A SYSTEMATIC REVIEW ON

evidence of the impact on investment rates
of reforms to improve the enforcement of contracts

Diego Aboal, Nelson Noya, Andrés Rius

First Draft for Review
Montevideo, Uruguay - March 31, 2011
## Contents

Summary 2

1. Background 3
   1.1. Introduction 3
   1.2. Definitional and conceptual issues 4
   1.3. Research background 7
   1.4. Aims and review questions 9
   1.5. Type of review 9
   1.6. User involvement 9

2. Mapping: methodology 10
   2.1. Overview 10
   2.2. Defining relevant studies: inclusion and exclusion criteria 10
   2.3. Identification of potential studies: search strategy 11
   2.4. Screening studies: applying inclusion and exclusion criteria 14
   2.5. Characterising included studies 14
   2.6. Identifying and describing studies: quality-assurance process 15

3. Systematic map results 16
   3.1. Overview 16
   3.2. Identifying studies 16
   3.3. Summary of findings from the systematic map 16

4. Synthesis methodology 19
   4.1. Overview 19
   4.2. Detailed description of studies in the synthesis 19
   4.3. Assessing quality of studies and weight of evidence for the review question 20
   4.4. Synthesis: quality-assurance process 21
   4.5. Synthesis of evidence 21

5. Synthesis results 22
   5.1. Overview 22
   5.2. Characteristics of studies included in the synthesis 22
      5.2.1 Characteristics of the selected studies 25
   5.3. Weight of Evidence 26
   5.4. Synthesis of evidence 28

6. Conclusions and implications 29

References 30

Appendix 1: Coding tool 33
Summary

This systematic review focuses on the evidence about one specific causal mechanism: from better enforcement of contracts to higher rates of capital accumulation, either directly or indirectly (e.g., through availability of financing).

The rationale for the review does not rest exclusively on the (still debated) link from investment to growth, but also on the fact that donors and governments do invest resources and political capital in improving the business environment, and in particular in seeking to improve the enforcement of contracts. While some of these efforts could be simply justified on grounds of promoting the rule of law, the underlying assumption for many of those reform efforts is that investments will be unleashed by them, so analyzing systematically the evidence in favour of that assumption may eventually help in deciding what priority those reforms should have.

Our systematic review adopts an approach to synthesis that shares with “realist synthesis” literature a concern with uncovering patterns of context-mechanism-outcome. This means that we are not only interested in discerning whether enforceability reforms have been followed by increases in investment, but also in the mechanisms that may cause changes in enforceability to influence investment decisions and aggregate investment rates. Therefore, in our review we put a premium on studies that explore various causal linkages.

The systematic review followed strict methodological guidelines made explicit in the systematic review protocol reviewed and extensively commented by 3 anonymous referees. Additionally it had the support of EPPI-centre’s experts.

Initially, inclusion-exclusion criteria were applied to the results of the comprehensive and systematic searches of relevant scholarly databases and search engines. We screened titles and abstracts of 2229 non-duplicated studies. Titles and abstracts were double screened by two different reviewers. 114 studies were classified as included for full text screening and the other 2115 were excluded. From the 114 studies included we were able to retrieve 88 studies for full text screening. From these 88 studies, and after applying the inclusion and exclusion criteria to the full text by a single reviewer, 70 were excluded and 18 were included. The dubious cases were double screened by a second reviewer and decisions were made jointly. From these 18 papers 3 are surveys of the literature and were used only as a source of additional bibliographical references. The other 15 are the ones selected for synthesis.

To assess the quality of the included studies, the review team used a critical appraisal approach based on a multi-dimensional concept of quality in research. This approach covers quality of reporting, methodological rigour, conceptual depth and breadth, and relevance. The quality assessment is done with two purposes: first, to exclude studies that clearly do not meet minimum professional/academic standards; second, to generate quality ratings to qualify synthesis results.

Cross-country evidence seems to dominate the relevant body of research (50% of the quantitative studies). This is reflected in geographic scope of the studies as well as unit of analysis (which mirrors in numbers but doesn’t coincide mechanically with the former). The prevailing research designs are cross-section and panel data regressions, with estimation methods customized to the specificities of data sets, variables, and/or reflecting trends in the econometric “best practice”.

Overall, the evidence gathered through this systematic review provides weak support for the claim that more effective contract enforcement promotes higher levels of investment. First, there is no study that unambiguously links an intervention or reform to enhance contract enforcement to changes in investment patterns. Second, few of the studies go beyond a generic discussion of direct and indirect effects to actually test the plausible indirect causal channels. Third, almost all the studies do very little or nothing in terms of robustness checks, or the strenuous but necessary attempts to rule out alternative explanations for the empirical findings.

This means that there is ample scope, and need, to enhance the evidentiary basis for a “conventional wisdom” hypothesis that seems weakly supported by evidence.
1. Background

1.1. Introduction

The key role of capital accumulation in promoting economic development has been almost a truism in Economics since the classical economists (Smith, Marx, Ricardo). Investment was one of the obvious ways to promote economic growth in the basic “modern” growth models (Solow, 1956), and the specific circumstances of underdeveloped economies in this regard have been more systematically explored since development economics became a recognizable sub-discipline (see, e.g., Rostow, 1960; Hirschman, 1958; and various contributions in Meier and Seers, 1984).

Even if development is conceptualized to be broader than economic growth, the growth of incomes and wealth are generally recognized to be powerful instruments to expand opportunities and reduce deprivation (e.g., Sen, 1999). For most development agencies, poverty reduction is the primary concern. Since 2000, the broader Millennium Development Goals have guided the efforts of those agencies, as well as those of most of the 180+ governments that adopted them. It is for these reasons that many development agencies are interested in strategies to promote sustained growth, and they have been promoting regulatory and policy reforms to attain that (White, 2008).

Growth is generally assumed to depend on sustaining high rates of investment, not exclusively but fundamentally by the local private sector (DFID, 2009). The investment-growth assumption is straightforward for many practitioners and agencies, and supported by some evidence (Levine and Renelt, 1992; Haussman, Pritchett, and Rodrik, 2005), but it has also been challenged (e.g., Dollar and Easterly, 1999; Devarajan, Easterly and Pack, 2001). The latter group of scholars would claim that investment is at least partly endogenous—i.e., growth promotes investment—and that, particularly in low income countries, both low growth and low investment can be symptoms of other underlying factors.

Among the factors that could determine growth and/or investment performance, institutions have become more prominent in the scholarly literature in recent years (North, 1990; Rodrik, 2000; Acemoglu and Johnson, 2005). The literature identifies a host of growth-and investment-relevant institutions, and their effects on either growth or investment can be direct or through some indirect channel. Salient among them are the institutions that protect investors from expropriation and those that determine how contracts are enforced. Douglass North argues that with increased specialization, larger numbers of trading partners, and geographic dislocation of transactions, more complex contracts became necessary, and therefore the institutions that reduce uncertainty about their execution became more crucial (North, 1990).

This systematic review focuses on the evidence about one specific causal mechanism: from better enforcement of contracts to higher rates of capital accumulation, either directly or indirectly (e.g., through availability of financing). The review does not assume necessarily that investment strictly causes growth (though it is justified, to some degree, by the assumption that facilitating investment will somehow benefit the growth process), nor does it examine all the linkages from quality of institutions to investment (e.g., we do not review studies focused on the impact of ambiguous property rights on investment).
The rationale for the review does not rest exclusively on the (still debated) link from investment to growth, but also on the fact that donors and governments do invest resources and political capital in improving the business environment, and in particular in seeking to improve the enforcement of contracts. While some of these efforts could be simply justified on grounds of promoting the rule of law, the underlying assumption for many of those reform efforts is that investments will be unleashed by them, so analyzing systematically the evidence in favour of that assumption may eventually help in deciding what priority those reforms should have.

1.2. Definitional and Conceptual Issues

The New Institutional Economics (NIE) has highlighted the role of institutions in shaping economic agents’ decisions and, ultimately, shaping also aggregate economic performance. It is then apposite to look to one of its founders for guidance on key definitional issues. Douglass North (1990) provides a conceptual framework that has been largely followed by many or most of the studies that are reviewed.¹

According to North, economic exchanges inevitably involve transaction costs and asymmetries of information, and it is to make these manageable (and the fundamental exchange viable) that economic agents devise institutions. Institutions are “the humanly devised constraints that shape human interaction” (North, 1990, p. 3; all the following page citations refer to the same volume). They can be formal or informal, and their main difference is that formal institutions are written ones (and there may be a written process code for how to obtain their enforcement or allocate pre-defined penalties). In fact, other than their written character, formal institutions only differ in degree from informal ones (p. 46), and they emerge as the increasing complexity of society raises the benefits from the formalization of constraints.

“Formal rules include political (and judicial) rules, economic rules, and contracts” (p. 47). This distinction is important: the review focus on enforcement of contracts, and will exclude the enforcement of general economic rules. While the latter are typically associated with some enforcement mechanism, to the extent that they represent unilateral impositions of the State and affect a broader population of agents that are expected to abide by them, we will not include them in our study (contracts is therefore reserved for mostly bilateral and voluntary agreements). As with regards to informal institutions, they can be “(1) extensions, elaborations, and modifications of formal rules, (2) socially sanctioned norms of behaviour, and (3) internally enforced standards of conduct.” (p. 40). They may also stipulate enforcement mechanisms, as we will see.

The whole theory of institutions and economic performance advocated by North rests on the adoption of new institutional arrangements as the agents’ response to the increasing complexity of economic transactions. In pre-modern societies, transactions were essentially personalized exchanges among “neighbours”, and production and trade was in small scales. Reputation and the risk of isolation from a

¹ North’s and the NIE’s framework are not free of conceptual problems (for some difficulties with North’s, see Field, 2006; for an appraisal of the NIE see Rutherford, 1994). However, it is appropriate to borrow basic definitions from the original source, given that these definitions are widely used, and the framework provides a basic benchmark to which complications or refinements can be compared.
community could function effectively to prevent or address opportunism. Gradually, impersonal exchanges among more distant parties became more frequent and economically significant, which led to the emergence of informal institutions with more explicit enforcement arrangements (these would include, for example, the ostracism of those who violated agreements, stipulated in unwritten codes of commercial conduct; p. 43).

With complex contracts that contain many hard-to-measure attributes about exchanged goods and services, and that are plagued by information asymmetries; and with the expansion of the reach of trade and the chances that transactions may never be repeated between the same two parties, it became necessary to devise third-party enforcement. In fact, it would be more appropriate to say that in modern societies the three forms of exchanges (and enforcement arrangements) co-exist, and even archaic and seemingly dysfunctional informal rules can have major impacts, as demonstrated by the evidence that the same formal rules imposed on different societies produce different outcomes (p. 36).

Our review focuses on one specific set of impacts of different forms of contract enforcement.

“Enforcement poses no problem when it is in the interests of the other party to live up to agreements. But without institutional constraints, self-interested behaviour will foreclose complex exchange, because of the uncertainty that the other party will find it in his or her interest to live up to the agreement.” (p. 33)

Enforcement can come from societal sanctions, from second-party retaliation or from a coercive third party (typically, the State), and the long-range economic history of the world shows each of these forms prevailing at some set of space-time coordinates. We do not exclude informal (multilateral, bilateral, or third party) enforcement from our review, even though most of the empirical literature focus on formal, third-party (state-backed) enforcement. (Some carefully executed, analytically rich and extremely interesting studies on informal institutions, such as Besley, 1995, did not meet our eligibility criteria due to their exclusive attention to definition of property rights rather than contract enforcement.)

As noted by Acemoglu and Johnson (2005), the NIE has persuaded many economists and political scientists that institutions are a primary determinant of economic performance. However, in much of the literature there has been a tendency to conflate a variety of economic institutions in a “cluster” that presumably defines a favourable business environment (on “business environment”, see next section). In their work, Acemoglu and Johnson distinguish contracting institutions, which are the institutions supporting private contracts, from property rights institutions, which are the institutions constraining government and elite expropriation.

Like these authors, we also note that there is much overlap between the two types (inadequate enforcement of private contracts could result in some form of expropriation, and constraints on governments’ expropriatory powers could be contained in seemingly voluntary contracts). In this review, we adopt a definition of relevant contracts that includes private contracts (as in Acemoglu and Johnson, 2005) but also the bilateral, mostly voluntary agreements that can be established between states and private parties (such as, e.g., when a private company gets a concession through public bidding, and agrees to certain terms that either the state or the private party could be tempted to violate, thus requiring third-party enforcement). We exclude from our review, in turn, general economic rules imposed unilaterally by the state, as these lack the voluntary aspect.
Our review focuses on the impact of reforms affecting the enforcement of contracts on rates of investment. We define investment as the accumulation of productive assets. These can be tangible (such as buildings, equipment or permanent plantations) or intangible (such as productive methods, or commercial patents). We are interested in investments that enhance productive capacities and are either made by domestic or foreign agents; we are not interested in foreign “direct investment” that simply acquires existing companies (or parts of them) without adding to the stock of productive assets (although the distinction is sometimes hard to make in practice).

There is a range of possible reforms that may directly or indirectly impact on the enforcement of contracts. Typical donor-funded reform programs, for example, tend to tackle simultaneously a number of “problems” in the laws and their enforcement (for example, rectifying identified flaws in the letter of commercial or civil laws, creating non-judicial arbitration mechanisms, facilitating access by aggrieved parties to the judicial system, reducing various costs of litigation, strengthening the capacities of the courts and judges, etc.). All these have some bearing on the speed and effectiveness of contract enforcement, and more broadly on “the rule of law” (see, e.g., World Bank, 2001). Moreover, other policies not directly connected to the contents or enforcement of written laws, may directly impact on contract enforcement. Woodruf (1998), for example, identifies a more or less direct effect of trade liberalization on informal contract enforcement; substantial investments in information technologies in the judiciary –i.e., public investments—may reduce litigation times and procedural mistakes, and therefore improve enforceability; etc.

This wide variety of possible “interventions” has important consequences for our review:

(a) One cannot expect to find a reform (“intervention”) that removed swiftly a single problem, leaving everything else unchanged and enhancing the quality of contract enforcement;

(b) Additionally, it is extremely difficult to find a study that could identify a control group. This has to do with the characteristics of most reforms (wide ranging reforms), that are related to the “public good” nature of justice by the courts and the formal equality of citizens before the law, although some “reforms” may have been deployed gradually, thus generating within-country variety of “exposure” over time; and other countries may serve as “controls”.

(c) In general, interventions are complex sets of simultaneous (and gradual) changes, taking place over more or less extended periods of time.

(d) “Reforms” can be expected to be reflected in bigger or smaller “jumps” in some of the usual international indicators for quality of contract enforcement, but may be indistinguishable from other, unintended, unplanned, uncontrolled, and/or endogenous changes in institutions. Our typical quantitative study treats all variations in enforcement recorded by the usual indicators similarly, regardless of whether actual “reform processes” are behind them.

(e) The “interventions” likely reflect very differently in the usual international indicators, and the existence and size of a variation in some of them may be affected by the nature, support, ideology, politics or other spurious aspects of the intervention, as seen by stakeholders or key informants, more than by its practical/objective effects on enforceability.

Analytically, weak enforcement of contracts has been found to impact on investment through a number of channels. First, it could most directly influence the uncertainty surrounding an investment project, or some of its critical activities or dimensions,
and therefore influence investors’ decisions by increasing the project’s costs or reducing its expected returns (note that we are not considering here the risk of expropriation of assets created by the investment, which would pertain to the effects of “property rights institutions”). Second, weak enforcement could inhibit lending, or otherwise influence financial markets in a way that hinders investment (Acemoglu and Johnson, 2005). In this line, some authors have found analytical grounds for the idea that “limited enforceability” not only affects the level of firms’ investments but it also increases its “sensitivity to the arrival of new technologies and generates greater macroeconomic volatility” (Cooley, Marimon and Quadrini, 2004). To the extent that aggregate (output) volatility influences investment (a simple accelerator model could show this), there is here another causal channel from enforcement of contracts to rates/levels of capital accumulation. Others have argued that imperfect enforcement influences, through financial contracts, the size distribution and overall heterogeneity of firms, which could reflect on the level or rate of investment (Monge-Naranjo, 2009). These are just an illustration rather than a complete listing of all relevant impact channels. In section 5, we will discuss the empirical evidence for alternative causal chains.

1.3. Research background

Research on the effects of institutions on economic performance has grown very rapidly since the early 1990s. Theoretical developments (e.g., North’s work) have prompted the generation of indicators and proxies for use in empirical analysis (see, e.g., Knack and Keefer, 1995; Kauffman et al., 2004). At the same time, proliferation of databases that include measures of institutional configurations have also stimulated empirical investigation of new or suitably modified research questions, but also “data-driven” research that has not shed much light on the causal chains or the robustness of meaningful theoretical hypotheses (Aaron, 2000; Keefer, 2004; Williams and Siddique, 2008).

As mentioned above, the enforcement of contracts can be “private” (Hamish et al., 2000). In these cases, it tends to be informal and will be affected by reforms to the formal enforcement mechanisms, but also by other policy reforms (as, e.g., in Woodruf, 1998). That said, countries will have some form of third-party, formal enforcement mechanism, and that is why quality of contract enforcement is usually taken to be an attribute of nation-states or sub-national jurisdictions (Djankov et al., 2003; Acemoglu and Johnson, 2005). For that reason, the quantitative operationalization of quality or effectiveness of enforcement has largely been done by creating indicators that rate such quality for pairs of country-time coordinates. At the same time, the nature of the problem itself leaves limited room for experimental designs in the implementation of reforms: principles of territoriality of the law and equal treatment of all citizens (in addition to the typical complexity of legal reforms) make this an unfriendly territory for randomized control trials and other quasi-experimental designs.

Three broad approaches have been followed to generate cross-country and longitudinal evidence:

- **Indirect measures such as the use of “contract-intensive” money**: The relative use of currency in comparison with contract-intensive money is taken as an indicator of inadequate/weak contract enforcement (Clague et al., 1999)

- **Experts’ assessment**: effectiveness, efficiency and/or fairness of the formal enforcement mechanism is assessed by practitioners and other key informants
and conveyed and aggregated through surveys (Knack and Keefer, 1995; but also La Porta et al. 1997, and Berkowitz et al., 2003; Staats et al., 2000)

- **Quantification of time and pecuniary costs to enforce standard contracts:** Legal experts are not asked about their opinions but to estimate the time and financial costs incurred by a private party to enforce some rather simple economic contracts (e.g., collect a bounced check, or evict a delinquent tenant; Djankov et al. 2003), or more complex lending contracts (Djankov et al. 2008)

A significant body of research has focused on exploiting these data sets and the variations over space and time of institutional factors and economic outcomes. This explains why most of the studies reviewed here are cross-section, panel or longitudinal analysis, for countries, firms or sub-national jurisdictions as the units of analysis. These studies are the ones that yield estimates of quantitative effects. Le (2004) is a good example of the conventional approach. One caution, however, is that the meaning of “reform” in these studies is not the customary one, of a deliberate change in formal institutions, but the studies actually examine more generally the effects of all recordable variations in certain attributes of institutional arrangements.

However, there are various reasons to include other research designs, even if they are sometimes less amenable to offering quantitative estimates of aggregate effects on investment. In fact, some authors have argued that the time series or panel approach is already showing diminishing returns, and that other sources of variation in the quality of institutions need to be exploited to understand their effects (Pande and Udry, 2005). This is so because, to have extensive international and time coverage, available indicators of institutional arrangements are mostly based on expert judgement or just the variation in formal procedures as stipulated in the law, so they may not really measure the variations of interest (i.e., they have validity problems; see Pande and Udry, 2005; Shirley, 2008). Moreover, institutions being partly endogenous to the process of development, the causal analysis of their effects based on time series requires ever more sophisticated and scarce “instruments” (that is, variables that are correlated with the institutional quality indicator but can logically be assumed to be exogenous to the economic outcome being “explained”; see Pande and Udry, 2005; Rodrik, 2005; Rehme, 2007). These two problems are hard to circumvent and provide the justification for alternative approaches that may have other limitations but are better equipped to discern causality in various contexts.
1.4. Aims and review questions

The objectives of the review are to provide a synthesis of the evidence about the impact of policies to enhance contract enforcement on investment. Investment is defined as physical capital accumulation or intangible productive asset accumulation. The review question is: What is the evidence of the impact on investment rates of reforms to improve the enforcement of contracts?, but with “reforms” broadly understood as above.

1.5. Type of review

Our systematic review adopts an approach to synthesis that shares with “realist synthesis” literature a concern with uncovering patterns of context-mechanism-outcome (Pawson et al., 2004). This means that we are not only interested in discerning whether enforceability reforms have been followed by increases in investment, but also in the mechanisms that may cause changes in enforceability to influence investment decisions and aggregate investment rates. Therefore, in our review we put a premium on studies that explore more complex causal linkages.

1.6. User involvement

This systematic review is undertaken thanks to a DFID’s grant, awarded through a call for proposals on pre-defined themes. One of the broad research questions was “What is the evidence of the impact on investment rates of implementing the following investment climate reforms: starting a business, protecting investors and enforcing contracts?”, and DFID indicated that questions could be partitioned into some of its components for the purpose of submitting an application (to make the review manageable, we applied to address the last of the three sets of reforms identified in the question). As mentioned before, DFID specifically identifies creating an environment that favours the development of businesses as a key objective of its Private Sector Development strategy.

It is therefore safe to assume that informing policy and practice (rather than contributing to scholarly literature or academic teaching) is the main purpose of the whole initiative, and therefore of this review. From such “revealed preference” of the funding donor, we chose to undertake a narrative synthesis, which seeks “to contribute to policy-makers’ and practitioners’ ‘sense-making’ -the way they understand and interpret the situations they encounter and the interventions they deploy” (Popay et al., 2006), rather than attempting quantitative meta-analysis.

The user has been interacting with the review team through comments on the draft protocol and intermediate outputs. We expect to continue this interaction through the review and commentary of this output.

---

2 The call invited applications to participate in “a cutting-edge pilot to increase the use of evidence in policy and contribute directly to shaping international development policy and practice”.
2. Mapping: methodology

2.1. Overview

In this section we describe the methodology used to map the relevant studies for this review, i.e., those studies that address our review question: What is the evidence of the impact on investment rates of reforms to improve the enforcement of contracts? In the following sub-sections we describe the different stages of the mapping exercise.

2.2. Defining relevant studies: inclusion and exclusion criteria

The following inclusion and exclusion criteria were developed in the protocol of this systematic review and were the result of a process of discussion with 3 anonymous referees. The inclusion and exclusion criteria established the limits of the mapping exercise.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>Relevance to the general question</strong>: We will include studies that address empirically the causal chain from quality of contract enforcement to levels of investment. A complete causal chain (i.e., going from empirical indication of changes in enforceability to variation in investment) must be empirically assessed in the study for it to be included. (See Appendix 1) While this includes both studies that enable some attribution of impact and others that can only detect correlation, the two will be analyzed separately (the latter group will be treated as a separate population, and will be examined mainly to contribute to the discussion of contexts-mechanisms-outcomes).</td>
</tr>
<tr>
<td>2. <strong>Accessibility</strong>: There is sufficient information available to allow screening, or it is possible to retrieve the full-text</td>
</tr>
<tr>
<td>3. <strong>Languages</strong>: English (Spanish, French and Portuguese, provided studies are abstracted in the indexed databases and key websites).</td>
</tr>
<tr>
<td>4. <strong>Publication date</strong>: on or after 1990;</td>
</tr>
<tr>
<td>5. <strong>Temporal coverage</strong>: studies must document changes in enforcement (and investment) occurred in the 20th century.</td>
</tr>
<tr>
<td>6. <strong>Geographic coverage</strong>: all of the world (despite the focus of the review on reforms in developing countries, it is considered that studies of developed countries, or of countries that have not undertaken reforms, may still shed light on causal mechanisms, or serve comparative purposes)</td>
</tr>
<tr>
<td>7. <strong>Units of analysis</strong>: individual investors, individual firms, industries, regions or countries;</td>
</tr>
<tr>
<td>8. <strong>Type of contracts</strong>: (i) among private parties (including labour contracts), (ii) between private parties and government agencies or branches.</td>
</tr>
<tr>
<td>9. <strong>Type of investment</strong>: those affecting gross capital formation; foreign direct investment (FDI), included if effects on “green field” FDI or capital accumulation by foreign subsidiaries (as opposed to “mergers and acquisitions”) can be discerned.</td>
</tr>
</tbody>
</table>
Exclusion criteria

<table>
<thead>
<tr>
<th>No.</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>General studies of quality of institutions and growth, if they do not allow to discern the marginal effects of enforceability on investment. Either the quality of institutions variables do not distinguish quality of contract enforcement, and/or investment is not assumed at least partly explained by it (or the assumption not tested with empirical evidence).</td>
</tr>
<tr>
<td>2.</td>
<td>Full text not available/accessible.</td>
</tr>
<tr>
<td>3.</td>
<td>Language other than English, Spanish, French or Portuguese.</td>
</tr>
<tr>
<td>4.</td>
<td>Published before 1990.</td>
</tr>
<tr>
<td>5.</td>
<td>Historical studies of pre-XXth century institutions and investment.</td>
</tr>
<tr>
<td>6.</td>
<td>Studies focused on the enforcement of general economic rules; studies focusing on regulations (unless the effect of regulation on “voluntary contract” enforcement, and of this on investment, can be discerned empirically).</td>
</tr>
<tr>
<td>7.</td>
<td>FDI focused exclusively on mergers and acquisitions (M&amp;A)</td>
</tr>
<tr>
<td>8.</td>
<td>Magazine and newspaper articles, editorials, letters, comments/opinion articles</td>
</tr>
</tbody>
</table>

2.3. Identification of potential studies: search strategy

The following databases were comprehensively searched applying the date parameters and other search criteria discussed below:

- Econlit
- RePEc (Research Papers in Economics; www.repec.org)
- Scopus
- JSTOR
- Citeulike (www.citeulike.org)
- Academic Search Complete

Accessibility constraints determined by our institutional subscriptions prevented us from full text access to Reuter’s Web of Science, which includes the widely used Social Science Citation Index. We note, however, that there is evidence of significant overlap among some of the leading scientific databases, and that Web of Science is far from being the most comprehensive, or unambiguously selective, for the social sciences and humanities (Hicks and Wang, 2010). Our own experience shows that, starting from EconLit and Scopus, and expanding to other databases, revealed diminishing returns: the first two databases identified 12 of the final 15 studies, and the additions to EconLit and Scopus produced a very high number of duplicates.

We started to search, but abandoned the search run on “theses” databases, as there was widely uneven access to full texts from our institutional subscriptions, therefore risking the introduction of significant bias of an uncertain nature.

We made extensive use of Google Scholar, for tracking back or forward citations online, as well as for following various leads from the early unsystematic review that led to the proposal and informed the protocol. Overall, we considered the coverage of institutional websites to be quite comprehensive based on the previous search
tools, so we did not perform hand searches of World Bank or other international organizations’ sites.

200 of the most recent entries in the New Institutional Economics e-library of the Social Sciences Research Network (SSRN) were scanned, yielding no new studies. This does not exhaust the scope for improving our “success rates” through hand-searching and snowballing, but suggests that our review may not be seriously biased as it refers to the broader universe of “English-language abstracted” items.

Searches started with broad parameters derived from our research question and inclusion/exclusion criteria. These searches were gradually narrowed under guidance from the three reviewers. All searches were stored to ensure replicability.

The EPPI-Reviewer software was used throughout the review to record searches, manage references, generate reports, record decisions, analyse data and report results.

The parameters of the first search consisted of the Boolean union of type 1 and type 2 keywords (see table of keywords below).
<table>
<thead>
<tr>
<th>Specific policy interventions and reforms</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>enforceability of contracts</td>
<td>Investment</td>
</tr>
<tr>
<td>contract enforcement</td>
<td>capital investment</td>
</tr>
<tr>
<td>contracting institutions</td>
<td>capital accumulation</td>
</tr>
<tr>
<td>enforcement of contracts</td>
<td>capital formation</td>
</tr>
<tr>
<td>enforcement costs</td>
<td>capital stock</td>
</tr>
<tr>
<td>contracting institutions</td>
<td>R&amp;D investment</td>
</tr>
<tr>
<td>contractual practices</td>
<td>R&amp;D expenditure</td>
</tr>
<tr>
<td>contract non-performance</td>
<td>fixed assets</td>
</tr>
<tr>
<td>enforceability of agreements</td>
<td>machinery and equipment</td>
</tr>
<tr>
<td>contractual arrangements</td>
<td>Infrastructure</td>
</tr>
<tr>
<td>dispute resolution systems</td>
<td>growth</td>
</tr>
<tr>
<td>contractual unreliability</td>
<td>development</td>
</tr>
<tr>
<td>contractual reliability</td>
<td>business start up</td>
</tr>
<tr>
<td>contract hold-ups</td>
<td></td>
</tr>
<tr>
<td>contract enforceability</td>
<td></td>
</tr>
<tr>
<td>contract unenforceability</td>
<td></td>
</tr>
<tr>
<td>judicial quality</td>
<td></td>
</tr>
<tr>
<td>court enforcement</td>
<td></td>
</tr>
<tr>
<td>contract intensity</td>
<td></td>
</tr>
<tr>
<td>contractibility</td>
<td></td>
</tr>
<tr>
<td>third part enforcement</td>
<td></td>
</tr>
<tr>
<td>state enforcement</td>
<td></td>
</tr>
<tr>
<td>judicial enforcement</td>
<td></td>
</tr>
<tr>
<td>informal enforcement</td>
<td></td>
</tr>
<tr>
<td>relational contracting</td>
<td></td>
</tr>
<tr>
<td>formal enforcement</td>
<td></td>
</tr>
<tr>
<td>limited enforcement</td>
<td></td>
</tr>
<tr>
<td>limited enforceability</td>
<td></td>
</tr>
<tr>
<td>ability to enforce contracts</td>
<td></td>
</tr>
<tr>
<td>contract intensive money</td>
<td></td>
</tr>
</tbody>
</table>
We searched for items dated between January 1990 and December 2010.

2.4. Screening studies: applying inclusion and exclusion criteria

Initially, studies were screened using pre-defined inclusion and exclusion criteria. Two reviewers applied the criteria independently, first to titles and abstracts only, for all items thrown out by the database searches. Studies were then (i) excluded, (ii) included, or (iii) marked as “pending” if the reviewer was unsure about their inclusion. The two independent reviews were compared and contradictory judgements or “pendings” were decided, moving the “included” ones to the next phase of review of full texts.

Once full texts were retrieved, the reviewers applied inclusion and exclusion criteria, based on assessments of the full texts. Dubious cases were reviewed by two reviewers and discussions were held until consensus was reached. A record was kept in EPPI Reviewer of all decisions, of criteria that had been used, and of those studies that were eventually excluded.

2.5. Characterising included studies

Once the review team was satisfied that the search and identification were not yielding significant new additions to the population of “included” studies, the reviewers filled out the descriptive portion of the coding and appraisal form.

The coding tool, consisting of a series of questions and a checklist (see Appendix 1), was inspired by different sources: a non-published document recommended by one of the reviewers of the protocol (Dickson, personal communication), the EPPI-Centre Data Extraction and Coding Tool for Education Studies V2 (“Coding Studies and Extracting Data for a Review”, EPPI, 2007); Cochrane Qualitative Research Methods Group, 2010, (draft of chapter 6); and Spencer, Ritchie, Lewis and Dillon (2003).

We are using the information extracted using the coding and appraisal form to describe the population of studies, and provide in section 4 some key general statistics for included studies (for example, geographic coverage, type of methods used, types of reforms or institutional variations examined).
2.6. Identifying and describing studies: quality-assurance process

The following mechanisms were part of the quality assurance methods:

- Search strategy controlled by senior reviewers
- Screening (title/abstract) conducted by more than one reviewer for each reference found
- Double-screening of dubious cases (looking at full text) and disagreements discussed by review team.
- Moderation exercises undertaken by senior reviewers, with papers screened and coded (depending on the stage of the review) by at least 2 members of review team. Results compared and discussed until consensus was reached.
3. Systematic map results

3.1. Overview

This section describes the results of the systematic mapping.

3.2. Identifying studies

Initially, the inclusion-exclusion criteria were applied to the results of the searches (2229 non-duplicated studies) based on titles and abstracts only. All titles and abstracts were double screened by two different reviewers.

Studies were classified by each reviewer as (i) excluded, (ii) included, or (iii) marked as “pending” if the reviewer was unsure about their inclusion.

There were 263 disagreements among reviewers. After arbitrating differences, based on abstract and title screening, 114 studies were classified as included for full text screening and the other 2115 were excluded. Most of the studies excluded in this phase were so based on inclusion criterion 1, i.e. they were irrelevant to answer the review question.

From the 114 studies included we were able to retrieve 88 studies for full text screening. From these 88 studies, and after applying the inclusion and exclusion criteria to the full text by a single reviewer, 70 were excluded and 18 were included. The dubious cases were double screened by a second reviewer and an agreement was made with the initial reviewer.

From these 18 papers 3 are surveys of the literature and were used only as a source of additional bibliographical references. The other 15 are direct inputs for the systematic review.

3.3. Summary of findings from the systematic map

From the systematic search and screening, 15 studies were identified that could provide evidence on the research questions. This includes one qualitative study that does not attempt to estimate quantitative impact of contract enforcement on investment but directly tackles the issue with qualitative evidence and was retained as it might contribute to understanding causal mechanisms (this treatment for qualitative studies was proposed in the protocol).

All the analyzed studies come (or were “confirmed”) from the electronic database searches, are published (as refereed journals or items from established working papers series), and are available in English. “Confirmed” studies are those that had been “unsystematically” identified in the early stages of this project but were nonetheless picked up by the database searches.
While the majority of studies focus specifically on contract enforcement and investment, some evidence valuable to answer the research question comes (and will come) from studies that have slightly different foci.

Cross-country evidence seems to dominate the relevant body of research. This is reflected in geographic scope of the studies as well as unit of analysis (which mirrors in numbers but doesn’t coincide mechanically with the former).

The prevailing research designs are cross-section and panel data regressions, with estimation methods customized to the specificities of data sets, variables, and/or reflecting trends in the econometric “best practice”. One study (Clague et al., 1999) also uses narratives of seven country stories (cases) to support its claims, but it also uses cross-country regression as its main analytical tool.
materials and finished goods. Our selected studies show a variety of approaches to measurement or classification of variation in the quality of contract enforcement, which is characteristic of the broader literature on institutions and economic development.

**Table 3.5: Type of investment**

<table>
<thead>
<tr>
<th>Type</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed assets in general</td>
<td>11</td>
</tr>
<tr>
<td>R&amp;D</td>
<td>1</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 3.6: Measure of contract enforcement**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert judgments-ordinal</td>
<td>3</td>
</tr>
<tr>
<td>Business surveys-ordinal</td>
<td>3</td>
</tr>
<tr>
<td>Business surveys-other</td>
<td>2</td>
</tr>
<tr>
<td>Contract Intensive Money</td>
<td>3</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
</tr>
</tbody>
</table>
4. Synthesis methodology

4.1. Overview

This section describes the methodology used to synthesize the information contained on the studies selected on the mapping stage. In particular, we describe the methods used to describe the studies, to assess their quality, to make a synthesis of their findings and to derive conclusions and implications.

4.2. Detailed description of studies in the synthesis

For the group of studies attempting to estimate a quantitative measure of “effect”, we tabulate and examine the following information:

i. information on the study setting, main characteristics of the indicator to measure the independent variable (or “intervention”), the study methods, the units of analysis (countries, sub-national jurisdictions, firms, etc.), and the study findings (narrative summary of findings).

ii. results of the critical appraisal for each study.

iii. Weight of Evidence ratings for each study (individual components A to C, and aggregate rating D).

Vote counting of studies is used to describe the findings. Results of vote-counting are qualified taking into account the critical appraisal of the studies and the heterogeneity of the research designs.

We use Weight of Evidence rating to summarize the reviewer’s critical appraisal and to examine robustness of findings. We explore if the results of the synthesis change in a meaningful way if we progressively include studies from top to bottom of this scale (particularly for components A and summary rating D of the WoE ratings).

Because all the studies that quantify the impact fall in the lowest category of component B of the WoE rating (due to their non experimental nature) we use our specific evaluation of methodological quality to carry out some sensitivity analysis.
The review team used a critical appraisal approach based on a multi-dimensional concept of quality in research. This approach covers quality of reporting, methodological rigour, conceptual depth and breadth, and relevance.

The quality assessment is done with two purposes: first, to exclude studies that clearly do not meet minimum professional/academic standards (e.g., omit sources of data, report mere opinions, clearly choose an inadequate econometric method); second, to generate quality ratings to qualify synthesis results.

With regard to the quantitative studies, we consider useful to clarify some criteria in evaluating quality.

First, it is important to note that the macro social and complex nature of the “intervention” to be analyzed (i.e., reforms that affect the quality of contract enforcement), generally prevents the existence of studies based on experimental designs. Judicial reforms are typically deployed simultaneously within a national jurisdiction, and it is hard if not impossible to exclude citizens from their effects. For this reason, all studies included in this systematic review are based on observational data.

Second, we are considering some standard criteria of research quality appraisal in quantitative studies. Studies that use only statistical correlation analysis without any attempt to establish causality will be rated as low quality. Other classic criteria are sample size and design, consideration of omitted variables, treatment of heteroskedasticity, and, when time series are used, proper analysis of stationarity conditions.

We are considering with particular care the analysis of causality and the treatment of endogeneity/exogeneity of independent variables. Even though there is not a “golden standard” to treat these problems with observational data, econometric theory has developed many techniques: instrumental variables estimations, two or three stage ordinary least squares, simultaneous testing of systematic policy reaction functions (implicitly -e.g., reduced form specifications, in vector autoregressions- or explicitly -structural specification, etc.), Granger-exogeneity tests in time series, some Bayesian techniques, etc. (Stock and Watson, 2006, Chapter 9). When instrumental variables are used, particular care must be taken to judge the adequacy of instruments.

Third, the appropriate use of control or confounding variables, based on the most generally accepted theories of the determinants of investment, has to be considered.

We are aware of the existence of a strong debate coming from the empirical growth literature, about the validity of cross country regressions, and panel data analysis for countries. This debate has extended, in particular, to the question of interaction between growth and institutions (Rodrik 2005). While we are not covering exhaustively this literature in our review, some of its key references are useful to evaluate the strengths and weakness of the research reviewed using regression analysis with similar kind of data (Rodrik 2005; Clemens & Bazzi 2009; Rehme 2007; Deaton 2010).

Following Gough (2007) and the example of Tripney et al (2009), we apply a “weight of evidence” approach to rate quality of studies (see questions 3.18-3.21 of the coding tool). Pairs of review team members filled out those sections of the coding tool. In case of disagreements, discussions were held until consensus was reached.
4.4. Synthesis: quality-assurance process

Each reviewer was assigned a sub-set of the references included in this systematic review, and all the substantive portions of the coding tool (sections 2 to 4) were doubly-reviewed for possible mistakes. In case of disagreement, discussions were held until consensus was reached.

4.5. Synthesis of evidence

Besides synthesizing the findings regarding the direction and size of effects (of contract enforcement on investment), we examined the studies seeking to identify various plausible causal mechanisms, and how they interact with context to determine outcomes.

The reasons why we chose these narrative methods of synthesis are:

- the complex and multiple embedded nature of the social “intervention” to study, i.e., reforms of judiciary and other reforms to improve contract enforcement. We want to understand the contexts and mechanisms that are part of the explanation of the outcomes of a given intervention.

- the interest in evaluating if and when the findings can be generalized to other areas (countries, regions, etc.), times, societies, etc., from a pool of very heterogeneous designs

The small number of relevant studies, and the heterogeneity in study designs, made it impossible to undertake a standard meta-analysis (Petticrew and Roberts, 2006).

Even though some of the information that we are extracting with the coding and extraction tool give us some indication about the contexts and mechanisms implicit in each intervention/study, we needed to dive deeper in each of the selected studies to identify them (that is, go back to full text of key articles, and discuss among three reviewers).
5. Synthesis results

5.1. Overview

This section describes the studies that were selected for synthesis, the quality appraisal and the synthesis itself.

5.2. Characteristics of studies included in the synthesis

As discussed above, the nature of the “innovations” for which “impact” is sought determines the nature and quality of the evidence most readily available, and therefore the inferences that can be drawn and the conclusions that can be backed with such evidence.

In general, the studies selected for synthesis are all statistical analyses of non-experimental data. No study evaluates an “enhancement contract enforcement” reform, neither with an ex-ante experimental design or an ex-post perspective.

In principle, there is no reason why a “natural experiment” couldn’t occur that would allow, at least, a before-after analysis of the behavior of economic agents (i.e., their investment behavior), in an otherwise “constant” environment altered only by some innovation in contract enforcement mechanisms. However, either these “natural experiments” are not frequent enough, or researchers have not taken advantage of them to study the effects of changes in enforceability on investment (analysts of institutions and development have indeed recognized the potential of natural experiments -see, e.g., Pande and Udry, 2005–but some studies that took that approach and came up in our searches did not meet some of the key inclusion criteria).

Studies also vary broadly in terms of the units of analysis, thus making it harder to draw robust conclusions applicable with confidence to some of the possible levels of analysis. Eight out of the fourteen studies have “countries” as the unit of analysis, followed by studies that look at micro data for firms (five) and just one that exploits variations in institutional performance among subnational jurisdictions within a single country (a second study uses variations among cities in a sample of firms, but the latter remain the units of analysis). This echoes the concerns of authors like Pande and Udry (2005) that too much emphasis has been placed by the “institutionalist” approach in analyzing cross-country differences, producing diminishing returns in terms of new insights, while other sources of institutional variation remain under-studied.
The following box contains the full references for the studies synthesized here.

**Box 1: Studies selected for synthesis**

<table>
<thead>
<tr>
<th>Quantitative studies</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Qualitative study</th>
</tr>
</thead>
</table>

The clustering of, and interrelations among the selected studies can be seen from various angles. The quantitative studies include two analyses (by overlapping groups of co-authors) of the same firm-level survey (items 6 and 9), and members of the same team are the authors of the single qualitative case study that met the eligibility criteria. These studies share a problematic issue, as they measure breaches of contracts (in the form of delayed payments in contracts between farmers and food processor companies) based on businessmen’s subjective rating of the “seriousness” of certain events as obstacles to the growth of business. On the other hand, at least one (study number 9) is among the most sophisticated in terms of the analytical
model, allowing for the indirect influence of contract enforcement on investment through the frequency (subjectively assessed) of the breach of contracts.

A separate set of three studies use the same “constructed” measure of quality of contract enforcement, including the originator of the indicator (study 3, which informed the work of 10 and 12). They acknowledge and cite study 11 (with study 2, probably the earliest studies in our group to have been written), and present themselves as avoiding some of the problems with its approach. These studies measure quality of enforcement indirectly through the concept of “contract intensive money” (that is, the ratio of noncurrency money to the total money supply), arguing that “the same governmental deficiencies that require self-enforcement of transactions also lead economic actors to prefer currency” (Clague et al., 1999, p. 188).

This is perhaps the least direct measure of contract enforcement used in the selected studies. The vast majority (ten of the fourteen studies) follow the pioneer work represented by 2 and 11 and use ratings of the quality of enforcement or rule of law based in some way on the views of experts and/or businessmen. These include studies exploiting data generated by World Bank-led efforts to understand firm behavior and the effects of the business environment (2, 7, 8).

Only one of the included studies (study 14) uses some “objective” measure of enforceability: Kohling (2000) takes into account backlog of cases awaiting review and the decisions made within a year at high courts of states or unions in India (as an indicator of the speed of law enforcement), as well as the ratio of appeals dismissed by the Supreme Court (as an indicator of predictability of enforcement). This is also the only study to exploit variations in quality of enforcement among subnational jurisdictions.

Perhaps less problematically, the studies also vary in terms of the measure of investment used. Ten studies measure investment as the rate of a period’s flow over GDP (2, 3, 7, 10, 12) or over the stock of capital (6, 8, 9, 11, 13). One study measures the share of profits that is reinvested (5), another the effect of contract enforcement on overall stock of capital (14) and a third on the level of investment in fixed assets (1). One study is in a class of its own as it focuses on the effects on the flow of investment in the intangible R&D specifically (4). The choice of measures is not unrelated to the choice of unit of analysis and methods (for example, the contract intensive money indicator would not make sense at the firm level, or would not have any simple and relevant interpretation anyway). Notably, all included studies examined considered domestic investment of some kind as their outcome variable, and none focused on foreign investment.

Generally, and even in a broader context of economic research that tends to use mostly non-experimental data, the fourteen studies are notable for the heavy reliance of cross-section analyses (seven of the fourteen studies are based on cross-section regressions). To the limitations of ex-post statistical analysis, these add the complications posed to any attempt to establish sequence or precedence in variations, that might be claimed to address some exogeneity issues.

In some cases (e.g., study 2), single measures for different time-periods of each explanatory or control variable imperfectly substitute for the possibility of estimating dynamically more appropriate models but, to illustrate the ambiguities involved, we found the authors’ interpretation of “causality” channels quite problematic. In other cases (e.g., study 5 uses subjective qualitative indicators from firm surveys that are “time invariant”, and quantitative data that reflects firms’ economic conditions in
the three previous years), “recall” data is collected from firms, with its potential for bias.

There is no straightforward correlation between form of quantitative analysis and the basic unit of analysis (e.g., the panel data approach is more or less evenly distributed among studies that have countries and firms as unit of analysis, and also includes the sole study based on subnational jurisdictions).

5.2.1 Summaries of the selected studies

Cungu et al (2008) examine evidence from a survey of farmers in Hungary to show that the severity of contract hold-ups (i.e., payment delays) deters investment in a non-linear manner. The previous study by two of the co-authors (Gow and Swinnen, 2002) had explored the same data with a more elaborate econometric strategy that seemed to make better justice to the indirect effect of enforcement on investment, but one could think that it did not satisfy journal reviewers, as a simpler but less easily interpretable specification was the one published in 2008. The “qualitative” study retained (Gow et al, 2000) serves to show the group’s early interest in the “hold-up” problem (in the form of long payment delays by food processors with significant market power), after economic reforms in former communist countries weakened contract enforcement and created disorganized markets with significant power asymmetries. The authors show that greater contract “self-enforcement” (promoted by a big player in an agrifood value chain) facilitates investment (particularly but not only in relationship-specific assets). In brief, this cluster of studies provides firm-level evidence consistent with a causal chain from contract enforcement, through prevalence of hold-up problems, to investment levels.

After developing the “indirect” but “objective” contract enforcement indicator based on cash as a proportion of total money (that is, the proportion of Contract-Intensive Money or CIM), Clague et al. (1999) examine the recent history of 7 countries (discussing some crisis and reform episodes), and run some cross country regressions to show a positive relation between quality of enforcement (indirectly measured by CIM) and investment. Prados de la Escosura and Sanz-Villarroy (2009) use the same indicator and a structural equations, time series approach to show a similar effect for the case of Argentina. After estimating a system of equations with a “simingly unrelated regression” methodology, the authors conclude from the investment equation that

“the larger the increase in CIM, the higher the investment rate. In other words, (…) one standard deviation increase in the level of CIM five periods earlier would raise the rate of investment by 2.6%. Thus, the improvement in contract enforcement, captured by CIM levels, appears as a major determinant of capital formation in Argentina.” (Prados and Sanz, 2009, p. 13).

The CIM indicator also shows up as a significant explanatory factor (with the expected sign) in the panel data investment functions estimated by Le (2004), which include a few other economic and political determinants. The variance of CIM, however, does not seem to generate the expected effect (i.e., the published tables show it with a positive coefficient in the investment equations, while it was supposed to reduce investment).
Brunetti et al (1998), with data from a private sector survey conducted in 73 countries and covering more than 3,800 enterprises, estimate standard cross-country growth and investment regressions that include institutional quality indicators. In particular, predictability of judiciary enforcement is significant and positively related with investment performance. With similar data, Dao (2005) shows that courts constraint, as measured by the share of senior managers in World Bank surveys that ranked “courts and dispute resolution systems” as a major or very severe constraint, linearly affect the share of gross capital formation in the GDP of a developing country (based on a cross-section analysis). We interpret the results of Honorati, M. and T. Mengistae (2007) as showing that weaker contract enforcement (reflected in greater exposure to corruption) reinforces the already negative effect on firms’ willingness to invest that is caused by inadequate labour regulations, power outages and financial constraints in India. These results must be taken with great caution as they are based on the reported significance of institutional malfunctioning, rather than on its observed manifestations. The possibility of politically-related bias (that is., a tendency to see all black or white, depending on one’s overall assessment of the institutional and policy context) is very real when using this type of data without the possibility of triangulation with more “objective” indicators.

Levine (1998) was the early predecessor to many of these studies. Consistently with his longer term program, however, he focuses on the determinants of financial depth. The identified study is relevant to this systematic review because it is the only one that tests the frequently assumed causal chain from contract enforcement, through financial depth, to investment (the paper is also unique in the series co-authored by Levine and collaborators in taking the analysis all the way to investment outcomes, rather than stopping at indicators of financial depth or sophistication).


Clarke, G. (2001) finds evidence of a negative relation of risk of expropriation and (positive relation) of the rule of law (plus other explanatory variables) to R&D expenditures, using cross-country data. Banerjee et al (2006) find that, controlling for relevant factors, better contract enforcement is associated with greater private investment in infrastructure (they have a panel data structure but estimate their results using pooled OLS).

5.3. Weight of Evidence

The double-coding and discussion of disagreements between reviewers yielded a judgement about the quality of reporting in each study, which we summarized using the “weight of evidence” framework (details of the methods are explained in Gough, 2007, and illustrated, for example, in Tripney et al, 2009). The WoE judgements for all the studies that reached the synthesis stage are shown in Table 5.1.

For reasons discussed before, all studies were rated “low” in dimension B, but this was not the only reason why no studies reached the “medium” category overall (the highest rating in dimension D was “low/medium”. A lot to do with these ratings has
the fact that most studies fail to examine and control for alternative explanations that could account for the observed partial correlations. In the overall rating, the studies are split in halves among those that have a “low” and those having a “medium-low”. 

Table 5.1

<table>
<thead>
<tr>
<th></th>
<th>WoE A</th>
<th>WoE B</th>
<th>WoE C</th>
<th>WoE D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>medium</td>
<td>low</td>
<td>medium</td>
<td>low/medium</td>
</tr>
<tr>
<td>2</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>3</td>
<td>medium</td>
<td>low</td>
<td>medium</td>
<td>low/medium</td>
</tr>
<tr>
<td>4</td>
<td>medium</td>
<td>low</td>
<td>high</td>
<td>low/medium</td>
</tr>
<tr>
<td>5</td>
<td>medium</td>
<td>low</td>
<td>medium</td>
<td>low/medium</td>
</tr>
<tr>
<td>6</td>
<td>low</td>
<td>low</td>
<td>medium</td>
<td>low</td>
</tr>
<tr>
<td>7</td>
<td>low</td>
<td>low</td>
<td>medium</td>
<td>low</td>
</tr>
<tr>
<td>8</td>
<td>low</td>
<td>low</td>
<td>low</td>
<td>low</td>
</tr>
<tr>
<td>9</td>
<td>medium</td>
<td>low</td>
<td>medium</td>
<td>low/medium</td>
</tr>
<tr>
<td>10</td>
<td>low</td>
<td>low</td>
<td>medium</td>
<td>low</td>
</tr>
<tr>
<td>11</td>
<td>medium</td>
<td>low</td>
<td>high</td>
<td>low/medium</td>
</tr>
<tr>
<td>12</td>
<td>low</td>
<td>low</td>
<td>medium</td>
<td>low</td>
</tr>
<tr>
<td>13</td>
<td>medium</td>
<td>low</td>
<td>medium</td>
<td>low/medium</td>
</tr>
<tr>
<td>14</td>
<td>low</td>
<td>low</td>
<td>medium</td>
<td>low</td>
</tr>
</tbody>
</table>
5.4. Synthesis of evidence

We used an additional indicator to record our assessment of the extent to which findings and conclusions related to the studies’ questions, and the strength of the reported conclusions. The five point ordinal scale (borrowed from other studies) is:

1. Not clear conclusions
2. Results ambiguous but a trend
3. Conclusions based on results
4. Results clear and likely to be true
5. Results are unequivocal

Table 5.2 presents the rating for the fourteen quantitative studies. No studies reached the highest rating (5). Even when we opted for the highest when two ratings could be justified, only two studies appeared analytically systematic and empirically thorough enough to merit the fourth point in the scale.

<table>
<thead>
<tr>
<th></th>
<th>Strength of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Clarke, G. (2001)</td>
</tr>
<tr>
<td>7</td>
<td>Dao, M. (2005)</td>
</tr>
<tr>
<td>10</td>
<td>Le, Q. (2004)</td>
</tr>
</tbody>
</table>

Overall, the evidence gathered through this systematic review provides weak support for the claim that more effective contract enforcement promotes higher levels of investment. First, there is no study that unambiguously links an intervention or reform to enhance contract enforcement to changes in investment patterns. Second, few of the studies go beyond a generic discussion of direct and indirect effects to actually test the plausible indirect causal channels. Third, almost all the studies do very little or nothing in terms of sensitivity checks, or the strenuous but necessary attempts to rule out alternative explanations for the empirical findings.
6. Conclusions and implications

The Systematic Review reveals that the evidence on the impact of improvements in contract enforcement on investment is spotty, comes from a rather disjoint body of literature, and does not meet the usual standards to back causal inference. Some of the hypothesized mechanisms seem consistent with available evidence (for example, the causal channel through financial development, or through provisions required for mitigating the costs of breach of contract and the effect on the value of investment projects). However, much remains to be done in terms of research to back policy strategies. It is unfortunate that some of the most promising data (such as, e.g., the “objective” indicators of quality of enforcement available for subnational jurisdictions in India) have not been more systematically analyzed, to yield more robust statistic results. In other cases, the analytical approach has been sound but the available indicator of quality of enforcement or breach of contracts has questionable validity.

While it seems possible to enhance the overall robustness of the empirical evidence by further statistical analysis of “panel of countries” (and the development of new indicators of institutional quality will continue to feed that “industry”), panels of subnational jurisdictions and/or firms would seem to hold more promise. The identification of “natural experiments”, and the development of the appropriate indicators to analyze them retrospectively would also seem as a neglected strategy worth exploring.

For policy makers and donor agencies, it seems that too much confidence has been put on plausible but unproven causal arguments. Improving contract enforcement may be a valuable objective in itself, as it would seem as a contributor and constituent of the “rule of law” more broadly. Quite another thing is to advocate institutional reforms that will demand investment of political capital and other resources, to improve the expediency, predictability, or fairness of judicial rulings, in the name of economic gains that are not yet proven in the scholarly literature. If evidence is to guide policy, policy makers and donors should continue investing resources in researching these causal links, perhaps with a more “Popperian” approach of trying to falsify the hypotheses of “conventional wisdom Institutionalism”, and staying with those that resist falsification rather than promising benefits based on (very weak) “confirmations”.
References


White, S. 2008. “A conceptual framework to guide research on private sector development in developing countries”, IDRC Working Papers on Globalization, Growth and Poverty, Number 6, Ottawa, Canada: IDRC.


Appendix 1

Coding tool - Systematic Review on The Impact of Contract Enforcement on Investment

0 Unique identifier code
1 General descriptive information
   1.1 Type of publication
       Institution/Government Report
       Refereed journal
       Non refereed journal
       Book Chapter
       Working paper series
       Conference paper
       Other
       Unknown
   1.2 Publication status
       Published
       Forthcoming
       Unpublished
       Unknown
   1.3 Source
       Electronic database
       Handsearch
       Citation
       Website
       Unknown
   1.4 Language
       English
       French
       Portuguese
       Spanish

2 Study Aims and Rationale
| 2.1 Topic focus/foci of the study (tick as many as necessary) | Contract enforcement and investment |
| - | Contract enforcement and economic growth |
| - | Institutions and investment |
| - | Institutions and economic growth |
| - | Finance and growth |
| - | Finance and investment |
| - | Contract enforcement and international trade |
| - | Contract enforcement and finance |
| - | Contract enforcement and vertical or horizontal integration |
| - | Contract enforcement and substitutes or complements (social capital, trust, etc.) |
| - | Reform of the judiciary |
| - | Investment climate reform |
| - | Private sector development |
| - | Formal and informal enforcement |
| - | Other (specify) |

| 2.2 Is this report linked to one or more other reports (included in the searching results) in such a way that they also report the same study? | Linked (specify) |
| - | Not linked |

| 2.3 When was the study carried out? If the authors give a year, or range of years, then put that in. If not, give a 'not later than' date by looking for a date of first submission to the journal, or for clues like the publication dates of other reports from the study. | Explicitly stated (please specify) |
| - | Implicit (please specify) |
| - | Not stated/unclear (please specify) |

| 2.4 Geographic scope | Individual country (specify) |
| - | Group of countries (specify) |
| - | Individual region inside a country (specify) |
| - | Group of regions inside a country (specify) |
2.5 Research design
Quantitative studies
- Cross section regression
- Panel data regression
- Time series regression
- Others (specify)
Qualitative studies
- Historical
- Narrative
- Observation
- Survey
- Audit
- Action-based
- Case series
- Expert opinion
- Focus group
- Other (specify)

2.6 Type of investment
Fixed assets in general
Machinery and equipment
R&D expenditure
Infrastructure
Other (specify)
Not stated

2.7 Origin of investment
Domestic
Foreign
- General
- Group of countries (specify)
- Country (specify)

2.8 Unit(s) of analysis
Country(ies).
Sub-national jurisdiction(s)
Industry(ies).
Firm(s) (any size and legal status)
Households or individuals
### 3 CRITICAL APPRAISAL QUESTIONS

#### Quality of study - reporting

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes/No (Specify)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Is the context of the study adequately described?</td>
<td>Yes (please specify)</td>
</tr>
<tr>
<td></td>
<td>No (please specify)</td>
</tr>
<tr>
<td>3.2 Are the aims of the study clearly reported?</td>
<td>Yes (please specify)</td>
</tr>
<tr>
<td></td>
<td>No (please specify)</td>
</tr>
<tr>
<td>3.3 Is there an adequate description of the sample used in the study and how the sample was identified?</td>
<td>Yes (please specify)</td>
</tr>
<tr>
<td></td>
<td>No (please specify)</td>
</tr>
<tr>
<td>3.4 Is there an adequate description of the methods used in the study to collect data?</td>
<td>Yes (please specify)</td>
</tr>
<tr>
<td></td>
<td>No (please specify)</td>
</tr>
<tr>
<td>3.5 Is there an adequate description of the methods of data analysis?</td>
<td>Yes (please specify)</td>
</tr>
<tr>
<td></td>
<td>No (please specify)</td>
</tr>
<tr>
<td>3.6 Do the authors avoid selective reporting bias?</td>
<td>Yes (please specify)</td>
</tr>
<tr>
<td></td>
<td>No (please specify)</td>
</tr>
</tbody>
</table>

#### Quality of the study - methods

<table>
<thead>
<tr>
<th>Question</th>
<th>Yes/No (Specify)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.7 Are there ethical concerns about the way the study was done?</td>
<td>Yes, some concerns (please specify)</td>
</tr>
<tr>
<td>Consider consent, funding, privacy, etc.</td>
<td>No (please specify)</td>
</tr>
<tr>
<td>3.8 Were potential users of the research appropriately involved in the design or conduct of the study?</td>
<td>Yes (please specify)</td>
</tr>
<tr>
<td></td>
<td>No (please specify)</td>
</tr>
<tr>
<td>3.9 Was the choice of research design appropriate for addressing?</td>
<td>Yes (please specify)</td>
</tr>
<tr>
<td>Question</td>
<td>Rating</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>3.10 Have sufficient attempts been made to establish the repeatability or reliability of data collection methods or tools?</td>
<td>No (please specify)</td>
</tr>
<tr>
<td>3.11 Have sufficient attempts been made to establish the repeatability or reliability of data analysis?</td>
<td>Yes (please specify)</td>
</tr>
<tr>
<td>3.12 To what extent are the research design and methods employed able to rule out any other sources of error/bias which would lead to alternative explanations for the findings of the study?</td>
<td>A lot (please specify)</td>
</tr>
<tr>
<td>3.13 How generalisable are the study results?</td>
<td>Details</td>
</tr>
<tr>
<td>3.14 In light of the above, do the reviewers differ from the authors over the findings or conclusions of the study? Please state what any difference is.</td>
<td>Not applicable (no difference in conclusions) Please state what any difference is.</td>
</tr>
<tr>
<td>3.15 What is the overall quality of the study? (taking into account all the quality assessment issues)</td>
<td>High (quality)</td>
</tr>
<tr>
<td></td>
<td>Medium (quality)</td>
</tr>
<tr>
<td></td>
<td>Low (quality)</td>
</tr>
</tbody>
</table>

**Relevance**

<table>
<thead>
<tr>
<th>Question</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.16 Can the study deliver inferences about the mechanisms at work in the review question?</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>3.17 Can the study deliver inferences about the effectiveness of the intervention/independent variable of the review question?</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>
Weight of evidence

3.18 Weight of evidence A: High trustworthiness
Taking account of all quality assessment issues, can the study findings be trusted in answering the study question(s)? How good is the execution of the study? In some studies it is difficult to distinguish between the findings of the study and the conclusions. In those cases, please code the trustworthiness of these combined results/conclusions.

3.19 Weight of evidence B: High: RCTs
Appropriateness of research design for allowing causal inference Medium: non-randomly allocated prospective evaluations
Low: all other study designs

3.20 Weight of evidence C: High
Relevance of particular focus of the study for addressing the question, or sub-questions, of this specific systematic review
Medium
Low

3.21 Weight of evidence D: High
Overall weight of evidence. Taking into account quality of execution, appropriateness of design and relevance of focus, what is the overall weight of evidence this study provides to answer the question of this specific systematic review?
Medium
Low

4 Results and conclusions

4.1 Outcome: impact Quantitative studies
evaluation

- Average impact coefficient (preferred specifications)
- Standard deviation of average impact
- Significance
- Upper limit of the 95% interval of confidence
- Lower limit of the 95% interval of confidence

Qualitative studies
- Strong impact
- Low impact
- No impact
- Not applied

4.2 Mechanism: what and how channels of intervention work to enhance the quality of contract enforcement
- Contract enforcement and finance (describe)
- Contract enforcement and international trade (describe)
- Contract enforcement and vertical or horizontal integration (describe)
- Other (specify and describe)

4.3 Mechanism: what and how channels of intervention do NOT work to enhance the quality of contract enforcement
- Contract enforcement and finance (describe)
- Contract enforcement and international trade (describe)
- Contract enforcement and vertical or horizontal integration (describe)
- Other (specify and describe)

4.4 Measurement of quality of contract enforcement used
- Expert opinion
  - Ordinal
  - Cost and time of judicial procedures
  - Other (specify)

- Business survey
  - Ordinal
  - Cost and time of judicial procedures
  - Other (specify)

- Contract intensity money (CIM)
- Informal enforcement
- Others (specify)

4.5 Measurement of investment
- Level
- Rate (investment/GDP)
Rate (flux/stock)

Stock (i.e. accumulated investment; level of asset)

Others (specify)

4.6 Do the author(s) refer to other previous studies and refute or confirm their results?

No

Not clear

Yes, to confirm (included in our review; specify)

Yes, to confirm (not included in our review)

Yes, to refute (included in our review; specify)

Yes, to refute (not included in our review)

Yes, ambiguous results (included in our review; specify)

Yes, ambiguous (not included in our review)

4.7 Strength of findings

No clear conclusions can be drawn. Not significant

Results ambiguous, but there appears to be a trend.

Conclusions can probably be based on the results.

Results are clear and very likely to be true.

Results are unequivocal.