend only on distance (not on commodity group). Under these conditions, the Appendix shows that a gravity equation may still have some promise of providing efficiency gains which dominate bias. The *CES* demand functions may be estimated as above in a second stage using the instrument $\phi_j \hat{Y}_j$. In principle, a trade flow system in the gravity model style capable of dealing with tariffs and even possibly with policy-induced change in shares can be developed.

[17] Disregarding bias, the procedure for solving out the parameters of the $m(\ )$ and $F(\ )$ functions are essentially the same as in Section II. The only new factor is the presence of $f(d_{ij})$. If we adopt the *log*-linear form, (16) assures us that any constant term it possesses is cancelled out (i.e., if $f(d_{ij}) = k_d d_{ij}^{\mu}$, the k_d term would not appear in the general constant term of the estimator of (16)). The constant term k_d can be identified by noting that

$$\text{(e)} \quad \frac{\hat{M}_{ij}}{\hat{M}_{ii}} = \frac{Y_j^{\hat{\phi}_y + 1} N_j^{\hat{\phi} N}}{Y_i^{\hat{\phi}_y + 1} N_i^{\phi \hat{N}}} \frac{1}{k_d d_{ij}^{\hat{\mu}}}$$

Equation (a) of fn. 9 becomes

$$\text{(a')} \quad \hat{k}_m \hat{k}_\phi \hat{k}_d = [\sum_{j=1}^{\hat{}} M_{ij} + \hat{M}_{i,n+1}] \, Y_i^{-(1+\hat{\phi}_y+\hat{m}_y)} N_i^{-(\hat{\phi}_N+\hat{m}_N)} d_{ij}^{-\hat{\mu}}$$

Equations (a'), (e), as well as (b) and (c) of fn. 9 can be solved for all constant term estimates. Other distance functions will require other identifying restrictions.

V. Conclusion

The gravity equation can be derived from the properties of expenditure systems. In this interpretation it is an alternative method of doing cross-section budget studies, and one with potentially important efficiency properties. Its use is at the widest limited to countries where the structure of traded-goods preference is very similar and, subsidiarily, where trade tax structures and transport cost structures are similar. In future work, it would be desirable to learn more about the tradeoff between bias and efficiency involved in the gravity equation. Other extensions include building an intertemporal version and identifying the trade-share function.

Appendix: The *CES* Case

The *CES* traded goods utility indicator is

$$U_j = [\sum_i \sum_k \beta_{ik} M_{ijk}^{-\rho}]^{-1/\rho}$$

where M_{ijk} is the quantity of good k from country i consumed in country j. Good k is a different commodity in each country due to differentiation. The elasticity of substitution is $\sigma = 1/(1+\rho)$. Expenditure shares derived from such a utility function are

$$\text{(A1)} \quad \Theta_{ijk} = \frac{\beta_{ik}(P_{ijk})^{1-\sigma}}{\sum_i \sum_k \beta_{ik}\sigma(P_{ijk})^{1-\sigma}}$$

where Θ_{ijk} is the traded-goods expenditure share of country j for good k of country i, and P_{ijk} is the price of good k from country i landed in country j. The denominator of (A1), when raised to the power $1/(1-\sigma)$, gives the "true cost-of-living" index for the *CES* function. The demand for imports is

$$\text{(A2)} \quad M_{ijk} = \Theta_{ijk} \phi_j Y_j \frac{1}{P_{ijk}}$$

Derivations are standard, so omitted.

Assume now that the transit cost factors are based on two components. The first is t_{ik}, a tax on good k of country i levied by all countries in the group. The second is a transit cost factor, common to all goods and dependent on distance, $h(d_{ij})$.

$$\text{(A3)} \qquad \frac{P_{ijk}}{P_{ik}} = t_{ik}h(d_{ij})$$

Define the "free trade" share:

$$\text{(A4)} \qquad \Theta_{ik} \equiv \frac{\beta_{ik} P_{ik}^{1-\sigma}}{\sum_i \sum_k \beta_{ik}^{\sigma} P_{ik}^{1-\sigma}}$$

Using (A3) and (A4) in (A1) and (A2) the demand equation can be written

$$\text{(A5)} \quad M_{ijk} = \Theta_{ik} \left[\frac{\dfrac{[t_{ik}h(d_{ij})]^{1-\sigma}}{\sum_i \sum_k \beta_{ik}^{\sigma} P_{ik}^{1-\sigma}(t_{ik}h(d_{ij}))^{1-\sigma}}}{\sum_i \sum_k \beta_{ik}^{\sigma} P_{ik}^{1-\sigma}} \right] \cdot \phi_j Y_j \frac{1}{P_{ik} t_{ik} h(d_{ij})}$$

The denominator of the large square bracket term is a weighted average of the transit cost factors, and equals a transit cost true cost-of-living index for country j raised to the power $1 - \sigma$. Denote this as $g_j^{1-\sigma}$. Simplifying (A5) and using the convention that free trade prices are unity:

$$\text{(A5}'\text{)} \quad M_{ijk} = \Theta_{ik} g_j^{-(1-\sigma)} (t_{ik}h(d_{ij}))^{-\sigma} \phi_j Y_j$$

The trade balance requirements are

$$\text{(A6)} \quad m_i \phi_i Y_i = \sum_j \sum_k \Theta_{ik} g_j^{-(1-\sigma)} \cdot (t_{ik}h(d_{ij}))^{-\sigma} \phi_j Y_j$$

Aggregate trade flows between i and j are

$$\text{(A7)} \quad M_{ij} = \phi_j Y_j g_j^{-(1-\sigma)} \; h(d_{ij})^{-\sigma} \sum_k \Theta_{ik} t_{ik}^{-\sigma}$$

The gravity equation substitution replaces $\Sigma_k \Theta_{ik} t_{ik}^{-\sigma}$ in (A7) with $m_i \phi_i Y_i / \Sigma_j \phi_j Y_j$. The proper substitution yields

$$\text{(A8)} \quad M_{ij} = \left[\frac{g_j^{-(1-\sigma)}}{\sum_j \dfrac{\phi_j Y_j}{\sum_j \phi_j Y_j} g_j^{-(1-\sigma)} h(d_{ij})^{-\sigma}} \right] \cdot \frac{m_i \phi_i Y_i \phi_j Y_j}{\sum_j \phi_j Y_j} [h(d_{ij})]^{-\sigma}$$

The gravity equation run on the stochastic version of (A8) omitting the square bracket term has a chance of reasonably small bias in the estimator if there is little variation in the square bracket term as we move across i and j. Note that with free trade prices of unity:

$$\text{(A9)} \quad g_j^{1-\sigma} = \frac{\sum_i \sum_k \beta_{ik}^{\sigma} (t_{ik})^{1-\sigma} h(d_{ij})^{1-\sigma}}{\sum_i \sum_k \beta_{ik}^{\sigma}}$$

The denominator of the square bracket is a weighted sum of $h(d_{ij})^{-\sigma}$ across j for a given i. The numerator is a weighted sum of $h(d_{ij})^{-\sigma}$ across i for a given j. Changing origin points i and destination points j may well create changes which wash out (as in the related square bracket term of (16) in the text). This is more likely for countries geographically distributed in a polygon (and impossible for countries distributed on a line). We might thus hope to gain more in efficiency than we lose in bias by estimating the stochastic form of (A8) omitting the bracket term. The parameters of the $m(\ )$ and $F(\ )$ functions can be identified as in the text, and the instruments $\hat{\phi}_j Y_j$ used to attack the stochastic form of the demand equation (A5′). Note that in estimating (A5′) we might again appeal to the lack of variation in $g_j^{1-\sigma}$ to produce simpler estimation techniques capable of producing asymptotically "unbiased" estimates of σ.

Other alternatives include approximation of g_j in the numerator of (A8) with a Laspeyres traded-goods price index, and full non-linear estimation of the stochastic form of (A8).

REFERENCES

N. D. Aitken, "The Effect of the *EEC* and *EFTA* on European Trade: A Temporal Cross-Section Analysis," *Amer. Econ. Rev.*, Dec. 1973, *63*, 881–92.

H. B. Chenery, "Patterns of Industrial Growth," *Amer. Econ. Rev.*, Sept. 1960, *50*, 624–54.

A. R. Gallant, "Three Stage Least Squares Estimation for a System of Simultaneous, Non-Linear, Implicit Equations," *J. Econometrics*, Feb. 1977, *5*, 71–88.

P. Isard, "How Far Can We Push The Law of One Price?," *Amer. Econ. Rev.*, Dec. 1977, *67*, 942–48.

Edward E. Leamer and Robert M. Stern, *Quantitative International Economics*, Boston 1970.

———, "The Commodity Composition of International Trade in Manufactures: An Empirical Analysis," *Oxford Econ. Pap.*, Nov. 1974, *26*, 350–74.