Study of Corruption Data and Using the Internet to Reduce Corruption

H. D. Vinod[?]

Economics Department, Fordham University, Bronx, NY 10458

E-mail: vinod@fordham.edu

Key Words: Cp, AIC, Trimmed Correlation, Income inequality, Schooling, Red tape, Efficient judiciary.

Abstract

I briefly review some of the micro and macro theory of corruption and relate it to various socio-economic and political variables. To quantify the burden of corruption I use Harrison and Vinod's (1992) 95% confidence interval for the marginal excess burden (MEB) of taxation. In developing countries, one dollar of corruption is estimated to impose a burden of $1.67, which becomes very large when compounded over time. A cross-sectional data-analysis reveals the relevance of "red tape" and "efficiency of judiciary." A subset regression using sophisticated (Cp and Akaike) criteria reveals other important factors. I argue that Corruption's economic harm is widely underestimated and bribery is recognised only "in the abstract," but individual criminals are not exposed. I suggest international aid and co-operation in exposing and fighting corruption. To this end, I describe a new Internet-based information storage and retrieval project.

1. Introduction and Dimensions of the Corruption Problem

This section introduces the dimensions of the corruption problem by noting that corruption hurts all countries and almost everyone. The points covered here are well summarised by the titles of following sections. (2) Theoretical models of corruption: A review. (3) Measuring the burden of corruption: A new approximation. (4) Description of interdisciplinary data. (5) Cross-sectional correlations and subset regressions. (6) Standardised corruption database on the Internet. (7) Conclusions.

Corruption scandals regularly appear on front pages of newspapers everywhere. When a corrupt official in a poor country chooses an economic development project simply because of a bribe, this wastes scarce resources. When that official favours incompetent domestic monopolists by imposing high tariffs on imports, this can reduce markets for exporting countries, cause unemployment and slower world economic growth. The terrible burden of slow economic growth in poor countries is indirectly shared by rich investors. Hence, corruption in poor countries deserves far greater international attention and exposure than it receives. Unfortunately, international co-operation in fighting corruption is missing due to jurisdictional issues. Since bribery is an internal crime against property, international law discourages foreign interference. This means that bribery "in the abstract"

[?] I thank Lionel Lucero for major help with data, and Rick Geddes and Sunil Deshmukh for comments. A version has been posted on www.fordham.edu/economics/vinod since February 23, 1999. This is prepared for presentation at the 9th International Anti-Corruption Conference, Durban, South Africa, 10-15 October, 1999.

is recognised, but no one is held responsible or blamed for it. The new world order should include innovative solutions to the jurisdictional problems in identifying those responsible for significant corruption. We need international technical assistance and co-oeration in apprehending and punishing the corrupt.

A fight against corruption involves fighting human nature. A bribe occurs when both parties benefit and it is generally difficult to balance a large potential benefit enjoyed by both parties with appropriate punishments and disincentives. Why have tough laws so far failed to reduce corruption in poor countries? Three fundamental problems with corruption are: (a) It is hidden, (b) Incentives for fighting it are improperly aligned, and (c) International jurisdiction is missing. We claim that all three problems can be simultaneously attacked by using the Internet. Dissemination of information about human rights abuses by Amnesty International and others has recently helped reduce human rights abuses. Similarly, the burden of corruption can be reduced by dissemination of information on the Internet, provided we name the criminals.

2. Theoretical Models of Corruption: A review

Corruption is sometimes defined as selling of government services or property by bureaucrats (bribe-takers) for personal gain. Since private and non-profit sectors can also have corruption, the definition should include unauthorised selling in those sectors also. In general, any criminal behaviour has several facets including social, economic, racial, psychological, political, legal, moral, religious, sexual, and medical. Hence, an interdisciplinary approach is generally followed in criminology, as can be seen from various criminology journals. However, since corruption involves exchange of property, it has attracted the attention of many economists. The corruption literature in economics is surveyed by Bardhan (1997). He provides detailed discussion of why the incidence of corruption is so different in different countries by using Schelling diagrams.

It is generally agreed that free market competition and democracy will bring peace and prosperity. However, bringing true competition in traditional societies of poor countries is not easy. Corruption and unfair enforcement of contract laws seriously hurts the honest entrepreneur. Equality of opportunity to all entrepreneurs and good government are needed for true competition to take root. This paper focuses on the measurement of the cumulative economic burden of corruption, and provides an empirical ranking of most relevant action variables.

2.1 Microeconomic Game Theoretic Viewpoint

I start with the micro viewpoint of the individual bribe-taker. Becker (1968) considers a rational individual choosing between criminal and legal behaviour in light of the probability of being detected and punished. Some refinements to Becker's model in the literature incorporate: (i) utility functions, (ii) risk neutrality associated with being caught, and (iii) the effect of the wealth of the rational individual on the decision. Game theory provides powerful tools for situations when some bargaining between economic agents is involved. For example, we can assume that there is recursive “Nash bargaining” between the bribe-taker and bribe-giver. A game-theoretic computation of an equilibrium bribe is made in Basu et *al* (1992). The viewpoint of individual bribe-taker is used in Rijckghem and Weder's

(1997) cross section study, which finds a negative relation between corruption and wages, implying a need to raise wages of bureaucrats.

Result 1: (Absentee Owner Hazard): A failure of incentives occurs because the cooperative Nash equilibrium between the bribe-taker and the bribe-giver ignores the public interest. The public is the absent third party, which owns the asset, not represented by anyone in the Nash bargain.

This is similar to the "moral hazard," arising from asymmetric information (See Sandmo, 1999), and agency costs arising from the "principal- agent problem." Mookherjee and Png's (1995) theory of "delegated enforcement" is also similar. To avoid confusion, I call this the "absentee owner hazard," where the public gets the worst deal. The bribe-taker is "selling" something that belongs to the absent third party. Unless detected, a seller's cost is near zero. The bribe-giver gets a resource at a fraction of the real price. Hence, there is a great incentive among both to keep the deal secret at the cost of the absent third-party owner. The hazard merely summarises all such reasons, which make a fight against corruption difficult.

2.2 Can Corruption Be Beneficial?

It seems intuitively plausible that a bribe-taking bureaucrat will work harder on his job. Hence, a question to the economist is whether costs associated with bribes exceed these potential benefits. Leff (1964) suggested that corruption may help economic growth, since it may be the "speed money," which reduces bureaucratic delays and induces hard work by bribe-taking government workers. However, Santhanam (1964) notes that, in practice, the opposite is true. The bureaucrats deliberately cause delays to extract the largest bribes that the "market" can bear. See Bardhan (1997) and Susan Rose-Ackerman (1996) for a recent overview of this literature. For examples of regressive impact, see Susan Rose-Ackerman and Andrew Stone (1996).

Mookherjee and Png (1995) start with the assumption of no social cost of bribery, ignore all transfer payments and still are able to prove the following proposition. "For every outcome when bribery is profitable, there exists another in which bribery is not profitable, that yields higher welfare." Thus, corruption can be proved to reduce social welfare, and cannot be beneficial in the long run.

2.3 Arguments for Compounding the Burden of Corruption

The aggregative dynamic macroeconomic view of corruption in this subsection hopes to show that compounding of burden is appropriate. It is well known that capital accumulation is the main engine of growth. Consider an aggregative closed economy producing output Y(t) at time t using two homogeneous inputs: capital K(t) and labour L(t). Let these inputs include the input of government services and public capital. The income identity states that the income can be either consumed or invested: $Y(t)=C(t)+I(t)$, where C(t) denotes consumption, and I(t) denotes investment. $I(t)=(d/dt)K(t)+\delta K(t)$ is called the investment identity which states that economic investment I(t) equals change in capital stock measured by the time derivative (d/dt) of K(t) plus a depreciation term $\delta K(t)$ proportional to the capital stock. Besides the two identities above, production function is the third pillar of neoclassical growth theory. It is customary to define $y(t)=Y(t)/L(t)$, $k(t)=K(t)/L(t)$ and

c(t)=C(t)/L(t), which leads to a univariate production function y=f (k), by measuring output and capital, per unit of labour and omitting the postscript (t).

The fundamental differential equation of neoclassical economic growth, Intriligator (1971, p.402) or Barro and Sala-i-Martin (1995, p.18), is:

$$(dk/dt)= s\, f(k) - (n+\delta)\, k, \qquad (1)$$

where n denotes the growth rate of population (labor force) and s is the savings rate. Dividing both sides by k gives (dlog k /dt) on the left side. The solution of the differential equation is

$$k=\exp[s\, t\, f(k)/k - n\, t - \delta\, t\,], \qquad (2)$$

where the *compounding* over time arises because of raising to the power of e (= 2.7183, approximately) in the solution (2). To verify (2) take the log of both sides and differentiate with respect to t.

Result 2: Corruption reduces savings rate, efficiency of capital in producing output per employee and increases depreciation of inputs. This exponentially reduces the rate of capital accumulation over time. Hence, the burden of corruption should be compounded.

Economic growth refers to growth of output f(k), which depends on capital accumulation k, the capital per employee. Equation (2) shows that accumulation of capital k is exponentially related to savings rate s, f(k)/k, and δ. One can verify the Result 2 by considering these three terms as follows. The bribe money needs to be hidden or quickly spent on conspicuous consumption. Only a small fraction is saved. It diverts productive resource from the entrepreneur to the corrupt official. Hence, bribes reduce the savings rate s. Misappropriation of property by bureaucrats reduces the efficiency of both labour and capital inputs. This degradation reduces the efficiency of capital term f(k)/k, and increases the depreciation term δ .

The above growth theory arguments use several simplifying assumptions, which can be relaxed at the cost of models that are more complicated. See the literature surveyed in Barro and Sala-i-Martin (1995). For example, optimisation by economic agents, open economies, multiple sectors, endogenous growth, etc. can all be adjusted for the effects of corruption. Such extensions can provide greater realism to the basic result that the harm from corruption is large should be compounded. A further extension is to use a simple modification of the usual theory for input augmentation by technological change, (See Barro and Sala-i-Martin, 1995, p.34). The modification is that corruption causes abridgement or contraction of inputs instead of augmentation. A still further extension for “open” economies is to note that corruption discourages foreign direct investment in poor countries. This can cause misallocation of trading patterns and inefficient prices paid for internationally traded goods and services, hurting both the rich and poor countries.

A bribe can saddle a poor country with a white elephant project, where economic costs far exceed the corrupt payment. A corrupt policy environment makes a value-subtracting

investment appear nominally profitable, but ultimately reduce capital accumulation. Mauro's (1995) international cross sectional study also finds that corruption lowers investment and hence economic growth.

2.4 Corruption as Illegal Taxation

I have already listed several sources of the burden of corruption. This subsection continues that discussion and develops the theory needed for the numerical estimation of the burden developed in the following section. Shleifer and Vishny (1993) take an interdisciplinary view of corruption suggesting a vital role of government institutions and political processes, and focus on the effect of corruption on resource allocations. They consider a simplified model of a government produced good, such as an import license or a passport. Next, they assume that a corrupt official can restrict the supply of the good without being detected. Shleifer and Vishny use familiar diagrams from microeconomic theory of monopoly pricing to establish the analogy between commodity taxes and bribes. Wei (1997) considers bribery as a tax from the viewpoint of international investors.

Shleifer and Vishny suggest that we should introduce competition between bureaucrats to reduce corruption. This would involve giving jurisdiction to more than one bureaucracy to issue permits, perhaps with free entry. Although it is hard to imagine completely free entry of official permit-givers, they should have more open record keeping and greater supervision. Using the Internet can create greater transparency.

The graphics from economics textbooks describing Chamberlin's theory of monopolistic competition may be used to extend Shleifer and Vishny's (1993) results. The extension shows that corruption imposes layers of monopolies or quasi monopolies creating contrived scarcities leading to higher average costs, misallocation of resources and inefficiencies. Shleifer and Vishny also note the following very important result.

Result 3: Since corruption must be kept secret to avoid detection and punishment, it is more distortionary than taxes.

Shleifer and Vishny give specific examples of distortions, where the official encourages substitution in favour of those items, which yield larger and/or less easily detected bribes. Bribe-takers can sometimes favour domestic monopolists by imposing tariffs and taxes on imports, follow rent-seeking activities to waste resources and capital in money laundering activities, demand fancier equipment than needed, and so forth.

3. Measuring the Burden of Corruption: A New Approximation

This section exploits Shleifer and Vishny's (1993) model, establishing the analogy between commodity taxes and bribes, to estimate the marginal excess burden (MEB) of corruption defined later in equation (3) of this section. Since neoclassical growth theory is too aggregative, a more comprehensive and detailed empirical study of distortions due to corruption can be based on a careful computation of the MEB. Following Browning (1987) and Stuart (1984) let us define MEB as the “per dollar surcharge that must be borne due to reduction in social welfare when tax revenue increases by one dollar.” Any tax system imposes some distortions and burdens on a productive economy. The welfare loss to the households occurs when taxation forces the consumer to a lower utility level.

Let T_m denote household's total labour time of which L_n is devoted to taxed uses in normal employment and let H_m refer to home production or leisure. Thus, by definition, $T_m = L_n + H_m$. Let w denote the average wage rate and let τ denote average tax rate on labour income wL_n. The government budget is assumed to balance with revenue equal to expenditures. That is, $\tau\ wL_n = G+R$, where G denotes government consumption and R denotes government transfer payments. Further assume that we have a linear relation: $G=a+b(\tau\ w\ L_n)$, whose slope b is used later in eq. (3).

Equilibrium is defined by Stuart (1984) from a numerical solution of ten equations of his complete model involving production and utility functions and first order conditions for static utility maximisation. Let ΔC denote the numeraire value of the reduction in household utility because taxes transfer money from the household sector to the government, leaving less for the household to consume. The ΔC is measured as the amount of taxed-sector output that would just restore the household to the original utility level. Let $\Delta(\tau\ w\ L_n)$ denote a change in tax revenue. Stuart then defines a static general equilibrium measure of the burden from legal taxation as

$$MEB= \{\Delta C / \Delta(\tau\ w\ L_n)\} - b \qquad (3)$$

Browning (1987) uses extensions of his own earlier static models to suggest that the benefits from taxation should be in the range of $1.15 to $1.50 to justify a one dollar of distortion caused by every additional (perfectly legal) tax dollar. Stuart's general equilibrium calculations are essentially similar and consistent with Browning's estimates.

Applied general equilibrium (AGE) analysis models, surveyed by Shoven and Whalley (1984), can obviously refine the MEB calculation with much greater detail, richer choice of functional forms and superior dynamic optimisation algorithms. Accordingly, Ballard *et al*(1985) developed a more sophisticated AGE model to estimate the MEB of the US tax system. A typical AGE model considers the effects of a 1% increase in all tax rates; determines the resulting increase in government revenues and reduction in household welfare. These are measured relative to a common numeraire. Ballard *et al*'s (1984) estimate is 0.332.

A comparable estimate in Harrison and Vinod (1992) is 0.45 along with a 95% confidence interval from 0.33 to 0.67. Recall Result 3 that due to secrecy, corruption is “more distortionary than taxes.” Hence, I calculate a lower limit on the distortion cost of corruption by measuring corruption in numeraire units of government tax income. Then, a one-dollar increase in corruption will create a burden equal to $1.45. This result, using Harrison and Vinod's ‘point estimate’ for US data, gives a conservative lower limit on the economic cost of corruption for the US, and seems to be reasonable for developed countries. In developing countries, the burden of corruption is known to be much larger. Hence, I propose using Harrison-Vinod's upper limit of their 95% confidence interval. Ideally, we need a separate AGE study of each poor country's corruption burden. However, such studies are too expensive, if at all possible, since they require data about secret activities where no records are kept.

Thus, a dollar reduction in corruption will benefit developing countries at least $1.67. These AGE estimates of the burden suggest a numerically important negative impact of

corruption. Result 2 of Section 2.3 has shown that we must use compounding to assess the long-term effect of the burden of corruption. Compounding raises $1.67 to a huge burden in just a few years. The estimation of the burden with reference to AGE models for MEB and its compounding are new concepts in this paper.

4. Description of Interdisciplinary Data

This section describes interdisciplinary international data for an empirical study of corruption. I include all relevant variables suggested by researchers in various fields. As with most empirical work in the social sciences, one cannot avoid using some proxies and indices, which only approximate the true underlying concepts.

Measuring corruption is a difficult task, since it is an illegal activity and records of such activities are never kept, and besides, if any information does surface, the bribe-takers try to quickly destroy and discredit it. Some even proceed to punish any informant, his family and his business interests. Corruption must be measured indirectly. I use Transparency International's corruption perception index released in Berlin on September 22, 1998. The index reports perceptions of the degree of corruption as seen by business people, professional risk analysts and the public. It ranges from 0 (highly corrupt) to 10 (highly clean). At least three surveys are used and larger countries receive up to twelve surveys. It is convenient to think of this as a corruption 'purity index,' rather than stick with the Transparency International's CPI terminology, which is confused with the consumer price index. My term purity index "PI" is consistent with the intuition that a higher number (10) is associated with greater purity or less perceived corruption.

For descriptive statistics, I follow the convention of reporting six items in parentheses separated by commas: (Q1, Median, Mean, Q3, N and Standard deviation), where Qi denotes i-th quartile and N is the number of data points. For example, descriptive statistics for "PI" are (2.925, 4.05, 4.778, 6.4, N=78, 2.403), where the convention is to explicitly identify N only. The first number is always Q1(=2.925) and the fourth is Q3(=6.4). Now I list fifteen variables, which are potentially related to corruption and for which data re available. I use UNDP (1998) data for my first ten variables.

1) Lpcgnp=log of per capita GNP in 1997.(1118, 2965, 9108, 16310, N=78, 11490)
2) Grwth=average annual growth rate between 1996 and 1997. (0.975, 2.8, 2.944, 4.825, N=68, 3.459)

3) Lgnp2=log of per capita GNP in 1997 measured in purchasing power parity (PPP) exchange rates and units of 100. (33.55, 64.2, 98.34, 168, N=76, 83.19)

4) Govt=general government consumption as a percent of GDP. (10, 14, 14.71, 18, N=77, 5.304)

5) Trade=trade as a percent of GDP. (46, 62, 75.49, 92.5, N=78, 52.85). Note that for countries like Hong Kong this number is 356, much larger than 100.

6) Exports=exports as a percent of GDP. (22, 30, 36.56, 43.5, N=78, 26.42)

7) Litrcy=adult literacy rate. (81.58, 92.95, 85.66, 99, N=78, 17.15)

8) School=mean years of schooling. (4, 6.45, 6.553, 9.2, N=78, 3.393)

9) Urban=urban population as a percent of total population. (45, 63.5, 61.56, 77, N=78, 21.25)

10) Ineq=percent of income (or consumption) enjoyed by the richest 20% earners. (39.3, 44.7, 45.83, 51.8 N=69, 8.894)

11) EcFree=index of economic freedom defined over the range 0 to 10, where 10 means most free based on 1996 rankings of countries in the *Wall Street Journal*, Dec. 3, 1998. The index goes from 1 for Hong Kong to 152 for the least free Vietnam. Descriptive statistics for the original data are (25.5, 54, 57.41, 88.75, N=78, 38.14). I make them comparable to my dependent variable purity index "PI" as follows. I construct a new index, y=ax+b, where x denotes the original index. Now y should range from 0 to 10. I simply solve two equations in two unknowns a and b such that, for example, y=0 for Vietnam and y=10 for Hong Kong. The solution is: a= (–10/151) and b=10+(10/151). In short, the data are such that EcFree=10 for the country having the greatest economic freedom from Government controls and EcFree=0, for the controlled economy. The new descriptive statistics are (4.189, 6.49, 6.264, 8.377, N=78, 2.526).

The following data are from Mauro (1995), and detailed descriptions are omitted for brevity. The descriptive statistics are reported as before.

12) Redtp= red tape or the degree to which regulatory environment forces approvals and permits and create an obstacle to business. (4.5, 6.33, 6.434, 8.5, N=53, 2.275)

13) Effjud= efficiency of the judiciary, as it affects business. (6, 7.25, 7.548, 10, N=53, 2.031)

14) PolStb= political stability and the degree to which the process becomes violent or tends to disintegrate. (6, 7.25, 7.548, 10, N=53, 2.031)

15) EthLF= ethno-linguistic fractionalisation index, which equals

$$\text{EthLF} = 1 - \Sigma^{I}_{i=1} (n_i / N)^2, \qquad (4)$$

where i=1, …, I ; n_i is the number of people in the i-th ethno-linguistic group, and I is the number of such groups. One of the novelties in Mauro (1995) is the use of eq. (4). This index equals 89 for a diverse country like India, 75 for Canada, 50 for United States, 32 for UK, 7 for Brazil and only 1 for a homogeneous society like Japan. The descriptive statistics are: (4.5, 6.33, 6.434, 8.5, N=53, 2.275).

It is convenient to use the above abbreviations of the fifteen variables throughout the paper. The relevance of some in a study of corruption is already discussed or obvious. For example, the first three per capita income and growth variables are obvious choices for economists and have been used before in the literature. However, some further discussion is appropriate at this point.

From the discussion of MEB, it is clear that "Govt" belongs in the list. The trade and export variables are included partly because economic literature suggests that "open" economies promote competition and improved allocation of resources across national boundaries. The empirical results in Rodrik (1998) suggest that open economies with greater participation in international trade tend to have bigger governments. Rodrik argues

that greater reliance on international markets involves greater risks to the national economy and bigger governments are needed to reduce this risk. The presence of foreign competition encourages efficiencies, the need for risk reduction requires a larger bureaucracy, and greater dealings with foreigners offers greater opportunities for easily concealed bribes in a foreign currency paid abroad. Thus, it is an empirical question whether open economies have greater or less corruption.

I include the "Litrcy, School and Urban" as sociological variables often mentioned in the popular discussion of corruption. Recall that I have defined corruption as a property crime somewhat similar to burglary of government resources. Chiu and Madden (1998) have several theoretical arguments showing that increases in income inequality lead to increases in burglary. Since one hypothesis of interest is whether income inequality is positively correlated with corruption, I include the "Ineq" in my list. The inclusion of "EcFree" variable is motivated by references in the popular press including the *Wall Street Journal* that economic freedom leads to greater efficiency, transparency and hence reduced corruption. The "Redtp" variable is analogous to "EcFree" .

The "Effjud" affects the probability that a bribe-taker is caught and punished, which is important in the micro economic discussion above. The World Bank (1997) report discusses why "separation of powers" between the executive and judiciary is needed, and how anti-corruption laws have been abused in some Latin American countries for party politics. The same report also discusses management and procedural reform, reducing of *ex parte* communication between judges and litigants, availability of small claims courts, alternative dispute resolution methods, legal aid to the poor, support of legal education, prevention of bribes extracted by court staff who are gatekeepers to the legal process, etc.

The "PolStb" variable is a proxy for political legitimacy of the current government, which in turn depends on equity, fairness and whether the rules of the game and transaction costs of rent-seeking are known beforehand. Political leaders can use corruption to reward supporters, buy the support of key groups, silence some opponents, manage ethnic diversity, or simply as a resource to fight elections. If this is the case, only a popular outcry against corruption, which challenges the power structure, can reduce it.

5. Cross-sectional Correlations and Subset Regressions

Having chosen the fifteen variables, I first study the simple correlation coefficients between all pairs of variables. I focus on correlations with the corruption purity index "PI". If N denotes the number of observations, define the sample correlation coefficient with the denominator (N−1). In the available international data sets, we acknowledge the common problem of missing data for different countries and for different variables. For computation of correlation coefficients, it is possible not to waste any information by using information from all available pairs of data. Of course, this means that different country pair correlations are based on different (N−1) values in the denominator.

Countries have a unique history, geography and various special circumstances. Hence, international data are often subject to a serious problem of "outliers." These outliers can hide or distort important general relationships among the majority of countries. However, generally accepted characterisation of individual observations as outliers is not available. The object of "trimming" is to eliminate the influence of a certain pre-specified percent of

extreme observations from the computations, without explicitly calling them as outliers. Such trimming does focus attention on the main relationships. I trim respectively 5, 10 or 20 percent of extreme observations.

I use the S-Plus software (Version 4.5), which does the trimming only after removing missing data, to ensure that the sample size N for different variables is the same. Table 1 reports ranked correlation coefficients between PI (corruption purity index) and 15 listed variables. There are separate columns for correlations after 5%, 10% and 20% trimming. Note that these simple correlations suggest the most important bivariate relations. From the largest to the smallest absolute values of correlation coefficients the listing of the variables from the most important to the least important in this sense is: Redtp, Effjud, Lpcgnp, Lgnp2, EcFree, School, PolStb, Urban, Litrcy, Govt, EthLF, Ineq, Export, Trade and Grwth. Each column of Table 1 has the superscript (1) for the variable with the highest correlation with PI. Superscripts (2), (3) and (15) are respectively for the second highest and third highest and the lowest absolute correlations. The correlation rankings and values change from one column to the next. For example, economic freedom (EcFree) has a correlation of 0.72 with the corruption purity index (PI), which increases to 0.88 upon trimming of 20% of extreme data values.

A clear policy implication of Table 1 ranking is that countries could reduce corruption by reducing redtape, making their judiciary more efficient and increasing economic freedom and schooling. It is also tempting to conclude that increasing per capita GNP will reduce corruption, except that previous studies, including Mauro (1995), have shown that income may be an endogenous variable. Another way of thinking about this is that there is a vicious cycle, where corruption cannot be reduced without increasing the income, and the income cannot be increased without reducing corruption. A structural model for corruption would have to regard income and corruption as jointly dependent variables, subject to the endogeneity problem. In any case, the data suggest that rich countries tend to have less corruption.

In Table 2a note that “School, PolStb, Urban and Litrcy” have a higher correlation with “PI” than “Govt, Ineq, Export, Trade and Grwth.” This hints at some interesting results. For example, social and political variables deserve greater attention than some economic variables. A big government *per se* may not necessarily encourage corruption, if it is well run, with an efficient judiciary and reduced red tape. At first glance, these low correlations cast doubt on Chiu and Madden's (1998) theory showing that increased income inequality leads to increased burglary, since corruption is also a property crime. However, we shall see later that “Ineq” contributes important independent information to the set of variables in my subset regression model.

The correlation between “PI” and “School” is 0.7101 for the entire data and 0.9371 (or the largest) for the 20% trimmed data. This suggests a great potential importance of schooling in reducing corruption, and that this relation is subject to outliers or major exceptions. In general, 20% trimming seems to increase the absolute value of correlation coefficients, except that the correlation with “Trade and Export” decreases after trimming.

Tables 2a and 2b report a 16 x 16 matrix of correlation coefficients among the “PI” and the 15 variables listed above. These matrices are symmetric and this wastes space in re-

porting. Hence, the lower triangular matrices in Tables 2a and 2b have simple correlation coefficients using all available data for each pair, where the calculations are made to ensure unbiased estimators. The upper triangular portions of Tables 2a and 2b contain 20% trimmed correlation coefficients. These are reported for completeness and for the convenience of the reader who may be interested in other pairs of relationships.

5.1 A Subset Regression Model

Now I propose a more sophisticated subset regression model for these data. Since these are interdisciplinary data, the theory from any particular discipline is not likely to be appropriate. Hence, a formal search for a reasonable regression model is based on Mallows' (1973) Cp statistic, which is popular among scientists and engineers, but much less frequently used in social sciences. Akaike Information Criterion (AIC) is related to the likelihood version of Cp by the relation: AIC = s^2(Cp + N), where s^2 is the usual unbiased estimate of residual variance σ^2 in the regression model.

The intuition behind AIC is to maximise the 'information content' of the final subset, and impose a penalty for including too many regressors. The Cp criterion seeks a subset regression, which is closest to the true regression in terms of "mean squared error," or expected squared Euclidean distance between estimates and the true values of a parameter vector β. Before using subset methods, Section 4 has already discussed why the fifteen variables are relevant. Only after including relevant variables, I let a subset algorithm help choose a model according to certain well established criteria and steps. Clearly, these criteria do not favour one discipline over another.

The most general (encompassing) model for the algorithm is:

PI = f { Lpcgnp, Grwth, Trade, Export, Litrcy, School, Urban, Govt, Lgnp2, EcFree, Effjud, Redtp, PolStb, EthLF, Ineq}.

Given the encompassing set, the subset selection algorithm chooses the optimal subset, by the Cp and AIC criteria, leading to the "best" model, generally making sure that the chosen regressors do not have too low t-values. The initial AIC= 38.5924. The algorithm reports Cp for each step and the term with the lowest Cp is dropped first. In this case, "Urban" had degrees of freedom (df=1), Residual sum of squares = 16.93833 and Cp= 37.25011, which is the lowest. Single term deletions and additions are attempted. Extreme versions where only one or two variables are present (besides the intercept) are also considered. The "Trade and Urban" variables are deleted early. The algorithm deletes and adds regressors until it finds the "best" model.

The "best" model parsimoniously explains the variation in the dependent variable "PI" and reported in the first four columns of Table 3. The regressors in the first column yield AIC= 30.5316, which is the lowest achieved by any subset from the complete set of fifteen regressors listed in Section 4.

One problem with this subset is that it includes the "Lpcgnp" (income level) variable as a regressor, which is not exogenous. Hence, I consider a second encompassing set, which excludes "Lpcgnp and Lpgnp2" variables:

PI = f { Grwth, Trade, Export, Litrcy, School, Urban, Govt, EcFree, Effjud, Redtp, PolStb, EthLF, Ineq}.

The "best" model for this encompassing set is similar to the earlier one, except that the "PolStb" is now excluded. The reduced encompassing set raises the ultimate AIC by 0.818 to 31.3496. The resulting coefficients, standard errors and t values are reported in the last three columns of Table 3. If we interpret the absolute size of t values as indicating the importance of the variable, the top three are "EcFree, Redtp, Govt" from the last column.

It is interesting that after including the top two socio-political variables (Effjud and Redtp), other socio-political variables such as "School, Urban, EthLF" do not contribute independent additional explanatory power. Similarly, international trade variables "Trade and Export" are eliminated. The respective correlations of "Grwth" with "Lpgnp2 and Lpcgnp" are only 0.111 and 0.085, in Table 2a. This may be because a poor country with low per capita income *level* can grow faster than a rich country. Hence, I do not initially treat "Grwth" as an endogenous variable. However, when I do exclude "Grwth," the optimal subset of regressors in decreasing order of absolute t-values has "Redtp, EcFree, Govt, and Ineq," where "Effjud" is now absent. For brevity, I omit further details.

Thus, the subset regression model supplements the results of correlation analysis. Simple correlation of "Govt" with "PI" is 0.4756, which becomes 0.7090 after 20% trimming in Table 1. This suggests that bigger government can be helpful in reducing corruption. The optimal subset regressions also include the "Govt" variable. Rodrik (1998) explains the positive role of "Govt." It is interesting to note that "Trade and Export" variables do not enter the optimal regression model and their simple correlations with "PI" *decrease*, rather than increase after 20% trimming. Some authors have argued that international trade and export-oriented "open" economies have to compete in world markets, and are therefore less corrupt. My empirical results do not support this view.

5.2 Implications For a Fight Against Corruption

After combining the results of both correlation and regression analyses the top five actions recommended for reducing corruption in order of importance are: (1) reduce red tape, (2) increase efficiency of the judiciary, (3) increase per capita GNP, (4) increase economic freedom and schooling, and (5) reduce income inequality. To reduce red tape one requires elimination of all unnecessary regulations, government licenses, and permits. Poor countries suffer from chronic shortages of almost everything, and it is tempting to impose rationing, so that most vulnerable people do not unduly suffer from shortage of vital supplies. Unfortunately, higher prices are needed to match supply with demand, to create incentives for greater production and for reduced waste. Rationing becomes politically expedient and creates several regulations that last far too long, and actually end up hurting the most vulnerable, as well as, the honest. A recent World Bank report mentions how well intentioned policies can produce unintended opportunities for corruption. Wherever possible, increased competition among bureaucrats should be injected to reduce red tape.

The second action item above is improving the efficiency of the judiciary. This requires better enforcement of existing laws. That is, the probability of being punished has to increase for both the bribe-giver and bribe-taker.

The third action item in fighting corruption is raising the per capita GNP. I have noted above a vicious cycle when low per capita income itself causes corruption and makes the burden of corruption even greater. A well-designed foreign aid program can break the vicious cycle.

6. Standardised Corruption Database on the Internet

In this section, I describe the details of a proposal for using the Internet to increase incentives for better law enforcement. In light of the 'absentee owner hazard,' (See Result 1) fighting corruption needs consistent and adequate incentives.

The standardised database will contain publicly available information from published and official law-enforcement sources and court documents on convictions and indictments. We will not include any anonymous or private information, which might invite libel problems. In addition, there will be adequate disclaimers and only those who explicitly agree with the disclaimers will get access to the web site. The standardised database categories, subject to revision as we learn more, are:

1) The size of the bribe in local currency and US dollars. Initial focus will be on large bribes.

2) Details about the bribe including (i) accused bribe-taker's name and location, (ii) accused bribe-giver's name and location, (iii) date of the bribe, (iv) nature of the cash and non-cash benefits exchanged (v) the basis for estimating the dollar amount (vi) response or explanation provided by the accused parties.

3) Details about sources of information, including (i) date of publication in a newspaper (ii) page numbers, (iii) copy of the original report with an English translation, etc. (iv) contact information, (v) published opinions on reliability of alleged evidence, etc.

4) The status of prosecution (names and location of prosecutors and their superiors) and court(s) of jurisdiction with key dates, reasons for delay, etc.

5) Names and location of the persons who exposed the corruption, and whether they wish to be named. Details of honours received by them, if any.

The database will help empirical researchers and policy makers in international agencies (World Bank, IMF, etc.). Moreover, the very presence of such data on the Internet will help focus the attention of local corruption fighters, government prosecutors, and the media on important cases. The glare of the Internet is expected to create disincentives for corruption and bring pressure to reduce corruption.

7. **Conclusions**

I use available studies and international cross sectional interdisciplinary data on corruption and fifteen related variables for seventy-eight countries. I estimate that costs of corruption are huge. In many developing countries, a dollar's worth of corruption causes a $1.67 worth of a burden on the economy. Neoclassical economic growth theory is used in Sections 2 and 3 to show that this burden of corruption *compounds* over time to become rather large. Thus, even a small reduction in corruption has a significant payoff for economic development. Since corruption in poor countries is rising in recent years, Section 6

discusses incentive failures and a novel database project for fighting corruption. Using simple, as well as, 5% to 20% trimmed correlation coefficients I determine which variables are related to corruption. I also use information theoretic and statistical criteria (AIC and Cp) to suggest an optimal subset of regressor variables for explaining international variation in corruption. After combining the results of correlation and regression analysis the top five actions recommended for reducing corruption in order of importance are: (1) reduce red tape, (2) increase efficiency of the judiciary, (3) increase per capita GNP, (4) increase economic freedom and schooling, and (5) reduce income inequality.

Mere dissemination of information about human rights abuses such as torture of political dissidents or child labour has helped reduce the abuses in recent years. The taxpayers in developed countries who help give the foreign aid and shareowners of multinational corporations need to be made more aware of the rampant corrupt practices in many countries. Section 6 includes details about a database proposal to use the Worldwide Web to expose corruption and increase the probability of removal of a corrupt entity from power. The long-term benefits of competitive politics, active media, informed civil society and better economic allocations are large and compounded over time.

REFERENCES

Ballard, Charles L., John B. Shoven and John Whalley (1985) General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the Unites States, *American Economic Review* 75, 128-138.

Bardhan, Pranab (1997) Corruption and Development: A Review of Issues *Journal of Economic Literature* 35, 1320-1346.

Barro, R. J. and Xavier Sala-i-Martin (1995) Economic growth. McGraw Hill, New York.

Basu, Kaushik, Sudipto Bhattacharya and Ajit Mishra (1992). Notes on Bribery and the Control of Corruption. *Journal of Public Economics* 48, 349-359.

Becker, Gary (1968) Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76, 169-217.

Browning, Edgar K. (1987) On the Marginal Welfare Cost of Taxation. *American Economic Review* 77, 11-23.

Chiu, W. Henry and Paul Madden (1998) Burglary and income inequality. *Journal of Public Economics* 69, 123-141.

Harrison, G. W. and H. D. Vinod (1992) The Sensitivity Analysis of Applied General Equilibrium Models: Completely Randomized Factorial Sampling Designs. *Review of Economics and Statistics* 74, 357-362.

Intriligator, Michael D. (1971) *Mathematical Optimization and Economic Theory*. Prentice Hall: Englewood Cliffs, NJ

Kaufman, Daniel. (1997) Corruption : The facts. *Foreign Policy*, summer issue, 114-131.

Klitgaars, Robert. (1988) *Controlling Corruption*. University of California Press, Berkeley, California.

Mallows, C. L. (1973) Some comments on Cp. *Technometrics* 15, 661-675.

Leff, N. (1964) Economic Development Through Bureaucratic Corruption. *American Behavioral Scientist*, 8-14.

Mauro, Paolo (1995) Corruption and Growth, Quarterly *Journal of Economics* 110, 681-

712.
Mookherjee, Dilip and I. P. L. Png (1995) Corruptible Law Enforcers: How Should They Be Compensated? *Economic Journal* 105, 145-159.
Rodrik, Dani (1998) Why Do More Open Economies Have Bigger Governments? *Journal of Political Economy* 106, 997-1032.
Rijckghem, C. and B. Weder (1997) *Corruption and the rate of temptation: Do low wages in the civil service cause corruption*. IMF Working paper WP/97/73, June.
Rose-Ackerman, Susan (1996) *When is Corruption Harmful?* A background paper for World Development Report 1997, World Bank, August.
Rose-Ackerman, Susan and Andrew Stone (1996) *The Costs of Corruption for Private Business: Evidence from World Bank Surveys*. World Bank, May.
Santhanam Committee, (1964) *Report by the Committee on Prevention of Corruption.* Government of India, Ministry if Home Affairs, New Delhi.
Shleifer, Andrei and Robert W. Vishny (1993) Corruption. *Quarterly Journal of Economics* 108, 599-617.
Shoven, John B. and John Whalley (1984) Applied general equilibrium models of taxation and international trade: An introduction and survey. *Journal of Economic Literature* 22, 1007-1051.
Stuart, C. E. (1984) Welfare Costs Per Dollar of Additional Tax Revenue in the United States. *American Economic Review* 74, 352-362.
UNDP, (1998) *Human Development Report and The World Bank.* The World Development Report, United Nations Development Program, 1998/99.
Sandmo, Agnar (1999) Asymmetric Information and Public Economics: The Mirrlees-Vickrey Nobel Prize. *Journal of Economic Perspectives* 13(1), 165-180.
Wei, Shang-Jin (1997) *How Taxing Is Corruption on International Investors?* National Bureau of Economic Research Working Paper 6030.
The World Bank (1997) *Helping Countries Combat Corruption: The Role of the World Bank.* September report of a task force on Corruption Action Plan Working Group, coordinated by Mike Stevens.

TABLE 1 CORRELATION COEFFICIENTS BETWEEN PI AND OTHER VARIABLES

Variable	All data	5% trim	10% trim	20% trim
EthLF	-0.4125	-0.5250	-0.5680	-0.7240
Ineq	-0.3843	-0.5230	-0.5748	-0.6905
Grwth	0.0469(15)	0.2442	0.2272	0.2167
Trade	0.2609	0.0969(15)	0.0856(15)	0.1422(15)
Export	0.3090	0.1962	0.1578	0.1817
Govt	0.4756	0.5705	0.6415	0.7090

Litrcy	0.4827	0.6686	0.7321	0.8274
Urban	0.5422	0.7625	0.7917	0.7548
PolStb	0.6319	0.7015	0.7092	0.7324
School	0.7101	0.8861(3)	0.9087(2)	0.9371(1)
EcFree	0.7168	0.8428	0.8626	0.8761
Lgnp2	0.7895	0.8616	0.8795	0.9066
Lpcgnp	0.8268(3)	0.8925(2)	0.9083(3)	0.9269(3)
Effjud	0.8541(2)	0.8655	0.8953	0.9224
Redtp	0.8947(1)	0.9123(1)	0.9229(1)	0.9330(2)

Superscripts (1), (2), (3) and (15) indicate within column ranks.

TABLE 2a COMPLETE CORRELATION MATRIX

Variable	**PI**	**EthLf**	**Ineq**	**Grwth**	**Trade**	**Export**	**Govt**	**Litrcy**
PI	1.000	-0.722	-0.690	0.218	0.141	0.181	0.710	0.829
EthLF	-0.411	1.000	0.410	0.098	0.118	-0.252	-0.362	-0.527
Ineq	-0.384	0.178	1.000	-0.179	0.238	0.210	-0.590	-0.751
Grwth	0.047	-0.181	0.084	1.000	0.314	0.396	-0.131	0.163
Trade	0.261	-0.103	-0.053	0.173	1.000	0.974	0.210	-0.316
Export	0.309	-0. 101	-0.091	0.177	0.985	1.000	0.368	-0.079
Govt	0.476	-0.268	-0.414	0.018	0.034	0.018	1.000	0.130
Litrcy	0.483	-0..522	-0.334	0.081	0.166	0.192	0.384	1.000
Urban	0. 541	-0.590	-0.282	0.151	0.221	0.261	0.319	0.609
PolStb	0.631	-0.310	-0.502	0.119	0.307	0..369	0.149	0.362

School	0.710	-0.532	-0.559	0.021	0.021	0.062	0.496	0.794
EcFree	0.717	-0.308	-0.162	-0.015	0.245	0.267	0.210	0.462
Lgnp[(2)]	0.790	-0.570	-0.369	0.111	0.201	0.261	0.330	0.702
Lpcgnp	0.827	-0.609	-0.400	0.085	0.178	0.242	0..378	0.687
Effijud	0.854	-0.310	-0.453	0.055	0.333	0..360	0.453	0.409
Redtp	0.895	-0.210	-0.304	0.093	0.358	0.410	0.392	0.514

Lower triangle has all data correlations.

Upper triangle has correlations after 20% trimming.

See Table 2b for the remaining columns.

TABLE 2b COMPLETE CORRELATION MATRIX

Variable	Urban	PolStb	School	EcFree	Lpgnp2	Lpcgnp	Effjud	Redtp
PI	0.750	0.731	0.940	0.880	0.908	0.929	0.921	0.930
EthLF	-0.738	-0.578	-0.728	-0.669	-0.805	-0.822	-0.565	-0.648
Ineq	-0.410	-0.721	-0.660	-0.490	-0.659	-0.722	-0.578	-0.580
Grwth	0.192	0.228	0.281	0.181	0.175	0.128	0.240	0.259
Trade	-0.160	-0.057	0.161	0.111	-0.134	-0.084	0.092	0.288
Export	0.122	0.292	0.308	0.261	0.000	0.055	0.291	0.322
Govt	0.370	0.329	0.680	0.370	0.440	0.520	0.412	0.540
Litrcy	0.720	0.808	0.940	0.852	0.857	0.876	0.804	0.702
Urban	1.000	0.649	0.820	0.870	0.830	0.831	0.758	0.800
PolStb	0.450	1.000	0.839	0.839	0.813	0.839	0.890	0.740
School	0.708	0.522	1.000	0.900	0.939	0.958	0.921	0.900
EcFree	0.511	0.472	0.606	1.000	0.909	0.883	0.923	0.880
Lpgnp 2	0.730	0.600	0.819	0.758	1.000	0.991	0.898	0.889
Lpcgnp	0.748	0.671	0.840	0.769	0.968	1.000	0.936	0.892

Effjud	0.548	0.643	0.702	0.502	0.621	0.677	1.000
0.923							
Redtp	0.601	0.609	0.708	0.627	0.691	0.735	0.787
1.000							

Lower triangle has all data, upper triangle has correlations after 20% trimming. See Table 2a for the remaining columns.

TABLE 3: REGRESSION OF PURITY INDEX ON OTHER VARIABLES

Variable	Coeff.	Std.Error	t value	Coeff.	Std.Error	t value
Intercept	-1.4820	1.6081	-0.9216	-11.1173	1.2796	-0.8732
Lpcgnp	0.3954	0.1785	2.2152	-	-	-
Grwth	0.1014	0.0504	2.0127	0.1258	0.0517	2.4324
Ineq	-0.0390	0.0177	-2.2020	-0.0422	0.0178	-2.3783
Govt	0.0922	0.0296	3.1159	0.1125	0.0296	3.8012
EcFree	0.2203	0.0959	2.2979	0.3529	0.0769	4.5894
Effjud	0.2685	0.1280	2.0969	0.1838	0.1155	1.5911
Redtp	0.3913	0.1034	3.7828	0.4347	0.1054	4.1229
PolStb	-0.2885	0.1759	-1.6402	-	-	-

Left Panel: Residual standard error = 0.7571, R-Squared = 0.9357,

F(8,32) = 58.25, with the p-value = 0.0

Right Panel: Residual standard error = 0.7952, R-Squared = 0.9247,

F(6,34) = 69.57, p-value = 0.0