DAUPHINE UNIVERSITÉ PARIS

DIAL
Développement Institutions & Mondialisation
UMR 225 IRD - Paris-Dauphine

IRD
Institut de recherche
pour le développement

DOCUMENT DE TRAVAIL   DT/2015-04

Supervision and Project Performance: A Principal-Agent Approach

Lisa CHAUVET
Paul COLLIER
Andreas FUSTER

SUPERVISION AND PROJECT PERFORMANCE:
A PRINCIPAL-AGENT APPROACH

Lisa Chauvet
IRD, LEDa, DIAL UMR 225, France
PSL, Université Paris-Dauphine, LEDa, DIAL UMR 225
chauvet@dial.prd.fr

Paul Collier
Professor of Economics and Public Policy, Blavatnik School of Government, Oxford
Director, International Growth Centre
paul.collier@bsg.ox.ac.uk

Andreas Fuster
Federal Reserve Bank of New York[1]
andreas.fuster@ny.frb.org

First draft: 2006; last revised: 2013.

Abstract

This paper extends and applies principal-agent theory to the performance of donor projects. There is variation in the degree of divergence between the interests of the donor (the principal) and the recipient government (the agent). Further, the effort expended on observation of the agent is a control variable. We show that in a wide range of circumstances an implication of principal-agent theory is that the principal should put greater effort into observation the wider is the divergence of interest with the agent. We then test this prediction using data on World Bank project performance. We measure the degree of divergence between donor and recipient interests, as perceived by the donor, through a donor classification system of recipient governments. Consistent with the theory, we find that donor supervision of projects is significantly more effective in improving project performance where interests are widely divergent. However, donors do not put more effort into the supervision of projects in such cases.

Key words: Principal-Agent theory, Aid projects, Supervision, Difficult partnerships

Résumé

Cet article étend et applique la théorie Principal-Agent à la performance des projets d'aide. Les intérêts du donneur (le principal) et du gouvernement receveur (l'agent) peuvent différer de manière

---

[1] The views expressed in this paper are solely those of the authors and not necessarily those of the Federal Reserve Bank of New York or the Federal Reserve System.

importante. Dans le modèle, l'effort mis en œuvre pour observer l'agent est une variable de contrôle. Nous montrons qu'une implication du modèle principal-agent est que le principal devrait faire d'autant plus d'effort pour observer l'agent quand ses intérêts divergent de ceux de l'agent. Nous testons ensuite ces prédictions en utilisant les données de performance des projets d'aide de la Banque mondiale. Nous mesurons le degré de divergence entre les intérêts du donneur et du receveur, telle que perçue par le donneur, par la classification des receveurs comme 'partenariats difficiles'. Comme prédit par le modèle, nous trouvons que la supervision des projets d'aide par le donneur permet d'autant plus d'assurer le succès des projets que les intérêts du donneur et du receveur diffèrent. Toutefois, le donneur ne semble pas faire plus d'effort de supervision dans les partenariats difficiles.

Mots Clés :   Théorie Principal-Agent, Projets d'aide, Supervision, Partenariats Difficiles.

JEL Code :   D86 - F35 - O19 - O22

I. INTRODUCTION

In many situations a principal must contract business with a range of agents whose interests are known to diverge to varying degrees from those of the principal. We analyze whether the principal should set her level of effort expended on supervision of the agent purposively so as to compensate for such differences in intrinsic motivation. We consider one particularly clear such situation, namely that in which a donor agency is required to finance development projects globally which are then implemented by recipient governments. The degree to which the interests of donor agencies and recipient governments are congruent varies radically between countries. The ideal situation is recognized to be one in which interests are coincident: the official language now used by donors to describe this is 'partnership'. While coincidence of interests is probably rare, in some situations it is manifestly unrealistic. Donor agencies, working together in the Development Assistance Committee of the OECD, have classified their dealings with a group of recipient governments as 'Difficult Partnerships'. These are, by definition, situations in which the donor (the principal) perceives an unusually wide divergence of interests between itself and the recipient government (the agent).

The principal-agent problem arises from the conjunction of non-congruent interests with the limited observability of agent effort. However, in most situations the degree of observability is not a given but is to an extent under the control of the principal. Since enhanced observation is costly, the principal must decide how much to spend on it for each agent. This is indeed the case in the context of a donor-financed project implemented by a recipient government. Donors supervise projects during implementation, and the degree of effort put into supervision is an important allocative decision for the managers of donor agencies.

This paper investigates whether expenditure on supervision should be related to prior information about the degree of divergence between the interests of the principal and the agent. That is, can more precise observation of agent effort offset divergent interests? In Section II we develop the theory, based on a simple principal-agent model. We introduce a measure for the degree of divergence between the interests of the principal and the agent, and a control variable through which the principal can, at a cost, increase the precision of her observation of the agent's effort. We show that under reasonable assumptions precise supervision is indeed a substitute for congruence of interests as long as the principal chooses optimally among the set of admissible contracts. With optimal choice of contract, incentives are higher-powered the better the precision of supervision and lower-powered the more congruent the interests of the two parties.

In Section III we test the model empirically using data on donor projects. The data cover all World Bank projects evaluated by its Independent Evaluation Group (IEG) over the years 1977 to 2002. The IEG rates completed projects by their degree of success, and also evaluates the degree of preparation and supervision by the World Bank of the projects.

Our main question of interest is whether the impact of supervision on project success is related to the degree of divergence of interests between the donor and the recipient country. The supervision effort put into a project is at the discretion of World Bank Country Directors who control operational budgets: in practice, the vagaries of budgeting and management produce wide variations in supervision effort. Corresponding to the OECD concept of 'Difficult Partnerships', the World Bank has its own classification of those recipient governments with which interests are likely to be most divergent, termed 'Low-Income Countries Under Stress', or 'LICUS' (World Bank, 2002). We thus have information on the performance of projects, the supervision effort put into each project, and the degree of divergence between the interests of the donor and the recipient government as perceived by the donor. This enables us to test both whether supervision is an effective substitute for

congruence of interests, and whether allocative decisions on supervision effort are set consequentially.

We find that while projects are considerably more likely to fail in countries where the government has widely divergent interests, supervision is differentially effective in increasing the probability of success. This is the case whether or not supervision is instrumented for in our econometric estimation. Hence, consistent with the theory, supervision is an effective substitute for congruent interests. However, while it might be expected that managers would allocate more supervision effort into those countries where interests are least congruent, in fact we find that they do the opposite. We consider why the incentive environment facing managers might generate this apparently perverse outcome.

A number of studies have used the IEG data to disentangle the respective contributions of country and project-level characteristics to project success. For instance, Isham et al. (1997) and Isham and Kaufmann (1999) focus on country characteristics such as civil liberties and sound macroeconomic policy and show that both positively affect the economic rate of return of World Bank projects. Recent work by Dreher *et al.* (2013) explores country-level political economy determinants of project success with a focus on the effects of politically-motivated aid (e.g. to countries that hold a non-permanent seat on the UN Security Council or an Executive Directorship at the World Bank). The authors find no evidence of a negative effect of politically-motivated aid on project performance, except when recipient countries are economically vulnerable (higher short-term debt).

Existing work using the IEG data on project preparation and supervision has found somewhat conflicting results. Dollar and Svensson (2000) focus on the success of structural adjustment programmes, and find that preparation and supervision have no impact on the success of reform; instead, success depends mainly on political-economy factors such as political instability, ethnic fragmentation, or democracy. However, their results are not confirmed by Kilby (2000) who finds that the supervision effort is effective in raising the

probability of success of World Bank projects, notably supervision provided at the early stage of projects. Kilby (2012) finds that the preparation of projects by the World Bank also significantly increases the likelihood of a satisfactory outcome rating. The discrepancy between Kilby's results (which are in agreement with our findings) and those of Dollar and Svensson (2000) is likely due to different treatments of the endogeneity of supervision and preparation with respect to project outcome.[i] Finally, Denizer *et al.* (2011) examine the impact on project performance of various previously unexplored projects characteristics. They notably look at the correlation between projects' outcomes and the quality of the task manager. The quality of the task manager is proxied for by the average rating of the other projects he managed. They find evidence that the quality of the task manager matters a lot for project performance, at least as much as the quality of policy.

II. A MODEL OF DIVERGENT INTERESTS AND SUPERVISION

II.A. The Set-Up

This section presents a stylized model whose goal it is to provide a theoretical foundation for the empirical investigation in the remainder of the paper. The model is adapted from Baker, Gibbons and Murphy (1994)[ii], which is in turn based on Baker (1992). It analyses the optimal contract between a risk-neutral principal (the donor, D; "she") and a risk-neutral agent (the recipient, R; "he"). The donor wants to implement a project in the recipient country. The outcome y of this project can be either 0 (failure) or 1 (success). The probability that it is a success is determined by the recipient's effort a: $\Pr(y = 1 \| a) = a$. One of the basic assumptions of the model is that, due to its complexity, y is not objectively measurable and therefore not contractible. It is also impossible to write contracts based on a.

There does, however, exist a verifiable performance measure p (for instance based on a report by a project supervisor appointed by D) on which contracts can be based; this performance measure takes the values 0 or 1, and $\Pr(p = 1|a) = \mu a$. μ is a random variable which is only observed by the recipient (before he chooses his effort level); its expected value equals 1 and its variance ε^{-1} ($\varepsilon = \dfrac{1}{\text{var}(\mu)}$ is the *precision* of the performance measure; it can be seen as measuring the *quality of supervision*). Thus, the performance measure is on average unbiased, but it is nevertheless distortive because it varies, and so does the recipient's effort as a consequence, despite the fact that the link between effort and contribution to project outcome is always the same. The fact that μ varies around 1 can be interpreted in this context as saying that there are projects where high effort increases both y and p (μ around one), projects where high effort increases y but not p (μ small), and projects where high effort increases p but not y (μ large). The fact that the recipient observes μ before choosing an effort level reflects the assumption that he observes the way in which supervision of the project will take place and therefore knows whether or not a high effort will be necessary to 'please' the supervisor (i.e. obtain $p= 1$).

We now assume that every project has a fixed overall cost of $c<1$, and that it is financed by the donor in two tranches, of sizes $(1-b)c$ and bc, respectively (where b is between 0 and 1). The payment of the second and final tranche of money, bc, can be made conditional on $p=1$. An implication is that p is the result of some interim evaluation, before the end of the project. As the model assumes that p and y are determined simultaneously, this means that when the evaluation takes place, it is already determined whether the project is a success or not, even though perhaps it cannot yet be observed by D.

The donor can choose b, the proportion of the money that she wants to be conditional on a positive evaluation, and ε, the precision of supervision (which comes at a cost).[iii] The timing of the game is as follows:

1. *D* proposes a project contract $(b; \varepsilon)$;

2. *R* accepts or rejects; if he accepts, he receives the first tranche of *(1-b)c* from *D*; if he rejects, receives his reservation payoff of *0*;

3. If he accepts the contract, *R* observes μ and chooses *a* at personal cost *c(a)* (both μ and *a* are unobservable to the donor);

4. *y* and *p* are realized; if *p=1*, *D* pays *R* the second tranche, *bc*.

Concerning the components of the recipient's utility function, we assume that his cost of effort is *c(a)* $= \gamma a^2$. Furthermore, we assume that he receives non-monetary utility of θy, with $\theta \in [0,1]$. This reflects the utility he derives from a successful outcome of the project, where θ measures the degree to which the interests of *D* and *R* are divergent or congruent.[iv] If θ=0, the recipient does not care at all about the outcome of the project; he only cares about the money he receives (and the effort he needs to exert to obtain it).[v] If θ=1, he cares about the outcome to the same degree as the donor. It is assumed that θ is common knowledge and exogenously given.

The cost of precision to the donor is denoted by $C(\varepsilon,\theta)$, with $C_\varepsilon > 0$, $C_{\varepsilon\varepsilon} > 0$ and $C_{\varepsilon\theta} \leq 0$. Thus, the marginal cost of precision is (at least weakly) higher in an environment with divergent interests.

II.B. Results

D's objective function is given by:

$$\pi(b,\varepsilon;\theta) = E_\mu \left[a \cdot 1 - (1-b)c - \mu a \cdot bc \right] - C(\varepsilon;\theta) \equiv V(b,\varepsilon;\theta) - C(\varepsilon;\theta),$$

where the last part within the square brackets reflects the idea that the second tranche (of size *bc*) is only paid if *p=1*, which happens with probability μa^*. $\pi(b,\varepsilon;\theta)$ is what the

donor maximizes with respect to b and ε, taking into account R's expected reaction to the contract.

The recipient's utility function (once he has observed μ)[vi] is given by:

$$U = a \cdot \theta + (1-b)c + \mu a \cdot bc - \gamma a^2.$$

All of the following propositions obtained from solving the model are proved in Appendix 1a). Starred letters denote optimal values.

PROPOSITION 1. The optimal size of the second tranche, b^*c, decreases in θ and increases in ε (we consider ε as given for the moment). (b^*c increases in γ.)

The trade-off is evident: a higher b provides incentives for R to exert effort (which is good for D), but it also has a cost, as the available performance measure is distortive. This cost is however lower the higher is precision ε, such that a higher b is optimal. Furthermore, when interests are more congruent (θ high), there is less need for monetary incentives, as θ and b both lead to higher (expected) effort of R.

For what follows, we denote by $V^*(\varepsilon;\theta) \equiv V(b^*,\varepsilon;\theta)$ the value of the contract to D for a given level of precision ε (exclusive of the cost of precision, $C(\varepsilon,\theta)$), and by $\pi(\varepsilon;\theta) = V^*(\varepsilon;\theta) - C(\varepsilon;\theta)$ the overall objective of the donor.

PROPOSITION 2. $V^*(\varepsilon;\theta)$ as well as the expected probability of project success increase in ε and θ. The second cross-partial derivatives $\dfrac{\partial^2 V^*}{\partial\varepsilon\partial\theta}$ and $\dfrac{\partial^2 \Pr(success)}{\partial\varepsilon\partial\theta}$ are negative.

The first part of the proposition is unsurprising: both precision of supervision and congruence of interests have a positive impact on the probability of project success and on the value of the project to the principal if the optimal b is chosen.

The second part of the proposition displays the central result of this section. The negative cross-partials signify that the (marginal) positive effect of increased precision is stronger the more divergent the interests of D and R; ε and θ are substitutes. Hence, we expect that increased precision has a stronger impact on the likelihood of project success in 'difficult partnerships'. Importantly, this prediction arises independently of whether the level of precision ε is chosen optimally; the proposition simply requires that for given precision, the size of the second tranche is set optimally (as described in Proposition 1).

The above proposition does not necessarily imply that the donor should choose higher precision when interests are not congruent. As the next proposition shows, whether this is the case depends on the shape of the cost-of-precision function:

PROPOSITION 3. The optimal choice of precision, $\arg\max_{\varepsilon} \pi(\varepsilon;\theta)$, is (weakly) decreasing in

θ as long as $\dfrac{\partial^2 C}{\partial\varepsilon\partial\theta} \geq \dfrac{\partial^2 V^*}{\partial\varepsilon\partial\theta}$, and increasing in θ otherwise.

Thus, as long as the cross-partial of the cost-of-precision function is not 'too negative' (a negative cross-partial means that a marginal increase in precision is more expensive the more divergent the parties' interests), we would expect that the donor chooses a higher precision level in circumstances where interests are divergent than in circumstances where they are congruent. Only if precision were much more costly in 'difficult partnerships' might it be optimal for D to choose a lower ε despite its differential effectiveness.

11

III. An Application to the Performance of Donor Projects

III.A. Context and Data

The World Bank is required by its mandate to undertake development projects in a large majority of the world's low-income countries. Only in the most extreme environments such as Somalia does it actually suspend its project operations. The World Bank does not itself implement projects. Its normal mode of operation can be decomposed into a series of discrete phases. In the first phase, *preparation*, agreement is reached with a recipient government on the content and design of the project. Once approved by the Bank's Board, the project then enters the *implementation* phase. During implementation, which is undertaken by the government, the project is financed in tranches released by the World Bank. Each tranche requires the authorization of World Bank management and so provides a review point. To inform the management review the World Bank undertakes supervision, a report being prepared ahead of each tranche release. If the project is judged to be seriously off-track it can be aborted or scaled down on the decision of World Bank management. The frequency and scale of the tranches is determined during the preparation phase, but since the finance is heavily consessional the Bank has effective power of decision. The wider is the perceived divergence of interests with the recipient government the more is the project liable to be 'back-loaded', with less money released in the first unconditional tranche. This accords both with the theory of Section II, in which b is varied according to θ, and also with the natural risk aversion common to public bureaucracies. The extent of effort put into supervision at each stage is a choice of World Bank management: Country Directors are assigned overall administrative budgets and can choose how to allocate these across a wide range of functions. Once completed or aborted, the project is evaluated. The evaluation is completely independent of the department which is responsible for the project. The IEG

reports directly to the Board of the World Bank and its staff are not permitted to move to positions in other parts of the Bank. IEG evaluates the performance of the project in discrete categories: highly successful, successful, partially successful, partially unsuccessful, unsuccessful, highly unsuccessful. In practice, the key distinction is between the three former categories and the three latter, so that in our empirical analysis we will treat the success variable as binary (data and variables are presented in Appendix 2). IEG also separately evaluates the supervision effort by the World Bank and the preparation effort by the recipient government (highly satisfactory, satisfactory, unsatisfactory, highly unsatisfactory).

Our data from IEG covers more than 2,000 projects in 102 countries and is comprehensive for the time period 1977-2002. Of these, almost 400 projects were in 'LICUS' countries, that is in countries which the Bank itself regarded as having governments whose interests were particularly divergent from those of the World Bank. The LICUS indicator is derived from the "Country Policy and Institutional Assessment" (CPIA), an internal rating of sixteen different aspects of policy by the World Bank. A country is assigned LICUS status if the CPIA averaged over the duration of the project is less than 3 and if the country was classified as low income at least one year during the project. The aim of the LICUS definition is to distinguish countries that are failing to provide an adequate environment for economic development opportunities. The CPIA intends to assess *governments' choices* in terms of policy and reforms rather than the mere economic performance of developing countries. As such the LICUS dummy is likely to be a close approximation to the concept of 'Difficult Partnership' and hence to the set of countries whose governments at the time of the project were known to have interests that were particularly divergent from those of the World Bank.[vii]

The LICUS classification used by the World Bank only classifies countries since 2002. However, that classification is based on a combination of income and CPIA data and so it is

possible to generate an imputed classification of LICUS-type status for all previous years. We use these criteria to generate a project-specific dummy variable that reflects whether, at the time during which the project was in its implementation phase, the country met the criteria for being a LICUS.[viii]

Table 1 shows that in LICUS countries, projects tend to be relatively less successful than in non-LICUS countries. While only 29% of the projects are failing in non-LICUS countries, almost 43% are failing in LICUS countries. Supervision and preparation are slightly worse in LICUS than in non-LICUS, something we will discuss in more detail in section III.D. below. Finally, Table 1 confirms that LICUS countries are poorer and have a lower CPIA than non-LICUS. On the other hand, the average expected duration at the beginning of the project and the time spent by the leader in office are not significantly different in the LICUS and non-LICUS subsamples.

[Table 1 here]

III.B. Empirical Specification and Results

Given the binary nature of the dependent variable, our main estimation equation is a probit model of the following form:

$$\Pr(Success_{j,i,T} = 1) = \alpha' X_{i,t_0} + \beta' Y_{j,i,t_0} + \gamma Preparation_{j,i,T} + \delta Supervision_{j,i,T}$$
$$+ \eta\ LICUS_{i,t} + \lambda(Preparation_{j,i,T} \times LICUS_{i,t}) + \mu(Supervision_{j,i,T} \times LICUS_{i,t}) + \varepsilon_{j,i,t}$$

where j denotes projects, i country and T the evaluation date, t_0 the date at which the project started and t the time during which the project was implemented. X_{i,t_0} is a set of variables controlling for the initial conditions in country i (income per capita, time the leader was in office, and the CPIA), while Y_{j,i,t_0} captures a set of project level characteristics (was

the project financed through IDA, was it an investment project, and the expected duration of the project at the time it started).

Our interest is in the interaction of the dummy variable for countries regarded as having particularly divergent interests, $LICUS_{i,t}$, with the measures of supervision and preparation.

In the first two columns of Table 2 we report the results from probit regressions with project success as the dependent variable. As found previously by Isham and Kaufmann (1999) and Denizer *et al.* (2011), the better is policy - measured by the CPIA - the higher is the probability that the project will be a success. The level of income also affects the probability of success (although less significantly so), with projects working better in higher income countries. Both preparation and supervision as assessed by IEG also significantly improve the probability of success, consistently with the results of Kilby (2000, 2012).

We introduce the LICUS dummy variable for divergent interests directly into the probit and also interact it with both preparation and with supervision. The direct effect is insignificant. This does not imply that divergent interests do not affect performance: since both the level of income and the CPIA score are included in the regression, the dummy variable which is derived from them adds no information.[ix] We now consider the interaction between divergent interests and donor choices. First, we consider the interaction between divergent interests and project preparation. The direct effect of project preparation is significant and positive: a better prepared project is more likely to be a success. This is unsurprising: the function of project preparation is to improve the technical design of the project and thereby raise its rate of return. Now consider how this might interact with divergent interests. As the expected rate of return on the project is increased, this should tend to widen the performance gap between governments with coincident and divergent interests. The government with interests coincident with those of the donor has an enhanced incentive to implement the project well if the expected rate of return is higher. In contrast, a government with widely divergent interests has no intrinsic interest in the success of the

project and so its performance will be unaffected by an increase in the rate of return. Hence, we would expect that the interaction between project preparation and the dummy for divergent interests would be negative.[x]

Now consider the interaction between divergent interests and project supervision. The prediction of the theory developed in Section II is that this interaction should be positive. Indeed, were the dummy variable fully to capture the cases in which interests were divergent, the direct effect of project supervision would become insignificant since with coincident interests the donor does not need the information that supervision provides.

[Table 2 around here]

These predictions are supported by our results. The interaction of divergent interests and project preparation is negative, although insignificant, and that with project supervision is positive and significant. Thus, supervision is differentially effective in situations where interests are divergent, and conversely, where interests are divergent it is supervision, not preparation, that is differentially important.[xi]

The magnitude of the probit coefficients in Table 2 is not directly interpretable (e.g. Ai and Norton, 2003), so we now consider the economic significance of the detected differential effect of supervision in LICUS countries. On average, the predicted probability of project success is 0.69. However, when all variables are taken at mean values of LICUS countries, the average probability drops down to 0.58. This is the case notably because LICUS countries have lower GDP per capita, lower CPIA, a higher expected duration of projects. When all variables are taken at the mean values for non-LICUS countries, the average predicted probability is equal to 0.77. Now, we vary the level of supervision. For LICUS countries, the predicted probability of success with a supervision level of 2 is 0.22, while with a supervision level of 3 it is 0.72 (holding all other variables constant at their mean for these countries). For non-LICUS countries, the respective predicted probabilities are 0.46 and 0.82. Increasing the level of supervision from 2 to 3 thus has a 14 percentage point

(=(0.72-0.22)-(0.82-0.46)) larger effect on success probability for LICUS countries than for non-LICUS countries, which is highly economically significant.

In Table 2, column 2 we show that the results from this regression are robust to the addition of controls for the sector in which the project is undertaken and of interactions between the sector and the LICUS dummy, none of which are significant.[xii] In column 3 we run the same estimation using simple OLS, which leads to similar results as in columns 1 and 2.

Finally, we look at an alternative measure for the divergence of interests between the donor and the recipient. The LICUS dummy we used so far is meant to capture the difficulty of donors' intervention when the economic and institutional environment of the receiving country is weak. It is based on the CPIA computed by the World Bank which assesses governments choices in terms of macro, structural, social policies. As an alternative, we use the Public Investment Management Index (PIMI) which was computed for the 2007-2010 period by Dabla-Norris et al. (2011). The PIMI index aims at measuring the efficiency of the public investment process and management. Among its components the one that is pertinent for our purposes is APPRAISAL. This is defined as the strategic guidance and project appraisal needed to ensure that "investments are chosen based on development policy priorities [and that projects] undergo further scrutiny of their financial and economic feasibility and sustainability to avoid wasteful "White elephant" projects" (Dabla-Norris et al., 2011). In the raw data, APPRAISAL is rated from 0 to 4. We transform it into a dummy variable which is equal to one for values below the median of the sample. Thus when the dummy is equal to one, the quality of broad strategic guidance and project appraisal is low. Unlike the LICUS dummy, APPRAISAL does not vary through time. In column 4 of Table 2, we replace the LICUS dummy by the APPRAISAL dummy. The results are very similar to those of column 2, both in terms of significance and magnitude of the effect.[xiii] Supervision is again a substitute for convergent interests.

III.C. Instrumental Variable Specifications

Next, we try to account for the likely endogeneity of *Preparation* and *Supervision*. We use as instruments some supply-side determinants of the amounts of aid received. We follow Tavares (2003) and use as instruments the total amount of aid of the five biggest donors (US, Japan, UK, France and Germany) weighted by cultural and geographical proximity to the recipient (bilateral distance and dummy variables for same main religion and same official language). This gives us three exogenous instruments. The idea of this kind of instruments is to use the economic conditions in the donor countries as a source of exogenous variation for supervision and preparation, which reflect donors' efforts and involvement in the success of aid project.

We also need to instrument *Preparation* x *LICUS* and *Supervision* x *LICUS*. To do so, we start by running the first-step estimations for *Supervision* and *Preparation* and then introduce the predicted values in levels and in interaction with the dummy *LICUS*. To account for predicted regressor issues, we bootstrap the standard errors.

The IV results are shown in Table 3. In columns 1 and 2, we present the first-step estimations from which we predict *Supervision* and *Preparation* in order to interact them with the dummy *LICUS*.[xiv] Among the three instruments, only two are individually significantly correlated with both *Supervision* and *Preparation*.[xv] The three instruments are however jointly significant in both regressions 1 and 2 and the F-statistics for *Supervision* and *Preparation* are respectively 9.0 and 4.35. Based on the usual rule-of-thumb that the first-stage F-statistic should be at least 10, we are thus somewhat confident in our instrumentation for *Supervision* but less so for *Preparation*.

[Table 3 around here]

In Columns 3 and 4, *Preparation* and its interaction with *LICUS* are not significant, either individually or jointly. *Supervision* is positive and significant (at $p<0.1$) in column 3 only. The interaction term *Supervision* x *LICUS* is positive and not significant in both columns 3 and 4, but is jointly significant with *Supervision* (this is shown in the last rows of Table 3). In Columns 5 and 6, we drop the interaction term of *Preparation* with the dummy *LICUS* which was never significant in any of the previous estimations of Tables 2 and 3. In both columns, our main coefficient of interest, *Supervision* x *LICUS,* is statistically significant, with $p<0.05$ and $p<0.1$, respectively. It bears noting that the magnitude of the coefficient is much larger than in the earlier (uninstrumented) regression, suggesting that the excludability condition may not be respected and that the IV results should be interpreted cautiously.

Preparation and *Supervision* are ordinal variables, taking the values one to four. Instrumented as they are in Table 3, they are assumed to be continuous, and a possible bias can derive from that. Table 4 explores an alternative way of treating the endogeneity of *Preparation* and *Supervision* through a recursive multivariate probit model.[xvi] *Preparation* and *Supervision* are transformed into binary variables (see Appendix 2). When the four equations are simultaneously estimated, *Supervision* and *Preparation* are significantly positive. *Supervision* x *LICUS* is also positive and significant at 5%.[xvii] The last row of Table 4 suggests that the error terms of the four simultaneously estimated equations are correlated.

[Table 4 around here]

Overall, we conclude that while the results from our IV analysis should be interpreted with some caution, the results are qualitatively consistent with our main finding from section III.B., namely that supervision is more effective in LICUS countries where the interests of donors and recipients tend to diverge.

III.D. Is Supervision Higher in LICUS Countries?

Since supervision seems to be differentially effective in the context of divergent interests, it might be expected that the management of the World Bank would allocate resources to supervision accordingly. In fact, this is not the case. On the contrary, where interests are most divergent supervision is significantly lower. Table 5 shows some descriptive statistics of the quality of supervision in the context of divergent interests, that is in LICUS countries, and in the context of congruent interests. Supervision is less likely to be satisfactory in countries where donor-government interests diverge. The proportion of projects with unsatisfactory and highly unsatisfactory supervision is significantly higher in the context of divergent interests (p-value=0.002). This is confirmed by ordered probit estimations of supervision as a function of the LICUS dummy and some project characteristics (Table 6). The dummy for divergent interests is significantly negative, suggesting that in difficult partnerships the quality of supervision is lower.

[Table 5 and Table 6 around here]

There are various possible explanations for such a pattern of behaviour. It might be the case that supervision is more costly in conditions of divergent interests. As discussed in Section II, this would produce an offsetting effect such that the efficient response of supervision would be *a priori* ambiguous. However, as noted by Kilby (2000), the pay-off to enhanced project success is so large relative to the costs of supervision that this explanation seems implausible. Alternatively, the management of the World Bank might themselves not face sufficiently strong incentives to achieve project success. Staff performance is assessed annually by indicators of performance in that year, most notably the disbursement on loans. The World Bank makes an overall commitment to disburse the IDA money that is provided to it every three years by its OECD members, and this in turn generates annual disbursement targets for the management team in each region. In contrast, because of the long lags

20

between the decision to propose a project to the Board and the eventual performance of the project, incentives actually to abort projects are weak. This encourages a 'culture of disbursement' rather than an emphasis upon project success. Hence, it may be that while some individual managers correctly match a high value of b with a high level of supervision, overall managers are more concerned to avoid evidence that would indicate that disbursements on a project should be suspended.

IV. CONCLUSION

In many situations the precision with which the principal can observe the behaviour of an agent can be increased at a cost. Further, the principal often has prior information concerning the degree of divergence of her interests from those of the agent. In this paper we have investigated whether the wider the divergence of interests the more should be spent to increase the precision of observation. In our model we have shown that conditional upon the principal being free to structure the contract so as to discriminate according to prior information about the degree of divergence of interests, greater effort to monitor the agent is indeed warranted where interests are more divergent. We then apply the model to the situation of donor projects that are implemented by a recipient government. Using World Bank data on project performance and donor assessments of the degree of congruence of donor-government interests, we have found that, consistent with the theory, supervision is differentially effective in improving the performance of projects where interests are least congruent. However, while this would lead us to expect that the donor would accordingly expend greater effort in supervision in situations of widely divergent interests, in fact donors do the opposite. This suggests that either there are offsetting costs of supervision, or the incentives facing donor management to allocate administrative budgets are not well-aligned with the objective of project success.

REFERENCES

Ai, Chunrong and Edward C. Norton (2003), "Interaction Terms in Logit and Probit Models." *Economics Letters* 80: 123-129.

Athey, Susan, Paul Milgrom, and John Roberts (1998), "Robust Comparative Statics", Draft Monograph.

Baker, George P. (1992), "Incentive Contracts and Performance Measurement", *Journal of Political Economy*, 100(3): 598-614.

Baker, George, Robert Gibbons, and Kevin J. Murphy (1994), "Subjective Performance Measures in Optimal Incentive Contracts", *Quarterly Journal of Economics*, 109(4): 1125-1156.

Barrett, David B. (ed.) (1982), *World Christian Encyclopedia,* Oxford University Press, Oxford.

Besley, Timothy, and Maitreesh Ghatak (2005), "Competition and Incentives with Motivated Agents", *American Economic Review*, 95(3): 616-636.

Bienen, Henry, and Nicolas Van de Walle (1991), *Of Time and Power: Leadership Duration in the Modern World,* Stanford University Press, Stanford.

CIA (2003), *The World Factbook*, 2003, Central Intelligence Agency, Washington.

Collier, Paul, Anke Hoeffler, and Catherine Pattillo (200) "Africa's Exodus: Brain Drain and Capital Flight", *Journal of African Economies,* 13: ii15 - ii54.

Dabla-Norris, Era, Jim Brumby, Annette Kyobe, Zac Mills, and Chris Papageorgiou (2011), *Investing in Public Investment: An Index of Public Investment Efficiency*, IMF Working Paper 11/37, Washington.

Denizer, Cevdet, Daniel Kaufmann and Aart Kraay (2011), *"Good Countries or Good Projects? Macro and Micro Correlates of World Bank Project Performance"*, World Bank Policy Research Working Paper 5646, Washington DC.

Dollar, David and Jakob Svensson (2000), "What Explains the Success or Failure of Structural Adjustment Programmes", *The Economic Journal* 110 (October): 894-917.

Dreher, Axel, Stephan Klasen, James Raymond Vreeland, Eric Werker (2013),The Costs of Favoritism: Is Politically Driven Aid Less Effective? *Economic Development and Cultural Change* (forthcoming).

Gurr, Ted Robert, Barbara Harff, and Monty G. Marshall (2003), *State Failure Task Force Report, Phase 3: Findings*, Center for International Development and Conflict Management, University of Maryland.

Isham, Jonathan, Kaufmann Daniel and Pritchett Lant H. (1997), "Civil Liberties, Democracy, and the Performance of Government Projects", *The World Bank Economic Review* 11(2): 219-242.

Isham, Jonathan, and Daniel Kaufmann (1999), "The Forgotten Rationale for Policy Reform: The Productivity of Investment Projects", *Quarterly Journal of Economics* 114 (1): 149-184.

Kilby, Christopher (2000), "Supervision and Performance: the Case of World Bank Projects", *Journal of Development Economics*, 62: 233-259.

Kilby, Christopher (2012), *"Assessing the Contribution of Donor agencies to aid effectiveness: The Impact of World Bank Preparation on Project Outcomes"*, Villanova School of Business Economics WP #20, Villanova University, USA.

Tavares, José (2003), "Does Foreign Aid Corrupt?", *Economics Letters,* 79(1): 99-106.

WDI (2004), *World Development Indicators*, Cd Rom.

World Bank (2002), *Low-Income Countries Under Stress*, Report of a Task Force, World Bank, Washington DC.

TABLE 1. Summary statistics (sample = 2023 projects, in 102 countries).

	Non-divergent interests LICUS = 0		Divergent interests LICUS = 1		All	
Projects characteristics						
Successful projects	1,163	(71%)	224	(57.5%)	1,387	(69%)
Failed projects	470	(29%)	166	(42.5%)	636	(31%)
Satisfactory supervision	1,251	(77%)	275	(70.5%)	1,526	(75%)
Poor supervision	382	(23%)	115	(29.5%)	497	(25%)
Satisfactory preparation	1,291	(79%)	271	(69.5%)	1,562	(77%)
Poor preparation	342	(21%)	119	(30.5%)	461	(23%)
Expected duration of project	5.1 years		5.4 years		5.1 years	
Countries characteristics						
Average income p.c.	1,478 $		319 $		1255 $	
Average duration of the leader	8.8 years		8.7 years		8.8 years	
Average CPIA	3.4		2.6		3.2	

Note: Income per capita, the duration of the leader in office, and the CPIA are measured, for each country, at the beginning of the project and then averaged over each sample and sub-sample (LICUS and non-LICUS). Supervision and preparation are scaled 0-1 for clarity (1-4 scale in Table 2).

TABLE 2. Probability of success of projects, benchmark estimations.

Dependent variable: Success = 1	Probit	Probit	OLS	Probit
Measure of divergence of interests:	LICUS	LICUS	LICUS	APPRAISAL
	(1)	(2)	(3)	(4)
Preparation	1.008***	1.013***	0.252***	1.071***
	(0.0990)	(0.105)	(0.0248)	(0.134)
Preparation x 'DIVERGENCE'	-0.137	-0.107	-0.0218	-0.0522
	(0.211)	(0.220)	(0.0520)	(0.181)
Supervision	1.032***	1.019***	0.260***	1.039***
	(0.0834)	(0.0838)	(0.0246)	(0.108)
Supervision x 'DIVERGENCE'	0.325**	0.349**	0.111**	0.335**
	(0.152)	(0.159)	(0.0422)	(0.162)
Dummy 'DIVERGENCE' = 1	-0.568	-0.606	-0.246*	-1.248
	(0.671)	(0.759)	(0.145)	(0.901)
Ln GDP pc, initial	0.119*	0.0958	0.0167	0.0652
	(0.0627)	(0.0594)	(0.0166)	(0.0945)
Duration leader in office, initial	-0.0127**	-0.0143***	-0.00336**	-0.0133**
	(0.00531)	(0.00552)	(0.00159)	(0.00676)
CPIA, initial	0.298***	0.302***	0.0685***	0.257***
	(0.0663)	(0.0674)	(0.0156)	(0.0729)
Dummy IDA = 1	0.240*	0.206*	0.0451	-0.00269
	(0.123)	(0.117)	(0.0369)	(0.176)
Dummy Investment proj. = 1	0.403***	0.366**	0.0828**	0.219
	(0.142)	(0.154)	(0.0384)	(0.190)
Duration of project	-0.0468**	-0.0489*	-0.0117*	-0.0924***
	(0.0234)	(0.0250)	(0.00603)	(0.0286)
Constant	-6.956***	-6.775***	-1.100***	-6.111***
	(0.568)	(0.558)	(0.145)	(0.905)
Observations	2023	2023	2023	1187
Log-likelihood	-873.9	-857.1		-509.1
R2			0.337	
Countries	102	102	102	52
Sector dummies	N	Y	Y	Y
Sector x DIVERGENCE dummies	N	Y	Y	Y
Likelihood-ratio test of rho=0[1]				
Joint significance [2]				
Supervision & Sup. x DIVERGENCE	0.000	0.000	0.000	0.000
Preparation & Prep. x DIVERGENCE	0.000	0.000	0.000	0.000

Robust standard errors clustered at the country level in parentheses. *** p<0.01, ** p<0.05, * p<0.1. [1] chi2 statistics, p-value in parentheses. [2] p-values.
Columns (1) and (2) estimated with probit. Column (3) estimated with OLS.

TABLE 3. Correcting for the endogeneity of preparation and supervision.

	First-step - OLS		OLS	Probit	OLS	Probit
Dependent variable	Supervision	Preparation	Success	Success	Success	Success
	(1)	(2)	(3)	(4)	(5)	(6)
Supervision predicted			0.746*	1.949	0.739*	1.993*
			(0.400)	(1.221)	(0.381)	(1.128)
Supervision predicted x LICUS			0.421	1.242	0.445**	1.087*
			(0.431)	(1.227)	(0.188)	(0.605)
Preparation predicted			0.284	1.259	0.291	1.207
			(0.534)	(1.694)	(0.519)	(1.573)
Preparation predicted x LICUS			0.0400	-0.268		
			(0.643)	(1.794)		
Dummy = 1 for LICUS	-0.0854**	-0.0739*	-1.243	-2.580	-1.198**	-2.883*
	(0.0408)	(0.0375)	(0.888)	(2.590)	(0.515)	(1.650)
Dummy IDA = 1	-0.00358	-0.0218	0.0565*	0.180**	0.0567*	0.178**
	(0.0324)	(0.0374)	(0.0306)	(0.0901)	(0.0297)	(0.0897)
Dummy Investment project = 1	-0.0703*	0.0374	0.128**	0.370**	0.128**	0.371**
	(0.0422)	(0.0505)	(0.0565)	(0.175)	(0.0549)	(0.171)
Duration of project	-0.0459***	-0.0427***	0.0157	0.0457	0.0156	0.0463
	(0.00709)	(0.00753)	(0.0100)	(0.0327)	(0.0108)	(0.0327)
Ln GDP pc, initial	-0.0196	0.00542	0.0193	0.0526	0.0192	0.0531
	(0.0209)	(0.0253)	(0.0196)	(0.0596)	(0.0190)	(0.0570)
Duration of leader in office, initial	-0.00412***	0.000223	-0.000944	-0.00388	-0.000961	-0.00377
	(0.00146)	(0.00153)	(0.00196)	(0.00586)	(0.00195)	(0.00564)
CPIA, initial	0.00801	0.0500	0.0551*	0.161*	0.0551*	0.161*
	(0.0270)	(0.0303)	(0.0291)	(0.0956)	(0.0301)	(0.0927)
Constant	3.230***	2.759***	-2.727***	-9.999***	-2.728***	-9.984***
	(0.160)	(0.210)	(0.510)	(1.636)	(0.532)	(1.651)
Sum of total budget of five main donors weighted by of distance	0.0490**	0.0564*				
	(0.0231)	(0.0291)				
Sum of total budget of five main donors weighted by same language	-1.937***	-1.226***				
	(0.411)	(0.420)				
Sum of total budget of five main donors weighted by same religion	-0.277	-0.0703				
	(0.347)	(0.377)				
Observations	2023	2023	2023	2023	2023	2023
Countries	102	102	102	102	102	102
Partial R-squared / R-squared	0.063	0.049	0.077		0.077	
F-test	9.00	4.35				
Joint significance (p-values) Sup & Sup x LICUS			0.0468	0.0777	0.0027	0.0201
Prep & Prep x LICUS			0.8542	0.7581		

*** p<0.01, ** p<0.05, * p<0.1. Col (1) and (2): Robust standard errors clustered at the country level in parentheses. Col (3)-(6): Bootstrapped standard errors in parentheses. Note: the Sargan over-identification test for the second step after the first step presented in columns (1) and (2) has a p-value of 0.20.

TABLE 4. Correcting for the endogeneity: multivariate probit estimations.

Dependent variable	Success	Supervision	Superv. x LICUS	Preparation
	(1)	(2)	(3)	(4)
Supervision	1.769***			
	(0.185)			
Supervision x LICUS	0.329**			
	(0.167)			
Preparation	1.144***			
	(0.125)			
Dummy = 1 for LICUS	-0.288**	-0.0763	6.685***	-0.142
	(0.144)	(0.114)	(0.192)	(0.104)
Dummy IDA = 1	0.247**	-0.00849	0.0536	-0.0609
	(0.0983)	(0.0927)	(0.0917)	(0.129)
Dummy Investment project = 1	0.346**	-0.111	0.200	0.174
	(0.135)	(0.115)	(0.165)	(0.125)
Duration of project	-0.0216	-0.109***	-0.133***	-0.114***
	(0.0230)	(0.0172)	(0.0288)	(0.0209)
Ln GDP pc, initial	0.134***	-0.0797	-0.105	-0.0164
	(0.0515)	(0.0561)	(0.0909)	(0.0740)
Duration of leader in office, initial	-0.0102**	-0.0113***	-0.00993	-0.000257
	(0.00434)	(0.00434)	(0.00921)	(0.00489)
CPIA, initial	0.269***	0.0424	0.124	0.0839
	(0.0593)	(0.0671)	(0.114)	(0.0801)
Constant	-3.604***	1.789***	-5.331***	0.820
	(0.444)	(0.455)	(0.847)	(0.601)
Sum of total budget of five main donors weighted by distance		0.135**	0.204**	0.191***
		(0.0640)	(0.102)	(0.0615)
Sum of total budget of five main donors weighted by same language		-5.207***	-5.938***	-3.283***
		(1.042)	(1.141)	(1.187)
Sum of total budget of five main donors weighted by same religion		-1.117	-2.423**	-0.757
		(1.022)	(1.229)	(0.955)
Observations	2023			
Countries	102			
Log-Likelihood	3033.5			
Probability that all $\rho = 0$	0.0000			

*** p<0.01, ** p<0.05, * p<0.1. Standard errors clustered at country level in parentheses. *Supervision* and *Preparation* are transformed into binary variables (see Appendix 2). By construction, *Supervision* x *LICUS* is also a binary variable.

TABLE 5. Quality of supervision in LICUS and non-LICUS countries.

In % of rated projects	Divergent interests (LICUS)	Non-divergent interests
Highly satisfactory	3	6
Satisfactory	68	72
Unsatisfactory	27	21
Highly unsatisfactory	2	1
	100	100
Not rated (in % of total)	37	23

Source: IEG, authors' calculations.

TABLE 6. Supervision as a function of project characteristics and dummy LICUS

Supervision	(1)	(2)	(3)
Dummy LICUS	-0.229***	-0.193**	-0.167*
	(0.0805)	(0.0844)	(0.0940)
Duration		-0.113***	-0.0986***
		(0.0156)	(0.0151)
Capacity = 1		0.217***	0.214***
		(0.0670)	(0.0661)
IDA = 1			-0.0496
			(0.0740)
Investment = 1			-0.146
			(0.103)
NGO = 1			0.0664
			(0.143)
Observations	2023	2023	2023

Ordered probit estimations. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Robust standard errors clustered at the country level.

APPENDIX 1

1.a. Proofs of the results in Section II

D's problem can be written as

(1) $$\max_{b,\varepsilon} E_\mu[a^* \cdot 1 - (1-b)c - \mu a^* \cdot bc] - C(\varepsilon; \theta)$$

(2) s.t. $$a^* \in \arg\max_a \; a \cdot \theta + (1-b)c + \mu a \cdot bc - \gamma a^2$$

(3) $$b \in [0,1]$$

where (2) denotes R's incentive compatibility constraint. R's optimal effort level is then given by

(4) $$a^* \equiv a(b, \theta, \mu) = \frac{\theta + \mu bc}{2\gamma},$$

such that D's problem becomes

(5) $$\max_{b,\varepsilon} E_\mu \left[\frac{\theta + \mu bc}{2\gamma} - (1-b)c - \mu \cdot bc \cdot \frac{\theta + \mu bc}{2\gamma} \right] - C(\varepsilon; \theta)$$

(6) $$= \max_{b,\varepsilon} \left[\frac{\theta + bc - (\theta bc + (1 + \varepsilon^{-1})b^2 c^2)}{2\gamma} - (1-b)c \right] - C(\varepsilon; \theta)$$

(using $E(\mu) = 1$ and $E(\mu^2) = 1 + \varepsilon^{-1}$).

The first-order conditions are then given by

(7) $$/b \; : \quad \frac{c - \theta c - 2(1 + \varepsilon^{-1})bc^2)}{2\gamma} + c = 0,$$

(8) $$/\varepsilon \; : \quad \frac{\varepsilon^{-2}b^2 c^2}{2\gamma} - \frac{\partial C(\varepsilon; \theta)}{\partial \varepsilon} = 0.$$

From (7), we get that[2]

(9) $$b^* = \frac{(1 - \theta + 2\gamma)\varepsilon}{2c(1 + \varepsilon)},$$

which implies Proposition 1. Plugging (9) into (8), the following condition for optimal precision ε^* is obtained:

(10) $$\frac{\partial C(\varepsilon^*; \theta)}{\partial \varepsilon}(1 + \varepsilon^*)^2 = \frac{(1 - \theta + 2\gamma)^2}{8\gamma}.$$

---

[2] We assume that γ takes a value such that $a^* \in [0,1]$ and $b^* \in [0,1]$. In particular, this means that we require $\gamma \leq \dfrac{2c(1 + \varepsilon) - \varepsilon(1 - \theta)}{2\varepsilon}$.

(It is easy to verify that the second-order conditions for a maximum are satisfied).

To prove Proposition 2, plug (9) into (4) and take the expectation with respect to μ, to obtain

$$(11) \qquad \Pr(success) = E_{\mu}[a^*] = \frac{1}{2\gamma}\left[\theta + \frac{(1-\theta+2\gamma)\varepsilon}{2(1+\varepsilon)}\right].$$

Likewise, $V^*(\varepsilon, \theta)$ as defined in the text can easily be found and is given by

$$(12) \qquad V^*(\varepsilon, \theta) = \frac{1}{2\gamma}\left[\theta + \frac{\varepsilon(1-\theta+2\gamma)(1-\theta-2\gamma)}{4(1+\varepsilon)}\right] + \frac{(1-\theta+2\gamma)\varepsilon}{2(1+\varepsilon)} - c.$$

From this, we obtain

$$(13) \quad \frac{\partial \Pr(success)}{\partial \theta} = \frac{1}{2\gamma}\left[1 - \frac{\varepsilon}{2(1+\varepsilon)}\right] > 0;$$

$$(14) \quad \frac{\partial \Pr(success)}{\partial \varepsilon} = \frac{1}{2\gamma}\left[\frac{(1-\theta+2\gamma)}{2(1+\varepsilon)^2}\right] > 0;$$

$$(15) \quad \frac{\partial^2 \Pr(success)}{\partial\theta\partial\varepsilon} = \frac{1}{2\gamma}\left[-\frac{1}{2(1+\varepsilon)^2}\right] < 0;$$

$$(16) \quad \frac{\partial V^*(\varepsilon, \theta)}{\partial \theta} = \frac{1}{2\gamma}\left[1 - \frac{\varepsilon(1-\theta)}{2(1+\varepsilon)}\right] - \frac{\varepsilon}{2(1+\varepsilon)} > 0;$$

$$(17) \quad \frac{\partial V^*(\varepsilon, \theta)}{\partial \varepsilon} = \frac{1}{2\gamma}\left[\frac{(1-\theta+2\gamma)(1-\theta-2\gamma)}{4(1+\varepsilon)^2}\right] + \frac{(1-\theta+2\gamma)}{2(1+\varepsilon)^2} > 0;$$

$$(18) \quad \frac{\partial^2 V^*(\varepsilon, \theta)}{\partial\theta\partial\varepsilon} = -\frac{1}{2(1+\varepsilon)^2}\left[\frac{(1-\theta+2\gamma)}{2\gamma}\right] < 0;$$

((16) holds because of the assumption that $\gamma \leq \dfrac{2c(1+\varepsilon)-\varepsilon(1-\theta)}{2\varepsilon}$; (17) holds as the

expression attains its minimum at $\theta = 1$, and its minimal value is given by $\dfrac{\gamma}{2(1+\varepsilon)^2} > 0$.)

Finally, Proposition 3 follows from standard comparative statics results (see for instance Athey, Milgrom and Roberts 1998). $\arg\max\limits_{\varepsilon} \pi(\varepsilon; \theta)$ is non-increasing in θ if and only if

$\pi(\varepsilon; \theta) = V^*(\varepsilon; \theta) - C(\varepsilon; \theta)$ has decreasing differences in $(\varepsilon; \theta)$, which in this context

means that $\dfrac{\partial^2 \pi}{\partial\theta\partial\varepsilon} = \dfrac{\partial^2 V^*}{\partial\theta\partial\varepsilon} - \dfrac{\partial^2 C}{\partial\theta\partial\varepsilon} \leq 0.$

1.b. Two-dimensional contracts

We will now briefly explore what happens if the contract proposed by D can contain a fixed component as well as a variable component (based on whether $p=1$); in other words, we allow for two-dimensional contracts and not just one-dimensional ones as in the main part of the paper.

Formally, D now proposes a contract $(f, \tilde{b}, \varepsilon)$, where f is a fixed payment that R receives as soon as accepting the contract (like $(1-b)c$ in the main part of the paper), $\tilde{b}$ is the 'bonus' he receives if $p=1$ (corresponding to bc in the main part), and ε is precision. It is assumed that these two payments together must at least cover the cost of the project, c.

As we did in the main part, we will assume that R's participation constraint is satisfied, such that D's problem can now be written as

$$(19) \qquad \max_{b,\varepsilon} \ E_\mu[a^* \cdot 1 - f - \mu a^* \cdot \tilde{b}c] - C(\varepsilon; \theta)$$

$$(20) \qquad \text{s.t.} \qquad a^* \in \ \arg\max_{a} \ a \cdot \theta + f + \mu a \cdot \tilde{b}c - \gamma a^2$$

$$(21) \qquad f + \tilde{b} \geq c$$

It is obvious that (21) must bind at the optimum, as otherwise it would be profitable for D to reduce f (which would increase her profit without affecting R's effort decision). Thus, $f^* = c - \tilde{b}^*$.

Solving for R's optimal effort decision and substituting it into (19), we can then rewrite the problem as

$$(22) \qquad \max_{b,\varepsilon} \ E_\mu \left[\frac{\theta + \mu\tilde{b}}{2\gamma} - (c - \tilde{b}) - \mu\tilde{b} \cdot \frac{\theta + \mu\tilde{b}}{2\gamma} \right] - C(\varepsilon; \theta).$$

This is identical to (5) except that $\tilde{b} = bc$; thus, $\tilde{b}^* = b^*c = \dfrac{(1-\theta+2\gamma)\varepsilon}{2(1+\varepsilon)}$, and all the other results remain the same. Note that $\tilde{b}^*$ is independent of c, as R's effort choice and D's benefit from a successful project are both independent of c.

Yet, there is a caveat to this analysis: the assumption that the recipient's participation constraint is satisfied is no longer innocuous. R's ex-ante expected utility from a contract

$(f, \tilde{b}, \varepsilon)$ equals $f + E_\mu \left[\dfrac{(\theta + \mu \tilde{b})^2}{4\gamma} \right]$, which must be bigger or equal to R's outside option

(which we assumed to equal zero) to induce participation. However, with the optimal contract

derived above, $f^* = c - \tilde{b}^* = c - \dfrac{(1 - \theta + 2\gamma)\varepsilon}{2(1 + \varepsilon)}$ could be negative for small c, which could in

turn lead to a negative expected utility for some parameter values and thus to a rejection of the

contract. Thus, the equivalents of Proposition 1-3 remain valid without further assumptions only

if $c \geq \bar{c}(\theta)$. As ex-ante expected utility increases in θ, the lowest threshold value is $\bar{c}(0)$,

which can be shown to equal $\dfrac{(1 + 2\gamma)\varepsilon}{2(1 + \varepsilon)} \left(1 - \dfrac{1 + 2\gamma}{8\gamma} \right)$.

Success of the project:
Defined according to the 'Outcome' variable which assesses the extent to which the project's major relevant objectives were achieved, or are expected to be achieved, efficiently (Source: IEG). Outcome is assessed on a 6-point scale: highly satisfactory (6), satisfactory, moderately satisfactory, moderately unsatisfactory, unsatisfactory and highly unsatisfactory (1). 'Success' is a dummy variable equal to one for the three highest ratings of outcome and zero otherwise.

Preparation:
This variable assesses the government / implementing agency performance in the preparation of the project. It considers specifically whether the government / implementing agency took account of economic, financial, technical, policy, and resource considerations, and ensured participation of major stakeholders in preparing the project (Source: IEG). It is rated on a 4-point scale: highly satisfactory (4), satisfactory, unsatisfactory, highly unsatisfactory (1). When this variable is transformed into a binary variable, it is equal to one when preparation is highly satisfactory and satisfactory, and zero otherwise.

Supervision:
This variable assesses the extent to which services provided by the World Bank supported implementation through appropriate supervision (Source: IEG). Two kinds of factors are considered to assess supervision. The first set of factors focus on development impact (timely identification of problems, appropriateness of solutions, effectiveness of World Bank supervision actions), while the second set of factors refers to the adequacy of supervision inputs and processes (adequacy of supervision resources, reporting quality, attention to fiduciary aspects). Each factor is rated as follows: (i) high (Bank performed all supervision actions with no shortcomings); (ii) substantial (Bank performed supervision actions generally well but with some shortcomings); (iii) modest (Bank supervision had major shortcomings); negligible (Bank largely failed to perform supervision). Overall supervision is rated on a 4-points scale: highly satisfactory (the project was rated at least 'substantial' on all factors, and 'high' on some), satisfactory (the project was rated at least 'substantial' on most factors), unsatisfactory (the project was rated less than 'substantial' on most factors), highly unsatisfactory (the project was rated 'negligible' on most factors). When this variable is transformed into a binary variable, it is equal to one when supervision is highly satisfactory and satisfactory, and zero otherwise.

IDA: is a dummy variable equal to one if the project is financed by IDA and zero if it is financed by IBRD (Source: IEG).

Investment: Dummy variable referring to the type of lending instrument. Lending instruments can be either 'investment' (dummy equals one) or 'adjustment' (dummy equals zero) (Source: IEG).

Duration: duration of the project. This variable corresponds to the duration between the starting date of the project (signature) and the original closing date of the project (Source: IEG).

GDP pc: Logarithm of initial GDP per capita (in constant dollars) (Source: WDI, 2004).

Duration leader in office: Number of years the national leader had been in office. '0' indicates transition year. Source: Gurr, Harff and Marshall ("State Failure Task Force", 2003) and Bienen and van de Walle ("Of Time and Power: Leadership Duration in the Modern World", 1991, Center of International Studies, Princeton University).

CPIA: Country Policy and Institutional Assessment. It has 16 equally weighted components, divided into four categories (6-point scale): (1) Macroeconomic management and sustainability of reforms; (2) Structural policies for sustainable and equitable growth; (3) Policies for social inclusion; (4) Public sector management. The initial value of the CPIA is introduced (starting year of the project). Source: World Bank.

LICUS countries: Dummy equals to one when the CPIA (averaged over the duration of the project) is less than 3 and when the country was a LIC for at least one year during the project (Source: World Bank).

APPRAISAL dummy: This dummy is equal to one when the first sub-component of the PIMI indicator is below the median of the sample. The first subcomponent - APPRAISAL (Strategic Guidance and Project Appraisal) - is assessed on the grounds of:
- Nature of strategic guidance and availability of sector strategies;
- Transparency of appraisal standards;
- Observed conduct of ex ante appraisals;
- Independent review of appraisals conducted.
Source: Dabla-Norris et al. (2011)

Instruments for supervision and preparation - Distance and supply-side variables:

Same language as donor i : dummy taking the value of one if the donor country and the recipient country share a common language [from Collier, Hoeffler and Pattillo (2004), source: CIA factbook (2003)].

Same religion as donor i : dummy variable taking the value of one if 30 percent or more of the population belong to one religious group in the donor as well as in the recipient country [from Collier, Hoeffler and Pattillo (2004), source : Barrett (1982)].

Distance from capitals: it is measured as the distance in kilometres between the capitals of the recipients and Washington D.C., Tokyo and Brussels [Collier, Hoeffler and Pattillo (2004), source: data made available by the World Bank]

Total aid budget of donor i: total net disbursements of ODA by donors *i*, in constant prices 2001 (*i* = France, Germany, Japan, UK, USA) (Source: OECD).

[i] Dollar and Svensson (2000) use regional dummies, per capita income and population as well as some project financial characteristics such as the number of conditions or loan size to instrument for preparation and supervision. Kilby (2000) tries to circumvent the issue of endogeneity by examining the relationship between supervision in a given year and the subsequent intermediate measure of project performance. Kilby (2012) constructs a predicted duration of project preparation using a stochastic frontier model and geopolitical variables (votes at the UN, military aid, UN Security Council non-permanent member) as sources of exogenous variation for the duration of preparation. Our instrumentation strategy is close to Kilby (2012) in that we use the geographical and cultural proximity of donors and recipient as proxies for the relative importance of recipients for the five main donors (see Tavares, 2003). We then weight the aid budget of the five main donors by these proximity variables. The objective of this identification strategy is to avoid using recipient or project characteristics as instruments, since those are unlikely to be exogenous to the success of aid projects.

[ii] Unlike them, we only consider a one-shot setting; thus, reputational contracts (which are the focus of their paper) are ruled out.

[iii] It may be natural to wonder how much the results in this section depend on the apparently artificial restriction to a one-dimensional space of contracts. This question is addressed in Appendix 1b).

[iv] This is related to Besley and Ghatak (2005), where θ measures the extent to which a worker identifies with the mission of the organization he works for.

[v] This is the assumption usually made in 'traditional' principal-agent theory.

[vi] His ex-ante utility (before he has observed μ), relevant for the participation decision, is $E_\mu(a^*\theta + (1-b)c + \mu a^* bc - \gamma a^{*2})$. As it is assumed that his outside option (what he gets if he rejects the project proposal) is 0, it is however easy to show that his participation constraint is always satisfied, such that his ex-ante utility plays no role.

[vii] Unfortunately the list of countries considered as 'Difficult Partnerships' by the OECD is not available.

[viii] Since the actual implementation phase is potentially endogenous to performance, we use the closing date for the project that was anticipated at the time of the presentation of the project for approval by the Board of the World Bank.

[ix] When income and the CPIA are dropped from the regression the direct effect of the dummy variable for divergent interests is highly significant and negative.

[x] In terms of the model of Section II, the issue concerns the cross-derivative of Pr(success) with respect to project preparation and θ. Let better preparation increase the value of the successful project, v. We then find that the optimal b increases in v, as does Pr(success), and $\frac{\partial^2 \Pr(success)}{\partial v \partial \theta} > 0$. Thus, preparation is predicted to have a higher marginal positive impact on the likelihood of project success the more congruent are the interests of D and R.

[xi] Overall, regression 1 of Table 2 has a good predictive power: out of the 2023 observations, only 19.4% are wrongly predicted for a cut-off point equal to 0.5.

[xii] When we introduce only the sector dummies, the results are very similar to those obtained with sector dummies and their interaction with *LICUS*.

[xiii] The sample of projects/countries is divided by two when we use the APPRAISAL dummy. We ran the regression of column 2 on this restricted sample and the results are unchanged.

[xiv] It is worth noting that estimating the first-step using an ordered probit yields very similar results.

[xv] As we instrument two variables, we need three instruments to compute the Sargan test for over-identification (p=0.20). Therefore, even though one of the instruments is not individually significant, we keep it as instrument in the first stage.

[xvi] We also estimated the first-steps of Table 3 using ordered probit and the results are very similar to those obtained using OLS.

[xvii] Unfortunately, we do not have enough identifying variables to estimate the model with a fifth equation for *Preparation* x *LICUS*.