Supervision and Performance: The Case of World Bank Projects

Christopher Kilby
Center for Economic Research
March 1995

I received invaluable input from Frank Wolak and Julie Schaffner of Stanford University. Among the many staff members of the World Bank who assisted me, I am especially grateful to Michel Pommier, Rene Ruivivar, Ying Qian, and Jian-Ping Zhou. I am also indebted to Bill Battaile, Wilfred Chow, Bernie Harragan, Carl Jayarajah, Ty Mitchell, Joanne Salop, Chris Stomberg and the staff of OPR and OED at the World Bank. The content of the paper benefits from comments during seminars at Stanford University and Tilburg University. I gratefully acknowledge a CentER Postdoctoral Fellowship which supported me while I completed this research. However, all opinions expressed and errors are my own.

Tilburg University, P.O. Box 90153, 5000 LE Tilburg
The Netherlands
and
Vassar College, Box 382, Poughkeepsie, NY 12601
U.S.A.

Abstract

This paper explores empirical aspects of the relation between supervision and project performance. I focus on development projects funded by the World Bank and on supervision done by the World Bank. The World Bank is the preeminent international development organization both in terms of money lent and leadership; furthermore, data measuring project performance and supervision are relatively comprehensive. The link between supervision and performance is of theoretical interest because it illuminates one side of World Bank-borrower interaction and of practical interest because supervision is an instrument controlled by the World Bank which may improve project performance.

Data are from 1426 World Bank-funded projects completed between 1981 and 1991. Analysis of the influence of World Bank supervision on project performance uses annual supervision and annual interim performance ratings. The annual updating process which generates the discrete interim ratings is described by an ordered probit likelihood function. Maximum likelihood estimates indicate a positive impact of early supervision on performance; late supervision has significantly less influence. The estimation predicts that a significant and persistent increase in the level of supervision may lead to a gain of several percentage points in the economic rate of return. Because of the size of World Bank-funded projects, the potential gains from increasing supervision far outweigh the costs.

Introduction

This paper examines supervision and its impact on project performance using data from the World Bank. Much of development assistance is via investment projects. The World Bank is the largest international agency involved in project lending with a portfolio of 1,850 projects under implementation presenting a total investment of 300 billion dollars.[1] The importance of World Bank supervision is clear since it may have a role in determining the benefits derived from this vast sum of investment. The primary goal of this paper is to measure the impact of World Bank supervision and to determine under what conditions the impact is greatest.

The econometric analysis draws on data from 1426 projects completed between 1981 and 1991. The performance measures are discrete ratings generated by World Bank project managers and evaluators while the supervision measure is World Bank staff weeks devoted to project supervision. Data are available for both final performance (one rating per project) and interim performance (several annual ratings for each project).[2] Use of the interim measures is preferable because estimation of the impact of supervision on performance is complicated by a feedback relation: supervision influences performance which in turn influences subsequent supervision allocation decisions. A probit estimation of the relation between supervision and performance using aggregate ex post measures confounds these causal links. The relationship it measures could not be interpreted as the

---

[1] Figure derived from World Bank (1993) on the assumption that World Bank funding accounts for forty-seven percent of total project cost. This is a nominal figure.

[2] General interest in these measures of performance rests on the assumption that the World Bank’s assessment does not differ dramatically from some abstract ideal.

Project managers also generate a supervision rating following supervision missions to the borrowing country. These missions are roughly semi-annual. The annual rating is most often the same as the last supervision rating but may differ if opinion has changed since the supervision mission.

impact of supervision on performance but rather reflects both the impact of supervision on performance and the impact of performance on supervision allocation. However, because the feedback is over time, annual data allow estimation of an ordered probit model which imposes sufficient structure to isolate the impact of supervision. The model estimated relates lagged annual supervision to annual changes in interim performance.

Interim performance, which in principle is continuous, must be treated as a latent variable because only a discrete performance rating is observed. Furthermore, changes in performance rather than levels must be the focus because the World Bank's rating process annually updates ratings so that changes in ratings directly reflect one year changes in the latent performance variable. This suggests an ordered probit model for performance where the annual *change* in performance is indexed by the annual *change* in the rating.

The estimated performance equation finds past supervision positively related to improvement in performance. Supervision is most effective early in the project and in smaller projects. Subsidiary calculations suggest that the benefits of supervision greatly outweigh the costs. Furthermore, the feedback pattern in the annual equations offers a explanation for the negative correlation observed between cumulative supervision and final performance.

The paper is organized as follows. Section I presents the data, providing summary statistics for ex post and interim performance ratings and for supervision. Possible problems with interim ratings, such as incentives for misreporting, are addressed; ex post and interim ratings are compared.

Section II investigates the feedback relation between supervision and performance. A naive performance equation is estimated with final performance as the dependent variable and aggregate supervision as an explanatory variable; a negative estimated coefficient on supervision suggests a correlation with the error term introduced by aggregation over time. Estimation of a simple supervision allocation equation reinforces the feedback explanation.

Section III constructs an econometric model of interim performance as a function of annual supervision. The relation between performance and performance ratings motivates an ordered probit model in terms of differences rather than levels. Other considerations -- boundary constraints on ratings, a change in the rating system in 1986, and heteroskedasticity -- are incorporated in the likelihood function. The change in performance depends on the previous year's supervision and exogenous variables.

Section IV presents estimation results, focussing on supervision. The results show that early supervision has a much stronger influence than later supervision and that, when translated into dollar terms, the marginal benefit of early supervision is two orders of magnitude greater than the marginal cost. Finally, a simulation demonstrates the consistency of the annual interim performance and final performance estimates.

Section V summarizes results and conclusions and suggests directions for further research.

I. Data

The central variables in this analysis are measures of performance and the number of staff weeks of World Bank supervision. This section describes available performance measures and provides details on supervision activities. Other variables relevant to performance are also presented.

I.1 Performance Measures

Two types of performance ratings are used by the World Bank, final performance ratings and annual interim ratings. Both are discrete and can be thought of as indices reflecting a latent variable -- project performance -- which is best thought of as the expected economic rate of return (calculated using shadow prices). In some cases, this expected economic rate of return is calculated explicitly; in other cases, it is not because the value of benefits is difficult quantify. In the majority of cases, projects are evaluated on cost effectiveness criteria.

Final Performance Ratings

The Operations Evaluation Department of the World Bank issues a final performance rating of Satisfactory / Unsatisfactory (1/0) for each project.[3] Table 1 describes the final performance ratings for 1426 projects completed between 1981 and 1991.[4] The performance of the sample closely matches

---

[3] Completion ratings are initially given by the operational staff (project managers); Operations Evaluation either issues this rating as the official rating or, in a few cases, reverses the rating. Reversals happen almost exclusively with projects rated satisfactory by the operational staff.

The delay between project completion and rating averages two and one half years with reports on unsatisfactory projects having the longer delays. As a result, there has been a surge in the percentage of unsatisfactory projects recently as the evaluation backlog has been reduced. Since this backlog was largely created and eliminated in the period studied, it should not cause a sampling problem.

[4] The sample was determined by data availability: final ratings must be before 1992; at least two consecutive interim ratings must be available (the interim rating system started in fiscal 1980); supervision data must exist for the life time of the project (supervision data begin in fiscal

that of the population in this period: 71.2 % of the projects were rated satisfactory. A regional break-down exposes some heterogeneity with projects in the Africa region performing the worst and those in the Europe, North Africa, and the Middle East region performing best.[5] Among the ten major types of projects, technical assistance and agricultural projects had the lowest proportion of satisfactory outcomes while energy and urban projects had the highest. Projects with below average size loans were more likely to fail than those with above average size loans while projects with below average supervision were less likely to fail than those with above average supervision. However, these correlations need not be causal.[6]

Interim Performance Ratings

World Bank project managers generate interim performance ratings on a scale of 1 (good), 2, 3, 4 (bad).[7] The ratings are collected as part of the Annual Review of Portfolio Performance[8]. Ratings are intended to be relative to initial project goals and, inter alia, are based on the expected economic

---

1972); control variables (notably staff weeks of preparation and macroeconomic indicators) must also be known. In practice, the first two requirements largely determine the sample.

[5] The regional division used in this paper reflects the structure of the World Bank during this period. Subsequently, these four regions have been reorganized into six regions.

[6] The correlation between the final performance rating and loan amount is 0.02 and between the final performance rating and cumulative supervision is -0.18.

[7] Prior to fiscal 1986, the ratings were from 1 to 3. In fiscal 1986, the rating 3 was subdivided into 3 and 4 where 3 represents serious problems being addressed by borrower and 4 represents serious problems not being addressed by borrower. This change is accounted for in the likelihood function presented in the appendix. Official guidelines for ARPP ratings are specified in the March 1989 OD 13.05.

[8] ARPP -- formerly ARIS: Annual Report on Implementation and Supervision.

rate of return.[9] Table 2 describes the annual interim ratings for the same 1426 projects as in Table 1; the sample size of 7461 annual observations indicates roughly 5 observations per project.[10] Average interim ratings indicate a relatively high percentage of satisfactory ratings as 87.8 percent of the observations are 1 or 2. Regional averages for interim performance ratings follow a pattern similar to that for final performance ratings with the top two and the bottom two regions the same as in Table 1 (though within these groups the order is reversed). Sectoral ratings differ dramatically, perhaps reflecting the relative nature of interim ratings. The large loan / small loan and high supervision / low supervision dichotomies follow the same pattern as with final performance: performance is better in projects with large loans and little supervision. Note that supervision is lagged by one year since it is the lagged value which may influence performance. Again, the negative relation does not indicate a causal direction since

---

9 Generally, there is no recalculation of an expected economic rate of return. Hence, this must be interpreted as a rating which reflects the same factors which would influence the rate of return. The rating also depends on other variables: development impact, availability of counterpart funds, procurement performance, etc. This clouds the World Bank's notion of economic analysis to some degree. In theory, a rating based on economic analysis is the best prediction of final performance (the project's contribution to social welfare) given all information available. With this interpretation, the only other dimension is on the World Bank's side -- how smoothly the administration is proceeding, independent of its effects on project performance. However, the evaluation form includes other ratings related to project performance separately from economic analysis. For this reason, I use the "overall" rating rather than the economic rating. The overall rating is also the focus of World Bank management.

10 This figure is less than the average length of projects in the sample (7.4 years) because some projects began before 1980 and a few projects are missing data (interim ratings or macroeconomic data for specific years). The data set includes observations for: 1) projects canceled before completion (the years before cancellation); 2) projects completed as scheduled; and 3) projects extended beyond the planned closing date (including the years after the planned closing date). Note that project length may be endogenous since the decision to cancel or extend project implementation depends on performance.

performance ratings are serially correlated.[11]

Table 3 mirrors Table 2 but describes the annual *change* in the performance rating and hence the sample is reduced to 6027.[12] For ease of interpretation, I have defined the change in performance to be *positive* if performance *improves*. Transitions of from -3 to +3 are possible though the extremes are not observed. The pattern by region and sector is different from both previous tables. There is no apparent link between loan size and the change in performance rating but the link between supervision and change in performance rating is reversed: projects with more supervision are more likely to show improvement.[13] Table 3 also describes the depth of the time series in this unbalanced panel data. Somewhat more than half the projects would have to be dropped from the sample to have a moderately long (5+) time series for each project.

Reliability of Ratings

Final performance ratings are likely to be the most reliable of the measures of project performance. The Operation Evaluation Department which is ultimately responsible for these ratings is autonomous and specializes in project evaluation. The department was founded nearly a decade before the starting date of the data set; its staff have considerable experience in World Bank operations prior to joining the department. This autonomy and experience coupled with relatively clear rating procedures promotes consistent evaluation. While there may be a general

---

[11] The correlation between interim performance ratings and loan amount is −0.10, between interim performance ratings and annual supervision 0.20 and between performance and supervision the previous year 0.17. The serial correlation of interim performance ratings is 0.65.

[12] Eight of the 1426 projects are missing one interim performance rating midway in their time series. First differencing the ratings therefore reduces the sample by 1434 observations.

[13] The correlation between the change in interim performance rating and loan amount is 0.01 and between the change in interim performance rating and annual supervision is 0.03. The serial correlation of changes in interim performance rating is -0.19, driven largely by boundary constraints.

bias (upward or downward) relative to some other independent evaluator, there is no reason to suspect bias as a function of some project characteristic (region, sector, level of supervision).

However, interim ratings by project managers deserve closer scrutiny. Three potential sources of bias must be considered: 1) Gaming -- up-grading a portfolio gradually to display improvement; 2) Shirking -- underestimating or ignoring performance problems to lighten the supervision workload; and 3) Ex post justification -- reporting improvement after intensive supervision to justify expenditure or personal effort.

Possibilities for gaming depend on the career incentives for project managers. Project managers are generally rotated every three to five years to promote professional development and maintain objectivity. One view of the interim rating process is the "gaming" by a new project manager of the rating system by giving projects low ratings initially and then gradually improving them regardless of actual events. However, two factors operate against this scenario: first, ratings are reviewed by higher management and country teams so that significant gaming is discouraged; and second, until recently, the career incentive system placed significantly more weight on new lending than on supervision performance. The rarity of gaming is supported by available data; for projects under implementation in both 1991 and 1992, a dummy variable for a change in Bank project management was statistically insignificant at the 90 % confidence level.[14]

---

[14] The test was performed using the same form of the likelihood function as is reported in the appendix. The coefficient was of the expected sign (i.e., a change in management is linked with a worsening of the ratings) but was small and statistically insignificant.

Details of the institutional arrangements also mitigate shirking issues in rating. A priori, one might expect a project manager to underestimate if not ignore performance problems to lighten the supervision workload. However, a project manager is involved with the supervision of both his own and other projects. For his own projects, the project manager coordinates, drawing on a team of managers and consultants particularly during supervision missions to the borrowing country. Likewise, he may assist in the supervision of projects not in his own portfolio. As a result, the amount of supervision done by a project manager need not depend on the supervision allocated to his own portfolio. This breaks the link between a manager’s portfolio’s performance and his supervision workload and reduces the incentive to inflate ratings.[15]

Finally, one might expect an ex post justification bias whereby intensive supervision or investment of personal effort automatically results in improved ratings. However, the potential for bias is reduced by the number of participants in the process and the possibility of a management review of the rating. Furthermore, as we shall see in Section IV, ex post justification bias is inconsistent with the estimation results.

Interim ratings are likely to be noisier than final ratings. They are generated by a large, diverse group of people with less rating experience than the Operations Evaluation staff and their rating guide-lines are not as clear. However, interim performance ratings prove to be necessary for assessing the impact of supervision.

---

[15] However, since the 1987 reorganization, project managers are more likely to supervise a project alone or with consultants. Given the availability of consultants, the project manager’s workload is not necessarily effected by the level of project supervision.

Comparing Interim and Final Performance Ratings

The simplest method of comparison is to calculate correlations. The sample correlation between the last interim and the final performance ratings is 0.34. Another approach is to examine transition frequencies. Table 4A presents transition frequencies between interim ratings with the ratings of 3 and 4 combined.[16] No change is the most likely event regardless of the year or the original rating. The large number of 1's relative to 3's (and 4's) makes the average change downward (see Tables 2 and 3) despite the apparent frequency of upward transitions.

The transition frequency matrices 4B and 4C allow comparison of interim and final ratings. Table 4B groups interim ratings into 1,2 and 3,4 for comparison with final ratings.[17] The final matrix summarizes transitions from the last interim rating (grouped as 1,2 and 3,4) to the final performance rating. Initially unsatisfactory projects behave in a similar manner in both transition matrices but projects rated satisfactory are more likely to switch to unsatisfactory in the final transition than in interim transitions. In other words, interim and final ratings agree as closely as can be expected for projects with an unsatisfactory interim rating but too many projects receive satisfactory interim ratings (according to the final performance measure).[18] This describes the "bias" which may be present in

---

[16] Otherwise, three matrices must be examined: a pre-FY86 3 by 3, an FY86 3 by 4, and a post-FY86 4 by 4. There are no dramatic differences between these matrices though comparing the different transition frequency matrices does support the stated split of 3 into 3 and 4: collapsing the 4 by 4 post-FY86 matrix into a 3 by 3 yields frequencies similar to those in the pre-FY86 matrix.

[17] Table 4B shows transitions for all interim ratings though the matrix describing only the transition from the second-to-last to the last interim rating is similar.

[18] These observations can be formalized with a generalized likelihood ratio test. Define $P_x = Pr(p_{i,t} = 1 \text{ or } 2 | p_{i,t-1} = 1 \text{ or } 2)$, $Q_x = Pr(p_{i,t} = 1 \text{ or } 2 | p_{i,t-1} = 3 \text{ or } 4)$, $P_y = Pr(p_i = 1 | p_{i,T} = 1 \text{ or } 2)$, and $Q_y = Pr(p_i = 1 | p_{i,T} = 3 \text{ or } 4)$ where $p_{i,t}$ is the interim rating for project i in year t, $p_{i,T}$ is the last interim rating for project i, and p_i is the ex post final rating for project i. The test of the hypothesis H_0: $P_x = P_y$ v. H_1: $P_x \neq P_y$ using the

interim ratings and suggests that upward and downward transitions in interim ratings may not be symmetrically determined.

I.2 Supervision Data

Data for supervision include all World Bank staff time recorded for a particular project during its implementation.[19] Supervision encompasses three activities: monitoring, management advising, and technical assistance. Although no activity breakdown is available, one can distinguish between staff types, the two main groups being World Bank regular staff and consultants. Staff tend to do more monitoring and management advising but less technical work. Conversely, consultants do most of the technical assistance, some management advising and less monitoring.[20]

Summary statistics for the sample of projects considered in this study for supervision and other continuous independent variables are presented in Table 5. The first row is for cumulative supervision over the life of the project. The average of 81.7 staff weeks reflects an average project implementation period of 7.4 years. As is illustrated by Figure 1, the maximum value is a significant outlier: only 5 % of the observations have above 182 staff weeks. On the low end, 5 % of the observations have less than 22 staff weeks (Figure 2). Annual supervision averaged 11.8 staff weeks per project with little variation in the group average between the first half and the second half of implementation (although individual projects may

---

GLRT for difference of means of binomials (Larsen and Marx, p. 380) yields a p-value of 0.0000. The GLRT test of the hypothesis H_0: $Q_x = Q_y$ v. H_1: $Q_x \neq Q_y$ yields a p-value of 0.84. These tests confirm that transitions from unsatisfactory ratings are similar between interim and final ratings while transitions from satisfactory ratings are significantly different.

19 The supervision variable excludes time which is not allocated to a specific project and time spent on preparation of the project completion report. The first cannot be allocated consistently and the second is functionally different from supervision.

20 The data do not allow one to distinguish between long-term and short-term consultants. Long-term consultants perform much the same tasks as regular staff; it is the short-term consultants who are generally contracted to carry-out more specialized technical tasks.

have considerable variation).[21] The upper 5 % of projects have 30 or more staff weeks of supervision per year in the early period and 28 in the second period. The lower figures are 1.8 and 1.0 staff weeks. See Figures 3 and 4 for the distribution of annual supervision.

I.3 Other Variables

Other control variables included in the estimations are: administrative region, economic sector, loan amount, World Bank contribution to project preparation, macroeconomic indicators, and institutional variables. The administrative regions are: Africa; Asia; Europe, Middle East and North Africa; and, Latin America and the Caribbean. These correspond to the major operational divisions of the World Bank during the period studied and may capture both administrative / organizational and geographic differences. Economic sectors broadly indicate the type of project.[22] Loan amount is in 1990 US dollars where the conversion from nominal figures uses a US GDP deflator and is based on the third year after board approval, roughly the middle of the project.[23] The loan amount is typically a fraction of the total project cost. Preparation data are in staff weeks and are intended to capture the complexity of the project relative to borrower capabilities. Macroeconomic indicators reflect the economic environment external to the project which may influence

---

[21] Note that supervision is lagged by one year so that supervision in the final year (typically only a fraction of the year) is excluded.

This figure is not representative of current levels as supervision has increase significantly in recent years.

[22] Sectors are listed in Table 1. I follow divisions established by the Operations Evaluation Department except: 1) Sectoral Adjustment Loans have been grouped with Structural Adjustment Loans due to the small number of the former; 2) some Technical Assistance projects and Sectoral Adjustment Loans which were included as subsectors of other sectors were reclassified; 3) Disaster Relief and Multisector projects have been pooled as "Other" because of their small numbers.

[23] I use the first year of projects with a planned length of two years, the second year for projects with planned length of 3 or 4 years, and the third year for all other projects.

performance via relative prices or general externalities. They may also proxy for unobservables such as differences in human capital, "level of development," or the influence of government policy. Institutional variables are constructed by administrative department and measure the availability of supervision resources and the competing demands for these resources. These may be taken as a rough measure of the opportunity cost of supervision.

II. The Role of Supervision

With the data available, we can either examine the relation between final performance and cumulative supervision or interim performance and annual supervision. A priori, the aggregate model is more appealing since final performance ratings may be somewhat more consistent. However, if the World Bank allocates supervision based on past performance, then cumulative supervision will be correlated with the error term in an final performance equation and the estimated supervision coefficient will not have a meaningful interpretation as it confounds the impact of supervision on performance with the allocation of supervision.

This section uses a simple model of annual performance and annual supervision allocation to demonstrate the potential endogeneity of cumulative supervision in a final performance equation. Estimation of the final performance equation results in a negative coefficient on cumulative supervision, apparently due to this endogeneity. Finally, estimation of an annual supervision allocation function confirms the importance of past performance in the allocation of supervision. These results point to estimating the performance equation on an annual level, the task pursue in the next section.

If the World Bank views supervision as a means of improving performance, the allocation of supervision to a project in a given year may depend on the project’s performance. In particular, one would expect the World Bank to allocate more supervision to problem projects than to exemplary ones. If performance responds to supervision, the result is a feedback relationship between supervision and performance. According to this story, estimating the relation between final performance ratings and cumulative supervision suffers from an endogeneity problem since both final performance and cumulative supervision depend on interim performance. The coefficient on supervision combines the effects of supervision on performance with the impact of the allocation mechanism.

A simple model demonstrates this point. The model has two equations, an annual change in performance equation and an annual supervision allocation equation. From these I derive an aggregate equation relating final performance to cumulative supervision and demonstrate that cumulative supervision is correlated with the error term.

Let $P^*_{i,t}$ be the performance of project i at time t (where $P^*_{i,t}$ *increases* as performance improves), $S_{i,t}$ the number of staff weeks of supervision of project i in year t, $\mathbf{X}_{i,t}$ a vector of fixed or exogenous factors influencing project performance, and $\mathbf{Y}_{i,t}$ a vector of fixed or exogenous factors influencing the allocation of supervision. Projects start at t=1 with an exogenously given initial performance, $P^*_{i,1}$. Project i continues for T_i periods: t = 1 to T_i. Define $\Delta P^*_{i,t} = P^*_{i,t} - P^*_{i,t-1}$. The equations in the model are:

$$\Delta P^*_{i,t} = \alpha S_{i,t-1} + \mathbf{X}'_{i,t}\beta + \upsilon_{i,t} \qquad (2.1)$$

$$S_{i,t} = \gamma P^*_{i,t} + \mathbf{Y}'_{i,t}\lambda + \nu_{i,t} \qquad (2.2)$$

where $\upsilon_{i,t}$ i.i.d. $(0,\sigma^2_\upsilon)$

$\nu_{i,t}$ i.i.d. $(0,\sigma^2_\nu)$

and $E(\upsilon\nu') = \mathbf{0}$ (i.e., uncorrelated errors)

I specify the equation in terms of the change in performance, anticipating the results of the next section. The relevant supervision variable is supervision in year t–1 since the dependent variable is the change between years t–1 and t. Why annual and not cumulative past supervision? Performance in any given year depends on cumulative past supervision (among other things) because physical and institutional project components change slowly (barring fires, wars, etc.) and hence performance does not "start anew" each period but rather depends on past events. Therefore, performance depends on cumulative supervision and the *change* in performance depends on the *change* in cumulative supervision, i.e., the annual change in performance depends on annual supervision:[24]

$$S_{i,t-1} = \sum_{r=1}^{t-1} S_{i,r} - \sum_{r=1}^{t-2} S_{i,r}.$$

Returning to equations (2.1) and (2.2), we can derive an equation for final performance:

$$
\begin{aligned}
P^*_{i,T_i} &= P^*_{i,T_i-1} + \Delta P^*_{i,T_i} \\
&= P^*_{i,T_i-1} + \alpha S_{i,T_i-1} + \mathbf{X}'_{i,T_i}\beta + \upsilon_{i,T_i} \\
&= P^*_{i,1} + \alpha \sum_{t=2}^{T_i} S_{i,t-1} + \sum_{t=2}^{T_i} \mathbf{X}'_{i,t}\beta + \sum_{t=2}^{T_i} \upsilon_{i,t} \\
&= P^*_{i,1} + \alpha S_i + \mathbf{X}'_i\beta + \upsilon_i \qquad (2.3)
\end{aligned}
$$

Where S_i, $\mathbf{X}_i$, and υ_i are the summations over t. Although final performance is a function of cumulative supervision (S_i), cumulative supervision is correlated with the error term. To see this, we can calculate the covariance between S_i and υ_i, starting with the individual components:

$$
\begin{aligned}
E(S_{i,t}\upsilon_{i,r}) &= E([\gamma P^*_{i,t} + \mathbf{Y}'_{i,t}\lambda + \nu_{i,t}]\upsilon_{i,r}) \\
&= E(\gamma P^*_{i,t}\upsilon_{i,r}) \\
&= E(\gamma[P^*_{i,1} + \alpha \sum_{j=2}^{t} S_{i,j-1} + \sum_{j=2}^{t} \mathbf{X}'_{i,j}\beta + \sum_{j=2}^{t} \upsilon_{i,j}]\upsilon_{i,r})
\end{aligned}
$$

---

[24] One could argue that there are lagged effects of supervision so that more than one supervision term should be included. This is equivalent to a weighted average specification for cumulative supervision in a performance equation. However, additional supervision terms turn out to be insignificant.

$$= E(\gamma[\alpha\sum_{j=1}^{t-1} S_{i,j} + \sum_{j=2}^{t} \upsilon_{i,j}]\upsilon_{i,r})$$

$$= \gamma\alpha E(\sum_{j=1}^{t-1} S_{i,j}\upsilon_{i,r}) + \gamma E(\sum_{j=2}^{t} \upsilon_{i,j}\upsilon_{i,r})$$

$$= \gamma\alpha\sum_{j=1}^{t-1} E(S_{i,j}\upsilon_{i,r}) + \gamma E(\upsilon_{i,r}\upsilon_{i,r}) \qquad \text{if } t\geq r$$

$$= \gamma\alpha\sum_{j=1}^{t-1} E(S_{i,j}\upsilon_{i,r}) + \gamma\sigma_{\upsilon}^{2} \qquad \text{if } t\geq r \qquad (2.4)$$

$$= 0 \qquad \text{else}$$

Solving this recursively and summing over r and t gives

$$E(S_i\upsilon_i) = -(\sigma_{\upsilon}^{2}/\alpha)*(T_i-1+[1-(1+\gamma\alpha)^{T_i-1}]/\gamma\alpha) \qquad \gamma\alpha\neq 0$$

which is generally non-zero so that S_i is correlated with the error term.[25] The impact of this correlation is clear in the following estimation of equation (2.3).

Probit Model of Final Performance:

Expanding equation (2.3)[26]

$$P^{*}_{i,T_i} = \alpha S_i + \beta_1 + \beta_2 RD1_i + \ldots + \beta_4 RD3_i + \beta_5 SD1_i + \ldots + \beta_{14} SD10_i + \beta_{15} L_i + \beta_{16} PR_i + \beta_{17} G_i + \beta_{18} O_i + \upsilon_i \qquad (2.5)$$

υ_i i.d. $N(0, T_i\sigma_{\upsilon}^{2})$

---

[25] See Appendix 2 for a derivation of this result.

[26] Note that $P^{*}_{i,1}$ is omitted from (2.5). This introduces a problem since $S_{i,1}$ may depend on $P^{*}_{i,1}$. However, $P^{*}_{i,1}$ is unobservable. One solution is to use the first interim rating as a measure of $P_{i,1}$. This would introduce a conceptual complication (interim and final ratings are not directly comparable) and a technical complication (including an indicator in place of the latent variable). For these reasons, $P^{*}_{i,1}$ is assumed to be constant for this estimation.

where[27]

S_i:	cumulative number of staff weeks of World Bank supervision on project i.
$RD1_i$ to $RD3_i$:	region dummies (Africa omitted).
$SD1_i$ to $SD10_i$:	sector dummies (Agriculture omitted).
L_i:	loan amount for project i in US$.
PR_i:	number of staff weeks of World Bank preparation for project i.
G_i:	average growth rate of GDP per capita for country in which project i takes place.
O_i:	average index of openness for the country in which project i takes place: (Exports+Imports)/GDP.

As noted in the previous section, the Operations Evaluation Department issues a satisfactory or unsatisfactory rating based in part on the expected economic rate of return (performance); if it is above 10 %, the project is generally rated satisfactory.[28] Since the specification includes a constant, the critical value can be set to 0 without loss of generality and the performance rating is defined as an index:

$$P_i = \begin{cases} 1 & \text{if } P^*_{i,T_i} > 0 \\ 0 & \text{if } P^*_{i,T_i} \leq 0 \end{cases}$$

With the additional assumption that $\upsilon_{i,t}$ are normally distributed, this results in the standard probit likelihood function except that υ_i are heteroskedastic with variance $T_i\sigma^2_\upsilon$. Heteroskedasticity follows from the derivation of Equation (2.3). Since a project's performance is influenced by random shocks, the

[27] The importance of region, sector, project size, and macroeconomic variables is suggested by previous research on the expected economic rate of return. See Kaufmann (1991), Kaufmann and Wang (1991), OED (1988) and (1990), Pohl and Mihaljek (1992), and Wallace and Silver (1991).

[28] This subsumes two other issues: reported values of the expected economic rate of return and projects with no reported expected economic rate of return value. In the case where inflated values are reported, I assume that the rating is based on the true value (implicitly). For projects with no economic rate of return calculated, I assume that the evaluation process used has a direct mapping into economic rates of return. If the evaluation method is consistent but does not map directly into the 10 percent expected economic rate of return rule, the estimation method is still valid with the caveat that interpretation of the sectoral constant is unclear.

longer the project has been running, the more random shocks it may have received and the greater its unexplained variance will be. The correction is to divide S_i and $\mathbf{X}_i$ by the square-root of T_i (where σ_υ^2 is assumed to be 1).

Maximum likelihood estimates are presented in Table 6 and Figure 5. The striking feature is the negative coefficient estimate for cumulative supervision (equivalently, the downward sloping curve in Figure 5). Over the typical range in the data (from minus one to plus one standard deviation from the mean level of supervision, 26 to 138 staff weeks -- see Figures 1 and 2), the predicted probability of success ranges from 84 percent for low supervision to 61 percent for high supervision. Equally striking though not evident in the table is the robustness of the negative relation.[29]

The intuition behind this negative relation follows directly from the endogeneity argument. If project performance is low, a project will receive added supervision, making it a "high" supervision project. If this extra supervision is only partly successful in raising project performance, many highly supervised projects will still have major problems at project completion. Conversely, if project performance is high, a project will receive less supervision, making it a "low" supervision project. Performance may deteriorate somewhat (or not improve as much as if supervision were high) but is likely to remain above the threshold level.

In general, when supervision allocation is based on performance, the correlation between cumulative supervision and final performance will reflect a combination of the allocation rule and the influence of supervision on performance. If supervision allocation only responds weakly to performance, the correlation will be positive although it will understate the true

[29] The negative relation is robust to changes in the functional form, changes in the specification of supervision, excluding the highest and lowest twenty percent of supervision values, omitting the heteroskedasticity correction, etc.

impact of supervision on performance. If the allocation rule roughly matches the impact of supervision, there may appear to be no aggregate relation between the two. Finally, if supervision allocation is more responsive to performance than performance is to supervision, the correlation will be negative. This last case may explain the probit results.

Supervision Allocation Equation:

The above interpretation of the probit results and the endogeneity argument are based on the assumption that supervision is allocated according to performance. This proposition can be investigated by estimating the supervision allocation equation (2.2). However, in the estimated equation, supervision is allocated on the basis of interim ratings rather than actual performance. Because of the timing of World Bank supervision allocation decisions and of interim performance ratings, ratings in years t-1 and t may influence supervision in year t. Again, the simplest model will serve our purposes.[30]

Expanding equation (2.2)

$$
\begin{aligned}
S_{i,t} = \; & \gamma_0 + \gamma_1 P2_{i,t} + \gamma_2 P34_{i,t} + \gamma_3 P3_{i,t} + \gamma_4 P4_{i,t} + \\
& \gamma_5 P2_{i,t-1} + \gamma_6 P34_{i,t-1} + \gamma_7 P3_{i,t-1} + \gamma_8 P4_{i,t-1} + \\
& \lambda_1 (t/T0_i) + \lambda_2 (t/T0_i)^2 + \lambda_3 RD1_i + \ldots + \lambda_5 RD3_i + \\
& \lambda_6 SD1_i + \ldots + \lambda_{15} SD10_i + \lambda_{16} L_i + \lambda_{17} L_i^2 + \lambda_{18} PR_i + \\
& \lambda_{19} G_{i,t} + \lambda_{20} GR_{i,t-1} + \lambda_{21} O_{i,t-1} + \lambda_{22} Z1_{i,t} + \lambda_{23} Z2_{i,t} + \\
& \lambda_{24} Z3_{i,t} + \lambda_{25} Z4_{i,t} + \lambda_{26} Z4_{i,t}^2 + \lambda_{27} Z5_{i,t} + \nu_{i,t}
\end{aligned}
\qquad (2.6)
$$

---

[30] Since the method of estimation is least squares, it is a minor issue whether we view supervision as a function of performance ratings or of actual performance. In the latter case, ratings proxy for actual performance. I include both the previous year's and the current year's rating because of the timing of events. Supervision budgeting happens at the same time as ratings are collected (P_{t-1}) while the actual amount of supervision may differ from the budgeted level, presumably based on new information (P_t). Including both P^*_{t-1} and P^*_t in equation (2.2) would complicate but not fundamentally alter the endogeneity result.

where

$P2_{i,t}$: 1 if $P_{i,t} = 2$, 0 otherwise.
$P34_{i,t}$: 1 if $P_{i,t} = 3$ and year < 86, 0 otherwise.
$P3_{i,t}$: 1 if $P_{i,t} = 3$ and year $\geq$ 86, 0 otherwise.
$P4_{i,t}$: 1 if $P_{i,t} = 4$, 0 otherwise.
$t/T0_i$: fraction of project i's planned implementation period completed. This may be greater than 1 if the project finishes behind schedule.[31]
$GR_{i,t-1}$: GDP/capita in 1987 US $
$Z1_{i,t}$ to $Z5_{i,t}$: institutional variables. $Z1_{i,t}$ to $Z3_{i,t}$ measure the availability of staff and consultants for supervision within the department managing project i in period t and $Z4_{i,t}$ to $Z5_{i,t}$ measure the supervision workload of the department.
See (2.5) for other variables.

Project performance ratings are represented by dummy variables to allow flexibility. Other variables which influence the allocation of supervision, such as phase of implementation (timing), region, sector, loan amount, project complexity relative to borrower capabilities (proxied by preparation), macroeconomic conditions, and the opportunity cost of supervision (measured by institutional variables), are included as controls.[32]

Table 7 and Figure 6 present regression results. Evidently, interim performance is a major consideration in the allocation of supervision; Figure 6 illustrates that a project with consistently poor ratings (3,3) receives 50 % more supervision than a model project (1,1). However, projects with major problems which are not being corrected by the borrower (4,4) receive less supervision possibly because some of them are

---

31 Although t/T_i (the fraction of the actual implementation period completed) may seem like a more appropriate measure of implementation progress, T_i must be treated as endogenous.

32 Institutional factors are intended to reflect internal World Bank conditions which may affect the allocation of supervision. The variables are constructed by department since supervision resources may be transferred more easily within departments than between them. Variables measure the total amount of supervision available to the department and the number of other projects competing for that supervision. The number of projects per department is sufficiently large so that endogeneity is not a serious issue.

inactive.[33]

These results support the conjecture that the allocation of supervision to problem projects contributes to the negative coefficient on cumulative supervision in the final performance equation. The estimation with final performance fails to identify the impact of supervision on performance because it does not control for the influence of prior performance on supervision allocation. To include this consideration, interim performance data are needed. The next section introduces a statistical model for estimating the impact of annual supervision on performance using interim ratings.

---

[33] Figure 6 summarizes the regression coefficients. The x-axis is the sum of P_{t-1} and P_t. The y-axis is the sum of the two coefficients on these dummy variables. When more than one combination of ratings yields the same x-value (e.g., 4: 1+3, 2+2, 3+1), the average of the relevant coefficient sums is plotted.

The responsiveness of supervision allocation to performance varies by administrative region with the Asia department being the most responsive and Africa the least.

III. Annual Performance Equation

This section develops a statistical model relating annual supervision to interim performance. As with final performance, interim performance is not directly observed but is measured by a discrete rating. Therefore, the key issue in constructing the likelihood function is the manner in which ratings reflect performance. When examining final performance, a threshold model linking performance to ratings is straightforward. However, with a sequence of interim ratings, a number of models are reasonable ex ante. The starting point is a general first order markov process but, for reasons of efficiency and interpretation, I consider two reasonable restrictions on the rating process termed "recalculate from scratch" and "annual update." Institutional and empirical evidence, however, support annual update more strongly. Other issues -- boundary constraints on the rating scale, a change in the rating system in 1986, and the covariance structure of the data -- also shape the likelihood function. The resulting model is a restricted first order markov process.

Model of Rating

Since interim performance ratings are made annually and -- at least in theory -- reflect the expected performance of a project conditional on available information, rating can be viewed as a conditional first order markov process. A transition probability matrix $\{q_{AB}\}$ gives the probability of a rating B in year t, conditional on having a rating A in year t-1 and other characteristics $\mathbf{X}_{i,t}$. The general form of such a matrix is simply:

		P_t 1	2	3	4
P_{t-1}	1	q_{11}	q_{12}	q_{13}	q_{14}
	2	q_{21}	q_{22}	q_{23}	q_{24}
	3	q_{31}	q_{32}	q_{33}	q_{34}
	4	q_{41}	q_{42}	q_{43}	q_{44}

The only necessary restrictions are that each element is non-negative and that each row sums to 1. The conditional probabilities in each row can be estimated by ordered probit if we assume that ratings are ordered, that conditioning variables enter as a linear combination, and that there is a standard normal error term.

While an unrestricted first order markov process has the advantage of generality, it has the accompanying disadvantages of unclear interpretation and efficient loss relative to a model with correct restrictions imposed. An appropriate across row restriction may eliminate these problems. The interpretation issue centers on the meaning of estimating four different rating processes. The World Bank does not provide four different rating procedures as is implied by separate estimation of each row. While we may want to let some parameters depend on the previous performance rating (i.e., allow them the vary by row), it is doubtful that there is a useful interpretation of a model in which all the parameters vary.

The second consideration is efficiency. Row by row estimation means that only within row variation in the dependent variables contributes to the estimation of parameters. Any information in between row variance is lost. This is a particularly serious issue when there is a strong systematic difference across rows. As demonstrated by supervision allocation equation (2.6), the level of supervision has a strong dependence on the performance rating and hence its mean does systematically by row.

Other efficiency concerns have to do with the number of parameters as compared with the number of positive observations in rows 3 and 4. Row by row estimation would require grouping these categories because of the number of parameters to be estimated and the lack of positive observations on certain dummy variables within these groups. Even when grouped, 3 and 4 still contain relatively few observations during the early period of project implementation. One of the most interesting questions is how the impact of supervision varies between the early and late stages of implementation; this issue could not be address for these two rows.

These issues of interpretation and efficiency argue for imposing some reasonable across row restriction. The appropriate source for such a restriction is World Bank rating procedure. I consider two rating models which imply across row restrictions, the "recalculate from scratch" model and the "annual update" model.

The recalculate from scratch model is a threshold model of rating where the latent variable is project performance and the observable index is the interim performance rating. In this model, each year the World Bank evaluator observes or calculates the project's performance and then reports "1" if it is above a fixed threshold P_1, "2" if it is above a fixed threshold P_2, etc. Since the evaluator recalculates performance from scratch each year, the previous rating does not enter into the determination of the new rating. Returning to the transition probability matrix, this means that, conditional on $\mathbf{X}_{i,t}$, the probability of a rating B does not depend on the previous rating A: $q_{AB} = q_B$. Thus, all rows of the transition probability matrix are the same.

		P_t			
		1	2	3	4
P_{t-1}	1	q_1	q_2	q_3	q_4
	2	q_1	q_2	q_3	q_4
	3	q_1	q_2	q_3	q_4
	4	q_1	q_2	q_3	q_4

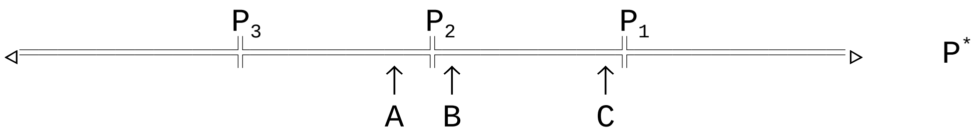
The annual update model follows from a different description of the rating process. In all but the first year of implementation, the previous rating (P_{t-1}) is known. After the first year, the project manager may simply calculate the change in performance since the last rating and update the rating only if there has been a large change. This is again a threshold model but in terms of changes rather than levels. The restriction for the transition probability matrix is that $q_{AB} = q_{A-B}$. Thus, for any top-left to bottom-right diagonal of the matrix, all the elements in that diagonal are identical:[34]

		P_t			
		1	2	3	4
P_{t-1}	1	q_0	q_{-1}	q_{-2}	q_{-3}
	2	q_1	q_0	q_{-1}	q_{-2}
	3	q_2	q_1	q_0	q_{-1}
	4	q_3	q_2	q_1	q_0

Examining the different implications of the two models in a few scenarios provides more intuition than eye-balling transition matrices. In the recalculate from scratch model, the interim performance rating is a direct indicator of the latent performance variable. We can think of each rating (4,3,2,1) as

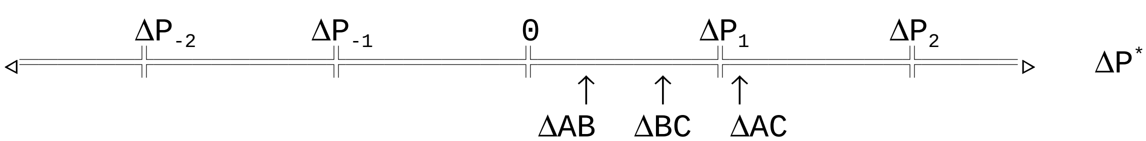
[34] Recall that the change in rating ($\Delta P_{i,t}$) has been defined so that positive values indicate improvement: $\Delta P_{i,t} = P_{i,t-1} - P_{i,t}$. This corresponds to having $q_{AB}=q_{A-B}$ rather than q_{B-A}. Only rating changes of -2 to 2 are observed in the sample; I set $q_3=q_{-3}=0$ henceforth.

corresponding to an interval along a "performance" number line with the numbers P_3, P_2, and P_1 indicating the boundaries between intervals:



A change in the rating in year t occurs when performance in years t-1 and t fall above and below a threshold value. If the two performance values closely straddle the threshold (e.g., A and B), a fairly small year to year change in performance will trigger a change the rating. However, the two may be farther apart but within the same interval and hence the rating remains unchanged (e.g., B and C). This model indicates that inferences about the general level of performance are possible but those about changes in performance may be misleading.

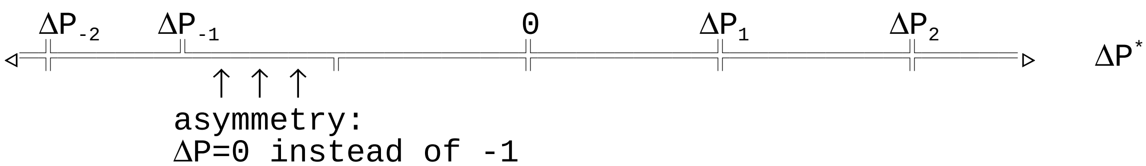
The annual update model has the exact opposite characteristic. A change in the interim performance rating indicates a sizeable one-year change in the latent performance variable. Each change in rating (-2,-1,0,1,2) corresponds to an interval along an "annual change in performance" number line with the numbers ΔP_{-2}, ΔP_{-1}, ΔP_1 and ΔP_2 indicating the boundaries between intervals:



A change in the rating in year t occurs when the difference between performance in years t-1 and t exceeds a threshold (e.g., ΔAC). Thus, a fairly small year to year change in performance will never trigger a change the rating (e.g., ΔAB). However, a series of gradual changes in performance will not be reflected in the rating if none of the annual changes are individually large enough to trigger a rating change (e.g., ΔAB, ΔBC). This model indicates that inferences about the general level of performance and possibly cumulative changes may be inaccurate but those about annual changes in performance are possible.

The choice between these alternative models should be based on rating procedures and empirical evidence. Fortunately, both criteria agree, favoring the annual update model. On the procedural side, the main evidence is from Form 590, the computer form for interim evaluation reports. This form can be characterized as an update: the previous project performance rating is provided next to the blank for the current rating; other fields likewise display previous values next to blanks for current values; and several comment fields request revisions and updates since the last filing of the form. In addition, World Bank staff in the Operations Policy Review Department describe the rating process as one of updating. That is, evaluators consider how much the project’s performance has changed since the previous rating and then update the previous rating accordingly.

The comparison of interim and final rating transitions presented in Section I and Table 4 also supports the annual update model. Recall that downward interim transitions are infrequent relative to downward final transitions while upward interim and final transitions occur with the same frequency. Since the annual update model allows upward thresholds to be asymmetric with downward thresholds, this asymmetry in frequencies could be the result of the location of thresholds. If the threshold for improving the rating is closer to 0 than that for worsening the rating, ratings will tend to go up but not down:



If ΔP_1 is at the "right" value but ΔP_{-1} is "too low," then upward transitions occur with the right frequency and downward transitions occur too infrequently.[35]

This asymmetry cannot be explained with the recalculate from scratch model. The division between satisfactory and unsatisfactory depends on P_2, the threshold between the ratings 2 and 3. If P_2 is too low, then downward transitions are too infrequent but upward transition are too frequent. Conversely, if P_2 is too high, then downward transitions are too frequent and upward transitions are too infrequent. Neither yields the observed asymmetric pattern.

This evidence supports selecting the annual update model as an identifying restriction. Since the annual change in performance is indexed by the annual change in rating, estimation is by ordered probit. However, additional considerations arise because the scale is the *change* in the observed rating.

Boundary Constraints:

Because the annual update model uses the change in the rating as the index variable, boundary constraints must be considered. For example, if $P_{i,t-1} = 1$, then $\Delta P_{i,t}$ may be 0, –1 or –2 but not 1 or 2. Therefore, the relation between the latent variable $\Delta P^*_{i,t}$ and the index variable $\Delta P_{i,t}$ depends on $P_{i,t-1}$. Continuing with the example, if the boundary constraint is not binding, we have:

$$\begin{aligned} \Pr(\Delta P_{i,t}=0) &= \Pr(\Delta P_{-1} \le \Delta P^*_{i,t} < \Delta P_1) = q_0 \\ \Pr(\Delta P_{i,t}=1) &= \Pr(\Delta P_1 \le \Delta P^*_{i,t} < \Delta P_2) = q_1 \\ \Pr(\Delta P_{i,t}=2) &= \Pr(\Delta P_2 \le \Delta P^*_{i,t}) = q_2 \end{aligned}$$

However, if $P_{i,t-1} = 1$, $\Delta P_{i,t}$ cannot be 1 or 2 regardless of the value of $\Delta P^*_{i,t}$ so that the appropriate probability statement is

$$\Pr(\Delta P_{i,t}=0 | \Delta P_{i,t-1}=1) = \Pr(\Delta P_{-1} \le \Delta P^*_{i,t}) = q_0+q_1+q_2$$

---

[35] That is, it takes a lot of bad news to lower a rating but only a little good news to improve it.

These constraints can also be represented in the probability transition matrix:

P_{t-1} \ P_t	1	2	3	4
1	$q_0+q_1+q_2$	q_{-1}	q_{-2}	0
2	q_1+q_2	q_0	q_{-1}	q_{-2}
3	q_2	q_1	q_0	$q_{-1}+q_{-2}$
4	0	q_2	q_1	$q_0+q_{-1}+q_{-2}$

These probabilities are the basis for constructing the likelihood function in Appendix 1.

There are three noteworthy points about this restriction on the likelihood function. First, this is once again a first order markov process since the probability of transition depends explicitly on both states rather than simply on their difference. However, estimation is not row by row since all estimated coefficients are the same across rows. With 5 q's and 14 cells in the matrix, there is still enough overlap that each q_{A-B} is identified. Second, conditional on the annual update model, the imposed boundary constraints are logically valid restrictions and hence improve the efficiency of estimation. Finally, boundary constraints avoid a bias in estimation.

Consider a project with a rating of 1 in year t-1 which, according to the supervision allocation function estimated in Section II, will receive relatively little supervision. If the project improves in spite of the low supervision (e.g., if supervision does not matter), no change will be observed because 1 is the highest rating. If the project's performance deteriorates sufficiently, a change of -1 or -2 may be observed. The result is that projects with a rating of 1 will generally be allocated less supervision while, for unrelated reasons, improvement is never observed. Symmetrically, projects with a rating of 4 will generally be allocated more supervision while deterioration is never observed. If these boundary effects are not incorporated in the likelihood function, the mechanism for allocating supervision will introduce an upward bias in the estimated coefficient on supervision. Ironically, supervision allocation can introduce a downward bias in the final performance estimation and an upward bias in the interim performance estimation.[36]

Modification of the Rating Scale:

Prior to 1986, the rating scale was from 1 to 3, where projects rated a 1 or 2 were considered satisfactory while projects with a 3 were unsatisfactory. The new scale left 1 and 2 unaffected but subdivided 3 into 3 and 4.[37] Because this change just divided an existing category, modelling it is quite straightforward -- all that is required is to modify what ratings changes are possible from each state. Prior to 1986, the transition probability matrix is just the 3 by 3 sub-matrix covering 1 to 3 with the probabilities from the fourth column added to the third column.[38]

Heteroskedasticity:

There are several reasons to suppose that the unexplained variance of the ratings may be related to observables. Variance may differ by region since these regions correspond to different administrative departments within the World Bank which may have slightly different procedures. In addition, certain parts of the world are be more volatile; this volatility may carry over to project performance. Variability may also depend on the type of project and thus depend on the sectoral dummies. It is

---

[36] This reinforces the usual arguments against a linear expectation model.

[37] Analysis of transition frequency matrices for pre-86 and post-86 ratings support the stated change: it does appear that 3 was subdivided and that the definitions of 1 and 2 were unaffected.

[38] The only subtlety is that changes from 3 in FY85 to 4 in FY86 may be spurious (i.e., there may be no change in performance). Therefore, a third matrix just for FY86 adds the q_{34} element to the q_{33} element, replacing q_{34} with 0.

conceivable that larger projects are less volatile than smaller ones or that macroeconomic changes are linked with variability in performance (e.g., high inflation might be good for some projects but bad for others). We might also think that the performance of heavily supervised projects is known more precisely and hence that the ratings of these projects are less prone to fluctuations.

A likelihood ratio test for heteroskedasticity compared a homoskedastic specification with a heteroskedastic specification of $\sigma^2(\mathbf{X}_{i,t})$ varying additively with regional and sectoral dummies and multiplicatively with the continuous independent variables. Heteroskedasticity is confirmed, though successive tests reveal heteroskedasticity by region and sector only. Thus, heteroskedasticity can be represented as $\sigma^2(\mathbf{X}_i)$ since the relevant variables are time-invariant.[39]

Other Specification Issues

Since the data set is a (unbalanced) panel, a fixed or random effect model was considered. Including a project fixed effect would capture any project-specific element of the change in performance. However, several time-invariant variables are included (region, sector, loan amount, preparation); the influence of these factors could not be estimated in a fixed effects model. Furthermore, the estimation procedure itself is problematic, both computationally and theoretically. Since the likelihood function is based on probit, no space or time saving algorithm is available. Although 175 projects which have only one observation on ΔP would drop from the sample, 1251 additional parameters would have to be estimated. More importantly, large sample consistency results do not apply to estimation of fixed effects parameters since T_i is small (ranging from 2 to 11). If the model can be transformed to one which is independent of these incidental parameters (e.g., first differencing in a linear

[39] Several functional forms were investigated; all had similar results.

model), consistent estimators of the remaining parameters may exist; however, no such transformation is known for the probit likelihood function.[40]

These considerations together with the intention of estimating population parameters rather than simply fitting the given sample argue for a random effects approach rather than fixed effects. However, with a random effects model, the covariance matrix ceases to be diagonal since there are non-zero covariances between observations within each project. The resulting likelihood function requires an additional integration over the distribution of the random effect.

Likewise, testing for other forms of autocorrelation, while theoretically desirable, lead to computational problems. The usual test statistics (Wald, Likelihood Ratio, and Lagrange Multiplier) are as computationally complex as the heterosketastic model. This is obvious in the case of the Wald and Likelihood Ratio statistics since they involve estimation under the alternative but is also true of the Lagrange Multiplier statistic. The score and the information matrix involve two integrations for adjacent error terms and thus the statistic does not simplify into the usual T^*R^2 from a regression of quantities readily obtainable from estimation under the null.

A more tractable approach in this situation is to follow the suggestion of Gourieroux, Monfort, Renault, and Trognon (1987) who provide a method for constructing simulated residuals to which standard testing techniques for continuous data can be applied. They prove that this procedure results in a conservative test. Correlograms (Figures 7 and 8) present autocorrelations and partial autocorrelations computed using simulated residuals from the estimation of Table 8. There is some evidence of negative autocorrelation but no identifiable

---

[40] For a discussion of the problems of a probit fixed effects model, see Hsiao (1986), pp. 161-164 or Maddala (1987), pp. 315-317. Although Hsiao notes a 1981 Monte Carlo study by Heckman which suggests the bias may be small, all the evidence taken together makes a random effects model more attractive than a biased fixed effects procedure.

pattern.[41] Given the limited extent of the time series available, I have not pursued estimation of the covariance structure.

Latent Variable Specification:

The annual change in performance equation estimated is a refinement of equation (2.2):

$$\begin{aligned}\Delta P^*_{i,t} = \; & \alpha_1(S_{i,t-1}*H1_{i,t}) + \alpha_2(S_{i,t-1}*H2_{i,t}) + \\ & \alpha_3(S_{i,t-1}{}^2*H1_{i,t}) + \alpha_4(S_{i,t-1}{}^2*H2_{i,t}) + \\ & \alpha_5(L_i*S_{i,t-1}*H1_{i,t}) + \alpha_6(L_i*S_{i,t-1}*H2_{i,t}) + \\ & \beta_1 H2_{i,t} + \beta_2(t/T0_i*H1_{i,t}) + \beta_3(t/T0_i*H2_{i,t}) + \\ & \beta_4((t/T0_i)^2*H1_{i,t}) + \beta_5((t/T0_i)^2*H2_{i,t}) + \\ & \beta_6 RD1_i + \ldots + \beta_8 RD3_i + \beta_9 SD1_i + \ldots + \beta_{18} SD10_i + \\ & \beta_{19} L_i + \beta_{20} PR_i + \beta_{21} G_{i,t} + \beta_{22} GR_{i,t-1} + \beta_{23} O_{i,t-1} + \upsilon_{i,t}\end{aligned} \quad (3.1)$$

$\upsilon_{i,t}$ i.d. $N(0,\sigma^2_\upsilon(\mathbf{X}_i))$

$\sigma^2_\upsilon(\mathbf{X}_i) = 1 + \theta_1 RD1_i + \ldots + \theta_3 RD3_i + \theta_4 SD1_i + \ldots + \theta_{13} SD10_i$

where

$S_{i,t-1}$: number of staff weeks of World Bank supervision in year t-1 on project i.
$H1_{i,t}$: = 1 if $t/T0_i < .5$, 0 otherwise.
$H2_{i,t}$: = 1 if $t/T0_i \geq .5$, 0 otherwise.
See (2.5) and (2.6) for other variables.[42]

As before, the central issue is the role of supervision. To allow for the possibility that the impact of supervision varies over the life of the project, supervision is divided into early and late supervision. This division is a simple method for capturing any time variation in the impact of supervision. To

[41] Random effects should result in positive autocorrelation; smoothing ratings (delaying bad news) should result in positive autocorrelation; gaming by raters is ambiguous in its effect.

[42] Period t-1 to t growth is used rather than t-2 to t-1 because performance data are by fiscal year while macro data are by calendar year.

avoid mistaking pure time differences in the evolution of performance for variations in supervision’s influence over time, other time variables are also included. Hence, the early/late division of supervision should capture only real differences in the impact of supervision on performance.

Interaction terms for supervision and loan amount are included. The effect of supervision on performance may vary by region, type of project, project size, level of preparation, etc. All these interaction terms were considered but, for simplicity, I include only significant interaction terms.

Other variables included are suggested by previous research on the expected economic rate of return. The likelihood function is given in Appendix 1.

IV. Estimation Results

This section presents results from maximum likelihood estimation of the annual change in performance equation specified in Section III. Implications of the estimates are explored with simulations. Tables 8 to 10 and Figures 9 to 11 present the results of the estimation and simulations. I focus primarily on the impact of supervision. As expected this impact is positive though small in the sense that a single week of supervision is unlikely to "turn-around" a troubled project. However, given the size of World Bank funded projects (an average total cost of over 100 million 1990 US dollars), small improvements in performance are large improvements in dollars terms and hence supervision appears to be very worthwhile.

IV.1 Coefficients

Table 8 presents the estimation results. The impact of supervision done during the first half of the project's planned implementation period is positive and significant.[43] The impact of later supervision (during the second half) is two orders of magnitude less than that of early supervision as measured by the expectation derivative.[44] Thus, during the first half of implementation, project performance was more likely to improve if previous supervision was high (all else being equal) while later on supervision has considerably less influence.

Figure 9 depicts the impact of early supervision on performance. As in Figure 5, each point represents the expected

---

[43] A likelihood ratio test of the joint hypothesis H_0: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$ against the alternative H_1: at least one of the coefficients is not 0, compares the LR statistic of 19.3 against a χ^2_6, yielding a p-value of 0.004. Thus, the hypothesis that the previous year's supervision is unrelated to the subsequent change in performance can be rejected at the 95 % confidence level.

[44] The likelihood ratio test of the joint hypothesis H_0: $\alpha_2 = \alpha_4 = \alpha_6 = 0$ against the alternative H_1: at least one of the coefficients is not 0, compares the LR statistic of 8.26 with a χ^2_3 yielding a p-value of 0.041. Therefore, the hypothesis that late supervision has no impact on performance can be rejected at the 95 % confidence level.

change in performance for a particular observation while the curve is the expected change in performance for a typical project as a function of the level of supervision in the previous year.[45] The graph clearly shows the positive relation between supervision and performance. As with any simulation, the extreme values must be interpreted with caution. For example, we would reasonably expect performance to diminish much more than indicated if a project continually received no supervision since this would indicate no checking of disbursement requests or validation of expenses.

The interaction of supervision with other variables was investigated; only the interaction with loan amount was significant. The negative estimated coefficients for these interaction terms in Table 8 indicate that the marginal impact of supervision falls as the size of the loan increases: more supervision is required to achieve a given increase in performance for a project with a large loan than for a project with a small loan. If two projects are similar except that one has a loan of 100 million dollars and the other a loan of 10 million dollars, the larger project requires 22 staff weeks of supervision per year in the early period to have the same influence as 12 staff weeks in the smaller project.[46]

Interaction terms between supervision and region / sector variables are not reported because they are insignificant. This implies that the influence of supervision is similar even across very dissimilar projects in varied environments. One

---

[45] The typical project is one with variables set equal to mean values as in Table 9. See the discussion of Table 9 below for details.

[46] This figure is derived using the estimates in Table 8 and the quadratic formula. The level of supervision S_2 makes the contribution of supervision in the two cases the same if it solves:

$$\hat{\alpha}_1 S_1 + \hat{\alpha}_3 S_1^2 + \hat{\alpha}_5 L_1 S_1 = \hat{\alpha}_1 S_2 + \hat{\alpha}_3 S_2^2 + \hat{\alpha}_5 L_2 S_2$$

where $S_1 = 12$, $L_1 = 10$, $L_2 = 100$, and α's are from Table 8. This does not include the direct effect of a larger loan; if instead we equate the expected change performance in the two projects, then $S_2 = 11.2$.

interpretation of this homogeneity is that the activity that supervision measures (e.g., monitoring) is similar across regions and sectors.

The remaining explanatory variables are ex ante project characteristics (region, sector, loan amount, and preparation) and macroeconomic conditions. These variables were allowed to enter both the conditional mean and variance; however, only region and sector dummy variables were significant in the variance.

The expected change in performance is relatively uniform across regions (with other factors held constant) indicating that observed regional differences in performance ratings are due to initial conditions or systematic differences in other covariates. However, the estimated variances in Asia and Europe, Middle East and North Africa are lower. Either performance changes in the these regions were more predictable than in other regions or the rating process was somewhat more consistent in these departments than in other departments of the World Bank. As noted above, the influence of supervision was uniform across regions.

There is somewhat more variation in the expected change in performance across sectors. Virtually all sectors were more likely than Agriculture to have improving performance though this difference was significant at the 90% level only in Education, Health, and Structural Adjustment operations. Several sectors had more unexplained variance than Agriculture: Development Finance Corporations, Education, and Transportation and Tourism. As noted above, the influence of supervision was uniform across sectors.

Once the interaction of loan amount and supervision is considered, loan amount has a direct positive link to performance. One interpretation of this is that large projects show improving performance. However, loan amount tends to be inversely related to the percentage of the project's total cost financed by the World Bank. Therefore, an alternative explanation is that the positive coefficient reflects better performance in projects with a greater percentage of domestic funding.

The number of staff weeks of World Bank time contributed to project preparation has a negative association with change in performance. As with supervision in the probit estimation, the negative coefficient does not indicate that World Bank inputs are detrimental to performance. Rather, preparation inputs are high when the project is complex or when the borrower's planning capabilities are limited. These same factors also contribute to lower project performance.

Of the macroeconomic factors, only the annual growth rate of GDP per capita has a significant impact on performance. The relevance of macroeconomic conditions to project performance has been noted by previous researchers and in World Bank annual reviews. The insignificance of both the level of GDP per capita and the degree of openness in the change in performance equation implies that the impact of these variables is either on initial performance (i.e., they classify the country as a high or low project performance country) or very gradual and not well captured by annual changes. Common wisdom at the World Bank is that macroeconomic fluctuations have immediate consequences for project performance. If so, these rapid changes are captured by the growth rate.[47]

Finally, the equation estimated does have time dependence as measured by the fraction of the planned implementation period

---

[47] The World Bank's focus on the consequences of macroeconomic conditions is reflected in the supervision allocation equation estimated in Section II in which all three variables are significant.

completed ($t/T0_i$). Upward changes in performance are more likely as the project nears completion. Various other dummy variables (**LAST** = 1 if last year of project, **LATE** = 1 if project past its planned closing date) were investigated but are dominated by relative time.[48] However, the relevance of time cannot be determined from the data. Performance may improve as old problems are solved and, as implementation winds-up, few new problems arise. Borrower and World Bank objectives may converge as project resources become less fungible and public consensus for the emerging project grows. However, there may be spurious reasons for this time dependence such as cohort effects.

Table 9 presents simulation results which illustrate the magnitude of the influence of the various factors on project performance. Table 9 includes all variables which were significant in Table 8 plus the Latin America and Caribbean regional dummy variable. The table presents expectation derivatives and expectation differences which are analogous to probability derivatives and differences in an ordinary probit estimation (see Appendix 1 for formulae). Expectation derivatives are calculated for continuous variables and reflect the impact of a marginal change in the variable on the expected performance of an average project. Expectation differences are calculated for discrete variables or for ranges of continuous variables (e.g., plus and minus one standard deviation from the mean) and reflect the impact of the discrete change in the variable on the expected performance of the project. Since the ordered probit is a nonlinear function, the values of the expectation derivatives and expectation differences depend on the point at which they are evaluated. To be representative of a typical project, time varying variables are set to the sample mean for the early implementation period and time invariant

[48] Four time measures were investigated; only the fraction of project implementation period completed ("relative time") is significant. The four measures are: year of the project (**t** = 1,2,...,$\mathbf{T}_i$), planned length of the project ($\mathbf{T0}_i$), fraction of planned implementation period complete (Relative time = $\mathbf{t/T0}_i$) and fiscal year (**YEAR** = 80,81,...).

variables are set to the project-level means.[49] Examination of all the variables reported in Table 9 is left to the reader; below I discuss only supervision.

IV.2 The Impact of Supervision

The above direct examination of the coefficient estimates provides a lot of information about the impact of supervision on performance. We learn that: 1) World Bank supervision has a measurable, positive impact on subsequent changes in performance; 2) early supervision is much more effective than later supervision; 3) there appear to be diminishing returns to supervision; 4) one staff week of supervision has more impact on the performance rating of a project with a small loan than on the performance rating of a project with a large loan; and 5) the impact of supervision is relatively homogeneous across regions, sectors and macroeconomic conditions. What is not immediately apparent is the magnitude of the impact of supervision on performance. Is it large enough in absolute terms to be of interest? At current levels of supervision, is the marginal benefit of supervision above or below the marginal cost?

The answer to the first question is found in Table 9 and Figures 9 and 10. The first five rows of Table 9 report the absolute and marginal impact of supervision on expected performance for different levels of supervision. Evaluated at

---

[49] Setting time invariant variables to project-level means rather than the early implementation sample means is of relatively little consequence as the means are not very different. However, time varying variables (especially those representing the percent of the planned implementation period completed -- t/T0 and $(t/T0)^2$) should be evaluated at their early period means, both for logical consistency and because the means in the later period are significantly and systematically different.

This rule is used for Early Supervision, Region, Sector, Loan Amount, and Preparation. However Late Supervision and (t/T0) are evaluated at different means. Late Supervision is evaluated with time invariant variables at project-level means but time varying variables at the late implementation sample means. This give the appropriate expectation derivative though makes comparison of expected changes in performance difficult. The time measures (t/T0) are evaluated at the means for the entire sample so that differences reflect only changes in (t/T0) rather than changes in the means.

the sample average (twelve staff weeks of supervision), the marginal impact of early supervision on expected performance is 0.01 on a scale of –2 to 2. Figure 9 presents these simulation results graphically (see discussion above).

The magnitude and meaning of a 0.01 point change in the performance rating is more apparent when it is translated into familiar terms, such as change in the economic rate of return or the net present value of the project. Translating from the change in performance to the change in the economic rate of return is straightforward -- all that is required is multiplication by a conversion factor. I use a conversion factor of 0.05 which is consistent with data for projects where both the final interim performance rating and the re-estimated economic rate of return are reported.[50] The resulting marginal impact of one staff week of supervision on the economic rate of return is 0.05 percentage points (evaluated at the sample mean).

Figure 10 presents the expected economic rate of return as a function of the level of early supervision (analogous to Figure 9). The level of supervision indicated on the X-axis is maintained for the first three years of this typical project.[51] The expected economic rate of return is fixed at the sample mean (15.7 percent) for the average project (12 staff weeks of early

---

[50] For projects in the sample with an expected economic rate of return at completion:

Last ARPP Rating	Average ERR	Number of Projects
1	18.80	210
2	15.95	319
3	8.49	70
4	3.19	14

For all projects with expected economic rates of return at completion in the Operations Evaluation Department’s Annual Review Database through 1991:

Final Rating	Average ERR	Number of Projects
1	19.8	1243
0	3.4	370

The averages for unsatisfactory projects (3/4 for ARPP; 0 for OED) are biased upward since reported rates of return are truncated at -5 percent.

[51] This calculation uses the same means as in Table 9.

supervision for 3 years); changes are relative to this point. The graph indicates that an average economic rate of return of 18% may be achieved if an average annual level of supervision of 35 staff weeks is maintained for the first three years project implementation. As before, an increase of one or two staff weeks has a small impact but a substantial and sustained increase in the average level of supervision may generate a noticeable improvement in the average economic rate of return. Is the extra supervision worthwhile?

If we translate the marginal benefit of supervision into dollars, we can compare it with the marginal cost. I first work through a simple example to illustrate how to convert from the economic rate of return to dollars via the net present value and what the type of results to expect. I then present results for a more realistic case.

Consider an average project in the sample but imagine that all the costs occur in the first year and that all benefits accrue in the second year. The total project cost (C) is 180 million 1990 US dollars, the economic rate of return (δ) is 0.157 and the marginal benefit of supervision in terms of the economic rate of return (δ') is 0.0005.

The level of benefits is implicit in the economic rate of return since we know the cost and the structure of the project. By the definition of the economic rate of return, $C = \frac{1}{1+\delta}B$ or $B = (1+\delta)C$ = 1.157*$180,000,000 = $208,260,000. This must be discounted to a present value: if the discount rate is 10 percent, the present value of benefits is $189,327,273. One additional staff week of supervision increases the economic rate of return to .1575 and the net present value of benefits to 1.1575*$180,000,000/(1.1) = $189,409,091. The extra staff week of supervision increases the net present value of the project by $81,818, many times the cost of an additional week of supervision.

This dramatic difference is driven by the size of World Bank projects. Although the impact of a week of supervision on overall performance is small, the cost of supervision is much smaller relative to the total cost of the project. In the calculation above -- if everything else remained constant -- the marginal benefit of supervision exceeds the marginal cost as long as the project total cost is above $12 million, a condition met by most World Bank-funded projects.[52]

This example demonstrates the procedure for converting the marginal benefit of supervision from economic rate of return terms to dollar terms. The example also points out that the marginal benefit of supervision in dollar terms is likely to be large relative to the marginal cost simply because of leverage due to the size of projects. The actual marginal benefit calculated is not correct, however, because the cost and benefit structures are over-simplified and do not represent the "typical" project.

The marginal benefit calculated with the cost and benefit structure of a typical project is illustrated in Figure 11. For this calculation, I assume that the typical project has a seven year implementation period followed by a ten year benefit period. The cost stream is front loaded as in real projects; I assume that the percentage of the total cost spent in each year is {.1, .25, .4, .1, .05, .05, .05}. The benefit stream is constant over the ten years. Furthermore, I assume that supervision increases benefits uniformly while having no impact on costs. See Appendix 2 for the exact formula for the marginal benefit and for a sensitivity analysis of the assumptions.

The marginal benefit curve in Figure 11 is even higher than that implied by the simple example and much larger than the marginal cost of supervision -- more than a hundred times larger at twelve staff weeks of supervision. The main point to take

---

[52] Of course, the estimated equation shows that the impact of supervision (δ') increases as the project size (loan amount) falls so that even a project below $12 million would have $MB_S > MC_S$.

away from these calculations is that, in both specifications, the marginal benefit of early supervision greatly exceeds the marginal cost at current levels of supervision. This result is quite robust to the assumptions made (though the exact magnitude of the marginal benefit does vary). For a broad range of specifications for cost and benefit streams, the marginal benefit of supervision at current levels exceeds the marginal cost. In fact, the first example provides an effective lower limit on the marginal benefit of supervision.[53]

For a number of reasons, the exact level of the marginal benefit curve and where it crosses the marginal cost curve in the diagram may be somewhat inaccurate and should be interpreted with caution. The nature of the estimation makes it more reliable near the sample average (twelve staff weeks of supervision); conclusions based on behavior at the extremes (near 0 and 80 staff weeks, for example) are less robust. In addition, the graph reflects the assumptions on the structure of costs and benefits discussed above. As the first example demonstrates, the particular structure of costs and benefits has a strong influence on the imputed dollar value of supervision. Finally, if reported economic rates of return are inaccurate (e.g., inflated), then the implied marginal benefit of supervision in terms of dollars will be somewhat lower. Rather than focusing on the exact level of supervision which equates the marginal benefit and marginal cost, it is more appropriate to underline that the benefits of supervision are substantial and that the marginal benefit of supervision greatly exceeds the marginal cost at current supervision levels.

Once again, it is the size of the project which magnifies the impact of supervision and translates a small change in project performance into a large change in dollar terms. For example in the project described in Figure 11, the present value

---

53 This result is "experimental" rather than analytical. The one period cost, one period benefit structure had the lowest MB_S among all the specifications tried.

of the benefits is $230 million at the average economic rate of return of 15.7 percent. If the economic rate of return increases by one percentage point to 16.7 percent, the present value of benefits increases to $248 million -- a gain of $18 million. Thus, the small influence of supervision on performance ratings translates into a large change in the dollar value of the project relative to the cost of supervision. This leverage due to size is similar to that in Margiotta and Miller (1993) where expenditures to mitigate incentive problems of top executives have over a hundred fold return in terms of the expected profits of the firm.[54]

IV.3 Final Performance Simulation

The relation between the estimates of the interim performance ordered probit reported in Table 8 and the final performance probit reported in Section II and Table 6 is not immediately clear. The general argument -- that the differential allocation of supervision to projects with poor interim performance might overwhelm the positive impact of supervision on performance and result in a negative correlation -- is clear; but do the equations estimated for annual performance and annual supervision have this implication? Furthermore, given the different performance measures used and the assumptions required to estimate the interim performance equation, how closely do the final and interim estimations agree? Exploring this issue may better explain the negative aggregate relation between supervision and performance and provide a rough consistency check for the interim performance estimation.

To address this question, I simulate aggregate data with the annual model and compare the results of a probit estimation using these simulated data to those using the actual data (Table 6). The procedure uses the estimated change in performance and

---

[54] Margiotta and Miller measure total benefits rather than marginal benefits since their model is discrete.

supervision allocation equations and the estimated annual update rule to simulate a series of annual ratings and supervision allocations based on initial conditions and exogenous factors. The last interim performance rating, together with project characteristics, is used to simulate final performance data. Summing annual supervision over the life of the project yields cumulative supervision.

The simulation proceeds by year, starting with the second year for which data are available (the first year provides starting values). There are three steps for each year. The first step simulates the change in performance by plugging the exogenous variables and the simulated supervision level for the previous year (or the starting value in the first round) into the estimated change in performance equation and adding a heteroskedastic random error term. The second step uses the annual update rule and the estimated threshold values to convert the change in performance to a change in the rating; the rating change is then added to the previous year's simulated rating (or the starting value in the first round). The third step simulates supervision allocation by plugging exogenous variables and simulated ratings into the estimated supervision equation and adding a (different) random error term. The process repeats until the last year of the project.

The last step is to convert the last interim performance rating into a final performance rating according to conditional transition probabilities. These conditional probabilities were estimated from the actual data with three separate probit equations (which included project characteristics as conditioning variables): one for projects with the last interim rating of 1, another for those with 2 and a third for those with 3 or 4. The estimated equations were used to generate predicted final performance $\hat{P}_i^*$ from the last simulated interim performance rating and project characteristics. The final rating is generated according to the rule: $\hat{P}_i$=1 if $\hat{P}_i^*\geq 0$, 0 otherwise.

Table 10 repeats the probit estimation of Table 6 using the simulated data $\hat{P}_i$ and $\hat{S}_i$. Since the data are simulated, different draws yield different data. To provide a table comparable to Table 6 but representative of simulation results, I repeated the simulation / re-estimation process forty times, ordered the results according to the magnitude of the first supervision coefficient and selected the twentieth one.

As with the true data, there is a negative relation between cumulative supervision and final ratings; supervision coefficients are also the same order of magnitude (–0.015 compared with –0.033 and 0.000014 compared with 0.000061). Agreement among other coefficients is not quite as good: three of seventeen coefficients differ in sign and magnitudes also vary.[55]

The simulation demonstrates consistency between the interim performance estimation and the final ratings. Together the estimated interim performance and supervision allocation equations (and the feedback relation embodied in them) do imply the negative aggregate relationship between final performance and cumulative supervision. In addition, the similarity of the coefficients suggests that the annual model describes the relation between supervision and performance fairly well. More specifically, the concern that interim ratings reflect an "ex post justification bias" which artificially inflates the impact of supervision is not supported by the comparison between final ratings. Rather, the difference between the supervision coefficients for actual and simulated data suggests either an

---

55 The signs of the first supervision coefficient is stable across all draws though there is variance in magnitude. In the forty repetitions, this coefficient ranged from -0.045 and to -.019 with the middle twenty values falling between -0.038 and -0.028. Other coefficients were more variable. Overall variability is demonstrated by the correlation between simulated and actual final ratings. In the example presented the correlation is 0.15; the range in the sample of forty draws is from 0.11 to 0.19.

Estimation based on a larger sample would stabilize the results but comparison with the actual data set would be more difficult. Note that these simulations condition on the starting values of the actual data rather than generating new starting values according to an estimated distribution. This approach is taken because the goal is to see if the interim model explains the final performance results.

underestimate of the impact of supervision in the interim performance equation or an overestimate of the responsiveness of supervision to performance in the supervision allocation equation.

V. Conclusion

This paper attempts to measure the impact of World Bank supervision on the performance of World Bank-funded projects and finds that supervision does improve performance. The model estimated indicates that the benefits from supervision far outweigh the costs simply because of the size of projects. Supervision is most effective early in the implementation period and in projects with smaller loans. While there is little evidence that a single additional week of World Bank supervision will dramatically change project performance, larger and more persistent differences in the level of World Bank supervision can have a significant impact. The average expected economic rate of return appears to be a few percentages points higher with the current level of supervision than if the World Bank did not supervise; substantial increases above the current level of supervision are predicted to have a similar effect. When translated into dollar figures, these small changes in the economic rate of return indicate large gains in dollar terms, gains that appear to justify increased supervision.

If more supervision is justified, what form should it take? Part of the answer to this question is clear: additional supervision would be more productive during the initial stages of implementation. But what type of supervision should be increased? World Bank supervision includes two functionally separate activities: monitoring and assistance. However, these activities are not recorded separately in World Bank data making it difficult to determine how much each activity contributes to the overall impact of supervision.

The difference between supervision as monitoring and supervision as assistance is explored in Kilby (1994). Several observations from that paper are relevant here. According to the argument presented, if monitoring is the more important element of supervision, then the impact of supervision on performance will be homogeneous across regions and sectors and will depend on the source of funds. The observed link between supervision and performance has both of these characteristics. Finally, because no substitute for World Bank monitoring exists, the impact of World Bank supervision as monitoring should be observable. In contrast, substitutes for World Bank assistance are available (e.g., international consultants) and may be purchased by the borrowing government when World Bank supervision is low. If this is the case, World Bank supervision is simply a residual which may be unrelated to the overall level of assistance and, as a result, the impact of World Bank supervision as assistance may not be observable.

Thus, the influence of supervision measured in this paper may be largely attributable to monitoring rather than assistance. While no conclusive evidence is currently available, this does provide a direction for future research -- to distinguish empirically between supervision as monitoring and supervision as assistance.

References

Baum, Warren C. 1982 *The Project Cycle*, The World Bank.

_______________, and Stokes M. Tolbert. 1985 *Investing in Development: Lessons from World Bank Experience*, The World Bank, Oxford University Press.

Dasgupta, Partha, Stephen Marglin, and Armartya Sen. 1972 *Guidelines for Project Evaluation*, Project Formulation and Evaluation Series, no. 2. New York: UNIDO.

Gittinger, J. Price. 1982 *Economic Analysis of Agricultural Projects*, 2nd Ed., The World Bank, Johns Hopkins University Press, Baltimore, MD.

Gourieroux, Christian, Alain Monfort, Eric Renault, and Alain Trognon. 1987 "Simulated Residuals." *Journal of Econometrics*, 34, pp.201-252.

Hsiao, Cheng. 1986 *Analysis of Panel Data*, Cambridge University Press, Cambridge.

Israel, Arturo. 1987 *Institutional Development: Incentives for Performance*, The World Bank, Johns Hopkins University Press, Baltimore, MD.

James, Jeffrey. 1995 (forthcoming) *The State, Technology, and Industrialization in Africa*, Mcmillan, London.

Kaufmann, Daniel. 1991 "The Forgotten Rationale for Policy Reform: The Productivity of Investment Projects." 1990 World Development Report Team, World Bank Mimeo.

Kaufmann, Daniel, and Yan Wang. 1991 "The Impact of Macroeconomic Policies on Project Performance in the Social Sectors." 1990 World Development Report Team, World Bank Mimeo.

Kilby, Christopher. 1994 *World Bank-Borrower Relations and Project Supervision*, Center for Economic Research Working Paper Number 9414, Tilburg, The Netherlands.

Larsen, Richard J., and Morris L. Marx. 1986 *An Introduction to Mathematical Statistics and Its Applications*, 2nd Edition, Prentice-Hall, Englewood Cliffs, New Jersey.

Little, I. M. D., and J. A. Mirrlees. 1969 *Manual of Industrial Project Analysis, Vol. II*, Paris: OECD Development Centre.

_______________. 1974 *Project Appraisal and Planning*. London: Heinemann.

_______________. 1991 "Project Appraisal and Planning Twenty Years On." *Proceedings of the World Bank Annual Conference on Development Economics, 1990*. Washington, DC, pp. 351-382.

Maddala, G. S. 1987 "Limited Dependent Variables Models Using Panel Data." *The Journal of Human Resources*, XXII (3), Summer, pp. 307-38.

Margiotta, Mary and Robert Miller. 1993 "Managerial Compensation and the Cost of Moral Hazard." Carnegie Mellon University, Graduate School of Industrial Administration Working Paper #1993-19.

Operations Evaluation Department. 1989 *Annual Review of Evaluation Results 1988*, The World Bank, Washington, DC.

_______________. 1991 *Annual Review of Evaluation Results 1990*, The World Bank, Washington, DC.

_______________. 1993 *Annual Review of Evaluation Results 1992*, The World Bank, Washington, DC.

Pohl, Gerhard, and Dubravko, Mihaljek. 1992 "Project Evaluation and Uncertainty in Practice: A Statistical Analysis of Rate-of-Return Divergences of 1,015 World Bank Projects." *The World Bank Economic Review*, May, 6:2, pp.255-277.

Ray, Anandarup. 1984 *Cost-Benefit Analysis: Issues and Methodologies*, The World Bank, Johns Hopkins University Press, Baltimore, MD.

Squire, Lyn and Herman G. van der Tak. 1975 *Economic Analysis of Projects*, The World Bank, Johns Hopkins University Press, Baltimore, MD.

Wallace, T. Dudley, and J. Lew Silver. 1991 "Evaluating the Evaluators: Analysis of Initial and Completed Evaluations of Agricultural Projects." Draft, Department of Economics, Duke University, Durham, NC.

World Bank. *The World Bank Annual Report, 1993*. Washington, DC.

Appendix 1: Likelihood Function

This appendix derives the likelihood function discussed in Section III and used in the change in performance estimation (Table 8). I also derive formulae for calculating expected values and expectation derivatives. These formulae are used to generate Table 9 and simulated values presented in the text.

The latent variable model for the change in performance is:

$$\Delta P^*_{i,t} = \mathbf{X}'_{i,t}\beta + \upsilon_{i,t} \qquad \upsilon_{i,t} \sim N(0,\sigma^2_\upsilon(\mathbf{X}_i)) \qquad t>1 \qquad \text{(A1.1)}$$

$$E(\upsilon_{i,t}\upsilon_{j,s}) = 0 \text{ for } i \neq j \text{ and for } t \neq s$$

where **X** includes supervision and β includes the coefficient for supervision. The relation between the latent variable ΔP^* and the observed index ΔP is that of an ordered probit though the threshold rules are complicated by a change in the rating system in fiscal year 1986 and by the limits of the rating scale:

<u>If $P_{i,t-1}=1$</u>

$$\Delta P_{i,t} = \begin{cases} -2 \text{ if } & \Delta P^*_{i,t} < \Delta P_{-2} \\ -1 \text{ if } & \Delta P_{-2} \leq \Delta P^*_{i,t} < \Delta P_{-1} \\ 0 \text{ if } & \Delta P_{-1} \leq \Delta P^*_{i,t} \end{cases}$$

<u>If $P_{i,t-1}=2$ and pre-FY86</u>

$$\Delta P_{i,t} = \begin{cases} -1 \text{ if } & \Delta P^*_{i,t} < \Delta P_{-1} \\ 0 \text{ if } & \Delta P_{-1} \leq \Delta P^*_{i,t} < \Delta P_{1} \\ 1 \text{ if } & \Delta P_{1} \leq \Delta P^*_{i,t} \end{cases}$$

<u>If $P_{i,t-1}=2$ and FY86 or later</u>

$$\Delta P_{i,t} = \begin{cases} -2 \text{ if } & \Delta P^*_{i,t} < \Delta P_{-2} \\ -1 \text{ if } & \Delta P_{-2} \leq \Delta P^*_{i,t} < \Delta P_{-1} \\ 0 \text{ if } & \Delta P_{-1} \leq \Delta P^*_{i,t} < \Delta P_{1} \\ 1 \text{ if } & \Delta P_{1} \leq \Delta P^*_{i,t} \end{cases}$$

<u>If $P_{i,t-1}=3$ and pre-FY87</u>

$$\Delta P_{i,t} = \begin{cases} 0 \text{ if } & \Delta P^*_{i,t} < \Delta P_{1} \\ 1 \text{ if } & \Delta P_{1} \leq \Delta P^*_{i,t} < \Delta P_{2} \\ 2 \text{ if } & \Delta P_{1} \leq \Delta P^*_{i,t} \end{cases}$$

<u>If $P_{i,t-1}=3$ and FY87 or later</u>

$$\Delta P_{i,t} = \begin{cases} -1 \text{ if } & \Delta P^*_{i,t} < \Delta P_{-1} \\ 0 \text{ if } & \Delta P_{-1} \leq \Delta P^*_{i,t} < \Delta P_{1} \\ 1 \text{ if } & \Delta P_{1} \leq \Delta P^*_{i,t} < \Delta P_{2} \\ 2 \text{ if } & \Delta P_{2} \leq \Delta P^*_{i,t} \end{cases}$$

<u>If $P_{i,t-1}=4$</u>

$$= 0 \text{ if } \quad \Delta P^*_{i,t} < \Delta P_1$$
$$= 1 \text{ if } \Delta P_1 \leq \Delta P^*_{i,t} < \Delta P_2$$
$$= 2 \text{ if } \Delta P_2 \leq \Delta P^*_{i,t}$$

These rules are derived by starting with an standard ordered probit relationship and then considering where the rating scale limits what is observed. One way of viewing this is to imagine an intermediate variable ΔP^{**}, the "desired" change in rating. The desired change is only observed when it can be accommodated by the rating scale, i.e., when boundary constraints do not bind.

An ordered probit likelihood function which follows directly from equation (A1.1) and the threshold rules outlined above:

$$L = \prod_{i=1}^{N} \prod_{t=2}^{T_i} (\sum_{j=1}^{4} (\sum_{k=-2}^{2} Y_{i,t,j,k} * \Pr(\Delta P_{i,t}=k | P_{i,t-1}=j, \mathbf{X}_{i,t}))) \qquad \text{(A1.2)}$$

$$= \prod_{i=1}^{N} \prod_{t=2}^{T_i} (Y_{i,t,1,-2}\Phi_{-2} + Y_{i,t,1,-1}(\Phi_{-1}-\Phi_{-2}) + Y_{i,t,1,0}(1-\Phi_{-1}) +$$
$$D86_{i,t}Y_{i,t,2,-2}\Phi_{-2} + Y_{i,t,2,-1}(\Phi_{-1}-D86_{i,t}\Phi_{-2}) + Y_{i,t,2,0}(\Phi_1-\Phi_{-1}) + Y_{i,t,2,1}(1-\Phi_1) +$$
$$D87_{i,t}Y_{i,t,3,-1}\Phi_{-1} + Y_{i,t,3,0}(\Phi_1-D87_{i,t}\Phi_{-1}) + Y_{i,t,3,1}(\Phi_2-\Phi_1) + Y_{i,t,3,2}(1-\Phi_2) +$$
$$D87_{i,t}Y_{i,t,4,0}\Phi_1 + D87_{i,t}Y_{i,t,4,1}(\Phi_2-\Phi_1) + D87_{i,t}Y_{i,t,4,2}(1-\Phi_2))$$

where :

$Y_{i,t,j,k}$ = 1 if $P_{i,t-1}=j$ and $\Delta P_{i,t}=k$, 0 otherwise

Φ_k = $\Phi((\Delta P_k - \mathbf{X}'_{i,t}\gamma)/\sigma(\mathbf{X}_i))$ $(k=-2,-1,1,2)$

$\sigma(\mathbf{X}_i)$ = $\mathrm{sqrt}(1+\mathbf{X}'_i\theta)$ = scale factor (standard deviation)

$D86_{i,t}$ = 1 if Fiscal Year ≥ FY86, 0 otherwise

$D87_{i,t}$ = 1 if Fiscal Year ≥ FY87, 0 otherwise

The Φ terms can be collected to simplify evaluation of the function:

$$L = \prod_{i=1}^{N} \prod_{t=2}^{T_i} (C_1 + C_2\Phi_{-2} + C_3\Phi_{-1} + C_4\Phi_1 + C_5\Phi_2) \qquad \text{(A1.3)}$$

where:

$C_1 = Y_{i,t,1,0}+Y_{i,t,2,1}+Y_{i,t,3,2}+D86_{i,t}Y_{i,t,4,2}$
$C_2 = Y_{i,t,1,-2}-Y_{i,t,1,-1}+D86_{i,t}Y_{i,t,2,-2}-D86_{i,t}Y_{i,t,2,-1}$
$C_3 = Y_{i,t,1,-1}-Y_{i,t,1,0}+Y_{i,t,2,-1}-Y_{i,t,2,0}+D87_{i,t}Y_{i,t,3,-1}-D87_{i,t}Y_{i,t,3,0}$
$C_4 = Y_{i,t,2,0}-Y_{i,t,2,1}+Y_{i,t,3,0}-Y_{i,t,3,1}+D87_{i,t}Y_{i,t,4,0}-D87_{i,t}Y_{i,t,4,1}$
$C_5 = Y_{i,t,3,1}-Y_{i,t,3,2}+D87_{i,t}Y_{i,t,4,1}-D87_{i,t}Y_{i,t,4,2}$

The expected rating conditional on $P_{i,t-1}$ and $\mathbf{X}_{i,t}$ is given by the formula:

$$E(\Delta P_{i,t}|P_{i,t-1},\mathbf{X}_{i,t}) = \sum_{j=1}^{4}\sum_{k=-2}^{2} k*Y_{i,t,j}*Pr(\Delta P_{i,t}=k|P_{i,t-1}=j,\mathbf{X}_{i,t}) \quad (A1.4)$$

$$\begin{aligned} &= Y_{i,t,1}(-2*\Phi_{-2}-1*(\Phi_{-1}-\Phi_{-2})+0*(1-\Phi_{-1}))+ \\ &\quad Y_{i,t,2}(-2*D86_{i,t}\Phi_{-2}-1*(\Phi_{-1}-D86_{i,t}\Phi_{-2})+0*(\Phi_1-\Phi_{-1})+1*(1-\Phi_1))+ \\ &\quad Y_{i,t,3}(-1*D87_{i,t}\Phi_{-1}+0*(\Phi_1-D87_{i,t}\Phi_{-1})+1*(\Phi_2-\Phi_1)+2*(1-\Phi_2))+ \\ &\quad Y_{i,t,4}(0*D87_{i,t}\Phi_1+1*D87_{i,t}(\Phi_2-\Phi_1)+2*D87_{i,t}(1-\Phi_2))\end{aligned}$$

$$\begin{aligned} &= Y_{i,t,1}(-\Phi_{-2}-\Phi_{-1})+Y_{i,t,2}(-D86_{i,t}\Phi_{-2}-\Phi_{-1}-\Phi_1+1)+ \\ &\quad Y_{i,t,3}(-D87_{i,t}\Phi_{-1}-\Phi_1-\Phi_2+2)+Y_{i,t,4}D87_{i,t}(-\Phi_1-\Phi_2+2)\end{aligned}$$

The marginal impact of a variable X_k is given by the conditional expectation derivative:

$$\frac{\partial E(\Delta P_{i,t}|P_{i,t-1},\boldsymbol{X}_{i,t})}{\partial \boldsymbol{X}_{i,t,k}} = \quad (A1.5)$$

$$\frac{\partial \boldsymbol{X}_{i,t}'\gamma/\sigma(\boldsymbol{X}_i)}{\partial \boldsymbol{X}_{i,t,k}} \; (Y_{i,t,1}(\phi_{-2}+\phi_{-1})+Y_{i,t,2}(D86_{i,t}\phi_{-2}+\phi_{-1}+\phi_1)+ Y_{i,t,3}(D87_{i,t}\phi_{-1}+\phi_1+\phi_2)+Y_{i,t,4}D87_{i,t}(\phi_1+\phi_2))$$

where:

$Y_{i,t,j}$ = 1 if $P_{i,t-1}=j$, 0 otherwise

ϕ_k = $\phi((\Delta P_k-\mathbf{X}'_{i,t}\gamma)/\sigma(\mathbf{X}_i))$

To get the expected value of ΔP and the expectation derivative conditional only on **X**, we multiply through by the unconditional probability of the initial state, $q_j=Pr(P=j)$. Dropping unnecessary subscripts, equations (A1.4) and (A1.5) become:

$$E(\Delta P|\mathbf{X}) = \sum_{k=-2}^{2} k*q_j*Pr(\Delta P=k|P_{-1}=j,\mathbf{X}) \tag{A1.6}$$

$$= q_1(-\Phi_{-2}-\Phi_{-1}) + q_2(-D86*\Phi_{-2}-\Phi_{-1}-\Phi_1+1) + q_3(-D87*\Phi_{-1}-\Phi_1-\Phi_2+2) + q_4D86(-\Phi_1-\Phi_2+2)$$

$$\frac{\partial E(\Delta p|\boldsymbol{X})}{\partial \boldsymbol{X}_k} = \frac{\partial \boldsymbol{X}'\gamma/\sigma(\boldsymbol{X})}{\partial \boldsymbol{X}_k}(q_1(\phi_{-2}+\phi_{-1}) + q_2(D86*\phi_{-2}+\phi_{-1}+\phi_1) + q_3(D87*\phi_{-1}+\phi_1+\phi_2) + q_4D86(\phi_1+\phi_2)) \tag{A1.7}$$

These formulae are used to calculate Table 9 and Figures 9 to 11. The parameters γ, θ, and ΔP_k and the probabilities q_j are replaced by their maximum likelihood estimates and, except where otherwise indicated, **X** is set equal to the sample mean. $\hat{\gamma}$, $\hat{\theta}$, and $\Delta\hat{p}_k$ are maximum likelihood estimates from equation (A1.3) and Table 8; $\hat{q}_j$ are simply the sample frequency:

$$\hat{q}_j = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i} Y_{i,t,j}}{\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{l=1}^{4} Y_{i,t,l}} \tag{A1.8}$$

The sample mean of **X** is used for expedience; reporting results for the different categorical variables separately would be unwieldy and un-enlightening. Similarly, use of $\hat{q}_j$ is acceptable when dealing with mean values.

For simulations, one may want to compute the expected value based on a sequence of independent variables $\{X_{i,t}\}$:

$$\tilde{q}_j(P_{i,1},\mathbf{X}_{i,2},\mathbf{X}_{i,3},...,\mathbf{X}_{i,t}) = f(j,t)$$

where f is defined recursively as

$$\begin{aligned} f(j,t) &= \sum_{i=1}^{4} g(j-i,i)*f(i,t-1) && \text{for } t > 1 \\ &= \Pr(P_1=j) && \text{for } t = 1 \\ g(k,i) &= \Pr(\Delta P=k \mid P_{-1}=i) \end{aligned}$$

If we condition on a known initial rating $P_{i,1}$, then $\Pr(P_{i,1}=j) = 1$ for one j and 0 for the other values of j. If we do not condition on an initial rating, then $\Pr(P_{i,1}|\mathbf{X}_{i,1})$ must be estimated. To do this, we first estimate an ordered probit using only the initial period data (the first ratings and the time-invariant variables) and then predict the probabilities $\Pr(P_{i,1}=j|\mathbf{X}_{i,1})$ for j=1,2,3,4. This method assumes that the processes generating the first rating and subsequent changes in performance are unrelated.

Appendix 2
Part I: Covariance Derivation

Equation (2.4) of Section II can be rewritten as:

$$f(t,r) = \rho\sum_{j=r}^{t-1} f(j,r) + m \quad \text{if } t\geq r$$
$$= 0 \quad \text{else}$$

This recursive condition is satisfied by

$$f(t,r) = m(1+\rho)^{t-r} \quad \text{if } t\geq r$$
$$= 0 \quad \text{else}$$

Summing over t = 1 to T-1 for a given r yields[1]

$$\sum_{t=1}^{T-1} f(t,r) = \sum_{t=r}^{T-1} f(t,r) = -m[1-(1+\rho)^{T-r}]/\rho \qquad \rho\neq 0$$

Summing over r from 2 to T yields

$$\sum_{r=2}^{T} -m[1-(1+\rho)^{T-r}]/\rho = -m(T-1+[1-(1+\rho)^{T-1}]/\rho)/\rho \qquad \rho\neq 0$$

In the notation used in Section II, $\rho=\gamma\alpha$, $m = \gamma\sigma_\upsilon^2$, and

$$E(S_i\upsilon_i) = -(\sigma_\upsilon^2/\alpha)*(T-1+[1-(1+\gamma\alpha)^{T-1}]/\gamma\alpha) \qquad \gamma\alpha\neq 0$$

If ρ and σ_υ^2 are non-zero, then $E(S_i\upsilon_i)\neq 0$ except possibly at a single value of ρ. I assume that supervision improves performance and more supervision is allocated to bad projects, i.e., $\alpha>0$, $\gamma<0$, => $\rho<0$. For all T, $E(S_i\upsilon_i)$ is negative for $\rho\in(-2,0)$ with $\lim_{\gamma\uparrow 0}E(S_i\upsilon_i)=0$ and $\lim_{\alpha\downarrow 0}E(S_i\upsilon_i)=\gamma\sigma_\upsilon^2(T-1)(T-2)/2<0$, $T>2$. For T odd, $E(S_i\upsilon_i)$ is strictly negative for all values of ρ; for

---

[1] $\rho=0$ if $\alpha=0$ or $\gamma=0$. If $\alpha=0$, the correlation between S_i and υ_i is irrelevant. If $\gamma=0$, $m=0$ and $f(t,r)=0$ $\forall$ t,r so that S_i and υ_i are uncorrelated.

T even, $E(S_i\upsilon_i)$ crosses the 0-axis once in the range [-3,-2) -- at ρ=-3 for T=4 and approaching ρ=-2 for T approaching infinity.

Part II: Converting to the Marginal Benefit of Supervision in Dollars

The following is the framework for converting the marginal benefit of supervision from economic rate of return terms to dollars terms. It is used to generate Figure 11. I also present a sensitivity analysis which demonstrates that the condition $MB_S > MC_S$ holds over most of the range of supervision values considered even if we relax the assumptions used to calculate Figure 11.

The economic rate of return δ is defined as the solution to:

$$\sum_{t=0}^{T} \left(\frac{1}{1+\delta}\right)^t B_t = \sum_{t=0}^{T} \left(\frac{1}{1+\delta}\right)^t C_t \qquad \text{(A2.1)}$$

where C_t and B_t are the cost and benefit in period t. Define $C_t = C*f_t$ where $\sum_{t=0}^{T} f_t=1$, $f_t\geq 0\ \forall t$; and $B_t = B*g_t$ where $\sum_{t=0}^{T} g_t=1$, $g_t\geq 0\ \forall t$. Equation (A2.1) can be rewritten as

$$B*\sum_{t=0}^{T} \left(\frac{1}{1+\delta}\right)^t g_t = C*\sum_{t=0}^{T} \left(\frac{1}{1+\delta}\right)^t f_t \qquad \text{(A2.2)}$$

C, the nominal total project cost, is known; and reasonable weights $\{f_t\}$, $\{g_t\}$ can be constructed.[2] Therefore, (A2.2) can be solved for B as a function of δ:

---

[2] An economic rate of return is unique if sign(B_t - C_t) switches no more than once (a "single crossing" property) and $\exists$ t s.t. $B_t \neq C_t$. These conditions are satisfied by imposing them directly on $\{f_t\}$ and $\{g_t\}$ (though, in general, there may be weaker conditions which also work). The examples I use satisfy the stricter conditions as do typical World Bank projects.

$$B = C*\frac{\sum_{t=0}^{T}\left(\frac{1}{1+\delta}\right)^{t} f_t}{\sum_{t=0}^{T}\left(\frac{1}{1+\delta}\right)^{t} g_t} \qquad (A2.3)$$

With an expression for B, we can calculate the net present value of the project evaluated in period i:

$$NPV_i = B*\sum_{t=i}^{T}\left(\frac{1}{1+R}\right)^{t-i} g_t - C*\sum_{t=i}^{T}\left(\frac{1}{1+R}\right)^{t-i} f_t \qquad (A2.4)$$

where R is the World Bank's discount rate. The marginal benefit of supervision evaluated in period i is $\frac{\partial NPV_i}{\partial S}$. Assume that $\frac{\partial C}{\partial S}=0$, $\frac{\partial f_t}{\partial S}=0$, and $\frac{\partial g_t}{\partial S}=0$ -- only the nominal level of benefits is affected by supervision. Some structure must be placed on these derivatives; setting them to 0 is the simplest solution.[3] With these assumptions, the marginal benefit of supervision evaluated in period i is:

$$MB_{S,i} = \frac{\partial NPV_i}{\partial S} = \frac{\partial NPV_i}{\partial B}*\frac{\partial B}{\partial S} \qquad (A2.5)$$

The first term simply discounts the change in benefits back to period i:

---

[3] Fixing B and allowing C to vary results in a larger marginal benefit of supervision when $R<\delta$. Hence, the method chosen gives a conservative estimate. Requiring f_t and g_t to be fixed means that any change is costs or benefits is proportional across years. If either f_t or g_t is allowed to vary, the pattern of variation must be specified since several different changes can cause the same change in economic rate of return but result in different changes in the NPV (and vice versa). In any case, it seems likely that supervision has some impact on the actual level of benefits (or costs) rather than just advancing or postponing them.

$$\frac{\partial NPV_i}{\partial B} = \sum_{t=i}^{T} \left(\frac{1}{1+R}\right)^{t-i} g_t = H_0(i,\{g_t\},R) \qquad \text{(A2.6)}$$

The second term in (A2.5) can be derived by differentiating (A2.3) with respect to S:

$$\frac{\partial B}{\partial S} = C*\left(\frac{-\sum_{t=0}^{T} t\delta'\left(\frac{1}{1+\delta}\right)^{t+1} f_t \sum_{t=0}^{T}\left(\frac{1}{1+\delta}\right)^{t} g_t + \sum_{t=0}^{T} t\delta'\left(\frac{1}{1+\delta}\right)^{t+1} g_t \sum_{t=0}^{T}\left(\frac{1}{1+\delta}\right)^{t} f_t}{\left(\sum_{t=0}^{T}\left(\frac{1}{1+\delta}\right)^{t} g_t\right)^2}\right)$$

$$= \delta' * C*\left(\frac{-H_1(\{f_t\},\delta)\,H_2(\{g_t\},\delta) + H_3(\{g_t\},\delta)\,H_4(\{f_t\},\delta)}{H_2(\{g_t\},\delta)^2}\right)$$

$$= \delta' * C * H(\{f_t, g_t\},\delta) \qquad \text{(A2.7)}$$

where δ' is $\partial\delta/\partial S$. Making the dependence on supervision explicit and combining (A2.6) and (A2.7), the expected marginal benefit of supervision measured in dollars evaluated at year i is given by:

$$MB_{S,i}(S) = \delta'(S)*C*H(\{f_t,g_t\},\delta(S))*H_0(i,\{g_t\},R) \qquad \text{(A2.8)}$$

Figure 11 uses equation (A2.8) with R=.1, $\{f_t\}$ = {.1, .25, .4, .1, .05, .05., .05}, $\{g_t\} = \{.1\}_{t=7}^{16}$ and S ranging from 0 to 80 staff weeks. $\delta'(S)$ is given by (A1.7) and (A1.8) using estimates in Table 8 and the conversion factor of 0.05 discussed in the text. $\delta(S)$ is given by (A1.6) and (A1.8) using estimates in Table 8 and the conversion factor but is converted from the expected *change* in the economic rate of return to the expected *level* of the economic rate of return. This "normalization" is achieved by adding the sample average economic rate of return and subtracting the expected change in the economic rate of return at the average level of supervision: $\delta(S) = \bar{\delta} + (\Delta\delta(S) - \Delta\delta(\bar{S})) = 0.157 + (\Delta\delta(S) - (-0.014))$. This has the desired property that the average project at the average level of supervision is expected to have the average economic rate of return. This is the same normalization used in Figure 10 to "locate" the expected economic rate of return curve.

Sensitivity Analysis:

We may be able to characterize how the simulated MB_S varies with δ', C, R, $\{f_t\}$, $\{g_t\}$, and δ and thus identify the direction and magnitude of bias if these variables specified inaccurately.

The first two variables enter (A2.8) linearly so that MB_S is proportional to them: an overestimate of δ' or C results in an overestimate of MB_S. An over-estimate of δ' may be the result of an incorrect conversion from the change in performance rating to the change in economic rate of return. The conversion factor may be incorrect simply because of the method used or because of a conditional or proportional bias in reported economic rates of return. In any case, it is hard to imagine this factor being outside the interval [.02, .10]. Errors in total project cost data may introduce errors of a lessor magnitude. Neither of these would reverse the inequality $MB_S > MC_S$ near the current level of supervision.

The impact of R is also straightforward:

$$\frac{\partial}{\partial R} MB_{S,i} = \left(\frac{\partial}{\partial R} \frac{\partial NPV_i}{\partial S} \right) * \frac{\partial B}{\partial S} = \frac{\partial B}{\partial S} * \sum_{t=i}^{T} -(t-i) * \left(\frac{1}{1+R} \right)^{t-i+1} < 0$$

if the marginal benefit of supervision is positive. Thus, picking a discount rate higher than the World Bank's actual discount rate introduces a downward bias in the simulated MB_S.

The magnitude of such a mistake is again relatively small. For example, $H_0(0,\{g_t\},R=0.05) = 0.58$, $H_0(0,\{g_t\},R=0.10) = 0.35$, $H_0(0,\{g_t\},R=0.15) = 0.22$ for $\{g_t\}$ described above.

The impact of an incorrect distribution of costs can also be characterized for some types of distributions. Any sequence $\{f'_t\}$ can be constructed from the sequence used $\{f_t\}$ and transfers of weight from f_s to f_r. For any transfer from f_s to f_r, $f_s = 1-\sum_{t\neq s} f_t$ and $\frac{\partial f_s}{\partial f_r} = -1$. The derivative of MB_S with respect to f_r is

$$\frac{\partial}{\partial f_r}MB_S = K*\left(-(ra^{r+1} - sa^{s+1})\sum_{t=0}^{T} a^t g_t + (a^r - a^s)\sum_{t=0}^{T} ta^{t+1}g_t\right)$$

$$= K' * \sum_{t=0}^{T}\left((t-r)\,a^r - (t-s)\,a^s\right)a^t g_t \qquad (3)$$

where $a = 1/(1+\delta)$ is between 0 and 1 and K, K' are positive constants. Assume that $r > s$ and $g_t = 0$ for $t \leq r$ (e.g., benefits begin after implementation). When g_t is non-zero, $(t-s)>0$, $(t-r)>0$, and $(t-s)>(t-r)$. Since a is between 0 and 1, $a^s \geq a^r$ and, finally, $(t-s)a^s > (t-r)a^r$. Each term in sum (3) is negative -- the simulated MB_S decreases as costs are shifted to a later period. Conversely, excessive front-loading of the cost structure introduces an upward bias in the simulated MB_S.

Once again, within the reasonable choices for $\{f_t\}$, the possible bias would not reverse the ordering of MB_S and MC_S in the relevant range. For example, if the current front-loaded distribution {.1, .25, .4, .1, .05, .05, .05} were replaced with the smoother distribution {.1, .15, .15, .15, .15, .15, .15}, the marginal benefit of supervision at twelve staff weeks is \$622,850 instead of \$786,663. The new MB_S curve crosses the MC_S curve at 78 staff weeks whereas the old curve crossed at 79 staff weeks.

For $\{g_t\}$, there is no general result for the sign of the bias:

$$\frac{\partial}{\partial g_r} MB_S \quad = \quad K''(b^r - b^s) + K''' * \sum_{t=0}^{T} \left((t-s+2\alpha)a^s - (t-r+2\alpha)a^r \right) a^t f_t$$

where K″, K‴ > 0, b = 1/(1+R) and

$$\alpha = \frac{\sum_{t=1}^{T} t a^t g_t}{\sum_{t=0}^{T} a^t g_t}, \quad 0 \le \alpha \le T$$

The first term in (4) is negative if r > s since 0 < b < 1 but the second term may be positive or negative. Hence, the sign of (4) depends on R, δ, $\{f_t\}$, $\{g_t\}$, r, and s. Here as well, the bias is typically too small to dramatically change the point at which the marginal benefit curve crosses the marginal cost curve.

If the assumed length of the cost and benefit streams is incorrect, a slightly larger bias is introduced. In general, the longer the actual stream of costs and benefits, the greater is the implied nominal benefit (B) and the greater will be the imputed MB_S. The increase in the imputed MB_S is roughly proportional to the length of the cost and benefit streams at the average level of supervision (as can be seen from the examples in Section IV) but the point at which $MB_S = MC_S$ is less variable. In the simple example in the text where all costs are in the first year and all benefits are in the second year, the marginal benefit curve crosses the marginal cost curve at 72 staff weeks of supervision.

An change in δ has an ambiguous but relatively small effect on the calculated value for MB_S. The sign depends on $\{f_t\}$, $\{g_t\}$, and the initial level of δ. For the values used in Figure 11, using δ = 0.10 instead of 0.157 gives a MB_S = $552,500 at 12 staff weeks of supervision; MB_S crosses MC_S at 78 staff weeks.

The conclusion of this sensitivity analysis is that the computed value of MB_S is somewhat sensitive to the assumptions made but the condition that $MB_S > MC_S$ over the relevant range of S is robust.

Appendix 3: Data Sources

This appendix lists data sources and variable definitions. Sufficient detail is provided for reconstructing the data set.[1] All data are from the World Bank. No disaggregate data can be released by the author without the express written permission of the World Bank.

Several databases were used to compile this data set. The Operation Evaluation Department's Annual Review Database (as of the end of 1991) is the source for final performance and some other time invariant project-specific data. This is referred to as OED. Some of the variables in the OED database (loan amount, total project cost, various dates) are originally from either the Financial Database (FDB) or the World Bank's Management Information System (OPMIS). Data for annual interim ratings, annual supervision, and preparation were taken directly from various parts of OPMIS. These parts include the Time Recording System (TRS) and the Annual Review of Portfolio Performance (ARPP). Macroeconomic indicators were constructed from data in the World Bank's Basic Economic and Social Indicators Database (BESD) according to standard definitions. Institutional variables in the supervision equation were constructed with TRS data and OED dates, using the organizational structure implied by Master Organization Codes (MOC) from the Personnel Database.

[1] With some exceptions. Some projects were reclassified into the TA and SECAL sectors because their project names conflicted with their original classification. Though relatively few projects were reclassified, they are too numerous to list here. In addition, the last five "institutional" variables are based on departmental divisions which changed over time; these divisions were reconstructed from the Personnel database but it is not a straightforward task.

Variable Name	Symbol	Description	Source
Final Performance Rating	P_i		OED
Interim Performance Rating	$P_{i,t}$	Overall Project Rating	ARPP
Annual Supervision	$S_{i,t}$	$HS_{i,t} + CS_{i,t}$	TRS
Annual World Bank Staff Supervision	$HS_{i,t}$	By Project, By Year; Taskid=SPN, Version=A, Stafftype=A,H,L,U,Y	TRS
Annual Consultant Supervision	$CS_{i,t}$	By Project, By Year; Taskid=SPN, Version=A, Stafftype= all except A,H,L,U,Y, and DS	TRS
Fraction of the planned implementation period completed	$t/T0_i$	Year of implementation / planned length of implementation in Years	Planned length derived from: fiscal year ("Original Closing Date") - fiscal year ("Board Date") + 1; OED

Variable Name	Symbol	Description	Source
Regional Dummies	RDK_i	Africa, Asia, EMENA, LAC	from OPMIS ProjCode
Sectoral Dummies	SDK_i	Following OED definitions except: all TA projects coded into TA sector; all SECAL's coded into SECAL sector. Sectors grouped if insufficient # of observations	OED
Loan Amount	L_i	Loan Amount in 1990 US $, converted from nominal using USGDP deflator series. Conversion based on the 1st year for projects with planned implementation length of 2 years, the 2nd year for 3-4 year projects, and the 3rd year for longer projects.	OED, BESD

Variable Name	Symbol	Description	Source
Original Total Project Cost	$TC0_i$	Nominal US $	OED
Preparation	PR_i	By project, Version=A, taskID=LENP,LENA,LENN, and ORD	TRS
GDP/Capita	$GP_{i,t}$	Real GDP in 1987 US $ / population	BESD
Growth Rate	$G_{i,t}$	$(GP_{i,t}-GP_{i,t-1})/GP_{i,t-1}$	BESD
Change in Degree of Openness	$O_{i,t}$	$(M_{i,t}+X_{i,t})/GDP_{i,t} - (M_{i,t-1}+X_{i,t-1})/GDP_{i,t-1}$	BESD
# of Staff in Department	$Z1_{i,t}$	# of Higher level staff in Department which supervised Project i in year t	Personnel Database

Variable Name	Symbol	Description	Source
Staff Supervision Resources	$Z2_{i,t}$	Total World Bank Staff supervision of projects managed by the same department as Project i in year t	TRS, Personnel Database
Consultant Supervision Resources	$Z3_{i,t}$	Total Consultant supervision of projects managed by the same department as Project i in year t	TRS, Personnel Database

Variable Name	Symbol	Description	Source
# of Planned Projects	$Z4_{i,t}$	Number of projects which are still in their planned implementation period (fraction ≤ 1) and are managed by the same department as Project i in year t	OED, TRS, Personnel Database
# of Unplanned Projects	$Z5_{i,t}$	Number of projects which are past their planned implementation period (fraction > 1) and are managed by the same department as Project i in year t	OED, TRS, Personnel Database

Tables and Figures

Table 1: Final Performance Rating

Observations = 1426

Variable	Count	Percent Satisfactory
Final Performance Rating		71.2
P_i = 1 (Satisfactory Rating)	1015	
P_i = 0 (Unsatisfactory Rating)	411	
Regions		
Africa	433	61.9
Asia	446	77.4
Europe, Middle East & North Africa	238	81.1
Latin American & Caribbean	309	67.6
Sectors		
Agriculture	467	61.5
Development Finance Corporations	124	73.4
Education	122	78.7
Energy	255	83.9
Health	20	65.0
Industry	68	63.2
Structural Adjustment Loan	33	60.6
Technical Assistance	45	57.8
Transportation and Tourism	213	76.5
Urban	67	83.6
Other	12	50.0
Loan Amount		
Small Loans	974	69.5
Large Loans	452	74.8
Supervision Level		
Low Supervision	860	75.8
High Supervision	566	64.1

Loan Amount in 1990 US$. Division between Large and Small Loans is sample project mean (US$ 58.2 million). Division between High and Low Supervision is sample project mean (81.7 staff weeks). Structural Adjustment includes Sectoral Adjustment. Other includes Disaster Relief and Multisector projects.

Table 2: Annual Interim Performance Rating

Observations = 7461

Variable	Count	Mean Rating
Interim Performance Rating		1.83
$P_{i,t}$ = 1 (Good)	2251	
$P_{i,t}$ = 2	4303	
$P_{i,t}$ = 3	852	
$P_{i,t}$ = 4 (Bad)	55	
Regions		
Africa	2189	1.87
Asia	2510	1.76
Europe, Middle East & North Africa	1175	1.82
Latin American & Caribbean	1587	1.88
Sectors		
Agriculture	2611	1.97
Development Finance Corporations	601	1.82
Education	661	1.71
Energy	1341	1.73
Health	114	1.85
Industry	308	1.81
Structural Adjustment Loan	77	1.64
Technical Assistance	210	1.60
Transportation and Tourism	1105	1.75
Urban	386	1.79
Other	47	1.49
Loan Amount		
Small Loan	5149	1.86
Large Loan	2312	1.75
Supervision Level[A]		
Low Supervision in year t-1	3649	1.79
High Supervision in year t-1	2378	1.99

Loan Amount in 1990 US$. Division between Large and Small Loans is sample project mean (US$ 58.2 million). Division between High and Low Supervision is sample mean (11.8 staff weeks). Structural Adjustment includes Sectoral Adjustment. Other includes Disaster Relief and Multisector projects.

[A] Based on 6027 observations to allow for lag.

Table 3: Annual Change in Interim Performance Rating

Observations = 6027 $\Delta P_{i,t} \equiv P_{i,t-1} - P_{i,t}$

Variable	Count	Mean Change in Rating
Change in Interim Performance Rating		-0.036
$\Delta P_{i,t}$ = -2 (Worse)	53	
$\Delta P_{i,t}$ = -1	786	
$\Delta P_{i,t}$ = 0	4538	
$\Delta P_{i,t}$ = 1	627	
$\Delta P_{i,t}$ = 2 (Better)	23	
Regions		
Africa	1755	-0.044
Asia	2064	-0.031
Europe, Middle East & North Africa	931	-0.054
Latin American & Caribbean	1277	-0.022
Sectors		
Agriculture	2144	-0.051
Development Finance Corporations	476	-0.057
Education	539	-0.009
Energy	1085	-0.025
Health	94	0.064
Industry	238	-0.113
Structural Adjustment Loan	45	0.067
Technical Assistance	165	-0.067
Transportation and Tourism	887	-0.016
Urban	319	-0.022
Other	35	0.000
Loan Amount		
Small Loan	4171	-0.037
Large Loan	1856	-0.034
Supervision Level		
Low Supervision in year t-1	3649	-0.049
High Supervision in year t-1	2378	-0.017

Length of Time Series

Years	Count	Years	Count	Years	Count
1	175	4	251	7	114
2	140	5	266	8	55
3	212	6	209	9-10	12

Loan Amount in 1990 US$. Division between Large and Small Loans is sample project mean (US$ 58.2 million). Division between High and Low Supervision is sample mean (11.8 staff weeks). Structural Adjustment includes Sectoral Adjustment. Other includes Disaster Relief and Multisector projects. There are 1434 individual time series because 8 of the 1426 projects have "broken" time series due to one missing rating observation each.

Table 4: Transition Frequencies

4A: Interim Ratings

$P_{i,t-1}$ \ $P_{i,t}$	1	2	3,4
1	70.5	27.3	2.2
2	10.7	80.7	8.6
3,4	2.8	34.0	63.2

4B: Grouped Interim Ratings

$P_{i,t-1}$ \ $P_{i,t}$	1,2	3,4
1,2	93.5	6.5
3,4	36.8	63.2

4C: Last Interim to Final Rating

P_{i,T_i} \ P_i	1	0
1,2	76.7	23.3
3,4	35.1	64.9

Figures indicate percentage of row state which become column state. Rows may not sum to 100 due to rounding error.

Table 5: Independent Variables

Variable	Mean	Std. Dev.	Min.	Max.	Count	Units
Supervision						
Cumulative: S_i	81.7	56.0	1.4	634	1426	Staff Weeks
Annual: $S_{i,t-1}$	11.8	9.8	0	165	6027	Staff Weeks
Other Ex Ante Characteristics						
Loan Amount	58.2	65.5	1.2	490	1426	'90 US$ $*10^6$ [A]
Preparation	97.8	66.7	0.05	580	1426	Staff Weeks
Time						
$t/T0_i$	0.95	0.36	0.18	4.5	6027	-
Macroeconomic Variables (Annual)						
Growth of GDP per Capita	0.007	0.05	-0.2	0.2	6027	%
GDP per Capita Level	958	966	107	11308	6027	1987 US$
Change in Index of Openness	0.52	0.29	0.03	1.76	6027	-
Institutional Variables (by Department)						
# of Staff	192	46	51	406	6027	Count[B]
Staff Supervision Resources	2683	451	334	3515	6027	Staff Weeks[C]
Consultant Supervision Resources	401	144	87	1053	6027	Staff Weeks[D]
# of Planned Projects	433	89	70	631	6027	Count[E]
# of Unplanned Projects	97	33	9	151	6027	Count[F]

**** Continued on next page ****

Table 5: Independent Variables -- Continued

Notes

[A] Converted from nominal to real terms using a US GDP deflator where the year of the deflator is approximate middle of the project: the World Bank board approval date plus 1 year for 2 year projects, 2 years for 3 to 4 year projects and 3 years for longer projects.

[B-F] Department refers to the World Bank department managing project i (e.g., East Africa, South Asia, etc.). Observations are annual.

[B] # of World Bank staff in department at end of fiscal year. Does not include consultants.

[C] Supervision on all department projects by World Bank staff.

[D] Supervision on all department projects by consultants.

[E] # of projects managed by department which are within their planned implementation period.

[F] # of projects managed by department which have overrun their planned implementation period.

Table 6: Final Performance Estimation
Method = Probit

Observations = 1426
Dependent Variable = P_i Mean = 0.71
Log Likelihood = -760.31

Parameter	Estimate	t-stat	
Cumulative Supervision			
S_i	-0.015	-3.76	**
$(S_i)^2$	0.000014	1.34	
Regions (base = Africa)			
Asia	0.70	2.35	**
Europe, Middle East & North Africa	0.50	1.42	
Latin American & Caribbean	-0.049	-0.18	
Sectors (base = Agriculture)			
Development Finance Corporations	0.59	1.55	
Education	1.61	3.84	**
Energy	1.76	5.36	**
Health	1.18	1.27	
Industry	-0.29	-0.63	
Multisector	-1.16	-1.16	
Structural Adjustment Loan	0.077	0.17	
Technical Assistance	-0.20	-0.37	
Transportation and Tourism	1.29	3.96	**
Urban	2.23	4.20	**
Other Ex Ante Characteristics			
Loan Amount	-0.0017	-1.09	
Preparation	-0.00030	-0.17	
Macroeconomic Variables			
Growth of GDP per Capita	0.21	4.42	**
Change in Index of Openness	0.017	3.38	**
Constant	1.41	4.04	**

* significant at the 90% confidence level
** significant at the 95% confidence level

Includes correction for Heteroskedasticity: each observation was divided by the square root of T_i, the length of the project (σ_υ^2 is assumed to be 1).

Table 7: Supervision Estimation
Method = Ordinary Least Squares

Observations = 6027 — Mean = 10.4
Dependent Variable = $S_{i,t}$ — Adjusted R^2 = 0.25

Parameter	Estimate	t-stat	
Performance Rating			
$P2_{i,t}$	1.84	6.13	**
$P34_{i,t}$	3.28	6.78	**
$P3_{i,t}$	2.31	2.98	**
$P4_{i,t}$	4.39	3.33	**
$P2_{i,t-1}$	1.40	4.68	**
$P34_{i,t-1}$	2.83	5.87	**
$P3_{i,t-1}$	0.29	0.30	
$P4_{i,t-1}$	-2.02	-1.14	
Time			
$(t/T0_i)$	-11.97	-13.41	**
$(t/T0_i)^2$	2.48	6.54	**
Regions (base = Africa)			
Asia	0.91	1.88	*
Europe, Middle East & North Africa	-0.67	-1.15	
Latin American & Caribbean	-0.68	-1.30	
Sectors (base = Agriculture)			
Development Finance Corporations	-1.93	-4.64	**
Education	-0.55	-1.41	
Energy	-0.86	-2.67	**
Health	7.26	8.56	**
Industry	-1.29	-2.34	**
Multisector	2.41	1.55	
Structural Adjustment Loan	8.62	7.14	**
Technical Assistance	4.39	6.53	**
Transportation and Tourism	-0.15	-0.45	
Urban	2.08	4.32	**
Other Ex Ante Characteristics			
Loan Amount	0.021	4.99	**
$(\text{Loan Amount})^2$	-0.000041	-3.30	**
Preparation	0.027	15.54	**
Macroeconomic Variables			
Growth of GDP per Capita	4.28	1.84	*
GDP per Capita Level	-0.00042	-3.10	**
Change in Index of Openness	-1.64	-3.80	**
Institutional Variables (by Department)			
# of Staff	0.0063	2.17	**
Staff Supervision Resources	0.0055	10.39	**
Consultant Supervision Resources	0.0044	4.10	**
# of Planned Projects	-0.019	-6.28	**
# of Unplanned Projects	-0.14	-5.25	**
$(\text{\# of Unplanned Projects})^2$	0.00050	3.73	**
Constant	13.11	11.78	**

* significant at the 90% confidence level
** significant at the 95% confidence level

Table 8: Interim Performance Estimation
Method = Ordered Probit

Observations = 6027 ΔP_t = -2,-1,0,1,2
Dependent Variable = ΔP_t Mean = -0.036
Log Likelihood = -4213.60

Parameter	**Estimate**	**t-stat**	
Annual Supervision			
Early $S_{i,t-1}$	0.034	2.65	**
Late $S_{i,t-1}$	0.0034	1.02	
(Early $S_{i,t-1})^2$	-0.00016	-0.60	
(Late $S_{i,t-1})^2$	0.000030	0.63	
Time (base = Early)			
Late	2.62	1.13	
(Early $t/T0_i$)	14.13	1.06	
(Late $t/T0_i$)	1.19	6.93	**
(Early $t/T0_i)^2$	-18.74	-0.99	
(Late $t/T0_i)^2$	-0.31	-4.60	**
Regions (base = Africa)			
Asia	0.048	0.99	
Europe, Middle East & North Africa	-0.065	-1.07	
Latin American & Caribbean	-0.018	-0.30	
Sectors (base = Agriculture)			
Development Finance Corporations	0.043	0.59	
Education	0.18	2.62	**
Energy	0.064	1.25	
Health	0.28	2.00	**
Industry	-0.10	-1.06	
Multisector	0.22	0.69	
Structural Adjustment Loan	0.36	1.72	*
Technical Assistance	0.031	0.28	
Transportation and Tourism	0.070	1.28	
Urban	0.11	1.34	
Other Ex Ante Characteristics			
Loan Amount	0.0018	4.15	**
Early $S_{i,t-1}$*Loan Amount	-0.00014	-2.55	**
Late $S_{i,t-1}$*Loan Amount	-0.000062	-2.83	**
Preparation	-0.00054	-1.90	*
Macroeconomic Variables			
Growth of GDP per Capita	0.79	2.11	**
GDP per Capita Level	0.000022	0.98	
Change in Index of Openness	0.093	1.35	

* significant at the 90% confidence level
** significant at the 95% confidence level

**** Continued on next page ****

Table 8: Interim Performance Estimation -- Continued

Parameter	Estimate	t-stat	
Heteroskedasticity Coefficients			
Asia	-0.13	-2.18	**
Europe, Middle East & North Africa	-0.13	-1.70	*
Latin American & Caribbean	-0.031	-0.40	
Development Finance Corporations	0.36	2.80	**
Education	0.29	2.50	**
Energy	0.086	1.22	
Health	0.16	0.75	
Industry	0.25	1.50	
Multisector	0.55	0.90	
Structural Adjustment Loan	-0.037	-0.12	
Technical Assistance	-0.071	-0.50	
Transportation and Tourism	0.16	1.85	*
Urban	0.070	0.60	
Cutoff Values			
Δp_{-2}	1.24	0.53	
Δp_{-1}	2.49	1.07	
Δp_{1}	4.71	2.03	**
Δp_{2}	5.93	2.55	**

* significant at the 90% confidence level
** significant at the 95% confidence level

Table 9: Expected Change in Performance and Expectation Derivatives

Variable	Expected Change in Performance	Expectation Derivative
Early Supervision = 0	-0.40	0.012
Early Supervision = 2[A]	-0.37	0.012
Early Supervision = 12[B]	-0.28	0.009
Early Supervision = 22[C]	-0.20	0.007
Early Supervision = 80	-0.0023	0
Late Supervision = 12[D]	-0.008	0.0002
(t/T0) = 3/7[E]	-0.34	
(t/T0) = 4/7[E]	-0.10	
Africa	-0.32	
Asia	-0.28	
Europe, Middle East & North Africa	-0.33	
Latin American & Caribbean	-0.32	
Agriculture	-0.32	
Development Finance Corporations	-0.34	
Education	-0.28	
Health	-0.22	
Structural Adjustment	-0.17	
Transportation and Tourism	-0.30	
Loan Amount = 1[F]	-0.31	
Loan Amount = 124[C]	-0.29	
Preparation = 31[A]	-0.29	
Preparation = 165[C]	-0.32	

All other time varying variables at mean for Early Period ($t/T0_i<.5$); all other time invariant variables at project-level mean. The variables presented are those which are significant at the 90% confidence level in Table 8 plus Latin American and Caribbean. Calculations based on Equations (A1.6), (A1.7), and (A1.8) from Appendix 1 and coefficient estimates from Table 8.

[A] Mean - 1 standard deviation
[B] Mean
[C] Mean + 1 standard deviation
[D] All other time varying variables at mean for Late Period ($t/T0_i \geq .5$).
[E] All other time varying variables at mean for entire data set.
[F] Lowest observation (used if (Mean-standard deviation)<0)

Table 10: Simulated Final Performance Estimation
Method = Probit

Observations = 1426
Dependent Variable = $\hat{P}_i$ Mean = 0.81
Log Likelihood = -614.51

Parameter	Estimate	t-stat	
Cumulative Supervision			
$\hat{S}_i$	-0.033	-4.56	**
$(\hat{S}_i)^2$	0.000061	2.39	**
Regions (base = Africa)			
Asia	1.79	5.26	**
Europe, Middle East & North Africa	1.62	3.77	**
Latin American & Caribbean	0.45	1.47	
Sectors (base = Agriculture)			
Development Finance Corporations	0.015	0.034	
Education	1.76	3.68	**
Energy	1.81	4.42	**
Health	1.07	1.06	
Industry	-1.57	-3.13	**
Multisector	-1.54	-1.52	
Structural Adjustment Loan	-1.25	-2.27	**
Technical Assistance	-0.35	-0.64	
Transportation and Tourism	1.70	4.31	**
Urban	0.35	0.69	
Other Ex Ante Characteristics			
Loan Amount	0.0033	1.62	
Preparation	0.0000046	0.00023	
Macroeconomic Variables			
Growth of GDP per Capita	0.21	3.92	**
Change in Index of Openness	0.0073	1.33	
Constant	3.00	5.85	**

* significant at the 90% confidence level
** significant at the 95% confidence level

Includes correction for Heteroskedasticity: each observation was divided by the square root of T_i, the length of the project.

Figure 1 -- Distribution of Cumulative Supervision

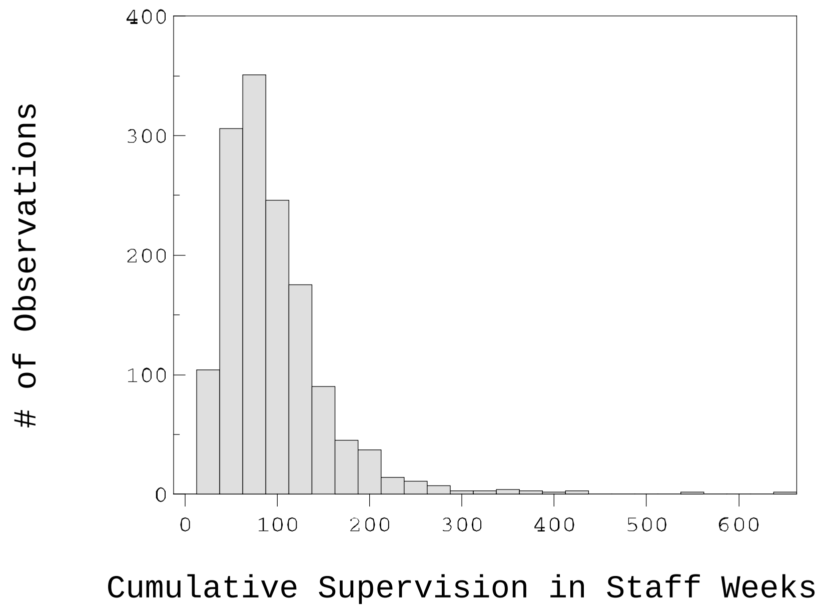


Figure 2 -- Detailed Distribution of Cumulative Supervision

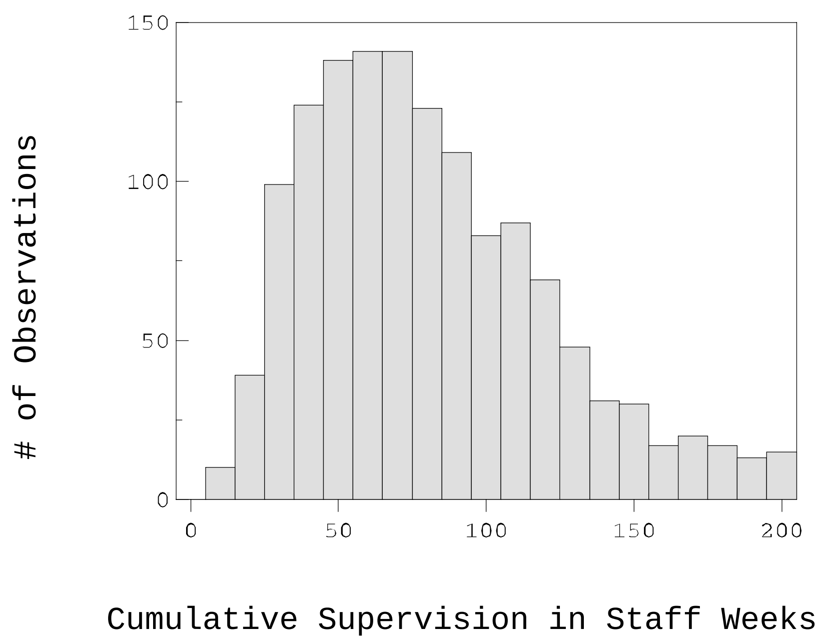


Figure 3 -- Distribution of Annual Supervision

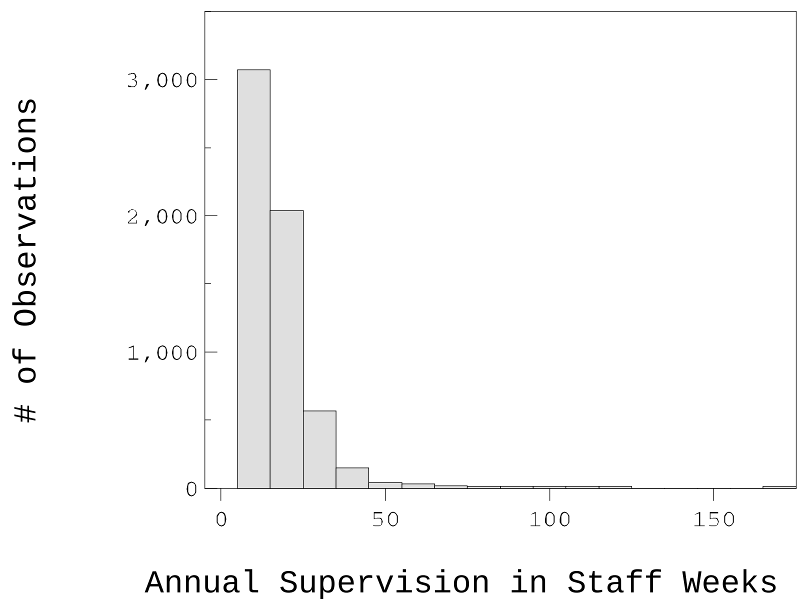


Figure 4 -- Detailed Distribution of Annual Supervision

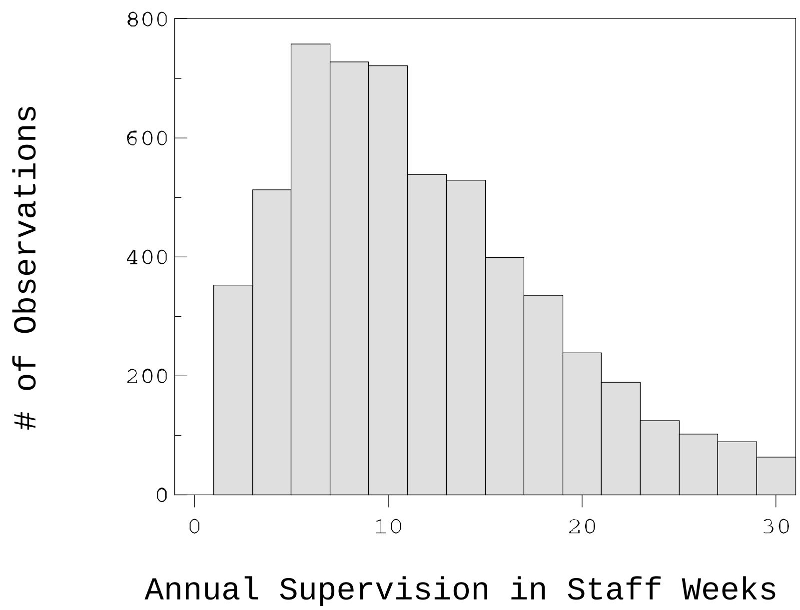
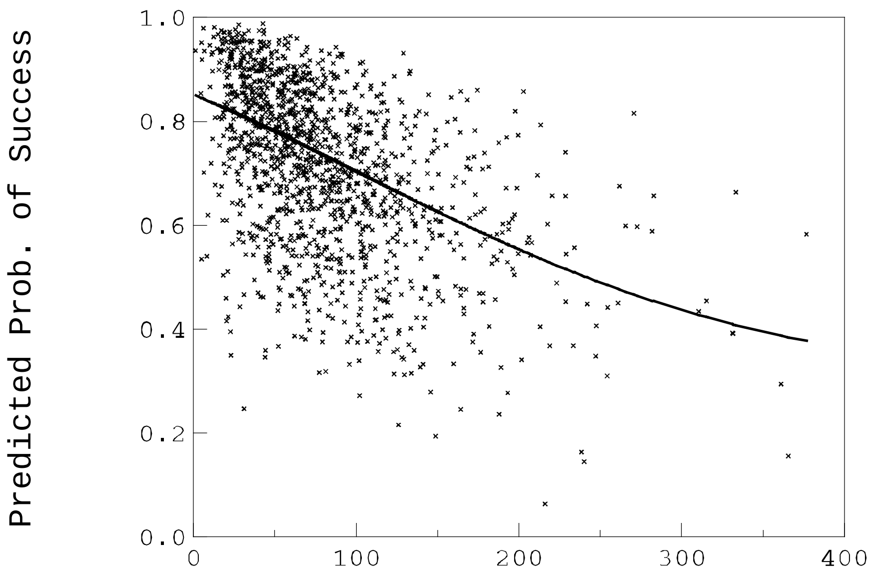


Figure 5 -- Expected Final Performance and Cumulative Supervision (Table 6)

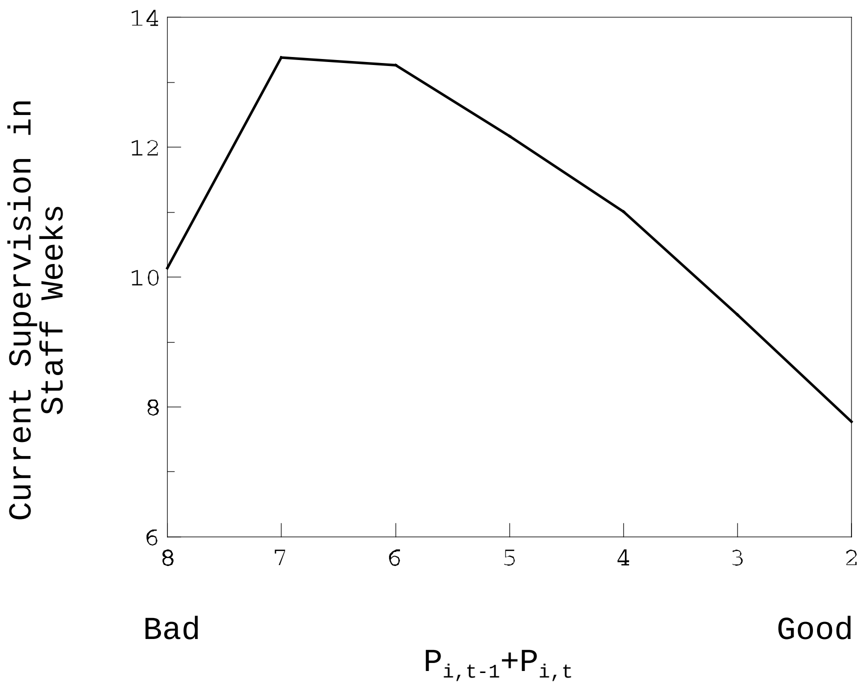


Cumulative Supervision in Staff Weeks

* = Predicted probability of a satisfactory final rating as a function of cumulative supervision with all explanatory variables at actual values.

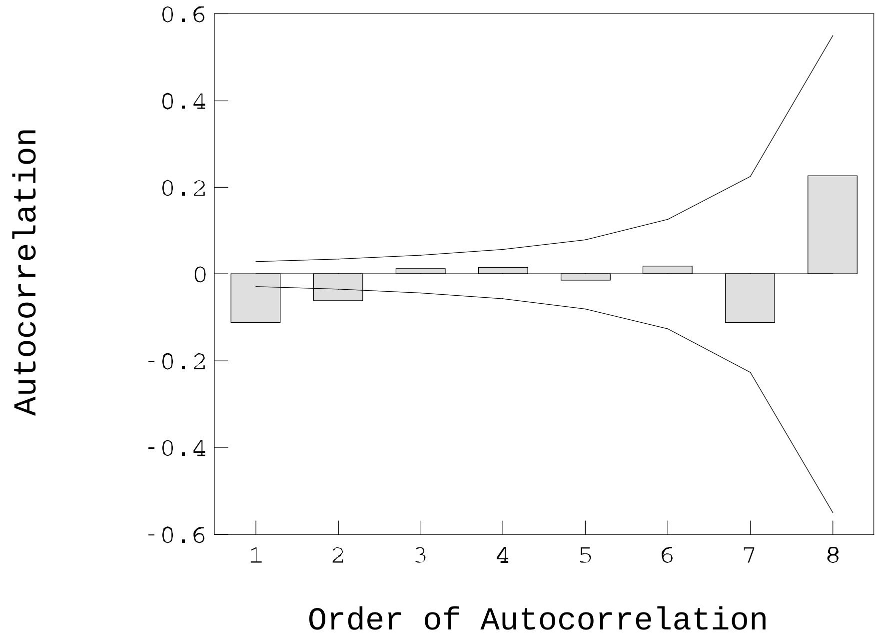
— = Predicted probability of a satisfactory final rating as a function of cumulative supervision for an average project (all explanatory variables except supervision set to sample mean values).

Figure 6 -- Supervision Allocation and Performance
(Table 7)



- = Predicted supervision allocation in staff weeks as a function of the sum of the current and previous interim performance ratings for an average project (all explanatory variables except performance ratings set to sample mean values). Where different rating combinations result in the same sum, the supervision allocation reported is a weighted average of the predicted supervision allocations for the individual rating combinations. Weights are derived from the sample frequencies of the combinations.

Figure 7 -- Autocorrelations For Simulated Residuals

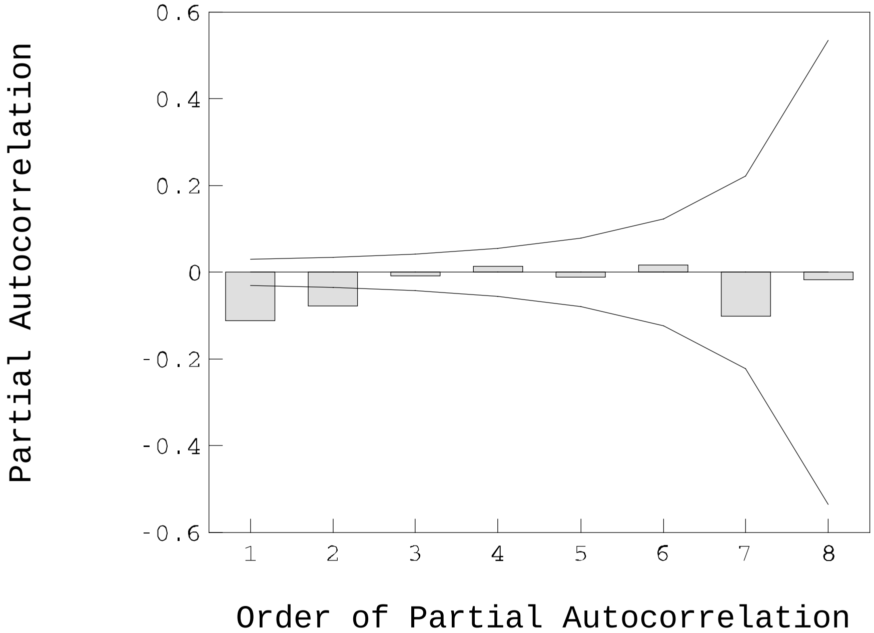


Lines indicate a 95% confidence interval for the hypothesis that $\rho(j) = 0$ for $j \geq i$ where $\rho()$ is the autocorrelation function and i is the x-axis. The approximation used for computing these bounds is

$$\pm 2 * SE(r(i)) \cong \pm 2 * \sqrt{\frac{1}{T}\left(1+2\sum_{m=1}^{i-1}(r(m))^2\right)}$$

where r() is the sample autocorrelation function and T is the number of observations used to compute r(i).

Figure 8 -- Partial Autocorrelations For Simulated Residuals

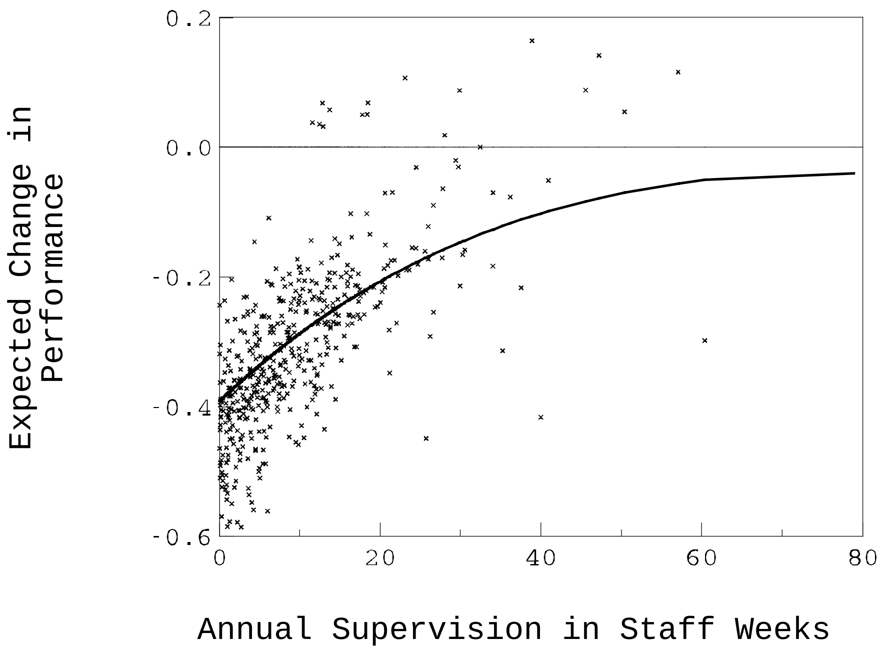


Lines indicate a 95% confidence interval for the hypothesis that $\phi_{jj} = 0$ for $j \geq i$ where ϕ is the partial autocorrelation function and i is the x-axis. The approximation used for computing these bounds is

$$\pm 2 * SE(\hat{\phi}_{ii}) \cong \pm 2 * \frac{1}{\sqrt{T}}$$

where T is the number of observations used to compute $\hat{\phi}_{ii}$.

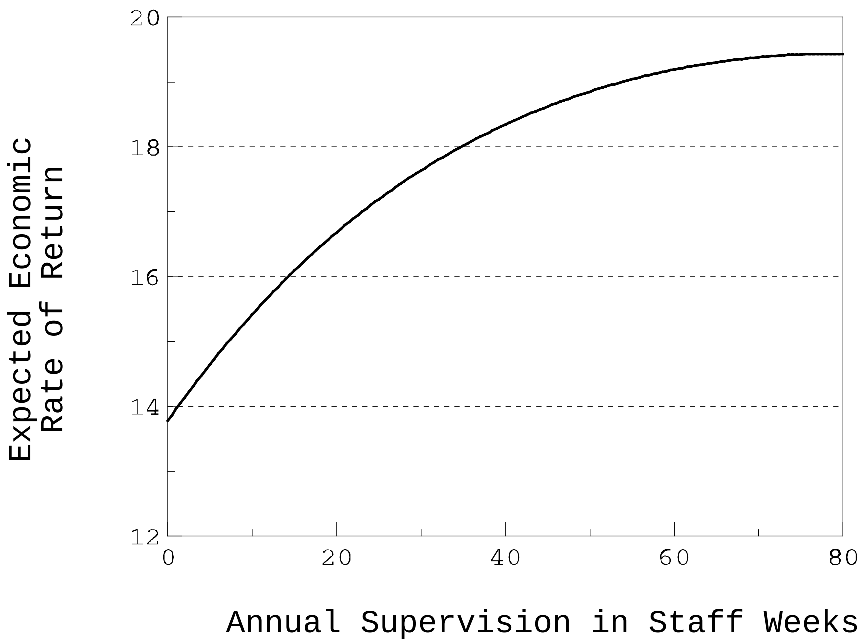
Figure 9 -- Expected Change in Performance and Annual Supervision -- Early Period (Table 8)



* = Expected change in performance on a scale of -2 to 2 as a function of supervision in the previous year in the first half of the planned implementation period with all explanatory variables at <u>actual values</u>.

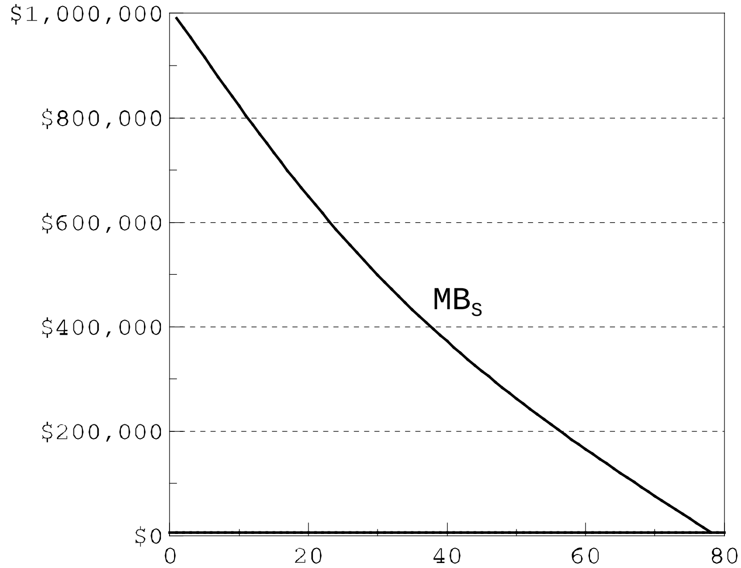
— = Expected change in performance on a scale of -2 to 2 as a function of supervision in the previous year in the first half of the planned implementation period for an <u>average project</u> (time varying explanatory variables set to mean for first half of the planned implementation period; time invariant explanatory variables set to mean for all projects).

Figure 10 -- Expected Economic Rate of Return and Early Supervision



Supervision level maintained for first three years of project implementation. Calculation based on average project (time varying explanatory variables set to sample mean for first half of the planned implementation period; time invariant explanatory variables set to sample mean for all projects).

Figure 11 -- Marginal Benefit of Early Supervision



Annual Supervision in Staff Weeks

Marginal Benefit of early supervision estimated for an average project (time varying explanatory variables set to sample mean for first half of the planned implementation period; time invariant explanatory variables set to sample mean for all projects; cost stream assumed to be {.1, .25, .4, .1, .05, .05, .05} in the first seven years; benefit stream assumed to be constant for subsequent ten years; World Bank discount rate set to 10 percent; supervision is in first year of implementation).