Notes on Gravity

James E. Anderson
Boston College and NBER

July 27, 2007

The gravity model is quite successful in explaining world trade patterns across countries. It is even more useful for inferring trade costs that are quite surprisingly large. These notes set out the gravity model and its implications. The model is applied to explain trade patterns and then to draw useful inferences about trade costs.

The main elements of the model are the description of demand and the requirement that markets clear. Goods originate at each origin (location) and tend to spread out to all destinations (locations). Even when the goods are similar, the pattern of shipments shows that countries or regions tend to buy from many sources. The assumption of the model is that there are tastes for goods that differ according to their origin. For example, consumers like California wine, Chilean wine, NZ wine, French wine and so forth. The model is completed by requiring that the world market for each type of wine must clear, total demand must equal the supply. The model is often applied at the aggregate level, so the good becomes GDP of California, Chile, NZ and so forth. These notes concentrate on the aggregate version.

The first section sets out the gravity model in a world without trade costs, a frictionless world. This is useful in isolating the powerful GDP and expenditure forces that drive trade even in the much more complex world of costly trade. The second section develops the full economic theory of gravity and draws the main implications. The third section draws together inferences about trade costs from recent intensive empirical investigation.

1 Frictionless Gravity

Each bilateral trade flow from origin i to destination j, the value of shipments of i's GDP exported to j, is denoted as X_{ij}. The demand assumption is that the expenditure on the country i good at each destination j is given by

$$X_{ij} = \alpha_i E_j.$$

Here α_i is a globally common spending share on goods from i, taken temporarily as a constant fraction. E_j is the expenditure of j. At the margin, an additional dollar of total expenditure results in α_i cents worth of spending on goods from country i. Let Y_i denote the GDP of i. If balanced trade is imposed (no international borrowing or lending), then $E_j = Y_j$.

The market clearance requirement implies that

$$\sum_j X_{ij} = Y_i.$$

Now derive the gravity model as follows. Replace X_{ij} with $\alpha_i E_j$ in the market clearance equation and solve for the expenditure share:

$$\alpha_i = Y_i / \sum_j E_j.$$

Next, use the world budget constraint: $\sum_j E_j = \sum_i Y_i = Y$. Substituting Y_i/Y for α_i in the demand equation yields the frictionless gravity equation:

$$X_{ij} = Y_i E_j / Y. \tag{1}$$

A slightly more intuitive form of (1) is useful in drawing out the implications. Define $s_i = Y_i/Y$, country i's share of world GDP. With balanced trade, $E_j = Y_j$. Then the frictionless gravity equation can be written

$$X_{ij} = s_i s_j Y.$$[1]

There are a number of important and useful implications of gravity in this form.

1. smaller countries are more naturally open. Country j's international trade to GDP ratio is given by

   $$\sum_{i \neq j} X_{ij}/Y_j = 1 - s_j$$

   which is decreasing in s_j.

2. Any country trades more with bigger partners.

3. Country pairs that are growing relative the the rest of the world take up an increasing share of the world's trade while pairs that are growing at lower than average rates take up a decreasing share.

   $$\frac{X_{ij}}{\sum_{i \neq j} X_{ij}} = \frac{s_i}{1 - s_j}.$$

   This expression is increasing in both s_i and s_j.

[1] With unbalanced trade, the frictionless gravity equation is written $X_{ij} = s_i e_j Y$ where $e_j \equiv E_j/Y$, country j's share of world expenditure. The implications are similar to those in the text.

This last implication is striking in thinking about China. Quite apart from any role of prices, cheap Chinese labor, devious export subsidies, 'unfair' trade practices, the model implies growth in bilateral trade with China as a share of global trade. China grows at more than 10 per cent per year in a world where countries like the US manage 3 per cent in a good year and where the US tends to be above the OECD average. China's share of world GDP is rising, that of the US is falling, and that of other rich countries is falling faster.

All three implications of the model are roughly in accord with the trade data. However, the data also strongly suggest that there is far too little trade for the frictionless model to be accurate. The US produces 25 per cent of world GDP, so according to the frictionless model, by implication 1, it should export 75 per cent of its GDP instead of some 10 per cent in recent years. (Some commentators argue that the ratio of merchandise exports to goods GDP is more meaningful. That ratio was 22 percent in 2000, and has grown much more quickly than the trade to GDP ratio. But it is still far too low to represent a frictionless world.) The natural place to look for an explanation is trade costs. This line of thinking is greatly encouraged by noting that countries close together come closer to the frictionless prediction than do countries far away.

The early empirical work on the gravity model related the actual trade flows to the theoretical prediction in (1) and tacked on variables that relate to bilateral trade costs. For example, incorporating the effect of bilateral distance D_{ij}, the econometric model becomes

$$X_{ij} = \frac{Y_i E_j}{Y} D_{ij}^{\delta}$$

Econometric estimates of the coefficient δ range from -0.7 to -1.2, centering around -1 in a variety of studies. Some basic calculations indicate that distance is far more powerful than its role in freight rates would explain. Evidently there are other more important types of trade costs that vary with distance. Without the ability to examine the books of firms that trade, it is not possible for the empirical economist to really discover the size and nature of these costs, so at this point the inferential work becomes a bit more tenuous.

Slightly more complex versions of empirical gravity allow for exponents on Y_i and E_j that need not equal unity (though estimates of these coefficients do often come close to one). More importantly, the more complete versions allow for other variables that intuitively might affect trade costs. For example, contiguity of countries increases their trade, all else equal. So does a previous colonial tie, common language, density of international communications links, immigrant ties (e.g., the proportion of population that is ethnic Chinese), membership in a currency union, membership in a free trade association, ... A reputation for corruption or for unfair court systems reduces bilateral trade. Presumably, foreigners are treated worse than locals, so there is a differential effect that acts in effect on the border.

One simple variant of empirical gravity collapses all the additional cost indicator variables into the presence or absence of an international border. In that case, borders are extremely powerful destroyers of trade, way bigger than trade policy can account for. This is known as the border puzzle, and was the cause of intense investigation following a famous study of the trade of US states and Canadian provinces. McCallum (1995) used the logarithmic variant of gravity with distance and the border as trade cost indicators. Let x, y, e, d denote the natural logarithms of X, Y, E, D. Then McCallum's regression was:

$$x_{ij} = ay_i + be_j + \delta d_{ij} + \beta B_{ij} + \epsilon_{ij}.$$

Here the variable B_{ij} has the value 1 if there is a border between region i and region j, and 0 otherwise. The the random error term ϵ is assumed to be log-normally distributed. McCallum calculated, based on his fitted regression that estimated values for a, b, δ, β, that trade between Canadian provinces was *22 times* larger than trade between a Canadian province and a US state at the same distance and with the same income! A huge puzzle indeed.

To begin to get a handle on what might be going on, it is natural for an economist to look more deeply into the economic theory that gravity only loosely reflects in this work. It turns out that this exercise, undertaken by Anderson and van Wincoop (2003) pays off big time.

2 The Economic Theory of Gravity

The delivered price of a good incorporates trade costs. Even though the factory gate price of goods from i is the same for buyers in all locations j, the trade costs differ and cause differences in the buyers' prices. The price for goods from i at the factory gate is denoted p_i, the trade cost from i to j results in a markup factor $t_{ij} > 1$ so the buyers' price is given by $p_i t_{ij}$.

Since the buyer's price differs by location, it is likely to cause differences in the expenditure shares in each location, depending on the substitutability of the products of the countries for each other. A very convenient representation of the substitution possibilities yields the expenditure share for goods from i in destination j as

$$\alpha_{ij} = \frac{(p_i t_{ij})^{1-\sigma}}{P_j^{1-\sigma}}, \tag{2}$$

where σ is the elasticity of substitution parameter and

$$P_j = [\sum_i (p_i t_{ij})^{1-\sigma}]^{1/(1-\sigma)}, \tag{3}$$

the cost of living (consumer price) index. Each good substitutes for the goods from all other locations in a symmetric fashion. There is good reason to believe that $\sigma > 1$.[2] The implication of (2) is that higher relative trade cost t_{ij}between i and j reduces the share, all else equal, while uniformly higher trade costs (all the t_{ij}'s , including t_{ii}, rise in the same proportion) leave the shares unchanged.

In deriving the gravity equation (1), it is no longer possible to substitute out the share parameter since this share α_{ij} differs in each location j. It *is* possible to substitute out the common element p_i using the market clearing condition. The end result (see the Appendix for details) is that the demand equation $X_{ij} = \alpha_{ij} E_j$ becomes the economic gravity equation

$$X_{ij} = \frac{t_{ij}^{1-\sigma}}{(\Pi_i P_j)^{1-\sigma}} \frac{Y_i E_j}{Y}.$$

Here, Π_i is an index of the trade costs that origin i faces on its shipments to every destination (outward multilateral resistance) while P_j is an index of the trade costs that destination j faces on its shipments from every origin (inward multilateral resistance). At the same time P retains its interpretation as a consumer price index.

Trade frictions modify the frictionless gravity model represented by the rightmost ratio on the right hand side of the equation. The bilateral friction t_{ij} enters the economic gravity model relative to the product of inward (P_j) and outward (Π_i) multilateral resistance. Intuitively, the bilateral cost of

[2]There is very good evidence that $\sigma > 1$. Gravity model investigations indicate a value ranging from 6 to 10.

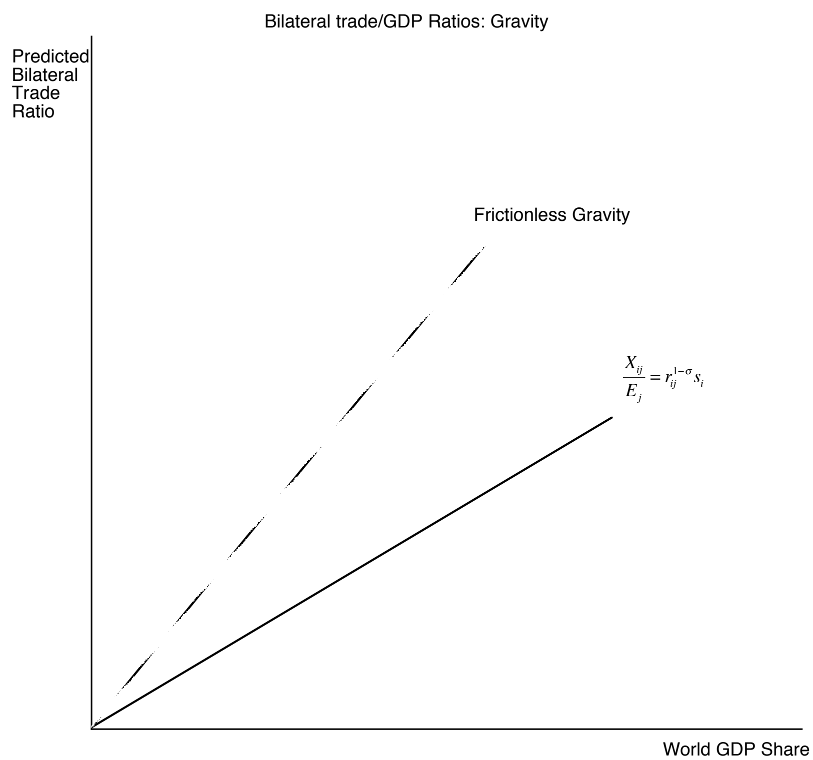
shipping outward from i to j affects the volume of bilateral trade relative to the average cost of shipping everywhere else, captured by outward multilateral resistance. A higher average cost of shipping will tip more trade into the link from i to j. Similarly, the bilateral cost impacts the demand side relative to the cost of shipping from all other origins; high average inward cost at j tends to shift more spending onto goods from i.

It can be shown that smaller countries tend to have higher multilateral resistance. Intuitively this is because a higher proportion of their trade must come from the rest of the world, from whom it is more expensive to buy. This factor tends to offset their greater natural openness arising in a frictionless world. The actual data suggests, however, that smaller countries continue to be more open. In some very special cases it is possible to prove mathematically that this must be so.

The gravity model can be neatly illustrated by considering the bilateral trade to GDP ratio. Let $r_{ij} = t_{ij}/\Pi_i P_j$. Then

$$\frac{X_{ij}}{E_j} = r_{ij}^{1-\sigma} s_i.$$

The diagram shows the predicted relationship referenced by the frictionless relationship where $r_{ij} = 1$.



If r_{ij} did not differ by trading partners i, then country j's data points would lie on the solid line. The dispersion of the data points away from the solid line based on an average value of r informs the econometric analysis about implicit trade costs. Locations i that are closer to j would presumably be above the line, for example, while further away locations would be below the line.

Relative trade flows are usefully related to relative GDP's in another useful manipulation of the gravity equation. For example, the exports of an individual EU country to the US are given by

$$X_{i,US} = \frac{Y_i E_{US}}{Y} r_{i,US}^{1-\sigma}.$$

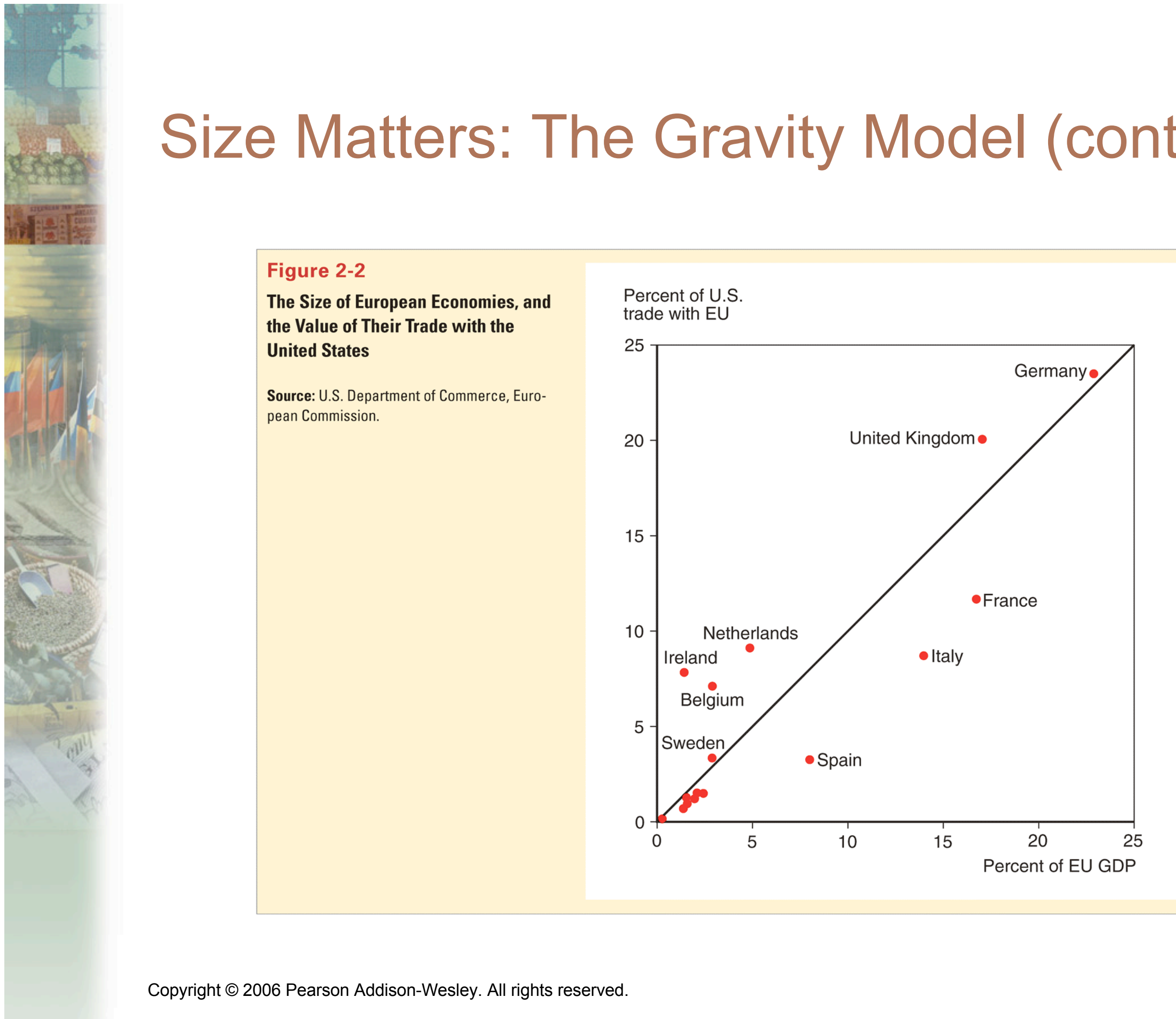
The total exports of the EU to the US are given by

$$X_{EU,US} = \frac{Y_{EU} E_{US}}{Y} \bar{r}_{EU,US}^{1-\sigma},$$

which is obtained by summing the preceding equation across EU countries and defining an average EU to US relative resistance. Taking the ratio of the second equation to the first yields:

$$\frac{X_{i,US}}{X_{EU,US}} = \frac{Y_i}{Y_{EU}} \frac{r_{i,US}^{1-\sigma}}{\bar{r}_{EU,US}^{1-\sigma}}.$$

This relationship is illustrated in Figure 2-2 in Krugman and Obstfeld (2005) for EU to US trade in 2003.



With uniform frictions over countries, the data points would lie on a 45 degree line. The deviations are due to relative resistance varying among the EU members. The US value of P_{US} is a constant for all EU members, so it cancels out of the calculation. Relative resistance varies in the EU to US trade because (i) bilateral resistance varies over EU to US pairs and (ii) outward multilateral resistance varies over the EU countries. As to bilateral resistance (i), countries with good infrastructure oriented toward the western coast of Europe are likely to have lower trade costs. As to multilateral resistance (ii), smaller countries have higher multilateral resistance, lowering their relative resistance. Now consider the outlying data points in Figure 2-2. Both (i) and (ii) operate to increase the EU to US trade share of Ireland, Netherlands and Belgium. Both (i) and (ii) operate to decrease the EU to US trade share of Spain, Italy and France. The two factors tend to offset for the EU to US trade share of the UK and Germany, with (i) apparently dominating (ii). For Sweden, the trade decreasing effect of factor (i) is offset by the trade increasing effect of factor (ii). (Sweden is economically oriented toward the Baltic Sea.)

The biggest consequence of the higher multilateral resistance of smaller regions is that it resolves the border puzzle. McCallum's regression is missing the multilateral resistance variables. This raises two issues. First, there is omitted variable bias in his regression. An appropriate method of controlling for multilateral resistance is to use fixed effects for regions, one each for inward and outward trade. Adding these regressors reduces the estimated border effect somewhat.

The second issue with McCallum's regression is one of useful interpretation of the border effect. The border coefficient β in McCallum's regression gives the effect of trading across the border relative to trade within borders. The theoretical gravity model implies that

$$\frac{X_{ij}}{X_{ik}} = \frac{t_{ij}^{1-\sigma}}{t_{ik}^{1-\sigma}} \frac{\Pi_i^{1-\sigma} P_k^{1-\sigma}}{\Pi_i^{1-\sigma} P_j^{1-\sigma}} \frac{Y_i E_j}{Y_i E_k}$$

for a bilateral pair of regions i and j within the same country (e.g. two Canadian provinces) relative to a bilateral pair of regions i and k in different countries (e.g., Canadian province i and a US state k). For pairs the same distance apart and with the same total expenditure levels, the right hand side of the preceding expression reduces through canceling like terms to

$$\frac{1}{\tau^{1-\sigma}} \frac{P_k^{1-\sigma}}{P_j^{1-\sigma}}.$$

Here τ is the implicit cost factor associated with crossing the border. McCallum's border effect thus combines both the effect of a 'pure' border cost and the effect of borders on relative multilateral resistance. Recall that Canada is about 1/10 the size of the US in GDP (and population). That means

Canadian provinces must naturally do far more international trade than do US states. This drives up the multilateral resistance of Canadian provinces relative to similar US states, raising the value $P_k^{1-\sigma}/P_j^{1-\sigma}$ in the preceding expression.

Anderson and van Wincoop (2003) use this setup to recalculate the effect of the border. In their results, τ is calculated to be around 1.4; equivalent to a 40 per cent 'tax' imposed by crossing the border. If $\sigma = 6$, a reasonable estimate in the context of the gravity model, then the effect of the border is to multiply within border trade relative to cross border trade by a factor 5.4. The effect of higher average multilateral resistance in Canada acts to roughly double this 'pure' effect of the border.

Is there still a border puzzle? A tax equivalent of 40 per cent on the open Canada-US border with free trade (mostly), a common language, common legal traditions, long familiarity and so forth does still seem puzzling. But the result along with many other gravity type investigations may also be taken to imply that international trade costs are 'surprisingly' high.

3 Estimating Trade Frictions

The t_{ij}'s can be inferred from the deviations of actual trade flows from the frictionless benchmark. These deviations can in turn be related statistically to such variables as distance, freight rates, tariffs, common language, common legal tradition, the security of property, contiguity, density of communications links and the like. The results can be converted to 'tariff equivalents'.

Anderson and van Wincoop (2004) contains a summary of the implications of the preceding 15 years of investigation on this topic.

Representative OECD total cost equivalent = 170%

- 21% transportation costs, (direct plus 9% time value of transport)
- 44% border related trade barriers
- 55% retail and wholesale margins
- total 1.7=1.21*1.44*1.55-1

Border Barriers

- 8% direct policy barrier

- 7% language barrier
- 14% currency barrier (from the use of different currencies)
- 6% information cost barrier
- 3% security barrier
- 44% total $=1.08 * 1.07 * 1.14 * 1.06 * 1.03 - 1$
- Overall range of g-model border barriers comparison (Table 7): 25-50%

Puzzles

- why does distance matter so much?
- why do borders matter so much?
- currency barrier puzzle

Heterogeneity of Trade Costs

- Bigger for non-OECD countries: Latin Am policy 16%, security 16%; Argentina domestic margin 80%
- Commodity line variation: Tables 2-6
- costs of trading firms?

4 References

Anderson, James E. and Eric van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle", *American Economic Review*, 93, 170-92.

Anderson, James E. and Eric van Wincoop (2004), "Trade Costs", *Journal of Economic Literature*, 42, 691-751.

Krugman, Paul R. and Maurice Obstfeld (2005), *International Economics: Theory and Policy*, 7th Edition, Addison-Wesley.

McCallum, John (1995), "National Borders Matter: Canada-U.S. Regional Trade Patterns", *American Economic Review*, 85, 615-23.

5 Appendix: Derivation of the Economic Gravity Equation

Use the market clearance equation to solve for the factory gate price as follows. The sum of expenditures equals the receipts of the seller (at destination prices) in equilibrium, meaning

$$Y_i = p_i^{1-\sigma} \sum_j \frac{t_{ij}^{1-\sigma}}{P_j^{1-\sigma}} E_j. \tag{4}$$

Divide both sides of (4) by Y. Then the market clearance condition becomes

$$\frac{Y_i}{Y} = (p_i \Pi_i)^{1-\sigma} \tag{5}$$

where by definition

$$\Pi_i^{1-\sigma} \equiv \sum_j \frac{t_{ij}^{1-\sigma}}{P_j^{1-\sigma}} \frac{E_j}{Y}. \tag{6}$$

Notice that Π_i is an index of the trade costs that origin i faces on all of its trade. For this reason it is called outward multilateral resistance.

The market clearance condition (5) can be solved for

$$p_i^{1-\sigma} = \frac{Y_i}{Y} \frac{1}{\Pi_i^{1-\sigma}}.$$

Now substitute the right hand side expression for p_i in the share equation (2). The demand equation $X_{ij} = \alpha_{ij} E_j$ becomes the economic gravity equation

$$X_{ij} = \frac{t_{ij}^{1-\sigma}}{(\Pi_i P_j)^{1-\sigma}} \frac{Y_i E_j}{Y}. \tag{7}$$

Just as Π is an index of outward trade costs, P is an index of inward trade costs and the two are systematically paired in equilibrium. This may be seen as follows. Substitute for p_i in the price index expression (3). This yields

$$P_j = \sum_i \frac{t_{ij}^{1-\sigma}}{\Pi_i^{1-\sigma}} \frac{Y_i}{Y}. \tag{8}$$

The expression on the right hand side of (8) is symmetric to (6). It is an index of the trade costs faced by buyers at location j from all origins. For this reason it is called inward multilateral resistance.

The multilateral resistances (for N countries there are $2N$ variables) can be solved from the system of (6), (8) (for N countries there are $2N$ equations), given the t's and the GDP and expenditure shares. A small technical issue arises because the P's and Π's are determined only up to a scalar by this system. The market clearance condition (5) can be summed on both sides to yield a normalization condition that further restricts the solution for the Π's, provided that the p's are known. Alternatively, one of the p's can be set equal to one, with (5) used to determine the corresponding Π.