OPTIMAL AGGREGATION OF LINEAR SYSTEMS

By

Edward E. Leamer
UCLA

Discussion Paper Number 240
April 1982

OPTIMAL AGGREGATION OF LINEAR SYSTEMS

Edward E. Leamer

The econometric analysis of highly disaggregated systems can produce estimated coefficients in such an abundance that consumers of research results find themselves inundated in details. An effective aggregation scheme reduces the complexity of such a system without however unduly reducing its accuracy. In this article it is assumed that an econometric system is to be used for forecasting, and the inaccuracy of the system is measured by the expected squared prediction error.

The increased inaccuracy due to aggregation has to be weighed against the benefits of increased simplicity, but a precise quantitative description of these simplicity benefits is not likely to be available in practice. We will consequently quantify the inaccuracy cost of aggregation but leave unquantified the simplicity benefit. This is analogous to reporting the reduction in the R^2 caused by the omission of a set of variables from a linear regression. Given a suitable list of assumptions, the R^2 measures the predictive accuracy of a model, and a researcher, who reports that a restricted model with a fewer number of explanatory variables has an R^2 almost as large as the complete model, is telling you that the inaccuracy cost is low, but is leaving unstated the simplicity benefit.

More formally the problem considered is how to aggregate a multivariate regression process given past data on the vector of m dependent variables and the vector of k explanatory variables. The object is to predict all m components given the future values of the k explanatory variables, with the inaccuracy of the prediction measured by the sum of squares of the component prediction errors. Assumptions are made which imply that the optimal conditional prediction is formed by regressing each of the m components on the k explanatory variables. This optimal conditional prediction is compared with the predictions generated by aggregated systems in which predictions of the components are functions of predictions of the aggregates. This in effect imposes restrictions across equations in the multivariate system.

The intent of the aggregation described here is to produce a simple representation of reality which works well for prediction purposes. The disaggregated system has k times m coefficients whereas the fully aggregated system has $k + m$ coefficients, k in the regression of the aggregate on the k explanatory variables plus m for translating the single aggregate into the m components. For example, a system of demand equations for related products (red Cherolets, blue Chevrolets, red Buicks, etc.) might be written as

$$D_1 = \alpha_1 + \sum_j \delta_{1j} P_j + \gamma_1 x + \beta_1 M$$

$$D_2 = \alpha_2 + \sum_j \delta_{2j} P_j + \gamma_2 x + \beta_2 M$$

$$\vdots$$

$$D_m = \alpha_m + \sum_j \delta_{mj} P_j + \gamma_m x + \beta_m M ,$$

where P_j $(j = 1,\ldots,m)$ are the prices of the m commodities, x is price of a composite alternative commodity and M is money income. A much simpler aggregated system which may work almost as well for prediction purposes is

$$D = \sum_j D_j = \alpha + \sum_j \delta_j P + \gamma x + \beta M$$

$$D_1 = \theta_1 D$$

$$D_2 = \theta_2 D$$

$$\vdots$$

$$D_m = \theta_m D ,$$

where D is the total demand and θ_j is the share of commodity j in total demand. The accuracy of the aggregated system depends on the coefficients in the disaggregated system. If the coefficients in the m disaggregated equations are proportional to each other, values for the coefficients in the aggregate system can be selected so that the two systems produce identical forecasts. Although it is not to be expected that the coefficients on the price variables, δ_{ij}, are proportional across equations, relative prices may vary so little that the aggregate system is nearly as accurate as the disaggregated system.

The aggregation method presented here has its source in a loss function which penalizes complexity. Another source of aggregation methods is prior information which asserts that the equations can be aggregated without error because the coefficient vectors of the m different equations are similar. In the terminology of Leamer (1978), the problem treated here is a "simplification search," whereas, were prior information the source of the desire to aggregate, the estimation difficulties could give rise to an "interpretive" search. In a series of papers, Chipman (1977a, 1977b) has presented results on aggregation as an "interpretive search" problem, including tests of the hypothesis that the system admits perfect aggregation.

The proportionality constraints across equations used to form a fully aggregated system are the same as the proportionality constraints in the MIMIC model of Joreskog and Goldberger (1975), with a latent variable equal to the aggregate and with each component being a linear function of the aggregate.

Most of the other econometric literature on aggregation deals with the consequences of the misspecifications induced by aggregating both independent and dependent variables, e.g., Theil (1954), Grunfeld and Griliches (1960) and Aigner and Goldfeld (1974). In this paper, the micro variables are the same in each equation and there is no misspecification caused by aggregation.

The econometric details are discussed in Section 1. An application to international trade data is presented in Section 2. The fifty-six two-digit SITC trade categories are aggregated into nine classes with only a seven percent increase in expected prediction error. In contrast, the nine one-digit SITC categories imply a fifty-five percent increase in the prediction loss compared with the fully disaggregated system.

1. The Aggregation Problem

The multivariate process is written as

$$\tilde{Y}_t = \tilde{B}\tilde{x}_t + \tilde{u}_t \qquad t = 0,\ldots,T \tag{1}$$

where $\tilde{Y}_t$ is a vector of m observables which depend on the vector of k observables $\tilde{x}_t$ and $\tilde{u}_t$ is a multivariate normal random vector with moments

$$E(\tilde{u}_t) = \tilde{0} \qquad t = 1,\ldots,T$$

$$E(u_t u'_{t^*}) = \begin{cases} \tilde{0} & t \neq t^* \\ \tilde{\Sigma} & t = t^* \end{cases}$$

The problem is to predict $\tilde{Y}_T$ given $\tilde{x}_T$ and $(\tilde{Y}_t, \tilde{x}_t)$, $t = 0,\ldots,T-1$, collected into the $T \times m$ matrix $\tilde{Y}$ and the $T \times k$ matrix $\tilde{X}$. The loss is assumed to be a function of the prediction $\hat{\tilde{Y}}_T$ and the actual $\tilde{Y}_T$

$$L = (\tilde{Y}_T - \hat{\tilde{Y}}_T)'(\tilde{Y}_T - \hat{\tilde{Y}}_T) \ .$$

The analysis to follow can be easily amended to deal with general quadratic loss functions. One which may be appropriate in some settings is the weighted sum of squares $(\tilde{Y}_T - \hat{\tilde{Y}}_T)'\ \tilde{D}\ (\tilde{Y}_T - \hat{\tilde{Y}}_T)$ where $\tilde{D}$ is a positive diagonal matrix. For example, if the vector $\tilde{Y}$ has components with non-comparable units such as heads of lettuce, cans of peas, quarts of orange juice, etc., the unweighted loss function seems to make little sense, and in any case the aggregation method would depend on whether you measured orange juice in quarts or gallons. A better approach would take as a prediction goal the value of each commodity, in which case the weights in $\tilde{D}$ would be the squares of commodity prices. Equivalently, equation (1) can be premultiplied by $\tilde{D}^{\frac{1}{2}}$. Another way to standardize the units is to weight by the inverses of the sampling variances.

Using the unweighted loss function, we can write the expected loss as

$$E(L|\hat{Y}_T, x_T, B, \Sigma) = E(Bx_T + u_T - \hat{Y}_T)'(Bx_T + u_T - \hat{Y}_T)$$

$$= (Bx_T - \hat{Y}_T)'(Bx_T - \hat{Y}_T) + \mathrm{tr}\Sigma$$

This expected loss is conditional on four arrays, all of which are unknown to the predictor; $\hat{Y}_T$, x_T, B, and Σ. The matrices B and Σ are clearly unknown, although the data (Y,X) contain information about them. The uncertainty about B and Σ is summarized in a posterior distribution. In particular, the posterior moments of B are

$$\bar{B} = E(B|(Y,X))$$

$$\bar{W} = E((B - \bar{B})'(B - \bar{B})|(Y,X)) = \sum_i \mathrm{Var}\,(\beta_i|Y,X)$$

where β_i is a row of B. Henceforth, all expectations are understood to be conditional on the data (Y_t, x_t), $t = 0,\ldots,T - 1$, although this conditioning statement is suppressed for ease of notation.

The prediction $\hat{Y}_T$ and the future explanatory variables x_T are also assumed to be unknown. Before observing x_T, we are required to select a predicting function $\hat{Y}_T(x_T)$ indicating the conditional predictions of Y_T. If this function can be selected without restriction, there is no cost to selecting it before x_T is observed. But the imposition of aggregation restrictions on the predicting function does involve costs which depend partly on what we expect x_T to be. It is important to understand that the problem considered is conditional prediction but unconditional aggregation. The real goal of this paper is the identification of a simple system $Y(x_T)$, and the prediction problem is selected as a sensible setting in which to measure the costs of simplicity. If aggregation as well as prediction were conditioned on $\underset{\sim}{x}_T$, we would have merely a problem of clustering a set of predictions $\hat{\underset{\sim}{Y}}_T$, and no aggregate system $\hat{\underset{\sim}{Y}}_T(\underset{\sim}{x}_T)$ would be sought.

We assume that $\underset{\sim}{x}_T$ without $\underset{\sim}{Y}_T$ does not contain information about $\underset{\sim}{B}$. We may therefore write the expected loss as

$$\begin{aligned} E(L|\hat{\underset{\sim}{Y}}_T,\underset{\sim}{x}_T,\underset{\sim}{\Sigma}) &= E(\underset{\sim}{B}\underset{\sim}{x}_T - \bar{\underset{\sim}{B}}\underset{\sim}{x}_T + \bar{\underset{\sim}{B}}\underset{\sim}{x}_T - \hat{\underset{\sim}{Y}}_T)'(\underset{\sim}{B}\underset{\sim}{x}_T - \bar{\underset{\sim}{B}}\underset{\sim}{x}_T + \bar{\underset{\sim}{B}}\underset{\sim}{x}_T - \hat{\underset{\sim}{Y}}_T) \\ &\quad + tr\underset{\sim}{\Sigma} \\ &= E(\underset{\sim}{x}_T'(\underset{\sim}{B} - \bar{\underset{\sim}{B}})'(\underset{\sim}{B} - \bar{\underset{\sim}{B}})\underset{\sim}{x}_T) + (\bar{\underset{\sim}{B}}\underset{\sim}{x}_T - \hat{\underset{\sim}{Y}}_T)'(\bar{\underset{\sim}{B}}\underset{\sim}{x}_T - \hat{\underset{\sim}{Y}}_T) \\ &\quad + tr\underset{\sim}{\Sigma} \ . \end{aligned}$$

This expected loss is minimized if the conditional prediction of $\underset{\sim}{Y}_T$ given $\underset{\sim}{x}_T$ is $\hat{\underset{\sim}{Y}}_T = \bar{\underset{\sim}{B}}\,\underset{\sim}{x}_T$. Generally, we restrict the prediction $\underset{\sim}{Y}_T$ to be a linear function of $\underset{\sim}{x}_T$

$$\hat{\underset{\sim}{Y}}_T = \hat{\underset{\sim}{B}}\underset{\sim}{x}_T \quad ,$$

and let the predictive moments of $\underset{\sim}{x}_T$ be

$$\underset{\sim}{S} = E(\underset{\sim}{x}_T\underset{\sim}{x}_T') .$$

Then the expected predicted loss becomes

$$E(L|\hat{\underset{\sim}{Y}}_T,\underset{\sim}{\Sigma}) = tr\underset{\sim}{W}\underset{\sim}{S} + tr(\bar{\underset{\sim}{B}} - \hat{\underset{\sim}{B}})'(\bar{\underset{\sim}{B}} - \hat{\underset{\sim}{B}})\underset{\sim}{S} + tr\underset{\sim}{\Sigma} \tag{2}$$

1.1 Full Aggregation

The expected loss (2) is minimized by setting $\hat{\underset{\sim}{B}}$ equal to $\bar{\underset{\sim}{B}}$. In a fully aggregated system the prediction is constrained to take the form

$$\hat{\underset{\sim}{Y}}_T = \hat{\underset{\sim}{\theta}}\hat{\underset{\sim}{\beta}}' x_T$$

where $\hat{\underset{\sim}{\beta}}$ is a $k \times 1$ vector chosen such that $\hat{\underset{\sim}{\beta}}'\underset{\sim}{x}_T$ predicts the aggregate, and $\hat{\underset{\sim}{\theta}}$ is an $m \times 1$ vector chosen such that $\hat{\theta}_i\hat{\underset{\sim}{\beta}}'\underset{\sim}{x}_T$ predicts the ith component. The addition to the expected loss due to this simplification is

$$\Delta E(L) = tr(\bar{\underset{\sim}{B}} - \hat{\underset{\sim}{\theta}}\hat{\underset{\sim}{\beta}}')'(\bar{\underset{\sim}{B}} - \hat{\underset{\sim}{\theta}}\hat{\underset{\sim}{\beta}}')\underset{\sim}{S}$$

$$= tr\bar{\underset{\sim}{B}}'\bar{\underset{\sim}{B}}\underset{\sim}{S} - 2\hat{\underset{\sim}{\theta}}'\bar{\underset{\sim}{B}}\underset{\sim}{S}\hat{\underset{\sim}{\beta}} + \hat{\underset{\sim}{\beta}}'\underset{\sim}{S}\hat{\underset{\sim}{\beta}}\hat{\underset{\sim}{\theta}}'\hat{\underset{\sim}{\theta}}$$

The vectors $\hat{\underset{\sim}{\beta}}$ and $\hat{\underset{\sim}{\theta}}$ may be chosen to minimize this quantity by setting to zero the derivatives

$$\partial\Delta E(L)/\partial\hat{\underset{\sim}{\beta}} = -2\underset{\sim}{S}\bar{\underset{\sim}{B}}'\hat{\underset{\sim}{\theta}} + 2\underset{\sim}{S}\hat{\underset{\sim}{\beta}}\hat{\underset{\sim}{\theta}}'\hat{\underset{\sim}{\theta}} = \underset{\sim}{0}$$

$$\partial\Delta E(L)/\partial\hat{\underset{\sim}{\theta}} = -2\bar{\underset{\sim}{B}}\underset{\sim}{S}\hat{\underset{\sim}{\beta}} + 2\hat{\underset{\sim}{\theta}}\hat{\underset{\sim}{\beta}}'\underset{\sim}{S}\hat{\underset{\sim}{\beta}} = \underset{\sim}{0} .$$

Assuming $\underset{\sim}{S}$ is invertible these can be rearranged to form

$$\hat{\underset{\sim}{\beta}} = \bar{\underset{\sim}{B}}'\hat{\underset{\sim}{\theta}}/\hat{\underset{\sim}{\theta}}'\hat{\underset{\sim}{\theta}}$$

$$\hat{\underset{\sim}{\theta}} = \bar{\underset{\sim}{B}}\underset{\sim}{S}\hat{\underset{\sim}{\beta}}/\hat{\underset{\sim}{\beta}}'\underset{\sim}{S}\hat{\underset{\sim}{\beta}} .$$

If the second equation is used in the first we obtain an equation for $\hat{\beta}$

$$\hat{\underset{\sim}{\beta}}\hat{\underset{\sim}{\beta}}'\underset{\sim}{S}\bar{\underset{\sim}{B}}'\bar{\underset{\sim}{B}}\underset{\sim}{S}\hat{\underset{\sim}{\beta}} = \bar{\underset{\sim}{B}}'\bar{\underset{\sim}{B}}\underset{\sim}{S}\hat{\underset{\sim}{\beta}}\hat{\underset{\sim}{\beta}}'\underset{\sim}{S}\hat{\underset{\sim}{\beta}} \quad , \quad \text{or}$$

$$(\bar{\underset{\sim}{B}}'\bar{\underset{\sim}{B}}\underset{\sim}{S} - \lambda\underset{\sim}{I})\hat{\underset{\sim}{\beta}} = 0$$

where $\lambda = \hat{\underset{\sim}{\beta}}'\underset{\sim}{S}\bar{\underset{\sim}{B}}'\bar{\underset{\sim}{B}}\underset{\sim}{S}\hat{\underset{\sim}{\beta}}/\hat{\underset{\sim}{\beta}}'\underset{\sim}{S}\hat{\underset{\sim}{\beta}}$. In words, $\hat{\underset{\sim}{\beta}}$ is an eigenvector of $\bar{\underset{\sim}{B}}'\bar{\underset{\sim}{B}}\underset{\sim}{S}$.

Using this condition, the addition to the expected loss can be written as a function of the eigenvalue λ,

$$\text{(3)} \qquad \Delta E(L) = \text{tr}\, \bar{B}'\bar{B}S - 2\, \hat{\beta}'S'\bar{B}'\bar{B}S\hat{\beta}/\hat{\beta}'S\hat{\beta} + \hat{\beta}'S\hat{\beta}\; \hat{\beta}'S\bar{B}'\bar{B}S\hat{\beta}/(\hat{\beta}'S\hat{\beta})^2$$

$$= \text{tr}\, \bar{B}'\bar{B}S - \lambda$$

which is minimized for λ maximized. Accordingly $\hat{\beta}$ is the eigenvector of $\bar{B}'\bar{B}S$ corresponding to the largest eigenvalue, a result which has its analog in Joreskog and Goldberger's (1975) maximum likelihood estimation of the MIMIC system. Since $\text{tr}\bar{B}'\bar{B}S$ is the sum of the eigenvalues, $\Delta E(L)$ is the sum of all but the largest eigenvalues of $V = \bar{B}S\bar{B}'$.

The problem with using an eigenvector c to form a model for the aggregate is that $c'x_T$ is not an optimal conditional prediction of any obvious aggregated quantity, and the interpretive simplicity of the resulting system is rather badly damaged. Accordingly, I now impose the restriction that $\hat{\beta}'x_T$ is the optimal predictor of the aggregate $1'Y_T$. This implies that

$$\hat{\beta} = \bar{B}'1 ,$$

$$\hat{\theta} = \bar{B}S\bar{B}'1/1'\bar{B}S\bar{B}'1 \qquad ,$$

and the increment to the expected loss is

$$\text{(4)} \qquad \Delta E(L) = \text{tr}\bar{B}'\bar{B}S - 1'\bar{B}S\bar{B}'\bar{B}S\bar{B}'1/1'\bar{B}S\bar{B}'1$$

$$= \text{tr}V - 1'V^2 1/1'V1$$

where

$$V = \bar{B}S\bar{B}'.$$

Both (3) and (4) are measures of the discordance among the rows of $\bar{B}$. If the coefficients are proportional across equations with $\bar{B} = \bar{\theta}\beta'$ both (3) and (4) take on the value zero, (3) because $\bar{B}'\bar{B}S$ has rank one and only one non-zero eigenvalue. The loss (4) can be thought to be an approximation to (3) since $\lim_{n\to\infty} 1'V^n 1/1'V^{n-1} 1 = \lambda_{max}$. To assist in interpreting (4) it is useful to consider the case of two equations with rows of $\bar{B}$ equal to $\bar{B}_1$ and $\bar{B}_2$, and with inner products $c_{ij} = \bar{B}_i' S \bar{B}_j$ weighted by the expected variability of the explanatary variables, S. Then a small amount of algebra yields

$$\Delta E(L) = 2(c_{22}c_{11} - c_{12}^2)/(c_{11} + c_{22} + 2c_{12}) \tag{5}$$

$$= 2(1 - \rho_{12}^2)\, c_{22}c_{11}/(c_{11} + c_{22} + 2c_{12})$$

where ρ_{12}^2 is the squared cosine of the angle between $\bar{B}_1$ and $\bar{B}_2$: $c_{12}^2/c_{11}c_{12}$. The corresponding value of the share vector is

$$\hat{\theta}' = (c_{11} + c_{12},\ c_{22} + c_{12})/(c_{11} + c_{12} + 2c_{12}). \tag{6}$$

What (5) reveals is that classes can be combined with little loss if they have similar coefficients (ρ_{12}^2 large, c_{12} positive) or if they have small coefficients (c_{11} and c_{22} small). A share $\hat{\theta}_i$ can be negative if c_{12} is negative, although the loss (5) creates a preference for positive c_{12}. A negative share causes interpretive difficulties and in the application below I have disallowed aggregation if it creates negative shares.

These measure of the discordance among the coefficient vectors include the constants in the regressions. For many applications it may be more meaningful to ignore the constants. We can eliminate the constants from consideration by letting the prediction vector have its own set of constants

$$\hat{Y}_T = \hat{\alpha} + \hat{B}\, x_T \quad .$$

Then $\hat{\alpha}$ can be selected to minimize the expected loss

$$E[(\bar{B}x_T - \hat{Y}_T)'\ (\bar{B}x_T - \hat{Y}_T)\,|\,Y_T,\Sigma]$$

$$= E((\bar{B} - \hat{B})x_T - \hat{\alpha})'\ ((\bar{B} - \hat{B})x_T - \hat{\alpha})).$$

Setting to zero the derivations of this expression with respect to $\hat{\alpha}$ yields

$$0 = -2(\bar{B} - \hat{B})\ E(x_T) + 2\hat{\alpha} \quad , \text{ or}$$

$$\hat{\alpha} = (\bar{B} - \hat{B})\bar{x}_T,$$

where $$\bar{x}_T = E(x_T|\Sigma) \quad .$$

When this is inserted back into the formula for expected loss, we obtain

$$E(L|Y_T,\Sigma) = \mathrm{tr}WS + \mathrm{tr}(\bar{B} - \hat{B})'(\bar{B} - \hat{B})\Omega + \mathrm{tr}\Sigma \tag{7}$$

where $\Omega = E(x_T - \bar{x}_T)(x_T - \bar{x}_T)'$. This is the same form as Equation (2) and the mathematics of aggregation need not be adjusted. The difference is that Equation (2) uses moments about zero S whereas Equation (3) uses moments around the means Ω. In particular, the constant has a zero predictive variance and all the constant terms drop out of (7).

An alternative choice for the share vector is $\hat{\theta} = 1$, which requires the prediction of each of the components to be exactly the same. For the applications I have in mind, this would not make much sense, basically because of the scale dependence. If you did restrict $\hat{\theta} = 1$, then the optimal value of $\hat{\beta}$ is $\hat{\beta} = \bar{B}'1/1'1$, which is the average value of the coefficients across equations. The corresponding increment to expected loss is $\text{tr}V - 1'V1/1'1$, which is greater than the loss incurred if only the restriction $\hat{\beta} = \bar{B}'1$ is imposed. The restriction $\hat{\theta} = 1$ is considered once again at the end of the next section.

1.2 Partial Aggregation

The algebra of partial aggregation makes use of aggregation matrices. An $n \times m$ matrix G of zeroes and ones is an aggregation matrix if there is a single one in each column and one or more ones in each row. For example, the matrix

$$G = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

aggregates the first and second elements of a three dimensional vector. Predictions of the aggregates GY_T are taken to be $G\bar{B}x_T$. These predictions are translated into predictions of the components via the $m \times n$ share matrix θ

$$\hat{Y}_T = \theta G\bar{B}x_T \ .$$

An additional assumption made for interpretive simplicity is that the components are predicted from only the subaggregates to which they belong. This implies that the non-zero elements of the share matrix $\underset{\sim}{\theta}'$ are located in the same places as the non-zero elements of $\underset{\sim}{G}$. For example, with $\underset{\sim}{G}$ defined by (8) we would have $\underset{\sim}{\theta}$ be

$$\underset{\sim}{\theta} = \begin{bmatrix} \theta_{11} & 0 \\ \theta_{21} & 0 \\ 0 & \theta_{32} \end{bmatrix} ,$$

$$\underset{\sim}{\theta}\underset{\sim}{G} = \begin{bmatrix} \theta_{11} & \theta_{11} & 0 \\ \theta_{21} & \theta_{21} & 0 \\ 0 & 0 & \theta_{32} \end{bmatrix}$$

and

$$\hat{Y}_T = \begin{bmatrix} \theta_{11}(\underset{\sim}{\overline{B}}_1' + \underset{\sim}{\overline{B}}_2')\underset{\sim}{x}_T \\ \theta_{21}(\underset{\sim}{\overline{B}}_1' + \underset{\sim}{\overline{B}}_2')\underset{\sim}{x}_T \\ \theta_{32}\ \underset{\sim}{\overline{B}}_3'\underset{\sim}{x}_T \end{bmatrix}$$

where $\underset{\sim}{\overline{B}}_i'$ is the ith row of $\underset{\sim}{\overline{B}}$.

The rows of the aggregation matrix $\underset{\sim}{G}$ will be denoted by $\underset{\sim}{G}_i'$, $i = 1,\ldots,n$, and the columns of the share matrix $\underset{\sim}{\theta}$ by $\underset{\sim}{\theta}_i$, $i = 1,\ldots,n$. It has been assumed by placement of zeroes that

$$\underset{\sim}{G}_i'\underset{\sim}{G}_j = 0 \quad \text{for} \quad i \neq j$$

$$\underset{\sim}{\theta}_i'\underset{\sim}{G}_j = 0 \quad \text{for} \quad i \neq j$$

$$\underset{\sim}{\theta}_i'\underset{\sim}{\theta}_j = 0 \quad \text{for} \quad i \neq j.$$

The share matrix $\underset{\sim}{\theta}$ is selected to minimize the expected loss (2) subject to $\hat{\underset{\sim}{B}} = \underset{\sim}{\theta}\underset{\sim}{G}\underset{\sim}{\bar{B}}$, and thus to minimize

$$\Delta E(L) = tr(\underset{\sim}{\bar{B}} - \underset{\sim}{\theta}\underset{\sim}{G}\underset{\sim}{\bar{B}})'(\underset{\sim}{\bar{B}} - \underset{\sim}{\theta}\underset{\sim}{G}\underset{\sim}{\bar{B}})\underset{\sim}{S}$$

$$= tr(\underset{\sim}{I} - \underset{\sim}{\theta}\underset{\sim}{G})'(\underset{\sim}{I} - \underset{\sim}{\theta}\underset{\sim}{G})\underset{\sim}{V}$$

where $\underset{\sim}{V} = \underset{\sim}{\bar{B}}\underset{\sim}{S}\underset{\sim}{\bar{B}}'$. This increment to expected loss can be rewritten as

(9)
$$\Delta E(L) = tr\underset{\sim}{V} - tr\underset{\sim}{G}'\underset{\sim}{\theta}'\underset{\sim}{V} - tr\underset{\sim}{\theta}\underset{\sim}{G}\underset{\sim}{V} + tr\underset{\sim}{G}'\underset{\sim}{\theta}'\underset{\sim}{\theta}\underset{\sim}{G}\underset{\sim}{V}$$

$$= tr\underset{\sim}{V} - 2tr\underset{\sim}{G}\underset{\sim}{V}\underset{\sim}{\theta} + tr\underset{\sim}{\theta}'\underset{\sim}{\theta}\underset{\sim}{G}\underset{\sim}{V}\underset{\sim}{G}'$$

$$= tr\underset{\sim}{V} - 2\sum_{i=1}^{n} \underset{\sim}{G}_i'\underset{\sim}{V}\underset{\sim}{\theta}_i + \sum_{i=1}^{n} \underset{\sim}{G}_i'\underset{\sim}{V}\underset{\sim}{G}_i\underset{\sim}{\theta}_i'\underset{\sim}{\theta}_i$$

because $\underset{\sim}{\theta}'\underset{\sim}{\theta}$ is a diagonal matrix with $\underset{\sim}{\theta}_i'\underset{\sim}{\theta}_i$ on the diagonal. The derivative of (9) with respect to $\underset{\sim}{\theta}_i$ is

$$-2\underset{\sim}{V}\underset{\sim}{G}_i + 2\underset{\sim}{G}_i'\underset{\sim}{V}\underset{\sim}{G}_i\underset{\sim}{\theta}_i.$$

Only the components corresponding to the non-zero elements of $\underset{\sim}{\theta}_i$ are set to zero. To do this, define the $m \times m$ diagonal matrix

$$\underset{\sim}{P}_i = \text{diag}\{\underset{\sim}{G}_i\} \quad ,$$

and let

(10)
$$\underset{\sim}{\theta}_i = \underset{\sim}{P}_i\underset{\sim}{V}\underset{\sim}{G}_i/\underset{\sim}{G}_i'\underset{\sim}{V}\underset{\sim}{G}_i.$$

Given these values of $\underset{\sim}{\theta}_i$, the increase in expected loss becomes

(11)
$$\Delta E(L) = tr\underset{\sim}{V} - \sum_i \underset{\sim}{G}_i'\underset{\sim}{V}\underset{\sim}{P}_i\underset{\sim}{V}\underset{\sim}{G}_i/\underset{\sim}{G}_i'\underset{\sim}{V}\underset{\sim}{G}_i$$

The preceding discussion has to be altered if the shares are restricted to be equal, that is, if each of the components of an aggregate has to have the same prediction. This amounts to the restriction that $\hat{\theta}$ is the transpose of an aggregation matrix, $\hat{\theta} = G'$. The prediction can then be written as $\hat{Y}_T = G'\hat{\beta}x_T$, where $\hat{\beta}$ is an $n \times k$ matrix selected to minimize

$$\begin{aligned} \Delta E(L) &= tr(\overline{B}-G'\hat{\beta})'(\overline{B}-G'\hat{\beta})S \\ &= tr(I-G'R)'(I-G'R)V \end{aligned}$$

where $\hat{\beta} \equiv R\,\overline{B}$. This can be written as

$$\begin{aligned} \Delta E(L) &= trV - trR'GV - trG'RV + trR'GG'RV \\ &= trV - 2trGVR' + trGG'RVR' \\ &= trV - 2\sum_i G_i'VR_i + \sum_i (G_i'G_i)R_i'VR_i \end{aligned}$$

since GG' is an $n \times n$ diagonal matrix with the number of components in each aggregate on the diagonal. The derivative of the expression with respect to R_i the $\underline{ith}$ row of R, is set of zero, $0 = -2\ G_i'V + 2(G_i'G_i)R_i'V$, to minimize the loss. Thus, with V assumed invertible, we have

$$\begin{aligned} R_i &= G_i(G'_i G_i)^{-1} \\ R &= (G'G)^{-1}G \\ \hat{Y}_T &= G'(G'G)^{-1}G\ \overline{B}\ x_T \end{aligned}$$

$$\Delta E(L) = trV - \sum_i G_i'VG_i/G_i'G_i \qquad (12)$$

1.3 Choice of Measures of Uncertainty

Last it is necessary to specify $\underset{\sim}{\Sigma}$, $\underset{\sim}{S}$ and $\underset{\sim}{W}$ in (2) and $\underset{\sim}{\Omega}$ in (7). These are moments of a posterior distribution which we will now form with the assumption that the prior is diffuse. For a similar treatment of the simplification of a single equation see Lindley (1968) or Leamer (1978, pp. 208-214). The T past observations of the m dependent variables are collected in the $(T \times m)$ matrix $\underset{\sim}{Y}$, and the T past observations on the k explanatory variables are collected in the $(T \times k)$ matrix $\underset{\sim}{X}$. Given a diffuse prior distribution, the posterior mean of $\underset{\sim}{B}$ is equal to the least squares matrix, which, assuming that $(\underset{\sim}{X}'\underset{\sim}{X})$ is invertible, can be written as

$$\underset{\sim}{\bar{B}} = \underset{\sim}{Y}'\underset{\sim}{X}(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$$

where each row of the $m \times k$ matrix $\underset{\sim}{\bar{B}}$ contains the regression coefficients from one of the m equations. The posterior variance of each row of $\underset{\sim}{B}$ is equal to the sampling variance

$$\mathrm{Var}(\underset{\sim}{\beta}_i) = \sigma_i^2(\underset{\sim}{X}'\underset{\sim}{X})^{-1}$$

where σ_i^2 is a diagonal element of $\underset{\sim}{\Sigma}$. Furthermore, if the vectors $\underset{\sim}{x}_t$ are assumed to come independently from a normal population, we would have the predictive moments approximately equal to

$$\underset{\sim}{S} = E(\underset{\sim}{x}_T\underset{\sim}{x}_T'|\underset{\sim}{X}) = \underset{\sim}{X}'\underset{\sim}{X}/T.$$

$$\underset{\sim}{\Omega} = E(\underset{\sim}{x}_T-E\underset{\sim}{x}_T)(\underset{\sim}{x}_T-E\underset{\sim}{x}_T)' = \underset{\sim}{X}'\underset{\sim}{X}/T - \underset{\sim}{X}'\underset{\sim}{1}\,\underset{\sim}{1}'\underset{\sim}{X}/T^2$$

Depending on whether the constants are constrained across equations, $\underset{\sim}{V}$ becomes either

$$\underset{\sim}{V} = \bar{B}S\bar{B}' = Y'X(X'X)^{-1}X'Y/T \quad ,$$

or

$$\underset{\sim}{V} = \bar{B}\Omega\bar{B}' = Y'X(X'X)^{-1}X'Y/T - Y'1\ 1'Y/T^2 \quad ,$$

where I have assumed that one of the columns of $\underset{\sim}{X}$ is the vector of ones, which implies $1'X(X'X)^{-1}X' = 1'$.

The minimum expected loss (2) with $\bar{B} = \hat{B}$ becomes

$$E(L|\Sigma) = tr(\sum_i \sigma_i^2 (X'X)^{-1})X'X/T + tr\Sigma$$
$$= \sum_i \sigma_i^2 \left(\frac{k}{T} + 1\right)$$

Again, using the diffuseness assumptions we would have

$$E(\sigma_i^2) = Y_i'(I - X(X'X)^{-1}X')Y_i/(T - k) \equiv s_i^2 \quad ,$$

where Y_i is the $i\underline{th}$ column of $\underset{\sim}{Y}$. Then the minimum expected loss is

$$E(L) = (\Sigma s_i^2)(\frac{k}{T} + 1).$$

1.4 Computer Algorithms

The number of $n \times m$ aggregation matrices is on the order of n^m, which for the problem considered in the next section with $m = 56$ and $n = 9$ is the huge number 2.7×10^{53}. The calculation costs of a global minimization of (11) or (12) for most problems will accordingly be unacceptably high. What can be done at reasonable cost is to find locally optimal aggregates by transfering components between aggregates one at a time to reduce (12) until all possibilities of further reductions are exhausted, as is done by Hartigan's (1975, p. 84) "K-means algorithm." The initial starting point for this local optimization algorithm is the set of n aggregates formed by sequentially coupling pairs of aggregates until only n remain. At each step, the pair of aggregates is coupled which causes the smallest increase in expected loss.

A final defect of the solution just presented is that the share vectors (10) are not necessarily positive, although they do add to one. A negative share can be interpreted to mean that the component behaves the opposite of the aggregate, and allowing negative shares therefore destroys the meaning of an aggregate. Accordingly, the local optimization procedure described in the previous paragraph is altered to allow only positive shares, with components shifted among aggregates when a negative share occurs.

2. Aggregation of Trade Data

This section reports an aggregation of the fifty-six two-digit SITC trade categories. The dependent variables are the net exports in thousands of dollars of these commodities in 1975 by thirty-four countries (OECD plus selected developing countries). The explanatory variables are:

Capital = capital stock in millions of 1966 dollars, formed by accumulating and discounting investment flows

L_1 = thousands of professional/technical workers

L_2 = thousands of literate nonprofessional workers

L_3 = thousands of illiterate workers, assuming the worker literacy rate is the same as the population literacy rate

T_1 = thousands of hectares of land area, tropical-rainy climate

T_2 = thousands of hectares of land area, humid mesothermal climate

T_3 = thousands of hectares of land area, humid micro-thermal climate

The capital stock figures are derived from World Bank Tables 1976, the labor variables from ILO, Labor Force Projections 1965-1985, and the land area variables from FAO, Production Yearbook and U.S. Air Force, Climatic Chart of the World. The theoretical underpinning of this equation comes from the Heckscher-Ohlin-Samuelson trade model with Vanek's (1968) assumption of identical homothetic tastes. If the number of immobile factors equals the number of commodities then trade is a linear function of the endowments. Otherwise, the function theoretically is non-linear and the estimated equation should be thought to be a linear approximation.

The percentage increase in expected loss due to aggregation is graphed as a function of the number of aggregates in Figure 1. If the system with ten aggregates is selected, the expected squared prediction error increases by only six percent. The marginal effect of further aggregation does not become too large until the number of aggregates is four or less. If a single aggregate is selected, the expected prediction loss increases by 172 percent.

The locally optimal system with nine aggregates can be compared with the one-digit SITC system. The one-digit SITC system with nine aggregates implies a fifty-five percent increase in expected loss, which compares very unfavorably to the seven percent increase implied by the locally optimal system. The composition of the locally optimal aggregates is indicated in Table 1, and the shares calculated by formula (10) for the SITC scheme are indicated in Table 2. Two of these SITC shares are negative, suggesting strongly that SITC 09 (Miscellaneous Food Preparations) and SITC 35 (Electric Energy) are misclassified.

The locally optimal scheme suggests many other misclassifications as well. SITC 11 and 12 are not important enough to form an aggregate by themselves. SITC 11, beverages, behaves most like the labor intensive manufactured products such as clothing and footwear. SITC 12, Tobacco, is more like cereals. SITC 41, 42 and 43 should also be split up, with 41, animal oils and fats, combining with cereals, and 42, fixed vegetable oils, fats, and 43, processed animal oils, moving to chemicals. SITC 25, pulp and waste paper, and SITC 64, paper and paperboard, should be combined. Actually, the one class which stays reasonably intact is SITC 5, chemicals, to which the locally optimal scheme adds SITC 71, nonelectrical machinery.

This optimal aggregation scheme imposes proportionality constraints on the constants as well as the slopes. The analysis has also been done with the constants ignored, that is with $\underset{\sim}{\Omega}$ replacing $\underset{\sim}{S}$ in the calculations. The resulting expected losses are only slightly less than the ones recorded in Figure 1. Also the nine aggregate classes are virtually the same as those reported in Table 1. The four commodities which are reallocated are indicated in Table 1 with brackets containing the aggregate to which they are assigned: SITC 29 to aggregate 1, SITC 42 to aggregate 9, SITC 83 to aggregate 5 and SITC 23 to aggregate 2.

At a formal statistical level the aggregation scheme reported in Table 1 does significantly better than the one-digit SITC scheme. But my interest in forming the aggregates was not of course to predict the net exports of some hypothetical randomly selected country, but rather to collapse the fifty-six two-digit commodities into a manageable number which reveal as clearly as possible, the salient aspects of international trade. I think this has been very well achieved.

The aggregates reported in Table 1 can be thought to be composed of commodities for which the sources of comparative advantage are similar. The attainment of this goal depends on the completeness of the list of factor endowments which are used as explanatory variables. Given that this list is not especially inclusive, the aggregation seems remarkably successful. There are two aggregates (1 and 2) of raw materials, two of farm products (3 and 4) and five of manufactured products (5 to 9). The raw materials are (1) petroleum and (2) wood and ores. (I am neglecting to mention the smaller components). The farm products are (3) meat, fruit, vegetables and sugar and (4) cereals, etc. The manufactured aggregates, arranged in rough order of technological complexity are (5) paper, pulp and non-ferous metals, (6) clothing and footwear, (7) iron, steel, textiles, (8) electrical machinery and transport equipment, and (9) chemicals and non-electrical machinery. The only obvious misclassifications of commodities of significant size are the assignment of (03) Fish and (07) Coffee to Aggregate 6, the labor intensive manufactured commodities. If a system of ten aggregates is selected, the only change that occurs is that SITC06, sugar, and SITC07, coffee, are combined to form the tenth aggregate. This leaves only SITC03 having a fishy assignment, though it is not obvious to which of the aggregates it can be sensibly assigned.

The least squares coefficients of the nine aggregates are reported in Table 3. Countries which are scarce in land tend to have comparative advantage in the last three manufactured aggregates (all the coefficients are negative). Two of the manufactured aggregates (5 and 6) tend to be exported by countries which are abundant in land, though the coefficients are small. The land with moderate climate (T_2) has generally the largest coefficients, especially so for the two farm products (3 and 4). Land with the cooler climate (T_3) does confer comparative advantage in cereals (4) but not fruits and vegetables (3). The first manufactured aggregate (5), paper and pulp, is associated with the cool land (T_3), which I take to reflect soft-wood forest resources.

Capital abundance confers comparative advantage in the last three manufactured aggregates. Abundance in professional/technical workers somewhat surprisingly is a disadvantage for the exports of all the manufactured aggregates. Cereals, especially, but also wood and ores, are associated with professional/technical workers. This, I suspect, is due to the "dumbbell" effect of the U.S., which is a cereal exporter and has an abundance of professional/technical workers. The abundance of literate nonprofessional workers (L_2) is associated with the exporting of the manufactured aggregates 6, 7, and 8. The relative importance of capital versus labor is measured by a comparison of the capital coefficients divided by the labor (L_2) coefficients. In that sense, aggregate (6), clothing and footwear, is the most labor intensive manufactured commodity since the capital coefficient is negative, and aggregate (9), chemicals and non-electrical machinery, is the most capital intensive because the labor coefficient is negative. Also (8), electrical machinery and transport equipment, is more capital intensive than (7), iron, steel and textiles (19.8/235 > 5.8/184). This conforms exceptionally well with capital intensities computed from input/output tables.

What seems to be the only puzzle in Table 3 is the set of coefficients for L_3, the illiterate work force. Most of the countries in the sample have similarly high literacy rates, and L_3 is almost a linear combination of L_1 and L_2. The coefficients on L_3 are accordingly very inaccurate and not much should be made of these values in Table 2.

To conclude, a regression analysis of the fifty-six two-digit SITC commodity classes produces a bewildering array of estimates. By setting up a formal aggregation problem, I have combined similar and small classes to form nine aggregates, the same number as the two-digit SITC aggregation scheme. This optimal scheme is much more accurate than the two-digit scheme. It also provides a rather clear picture of the sources of international comparative advantage.

It is possible to find fault with the aggregation scheme presented here, and some reallocations might be done if they improved the interpretive clarity of the aggregates without unduly increasing the expected loss. Reallocations of small categories should not affect the general conclusions about the sources of comparative advantage, however.

References

Aigner, D. I. and S. M. Goldfeld, "Estimation and Prediction from Aggregate Data when Aggregates are Measured More Accurately than Their Components," Econometrica, 42, (January, 1974), 113-134.

Chipman, J. S., "Estimation and Aggregation in Econometrics: An Application of the Theory of Generalized Inverses," M. Zahair Nashed, ed., Generalized Inverses and Applications (New York: Academic Press, 1976), 549-569.

Chipman, J.S., "Statistical Problems Arising in the Theory of Aggregation," P. R. Krishnaiah, ed., Proceedings of the Symposium on Applications of Statistics (Amsterdam: North-Holland Publishing Co., 1977), 123-140.

Chipman, J.S., "Towards the Construction of an Optimal Aggregative Model of International Trade: West Germany, 1963-1975," Annals of Economic and Social Measurement, 6, (1977), 535-554.

Grunfeld, Y. and Z. Griliches, "Is Aggregation Necessarily Bad?" Review of Economics and Statistics, 42, (1977), 1-13.

Hartigan, J., Clustering Algorithms, (New York: John Wiley, 1975).

Joreskog, K. and A. Goldberger, "Estimation of a Model with Multiple Indicators and Multiple Causes of a Single Latent Variable," Journal of the American Statistical Association, 70, (1975), 631-639.

Leamer, E.E., Specification Searches, (New York: John Wiley, 1978).

Lindley, D.V., "The Choice of Variables in Multiple Regression," J. Roy. Stat. Soc., B, 31, (1968) 31-66.

Vanek, J., "The Factor Proportions Theory: The N-Factor Case," Kyklos, 21, (1968) 749-756.

TABLE 1

Composition of Nine Aggregates

	SITC	Description	Share
Aggregate 1	33	Petroleum, Petroleum Products	1.000
Aggregate 2	00	Live Animals	.036
	24	Wood, Lumber, Cork	.305
	26	Textile Fibres	.153
	27	Crude Fertilizers, Materials	.056
	28	Metalliferous Ores, Scrap	.421
	29	Crude Animal, Vegetable Materials [1]	.030
Aggregate 3	01	Meat, Meat Preparations	.452
	02	Dairy Products, Eggs	.067
	05	Fruit, Vegetables	.222
	06	Sugar, Sugar Preparations, Honey	.232
	42	Fixed Vegetable Oils, Fats [9]	.009
	83	Travel Goods, Handbags, Etc. [6]	.016
Aggregate 4	04	Cereals	.457
	08	Feeding Stuff for Animals	.056
	12	Tobacco, Tobacco Manufactures	.073
	21	Hides, Skins, Furskins, Undressed	.037
	22	Oil-seeds, Oil-nuts, Oil-kernals	.193
	32	Coal, Coke, Briquettes	.150
	35	Electric Energy	.002
	41	Animal Oils, Fats	.021
	56	Fertilizers, Manufactured	.010

TABLE 1 (Continued)

	SITC	Description	Share
Aggregate 5	23	Crude Rubber [2]	.017
	25	Pulp, Waste Paper	.185
	34	Gas, Natural and Manufactured	.116
	64	Paper, paperboard	.280
	68	Non-ferrous metals	.402
Aggregate 6	03	Fish, Fish Preparations	.135
	07	Coffee, Tea, Cocoa, Spices, Etc.	.243
	11	Beverages	.087
	61	Leather, Dressed Furskins	.016
	63	Wood, Cork Manufactures	.064
	66	Non-metallic Mineral Manufactures	.099
	84	Clothing	.236
	85	Footwear	.120
Aggregate 7	62	Rubber Manufactures, NES	.045
	65	Textile Yarn, Fabrics, etc.	.253
	67	Iron, steel	.459
	81	Sanitary, etc., Fixtures, Fittings	.008
	82	Furniture	.027
	89	Misc. Manuf. Articles, NES	.209
Aggregate 8	69	Manufactures of Metal	.097
	72	Electrical Machinery	.310
	73	Transport Equipment	.592

TABLE 1 (Continued)

	SITC	Description	Share
Aggregate 9	09	Miscellaneous Food Preparations	.008
	43	Animal, Vegetable Oils, Fats, Proc.	.002
	51	Chemical Elements, Compounds	.076
	52	Mineral Tar, Crude Chemicals	.001
	53	Dyeing, Tanning, Coloring Mat.	.023
	54	Medicinal, Pharmaceutical Pro.	.032
	55	Essential Oils, Perfume Materials	.010
	57	Explosives, Pyrotechnic Products	.001
	58	Plastic Materials, Cellulose, Etc.	.073
	59	Chemical Materials, Products, NES	.040
	71	Machinery, Other than Electric	.650
	86	Prof., etc., Instruments, etc.	.084

TABLE 2

Share Vectors: SITC One-digit Aggregation

SITC	Share	SITC	Share	SITC	Share
00	.034	32	.029	61	.007
01	.273	33	.921	62	.032
02	.054	34	.052	63	.043
03	.082	35	-.001	64	.069
04	.089			65	.200
05	.129	41	.536	66	.082
06	.130	42	.408	67	.411
07	.183	43	.056	68	.060
08	.033			69	.096
09	-.006	51	.291		
		52	.005	71	.409
11	.967	53	.089	72	.216
12	.033	54	.125	73	.375
		55	.038		
21	.036	56	.028	81	.009
22	.100	57	.003	82	.052
23	.008	58	.265	83	.032
24	.244	59	.156	84	.329
25	.078			85	.181
26	.132			86	.043
27	.048			89	.355
28	.336				
29	.020				

TABLE 3

Estimated Coefficients of the Nine Aggregates

Aggregate	Capital	L_1	L_2	L_3	T_1	T_2	T_3
1	-3.5	344	-72	14	.5	3.5	1.3
2	-3.4	870	-132	15	1.5	5.5	3.0
3	1.9	-302	-27	17	1.0	8.6	-.3
4	.8	1044	-136	9	.8	7.5	1.8
5	-.2	-143	-12	9	.1	.2	4.1
6	-7.5	-401	89	-15	.6	.6	.4
7	5.8	-1693	184	-1.6	-1.0	-4.8	-1.7
8	19.8	-2777	235	17	-1.2	-8.0	-2.4
9	23.6	-985	-16	42	-.9	-11.4	-5.0

1. Petroleum
2. Wood and Ores
3. Meat, Fruit, Vegetables and Sugar
4. Cereals, etc.
5. Paper, pulp, and non-ferrous metals
6. Clothing and footwear
7. Iron, steel and textiles
8. Electrical machinery and transport equipment
9. Chemicals and nonelectrical machinery

TABLE 4

Coefficients for One-digit SITC Classes

SITC	CAP	L_1	L_2	L_3	T_1	T_2	T_3	Const.
0	1.16	35.9	-79.7	23.5	2.6	13.3	1.3	40991
1	-2.32	171.4	1.5	-6.3	-.1	.2	.2	-9193
2	-5.03	1249.2	-168.1	13.7	1.8	6.8	4.2	41647
3	-1.63	412.9	-104.4	22.4	.7	4.1	1.5	-105457
4	.39	-.8	-3.9	1.2	.1	.5	.0	-18050
5	7.08	-260.5	-16.0	15.0	-.3	-3.2	-1.0	37141
6	4.90	-1793.8	181.8	3.0	-1.0	-4.3	1.2	56200
7	32.6	-3090.6	201.9	40.4	-1.7	-14.5	-5.2	-123412
8	.1	-773.9	98.5	-6.3	-.7	-1.2	-1.3	-20373

0. Food and Live Animals Chiefly for Food
1. Beverages and Tobacco
2. Crude Materials, Inedible, Except Fuels
3. Mineral Fuels, Lubricants and Related Materials
4. Animal and Vegetable Oils, Fats and Waxes
5. Chemical and Related Products, N.E.S.
6. Manufactured Goods Classified Chiefly by Material
7. Machinery and Transport Equipment
8. Miscellaneous Manufactured Articles

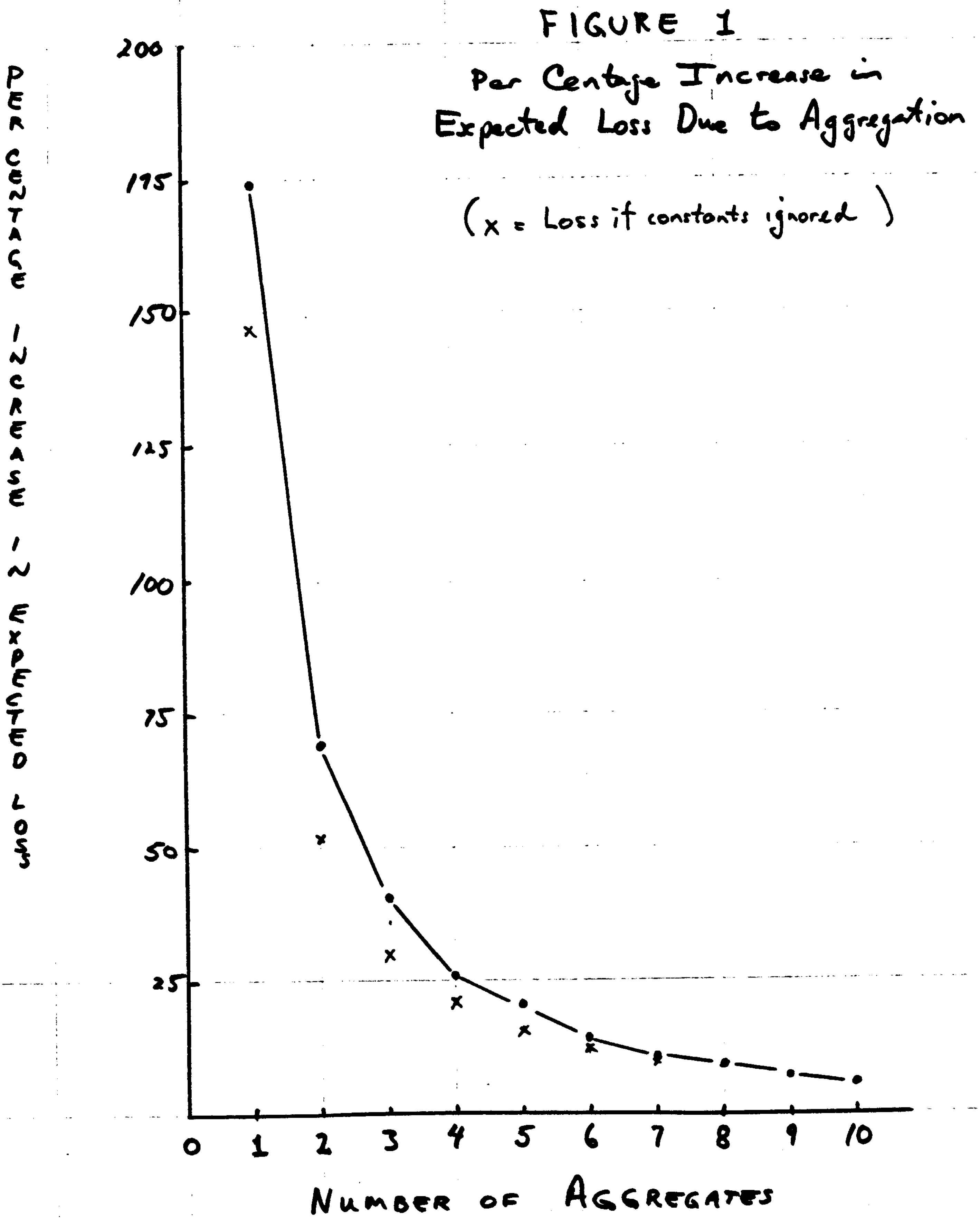
FIGURE 1
Per Centage Increase in
Expected Loss Due to Aggregation
(x = Loss if constants ignored)
PER CENTAGE INCREASE IN EXPECTED LOSS
200
175
150
125
100
75
50
25
0 1 2 3 4 5 6 7 8 9 10
NUMBER OF AGGREGATES