The Whys of Social Exclusion: Insights from Behavioral Economics
Karla Hoff and James Walsh

Generalization in the Tropics – Development Policy, Randomized Controlled Trials, and External Validity
Jörg Peters, Jörg Langbein, and Gareth Roberts

Privatization in Developing Countries: What Are the Lessons of Recent Experience?
Saul Estrin and Adeline Pelletier

Public-Private Partnerships in Developing Countries: The Emerging Evidence-based Critique
James Leigland
All over the world, people are prevented from participating fully in society through mechanisms that go beyond the structural and institutional barriers that rational choice theory identifies—poverty, exclusion by law or force, taste-based or statistical discrimination, and externalities from social networks differentiated by socioeconomic status. This paper discusses four additional mechanisms that can be explained by bounded rationality: (a) implicit discrimination, (b) self-stereotyping and self-censorship, (c) rules of thumb adapted to disadvantaged environments that are dysfunctional in more privileged settings, and (d) “adaptive preferences,” in which an excluded group comes to view its exclusion as natural. Institutions, if they are stable, come to have cognitive foundations—concepts, categories, social identities, and worldviews—through which people mediate their perceptions of themselves and the world around them. Abolishing or reforming a discriminatory institution may have little effect on the social categories it created; groups previously discriminated against by law may remain excluded through custom and habits of the mind. Recognizing new forces of social exclusion, behavioral economics identifies ways to offset them. Some interventions have had very consequential impacts.

As the precision of economics has increased through field, lab, and as-if-random “natural” experiments, researchers have uncovered socio-psychological barriers to upward mobility. In India, low-caste boys solved mazes just as well as high-caste boys when their caste was not publicly revealed, but solved 23 percent fewer mazes than high-caste boys when caste identity was revealed in mixed-caste groups (Hoff and Pandey 2014). In France, grocery store clerks of African origin were 9 percent more
productive than other clerks except on days when they were supervised by managers implicitly biased against minorities; on these days, they were of average productivity (Glover, Pallais, and Pariente 2017). A belief that a race, gender, caste, or other ascriptive group is inferior can affect how others treat members of the group and how members of the group feel about themselves, creating productivity differences that sustain the beliefs, although no inherent productivity differences exist (Hoff and Stiglitz 2010; World Bank 2015).

In the last 30 years, economics has taken a cognitive turn. Economists and psychologists have made breakthroughs in understanding how people make decisions, and a new field has emerged—behavioral economics. Camerer (2005) and Hoff and Stiglitz (2016) distinguish two strands. With insights from psychology, Strand 1 views the individual as a quasi-rational actor: he thinks clearly under ideal conditions, but under most real-world conditions his judgment and behavior are affected by seemingly irrelevant contextual factors in the moment of decision (Thaler 2016). With insights from social psychology, sociology, and anthropology, Strand 2 views the individual as a quasi-rational, enculturated actor: experience and exposure to social patterns have persistent effects on judgment and behavior by shaping the cognitive toolkit with which information is processed. The toolkit includes categories, concepts, causal narratives, and other mental models (or, equivalently, schemas; Douglas 1966, 1986; Bruner 1991; D’Andrade 1995; Strauss and Quinn 1997; Bicchieri 2006). An individual may have multiple, inconsistent mental models to interpret a situation; and cues in the environment will influence which one is activated (DiMaggio 1997, p. 275). There is a two-way relationship between individuals and institutions. Individuals create institutions, but institutions shape the mental models of individuals and what primes particular behaviors: “In an ongoing cycle of mutual constitution, people are socioculturally shaped shapers of their environment” (Markus and Kitayama 2010).

The purpose of this essay is to provide a perspective from behavioral economics on the forces that maintain social exclusion and on interventions to offset them. The essay is divided into five sections. We first provide evidence that institutions influence how people think. Then we discuss four mechanisms through which an institution that excludes an ascriptive group can have persistent effects long after the institution is reformed or abolished. The mechanisms for this exclusion are (a) implicit discrimination, (b) self-stereotyping and self-censorship, (c) the rules of thumb of individuals who try to live in two worlds—a narrow and insecure world of disadvantaged groups and an orderly world of school or work, and (d) “adaptive preferences,” in which the excluded group comes to view the exclusion as natural. We discuss interventions to offset each mechanism, in some cases with hugely consequential impact.

The “Schematizing Power of Institutions”

A common definition of institutions is “rules of the game.” In traditional economics, the rules (in particular, property rights) affect only the opportunity sets over which
people optimize; institutions have no impact on the way individuals think. In contrast, all other social sciences recognize that institutions have a “schematizing power”; they shape the knowledge structures (mental models or schemas) that an individual uses to process information—what he attends to, what he perceives, and how he interprets ambiguous signals. Bruner (1990) argues that the schematizing power of institutions creates and sustains social identities: the symbolic systems make the user a reflection of the community by influencing how others see him and how he sees himself.

Social groups that differ in their experiences and exposure have different mental models and behave in systematically different ways in the same situation (e.g., Henrich et al. 2001; Brooks, Hoff, and Pandey, 2017). Many scholars view shared mental models as a primary manifestation of culture (see Douglas 1986, especially pp. 46–48; Swidler 1986; and the definition in DiMaggio 1997, of “culture as a network of interrelated schemata”). Culture has a constitutive role, not merely a regulatory role. Economists and political scientists increasingly incorporate mental models or rules of thumb (based on a schematic view of a situation) as a variable to explain change or persistence of inequality; for example, Hoff, Fehr, and Kshetramade 2011 (impact of caste identity on altruistic punishment in India to protect an ingroup); Alesina, Giuliano, and Nunn 2013 (impact of plough cultivation in pre-industrial agriculture on modern gender roles); Acharya, Blackwell, and Sen 2016 (impact of the level of historical dependence on slave labor on contemporary racism in the United States); Bedolla and Miachelson 2012 and Carpenter and Foos 2017 (impact of “learned disengagement” of marginalized U.S. citizens on the response to get-out-the-vote activities).

Categories, one kind of mental model, lay the foundation for social stratification. The psychologist Gordon Allport (1950) argued that “[t]he human mind must think with the aid of categories. Once formed, categories are the basis for normal prejudgment. We cannot possibly avoid this process. Orderly living depends on it.” Institutions that create hierarchies of ascriptive groups (e.g., by race, gender, or ethnicity) impair the ability of others to learn things about the person that do not fit the category, since mental models filter information in a way that tends to preserve categorical beliefs. Psychologists are beginning to understand the neural basis of categorization and associative learning:

When neurons are consistently activated by co-occurring features of experience, physical changes in the neurons strengthen the connections between and among them... Thereafter, if one of those neurons is activated, it will be more likely to activate another in that group ... Growing up in an environment of a given cultured shape brings with it a distinctive pattern of experiences and corresponding neural changes.... The synaptic changes... cannot be erased like sentences from a text... Change in the world can lead to a new pattern of strong neural connections, but it does not completely destroy earlier learning (Strauss and Quinn 1997, 90; emphasis added).
A simple experiment by Bruner and Potter (1964) in visual identification provides suggestive evidence that categorization leads individuals to resist disconfirming evidence. Subjects were shown eight photographs of ordinary objects (e.g., a dog standing on grass), projected one at a time, which were gradually brought into focus. One group of subjects saw the photos starting from almost complete blur. Another group of subjects saw the photos starting from medium blur. And a third group saw the photos starting from light blur. For all groups, the picture being exposed was stopped at the same point of focus, regardless of its starting point. At this common terminal point, the subject was asked what the object was. The surprising finding was that subjects who had seen the longer video, starting at greater blur, were less likely to identify the object correctly; that is, despite more exposure, they learned less. Slightly less than one-fourth of the subjects recognized pictures when they began their viewing with a very blurred image, but more than half recognized them when viewing began with light blur. Bruner and Potter (1964) suggest that people have difficulty rejecting mental representations that they have constructed. Individuals “hang on” to false hypotheses: “at any particular clarity of the display, those who see it for the first time are more likely to recognize the object than those who started viewing at a less clear stage.” This is called an interference phenomenon.

Kahan et al. (2017) demonstrate a related finding in the political domain. The authors find that when people process scientific data that conflicts with their ingrained worldviews (e.g., that gun control increases crime), they often misinterpret the data. The interference problem is more severe, the more numerate they are. This indicates that the problem is not inadequate mathematical skill. The authors suggest that mathematical skill may actually enhance the ability to filter out unwanted information.

People may be capable of suppressing their biases in clear-cut cases but incapable of doing so in situations of ambiguity. The psychologists John Darley and Paget Gross (1983) investigated how the social class of a student influences others’ judgments of how well she is doing in school. The experimental subjects were randomly allocated to one of four groups. Group 1 saw a video of a nine-year old girl, called Hannah, in a low-income neighborhood and were informed that her parents had only a high school education. Group 2 saw a video of Hannah in a high-income neighborhood and were informed that her parents were college-educated. Groups 3 and 4 had the same information as groups 1 and 2, respectively, but in addition viewed a videotape depicting Hannah taking an oral test. There was only one version of the videotape. It depicted Hannah’s performance as inconsistent—she answered some challenging questions correctly and some easy questions incorrectly.

Traditional economics would predict that additional information could only increase the precision of participants’ assessments and narrow the gap in assessments between them, but the opposite was true. The first two groups, who had information only about Hannah’s socioeconomic background, differed very little in their assessment of how well she was doing in school. In contrast, groups 3 and 4
differed significantly in their assessments of how well Hannah was doing in school. The “rich” Hannah was judged to be more able than the “poor” Hannah. Expecting her to do better, viewers of the “rich” Hannah compared to viewers of the “poor” Hannah evaluated her performance in the oral test more favorably. The results supported the hypothesis that mental models play a role in information-processing that is distinct from their role in pre-judgment.

Alesina, Guiliano, and Nunn (2013) provide an example of the persistent effect, throughout the world, of the meanings that an historical institution gave to gender. Pre-industrial agriculture used either shifting cultivation or plough cultivation. Unlike the hoe or digging stick used to prepare the soil in shifting cultivation, the plough requires significant strength—either to pull it or to control the animal that pulls it. In areas topographically well-suited to crops for which plough agriculture is efficient (wheat, barley, and rye), men had an advantage in farming relative to women, and adoption of the plough created gendered occupations—men in the field, women at home—that have influence in modern times. In such areas, female labor force participation in the year 2000 was more than 20 percentage points lower than in other areas. This influence remains as individuals migrate: the gender norms of immigrants’ children who live in the United States and Europe are influenced by whether their ancestors were members of an ethnic group that used plough cultivation in the pre-industrial period.

Mental models that an institution of social exclusion creates can be deliberately strengthened after the institution is abolished in order to make the old social pattern persist. In U.S. southern counties that in 1860 had a high proportion of slaves, whites are more likely today to express racial resentment toward African Americans and to oppose affirmative action, compared to whites who live in otherwise similar areas that had lower population shares of slaves (Acharya, Blackwell, and Sen 2016). In order to hold down agricultural labor costs after slavery was abolished (as well as for social and political reasons), whites in counties that had relied heavily on slave labor reinforced racist norms and racial hostility. This shaped attitudes that were transmitted across generations through culture and institutions, such as Jim Crow. Anti-black attitudes faded earlier in areas with a low historical dependence on slave labor.

Drawing on many other social sciences, twenty-first century behavioral economics (which we have called Strand 2) introduces into economic theory a new variable for processing information—mental models. The new variable is shaped by institutions through experience and exposure, and activates four mechanisms of social exclusion discussed in the remainder of the paper.

**Barrier 1: Implicit discrimination**

Rational choice theory offers two explanations for discrimination: taste-based and statistical. Taste-based discrimination arises when individuals dislike members of a group. Becker (1957) argued that employers who discriminate based on social
identity knowingly incur higher costs. Taste-based discrimination would therefore not survive in perfectly competitive markets.

Statistical discrimination arises when individuals have imperfect information about an individual and assess his expected productivity based on his membership in a group (e.g., a race, social class, or gender) (Arrow 1973). Employers may refuse to hire a member of a group because of a rational determination that the person has lower expected productivity than job applicants who are members of another group.

A third kind of discrimination, left out of traditional economics, is implicit (unconscious) discrimination (Banaji and Greenwald 1995; Greenwald and Krieger 2006). Implicit discrimination differs from explicit discrimination in many ways: its sources, malleability, and effect on behavior. Explicit and implicit discrimination do not emerge from the same socialization process (Dovidio et al. 1997). Explicit discrimination is much easier than implicit discrimination to change (Wilson et al. 2000), which is consistent in that self-reports of explicit racism show a large decline among whites, whereas racially discriminatory behaviors remain common among them (Dovidio and Gaertner 2004). Implicit bias predicts important life outcomes. Nosek et al. (2009) find that cross-country variation in implicit attitudes against women in science predict gender-based achievement gaps in eighth-grade science and math. As we mentioned above and discuss more below, implicit bias by supervisors against minority staff can directly impair their performance.

Beaman et al. (2009) investigated how having women in leadership positions affected the attitudes and the behavior of constituents. The investigators studied the impact of a policy in India in which the government randomly reserved—in one-third of the villages—the position of village council leader for women candidates. The study revealed that the impact was very consequential. The investigators used the Goldberg paradigm, which is a common way to measure bias: they asked subjects in one group to evaluate a taped speech by a man, and asked subjects in another group to evaluate the identical speech made by a woman. In villages that had never had political quotas for women, both male and female respondents gave the male politician higher ratings than the woman for effectiveness. But in villages with political quotas for women for the previous seven years, men evaluated the woman’s speech as just as effective as the man’s. Discrimination by the measure of the Goldberg paradigm had been removed.

Goldin and Rouse (2000) find evidence of gender bias in hiring for symphony orchestras in the United States. Before 1980, none of the five highest-ranked U.S. orchestras had more than 12 percent women. Through the 1970s and 1980s, the share of women hired by the orchestras increased—from about 10 percent in 1970 to about 35 percent in the mid-1990s. During this time, most orchestras introduced screens that hid the identity and gender of applicants from the hiring panel when they auditioned. Using data from audition records, the investigators found that “blind” auditions increased the probability by 50 percent that a woman would advance from
preliminary rounds. The researchers attribute about 30 percent of the gain in the number of female musicians in orchestras to the advent of blind auditions.

Crimes by African Americans are understood differently than crimes by whites. For example, Pager, Western, and Bonikowski (2009) find evidence of the much greater cost to African American job applicants than to white job applicants of having a prison record. The investigators recruited African American and whites to apply for the same set of jobs with similar fictitious résumés, except that one group had résumés with a prison record and the other group did not. The participants applied in person for the jobs. African Americans were only half as likely as equally qualified whites to receive a callback or job offer. Moreover, African American applicants without a criminal record were no more likely to receive callbacks or job offers than white applicants with a criminal record. The study describes the experience of the participants (the “testers”):

In applying at an auto dealership ... testers met with very different reactions [by race]. Joe, the black tester, was informed at the outset that the only available positions were for people with direct auto sales experience. ... When the employer interviewed Keith, their white ex-felon test partner, he gave him a stern lecture regarding his criminal background. The employer warned, “I have no problem with your conviction, it doesn’t bother me. But if I find out money is missing or you’re not clean or not showing up on time I have no problem ending the relationship.” Despite the employer’s concerns, Keith was offered the job on the spot (p. 790).

Salience is a central theme of behavioral economics. If individuals were unboundedly rational, all facts about an individual and a situation would be equally accessible. But for boundedly rational individuals, situations are simplified, and how they are simplified reflects attitudes that can contribute to discrimination. Women are promoted at lower rates than men in science, technology, engineering, and math (STEM) fields. Sarsons (2017) shows that an important explanation for the low promotion rate of women economists is that coauthored research publications by female economists matter less for tenure than coauthored research publications by male economists. It is not known how much of this discrimination is implicit. But a recent University of California, Berkeley senior thesis by Wu (2017) found a way to measure attitudes towards women in economics that suggests the importance of implicit discrimination. In an online, anonymous message board that had more than one million posts, Economics Job Market Rumors, female economists were much more often than male economists perceived in ways unrelated to their professional roles. The three most common words used to describe female economists were “hotter,” “lesbian,” and “bb” (baby). For men, the most common words were “mathematician,” “pricing,” and “adviser.”

Implicit prejudicial attitudes can be self-fulfilling, just as explicit prejudice and discrimination can be (Akerlof 1976). In the empirical study of grocery store cashiers in France mentioned at the beginning of this essay, workers of African origin were
substantially less productive on the days they were supervised by implicitly biased managers, where implicit bias was measured by the Implicit Association Test (Glover, Pallais, Pariente 2017). Bertrand, Chugh, and Mullainathan (2005) argue that IATs tap unconscious attitudes, which activate a different part of the brain than the parts that engage in conscious deliberation. When the employees of African origin worked with unbiased supervisors, they were 9 percent more productive than clerks of non-African origin; but when they worked with biased supervisors, they had only average productivity. The authors find evidence that the productivity decline arose because biased managers avoided interactions with minorities. One reason for the low supervision was that the supervisors were worried that they would be accused of bias if they made an inappropriate remark in their interactions.

A two-stage experiment in Sweden provides additional evidence that implicit bias contributes to discriminatory behavior (Rooth 2010). In the first stage of the experiment, employers received equivalent applications to advertised jobs from applicants with Swedish names and from applicants with Arab-Muslim sounding names. All applicants were represented as male. Rooth contacted the recruiters to ask if they would participate in tests for explicit and implicit attitudes toward Arab Muslims. The tests demonstrated that employers had strong, explicit negative attitudes toward Arab-Muslims, though not on the basis of beliefs regarding lower productivity: A clear majority of employers (77 percent) stated that there were no performance differences between the two groups. Implicit attitudes predicted the difference in callback rates between applicants with Arab-Muslim sounding and Swedish names much more reliably than did explicit attitudes. The probability of a callback for Arab-Muslim job applicants declined by 5 percent for each one standard deviation increase in negative implicit association of Arab-Muslim men.

If prejudice reflects implicit thoughts, not conscious tastes or statistical discrimination, discriminatory beliefs and attitudes may be sensitive to subtle contextual cues. In one experiment, individuals were asked which group they preferred—a group of well-liked African American athletes or a group of disliked white politicians. Respondents preferred the first group when the context emphasized occupation, but preferred the second group when it emphasized race (Mitchell, Nosek, and Banaji 2003).

Context influences the salience and valence of categories. Shayo and Zussman (2011) investigate more than 1,500 judicial decisions in Israeli small claims courts, where cases are randomly assigned to Arab or Jewish judges. These authors find evidence of judicial in-group bias: a claim is 17 percent to 20 percent more likely to be accepted if assigned to a judge of the same ethnicity (Arab or Jewish) as the plaintiff. Consistent with the emphasis in behavioral economics on the effects of salience on attention, the ethnic bias increases with the population-adjusted number of fatalities in the year preceding the ruling that are from Palestinian politically-motivated attacks in the vicinity of the court. Terrorism leads Arab judges to favor Arab
plaintiffs and Jewish judges to favor Jewish plaintiffs. Shayo and Zussman conclude that ethnic conflict, by intensifying ethnic identities, can erode trust in the rule of law: “There is rather little ethnic ingroup bias in the Israeli courts except during periods in which political violence intensifies ethnic identification. In other words, by heightening identification, ethnic conflict can dramatically undermine the proper functioning of an ostensibly impartial institution like the court system” (2011).

**Interventions to Reduce Discrimination**

Identification of the impact of implicit discrimination leads immediately to the question of whether interventions can reduce it. Evidence suggests that information alone may not be enough to change beliefs, since individuals tend to resist corrective information that calls their priors into question (Nyhan and Reifler 2010; the defensive reaction is called *counter-arguing*). However, a longstanding body of work on prejudice reduction emphasizes the beneficial impact of interactions between social groups (Allport 1954; Pettigrew and Tropp 2006). Much of this work argues that prejudice results from a lack of experiential knowledge and understanding of outgroups, and that it can be offset by interactions between the groups. A recent quasi-natural experiment provides strong supportive evidence (Rao 2018). Many private schools in New Delhi lease land from the government at heavily subsidized rates. In exchange, the law requires that the schools admit randomly chosen poor children tuition-free. The law was not enforced until 2007, when a Delhi High Court decision ordered almost 400 private schools to reserve one-fifth of their seats for students from households with incomes of less than about $2,000 per year. The schools would be partially compensated by the government. The order applied only to new admissions of students, who usually enter in pre-school, and required the students to be integrated into the same classrooms as the non-scholarship students.

Rao discovered that exposure for four years to poor classmates and, in particular, regular interaction with them in cooperative groups, changed the social preferences and behavior of the rich students. When offered opportunities to come to school over the weekend to support fundraising efforts for a charity that served disadvantaged children, rich students in cohorts with poor classmates sent 10 percent more volunteers than those without poor classmates. In dictator games, rich students who had poor classmates shared 45 percent more of their endowment, and were much more likely to split their endowment equally with the other player.

To measure the impact on discriminatory behavior by the rich students, Rao devised an experiment involving a relay race, with a prize going to the winner. Before the relay teams were formed, each student ran a sprint individually, which revealed how fast he ran. Then participants selected their teammates in the relay race. As Rao explains, “By having participants choose between more athletic poor students and less athletic rich students, I create a tradeoff between ability and social similarity,”
(2018). In the relay race with the smallest prize (about $0.85), almost one-third of the rich students without poor classmates discriminated against poor students. Those with poor classmates discriminated 12 percentage points less.⁶

Recent research suggests that even a brief interaction between individuals in ingroups and out-groups can sometimes shift tastes and beliefs. Evidence that bears this out is an experiment on attitudes towards transgender people, who face high levels of prejudice and discrimination in the United States. In a randomized field experiment, voters were canvassed by activists for transgender rights. The canvassers, some of whom were transgender, told voters that they might be asked to vote on whether to repeal legal protections for transgender people. Voters were asked to discuss their views and were shown a video presenting the arguments on each side. After the video, the canvassers encouraged voters to engage in “perspective taking”—discussing an occasion when they felt judged negatively for being different and considering whether their own experience offered a window into the experiences of transgender people. The brief interaction had a large effect. As compared with a control group (voters who were canvassed on recycling issues), the intervention increased positive attitudes toward transgender people by 10 percentage points. To put this in perspective, the impact is larger than the increase in positive attitudes toward gay men and lesbians in the United States between 1998 and 2012 (Broockman and Kalla 2016).

There is also evidence that a brief intervention can change behavior in a high-stakes interaction. Disciplinary problems of students are a strong predictor of negative life outcomes; school sanctions early in life can set off a long-lasting negative trajectory (Rocque and Paternoster 2011). In an experiment in five racially diverse middle schools in California, a brief intervention encouraged math teachers to adopt an empathetic instead of a punitive mindset with regard to discipline (Okonofua and others 2016). The teachers were told that the purpose was to review “common but sometimes neglected wisdom about teaching and to collect their perspectives as experienced teachers on how best to handle difficult interactions with students, especially disciplinary encounters.” The teachers read an article about reasons why students misbehave (e.g., social and biological changes during adolescence, worries, stresses, and social anxiety), which discouraged teachers from labeling students as troublemakers. The article encouraged teachers to place value on students’ experiences and to develop and sustain positive relationships with the students. This was reinforced with stories from students. The intervention neither discouraged disciplinary actions nor encouraged the view that students’ perspectives were necessarily reasonable. The intervention reduced suspensions from school. Students of teachers who received the treatment were half as likely to be suspended over the academic year.

An alternative way to change behavior is to increase self-awareness. Publicizing the extent of implicit bias may alert people to discriminatory actions and thereby reduce, or even eliminate, the discrimination. Price and Wolfers (2007) reported that
fouls in NBA games from 1991 to 2002 were more likely to be called against African Americans when the referee team was made up only of whites, and were more likely to be called against whites when the referee team was made up only of African Americans. The New York Times covered the study on its front page. Other newspapers and TV stations, such as ESPN, covered it as well. Six years later, Pope, Price, and Wolfers (2013) showed that the discovery and publicity ended the discriminatory behavior of the referees. The new study compared the period of 2002 to 2007 (before the publication of the earlier study) to the period of 2007 to 2010 (after the publication). In-group bias persisted until 2007, but stopped from 2007 to 2010. The researchers explored whether the change occurred due to a shake-up in the line-up of referees, but found that of the 66 referees who had officiated at least 100 games in from 2003 to 2006, 55 officiated at least 100 games from 2007 to 2010. The researchers also considered whether the NBA changed the racial makeup of the teams, but found that the fraction of mixed-race referee teams actually decreased. The researchers also spoke to the NBA league administrators, who indicated that no policies had changed in response to the paper. Thus, no observed change in structural conditions, but only a change in awareness, seems to explain the end of discriminatory refereeing.

It is sometimes possible to make small changes in a decision-making environment to activate scripts that rely less on stereotypes and thereby reduce discrimination. Bohnet, Geen, and Bazerman (2016) constructed a game in which participants had to hire employees for mathematical or verbal tasks. Participants were paid based on the performance of the person they hired. Participants were given the following information about a candidate: (a) performance on a prior test, (b) gender, (c) his or her identity as an American student from the Boston area, and (d) the average performance of the pool from which the candidate was drawn. Participants could hire either someone offered to them or a randomly selected person from the pool. There were two experimental conditions. In one condition, participants were offered only one candidate. In the other condition, participants were offered a male candidate and a female candidate. All the offered candidates had on a prior test performed at average or slightly below average level of the pool from which they were drawn. IATs show that people have implicit beliefs that men have greater math skills than women, and that women have greater verbal skills than men (Nosek, Banaji, and Greenwald 2002; Plante, Theoret, and Favreau 2009). Consistent with this finding, the participants were more likely to choose the offered male candidate for a math task and the offered female candidate for a verbal task. When evaluated on their own, stereotype-advantaged individuals were chosen 66 percent of the time. However, the stereotype advantage was present only when participants were deciding on a single offered candidate and not present when they had a choice between a man and a woman, whom they could compare to each other. A fine adjustment to the hiring process eliminated gender discrimination.
The examples presented in this section reduced prejudice for reasons that may be heavily contingent on context. The findings may not be parsimoniously applicable to new settings. They may require adaptation to work in new contexts, or they may not work at all. Research in this area is growing rapidly to identify interventions that reduce discrimination.

Barrier 2: Self-stereotyping and Self-censorship

Up to now, we have discussed stereotyping of others. But individuals also stereotype themselves. The stigma of social exclusion can be so profound as to “get into people’s heads” and degrade their self-concept (Goffman 1963).

To function well, people need to understand themselves in terms that sustain their sense of personal integrity and value and that accurately reflect their abilities. Many institutions are designed to support the psychological and social needs of dominant groups by providing role models, narratives, and rituals to embolden individuals to make the effort needed to succeed. Groups at the margins of society, on the other hand, often must expend extraordinary effort to negotiate their place in environments not made for them or hostile to them. “Like a distracting alarm, psychological threat can also consume mental resources that could otherwise be marshaled for better performance and problem solving” (Cohen and Sherman 2014, p. 335). Individuals may spend time assessing whether they belong, whether they are wanted, whether they are good enough, whether they are worthy (Walton and Cohen 2007). For a given situation, two individuals—one in an in-group, the other in an out-group—may be engaging in very different amounts of mental energy.

Even small pressures on mental resources can reduce the ability to self-regulate. A trivial example demonstrates the point. Shiv and Fedorikhin (1999) asked a group of students to memorize a number and recall it a few minutes later in another room. They randomly assigned the students to two groups. One group was given a seven-digit number to memorize. The other group was given a two-digit number. When the students left the room, they had a choice of snack as reward for their participation—a bowl of fruit salad or a piece of chocolate cake. Compared to those asked to memorize the two-digit number, those asked to memorize the seven-digit number were 50 percent more likely to choose the cake, the less healthful choice—63 percent versus 42 percent.

How and when does a negative stereotype become a cognitive tax? The cues in a situation, the individual’s perception of the situation, and the relevance attributed to social identity in that situation all influence whether or not the stereotype is activated (Okamura 1981). For some individuals, a negative stereotype may be chronically activated. Pioneering studies on the effect of priming social identity find that merely checking a box to indicate race before taking an aptitude test lowers the performance of African American students, but not of white students (Steele and Aronson 1995, 1998). The “race prime” appears to raise the consciousness of
negative stereotypes among African Americans. Steele and Aronson (1998) call this effect “stereotype threat”: “Participants who experience stereotype threat spend more time doing fewer items less accurately.”

Stereotype threat has been documented in many contexts. In India, individuals are born into castes, which in each locality are ranked. High-caste individuals are traditionally considered socially and intellectually superior to low-caste individuals (called Scheduled Castes). While discrimination against low caste members is illegal, low-caste children nonetheless encounter the traditional order of caste and untouchability in the fables they learn and often in the continued insults, discrimination, and atrocities against upwardly mobile members of low castes. Hoff and Pandey (2006, 2014) assessed the effect of making caste identity public, and of caste segregation on the performance of junior high school boys in rural north India. Caste segregation is a mark of the civic privileges of the high castes, and the social exclusion and inferiority of the low castes (Jodhka 2002). The participants were asked to solve mazes and paid for each maze they solved. Participants were randomly assigned to one of three conditions: (a) anonymous, (b) caste revealed in mixed-caste groups, and (c) caste revealed in groups segregated by caste status (high or low). In the first condition, three high-caste and three low-caste boys were placed in a session and their identity and caste were not made public in the session. Since, in general, the children came from six different villages, their caste would not be known to the other children in the session. In the second condition, three high-caste and three low-caste boys were placed in a session and their identity and caste were made public at the beginning of the session. The third condition was the same as the second one, except that the six boys in a session were all from high castes or all from low castes.

The anonymous condition showed that low-caste boys solved mazes just as well as high-caste boys. However, publicly revealing caste in mixed-caste groups created a 23 percent caste gap in total mazes solved in favor of the high castes, controlling for other individual variables. A possible explanation is that the boys felt “I can’t or don’t dare to excel.” In the third condition, segregation depressed the performance of both high-caste and low-caste boys. If segregation evokes a sense of entitlement in the high caste, the high-caste boys may have felt, “Why try?”

The experiment to test the effect on maze-solving ability of making a stigmatized identity salient was replicated in Beijing, China, although in this experiment the identity treatment was stronger: in addition to revealing children’s social identity, students completed a pre-experiment survey that asked questions about their social identity and about the characteristics of groups with their own and other social identities (Afridi, Li, and Ren 2015). The subjects were elementary school children aged 8 to 12 drawn from two social categories: (a) households classified as urban Beijing households, a privileged category, and (b) households classified as rural non-Beijing, a disadvantaged category in Beijing. The household registration system in China,
known as *hukou*, classifies citizens based on the birthplace of their parents or grandparents, and favors those categorized as local urban residents in housing, jobs, access to schools, and public benefits. Unlike categories of gender, class, and caste, *hukou* is a transparent man-made creation. The experimental findings show that priming *hukou* shifts performance in ways that mirror the way the groups are ranked. In this sense, the social identity “makes up people,” and a group’s alleged inferiority becomes “an equilibrium fiction” (*Hoff and Stiglitz 2010*).

Most people have difficulty judging their own ability. *Coffman (2014)* implemented an experiment with U.S. college students that reveals that a person’s judgment of his ability in a given domain depends on the interaction between his gender and the gender stereotype associated with the domain. In the experiment, which minimized discrimination and fear of discrimination, female participants under-contributed their ideas in male-typed domains and vice versa; that is, they *self-censored*. If women and men are less likely to contribute their ideas, this will hinder their advancement; it is a self-administered kind of social exclusion. The variable that explains it is the mental model of gender.

A related mechanism that can lead to the replication of inequality after structural and institutional barriers have been removed is coordination on unequally rewarded tasks. In academic departments, individuals have some choice over how much of their time to spend on non-promotable tasks, for example, attending committee meetings, evaluating applicants, and advising undergraduates. There is extensive evidence in academia and industry that women spend more time on non-promotable tasks than men (see references in *Babcock et al. 2017*). While many factors could explain this—for example, gender differences in preferences and abilities—it could also be driven by shared expectations regarding the appropriate behavior of women and men. To investigate this, *Babcock et al. (2017)* ran controlled experiments. They examined how gender affects the allocation between men and women of a relatively poorly paid task. In each of ten rounds, participants—all seated in one large room and each with a computer—were randomly divided into groups of three persons. Members of the group had to make one decision—to volunteer, or not, to be the poorly-rewarded person in the group. Each round could last at most two minutes. To volunteer, a person clicked on his or her screen. As soon as a group member clicked, or two minutes had elapsed, the round ended. The incentives were as follows: each player in a group got $1 if nobody in the group volunteered to be the poorly rewarded person. If somebody volunteered, the volunteer got $1.25 and the other two people in the group each got $2.00.

When the participants in the room were roughly half men and half women, so that individuals knew that their group was very likely to be mixed gender, women volunteered twice as often as men. When the participants in the room were all men or all women, so that individuals knew that their group was single gender, men and women were equally likely to volunteer for the poorly-rewarded task. The results are
consistent with the hypothesis that mental models of gender have a large effect on behavior, and tend to replicate historical inequalities.

In a follow-up experiment, individuals played a similar game in which there was a fourth member of each group. His or her only role was to try to resolve the coordination problem by asking one member to volunteer to be the poorly-rewarded person. Women received more requests than men by a factor of 2.50. The gap increased as the game was repeated over ten rounds. When asked to volunteer, women were 49 percent more likely than men to agree. The results suggest that shared expectations based on traditionally unequal gender roles replicate the inequality when individuals coordinate anonymously in novel situations on unequally rewarded tasks. The traditional gender roles carry over to the novel situations even though there is no structural basis for the carry-over.

Another area in which gender roles affect behavior in ways that replicate historically imposed inequalities is in salary negotiation. Lab and field evidence suggests that men are more likely to ask for higher compensation than women (Small et al. 2007). But seemingly minor situational factors loom large in individuals’ decision-making. In a field experiment, List and Leibbrandt (2014) replicate the result in Small et al.: they find that when salaries are not explicitly made negotiable, women are 23 percent less likely to negotiate for higher salaries; but when it is clear that salaries are negotiable, women are 8 percent more likely than men to negotiate for higher salaries.

Interventions to Weaken the Influence of Negative Stereotypes on Self-assessment

Minor interventions can insulate socially excluded groups from the “threat in the air” created by social stigma. The next five paragraphs discuss experiments with disadvantaged students in the U.S. Then we discuss an intervention in India that reduced the legitimacy of domestic violence.

Experiments in the United States show that interventions that inculcate feelings of belonging can improve academic performance among non-traditional college students (Yeager et al. 2016). In one experiment, disadvantaged high school students who had been admitted to two- and four-year colleges were invited to participate in an online module designed to dispel the belief that disadvantaged students are the only group that has difficulty in college or the only ones who question whether they belong in college. One year later, 45 percent of the students who had participated in the intervention were enrolled full-time in school, compared to 32 percent of the control group. A similar experiment, in which participants had already entered college, reduced the enrollment gap between disadvantaged and advantaged students by 40 percent.

An intervention that has been tested in multiple contexts is a values affirmation exercise. In one experiment, an essay assignment was given to students through their normal coursework two times during the school year (Cohen et al. 2006). The
assignments asked students to think about their most important value. It took 10 to 15 minutes to complete an essay on this topic. Students knew their teachers would read their essays, but the teachers did not know the identity of the students in the treatment. The treatment improved the grade point average of socially excluded groups (African Americans and Latino Americans) in academic courses, and reduced by 50 percent the proportion of African Americans receiving a D or an F in their first term. The effects persisted over the two years in which participants’ grade point averages were tracked. African American students’ grades improved by 0.24 grade points. Among low-achieving African Americans students, performance improved even more: grade point averages increased by 0.41 points, and the rate of remediation or grade repetition was less than a third of that of the control group (Cohen et al. 2009).

Another way to help socially excluded groups improve performance is to frame the idea of intelligence as a *malleable* trait that grows in response to hard work, rather than as a *fixed* trait (Hong et al. 1999; Dweck 2006; Nussbaum and Dweck 2008). Disadvantaged groups may be more likely to believe that intelligence is fixed rather than malleable (Claro, Paunesku, and Dweck 2016). In a seminal study, Aronson, Fried, and Good (2002) tested the impact of fostering a “growth mindset” among African American college students. Students were taught the theory of malleable intelligence in three one-hour lab sessions and were encouraged to explain the ideas in a letter to an at-risk middle school student. The intervention increased the participants’ belief that intelligence was malleable, their enjoyment of academics, and their belief that academics are important; the intervention increased the participants’ semester grade point average from 3.05 to 3.32.

Many interventions to counter the belief that intelligence is a fixed trait have been effective in small or lab settings, but can such interventions be effective at scale and for a heterogeneous population? Recent and ongoing work investigates this. Paunesku et al. (2015) designed and delivered two online 45-minute-videotapes to over 1,500 students in 13 geographically diverse U.S. high schools. One video communicated the idea of malleable intelligence and growth mindset. The other video encouraged students to reflect on how working hard at school can help the students accomplish meaningful goals. One-third of participants were at risk of dropping out of high school.

The intervention was a success. Compared to a control group, both treatments raised students’ grade point averages in core academic courses. The fraction of at-risk students who satisfactorily completed core courses was 6.4 percentage points higher for the treatment group than for the control. The results show that interventions can be delivered at low-cost and at scale. The World Bank’s behavioral science unit, the Mind, Behavior, and Development Unit (eMBed), has delivered paper-based growth mindset interventions to approximately 40,000 students in Peru and 200,000 students in Indonesia. The preliminary results from Peru are promising.
Narratives are a source of shared mental representations. Some narratives enable people to see situations in a way that spurs activity. Other narratives constrain agency by representing a person as unable to influence outcomes outside a narrow domain. Famous examples of positive narratives that have shaped many Americans’ views of poverty are the rags-to-riches novels by Horatio Alger: no matter how dire the hero’s straits at the beginning, every hero in Alger’s stories escapes poverty by dint of effort, ability, and inner strength.

A Theater for Development program, *Jana Sanskriti*, active in villages in West Bengal, India since 1985, engages in fieldwork in which villagers describe incidents of domestic violence and other problems that they face. The artistic director then writes plays that are performed in the villages to dramatize the problems people experience in their daily lives. As Augusto Boal, the Brazilian writer and politician who created Theater for Development, explains,

> When the skit is over, the participants are asked if they agree with the solution presented. At least some will say no. At this point it is explained that the play will be performed once more, exactly as it was the first time. But now any participant in the audience has the right to replace any actor and lead the action in the direction that seems to him most appropriate... The other actors have to face the newly created situation, responding instantly to all the possibilities that it may present... Boal (1973).

The goal of *Jana Sanskriti* is to enable villagers to change their shared representations, for example, of domestic violence, and collectively rehearse social change. To evaluate the impact, Hoff, Jalan, and Santra (in progress) surveyed random samples of registered voters in villages where the plays had been performed and in matched villages where plays had never been performed. The study finds that exposure of a village to the plays reduced to less than 5 percent the fraction of both men and women who thought domestic violence was legitimate. Exposure reduced the percentage of households in which domestic violence had recently occurred from 32 percent to 26 percent. By providing an entry point for communities to collectively contest traditional social norms, the theater program may expand individuals’ cognitive toolkit for interpreting domestic relationships. In the new mental models, domestic violence is perceived as cruel and illegitimate.

**Barrier 3: The Challenge of Adapting to Two Worlds**

The title of Kahneman’s (2011) book, *Thinking, Fast and Slow*, captures a central tenet in behavioral economics and a major theoretical development in the understanding of human behavior: individuals have a dual cognitive process. “Fast” thinking is intuitive and automatic, and “slow” thinking is deliberate and effortful. Fast thinking is generally well-adapted to environments that one knows well: “Time and energy...
are saved, rumination and doubt are reduced, and nothing important is lost” (Ross and Nisbett 1991; see also Gigerenzer and others 1999). But in less familiar environments, fast thinking may lead to systematic mistakes. Economically and racially segregated neighborhoods put disadvantaged individuals in a difficult position: they do not access as young children the more privileged settings in which they can learn the norms and rules of thumb needed to thrive in such environments.

When the U.S. government offered people living in a poor neighborhood vouchers to move to a higher-income neighborhood, young children’s earnings later in life improved by $3,477 on average per year, an estimated $302,000 in a lifetime (Chetty, Hendren, and Katz 2016). A likely benefit of moving to better neighborhoods includes access to the social and cultural capital to function well in middle-income environments (Wilson 1987; Sampson 2012).

Individuals in out-groups must often navigate from a young age two very different cultural worlds—their home environment, with its epistemology and norms, and the world of the in-group, with different epistemology and norms. Children who live in areas plagued by high rates of crime must learn one set of rules of thumb to survive in their neighborhood and another set to thrive in school. School environments are heavily regulated by formal authority and norms of civility. In contrast, in high-crime neighborhoods, power and authority are fluid and negotiable. Anderson (1999) writes that “one of the most salient features of urban life in the minds of many people today is the relative prevalence of violence.” For people living in neighborhoods that lack institutions to prevent crime, conduct is typically regulated by the threat of violence—the “code of the street.” Coates (2015) wrote the following of his time growing up in Baltimore:

To survive the neighborhoods and shield my body, I learned another language consisting of a basic complement of head nods and handshakes. I memorized a list of prohibited blocks. I learned the smell and the feel of fighting weather. And I learned that “Shorty, can I see your bike” was never a sincere question, and “Yo, you was messing with my cousin” was neither an earnest accusation nor a misunderstanding of the facts. These were the summonses that you answered with your left foot forward, your right foot back, your hands guarding your face, one slightly lower than the other, cocked like a hammer.

Particularly for males who live in dangerous neighborhoods, there is generally no set of automatic responses to the assertion of authority that they can apply successfully both in their neighborhoods and in the school or workplace. An automatic response of compliance would endanger them in their neighborhoods, whereas an automatic response of non-compliance in school could lead to suspension, expulsion or termination. Middle-class youth normally do not face this problem. For them, the appropriate response to authority is the same in the home and school environment.
Interventions to Support Adaptive Strategies of Disadvantaged Groups

Traditional economics takes the person’s preferences as fixed. The standard policy prescription to deter disorder and violence is punishment. To encourage individuals to invest in their human capital, the standard policy prescription is to provide them information on the benefits or greater incentives are advised. In contrast, behavioral economics recognizes that the person can revise his mental models, behavior, and performance if given adequate social and psychological supports. This section discusses interventions that have increased individuals’ life skills and raised their aspirations.

One of the simplest ways for socially excluded groups to learn the rules of thumb necessary to succeed in more privileged settings is mentoring and coaching. Bettinger and Baker (2011) evaluated a mentoring and coaching service for non-traditional U.S. college students. Coaches worked with students to help them clarify their aspirations, connect their daily activities to long-term plans, and build skills such as time management and self-advocacy. Coached students were 14 percent more likely to persist in school after 24 months and four percentage points more likely to complete their degree within four years of receiving the treatment.

All-encompassing interventions are not always needed to change behavior in very positive ways. In some cases, rules of thumb lead to poor outcomes and can be changed. In South Africa, women who seek jobs generally do not ask for reference letters from former employers. When they do ask them to send letters, callback rates increase by 89 percent (Abel et al. 2017). One might suppose that this is because more capable women seek letters; their superior abilities, not the letters, are the reason for the higher callback rates. But this was not the explanation. A randomized controlled trial (RCT) that encouraged women in the treatment group to seek and use reference letters doubled their employment rates. In contrast, getting reference letters had no effect in the case of male applicants. Data collected three months after the intervention shows that the intervention closed the gender gap in job-search success.

Another strategy to lead people to adopt better rules of thumb in a particular domain is to encourage goal-setting. Deliberating and focusing on specific, challenging goals stimulates goal-directed behavior (Locke and Latham 1990). Goal-setting focuses attention on aspired states. It makes salient the losses incurred if one does not reach one’s goal, the relationship between steps necessary to achieve the goal, and the goal itself. An experiment in Canada recruited 85 low-performing university students and randomly assigned half of them to an intervention (Morisano et al. 2010). Students in the treatment group were invited to write down their aspirations, values, role models, priorities, and the ways that achieving their goals would affect the lives of other people. The students in the control group were asked to write about earlier positive experiences. Figure 1 shows that the students who participated in the goal-setting intervention had grades in the next semester almost half a point higher (on a
Figure 1. Low-performing Students Who Were Treated in the Goal-setting Intervention Made Gains in Their Grade Point Average Compared to the Control Group

Note: Morisano et al. (2010)

scale of four) than the control group. A related experiment with similar results can be found in Schippers et al. 2015.

One variant of the goal-setting approach is called “WOOP” (wish, outcome, obstacle, plan). This approach combines mental contrasting with detailed implementation intentions (Oettingen 2014). Mental contrasting entails visualizing your goals, the reasons they are important to you and the people around you, and considering the obstacles to achieving them. Implementation intentions involve making detailed “if... then...” plans to overcome obstacles (Gollwitzer 1999). Oettingen finds that contrasting their goals with the barriers to achieving them enhances motivation. Duckworth et al. (2013) implemented an intervention based on WOOP to eleven-year old children from disadvantaged backgrounds in the United States. The students were given a worksheet packet and asked to write down their most important wish or goal related to school work—“something that is challenging, but that you can achieve within the next few weeks or months,” the instructor explained. The children were also asked to write down “the one best outcome, the one best thing of fulfilling your wish or reaching your goal.” The children were given time to imagine the outcome they had written about, and then randomly assigned to one of two groups. In the treatment group, students were asked to imagine an obstacle they might face in achieving their wish and to create “if... then...” plans to overcome it. In the control group, students were asked to imagine a second positive outcome. The treatment group subsequently performed better than the control group: they had higher report card grades, higher attendance rates, and better conduct (Duckworth et al. 2013).

We next return to the central example in our discussion of barrier 3—automatic aggression in response to assertions of authority. Heller et al. (2017) evaluated RCTs of cognitive behavioral therapy programs to reduce automatic aggressive responses by disadvantaged male youths in the Chicago area. Two of the treatments were a program called “Becoming a Man.” The third treatment was a program in a Juvenile Temporary Detention Center.
A simple activity illustrates how the Becoming a Man program worked. The activity leaders used a provocative exercise to show how participants automatically followed one strategy rather than taking a moment to weigh their options. Activity leaders broke groups of participants into pairs and gave one person in each pair a ball. The other person was instructed to get the ball from him. He was given 30 seconds to do so. The automatic response of almost all the participants was to use force to take the ball. After the exercise, the activity leader pointed out that the participants could have asked for the ball. When prompted for an explanation as to why they had not done that, they usually responded that their partner would not have complied. The activity leader then asked the partner what he would have done if asked, and most partners said that they would have given the ball.

The participants in Becoming a Man had an average GPA of 2.0 (out of 4.0), had typically missed six to eight weeks of school in the year, and in many cases had a history of arrests. The intervention reduced participants’ interactions with the criminal justice system. The first intervention, delivered to boys in seventh to tenth grade, reduced violent crime by 44 percent and non-violent crime by 36 percent. Program participants also became more engaged in school, which the authors forecast could translate into increases in graduation rates of about 7 percent to 22 percent. The second intervention, delivered to boys in ninth and tenth grade, reduced arrests by 31 percent. The third intervention, delivered at the Juvenile Temporary Detention Center, reduced by 16 percentage points the re-admittance rate to the detention center within 18 months of release.

Cognitive therapy interventions have also been tested in Sierra Leone and Liberia. In Sierra Leone, Betancourt et al. (2014) evaluated a youth readiness intervention that delivered psychosocial supports to war-affected youth. The goal was to help them regulate their emotions and improve their problem-solving skills. The treatment provided all participants with an education subsidy and randomly allocated the psychosocial support treatment. Students who received the psychosocial supports performed better and were more likely to stay in school.

In Liberia, Blattman, Jamison, and Sheridan (2017) partnered with a local organization to provide group-based therapy and/or $200 cash (about three months’ wages) to almost 1,000 criminally-engaged men. Participants were randomly divided into four groups: one group received only the cash grant; a second group received eight weeks of therapy designed to foster self-regulation, patience, and a non-criminal identity; a third group received both the grant and therapy; and the control received neither. Those who had received only the cash transfer made no changes in criminal behavior or self-regulation. Among those who received only the therapy, violent and criminal behavior declined—in the short run, the individuals were 55 percent less likely to carry a weapon and 47 percent less likely to sell drugs. The effects were longer-lasting and stronger among those who had received the therapy followed by
The participants in this third group became less impulsive, had higher self-esteem, and were less likely to steal for at least a year after the intervention.

The studies discussed in this section show that short-term interventions can induce changes in behavior with long-run effects on well-being. One set of interventions shifted individuals from a “fixed” to a “growth” mindset of intelligence and helped them form goals, strategies to achieve them, and the emotional skills to follow through. Another set of interventions gave individuals who grew up in unsafe neighborhoods the mental tools to adapt to a non-violent environment and build productive lives.

Barrier 4: “Adaptive Preferences”

Traditional economics, as we have emphasized throughout, assumes fixed preferences, whereas Strand 2 of behavioral economics does not. In the past two decades, lab and field and “as-if-random” natural experiments (in which we include the work on gender and the plough) have provided evidence that is difficult to explain except through the impact of experience and exposure on preferences. Taking account of the malleability of preferences sheds light on a particularly perverse driver of social exclusion: the “adaptive preferences” of the oppressed (Nussbaum 2001). That is, a subjugated group may come to see its subjugation as natural, normative, or even preferred. Bourdieu (2000) writes that “the realistic, even resigned or fatalistic, dispositions which lead members of the dominated classes to put up with objective conditions that would be judged intolerable or revolting by agents otherwise disposed...help to reproduce the conditions of oppression” (emphasis added). Fatalistic attitudes and beliefs can be difficult to tackle in part because they call into question whether adults understand their own preferences. Duflo (2012) describes “hope” as a form of capability.

Examples of “adaptive preferences” in varying degrees of intensity exist in many contexts. For example, Guyon and Huillery (2014) find evidence of a large social class gap in aspirations of 14-year old students in the area of Paris. In many countries, as shown in figure 2, a high proportion of women report that a husband is justified in beating his wife for refusing sex. One reason that preferences adapt is that existing institutions serve as reference points—the baseline relative to which people imagine alternative realities. People have limited ability to imagine counterfactuals.

In parts of rural India, most girls leave school early, marry, and have children at a young age. The explanation in traditional economics would be that the behaviors reflect the preferred outcomes of the girls or their families given their opportunity sets. However, a recent RCT points also to the role of limited opportunities to imagine alternative ways of life. In a large-scale field experiment, Jensen (2012) provided three years of recruiting services to help women in remote villages in India get jobs in
the business process outsourcing industry. Although, on average, only three women per village were hired, young women in the treatment villages became less likely to be married or have children, and more likely to work or continue their education. The proportion of young women aged 15 to 21 who were married dropped from 71 percent to 66 percent, and the proportion with children dropped from 43 percent to 37 percent. In addition, the treatment closed 30 percent of the gap in body mass index between girls in the villages and girls in the wealthiest families in New Delhi. Was this simply a reflection of pre-existing preferences responding to new opportunities? Perhaps. But an equally plausible interpretation is that there was also a change in preferences: exposure to local village women who got good jobs helped young women imagine better lives for themselves.

One way for adaptive preferences to emerge is through a perverse trusting relationship that oppressed individuals may develop with their oppressors. This is sometimes called the Stockholm syndrome (Namnyak et al. 2008). The term emerged from a dramatic event in Sweden. On a summer morning in 1973, a prison-escapee entered a bank with a submachine gun and shot a police officer. In the failed bank robbery, he took four hostages and demanded that his prison mate be released from prison and join him. The government acquiesced. The two men barricaded themselves in the bank, with the hostages locked in the bank vault. Astonishingly, the hostages began to develop a bond with their captors and resisted cooperation with the police.
Examples related to the Stockholm Syndrome exist at the level of a whole society. In Sierra Leone, villages are ruled by “paramount chiefs” from families originally recognized by British colonial authorities. The number of ruling families varies across villages due to accidents of history. Villages with fewer ruling families have worse governance, child health, educational attainment, and incomes (Acemoglu, Reed, and Robinson 2014). One might expect that villages with poorer development outcomes would have less respect for authority and be less satisfied with the ruling families. But this is not the case. The researchers found that villages with fewer ruling families reported higher respect for authority.

In the Indian state of Maharashtra, local government is by a variety of objective measures more oppressive in villages in which the high castes own most of the land. To increase the extent to which the landless depend on them, the high castes block many national pro-poor programs. Yet the perceived legitimacy of village government is higher in high-caste-dominated villages: low-caste residents are 14 percent more likely to report trusting the landholders in the high-caste-dominated villages (Anderson, Francois, and Kotwal 2015).

Role Models

What can change dysfunctional “adaptive preferences”? By observing others, individuals may develop aspirations and more positively assess their prospect of achieving them (Bandura 1986, 1997). Role models who achieve success help individuals imagine new life paths and may boost self-efficacy beliefs.

A natural experiment in India shows the impact of creating role models by securing places for women in positions of power. As discussed above, for the position of village council leader (Pradhan), in the 1990s the government of India mandated quotas for women in a randomly selected one-third of all villages. Exposure to women Pradhans changed beliefs in at least four ways that reduced the social exclusion of women (Beaman et al. 2009; Beaman et al. 2012). (1) The experience of living under a woman Pradhan erased the prejudice, on average, of male villagers against women leaders by many measures—the evaluation of political speeches (discussed above), an IAT, and the assessment of the quality of actual village Pradhans. (2) In villages that had women leaders, parents’ aspirations for their teenage daughters—and teenage girls’ aspirations for themselves—were higher: and girls have gone to school somewhat longer and done somewhat fewer hours of housework. (3) After the quota program ended in a village, women had run for political office in higher percentages and in many cases won (see figure 3). (4) The greater presence of female political representatives produced an unexpected change in women’s reporting of crimes against women, and in the willingness of the police to record the reports (Iyer et al. 2012). This occurred even though Pradhans have no jurisdiction over the police. The increased reporting by female victims of crimes appears instead to reflect
Figure 3. The Fraction of Women Who Won Office in Free Elections

---

Note: Figure from World Bank (2015), data from Beaman et al. (2009).

a change in their perception of the costs—psychic and otherwise—of reporting the crimes.

Another way to change the “adaptive preferences” of disadvantaged groups is through documentaries of people who escaped poverty, and soap operas that expose viewers to new ways to build their lives. A field experiment conducted by Bernard et al. (2014) in Ethiopia examined the impact of showing a one-hour video with profiles of four Ethiopian villages that had escaped poverty or entered the middle class. The team randomly selected about 60 villages for the treatment and about 60 villages for the control group. The locations were remote—the researchers reached them by 4x4 vehicles or camels. Only 10 percent of the local populations watched TV at least once a week. The intervention raised aspirations and expectations, and the effect persisted. Five years later, aspirations were still significantly higher in the treatment than in the control group. The intervention also increased future-oriented behavior. School enrollment of children was 17 percent larger than in the control group six months after the intervention.

There is evidence that exposure to unfamiliar outcomes and social patterns even in fiction can give individuals new role models. The rise and popularity of soap operas in Brazil depicting agentic women with few or no children was a statistically significant factor in the recent decline in Brazil’s fertility rate. By taking advantage of the plausibly exogenous spread of television across Brazilian municipalities, La Ferrara, Hoff and Walsh
Chong, and Duryea (2012) showed that a fertility decline was caused by the soap operas in the year that a municipality first received the emissions. The decline was most pronounced for women who were close in age to a protagonist of one of the soap operas, which is consistent with the idea that the characters were role models. To put this effect in a comparative perspective, the impact was similar to the effect of two additional years in women’s education. Fertility fell by 11 percent of the mean among women aged 35 to 44. Organizations such as the BBC Media Action have brought this approach to scale across the world, tackling problems such as ethnic conflict, poor health, and oppressive gender norms.

Conclusion

A central theme of this essay is that institutions create concepts, categories, social identities, and other mental models that can persist long after the institutions are abolished. The mental models influence how boundedly rational people process information and what they want and aspire to become. Reforming or abolishing an institution to reduce social exclusion may not change the mental models that the institution has advanced. In that case, social exclusion will persist. The realization of equal substantive opportunity may require interventions that target socio-psychological barriers to social inclusion and upward mobility.

The interpretation of the causes of social exclusion in traditional economics contrasts sharply with the perspective in behavioral economics. If a social group remains at the bottom of the social ladder long after procedural equality of opportunity has been established, the implication in traditional economics would be that the group has fixed characteristics that impede upward mobility, or that network externalities keep it from rising: the group will move up in socio-economic status only if an event makes possible a coordinated change in behavior.

In contrast, behavioral economics shows that social exclusion is caused not only by structural and institutional barriers, but also by socio-psychological factors and further, that interventions can offset them. A stigmatized ascriptive identity affects its members in many ways besides procedural barriers to opportunity and explicit animus. Socially excluded groups face implicit bias. Negative stereotypes affect the group’s performance, self-concept, and aspirations and can also drive self-censorship. Growing up in a segregated, disadvantaged neighborhood can give individuals rules of thumb that are poorly adapted to success in school and work. The implication is that equality of outcomes between ascriptive social groups, such as those defined by race, gender, or ethnicity, should be a policy target along with formal equality of opportunity.

Procedural equality of opportunity as a target leaves unaddressed the schematizing power of the institutions that historically denied opportunity to certain groups. These institutions made the social inequalities appear normal and possibly
normative. A rapidly growing body of literature suggests that interventions can relax the constraints created by mental models that are a legacy of historical institutions. The impact of many of the interventions described in this paper are difficult to explain in rational choice theory, but are not difficult to explain under the quasi-rational, enculturated actor framework of behavioral economics.

Notes


1. The first three barriers are obvious, but the fourth may not be. An individual in a low-income class may rationally choose not to seek upward mobility—for example, not to work hard in school and not to delay child-bearing beyond adolescence—if doing so would create too great a distance from peers, connections to whom make a person happy (Akerlof, 1997) or if efforts to seek upward mobility (e.g., taking advanced courses in school or buying a computer) have a high return only if others within one’s social network make them, too (DiMaggio and Garip 2012). Peer group effects create a coordination problem that can explain the stability of class structure.


3. However, some evidence suggests that repeated interactions between ethnic groups in conflict with each other result in exclusionary attitudes towards the outgroup (Enos 2014).

4. As a result, within an affected school, the presence of poor children increased sharply in new cohorts but not in existing, older ones. Some schools delayed taking any action on the plan for a year because enrollment decisions had already been made. The variation—between cohorts within a school and between schools—in exposure of rich children to poor classmates makes it possible to identify the impact on the rich children of social interaction with poor children.

5. In a dictator game, there are two players—a dictator and a recipient. The dictator is given an endowment. In this experiment, it was 10 Indian rupees (20 U.S. cents). The dictator makes one decision—how to split the endowment between himself and the recipient. In this experiment, the students were invited to play as the dictator in two games. The recipients were anonymous but the dictator had information on the socio-economic status of their school. In one game, the recipient was from a school with poor children. In the other, the recipient was from an elite private school.

6. Boisjoly, Duncan, Kremer, Levy, and Eccles (2006) also find that exposure creates pro-social attitudes and behavior by in-groups toward out-groups. Compared to white students who were not randomly assigned African American roommates in college, white students who were assigned African American roommates were between one-third and one-half of a standard deviation more likely to endorse affirmative action. They reported several years later that they interacted more often and more comfortably with minorities.

7. This caused a preference reversal in sessions with both the math-based and verbal-based tasks. When candidates were made available separately, 65 percent of participants chose lower-performing males and 44 percent selected higher-performing females. When both a male and female were available for selection, only 3 percent of the participants chose the lower-performing male and 57 percent chose the higher-performing female.

8. White students also participated in the study. The intervention improved their performance but by a smaller amount than that of African American students.

9. The value of crime reduction alone is estimated to yield benefit-cost ratios that range from 5-to-1 up to 30-to-1 (Heller et al. 2017, p. 5).

10. School engagement increased by 0.14 standard deviations in the first year and by 0.19 standard deviations the second year.
References


Generalization in the Tropics – Development Policy, Randomized Controlled Trials, and External Validity

Jörg Peters, Jörg Langbein, and Gareth Roberts

When properly implemented, Randomized Controlled Trials (RCT) achieve a high degree of internal validity. Yet, if an RCT is to inform policy, it is critical to establish external validity. This paper systematically reviews all RCTs conducted in developing countries and published in leading economic journals between 2009 and 2014 with respect to how they deal with external validity. Following Duflo, Glennerster, and Kremer (2008), we scrutinize the following hazards to external validity: Hawthorne effects, general equilibrium effects, specific sample problems, and special care in treatment provision. Based on a set of objective indicators, we find that the majority of published RCTs does not discuss these hazards and many do not provide the necessary information to assess potential problems. The paper calls for including external validity dimensions in a more systematic reporting on the results of RCTs. This may create incentives to avoid overgeneralizing findings and help policy makers to interpret results appropriately. JEL codes: C83, C93

In recent years, intense debate has taken place about the value of Randomized Controlled Trials (RCTs). Most notably in development economics, RCTs have assumed a dominant role. The striking advantage of RCTs is that they overcome self-selection into treatment and thus their internal validity is indisputably high. This merit is sometimes contrasted with shortcomings in external validity (Basu 2014; Deaton and Cartwright 2016). Critics state that establishing external validity is more difficult for RCTs than for studies based on observational data (Moffit 2004; Roe and Just 2009; and Temple 2010; Dehejia 2015; Muller 2015; Pritchett and Sandefur 2015). This is particularly true for RCTs in the development context that tend to be implemented at smaller scale and in a specific locality. Scaling an intervention is likely to change the treatment effects because the scaled program is typically implemented by
resource-constrained governments, while the original RCT is often implemented by effective NGOs or the researchers themselves (Ravallion 2012; Bold et al. 2013; Banerjee et al. 2017; Deaton and Cartwright 2016).

This does not question the enormous contribution that RCTs have made to existing knowledge about the effectiveness of policy interventions. Rather, it underscores that “research designs in economics offer no free lunches—no single approach universally solves problems of general validity without imposing other limitations,” (Roe and Just 2009). Indeed, Rodrik (2009) argues that RCTs require “credibility-enhancing arguments” to support their external validity—just as observational studies have to make a stronger case for internal validity. Against this background, the present paper examines how the results published from RCT-based evaluations are reported, whether external validity-relevant design features are made transparent, and whether potential limitations to transferability are discussed.

To this end, we conduct a systematic review of policy evaluations based on RCTs published in top economic journals. We include all RCTs published between 2009 and 2014 in the American Economic Review, the Quarterly Journal of Economics, Econometrica, the Economic Journal, the Review of Economic Studies, the Review of Economics and Statistics, the Journal of Political Economy and the American Economic Journal: Applied Economics. In total, we identified 54 RCT-based papers that appeared in these journals.

Since there is no uniform definition of external validity and its hazards in the literature, in a first step we establish a theoretical framework deducing the assumptions required to transfer findings from an RCT to another policy population. We do this based on a model from the philosophical literature on the probabilistic theory of causality provided by Cartwright (2010), and based on a seminal contribution to the economics literature, the toolkit for the implementation of RCTs by Duflo, Glennerster, and Kremer (2008). We identify four hazards to external validity: (a) Hawthorne and John Henry Effects; (b) general equilibrium effects; (c) specific sample problems; and (d) problems that occur when the treatment in the RCT is provided with special care compared to how it would be implemented under real-world conditions.

As a second step, we scrutinized the reviewed papers with regard to how they deal with the four external validity dimensions and whether required assumptions are discussed. Along the lines of these hazards we formulated seven questions, then read all 54 papers carefully with an eye toward whether they address these seven questions. All questions can be objectively answered by “yes” or “no”; no subjective rating is involved.

External validity is not necessary in some cases. For example, when RCTs are used for accountability reasons by a donor or a government, the results are only interpreted within the evaluated population. Yet, as soon as these findings are used to inform policy elsewhere or at larger scale, external validity becomes a pivotal element. Moreover, test-of-a-theory or proof of concept RCTs that set out to disprove a general theoretical proposition speak for themselves and do not need to establish external
validity (Deaton and Cartwright 2016). However, in academic research most RCTs presumably intend to inform policy, and as we will also confirm in the review, the vast majority of included papers appear to generalize findings from the study population to a different policy population.2

Indeed, RCT proponents in the development community advocate in favor of RCTs in order to create “global public goods” that “can offer reliable guidance to international organizations, governments, donors, and NGOs beyond national borders,” (Duflo, Glennerster, and Kremer 2008). As early as 2005, during a symposium on “New directions in development economics: Theory or empirics?” Abhijit Banerjee acknowledged the requirement to establish external validity for RCTs and, like Rodrik, called for arguments that establish the external validity of RCTs (Banerjee 2005). Indeed, Banerjee and Rodrik seem to agree that external validity is never a self-evident fact in empirical research, and that RCTs in particular should discuss in how far results are generalizable.

In the remainder of the paper we first present the theoretical framework and establish the four hazards to external validity. Following that, the methodological approach and the seven questions are discussed. The results are presented in the next section, followed by a discussion section. The subsequent section provides an overview on existing remedies for external validity problems and ways to deal with them in practice. The final section concludes.

Theoretical Background and Definition of External Validity

**Theoretical Framework**

Understanding what external validity exactly is and how it might be threatened is not clearly defined in the literature. What we are interested in here is the degree to which an internally valid finding obtained in an RCT is relevant for policy makers who want to implement the same intervention in a different policy population. Cartwright (2010) defines external validity in a way that is similar to the understanding conveyed in Duflo, Glennerster, and Kremer (2008): “External validity has to do with whether the result that is established in the study will be true elsewhere.” Cartwright provides a model based on the probabilistic theory of causality. Using this model we identify the assumptions that have to be made when transferring the results from an RCT to what a policy maker can expect if she scales the intervention under real-world conditions.

Suppose we are interested in whether a policy intervention \( C \) affects a certain outcome \( E \). We can state that \( C \) causes \( E \) if

\[
P(E|C&K_i) > P(E|\bar{C}&K_i)
\]

where \( K_i \) describes the environment and intervention particularities under which the observation is made, and \( \bar{C} \) denotes the absence of the intervention. Assume this
causal relationship was observed in population A and we want to transfer it to a situation in which C is introduced to another population, A’. In this case, Cartwright points out that those observations, K_i, have to be identical in both populations A and A’ as soon as they interfere with the treatment effect. More specifically, Cartwright formulates the following assumptions that are required: (a) A needs to be a representative sample of A’; (b) C is introduced in A’ as it was in the experiment in A; (c) the introduction leaves the causal structure in A’ unchanged.

In the following, we use the language that is widely used in the economics literature and refer to the toolkit for the implementation of RCTs by Duflo, Glennerster, and Kremer (2008). Similar to the Cartwright framework, Duflo, Glennerster, and Kremer introduce external validity as the question “[…] whether the impact we measure would carry over to other samples or populations. In other words, whether the results are generalizable and replicable”. The four hazards to external validity that are identified by Duflo, Glennerster, and Kremer are Hawthorne and John Henry Effects, general equilibrium effects, the specific sample problem, and the special care problem. The following section presents these hazards to external validity in more detail. Under the assumption that observational studies mostly evaluate policy interventions that would have been implemented in every case, Hawthorne/John Henry Effects and the special care problem are much more likely in RCTs, while general equilibrium effects and the specific sample problem equally occur in RCTs and observational studies.

Potential Hazards to External Validity

In order to guide the introduction to the different hazards of external validity we use a stylized intervention of a cash transfer given to young adults in an African village. Suppose the transfer is randomly assigned among young male adults in the village. The evaluation examines the consumption patterns of the recipients. We observe that the transfer receivers use the money to buy some food for their families, football shirts, and air time for their mobile phones. In comparison, those villagers who did not receive the transfer will not change their consumption patterns. What would this observation tell us about giving a cash transfer to people in different set-ups? The answer to this question depends on the assumptions identified in Duflo, Glennerster, and Kremer’s nomenclature.

Hawthorne and John Henry effects might occur if the participants in an RCT know or notice that they are part of an experiment and are under observation. ³ It is obvious that this could lead to altered behavior in the treatment group (Hawthorne effect) and/or the control group (John Henry effect). ⁴ In the stylized cash transfer example, the recipient of the transfer can be expected to spend the money for other purposes in case he knows that his behavior is under observation. It is also obvious that such behavioral responses clearly differ between different experimental set-ups. If the experiment is embedded into a business-as-usual setup, distortions of participants’
behavior are less likely. In contrast, if the randomized intervention interferes noticeably with the participants’ daily life (e.g., an NGO appearing in an African village to randomize a certain training measure among the villagers), participants will probably behave differently than they would under non-experimental conditions.5

The special care problem refers to the fact that in RCTs, the treatment is provided differently from what would be done in a non-controlled program. In the stylized cash transfer example, a lump sum payment that is scaled up would perhaps be provided by a larger implementing agency with less personal contact. Bold et al. (2013) provide compelling evidence for the special care effect in an RCT that was scaled up based on positive effects observed in a smaller RCT conducted by Duflo, Kremer, and Robinson (2011b). The major difference is that the program examined in Bold et al. was implemented by the national government instead of an NGO, as was the case in the Duflo et al. study. The positive results observed in Duflo, Kremer, and Robinson (2011b) could not be replicated in Bold et al. (2013): “Our results suggest that scaling-up an intervention (typically defined at the school, clinic, or village level) found to work in a randomized trial run by a specific organization (often an NGO chosen for its organizational efficiency) requires an understanding of the whole delivery chain. If this delivery chain involves a government Ministry with limited implementation capacity or which is subject to considerable political pressures, agents may respond differently than they would to an NGO-led experiment.”

Vivalt (2017) confirms the higher effectiveness of RCTs implemented by NGOs or the researchers themselves as compared to RCTs implemented by governments in a meta-analysis of published RCTs. Further evidence on the special care problem is provided by Allcott (2015), who shows that electricity providers that implemented RCTs in cooperation with a large research program to evaluate household energy conservation instruments are systematically different from those electricity providers that do not participate in this program. This hints at what Allcott refers to as “site selection bias”, whereby organizations that agree to cooperate with researchers on an RCT can be expected to be different compared to those that do not, for example because their staff are more motivated. This difference could translate into higher general effectiveness. Therefore, the effectiveness observed in RCTs is probably higher than it will be when the evaluated program is scaled to those organizations that did not initially cooperate with researchers.

The third identified hazard arises from potential general equilibrium effects (GEE).6 Typically, such GEE only become noticeable if the program is scaled to a broader population or extended to a longer term. In the stylized cash transfer example provided above, GEE occur if not only a small number of people but many villagers receive the transfer payment. In this scaled version of the intervention, some of the products that young male villagers want to buy become scarcer, and thus more expensive. This also illustrates that GEE can affect non-treated villagers, as prices increase for them as well. Moreover, in the longer term if the cash transfer program is implemented
permanently, certain norms and attitudes towards labor supply or educational investment might change.\textsuperscript{7}

This example indicates that GEE in their entirety are difficult to capture. The severity of GEE, though, depends on some parameters like the regional coverage of the RCT, the time horizon of the measurements, and the impact indicators that the study examines. Very small-scale RCTs or those that measure outcomes after a few months only are unlikely to portray the change in norms and beliefs that the intervention might entail. Furthermore, market-based outcomes like wages or employment status will certainly be affected by adjustments in the general equilibrium if an intervention is scaled and implemented over many years. As a matter of course, it is beyond the scope of most studies to comprehensively account for such GEE, and RCTs that cleanly identify partial equilibrium effects can still be informative for policy. A profound discussion of GEE-relevant features is nonetheless necessary to avoid the ill-advised interpretation of results. Note that GEE are not particular to RCTs and, all else being equal, the generalizability of the results from observational studies is also exposed by potential GEE. Many RCTs, particularly in developing country contexts, are however, limited to a specific region, a relatively small sample size, and short monitoring horizon, and are thus more prone to GEE than country-wide representative panel-data based observational studies.

In a similar vein, the fourth hazard to external validity, the specific sample problem, is not particular to RCTs but might be more pronounced in this setting. The problem occurs if the study population is different from the policy population in which the intervention will be brought to scale. Taking the cash transfer example, the treatment effect for young male adults can be expected to be different if the cash transfer is given to young female adults in the same village or to young male adults in a different part of the country.

Methods and Data

\textit{Review Approach}

We reviewed all RCTs conducted in developing countries and published between 2009 and 2014 in the leading journals in economics. We included the five most important economics journals, namely the \textit{American Economic Review}, \textit{Econometrica}, \textit{Quarterly Journal of Economics}, \textit{Journal of Political Economy}, the \textit{Review of Economic Studies}, as well as further leading journals that publish empirical work using RCTs such as \textit{American Economic Journal: Applied Economics}, \textit{Economic Journal}, and \textit{Review of Economics and Statistics}.

We scrutinized all issues in the period, particularly all papers that mention either the terms “field experiment”, “randomized controlled trials”, or “experimental evidence” in either the title or the abstract, or which indicated in the abstract or the title...
that a policy intervention was randomly introduced. We excluded those papers that examine interventions in an OECD member country.\(^8\) In total, 73 papers were initially identified. Our focus is on policy evaluation and we therefore excluded mere test-of-a-theory papers.\(^9\) In most cases, the demarcation was very obvious and we subsequently excluded 19 papers. In total, we found 54 papers based on an RCT to evaluate a certain policy intervention in a developing country.\(^{10}\) The distribution across journals is uneven, with the vast majority being published in American Economic Journal: Applied Economics, American Economic Review and Quarterly Journal of Economics (see figure 1).

Figure 2 depicts the regional coverage of the surveyed RCTs. The high number of RCTs implemented in Kenya is due to the strong connection that two of the most prominent organizations that conduct RCTs have to the country (Innovation

---

**Figure 1. Published RCTs Between 2009 and 2014**

Note: A total of 54 studies were included, frequencies appear in bold.
for Poverty Action [IPA] and the Abdul Latif Jameel Poverty Action Lab [J-Pal]).
Most of these studies were implemented in Kenya’s Western Province by the Dutch
NGO International Child Support (ICS), IPA, and J-Pal’s cooperation partner in the
country.11

We read all 54 papers carefully (including the online supplementary appendix) to
determine whether each paper addressed seven objective yes/no-questions. An ad-
ditional filter question addresses whether the paper has the ambition to generalize.
This is necessary, because it is sometimes argued that not all RCTs intend to generate
generalizable results and are rather designed to test a theoretical concept. In fact, 96
percent of included papers do generalize (see next section for details on the coding
of this question). This is no surprise, since we intentionally excluded test-of-a-theory
papers and focused on policy evaluations. The remaining seven questions all address
the four hazards to external validity outlined in the first, and examine whether the

Note: A total of 54 studies were included, frequencies appear in bold.
“credibility-enhancing arguments” (Rodrik 2009) are provided to underpin the plausibility of external validity. Appendix A in the appendix shows the answers to the seven questions for all surveyed papers individually. In general, we answered the questions conservatively, that is, when in doubt we answered in favor of the paper. We abstained from applying subjective ratings in order to avoid room for arbitrariness. A simple report on each paper documents the answers to the seven questions and the quote from the paper underlying the respective answer. We sent these reports out to the lead authors of the included papers and asked them to review our answers for their paper(s). For 36 of the 54 papers we received feedback, based on which we changed an answer from “no” to “yes” in 9 cases (out of 378 questions and answers in total). The comments we received from the authors are included in the reports, if necessary followed by a short reply. The revised reports were sent again to the authors for their information and can be found in the online supplementary appendix to this paper.

Seven Questions

To elicit the extent the paper accounts for Hawthorne and John Henry effects, we first asked the following objective questions:

1. Does the paper explicitly say whether participants are aware (or not) of being part of an experiment or a study?

   This question accounts for whether a paper provides the minimum information that is required to assess whether Hawthorne and John Henry effects might occur. More would be desirable: in order to make a substantiated assessment of Hawthorne-like distortions, information on the implementation of the experiment, the way participants were contacted, which specific explanations they received, and the extent to which they were aware of an experiment should be presented. We assume (and confirmed in the review) that papers that receive a “no” for question 1 do not discuss these issues because a statement on the participants’ awareness of the study is the obvious point of departure for this discussion. It is important to note that unlike laboratory or medical experiments, participants in social science RCTs are not always aware of their participation in an experiment. Only for those papers that receive a “yes” to question 1 do we additionally pose the following question:

2. If people are aware of being part of an experiment or a study, does the paper (try to) account for Hawthorne or John Henry effects (in the design of the study, in the interpretation of the treatment/mechanisms, or in the interpretation of the size of the impact)?
The next set of questions probes into *general equilibrium effects*. As outlined in the first section, we define general equilibrium effects as changes due to an intervention that occur in a noticeable way only if the intervention is scaled or after a longer time period.

Two questions capture the two transmission channels through which GEE might materialize:

3. Does the paper explicitly discuss what might happen if the program is scaled up?
4. Does the paper explicitly discuss if and how the treatment effect might change in the long run?\(^{13}\)

For both questions, we give the answer “yes” as soon as the respective issue is mentioned in the paper, irrespective of whether we consider the discussion to be comprehensive. The third hazard is what Duflo, Glennerster, and Kremer call the *specific sample problem* and is addressed by question 5:

5. Does the paper explicitly discuss the policy population (to which the findings are generalized) or potential restrictions in generalizing results from the study population?

We applied this question only to those papers that explicitly generalize beyond the study population (see the filter question below). As soon as a paper discusses the study population vis-à-vis the policy population, we answered the question with “yes”, irrespective of our personal judgment on whether we deem the statement to be plausible and the discussion to be comprehensive.

The fourth hazard, special care, is accounted for by the last two questions.

6. Does the paper discuss particularities of how the randomized treatment was provided in demarcation to a (potential) real-world intervention?

As soon as the paper makes a statement on the design of the treatment compared to the potential real-world treatment, we answered the question with “yes”, again irrespective of our personal judgment of whether we deem the statement to be plausible and comprehensive. In addition, to account for the concern that RCTs implemented by NGOs or researchers themselves might be more effective than scaled programs implemented by, for example, government agencies, we ask:

7. Who is the implementation partner of the RCT?

The specific wording of the additional filter question is “Does the paper generalize beyond the study population?” Our coding of this question certainly leaves more room for ambiguity than the coding for the previous objective questions. We therefore answered this additional question by a “yes” as soon as the paper makes any generalizing statements (most papers do that in the conclusions) that a mere test-of-a-theory would not make.\(^{14}\) Note that in this question we do not assess the
Table 1. Reporting on External Validity in Published RCTs

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer is yes (in percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hawthorne and John Henry Effect: Does the paper</strong></td>
<td></td>
</tr>
<tr>
<td>1. explicitly say whether participants are aware of being part of an</td>
<td>35</td>
</tr>
<tr>
<td>experiment or a study?</td>
<td></td>
</tr>
<tr>
<td>2. (try) to account for Hawthorne or John Henry effects?</td>
<td>29</td>
</tr>
<tr>
<td><strong>General Equilibrium Effects: Does the paper</strong></td>
<td></td>
</tr>
<tr>
<td>3. discuss what happens if program is scaled up?</td>
<td>44</td>
</tr>
<tr>
<td>4. discuss changes to treatment effects in the long run?</td>
<td>46</td>
</tr>
<tr>
<td><strong>Specific Sample Problems: Does the paper</strong></td>
<td></td>
</tr>
<tr>
<td>5. discuss the policy population or potential restrictions to</td>
<td>77</td>
</tr>
<tr>
<td>generalizability? †</td>
<td></td>
</tr>
<tr>
<td><strong>Special Care: Does the paper</strong></td>
<td></td>
</tr>
<tr>
<td>6. cover particularities of how the randomized treatment was provided</td>
<td>20</td>
</tr>
<tr>
<td>in demarcation to a (potential) real-world intervention discussed?</td>
<td></td>
</tr>
</tbody>
</table>

*Note:* † indicates that question 3 only applies to those 19 papers that explicitly state that participants are aware of being part of an experiment. ‡ indicates that question 5 only applies to those 52 papers that explicitly generalize.

degree to which the paper generalizes (which in fact varies considerably), but only if it generalizes at all.

Results

Table 1 shows the results for the seven questions asked of every paper. As noted above, 96 percent of the reviewed papers generalize their results. This underpins the proposal that these studies should provide “credibility-enhancing arguments”. It is particularly striking that only 35 percent of the published papers mention whether people are aware of being part of an experiment (question 1). This number also reveals that it is far from common practice in the economics literature to publish either the protocol of the experiment or the communication with the participants. Some papers even mention letters that were sent or read to participants but do not include the content in the main text or the appendix.

Only 46 percent of all papers discuss how effects might change in the longer term and whether some sort of adjustments might occur (question 4). Here, it is important to note that around 65 percent of the reviewed papers examine impacts less than two years after the randomized treatment; on average, impacts are evaluated 17 months after the treatment (not shown in the table). While this is in most cases probably inevitable for practical reasons, a discussion of whether treatment effects might change in the long run, for example, based on qualitative evidence or theoretical considerations, would be desirable. Note that most of the papers that do discuss long-term...
effects are those that in fact examine such long-term effects. In other words, only a small minority of papers that only look at very short-term effects provides a discussion of potential changes in the long run.

Likewise, potential changes in treatment effects in case the intervention is scaled are hardly discussed (question 3, 44 percent of papers); 35 percent of the papers do not mention GEE related issues at all, that is, received a “no” for questions 3 and 4 (not shown in table 1). The best score is achieved for the specific sample problem: 77 percent of papers discuss the policy population or potential restrictions to generalizability.

As the results for question 6 show, only 20 percent discuss the special care problem. This finding has to be interpreted in light of the result for question 7 in figure 3: more than 60 percent of RCTs were implemented by either the researchers themselves or an NGO. For these cases, a discussion of the special care issue is particularly relevant. The
remaining RCTs were implemented by either a large firm or a governmental body—which may better resemble a business-as-usual situation.\textsuperscript{15}

Table A1 in the supplementary online appendix provides a further decomposition of the results presented in table 1 and shows the share of “yes” answers for the respective year of publication. There is some indication of an improvement from 2009 to 2014, but only for certain questions. For example, the share of “yes”-answers increases to over 50 percent for question 1 on people’s awareness of being part in a study and question 3 on the implications of scaling. For the specific sample dimension, the share of “yes” answers to question 5 is lower in 2014 than in the years from 2009 until 2013. For all other questions, we do not observe major differences. Overall, there is no clear trend towards a systematic and transparent discussion of external validity issues.

Discussion

In this section we consider some of the comments and arguments that have been put forward during the genesis of the paper. We would like to emphasize that for the sake of transparency and rigor, we only used objective questions and abstained from qualitative ratings. While we acknowledge that this approach does not do justice to every single paper, we argue that the overall pattern we obtain is a fair representation of how seriously external validity issues are taken in the publication of RCTs. Please note once again that we answered all questions very conservatively.

To summarize the results, we find that many published RCTs do not provide a comprehensive presentation of how the experiment was implemented.\textsuperscript{16} More than half of the papers do not even mention whether the participants in the experiment are aware of being randomized—which is crucial for assessing whether Hawthorne or John Henry effects could co-determine the outcomes in the RCT. It is true that in some cases it is obvious that participants were aware of an experiment, but in most cases it is indeed ambiguous. In addition, even in cases where it is obvious, it is important to know what exactly participants were told and thus, a discussion of how vulnerable the evaluated indicators are to Hawthorne-like distortions would be desirable.

Furthermore, our results show that potential general equilibrium effects are only rarely addressed. This is above all worrisome in the case that outcomes involve price changes (e.g., labor market outcomes) so that repercussions when the program is brought to scale are almost certain. Likewise, the special care problem is hardly discussed, which is particularly concerning in the developing country context, where many RCTs are implemented by NGOs that are arguably more flexible in terms of treatment provision than the government.

A number of good practice examples exist where external validity issues are avoided by the setting or openly addressed, demonstrating that a transparent discussion of “credibility enhancing arguments” is possible. As for Hawthorne effects,
in Karlan et al. (2014), for example, participants are not aware of the experiment, which is also clearly stated in the paper. In Bloom et al. (2013), in contrast, participants are aware, but the authors discuss the possibility of distorting effects intensely. For general equilibrium effects, Blattman, Fiala, and Martinez (2014) address potential adjustments in the equilibrium, which are quite likely in their cash transfer randomization. As for the specific sample problem, Tarozzi et al. (2014) openly discuss that their study might have taken place in a particular population. Good practice examples for the special care hazard are again Blattman, Fiala, and Martinez (2014), since their program is implemented by the government and therefore resembles a scaled intervention. Duflo, Dupas, and Kremer (2011a) reveal potential special care problems and acknowledge that a scaled program might be less effective. 17

We abstain from giving explicit bad practice examples (for obvious reasons), but indeed some studies are, we believe, negligently silent about certain hazards in spite of very obvious problems. In a minority of cases, this is even exacerbated by a very ambitious and broad generalization of the findings.

Some commentators argued that RCTs that test a theory are not necessarily meant to be generalized. Yet by design we concentrate our review on papers that evaluate a certain policy and hence the vast majority of papers included in this review do generalize results. In addition, a mere test-of-a-theory paper should in our views communicate this clearly to avoid misleading interpretations by policy makers and the public.

This is related to the question of whether in fact all papers are supposed to address all external validity dimensions included in our review. Our answer is yes, at least for policy evaluations that generalize their findings. One might argue that some of the reviewed papers are completely immune to a certain external validity hazard, but the cost of briefly establishing this immunity is negligible.

Potential Remedies

In an ideal world, external validity would be established by replications in many different populations and using different designs that vary the parameters which potentially codetermine the results. Systematic reviews can then compile the collective information in order to identify patterns in the effectiveness that eventually inform policy. This is the mission of organizations like the Campbell Foundation, the Cochrane Foundation, as well as the International Initiative for Impact Evaluation (3ie), and systematic reviews have indeed been done in a few cases. 18 In a similar vein, Banerjee et al. (2017) propose a procedure “from proof of concept to scalable policies.” The authors acknowledge that proof of concept studies are often intentionally conducted under “ideal conditions through finding a context and implementation partner most likely to make the model work”. These authors suggest an approach of “multiple iterations of experimentation”, in which the context that
co-determines the results is refined. Banerjee et al. (2017) also provide a promising example in India for such a scaling up process. Yet it is evident that this approach, as well as systematic reviews, require a massive collective research endeavor that will take many years and is probably not feasible in all cases.

It is this paper’s stance that in the meantime, individual RCTs with a claim to broader policy relevance have to establish external validity, reveal limitations, and discuss implications for transferability openly. To achieve this goal, the first and most obvious step is to include a systematic reporting in RCT-based publications following the CONSORT statement in the medical literature. This reference to the CONSORT statement as a role-model for economics research has already been postulated by Miguel et al. (2014) and Eble, Boone, and Elbourne (2017), for example. Some design features could be retrieved already in the pre-analysis plan, but at the latest during the peer-review process the checklist should be included and reviewed. Such a checklist ensures that the reader has all information at hand allowing her to make an informed judgment on the transferability of the results. Moreover, potential weaknesses should be disclosed, thereby automatically entailing a qualitative discussion to establish or restrict the study’s external validity. In addition, a mandatory checklist also creates incentives to already take external validity issues into account in the study’s design phase.

Next to more transparency in the publication of RCTs, a few instruments exist to deal with external validity hazards—some of which are post hoc, others of which can be incorporated in the design of the study. For Hawthorne and John Henry effects, the most obvious solution is not to inform the participants about the randomization, which of course hinges upon the study design. Such an approach resembles what Levitt and List (2009) refer to as a “natural field experiment”. In some set-ups, people have to be informed, either because randomization is obvious or for ethical reasons. The standard remedy in medical research—assigning a third group to a placebo treatment—is not possible in most experiments in social sciences. Aldashev, Kirchsteiger, and Sebald (2017) emphasize that the assignment procedure that is used to randomly assign participants into treatment and control groups affects the size of the bias considerably. These authors suggest that a public randomization reduces bias compared to a non-transparent private randomization.

Accounting for general equilibrium effects comprehensively is impossible in most cases, since all types of macro-economic adjustments can hardly be captured in a micro-economic study. In order to evaluate what eventually happens in the general equilibrium, one would have to resort to computable general equilibrium (CGE) models. Indeed, there are ambitions to plug the results of RCT-based evaluations into CGE models, as is done with observational data in Coady and Harris (2004).

The seminal work on GEE so far tests for the existence of at least selected macro-economic adjustments and spillovers by randomizing not only the treatment within clusters (e.g., markets), but also the treatment density between clusters. Influential
examples of this approach are Crépon et al. (2013) on the French labor market, and Muralidharan and Sundararaman (2015) for school vouchers in India. Using the same approach, Burke et al. (2017) randomizes the density of loan offers across regions to account for GEE. Moreover, randomizing the intervention on a higher regional aggregation allows for examining the full general equilibrium effect at that level (Banerjee et al. 2017). Muralidharan, Niehaus, and Sukh (2017), for example, examine a public employment program at a regional level that is “large enough to capture general equilibrium effects”. Attanasio, Kugler, and Meghir (2011) exploit the randomized PROGRESA roll-out on the village level to study GEE on child wages.

As for the specific sample problem, there is an emerging body of literature that provides guidance on extrapolating findings from one region to another. Pearl and Bareinboim (2014) develop a conceptual framework that enables the researcher to decide whether transferring results between populations is possible at all. Moreover, these authors formulate assumptions that, if they hold true, allow for transferring results from RCT based studies to observational ones (“license to transport”). Gechter (2016) takes a similar line and develops a methodology that calculates bounds for transferring treatment effects obtained in an RCT to a non-experimental sample. The key assumption here is that “the distribution of treated outcomes for a given untreated outcome in the context of interest is consistent with the experimental results,” (see Gechter 2016). Further contributions offer solutions for very specific types of RCTs. For example, Kowalski (2016) provides a methodology suitable for RCTs using an encouragement design (i.e., with low compliance rates), while Stuart et al. (2011) propose a methodology to account for selection into the RCT sample, which is often the case in effectiveness studies.

The degree to which scholars believe in the generalizability of results also hinges upon which part of the results chain they focus. One line of thinking concentrates on the human behavior component in evaluations, also referred to as “mechanism”, and assumes this to be more generalizable than what is found on the intervention as a whole (see, e.g., Bates and Glennerster 2017). The other viewpoint puts more emphasis on the treatment as a policy intervention. Here, the complexity of interventions and the context in which they happen are decisive. This camp calls for combining evidence from rigorous evaluations with case studies (Woolcock 2013) or “reasoned intuition” (Basu 2014; Basu and Foster 2015) to transfer findings from one setting to a different policy population.

This complexity feature is very much related to what we have referred to as special care in the provision of the treatment, which is arguably very heterogeneous across different policy environments. There seems to be a growing consensus that this is an important external validity concern (see, e.g., Banerjee et al. 2017), and some scholars have made recommendations on how to account for this. Both Bates and Glennerster (2017) and Woolcock (2013) provide frameworks that guide the transferability assessment, and special care is one important feature.
Bates and Glennerster (2017) suggest isolating the mechanism from other intervention-related features, while Woolcock (2013) argues that in many “developing countries [. . .] implementation capability is demonstrably low for logistical tasks, let alone for complex ones.” Hence, the higher the complexity of an intervention, the more implementation capability becomes a bottleneck, and, to use our wording, the more special care puts external validity at risk. Woolcock’s position is that for complex interventions—that is, the vast majority of policy interventions—generalizing is a “decidedly high-uncertainty undertaking”. Woolcock suggests including qualitative case studies into these deliberations.

**Conclusion**

In theory, there seems to be a consensus among empirical researchers that establishing external validity of a policy evaluation is as important as establishing its internal validity. Against this background, this paper has systematically reviewed published RCTs to examine whether external validity concerns are addressed. Our findings suggest that external validity is often neglected and does not play the important role that it is associated with in review papers and the general academic debate.

In a nutshell, our sole claim is that papers should discuss the extent to which the different hazards to external validity apply. We call for dedicating the same devotion to establishing external validity as is done when establishing internal validity. This thinking implies that papers published in top academic journals are not only targeted to the research community, but also to a policy-oriented audience (including decision-makers and journalists). This audience, in particular, requires all the information necessary to make informed judgments on the extent to which the findings are transferable to other regions and non-experimental business-as-usual settings. More transparent reporting would also lead to a situation in which more generalizable RCTs receive more attention than those that were implemented under heavily-controlled circumstances or in a very specific region only.

It would be desirable if the peer review process at economics journals explicitly scrutinized design features of RCTs that are relevant for generalization. As a starting point, this does not need to be more than a checklist and short statements to be included in an electronic appendix. The logic is that if researchers know already at the beginning of a study that they will need to provide such checklists and discussions, they will have clear incentives to account for external validity issues in the study design. Otherwise, external validity degenerates to a nice-to-have feature that researchers account for voluntarily and for intrinsic reasons. These internal incentives will probably work in many cases. But given the trade-offs we all face during the laborious implementation of studies, it is almost certain that external validity will often be sacrificed for other features to which the peer-review process currently pays more attention.
Jörg Peters is heading the research group “Climate Change in Developing Countries” at RWI, Germany and is Professor at University of Passau. All correspondence to be sent to: Jörg Peters, RWI, Hohenzollernstraße 1–3, 45128 Essen, Germany, e-mail: peters@rwi-essen.de, phone: 49-201-8149-247. Jörg Langbein is Researcher at RWI, Germany. Gareth Roberts is lecturer at University of the Witwatersrand and researcher at AMERU, Johannesburg, South Africa. The authors thank Maximilian Huppertz and Julian Rose for excellent research assistance. The authors are also grateful for valuable comments and suggestions by the editor Peter Lanjouw, three anonymous referees, Martin Abel, Mark Andor, Michael Grimm, Angus Deaton, Heather Lanthorn, Luciane Lenz, Stephan Klasen, Laura Poswell, and Colin Vance, as well as seminar participants at the University of Göttingen, University of Passau, University of the Witwatersrand, and Stockholm Institute of Transition Economics. We contacted all lead authors of the papers included in this review and many of them provided helpful comments on our manuscript. Langbein and Peters gratefully acknowledge the support of a special grant (Sondertatbestand) from the German Federal Ministry for Economic Affairs and Energy and the Ministry of Innovation, Science, and Research of the State of North Rhine-Westphalia. A supplemental appendix to this article is available at https://academic.oup.com/WBRO.

1. The title is an obvious reference to an important contribution to this debate, Angus Deaton’s “Instruments of Development: Randomization in the Tropics and the Search for the Elusive Keys to Economic Development”, published as an NBER working paper in 2009 (Deaton 2009). A revised version was published under a different title in the Journal of Economic Literature (Deaton 2010).

2. Note that our focus is on policy evaluation. In our protocol, we therefore excluded laboratory experiments, framed field experiments, and test-of-a-theory field experiments that are obviously not meant to evaluate a policy intervention.

3. The Hawthorne effect in some cases cannot be distinguished from survey effects, the Pygmalion effect, and the observer-expectancy effect (see Bulte et al. 2014). All of these effects, which generally might occur in observational studies, can be amplified by the Hawthorne effect and the experimental character of the study. See Aldashev, Kirchsteiger, and Sebald (2017) for a formalization of the Hawthorne and John Henry effect.

4. The John Henry effect describes the effect that being randomized into the control group can have on the performance of control group members. John Henry is a legendary black railroad worker, who—equipped with a traditional sledgehammer—competed with a steam trill in an experimental setting. Being aware of this exercise, he strived to outperform the steam drill. While he eventually succeeded, he died from exhaustion (see Saretsky 1972, for a very classic example of a John Henry effect).

5. See Bulte et al. (2014) and Simons et al. (2017) for evidence on strong Hawthorne effects in experiments in Tanzania and Uganda, respectively, and McCambridge, Witton, and Elbourne (2014) for a systematic review on Hawthorne effects in medical research. Cilliers, Dube, and Siddiqi (2015) provide evidence for the distorting effects of foreigner presence in framed field experiments in developing countries. See also Zwane et al. (2011).

6. See Crépon et al. (2013) for an example of such GEE in a randomized labor market program, in which treated participants benefited at the expense of non-treated participants.

7. Attanasio, Kugler, and Meghir (2011) observe a reduction in labor supply for child labor in the Mexican PROGRESA conditional cash transfer intervention, which is disbursed conditioned on children going to school.

8. The present study builds on an earlier paper that also included RCTs conducted in developed countries, see Peters, Langbein, and Roberts (2016).

9. See appendix B for the list of the excluded papers and the reason for exclusion.

10. A comprehensive list of included papers and their rating is found in Appendix A.

11. See Roetman (2011) for more information on the genesis of RCTs in Kenya and the role of ICS.

12. The filter question on whether the paper generalizes beyond the study population was added post-hoc, as a response to comments made by some authors.
13. The time period of a study is of course not only an external validity issue. See King and Behrman (2009) on the relevance of timing for impact evaluations.

14. We coded this question by “yes” in case the paper derives explicit policy recommendations for other regions or countries, and in case it makes statements like “our results suggest that this policy works/does not work” or “our results generalize to”.

15. It could of course be argued that NGOs can also be considered as “business-as-usual”, since many real-world interventions, especially in developing countries, are implemented by NGOs. However, for most of the 20 RCTs that were implemented by an NGO, the cooperating NGO was a rather small one and regionally limited in its activities. Thus, bringing the intervention to scale would be the task of either the government or a larger NGO with potential implications for the efficacy of the intervention.

16. This finding is in line with Eble, Boone, and Elbourne (2017) who review RCTs published between 2001 and 2011 for how they deal with different sorts of biases (also covering Hawthorne effects).

17. Details on these examples can be found in the review report on the respective paper in the online supplementary appendix.


19. See Moher et al. (2010) and Schulz, Altman, and Moher (2010).

References


**Appendix A: Reviewed Papers and Ratings**
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Aker et al. (2012)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>NGO</td>
<td>Yes</td>
</tr>
<tr>
<td>Alatas et al. (2012)</td>
<td>Yes</td>
<td>Participants are NOT aware</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Government</td>
<td>No</td>
</tr>
<tr>
<td>Ashraf et al. (2010)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>NGO</td>
<td>No</td>
</tr>
<tr>
<td>Ashraf et al. (2014)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Researcher</td>
<td>No</td>
</tr>
<tr>
<td>Attanasio et al. (2011)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Government</td>
<td>Yes</td>
</tr>
<tr>
<td>Baird et al. (2011)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>NGO</td>
<td>No</td>
</tr>
<tr>
<td>Barrera-Osorio et al. (2011)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Regional Public Authority Firm</td>
<td>Yes</td>
</tr>
<tr>
<td>Bertrand et al. (2010)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Firm</td>
<td>No</td>
</tr>
<tr>
<td>Björkman and Svennsson (2009)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>NGO</td>
<td>Yes</td>
</tr>
<tr>
<td>Blattman et al. (2014)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Government</td>
<td>Yes</td>
</tr>
<tr>
<td>Blimpo (2014)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Researcher</td>
<td>Yes</td>
</tr>
<tr>
<td>Bloom et al. (2013)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Researcher</td>
<td>Yes</td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------------------------------------------------</td>
<td>-------------------------------</td>
<td>------------------------------------------</td>
<td>-------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------</td>
<td>-----------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Burde and Linden (2013)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>NGO</td>
<td>No</td>
</tr>
<tr>
<td>Casey et al. (2012)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Government</td>
</tr>
<tr>
<td>Chinkhumba et al. (2014)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Researcher</td>
<td>Yes</td>
</tr>
<tr>
<td>Cohen and Dupas (2010)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Researcher</td>
<td>Yes</td>
</tr>
<tr>
<td>Collier and Vicente (2014)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>NGO</td>
<td>Yes</td>
</tr>
<tr>
<td>Das et al. (2013)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Researcher</td>
<td>Yes</td>
</tr>
<tr>
<td>de Mel et al. (2009a)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Researcher</td>
<td>No</td>
</tr>
<tr>
<td>de Mel et al. (2013)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Researcher</td>
<td>No</td>
</tr>
<tr>
<td>Drexler et al. (2014)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Firm</td>
<td>No</td>
</tr>
<tr>
<td>Duflo et al. (2011a)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>NGO</td>
<td>Yes</td>
</tr>
<tr>
<td>Duflo et al. (2011b)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>NGO</td>
<td>Yes</td>
</tr>
<tr>
<td>Duflo et al. (2012)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>NGO</td>
<td>Yes</td>
</tr>
<tr>
<td>Duflo et al. (2013)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Regional Public Authority</td>
<td>No</td>
</tr>
</tbody>
</table>

Burde and Linden (2013) No N/A No No Yes No No NGO No
Casey et al. (2012) No N/A No Yes Yes Yes No Government No
Chinkhumba et al. (2014) Yes No Yes No Yes Yes No Researcher Yes
Cohen and Dupas (2010) No N/A Yes Yes Yes Yes No Researcher Yes
Collier and Vicente (2014) No N/A No No Yes Yes No NGO Yes
Das et al. (2013) No N/A Yes Yes Yes Yes No Researcher Yes
de Mel et al. (2009a) Yes No No No Yes Yes No Researcher No
de Mel et al. (2013) No N/A Yes No No Yes Yes NGO Yes
Drexler et al. (2014) No N/A No No Yes No No Firm No
Duflo et al. (2011a) No N/A Yes No Yes Yes Yes NGO Yes
Duflo et al. (2011b) No N/A Yes No Yes No No NGO Yes
Duflo et al. (2012) Yes No No Yes Yes Yes Yes NGO Yes
Duflo et al. (2013) No N/A No Yes Yes Yes No Regional Public Authority Yes
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dupas (2011)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>NGO</td>
</tr>
<tr>
<td>Dupas and Robinson (2013a)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Firm</td>
</tr>
<tr>
<td>Dupas and Robinson (2013b)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Researcher</td>
<td>Yes</td>
</tr>
<tr>
<td>Dupas (2014)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Researcher</td>
<td>Yes</td>
</tr>
<tr>
<td>Feigenberg et al. (2013)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Firm</td>
<td>Yes</td>
</tr>
<tr>
<td>Field et al. (2013)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Firm</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Fujiwara and Wantchekon (2013)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Researcher</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Gné et al. (2010)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Firm</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Gné et al. (2012)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Government</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Glewwe et al. (2009)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>NGO</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Glewwe et al. (2010)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>NGO</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Hanna et al. (2014)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Researcher</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------------------------------------</td>
<td>-------------------------------</td>
<td>------------------------------------------</td>
<td>-------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>----------------------------------</td>
<td>-------------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Jensen (2010)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Researcher</td>
</tr>
<tr>
<td>Jensen (2012)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Researcher</td>
</tr>
<tr>
<td>Jensen and Miller (2011)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Government</td>
</tr>
<tr>
<td>Karlan et al. (2014)</td>
<td>Yes</td>
<td>Participants are NOT aware</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Government</td>
</tr>
<tr>
<td>Karlan and Valdivia (2011)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>NGO</td>
<td>No</td>
</tr>
<tr>
<td>Kremer et al. (2011)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>NGO</td>
<td>No</td>
</tr>
<tr>
<td>Kremer et al. (2009)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>NGO</td>
<td>Yes</td>
</tr>
<tr>
<td>Macours et al. (2012)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Government</td>
<td>No</td>
</tr>
<tr>
<td>Macours and Vakis (2014)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Government</td>
<td>No</td>
</tr>
<tr>
<td>Muralidharan and Venkatesh (2010)</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>NGO</td>
<td>No</td>
</tr>
<tr>
<td>Muralidharan and Venkatesh (2011)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>NGO</td>
<td>No</td>
</tr>
<tr>
<td>Olken et al. (2014)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Government</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------------------------------------</td>
<td>--------------------------------</td>
<td>------------------------------------------</td>
<td>-------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>Oster and Thornton (2011)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Researcher</td>
</tr>
<tr>
<td>Pradhan et al. (2014)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Firm</td>
</tr>
<tr>
<td>Robinson (2012)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Researcher</td>
</tr>
<tr>
<td>Tarozzi et al. (2014)</td>
<td>No</td>
<td>N/A</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Firm</td>
</tr>
<tr>
<td>Vicente (2014)</td>
<td>No</td>
<td>N/A</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Government</td>
</tr>
</tbody>
</table>
### Appendix B: Excluded Papers and Reason for Exclusion

<table>
<thead>
<tr>
<th>Author</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adhvaryu (2014)</td>
<td>Quasi-experiment</td>
</tr>
<tr>
<td>Armantier and Boly (2013)</td>
<td>Artefactual experiment</td>
</tr>
<tr>
<td>Ashraf (2009)</td>
<td>Behavioral Field experiment</td>
</tr>
<tr>
<td>Attanasio et al. (2012)</td>
<td>Artefactual experiment</td>
</tr>
<tr>
<td>Bauer et al. (2012)</td>
<td>Behavioral Field experiment</td>
</tr>
<tr>
<td>Beaman and Magruder (2012)</td>
<td>Artefactual experiment</td>
</tr>
<tr>
<td>Beaman et al. (2009)</td>
<td>Natural experiment</td>
</tr>
<tr>
<td>Besley et al. (2012)</td>
<td>Theoretical paper</td>
</tr>
<tr>
<td>Bursztyn and Coffman (2012)</td>
<td>Natural experiment</td>
</tr>
<tr>
<td>Cai et al. (2009)</td>
<td>Behavioral field experiment</td>
</tr>
<tr>
<td>Chassang et al. (2012)</td>
<td>Theoretical paper about RCTs</td>
</tr>
<tr>
<td>De Mel et al. (2009b)</td>
<td>Reply to a previously published article</td>
</tr>
<tr>
<td>DiTella and Schargrodsky (2013)</td>
<td>Natural experiment</td>
</tr>
<tr>
<td>Gneezy et al. (2009)</td>
<td>Artefactual experiment</td>
</tr>
<tr>
<td>Hjort (2014)</td>
<td>Natural experiment</td>
</tr>
<tr>
<td>Karlan and Zinman (2009)</td>
<td>Behavioral field experiment</td>
</tr>
<tr>
<td>Lucas and Mbiti (2014)</td>
<td>Quasi-experiment</td>
</tr>
<tr>
<td>Voors et al. (2012)</td>
<td>Artefactual experiment</td>
</tr>
</tbody>
</table>
Privatization in Developing Countries: What Are the Lessons of Recent Experience?

Saul Estrin and Adeline Pelletier

This paper reviews the recent empirical evidence on privatization in developing countries, with particular emphasis on new areas of research such as the distributional impacts of privatization. Overall, the literature now reflects a more cautious and nuanced evaluation of privatization. Thus, private ownership alone is no longer argued to automatically generate economic gains in developing economies; pre-conditions (especially the regulatory infrastructure) and an appropriate process of privatization are important for attaining a positive impact. These comprise a list which is often challenging in developing countries: well-designed and sequenced reforms; the implementation of complementary policies; the creation of regulatory capacity; attention to poverty and social impacts; and strong public communication. Even so, the studies do identify the scope for efficiency-enhancing privatization that also promotes equity in developing countries.

There is a large body of literature about the economic effects of privatization. However, since it was mainly written in the 1990s, there was typically limited emphasis on issues which have come to the fore more recently, as well as more recent developments in the evidence about privatization itself, much of it from developing economies. This motivated us to write this paper, which summarizes the evidence about the impact of recent privatizations, not only in terms of firms’ efficiency but also with regard to the effects on income distribution. In addition, we are particularly attentive to the process of privatization in developing countries, notably with respect to the regulatory apparatus enabling successful privatization experiences.

When governments divested state-owned enterprises in developed economies, especially in the 1980s and 1990s, their objectives were usually to enhance economic efficiency by improving firm performance, to decrease government intervention and
increase its revenue, and to introduce competition in monopolized sectors (Vickers and Yarrow 1988). Much of the earlier evidence about the economic impact of privatization concerned these topics and was based on data from developed countries and later, transition countries. These findings have been brought together in two previous surveys, by Megginson and Netter (2001) and Estrin et al. (2009) respectively. The former assesses the findings of empirical research on the effects of privatization up to 2000, mainly from developed and middle-income countries, while the latter concentrates on transition economies including China, over the 1989 to 2006 period. However, the experiences from the wave of privatizations that have occurred in developing countries before and since these studies warrant a new examination of the impact of privatization in the context of the development process.

The tone of the privatization debate has evolved in recent years in international financial institutions as privatization activity has shifted towards developing economies, and as a consequence of the difficulties of implementation and some privatization failures in the 1980s and 1990s (Jomo 2008). As a result, more emphasis in policy-making is now being placed on creating the preconditions for successful privatization. Thus, in place of a simple pro-privatization bias characteristic of the Washington consensus (Boycko, Shleifer, and Vishny 1995), it is now proposed that governments should first provide a better regulatory and institutional framework, including a well-functioning capital market and the protection of consumer and employee rights. In other words, context matters: ownership reforms should be tailor-made for the national economic circumstances, with strategies for privatization being adapted to local conditions. The traditional privatization objective of improving the efficiency of public enterprises also remains a major goal in developing countries, as does reducing the subsidies to state-owned enterprises (SOEs).

This article therefore reviews the recent evidence on privatization, with an emphasis on developing countries. The first section presents some stylized facts. The next section examines the effects of privatization in terms of firms’ efficiency and performance. In the following section, we go on to examine the distributional impacts of privatization. Policy recommendations are developed in the final section.

Privatization Trends: Stylized Facts

Privatization Trends Since the Late 1980s

The data on privatization prior to 2008 (with a regional breakdown) is sourced from the World Bank Privatization database but unfortunately this was discontinued in 2008 and no consolidated data is available after that date. Since we have not been able to find disaggregated data post-2008, we therefore present world aggregates, based on the Privatization Barometer database.

The early literature focused on developed economies and Western Europe represented roughly one-third of global privatization proceeds over the period 1977 to
2002 (Roland 2008). Even so, many of these deals only concerned minority stakes of SOEs (Bortolotti and Milella 2008). There were also spectacular numbers of privatizations during the transition process after 1990 in Central and Eastern Europe, with proceeds totaling $240 billion to 2008, in addition to widespread free or subsidized allocation of shares in former SOEs (Estrin et al. 2009). The revenues from privatization have been more limited in Africa, the Middle East and South Asia, with total proceeds below $50 billion for each (see figure 1). However, proceeds are on par with or above Europe once they are expressed as a percentage of GDP.

For the rest of Asia, the picture is rather different. While South Asia has experienced only a limited number of privatizations (especially India), this was not the case in East Asia, where total privatization proceeds represented 30% of the world’s total ($230 billion) over the 1988 to 2008 period. China, in particular, stands out. Over a 25-year period, the Chinese government has encouraged innovative forms of industrial ownership, especially at the subnational level, that combine elements of collective and private property (Brandt and Rawski 2008). New private entry and foreign direct investment have also been encouraged. As a result, by the end of the 1990s, the non-state sector accounted for over 60% of GDP and state enterprises’ share in industrial output had declined from 78% in 1978 to 28% in 1999 (Kikeri and Nellis 2004). The OECD estimated the state-owned share of GDP had further declined to 29.7% by 2006 (Lee 2009).

Finally, in Latin America and especially in Chile, large-scale privatization programs have been launched, especially in the infrastructure sector, starting in 1974 in Chile and peaking in the 1990s. Between 1988 and 2008, the total privatization proceeds in Latin America amounted to $220 billion (28% of total world proceeds).
One needs to be cautious, however, when interpreting the raw data because of differences in the size of economies. The differences between the privatization experience of Africa, Asia, and Europe become less striking when proceeds are normalized by GDP, though privatization revenue to GDP is high in Latin America, representing, on average, 0.5% of GDP over the period.

**Privatization Trends Since 2008**

The five years to 2015 have been marked by the predominant role of China in global privatizations, while the EU’s share has been below its long-term average of 45% of the world’s total proceeds, running at only one-third of worldwide totals, on average. According to the Privatization Barometer (PB) Report 2013–2014, global privatization total proceeds exceeded $1.1 trillion from January 2009 to November 2014, with $544 billion of divested assets between January 2012 and November 2014.\(^3\)

In addition, the 20-month period beginning in January 2014 witnessed privatizations totaling $431.4 billion (PB report 2015). This is far more than any comparable period since the beginning of the privatization programs in the U.K. in the late 1970s (see figure 2), though as noted below, a significant part of this was driven by the unwinding of positions taken in banks by governments during the financial crisis.

China has consistently been one of the top privatizers from 2009 to 2015; it was the second-largest privatizer in 2009 and the first in 2013, 2014, as well as the
8-month period of January to August 2015. Aggregate privatization deals in China totaled more than $40 billion in both 2013 and 2014 and a spectacular $133.3 billion in the first eight months of 2015 through 247 sales. The bulk of these privatization revenues came from the public and private placement offering of primary shares by SOEs (PB report 2015). However, the state’s equity ownership stake was generally only reduced indirectly, by increasing the total number of shares outstanding (PB report 2015). In fact, Hsieh and Song (2015) have shown that almost half of the state-owned firms in 2007 and nearly 60 percent of them in 2012 were legally registered as private firms. The term used in China for this ownership change is that the large state-owned firms are “corporatized” rather than privatized. The typical form this “corporatization” takes is that of a minority share traded in the stock market and merged into a large state-owned conglomerate, the controlling shareholder (Hsieh and Song 2015).

The next-leading country in terms of privatization proceeds after China is the United Kingdom, but it is far behind, with total proceeds of $17.2 billion in 2014 (against $7.8 billion in 2009).

In the EU as a whole, with countries addressing their government deficits post-2008, privatization proceeds rose to a five-year peak in 2013, to $68.0 billion, and a nine-year peak of $77.6 billion in 2014, while the annualized value of privatizations during 2015 (based on the first 8 months) reached $63.3 billion. This represents more than one-third of the global annual totals in 2014, but is only 20.0% of worldwide totals in the first 8 months of 2015, and lower than the long-run average EU share of about 44.6% (PB report 2015). This relative decline of EU privatization proceeds is also reflected in the fact that China alone generated revenues from privatization almost as great as did the EU countries combined during 2015 ($68.0 billion versus $77.6 billion for China; PB report 2015).

China and India were the two top emerging countries by total privatization revenues in 2015. The five largest single deals outside the developed world in 2014 were realized in China, with the recapitalization and primary share offering of CITIC Pacific Ltd, the private placement of BOE Technology Group, the primary-share initial public offering (IPO) of Dalian Wanda Commercial, and finally the primary-share IPO of CGN Power and of HK Electrical Investments Ltd.

In the following section, we focus on the privatization experience in Africa and South Asia. While the privatization programs in Eastern Europe, China, and Latin America are among the most important in terms of total proceeds, a rich literature already exists discussing them (see Estrin et al. 2009 on transition economies and Estache and Trujillo 2008 on Latin America). Moreover, while privatization in Latin America and Eastern Europe culminated in the 1990s, much privatization in Africa and South Asia is more recent (Roland 2008).
Privatization programs in sub-Saharan Africa (SSA) occurred in successive waves, with some countries privatizing much earlier than others (Bennell 1997). The first group to start such programs in the late 1970s to early 1980s was composed of francophone West African countries (e.g., Benin, Guinea, Niger, Senegal, and Togo) but their progress was limited. The second group, both Anglophone and Francophone countries (Ghana, Nigeria, Ivory Coast, Mali, Kenya, Malawi, Mozambique, Madagascar, and Uganda), started privatizing in the late 1980s. These programs were often influenced by pressure from the international financial institutions (Nellis 2008) though, as noted by Bennell (1997), no significant progress was made anywhere except Nigeria until the late 1990s. The final group, the “late starters”, did not begin to privatize until the early to mid-1990s. Among this group, Tanzania, Burkina Faso and Zambia have shown a strong political commitment to privatization, whereas in the other three countries (Cameroon, Ethiopia, and Sierra Leone), only minimal progress was made in the 1990s.

**Privatization in the 1990s: A Slow Start.**

Only a minority of SOEs in SSA were subject to privatization over the period 1991 to 2001, and very little privatization has taken place outside of South Africa, Ghana, Nigeria, Zambia, and Cote d’Ivoire (Nellis 2008). African states have privatized a smaller percentage (around 40%) of their SOEs than in Latin America and the transition economies (Nellis 2008). In addition, privatization has generally concerned smaller manufacturing, industrial, or service firms. Bennell (1997) also reports that smaller SOEs were usually targeted during the initial stages of privatization programs in SSA because they were easier to sell. Five industries in particular were prominent in most programs: food processing, alcoholic beverages, textiles, cement and other non-metallic products, and metal products. These industries accounted for 60% of the total proceeds from the sale of manufacturing SOEs during 1988 to 1995 (Bennell 1997), if we exclude the exceptional and large sale of ISCOR (Iron and Steel Industrial Corporation) in South Africa.

Bennell (1997) explains that the slow progress in privatization in the 1990s was due to a lack of political commitment compounded by strong opposition from entrenched vested interests (senior bureaucrats in ministries and SOEs themselves, as well as public sector workers concerned about their job security). For instance, in Cameroon, only five of the thirty SOEs scheduled for privatization were sold by the end of 1995. In other countries such as Nigeria, the privatization program started well but then stalled. Despite the fact that Nigeria’s program had been one of the most successful in SSA in the 1990s, it was suspended in early 1995 in favor of a mass program of “commercialization”. In Madagascar, the privatization program was also
suspended in mid-1993 due to serious mismanagement and its subsequent unpopularity. In addition, Bennell (1997) reports that there were nationalist concerns about the possible political and economic consequences of increased foreign ownership as a result of privatization.

However, in the late 1990s, certain political constraints lifted. First, a growing number of governments in SSA started to undertake significant economic reforms, under the aegis of the World Bank and the IMF, in which privatization was an integral part. Reforms and privatization were also progressively being accepted by the population. In addition, important political liberalization, with multi-party elections, broke with the previous statist policies, and created some room for maneuver to implement privatization programs. Finally, the weak financial position of SOEs in many SSA countries and their rapid deterioration, in conjunction with the fiscal crisis the state experienced in the 1990s, also opened the way for a sell-off of SOEs to raise government revenues and reduce expenditures.

Despite this stronger commitment, Nellis (2008) notes that there were actually only a few privatization deals in Africa in the 1990s, mainly in infrastructure, and even in these the state retained significant minority stakes; around one-third of the shares on average were retained. Between 1988 and 1999, the total proceeds from privatization in SSA amounted to $9.8 billion, with the manufacturing and services sector accounting for 36% of the total, infrastructure 28%, the energy sector 17%, the primary sector 14%, and the financial (and other) sector 6% (see World Bank Privatization Database).

**The Early to Mid-2000s; More Rapid Progress.**

There were some important privatizations in SSA between 2000 and 2008, and total proceeds increased to $12.654 billion (see World Bank Privatization Database). Nigeria comprised 51% of this amount, followed by Kenya (10%), Ghana (9%) and South Africa (6%). Infrastructure represented 73% of the total amount of the deals, followed by the manufacturing and services sector (17%), the financial sector (6%), energy (4%) and the primary sector (1%; see World Bank Privatization Database).

**Privatization Post-2008: A Slowdown.**

Privatization activity slowed in SSA with the economic downturn after 2008. One notable exception was Benin, with the privatization of the cotton and the public utility sectors. The concession for the operation of the container terminal of the Port of Cotonou and the majority stake in the cement company were awarded to a strategic private investor in September 2009 and March 2010, respectively, and the privatization of Benin Telecom was launched in 2009 (this is still ongoing: IMF 2010). Nigeria was also notable for its sale of 15 electricity-generating and distribution companies in 2013, raising $2.50 billion (see Megginson 2014). In Chad, the government
announced in 2015 that it was re-launching the sale of 80% of Société des Télécommunications du Tchad (Sotel-Tchad), after the previous attempt collapsed in 2010. Because the World Bank Privatization Database does not have data on privatization after 2008, one cannot compare the aggregated privatization proceeds post-2008 to those of earlier decades.

Privatization in South Asia: A Slow Opening

Privatizations in South Asia have traditionally been rare, despite the notable inefficiency of SOEs (Gupta 2008). The governments’ reluctance to privatize can be partly explained historically, with the government’s close involvement in the establishment of an industrial base in the postcolonial era, especially in India (Gupta 2008). Particular sectors had been reserved exclusively for SOEs, such as the infrastructure sector and capital goods and raw materials industries such as steel, petroleum, and heavy machinery. In addition, the government nationalized many loss-making private companies; more than half of the firms owned by the Indian federal government were loss-making in the 1990s.

Following the balance of payment crisis of 1991, the Indian government implemented a series of reforms under the Industrial Policy Resolution of 1991 to encourage private enterprise. Privatization was initiated mainly through two approaches: partial privatization and strategic sales. However, the former was very limited, with the government selling only minority equity stakes until 2000, and without transferring management control. Political uncertainty prevented the emergence of a coherent privatization policy. Majority stakes sales and the transfer of management control were only conducted after the elections of 1999, and even then, until 2004 the government retained an average ownership stake of 82% in all SOEs (Gupta 2008).

The stalled privatization program was revived in 2010 with a secondary offering of shares in National Thermal Power Corporation Ltd (NTPC), which owns 20% of India’s power generation capacity (Gupta 2009). However, the sale of the $1.85 billion block of shares only reduced the government’s stake by an additional 5%, leaving 85% still under government control. In addition, the process of privatization was viewed as poor, with the secondary offering subscribed only 1.2 times, and even this after assistance from government-owned financial institutions.

In summary, between 2000 and 2008, the proceeds of privatization in South Asia totaled $17.45 billion, the bulk being realized in India (see figure 3) (55%) followed by Pakistan (43%). Afghanistan, Bangladesh, Nepal, and Sri Lanka provided the remaining 2% (see World Bank Privatization Database). Between 2000 and 2008, the infrastructure sector represented 51% of the proceeds, followed by the energy sector (26%), the financial sector (12%), manufacturing and services (10%), and the primary sector (2%) (see World Bank Privatization Database).
The Effects of Privatization: Efficiency and Firm Performance

Overall, as we report below, the studies on developing economies show that a move from state to private ownership alone does not automatically yield economic gains. Rather, a number of factors have been found to influence the success of privatization, namely:

- Which firms are privatized; there can be a positive (or negative) selection effect.
- Whether privatization is total or partial; evidence suggests that the former is more beneficial.
- The regulatory framework, which in turn depends on the institutional and political environment.
- The characteristics of the new owners; foreign ownership has been associated with superior business performance post-privatization, especially relative to “insider” ownership (privatization to managers and workers).\(^9\)
- Effective competition. This has been found to be critical in bringing about improvements in company performance because it is associated with lower costs, lower prices, and higher operating efficiency.\(^{10}\)

In the following sub-sections, we introduce the estimation techniques that have been used to measure the impact of privatization on firms’ performance, and then examine privatization experiences in three sectors (banking, telecommunications, and...
utilities) in developing countries. We also provide an analysis of the robustness of the evidence in the literature about the impact of privatization.

Measuring Efficiency and Firms’ Performance Post-Privatization

As Megginson and Sutter (2006) note, researchers face numerous methodological problems when they analyze the economic effects of privatization. In particular, data availability and consistency, especially in developing countries, and sample selection bias—occurring, for example, if the “best” firms are privatized first—represent key issues. Other problems arise when using accounting data: the determination of the correct measure of operating performance, the selection of an appropriate benchmark and statistical tests are important challenges. These issues are germane to the interpretation of the results of the studies surveyed below.

A variety of methods have been used to measure the impact of privatization on firms’ post-privatization performance and efficiency, measured in a number of ways including return on equity, output growth, labor productivity and changes in cost and income. We distinguish between two different empirical approaches. The first consists of comparing the performance of government-owned firms to that of privately-owned firms. The second approach consists of comparing pre- and post-divestment performance for companies privatized via share issues (public offerings; Megginson, Nash, and van Randenborgh methodology).

Comparing Government-owned Firms to Privately-owned Firms

An obvious way to examine the impact of privatization is to compare the performance of government-owned to privately-owned firms. Studies in this tradition compare post-privatization performance changes with either a comparison group of non-privatized firms or with a counterfactual. However, important methodological issues arise, especially in the earlier studies. First, it is difficult to determine the appropriate set of comparison firms, especially in developing countries where the private sector is limited. Second, selection effects and endogeneity may bias the comparison, as factors determining whether the firm is publicly or privately owned are also likely to affect performance (Gupta, Ham, and Svejnar 2008).


One of the first studies to compare SOE and private firm performance is that of Ehrlich et al. (1994). These authors used a sample of 23 comparable international airlines (18 from developed countries and 5 from developing/emerging countries) of different ownership categories over the period 1973 to 1983 for which they have data on cost and output for comparable goods. These authors find a
significant association between ownership and firm-specific rates of productivity growth. Interestingly, the empirics also suggest that the benefits derive primarily from complete privatization of the firm, and that a partial change from state to private ownership has little effect on long-run productivity growth. Other studies have employed a similar approach examining differences in efficiency between private and government-owned firms within a specific country, such as Majumdar (1996) for Indian firms and Tian (2000) with Chinese firms. These authors both find that private-sector firms are more efficient. However, these results are not highly robust from the perspective of contemporary methods, as they do not directly address selection issues.

Concerning studies using a counterfactual approach, one can cite the influential study by Galal et al. (1994), which was sponsored by the World Bank. These authors compare the actual post-privatization performance of twelve large firms in the airlines and utilities industry in Britain, Chile, Malaysia, and Mexico to a counterfactual performance. Further, they estimate net welfare gains in eleven of the twelve cases considered, equaling on average 26 percent of the firms’ pre-divestiture sales. La Porta and Lopez-de-Silanes (1999) study privatization in Mexico and find that privatized Mexican SOEs rapidly close a large performance gap with industry-matched private firms that had existed prior to divestment. These authors find that output increases by over 50% and that the privatized firms reduce employment by half, while the remaining workers see a significant pay rise.

Cross-country, Multi-industry Comparisons of X-efficiency and Profitability Ratio of Private and Government-owned Firms.

Another approach has been to exploit a multi-industry, multi-national cross-sectional time series to analyze the effects of government ownership on efficiency. The advantage of this method is that it captures differences that are not apparent in single-country or single-industry series, and the results are therefore methodologically more soundly based. In their seminal work, Boardman and Vining (1989) use measures of X-efficiency and profitability ratios of the 500 largest non-U.S. manufacturing and mining corporations in 1983 (“The International 500”; Fortune 1983). Privately-owned firms are found to be significantly more profitable and productive than state-owned and mixed ownership enterprises, but mixed enterprises are no more profitable than SOEs. Another important study is that of Frydman et al. (1999), which compares the performance of privatized and state firms in the transition economies of Central Europe in 1994 using a fixed-effects model. To control for the possibility that better firms are selected for privatization, these authors compare the pre-privatization performance of managerially-controlled firms with those controlled by other owners. Frydman et al. (1999) find that privatized firms perform better than the state-owned firms but that the performance improvement is related to revenue improvement rather than cost reduction in privatized firms.
As noted, governments sequence privatizations strategically, often leading the most profitable firms to be privatized first (Gupta, Ham, and Svejnar 2008; Dinc and Gupta 2011). To control for selection and endogeneity biases, the latest studies have employed more advanced econometric techniques including differences in difference, triple differences matching methods, and instrumental variable methods.

For instance, Dinc and Gupta (2011) examine the influence of political and financial factors on the decision to privatize government-owned firms in India using data from the 1990–2004 period. They find that profitable firms and firms with a lower wage bill are likely to be privatized early and that the government delays privatization in regions where the governing party faces more competition from opposition parties. The results therefore suggest that firms’ financial characteristics have a significant impact on the government’s decision to privatize. This raises an identification issue for evaluating the effect of privatization on firm performance: if more profitable firms are more likely to be privatized, we may overstate the impact of privatization on profitability when we compare the performance of government-owned to that of privatized firms. The authors then proceed to use political variables as instruments for the privatization decision, adopting a two-stage least squares treatment effects regression. After addressing the selection bias, they find that privatization still has a positive impact on performance in India.

Comparing Pre-post Divestment Sales and Income Data for Companies Privatized by Public Share Offering

This set of studies examines the effects of privatization on firm performance by comparing pre- and post-divestment data for companies privatized via public share offerings. Each firm is compared to itself (a few years earlier) using inflation-adjusted sales and income data. The first study using this methodology is by Megginson, Nash, and van Randenborgh (1994). As Megginson and Netter (2001) note, this methodology suffers from several drawbacks, among which selection bias is probably the greatest concern, since privatizations through share sales—Share Issue Privatization (SIPs)—represent the largest companies sold during a privatization program. Another weakness is that the Megginson, Nash, and van Randenborgh methodology can only examine simple accounting variables (assets, sales, etc.), which is an issue when comparing accounting information at different points in time and in different countries. Most of the studies in this tradition also imperfectly account for macroeconomic or industry changes in the pre- and post-privatization window (see Megginson and Netter 2001, for a critique). These studies also cannot account for the impact on privatized firms of regulatory or market-opening initiatives that are often launched in parallel with privatization programs. However, the Megginson, Nash, and van Randenborgh methodology allows the analysis of large samples of firms from different industries, countries, and time periods and, while carrying the risk of
selection bias, SIP samples contain the largest and most (politically) important privatizations.

Most of these studies do identify a significant improvement in company performance, post-privatization, though methodological reservations remain. Research in this tradition has focused on specific industries (banking [Verbrugge, Owens, and Megginson 2000] and telecommunications [D’Souza and Megginson 2000]); has used data from a single country (Chile [Maquieira and Zurita 1996]) and employed multi-industry, multinational samples. However, the significance of many of the operating and financial improvements is not robust to adjustments for changes experienced by other firms over the study period.

A very recent work by Li et al. (2016) overcome the empirical limitations of the previous SIPs studies mentioned above by employing a triple difference approach. The authors are able to separate the pure privatization effect from the listing effect, using a database of 204 Chinese SIPs from 1999 to 2009 matched with otherwise comparable state-owned enterprises and privately-owned firms. The first double-difference compares the performance change of SIP firms before and after listing with the performance change of a control group of fully state-owned and unlisted SOEs to capture the combined “SIP effect” of going public and privatizing. The second double-difference compares the performance change of privately-owned firms before and after their listing with the performance change of a control group of privately-owned firms that remain unlisted. This captures the “pure listing effect”. These authors obtain the “pure privatization effect” by taking the difference between these two double differences. Interestingly, they continue to find a positive impact from privatization using this exacting methodology: they find a significant positive increase in profitability post-SIP in divested Chinese state-owned companies, even after the negative IPO listing effect is taken into account.

Empirical Evidence to Date in Developing Countries

In this section, we summarize the empirical evidence to date about the effects of privatization on firms’ performance and efficiency in developing countries, drawing on the discussion of methodology outlined above. The sectors covered include banking, telecommunications, and utilities. To examine the reliability of the evidence in drawing policy conclusions, we classify the papers reviewed into four categories depending on the quality of the sample and the robustness of the methods used.

The Banking Sector

The studies reviewed by Clarke, Cull, and Shirley (2005), which focus on developing countries and employ the Megginson, Nash, and van Randenborgh methodology or a stochastic frontier approach, find that bank performance usually improved
after privatization. For instance, Boubakri et al. (2005), applying the Megginson, Nash, and van Randenborgh methodology to analyze 81 bank privatizations in 22 low- and middle-income countries, find that some measures of performance improved after privatization, but that this pattern was not common across countries; environmental factors also played a role. The study by Beck, Cull, and Jerome (2005) in Nigeria shows that privatization can improve bank performance, even when the macroeconomic and regulatory environment is inhospitable and the government sells the weakest banks. However, Beck, Cull, and Jerome argue that an adverse macroeconomic and regulatory environment reduces the benefits of privatization. Azam, Biailis, and Dia (2004) also show (both theoretically and empirically) the benefits of having a strong, independent regulatory agency to ensure that privatized banks play an efficient role in financial development.

The studies surveyed by Clarke, Cull, and Shirley (2005) also find that bank privatization has a greater positive effect when it is total rather than partial. This result has been found in transition countries (Bonin, Hasan, and Wachtel 2005) as well as in Brazil (Beck, Crivelli, and Summerhill 2005) and Nigeria (Beck, Cull, and Jerome 2005). Furthermore, there is evidence that privatization boosts competition in the banking sector. For instance, Ochere (2005) examines share-issue privatizations in nine countries using the Megginson, Nash, and van Randenborgh methodology and finds that rival banks suffered abnormally negative returns following privatization announcements, which suggests that shareholders expected more intense competition and lower returns.

Thus, evidence suggests that performance improves more when the government fully relinquishes control; when banks are privatized to strategic investors rather than through share issues; and when bidding is open to all, including foreign banks (Clarke, Cull, and Shirley 2005; Megginson 2005). A more recent paper by Clarke, Cull, and Fuchs (2009), which examines the privatization of Uganda Commercial Bank (UCB) to the South African bank Stanbic, shows that these elements of best practice also apply when the banking sector is concentrated and under-developed. The government fully relinquished control to a strategic investor in an open sales process that allowed foreign participation, and the authors found that profitability improved post-privatization with no evidence that outreach declined. A similar impact of privatization to a foreign bank has been found in the case study of the privatization of Tanzania’s national bank of commerce to the Dutch Rabobank (Cull and Spreng 2011).

The Telecommunications Sector

One of the first telecom studies focused on developing countries, by Wallsten (2001), used a panel of 30 African and Latin American countries from 1984 to 1997 with a methodology similar to Megginson, Nash, and van Randenborgh. Overall, the author
finds that competition is significantly associated with increases in per capita access and decreases in costs. However, privatization alone is associated with few benefits, and is negatively correlated with connection capacity. In addition, privatization only improves performance when coupled with effective and independent regulation and increases in competition.

More recently, Gasmi et al. (2013) have examined the impact of privatization of the fixed-line telecommunications operator on sector performance, analyzing the outcomes of privatization reforms in a 1985 to 2007 panel dataset on a selection of 108 countries (including OECD countries, Asia, Africa, Latin America). These authors find that the impact of privatization on sector outcomes (fixed-line deployment, cellular deployment, labor efficiency, price of fixed-line) was positive in the OECD countries, Central America, and the Caribbean, and in resource-scarce coastal Africa and Asia. However, the impact was negative in South America and in African resource-scarce landlocked countries, and no significance was identified in resource-rich African countries.

Gasmi et al. (2013) note that countries with successful privatizations have developed their infrastructure through the creation of appropriate institutional structures which have improved the effectiveness of infrastructure policies, and that the coverage of networks increased thanks to the additional capital available with privatization. In contrast, privatization outcomes proved to be poor in South America, in both resource-scarce landlocked African countries and resource-rich African countries due to weak contractual design and inadequate enforcement of policies in the infrastructure sector, as well as insufficient aggregate demand. In the absence of strong state capacity, competition appeared to be a more effective instrument to foster performance than privatization.

The extent of infrastructure privatization also diverged across regions. While almost all OECD countries have privatized their telecommunications utilities, the rate of privatization is only around 70% in Latin American, Asian, and African resource-scarce coastal countries. In African resource-scarce landlocked and resource-rich countries, the percentage of privatized infrastructure in telecommunications is even lower, at around 40% and 30%, respectively. Overall, the study by Gasmi et al. (2013) shows that there were limited privatization effects on network expansion, and that productive efficiency did not increase in all the regions post-privatization. As such, the authors conclude that there is no unique model of reform for infrastructure sectors.

The Utilities Sector

Turning to water privatization, Estache and Rossi (2002) estimate a stochastic cost frontier using 1995 data from a sample of 50 water companies in 29 Asian and Pacific countries. These authors find that efficiency is not significantly different in
private and public companies. Kirkpatrick, Parker, and Zhang (2006) use a questionnaire survey on water utilities in Africa, covering 13 countries and 14 utilities that reported private sector involvement, and undertake data envelopment analysis and stochastic cost frontier techniques. These authors do not find strong evidence of performance differences between state-owned water utilities and water utilities involving some private capital. The authors consider that this result is related to the technology of water provision, the costs of organizing long-term concession agreements, and regulatory weaknesses. In particular, the authors argue that the nature of the product severely restricts the potential for competition and therefore the efficiency gains. This means rivalry under privatization must derive from the form of competition for the market—competition to win the contract or concession agreement. But, as the authors explain, transaction costs can be high in the process of contracting for water services provision; for example, the costs of organizing the bidding process, monitoring contract performance, and enforcing contract terms where failures are suspected. The importance of transparent competition for the market to achieve efficiency gains and prevent the grabbing of assets by political cronies was also evidenced by more recent research by Tan (2012) in the context of private participation in infrastructure (PPI) in water in Malaysia. The author shows that the efficiency gains of water privatization (measured by water loss and unit costs) were inconclusive over the period 2001 to 2008. Despite this, and the subsequent renationalization of water assets, PPI continues to be promoted—it is being recast in the form of management contracts—because it provides captive rents. This is also evidenced in the “cherry-picking” of segments and areas for privatization: private sector participation is concentrated in the more lucrative water treatment segment and higher income states, leaving the less profitable segments and (more rural) areas to the public sector.

In terms of privatizing electricity, the study of Zhang, Parker, and Kirkpatrick (2008) provides an econometric assessment using panel data for 36 developing and transition countries over the period 1985 to 2003. These authors examine the impact of these reforms on generating capacity, electricity generated, labor productivity in the generating sector, and capacity utilization. They find that, overall, the gains in economic performance from privatization and regulations are limited, while introducing competition is more effective to stimulate performance. In particular, they do not find that privatization leads to improved labor productivity or to higher capital utilization, or to more generating capacity and higher output, except when it is coupled with the establishment of an independent regulator. The authors conclude that when competition is weak, an effective regulatory system is needed to stimulate performance, while the regulation of state-owned enterprises without privatization is ineffective.

A more recent study by Balza, Jimenez, and Mercado (2013) examines the relationship between private sector participation, institutional reform, and performance of the electricity sector in 18 Latin American countries over the last four decades.
This also finds that, regardless of the level of private participation, well-designed and stable sectoral institutions are essential for improving the performance of the electricity sector. In particular, privatization is robustly associated with improvements in quality and efficiency, but not with accessibility to the service. In contrast, regulatory quality is strongly associated with better performance in terms of both quality and accessibility.

Summary

To bring together this evidence and evaluate its robustness as a basis for policy, we classify the papers reviewed in this section into four categories depending on the quality of the sample and the robustness of the methods used. Category I: single country data, basic statistics, or econometrics (or small sample). Category II: cross-country data, basic statistics, or econometrics (or small sample). Category III: single country data, more advanced econometric techniques. Category IV: cross-country data, advanced econometric techniques. The findings are reported in table 1 and taken together, provide qualified evidence that privatization can improve company performance, including from studies that use the most advanced econometric methods.

Thus, the evidence from empirical studies of privatization in developing countries suggests that the performance of banks improved significantly after privatization in many cases. However, the gains from privatization in the utilities sector (electricity and water) have tended to be limited. Finally, concerning the telecommunications sector, the impact of privatization on efficiency and coverage varies by region. It has been shown to be positive in Central America and in resource-scarce coastal Africa and Asia, but negative in South America and in African resource-landlocked countries. Thus, the impact appears to be context- as well as sector-specific. The main factors explaining this variation are regulatory quality (and behind that the quality of institutions), heterogeneity in effective competition, differences in the detail of contractual design, and in the characteristics of the new owners.

Privatization Process: Distributional Impacts

Thomas Piketty’s recent book (2014), which has highlighted the importance of income distribution in the growth process, also discussed the impact of privatization on capital accumulation. In principle, privatization need not affect the stock of wealth in an economy, nor its distribution. State-owned firms are public assets which earn a return for their owners. Provided the assets to be privatized are valued in such a way that their price represents the discounted sum of the profits to be earned from them, then privatization means that the state is replacing an income stream with its discounted capital value in its asset portfolio. At the same time, the private sector is purchasing an asset which generates its full value over time from its annual earnings.
Table 1. Methodology and Classifications of Empirical Papers

<table>
<thead>
<tr>
<th>Authors</th>
<th>Method</th>
<th>Data</th>
<th>Results</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beck, Crivelli, and Summerhill 2005</td>
<td>Measures of performance: ROE, ROA, overhead costs/assets</td>
<td>Brazil, unbalanced panel of 207 banks with quarterly data over the period January 1995 to September 2003, with a total of 4,864 observations.</td>
<td>Privatised banks increased their performance, but not restructured banks.</td>
<td>III</td>
</tr>
<tr>
<td>Authors</td>
<td>Method</td>
<td>Data</td>
<td>Results</td>
<td>Category</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Bonin, Hasan, and Wachtel</td>
<td>Measures of performance: cost and profit efficiency, ROA Four ownership types: foreign greenfield, domestic de novo, state-owned, privatised. Stochastic frontier analysis (SFA) to estimate bank efficiency.</td>
<td>Transition countries (Bulgaria, the Czech Republic, Croatia, Hungary, Poland, and Romania); 67 different banks from 1994 to 2002 (451 observations).</td>
<td>Foreign-owned banks are most efficient, and government-owned banks are least efficient. Voucher privatization does not lead to increased efficiency and early-privatised banks are more efficient than later-privatised banks (and no evidence of selection effect).</td>
<td>IV</td>
</tr>
<tr>
<td>Boubakri et al. (2005)</td>
<td>Measures of performance: ROE, net interest margin, credit risk. Examine three categories of controlling owners: foreign investors, local industrial groups, and the government itself. Megginson, Nash, and van Randenborogh methodology on a panel of banks. Period of seven years: three years prior to privatization and three years post-privatization, including the year of privatization itself.</td>
<td>81 bank privatizations occurring between 1986 and 1998, in 22 low-and middle-income countries.</td>
<td>Profitability increases post-privatization, but it depends on the type of owner (higher economic efficiency exhibited by banks owned by local industrial groups and foreign owners).</td>
<td>IV</td>
</tr>
</tbody>
</table>
Table 1. Continued

<table>
<thead>
<tr>
<th>Authors</th>
<th>Method</th>
<th>Data</th>
<th>Results</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otchere (2005)</td>
<td>Measures of performance: CAMEL criteria (Capital adequacy, Asset quality, Management efficiency, Earnings ability and Labor (employment levels and productivity). Stock market data. Megginson, Nash, and van Randenborgh methodology: 3 years pre-privatization operating performance data and 5 years post privatization. Examines pre- and post-privatization operating performance of the privatised banks relative to that of the rival banks.</td>
<td>Analyze 21 privatizations (and 65 rival banks) from middle- and low-income countries.</td>
<td>Statistically significant improvement in operating performance for the privatized banks in the pre- and post-privatization period, apart from reduction in loan loss provisions ratio. One reason for the lack of improvement might be the continued government ownership of these banks.</td>
<td>IV</td>
</tr>
<tr>
<td>Cull and Spreng (2011)</td>
<td>Measures of performance: ROA, NPL. Examines the privatization of National Bank of Commerce. Test whether the privatization of the two successor banks to the original National Bank of Commerce resulted in improved performance.</td>
<td>42 banks operating in Tanzania between December 1998 and December 2006.</td>
<td>Sale to a foreign strategic investor (Rabobank from the Netherlands) resulted in improved profitability and reductions in non-performing loans, along with an increase in the ratio of loans to total assets.</td>
<td>III</td>
</tr>
</tbody>
</table>
### Table 1. Continued

<table>
<thead>
<tr>
<th>Authors</th>
<th>Method</th>
<th>Data</th>
<th>Results</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Telecommunications</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Utilities - water</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estache and Rossi (2002)</td>
<td>Stochastic cost frontier</td>
<td>1995; 50 companies; 29 Asian-Pacific countries.</td>
<td>Efficiency is not significantly different in private companies than in public ones.</td>
<td>IV</td>
</tr>
<tr>
<td>Kirkpatrick, Parker, and Zhang (2006)</td>
<td>Stochastic cost frontier</td>
<td>2000; Africa; 76 observations, including 10 private-sector operations.</td>
<td>No strong evidence of differences in the performance of state-owned water utilities and water utilities involving some private capital in Africa.</td>
<td>IV</td>
</tr>
<tr>
<td>Authors</td>
<td>Method</td>
<td>Data</td>
<td>Results</td>
<td>Category</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td></td>
<td>water production capacity (the amount of water treated for distribution),</td>
<td>states.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>length of pipes. Case study (graphs and statistics). Different</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ownerships: public ownership, corporatized, public–private, private.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilities - electricity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhang, Parker, and Kirk-</td>
<td>Measures of performance: net electricity generation per capita of the</td>
<td>Panel data for 36 developing and</td>
<td>Competition seems to be most effective to increase performance. On their own privatization</td>
<td>II</td>
</tr>
<tr>
<td>patrick (2008)</td>
<td>population, installed generation capacity per capita of the population, net electricity generation per employee in the industry and electricity generation to average capacity (capacity utilization). The privatization variable used in the study was constructed as the percentage of generating capacity owned by private investors. Fixed effects (country and year) to deal with endogeneity.</td>
<td>transitional countries, over the period 1985 to 2003.</td>
<td>and regulation do not lead to significant improvement in performance.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balza, Jimenez, and Mercado</td>
<td>Measures of performance: real end-user prices for residential electricity (excluding taxes); percentage of households with access to electricity; electricity capacity generation; and electricity loss as a percentage of total electricity production. Privatization measured as the cumulative investment in the electricity sector as a percentage of average gross capital formation in the period 1984 to 2010.</td>
<td>1971 to 2012; 18 Latin American countries (panel of countries). Country-level analysis.</td>
<td>Countries with higher private investment tend to provide more efficient and better-quality electricity services.</td>
<td>II</td>
</tr>
</tbody>
</table>
Hence, privatization does not necessarily entail a net transfer of wealth between the public and private sectors.

However, the privatization process has not always followed these principles of public finance (Estrin et al. 2009). In the extreme, as in the programs in the Czech Republic or Russia, significant state assets were transferred to private hands at nominal or zero prices; in effect, a transfer of wealth from the state to the private sector. More generally, state assets have frequently been undervalued. This may have been in order to make the assets more attractive to the market, or because the SOEs were loss-making and the short-term requirement to balance the budget dominated long-term state asset portfolio criteria. In some cases, ideological arguments have also played a role; Margaret Thatcher and several of her admirers in transition economies viewed privatization as a policy mechanism for broadening the private ownership of shares in companies (Estrin 2002). Whatever the motivation, the undervaluation of state assets leads to a net redistribution of assets from state to private hands. Piketty argues that this was an important element in relatively larger growth of private wealth in Britain than in other Western European countries between 1970 and 2010. Furthermore, it was almost certainly a major factor in what he describes as the “considerable growth of private wealth in Russia and Eastern Europe.... which led in some cases to the spectacularly rapid enrichment of certain individuals (I am thinking of the Russian oligarchs),” (2014).

As the quotation from Piketty makes clear, the impact on income distribution of privatization depends on how the ownership of the assets is transferred from state into private hands; both the pricing and to whom the SOEs are privatized. In the extreme case when assets are transferred by voucher to each citizen equally from the state to private hands at a zero or nominal price, as in the Czech Republic, there is a transfer from public to private assets equal to the value of the privatized firms, but the impact on income distribution will be egalitarian because the process transfers shares to all citizens equally. In contrast, if assets are freely transferred to a single wealthy individual, the impact will be to severely worsen the distribution of income. In practice, state-owned assets that are transferred at below their market value are often also transferred to individuals who are already wealthy, leading to increasing inequality.

Political factors may play a significant role in this process, with corrupt elites seizing state assets for themselves, or using them to reward their cronies or political supporters. Thus, rather than being used to improve efficiency, privatization may be employed by the ruling group as a mechanism to redistribute wealth and resources. Acemoglu and Robinson (2012) point to the transfer of state assets into the hands of the governing elite (often associated with the deliberate continuation of monopoly power) as a mechanism of extractive political institutions; they cite the telecommunication privatization in Mexico and the huge amount of wealth accumulated by Carlos Slim ($47 billion in 2016 dollars) as an example.
But negative distributional effects may also occur for reasons of perceived efficiency enhancement, for example because the state believes that particular private individuals are those most likely to be able to improve company performance. This implies a trade-off between efficiency and equity objectives in the privatization process. Equity is supported by processes which engender dispersed ownership, while it is usually argued that efficiency is driven by concentrated ownership (Estrin 2002). The empirical evidence highlights this trade-off; improvements in the performance of privatized firms have been found to depend on subsequent ownership arrangements (Djankov and Murrell 2002). Notably, privatization to concentrated owners, such as to foreign firms or to small groups of strategic owners, yields greater improvements in performance than privatization to the general population via share offerings, or to managers and workers (Estrin et al. 2009).

Birdsall and Nellis (2003) place the issue of the distributional impact of privatization more formally into an efficiency/equity framework. The effect of privatization on income distribution between taxpayers and the new owners depends both on the initial price and on the post-sale stream of value produced. There is no unambiguous prediction about the distributional effects of privatization, which will instead depend on initial conditions, the privatization process and the post-privatization political and economic environment. Any assessment of the effects should be dynamic and highly country-specific, depending on the political and economic context and its history. However, they argue that there is scope for efficiency-enhancing privatization which also promotes equity in developing countries.

We review below the distributional impacts of privatizations through their effect on ownership, employment, prices and their fiscal effects (see table 2 for a summary).

A Review of the Distributional Impacts of Privatizations in the Last Decade

Ownership.

As Megginson (2000) notes, in countries that have privatized through asset sales, the process has frequently been non-transparent and plagued by insider dealing and corruption. Thus in Russia, the “loans for shares” programs enabled well-connected financiers to obtain controlling stakes in the country’s most valuable firms for a price well below their true value (Megginson 2000). Moreover, the distributional impact of voucher privatizations has also been disappointing; in Russia and the Czech Republic, the returns on the vouchers were much lower than anticipated, and very small in comparison to what a very few well-connected groups of people obtained in the privatization process (Birdsall and Nellis 2003).

Employment.

Privatization can also affect the distribution of income through its impact on employment. As public enterprises tend to be overstaffed prior to privatization, private
<table>
<thead>
<tr>
<th>Distributional impact</th>
<th>Progressive effect</th>
<th>Regressive effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ownership</td>
<td>If the sale is conducted in a transparent way, with a wide distribution of vouchers with positive returns.</td>
<td>If the asset is under-priced and rewards political cronynism. If the sale is non-transparent.</td>
</tr>
<tr>
<td>Employment</td>
<td>If newly-privatized firms become more efficient and dynamic, total employment might recover after the initial restructuring phase.</td>
<td>The restructuring and consequent disproportionate layoff of specific categories of worker.</td>
</tr>
<tr>
<td>Prices</td>
<td>Privatization can lead to a fall in prices if it is accompanied by increased competition. In addition, if private management leads to efficiency gains, some of the savings can be passed on to consumers.</td>
<td>Prices may increase if they were previously below cost-recovery level.</td>
</tr>
<tr>
<td>Access</td>
<td>Access may increase if the privatized business is expanded through investments.</td>
<td>If the private owner decreases its engagement in specific market segments that are beneficial to the poor. In addition, poorer consumers can see their access reduced if privatization is accompanied by the end of illegal water and electricity connections.</td>
</tr>
<tr>
<td>Fiscal</td>
<td>If it leads to increased access by the poor to government services funded by new tax flows.</td>
<td>Privatization may affect real income (net of taxes) if it reduces the tax burden differentially across households. Privatization transfers control rights to private interests and eliminate public subsidies, benefiting taxpayers but reducing consumers’ surplus if costs are increased.</td>
</tr>
</tbody>
</table>

Ownership can lead to restructuring and consequently disproportionate redundancies for specific categories of worker (low-skilled, for instance). The study by Chong and Lopez-de-Silanes (2002) based on a survey of 308 privatized firms (covering 84 countries) over the period 1982 to 2000 showed that employment was reduced in 78% post-privatization, likely worsening income distribution (Birdsall and Nellis 2003).

That being said, if the newly-privatized firm becomes more efficient, total employment might recover after the initial restructuring phase. In addition, government-owned firms that do not privatize may also have to reduce workforce size. Research by Gupta (2011) on privatization in India covering the 20-year period of 1989 to 2009 shows that privatization increases employment significantly and is not associated with a decline in employee compensation. Moreover, Gupta argues that an evaluation of the redistribution of wealth from the government to private
owners must also take account of the cost of subsidies to government-owned firms. However, the employment costs of privatization will be borne by specific groups of workers, while the benefits, in terms of reduced subsidies, are distributed across taxpayers. Hence, privatization may face opposition from organized interests who benefit from maintaining government ownership.

While Gupta’s (2011) work is a single-country study, it has the merit of using more advanced econometric methods to control for dynamic selection bias by applying firm fixed effects and comparing privatized firms to a control group of firms that have also been selected for privatization but have not yet been sold. In addition, the share of private ownership is introduced with a lag to reduce the possibility of simultaneity between privatization and performance.

**Prices and Access.**

Privatization can also have different impacts on income groups through prices and access to services. First, privatization can lead to a fall in prices if it is accompanied by increased competition. In addition, if private management leads to efficiency gains, some of the savings can be passed on to consumers. However, prices may increase if they were previously below cost-recovery level. The distributional impact depends on how the consumption of the firms’ goods and services varies by income levels. Access may increase if the privatized business is expanded through investments which could not be undertaken in public ownership. However, private owners may decrease their engagement in specific, low-return market segments, which may disproportionately affect the poor. Price increases are common following privatization in network or infrastructure industries, along with increases in the quality of services. On the one hand, subsidized services tend to benefit relatively wealthy consumers more than poorer ones; as such, they may be relatively more impacted than the lower-income segment by privatization. On the other hand, price increases following the privatization of electricity and water will increase the burden of poorer consumers, especially if it is accompanied by the end of illegal water and electricity connections (Birdsall and Nellis 2003).

Several studies in Latin America have shown that utility privatization has in fact led to network expansion and increased access to the service by the population, especially the rural poor (for Peru, see Torero and Pasco-Font 2001; for Argentina, see Chisari, Estache, and Romero 1999; Delfino and Casarin 2001, and Ennis and Pinto 2002; for Bolivia, see Barja and Urquiola 2001; for Mexico, see Lopez-Calva and Rosellon 2002). This increased network coverage has often been the consequence of market expansion enabled by private investment capital (see Clarke, Kosec, and Wallsten 2004).

When access has increased significantly without a steep rise in prices, privatization has had positive distributional effects (Birdsall and Nellis 2003). However,
increased access has often been accompanied by substantial price increases (Estache, Foster, and Wodon 2002). In addition, an important negative distributional impact has been realized through the elimination of illegal connections to electricity and water networks by lower-income people. A recent paper by Hailu, Guerreiro-Osorio, and Tsukada (2012) on water service privatization in Bolivia in the late 1990s and early 2000s shows how tariff increases required for full cost recovery may lead to adverse privatization outcomes; in this case, the eventual renationalization of the company. To examine the impact of privatization on access, the authors use a difference-in-difference approach comparing two groups: households in cities where the utility was privatized, and households in other cities, with two points in time, before (1996) and after (2001 and 2005) privatization. These authors find a positive relationship between access to water and living in cities where the water utility was privatized. However, the water sector was renationalized in 2006, partly because of popular movements against the tariff increases required for full cost recovery and the failure of the concessionaire to meet targets stipulated in the contract.

Finally, Austin, Descisciolo, and Samuelsen (2016) point to the limits of privatization in sectors with public goods characteristics. Examining the privatization of healthcare in 99 less-developed nations over the 1995–2000 period, they employ two-way fixed effects ordinary least squares regression models. The fixed effects allow them to deal with unmeasured, time-invariant variables that are excluded from a regression model. They regress tuberculosis prevalence per 100,000 on the log of private health expenditures, the log of public health expenditures and a set of controls (economic development, education, HIV prevalence and access to water and sanitation). They find that, while public health expenditures reduce tuberculosis rates in developing nations over time, this is not the case for private health expenditures.

Fiscal Effects.
The fiscal effects of privatization on income distribution are indirect and come through changes in revenues and expenditures. In particular, privatization may affect real income (net of taxes) if it reduces the tax burden differentially across households, or if it leads to increased access by the poor to government services funded by new tax flows. The study of Davis et al. (2000) on 18 developing and transition countries showed that the net fiscal effects of privatization were receipts in the order of 1% of GDP. In some countries, the main fiscal benefits of privatization have been to eliminate subsidies. Subsidies in critical infrastructure services have often led to the rationing of under-priced services, hardly affecting poorer households that often had little or no access to these services, while the non-poor enjoyed the underpriced access. To the extent that privatization stops these flows of subsidies, it produces indirect benefits in terms of increased retained revenues (Birdsall and Nellis 2003), which could indirectly benefit the poor.
Policy Implications

The traditional literature, primarily concerning developed economies, argued that privatization had largely positive effects on the economic and financial performance of the companies involved, as well as wider spillover benefits, for example, via technological diffusion from foreign ownership of former SOEs and enhanced efficiency from the privatization of utilities and other forms of infrastructure. Moreover, privatization programs also frequently achieved additional objectives, including the generation of revenues to relax state budget constraints and a broadening of share ownership amongst the population. On this basis, privatization became an important element of reform programs in transition and then developing economies from the 1990s. The experience of the past twenty years leaves some of these conclusions unchanged, but leads us to a more nuanced evaluation of the effects of privatization in the context of economic development.

In particular, though state sectors are often very large in developing economies, it has been hard to establish widespread privatization programs in many parts of the world, partly because of political opposition. This has arisen for a variety of reasons. First, the record of privatization as it spread to middle income and then transition economies (including China) was not always so positive as in developed economies. The lesson of the transition economy experience was that privatization was not always a panacea: if the mode of privatization was inappropriate or the market environment not competitive, privatization might not enhance the performance of the firms involved (Estrin et al. 2009). Moreover, privatization programs were associated with scandals: inappropriate valuations led to the emergence of extreme inequalities of wealth. Second, in developing economies where the institutional environment, particularly with respect to regulation of monopolies, was sometimes even weaker than in transition economies, the benefits of privatization were even less automatic, depending on the sector, and were contingent to a significant degree on the design of the privatization program. Third, distributional issues are especially significant in developing economies, so privatization programs also had to consider distributional impacts in ways that had been less relevant for developed economies; opposition rested on issues raised by the efficiency-equity trade-off. Finally, political economy issues are perhaps of even greater consequence for policy choices in developing economies, and privatization programs are especially open to manipulation by extractive political institutions and elites in fragmented political environments.

This long list of concerns has meant that the spread of privatization programs to developing countries has been limited, both geographically and with respect to sectoral reach. The slowdown in privatization has no doubt been exacerbated by the global recession of 2008 and the resulting flight from risk, which has particularly affected stock markets in developing economies. Moreover, the evidence about the effects of such privatizations of economic performance is quite nuanced. To be
successful, a privatization program needs to align its objectives with its methods of privatization, taking into account the sector in which the company operates and the national, institutional, and political context.

**Necessary Pre-conditions for Successful Interventions: Regulatory Agencies and Managerial Incentives**

As Lopez-de-Silanes (2005) notes, good rules and contracts are key for a smooth and beneficial privatization process. However, government restructuring of SOEs prior to their sale is likely to be fraught with political difficulties because officials may try to extract private benefits. Although restructuring could increase revenues from the sale, Lopez-de-Silanes suggests that restructuring policies do not lead to higher revenues. In addition, Lopez-de-Silanes (2005) notes the importance of policies to complement privatization; of particular importance is the need to set up an appropriate regulatory and institutional framework for the post-privatization period.

Indeed, several papers have shown how a strong and independent regulatory institution can help address the negative impact of corruption on the privatization process. Wren-Lewis (2013) uses a fixed-effects estimator on a panel of 153 electricity distribution firms across 18 countries in Latin America and the Caribbean from 1995 to 2007. He regresses the log of labor employed on a corruption indicator, independent regular authority dummies (including dummies for good and bad regulators), and private ownership dummies and interaction terms. Wren-Lewis employs firm fixed effects to control for time-invariant unobservables. Because each firm is present in only one country or province, the corruption and regulation terms are estimated based only on changes in these variables within countries/provinces. He also includes year fixed effects to take into account time effects. Wren-Lewis shows that greater corruption is associated with lower firm labor productivity, but this association is reduced when an independent regulatory agency is present. However, because of broader institutional weaknesses, developing countries face many challenges in establishing a strong regulator. One limit of this study is that there may be important (unobserved) parts of the reform package that also impact productivity. As such, it should not be assumed that the (observed) reform will have the same impacts elsewhere.

Gassner and Pushak (2014) have examined the impact that the UK regulatory model has had in developing and transition countries, and the extent to which they have successfully followed its key features; competition, independence and efficiency of service delivery through incentive-based regulation. The authors note that while regulatory agencies have spread rapidly, the success of the UK regulatory model has been only partial in middle and low-income countries. They argue that the context of developing countries, with below cost-recovery tariffs and continued
state-ownership, makes it more difficult to establish truly independent regulatory institutions.

Thus, developing countries face many regulatory challenges; they often start with important operational inefficiencies and insufficient revenue generation. In addition, a majority of firms in potentially regulated sectors are still publicly-owned because they are not attractive enough for private sector investors, and because governments do not want to cede control of essential services. Under these circumstances, incentive regulation for efficiency savings is difficult: given the low tariffs, not enough investment can be undertaken to improve service delivery, and without private profit motives there is not a strong incentive for managers to bring about efficiency. Under-pricing and poor operational performance are serious problems: according to the 2010 Africa Infrastructure Report published by the World Bank (2010), the under-pricing of electricity costs the sector at least $2.2 billion a year in forgone revenues (0.9% of GDP on average).

Recently, the concept of hybrid regulatory models has been proposed as a solution to the challenges in developing countries (Eberhard, 2007). In hybrid models, regulatory contracts and independent regulatory agencies coexist. In a context where the institutional capacity is low and/or regulatory commitment is weak, an independent regulatory agency is supplemented by contracting out or outsourcing certain regulatory functions. An illustration of this is the 20-year water and electricity concession contract in Gabon, which requires external experts to monitor the service provider’s performance in achieving coverage targets. The experts are paid from dedicated funds set aside from the concessionaire’s revenues and produce only nonbinding studies. This monitoring mechanism is aimed at strengthening the independence and competence of the ministerial department responsible for supervising the contract. Policymakers may also obtain regulatory assistance from regional regulators or from other countries through twinning arrangements. For example, the Eastern Caribbean Telecommunications Authority (ECTEL) serves the member countries of the Organisation of Eastern Caribbean States as a shared regulatory body (Tremolet, Shukla, and Venton 2004).

Taking into consideration local management and incentives is also important for successful privatization. Liu, Sun, and Woo (2006) identify the motives of local government leaders and the constraints that they face during a privatization process. These authors conclude that local governments’ motivation to privatize their SOEs depends on whether the ownership transfer sufficiently stimulates the growth of local tax revenues without sacrificing bureaucrats private control benefits. In addition, Dinc and Gupta (2011) in their study of privatization in India observed that no firm located in the home state of the minister in charge is ever privatized, which highlights the importance of local political factors in the privatization process.
What about Remaining SOEs?

To a certain extent, the recommendations about regulation and managerial incentives also apply to remaining SOEs. In fact, Bartel and Harrison (2005) argue that public-sector inefficiency is due to the softness of budget constraints and the degree of internal and external competition. This implies that efficiency gains in SOEs could be achieved by reducing or eliminating government financing for public enterprises, and/or increasing import competition.

Regarding agency-type problems, Hsieh and Song (2015) observed that one of the key reorganizations of state-owned “corporatized” firms in China was that the parent company (the controlling shareholder) of the firm incorporated as Limited Liability Corporation was to monitor the firm and be responsible for the compensation of the firm’s senior managers. These managers were held accountable for the firm’s bottom line, which reduced agency-type problems. The senior executives of the parent company, in turn, were directly appointed by the local government or by the Central Organization Department of the Communist Party.

Privatization to Foreign Owners

Work on transition economies established that when SOEs are privatized to foreign investors, the efficiency gains are particularly pronounced. The results on foreign ownership do seem, however, to be replicated in the developing economy context. Thus Du, Harrison, and Jefferson (2014) have found that foreign equity participation is associated with an improvement in productivity which is greater for SOEs than for non-SOEs in China’s manufacturing sector, suggesting that foreign firms can play an important role in improving SOE performance. The benefits of privatization via transfer to foreign firms have also been observed in the case of banking in Africa (see Clarke, Cull, and Shirley 2005).

Part of the reason that foreign ownership improves productivity can be found in the relation between foreign ownership and corporate risk-taking. Boubakri, Cosset, and Saffar (2013) found that foreign (state) ownership is positively (negatively) related to corporate risk-taking, and that this relation is stronger in countries with better institutions. To the extent that corporate risk-taking is an important driver of economic growth, privatization via the transfer of ownership to foreign owners should yield important economic benefits through a reorganization of prevailing incentive structures and changes in the degree of risk aversion. Jaslowitzer, Megginson, and Rapp (2016) also observe that risk aversion and financial conservatism are one of the reasons that state ownership is associated with inefficiency. Using a matched panel of 624 firms, these authors find that state ownership curtails firms’ responsiveness to investment opportunities. Despite these findings, in some developing countries the
sale of state assets to foreigners, which carries overtones of colonial legacies, can be a politically charged subject.

Concluding Comments

Privatization involves the transfer of productive assets from the state to private hands. Such transfers are, by their very nature, politically sensitive and subject to potential corruption and abuse. We outline below some important issues that policy makers in a developing country should consider when examining a proposed privatization. In so doing, we assume that the primary purpose of privatization is to enhance economic growth.

First, policy-makers need to examine and establish the preconditions for success, in terms of the business environment for competition, governance, and entry. The evidence suggests that privatization has greater benefits on firm performance in stronger business environments because the success of the process relies on effective corporate governance of the privatized entity, as well as effective market competition. Key issues at the national and sectoral level include:

- Depth and liquidity of the capital market (particularly important for privatization via IPO).
- Barriers to new domestic firm entry (formal entry costs, bureaucratic costs, possibilities for incumbents to restrict entry by the use of political relationships).
- Quality of the legal system concerning corporate governance, for example company accounting procedures, rules on minority shareholders, etc.
- Quality of business support, for example, legal firms, accounting firms, management consultants, recruitment firms.
- Openness to foreign direct investment, both via acquisitions (via privatization) or via greenfield (to create competition), and access to foreign portfolio capital.
- Depth and competitiveness of managerial market (pool of qualified managers).
- Strength and effectiveness of competition, and competition agency.
- Independence of anti-monopoly agency from state.

The quality and independence of the state’s administrative apparatus is particularly important. Privatization makes considerable demands on the capability of the state, both in ensuring that the process is not captured by local elites, and in managing the relationship between the government and the firm at arm’s-length post-privatization, for example, via regulation. Successful privatization requires competent government with low levels of corruption.

Turning to the privatization process itself, there is strong evidence that openness of bidding to all, including foreign firms, is a key factor of success.

Policy-makers also need to determine the appropriate privatization methods. Related to this, the pricing of the assets to be privatized are a crucial issue with respect to the transfer of assets from public to private hands, and the likely impact on the distribution of income and wealth. The chosen methods depend in part on the preconditions noted above. Countries with poorly developed capital markets are unlikely to be able to privatize through IPOs. The main methods of privatization, listed on the
basis of the evidence of the literature in order of likely favorable impact on economic growth and development are as follows:

- Sale to high-quality foreign firms.\(^{15}\)
- Sale on domestic capital market via IPO.
- Sale to domestic businesses or business groups (trade sale).
- Sale to existing managers and/or workers.
- Free distribution of shares to the population (mass privatization).

There are obvious trade-offs. Free distribution ensures equality in the allocation of assets around the population, but is likely to lead to weak corporate governance. Selling to foreign owners, with appropriate safeguards, can raise company efficiency but may lead to job losses.

Privatization seeks to improve company efficiency via corporate governance. However, as we have seen, a number of side-effects may impact other key policy targets and these need to be considered in advance.

**Social and Economic Side Effects.** Higher efficiency/profitability may be obtained through lower levels of employment, lower wages, reduced public service provision and higher product prices, with negative distributional and social effects.

**Competition Side Effects.** Especially if the government is concerned with selling to foreigners and/or maximizing revenues, competition effects may be negative and serious.

**Global Impact.** Selling key assets such as banks or resource companies to foreign firms may restrict the range of domestic policy and hinder long-term development.

**Political Side Effects.** Selling assets to elites may concentrate political power and economic wealth into fewer hands.

**Effects on Distribution of Income.** An enhanced focus on the profitability of firms may lead to increased prices of important products for poor households, as well as reduced pay, worse employment conditions, and fewer job prospects.

**Effects on Fiscal Balance.** In principle, this should be unchanged because if the asset is priced correctly, the price should reflect the future expected earnings from the company. In practice, pricing may be set low to achieve distributional targets or to support elites and friends. This would worsen the government’s balance sheet. At the same time, the new owners may be more productive than the state, and hence raise activity and profits, with a positive effect on GDP and government revenues.

**Notes**

Saul Estrin is a professor of management at the London School of Economics; correspondence to be sent to s.estrin@lse.ac.uk. Adeline Pelletier is a lecturer in strategy at the Institute of Management Studies, Goldsmiths College, University of London. This work was supported by the U.K. Department for International Development and the Overseas Development Institute. The authors would like to thank Tim Estrin and Pelletier.
Green, Alberto Lemma Deborah McGurk, Anne McKinnon, Bill Megginson, John Nellis, Jon Stern, and Jan Svejnar.

1. Kikeri and Nellis (2004) have also conducted a wide-ranging assessment of privatization.
2. Each of these three regions representing between 3% and 5% of total world privatization proceeds over the 1988 to 2008 period.
3. The privatization barometer database provides world aggregate data on privatization and a country breakdown for developed countries. We are not aware of an alternative database providing such information. This was also confirmed by several academic and practitioner experts on privatization whom we contacted during the course of this research.
4. Infrastructure includes transportation, water and sewerage, telecommunications, natural gas transmission and distribution, and electricity generation, transmission, and distribution.
5. The manufacturing and services sector includes agribusiness, cement, chemicals, construction, steel, hotels, tourism, airlines, maritime services and other sub-sectors that are not infrastructure or finance related.
6. The financial sector includes banks, insurance, real estate, and other financial services.
7. The energy sector includes the exploration, extraction, and refinement of hydrocarbons, oil, and natural gas.
8. The primary sector includes the extraction, refinement, and sale of primary minerals and metals such as coal and iron ore.
9. The ownership pattern resulting from privatization often depends on the mode of privatization chosen. Thus, private sales usually lead to concentrated strategic owners, while mass privatization usually generates widespread ownership, at least initially. The impact of mode of privatization on national economic performance in transition economies is explored in Bennett, Estrin, and Urga (2007).
10. Note, however, that in the utilities sector (particularly for water), the technology and the nature of the product restrict the possibility of competition in the market and therefore the efficiency gains following privatization. In this case, competition for the market (to win the contract or concession agreement) has to be organized. Given the ambiguous results of privatization in noncompetitive markets in terms of improving economic performance (Megginson and Netter 2001), regulation may prove to be more effective (Kirkpatrick, Parker, and Zhang 2006).
11. The performance of privatized banks in the seven countries of the West African Economic and Monetary Union from 1990 to 1997 improved in the first year after privatization, but not after that.
12. Improvements in performance in Nigeria were observed in fully-divested banks, but not in the ones where the government retained minority shareholdings.
13. Whereas competition is feasible in telecommunications markets, it is usually cost-inefficient in the market for water services, given the scale of the investment in network assets required to deliver the product.
14. Privatization is also not associated with the profitability and efficiency of government-owned firms.
15. Note, however, that this method may suffer from a trade-off with competition objectives since foreign firms may seek local monopoly power. Such sales may be accompanied by conditions with respect to technology transfer, domestic content of inputs, employment, environment, etc.

References


Public-Private Partnerships in Developing Countries: The Emerging Evidence-based Critique

James Leigland

Advocates of public-private partnerships (PPPs) for infrastructure services in developing countries have long battled criticism of these arrangements by civil society groups. The view among PPP advocates generally has been that these criticisms are mostly ideological polemics that mix opinion with selected but often misinterpreted facts. But over the last two decades, as the experience with PPPs has increased in both developed and developing countries, a different kind of critique has emerged, one that is based on non-ideological empirical research, and is sometimes expressed by PPP advocates. These studies often focus on individual aspects of PPPs, and usually do not claim to be “PPP evaluations” or express opinions on the overall value of PPPs. Taken together, a powerful, evidence-based critique of PPPs is emerging, but one that is more measured than much of the criticism of the last two decades. This new critique recognizes many cases in which PPPs have not been successful, but also some situations in which PPPs can generate value for money. Because of its critical tone, some of this research is now regularly cited by the civil society critics of PPPs, giving their arguments more weight than was the case a decade ago. This paper attempts to summarize some of the most compelling examples of this kind of emerging critique, and uses the summary to assess the practicality of the G20’s recent advocacy of large, “transformational” PPPs as tools for dealing effectively with infrastructure challenges in low-income countries.

Global support for public-private partnerships (PPPs) for infrastructure seems stronger than ever before. Discussions in G20 meetings over the last several years have increasingly focused on the need for a huge scale-up in infrastructure investment in developing countries, particularly low-income countries. G20 pronouncements talk about the advantages of realizing this scale-up via large, “transformational” projects involving private sector participation. By this they mean large, regional, or...
cross-border infrastructure projects involving private investment and management, which potentially have positive, transformational impacts on entire countries or regions. From time to time over this period, the G20 has considered efforts to help modify the mandates of national and international development banks so that these institutions will take the lead on such PPPs and crowd in the private sector.

As a result, several development banks have been considering adjustments to their business models to give more attention to regional infrastructure PPPs. During the negotiations in 2013 for the 17th replenishment of the International Development Association (IDA17), the World Bank proposed using IDA funding to help develop transformational PPP projects. Those proposals have now evolved into the Bank’s design of the Global Infrastructure Facility (GIF), an entity meant to coordinate the efforts of MDBs, private investors, and governments to prepare and structure PPPs. The BRICS countries, at their summit in Durban in March 2013, announced plans to create a new development bank (now known as the New Development Bank) that would focus on infrastructure, and do so in a way that would make up for the deficiencies of the existing international financial architecture and help catalyze the private sector investment needed in rapidly-growing BRICS economies (Republic of South Africa 2014). Perhaps the most ambitious and concrete commitment of this kind to date is the decision, announced by the African Development Bank (AfDB) in July 2013, to create a billion-dollar preparation and financing facility for large infrastructure projects in Africa, referred to as Africa50. The institution’s purpose “is to unlock private financing sources... and to accelerate the speed of infrastructure delivery in Africa,” (AfDB 2013).

But as this enthusiasm for PPPs is growing, so is a less widely-recognized body of research that takes a much more measured approach; it still represents a kind of advocacy, but one that incorporates a greater degree of critical analysis of PPP successes and failures. A recent example of this is evident in a World Bank working paper by Michael Klein (2015), an influential PPP advocate during the 1990s and early 2000s. Klein notes that despite more than two decades of use and refinement of the PPP mechanism, there are still no consistent geographical patterns of usage: “The general picture is one of waves of enthusiasm for PPPs followed by some disenchantment and consolidation. Different countries were caught up in the waves at different times.” What accounts for this lack of sustained enthusiasm? Klein says that evaluations show that PPPs can outperform public sector firms, and “are useful tools for reform of service delivery” (Klein 2015). But it is no longer clear that PPPs are consistently better run than public firms. “The evidence suggests that well-run public firms tend to match the performance of private firms in regulated sectors” (Klein 2015).

Klein’s comments are a reminder that a significant amount of evidence-based research on PPPs has accumulated since the late 1990s. But a good deal of it, particularly over the last decade, has not been uniformly positive about PPPs, at least not in the fashion of the largely promotional literature published by MDBs and donors in
the 1990s and early 2000s. Some of this research, for example by economist Antonio Estache, is now being used to bolster PPP criticisms prepared by civil society groups (e.g., Alexander 2013). Such groups have produced a broad collection of critical PPP studies: International Rivers (Bosshard, 2012); Public Services International (Hall 2015); Heinrich Boell Foundation (Alexander 2013); CEE Bankwatch Network (2008); Oxfam (Marriott 2014); the Bretton Woods Project (2016). These groups have long been critical of PPPs, but in the past their arguments against private participation have often seemed more ideological than evidence-based, and therefore not very compelling. But the growing use of evidence-based research reported on by respected social scientists like Estache, Klein, and others has added weight to their arguments, warranting more careful consideration by PPP advocates.

Much of this evidence-based critique of PPPs is dispersed in collections of working papers and academic articles, or focuses on individual aspects of PPP projects, and does not claim to be “PPP evaluations.” In any case, this body of research is having no noticeable impact on PPP discussions in organizations like the G20. This paper attempts to summarize some of the most compelling examples of this kind of emerging critique by organizing them into several key themes: (i) how prevalent is the usage of PPPs in the developing world; (ii) the costs and profits associated with these projects; (iii) the significant outlay involved in preparing these projects; (iv) their struggle to attract commercial financing; (v) how developmental they are; and (vi) the institutional and political problems that limit success. The discussion of each key theme also attempts to extract insights about some of the efforts to deal with these criticisms and establish conditions under which PPPs are likely to provide value for money. Finally, this paper uses this summary to (vii) assess the practicality of the G20’s recent promotion of “transformational” PPPs as mechanisms for dealing effectively with infrastructure challenges in low-income countries.

As there is no universally accepted definition of “public-private partnership” (PPP), this paper adopts Klein’s broad definition of PPP to mean “private participation” in infrastructure sectors like telecoms, transport, water and sanitation, waste management, and electricity: “Activities that fall under this umbrella may sometimes be characterized, for example, as ‘concession’ or ‘franchise’ or ‘build-operate-transfer’ deals” (Klein 2015). Where indicated, some of the data cited here covers “private participation in infrastructure,” which includes information on privatized utilities, as well as more traditional kinds of infrastructure PPPs.

### PPP Prevalence

The foundation of Klein’s argument is that although many countries use PPPs at least occasionally, the prevalence of usage surges in waves, often driven by fiscal problems or other ways in which the public system has “run into trouble,” (Klein 2015). But the waves inevitably recede, “... in seemingly random patterns,” (Klein 2015). Klein
attributes the shallowness of PPP popularity to a lack of clear and consistent evidence that PPPs perform better than public sector organizations. He estimates that as of 2015, while PPPs account for a share of the total infrastructure investment in low- and especially middle-income countries, it is normal for a country to use PPPs for only about 15 to 25 percent of total infrastructure investment.

Other sources of data suggest even lower levels of PPP prevalence. The World Bank’s Africa Infrastructure Country Diagnostic study (AICD), published in 2010, found that in total, the private sector accounted for an impressive level of infrastructure investment in sub-Saharan Africa (SSA) by contributing about 29 percent of total capital spending (Table 1).

But the AICD further qualified this data in several important ways: First, it demonstrated that private investment was heavily skewed in terms of countries, with about 60 percent of total SSA private sector investment shared equally by just two countries—Nigeria and South Africa. Second, private investment was heavily skewed in terms of sectors, with 77 percent of SSA’s private investments since 2000 going to telecommunications, mostly via build-own-operate projects (BOO). According to the AICD, the energy sector, which is arguably the most in need of urgent major capital investment, attracted only 10 percent of total private investment.

Other studies tend to support lower usage figures. Burger and Hawkesworth (2011) surveyed 22 countries (19 OECD countries and three middle-income countries) regarding value-for-money issues associated with PPPs. Of these, eighteen countries reported that less than 10 percent of public sector infrastructure investment took place via PPP arrangements. From 2000 to 2010, the UK’s Private Finance Initiative (PFI) probably averaged a higher annual percentage of total infrastructure investments via PPPs than most other OECD countries, at about 12 percent. The only two non-OECD countries surveyed, Mexico and Chile, reported that over 20 percent of their infrastructure investment occurred via PPPs. A number of the countries surveyed admitted informally that they did not foresee PPPs exceeding 15 percent of total public investment (Burger and Hawkesworth 2011).

| Source: Foster and Briceño-Garmendia 2010. |
Across the developing world, PPPs play a relatively small role in infrastructure investment, averaging between 15 to 20 percent according to the Independent Evaluation Group of the World Bank (Independent Evaluation Group 2014). In the poorest developing countries, the use of PPPs has been even more negligible. Figure 1 demonstrates this, using data from the World Bank’s PPI Project Database to show investments related to “private participation in infrastructure” (PPI) in countries eligible for support from the International Development Association (IDA; i.e., countries whose Gross Net Income per capita is below $1,215), and contrasts these against data from non-IDA developing countries (“blend” countries have been excluded). ¹ In its review of PPI activity in IDA countries since 2011, a World Bank report remarks: “The market for PPIs has not been expanding,” (Ruiz-Nunez 2016).

In the developing world, a share of infrastructure investment in the range of 15 to 20 percent does not mean that PPPs have failed to play a significant role in infrastructure. But it is far less than what was expected of PPPs in the 1990s when Klein and his colleagues at the World Bank were considering sharp reductions in infrastructure lending because they expected the private sector to eventually play a more dominant role in bridging the gap and financing and managing infrastructure services in that region of the world. ²

What does this information about PPP prevalence tell us about the conditions under which PPPs are likely to provide value for money? The message is simple: PPPs work much better in middle-income economies than they do in low-income

---

**Figure 1.** PPI Investment by IDA Status, 1995 to 2015, in $ millions

![Graph showing PPI Investment by IDA Status, 1995 to 2015, in $ millions](https://example.com/graph.png)

*Source: World Bank PPI Database as of Jan. 2017*
countries. This means that in most cases a complex, long-term, brownfield concession for retail water distribution, for example, requiring significant capital investment, should not be the first choice as the service delivery solution in a least-developed country (as such contracts often were in the early 1990s). This review suggests that the poorest countries can usually benefit more from traditional technical assistance and capacity building, or from hybrid projects that mix elements of PPP contracts with those of consulting or engineering, procurement, and construction (EPC) contracts to reduce risks for the private partners. Reforms to legal and regulatory frameworks within which PPPs eventually would be structured are also critical in these countries, along with help in improving government procedures for things like procurement and construction management.

Costs and Profits

PPPsinvolvemultiplestakeholders, all of whom have interests in how the project performs. The conventional view of PPPs is that, when compared with typical public projects, they should provide better service at lower costs. The private partner needs to make a profit, but the return should be reasonable. Projects that do not meet these expectations can be subject to public criticism from government, the news media, user groups, civil society, etc. But getting the balance right in a way that satisfies all key stakeholders is not an easy task, as the recent history of PPPs indicates.

The profitability of private firms involved in PPP projects, and the resulting costs for government and end-users, has long been a controversial subject in industrialized countries like the United Kingdom, where the PFI became one of the OECD’s best-known PPP programs starting in the early 1990s. Studies of the profitability of PFI deals in the U.K. health sector have found rates of return as high as 60 percent (Shaoul 2008).

Whether or not the cause is high profits, PPPs have often tended to be more expensive than traditional public procurement. A 2006 report by the European Investment Bank (EIB) reviewed the costs of 227 road projects in 15 European countries and concluded that projects done as PPPs (65 of the total), were 24 percent more expensive than those done via traditional public procurement (Blanc-Brude, Goldsmith, and Välilä 2006). In a 2015 review of effective interest rates on private finance projects, the U.K.’s National Audit Office found that these rates, at 7 to 8 percent, were double the rates on normal government borrowing, at 3 to 4 percent (U.K. National Audit Office 2015). In a 2016 review of PPP literature sponsored by the UN’s Department of Economic and Social Affairs, the authors concluded that “Overall, the evidence suggests that PPPs have often tended to be more expensive than the alternative of public procurement...” (Jomo et al. 2016).

Most industrialized countries try to anticipate project benefits and costs, including private profits, by requiring a “value for money” (VFM) analysis of PPP projects. VFM
analysis involves estimating project costs, including profits for the private partners, and measuring them against project benefits, including service quality, quantity, and prices for governments or end-users. Quantitative VFM assessment typically involves comparing the chosen PPP option against a “public sector comparator” (PSC). The PSC allows a comparison of the risk-adjusted cost to government of procuring the project through traditional procurement (the PSC), with the expected cost to government of the PPP (pre-procurement) or the actual PPP bids (post-procurement).

The survey by Burger and Hawkesworth (2011) found that 17 of 22 countries used “public sector comparators” to assess the value for money of PPPs. But ever since the technique was first refined and pioneered as part of the UK’s PFI program in the 1990s, it has been criticized for being inaccurate and subject to manipulation, leading some observers to conclude that it is often an expensive way of endorsing the pre-selected choice of private participation. A UK Audit Commission report (2003) concluded that “the PSC has lost the confidence of many people, and risks being seen more as a hoop to jump through on the way to government funding than a valuable exercise that can help ensure better VFM.”

Engel, Fischer, and Galetovic (2014), noted that in some countries PPPs are attractive to government not necessarily because they are expected to be less expensive, but simply because accounting rules allow project costs to be moved off government books in order to give the appearance of lower debt levels. Klein also mentions this and notes that when the United Kingdom changed its accounting rules in 2009, PFI projects became less attractive as a result. If true, this may account for some of the well-documented failures of the PSC to accurately forecast PPP project costs.

Of course, other kinds of quantitative comparisons can be done on an economic cost-benefit basis, using a wide variety of VFM methodologies (World Bank Institute and Public-Private Infrastructure Advisory Facility 2013). It is also true that regardless of the method used, huge mistakes are routinely made in the estimation of financial costs and benefits associated with infrastructure projects—whether or not PPPs are involved. In a survey of 58 rail projects (a mixture of public and private projects), Flyvbjerg (2005) found that costs were underestimated by an average of 45 percent and demand forecasts were overestimated by an average of 51 percent. Flyvbjerg concluded that such consistently large mistakes in cost-benefit estimations must be attributed to a combination of faulty techniques, as well as causal effects like “optimism bias” and “strategic misrepresentation.” Doing an infrastructure project as a PPP does not lead to any more accurate estimation of costs and benefits. Vassallo (2007) and Bain (2009) investigated the forecasting performance of privately-financed toll roads; their findings were similar to Flyvbjerg’s. These authors found that PPP toll road traffic forecasts were typically characterized by large errors and considerable optimism bias.

However, in the developing world, VFM techniques like PSCs are arguably even more unreliable than in OECD countries. Even industrial countries have little objective
relevant data upon which to base cost estimates. Without such data, which is virtually non-existent in low-income countries, calculating with any accuracy how much a project will cost over 25 to 30 years of operation is almost impossible. As noted by Estache and Philippe (2012), officials in these countries often find out, after contracts have been signed, that original forecasts of project costs and profits were inaccurate. Indeed, projects often turn out to be unprofitable for private partners, triggering renegotiations:

“Experience has shown that besides the short-term subsidies sometimes needed to support privatization processes, the public sector often has eventually had to commit subsidies for the long-term as well—usually as one of the outcomes of a renegotiation.” (Estache and Philippe 2012).

A widely-cited study by Sirtaine et al. (2004), examined the profitability of PPP projects in Latin America during the late 1990s using a sample that included both brownfield and greenfield projects. The Sirtaine study suggested that on average, projects demonstrated profitability only after about 10 years. Up to that point, project shareholders earned negative returns on their investments, even when adding in management fees, estimated accumulated capital gains, and potential investment markups. The study further found that 40 percent of the sampled concessions did not have the potential to ever become profitable, with that proportion increasing to 50 percent for concessions in the energy and transport sectors. Indeed, Sirtaine estimated that over the history of these projects, on average they were unable to generate sufficient annual operating income after taxes to cover all of their financial obligations. Only by adding in “indirect forms of dividends” like investment mark-ups, transfer fees, and payments for capital appreciation paid at the end of the contract period could these concessions generate reasonable remuneration for their private partners. The study suggested that, at a minimum, these PPPs face constant cash flow problems over at least the first decade of their existence.

Another study, by Guasch (2004), reached conclusions consistent with Sirtaine’s findings. Guasch found high rates of PPP contract renegotiation after only a few years of operation, that is, long before Sirtaine says that private partners are able to confirm their projects’ long-term profitability. Using a sample that also included brownfield and greenfield projects in Latin America, the study concluded that PPI projects in that region registered a high incidence of renegotiation, about 42 percent, coming on average after only 2.2 years of operation. The results of renegotiation tended to favor operators, mostly with improvements to cash flow and profitability via compromises such as permitted delays in investment obligations (69 percent), reductions in investment obligations (62 percent), tariff increases (62 percent), and increased pass-through to tariffs of cost items (59 percent). Again, this suggested that PPP cash flow problems tended to be severe, generating stresses that operators and investors had to
live with over the long term, and which were critical in precipitating many renegotiations.

Estache and Philippe (2012) concluded that sectors like telecommunications and electricity generation in which cost-reflective tariffs seem less controversial have proven to be reasonably profitable and largely free of subsidies. But most other sectors in the developing world—electricity distribution and transmission, transport, and water and sanitation—often require subsidies to sustain cash flows or bring the rate of return close to the cost of capital. If, with careful analysis the subsidies are not built into the initial contracts, they are likely to result from renegotiation of those contracts. In the latter event, renegotiation increases the risk of a less than optimal reduction in the fiscal benefits that governments care most about, that is, the size and pace of investment. Renegotiations also tend to increase the most politically painful costs—tariffs or off-take payments—sometimes to unaffordable levels for governments.

How are governments and their development partners coping with the fact that PPPs are costlier and less profitable than assumed in the 1990s? One way is to rely more heavily on “blended finance” approaches to PPPs. Since the launch in the early 2000s of the multi-donor trust fund for output-based aid (the Global Partnership on Output-Based Aid), “blended finance” has become increasingly popular as a way of using concessional finance to catalyze private sector investment, particularly in infrastructure PPPs. The International Finance Corporation’s (IFC) Blended Finance Unit, launched in 2007, and the EU’s regional blending finance facilities, such as the EU-Africa Infrastructure Trust Fund, have all used subsidies to bring down the costs of various kinds of infrastructure PPPs (IFC 2012). The use of blended finance in this way creates a hybrid approach that combines PPP elements with those of more traditional public projects.

But blended finance involves subsidies, and the use of subsidies requires justification to ensure that it is really crowding in private finance rather than crowding it out. Economists typically recommend the use of cost-benefit analyses for such justifications to clearly identify any obstacles that reflect market failure and help determine whether subsidized finance can solve the problem. Theoretically, cost-benefit analyses can confirm that the likely development impacts of using subsidized finance far outweigh the distortions that may result. When this kind of analysis can be done, it almost certainly leads to more developmental projects. But the difficulties involved in this should not be underestimated. The World Bank’s Independent Evaluation Group found in 2010 that the use of cost-benefit analysis dropped from 70 percent of all World Bank projects in 1970 to 25 percent in 2008, largely because of problems with data and the difficulties in quantifying costs and benefits (Independent Evaluation Group 2010).
Preparation Costs

The OECD (2008) has indicated another important reason for the high costs associated with PPPs in many countries—the high cost of PPP project preparation, especially when compared with the costs of traditional public procurement. Preparation costs include the legal, financial, and technical costs incurred by both public and private sector actors in developing a PPP for commercial operation, and so include “transaction costs” associated with PPP procurement processes and contract negotiation, as well as (especially in some developing countries) “upstream” legal, regulatory, and policy preparation tasks that go well beyond normal transaction costs.

In their review of the PPP literature, De Schepper, Haeqendonck, and Dooms (2015) noted widespread agreement among practitioners and academics that PPP preparation costs are higher than preparation costs associated with traditional public procurement. In many cases, these costs are so high that they discourage potential bidders from competing for projects, and in some cases undermine the basic cost-effectiveness rationale of PPPs and negatively impact on the economic and financial viability of projects. De Schepper’s study of 172 public infrastructure projects in Belgium found much longer and more complex bidding processes associated with PPPs than with projects procured via traditional public methods.

There are several reasons why PPP preparation costs are so much higher than the costs of preparing projects involving traditional public procurement. The survey by Burger and Hawkesworth, which found widespread use of “public sector comparators” to ascertain value for money of PPPs, also noted that the process of preparing a PSC and requiring each bidder to prepare a projected PPP model is costly and time-consuming. Complex PSCs can take several months to finalize, and the resulting reduction in net benefits for potential private partners may cause them not to bid on projects that are deemed to be too small, with potential revenues not high enough to compensate for preparation costs. On the other hand, some projects are avoided because they seem too large and complex, leading to expensive bid preparation and time and cost overruns resulting from protracted contract negotiations. This may explain the findings of Zitron (2006), which showed that on average, only three bidders competed for each PPP contract in the United Kingdom, and that there were fewer than three bidders on one-quarter of the contracts. Governments can partially or completely compensate firms for their bidding costs, but of course this increases the cost of the PPP option relative to the traditional public procurement option.

Twelve of the countries in the Burger and Hawkesworth survey said they used some form of cost-benefit analysis to assess value for money for projects involving traditional public procurement, but these appear to be much less rigorous tests than PSCs and do not constitute the public-sector equivalent of “private sector comparators.” Only a few countries applied criteria to all prospective projects to establish which mode of procurement would result in the highest value for money. In other
words, for many countries traditional public procurement remains the lower cost, less complicated, default option for procurement, and the PPP choice depends on the discretion of departmental officials.

There is no single widely accepted metric for infrastructure project preparation costs. Costs may be measured as a percentage of initial capital costs, construction costs, or even the total net present value of the deal over its entire lifetime. Most published work has focused on “transaction” costs, usually synonymous with “downstream transaction costs,” and mostly procurement-related costs incurred by governments or their MDB partners. A sampling of different metrics follows below:

- In a widely-cited study of the procurement phase costs of bidding and negotiating contracts for 55 infrastructure PPPs in six sectors of the U.K. economy, Dudkin and Valila (2005) concluded that total transaction costs of this type averaged over 12 percent of the capital value of the projects, with the public-sector cost at about 3.5 percent, the winning bidder’s cost at about 3.8 percent, and costs to losing bidders totaling about 5 percent.
- The World Bank (Lin and Doemeland 2012) estimated that MDB and government preparation costs for the Nam Theun 2 hydropower project in Lao PDR (total investment: $1.4 billion) amounted to $124 million at project close in 2005, or 9 percent of total project investment.
- Farajian (2010) found lower average transaction costs in several case studies of large transport PPPs in the United States, but concluded that the transport agencies involved were not reporting all relevant costs. He also found that some individual state departments of transportation were requiring that project managers’ budget up to 10 percent of total investment for procurement and contract negotiations managed by state officials.
- Castalia Strategic Advisors (2010) estimated government transaction costs of PFI school projects in Australia and New Zealand to be 10 to 11 percent of total project construction costs, with significant premiums to be expected for first-time projects in either country.
- Bhattacharya, Romani, and Stern (2012), found that for projects in developing countries involving limited experience with the type of project or technology, or in a low-capacity country, preparation costs, including costs of design and arranging financing “can constitute up to 10 percent of overall investment costs.”
- The AfDB (2013) estimated that their preparation costs for large infrastructure projects in Africa can reach 10 to 15 percent of project capital costs.

These kinds of studies suggest that PPP project preparation costs in developing countries are much higher than they are in OECD countries because of the need to include costs for things like upstream preparation and premia for new or particularly complicated sectors like hydropower. These costs are also considerably higher than for projects involving traditional public procurement because of the need in many cases to carry out value-for-money analyses (which are typically less onerous for public procurement) and use more complex bidding processes.

Under what conditions can governments and their development partners deal effectively with the time and costs involved in project preparation? For many years, the default solution was to make as much of this preparation as possible the responsibility of the private sector. This attitude was an outgrowth of the notion that
implementing a PPP meant handing over government problems with infrastructure to private companies for solution. Probably the first notable example of this approach was the Buenos Aires water concession, signed in December 1992. This was one of the first, large brownfield infrastructure concessions. A “defining feature” of the tender process was a lack of information about the water system and its problems (Alcazar, Abdala, and Shirley 2000). Bidders were supposed to do all of their own due diligence, and the winning bidder was supposed to refine its implementation plan in the first few months after signing the contract. A protracted series of contract renegotiations would follow because latent defects made the project much more expensive than anticipated.

That approach has been frequently copied over the years (e.g., Kotze, Ferguson, and Leigland 2000), but by the early 2000s, donors and MDBs began to recognize that: (i) private partners cannot by themselves fully prepare PPP projects in a way that optimizes economic benefits; and (ii) a much more substantial effort by donors and MDBs to pay for and supervise preparation, before private partners become involved can result in more effective, sustainable, and pro-poor projects. The stronger role for governments and their development partners in identifying problems and designing solutions for private partners to implement is a characteristic of hybrid management contracts being developed or implemented in countries like Benin, Liberia, and Sierra Leone (Republic of Sierra Leone 2015). These contracts shift risks away from private partners, toward governments, donors, and MDBs, who are, theoretically, better able to mitigate those risks. Ultimately, this should make the contracts more productive and sustainable.

On a broader scale, a number of multi-donor project preparation facilities were established in the 2000s to deal with the high costs of preparation, but evaluations later showed that these facilities were too small and too bureaucratic to have much of a substantial impact (Cambridge Economic Policy Associates 2012). As noted in the introduction above, much larger preparation platforms have been created more recently to try to address the problem, but they have not yet had a chance to show results in terms of sustainable PPPs.3

Finance and Investment

PPP advocates expect, and sometimes promise, that such contracts will lead to improved efficiency and increases in investment from private sources, consequently leading to greater service capacity and coverage. But again, the evidence suggests that this is not consistently the case. As various studies demonstrate, private sources of finance have not played as strong a role in infrastructure PPPs in developing countries as hoped. In Africa, the long-standing practice in some sectors of government on-lending concessional finance to private concessionaires illustrates this fact. Many of the rail concessions in SSA benefited from this kind of debt, sourced as relatively
inexpensive sovereign guaranteed loans from bi-lateral donors or MDBs. These concessions often needed financial help because concessionaires could not or would not borrow to finance assets like rail lines, with operational life-spans being much longer than the terms of the concession contracts. Governments themselves often filled this gap by providing a large share of project financing. In the 1990s, this kind of on-lending was usually done with an interest rate mark-up to the concessionaire to cover government administrative costs and loan loss liability, but was still lower than the commercial rates that the concessionaire could obtain. As commercial lending for all kinds of brownfield concessions began to diminish following the Asian Crisis, the mark-ups were slashed to compensate for diminished global project finance lending, and by the mid-2000s, in some cases the mark-ups disappeared entirely (Pozzo di Borgo et al. 2006).

A study by Gassner, Popov, and Pushak (2009), examined whether water and electricity utilities involving private participation outperformed those run by governments. The study used a data set of more than 1,200 utilities in 71 developing countries. The results of the study showed that the private sector delivered higher labor productivity and operational efficiency, convincingly outperforming comparable companies that remained state owned and operated. But the study found mixed evidence regarding increases in investment and could not conclude that investment usually increases with private participation.

Gassner addressed the investment question in terms of different sectors and types of private participation. For electricity divestitures, investment per worker increased with private participation, as economic theory predicts. For lease and management contracts, which are more common for water and sanitation than for electricity, private partners were usually not obligated to make major capital investments, but Gassner’s results suggested that even if private participation generated operational improvements, the public utility companies involved did not increase investment. And for concession contracts, which also tend to generate improvements in operational efficiency, “... there is no conclusive evidence that investment increases.” (Gassner, Popov, and Pushak 2009). If efficiency improves, but is not supported by increases in public or private investment in the maintenance and expansion of utility networks, Gassner asked how operational improvements could be sustained over the long term.

A second study, by Marin (2009), undertook a global review of public-private partnerships for urban water utilities in developing countries. The study focused on 65 large urban water PPPs in operation for at least five years. Marin concluded that the principal benefit of these projects was operational efficiency rather than their ability to supply private sector finance. The author also acknowledged that private investment was the main attraction of PPPs in the water sector in the 1990s, but added that “experience has shown that this was largely the wrong focus,” (Marin 2009).

A host of other studies have found PPP-related efficiencies generated under different conditions, in different sectors, including electricity distribution,
telecommunications, and water distribution (Estache and Rossi 2004; Andrés, Schwartz, and Guasch 2013), and transport (Perelman and Serebrisky 2012). Estache and Philippe (2012) survey many other studies showing some measure of PPP efficiency in different sectors. But again, these studies do not consistently show a corresponding increase in investment.

Government on-lending to PPPs accelerated in some countries after the onset of the 2008 global financial crisis, which caused a collapse of the syndication market, inactive bond markets, and unusually high costs of funding. Governments and DFIs used concessional lending to meet funding shortfalls resulting from projects unable to raise sufficient commercial debt finance on acceptable terms. The U.K. government attempted to salvage its PFI with a proposed investment fund that would provide up to 100 percent of the financing needed for projects to reach financial closure (Farquharson and Encinas 2010). In developing countries such as Brazil, India, Mexico, and South Africa, government-owned development finance institutions provided funding or guarantees to kick-start PPP projects. Because countries like Brazil and India accounted for such a large share of private investment activity in developing countries, their use of DFIs and local public banks to fund large projects kept global PPI market investment numbers from declining as much as expected from 2008 to 2009 (Izaguirre 2009).

The global financial markets have still not fully recovered from the financial crisis and resultant increased financial regulation, at least not in the sense of returning to earlier levels of high liquidity, substantial risk appetites, and low cost of project finance debt provided by commercial banks. Governments, especially those in larger middle-income countries, continue to replace scarce commercial finance with public money from government budgets or donors and MDBs. Recent efforts by the World Bank’s PPI Database to collect detailed information on PPP financing sources in developing countries confirm that the private sector plays even less of a PPP financing role in low-income countries. In 2015, 57 percent of PPP funding in IDA countries came from governments, multilaterals, and bilateral institutions, with the balance coming from commercial sources. For all developing country regions in 2015, the percentage was 41 percent (table 2). In her analysis of this data, one World Bank staff member noted that, “One of the prevailing notions about PPPs is that upfront costs are wholly paid for by the private sector... However, this is a myth,” (Chao 2016).

What does this critique say about the conditions under which affordable private financing can be accessed for infrastructure PPPs? There clearly are factors that make a difference in terms of financing costs involving the particulars of a project like country, sector, type of contract, risk mitigation features of the contact, and the mix of financiers. The stronger a country’s sovereign credit ratings, the easier it seems to be for infrastructure PPPs to raise affordable private finance (IMF 2006). In terms of sectors, power generation is probably the easiest in which to raise private finance, largely because the supply of power is relatively commercial, with cost-reflective end-user
tariffs being less controversial for power than for services like water and sanitation. Electricity generation PPPs can also be purely private operations, often structured as Independent Power Producer (IPP) projects involving ownership of the assets by the private sector rather than by government as in traditional concessions. Private financing for power generation can also be relatively affordable because generation is not a natural monopoly service in the same way that some other infrastructure services are, such as water supply—multiple facilities can be built to feed power into national grids, so generation can be relatively competitive. This puts some downward pressure on financing costs.

IPPs may be popular, but affordable private finance for all kinds of large PPPs usually requires a host of risk mitigation mechanisms. Most IPP power purchase agreements (PPAs) must be backed by security arrangements such as escrow accounts, letters of credit, targeted subsidies, budget commitments, etc. In countries without domestic capital markets that can finance PPP projects, PPAs often must be denominated in hard currencies such as U.S. Dollars or Euros, indexed to currency baskets, or backed by foreign exchange liquidity facilities. Sovereign guarantees are also usually required to back various aspects critical to PPP project cash flows and profitability, including off-take commitments, fuel supply availability, currency convertibility and transferability, interest rates, exchange rates, tariff rates, and revenue levels. Donors, MDBs, as well as private institutions also provide guarantees or insurance products to cover risks that private lenders or investors are unable or unwilling to take. Finally, governments, donors, and MDBs have all increased their financing shares in large PPPs to reduce the size of private project finance and help make it affordable.

But when governments assume or share project investment risks, they need to manage conflicts of interest. For example, can governments act simultaneously as financiers interested in the financial sustainability of projects, as well as regulators charged with protecting the interests of end users? Will a government allow a concession company in which it has invested substantial amounts of capital to declare bankruptcy?

Table 2. Breakdown of PPI financing by region, 2015, in $ billions

<table>
<thead>
<tr>
<th>Region</th>
<th>Total Investment</th>
<th>Public Money</th>
<th>Donors/MDBs</th>
<th>Private Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin America</td>
<td>17.3</td>
<td>39%</td>
<td>14%</td>
<td>46%</td>
</tr>
<tr>
<td>East Asia</td>
<td>10.1</td>
<td>13%</td>
<td>4%</td>
<td>83%</td>
</tr>
<tr>
<td>SSA</td>
<td>6.2</td>
<td>13%</td>
<td>21%</td>
<td>64%</td>
</tr>
<tr>
<td>South Asia</td>
<td>5.4</td>
<td>30%</td>
<td>21%</td>
<td>49%</td>
</tr>
<tr>
<td>All Regions</td>
<td>39.0</td>
<td>27%</td>
<td>14%</td>
<td>59%</td>
</tr>
</tbody>
</table>

Source: Chao and Saha 2016.
Does it ultimately matter if private participation leads to operational efficiency, but not more investment? Lastly, if services become less costly and their availability increases, why should a lack of investment be a concern?

The problem is that PPPs tend to involve complex contracts that are difficult and costly to prepare. Most of the time this effort and expense is justified on the basis that more investment will be forthcoming. The absence of such investment substantiates Klein’s point that very little separates PPPs from well-run public companies, and argues for more attention to basic management improvements and better regulation before considering private participation.

Poverty Impacts

If a PPP generates investment, or even if it achieves only efficiency gains, it can make possible a host of other benefits like improved service quality, affordable tariffs, or expanded service access. In poor countries, these kinds of benefits are usually cited as justification for involving the private sector in infrastructure service delivery. But Estache and Philippe (2012) note that there is relatively little empirical evidence that focuses on the development impacts of small or large-scale PPPs. What evidence there is suggests that such projects do not provide the poor with enough affordable access to services.

Gassner’s study, cited above, shows why this topic can generate so much controversy. She compared average annual values for indicators measuring performance before and after private sector involvement in water and electricity utilities. Private involvement (via PPPs or privatization) generated a number of impressive efficiency gains, but they were not associated with an equitable distribution of project benefits. For example, the labor productivity gains were associated with reductions in staff numbers for both water and electricity. Employment fell by 24 percent in electricity and by 22 percent in water following the introduction of private participation. Utilities operated by governments used considerably more employees than privately operated utilities to achieve the same level of service. Moreover, the efficiency gains were not associated with changes in investment or average residential tariffs. Because the gains would normally translate into lower costs for the operator, Gassner speculated about reasons why there was no sign of the lower costs translating into greater investment or lower prices. One possible explanation was that services were already so underpriced that even huge efficiency gains could not justify price reductions. Another possibility was that the private operators took advantage of weak regulatory oversight to simply retain all of the gains as profits, passing on nothing to customers or to cover O&M costs. This latter possibility raises questions about the long-term sustainability of PPP efficiency gains.

In their survey of PPP research, Estache and Philippe (2012) concluded that investment and efficiency gains in the telecommunications sector (resulting mostly
from technological advances) have generally contributed to the wide distribution of benefits like increased access, improved affordability and better service quality. But in most other sectors, even when benefits result from improved efficiency or investment, they are not always shared with customers or governments. This is because regulation is not typically designed to pass on such gains to other stakeholders like residential users: “For the most successful projects, unless regulation works, efficiency gains become rents which fuel conflicts between government, users and operators,” (Estache and Philippe 2012).

How can PPPs be done in such a way that the likelihood of developmental benefits is strongly increased? PPPs that successfully generate benefits for the poor residents of developing countries usually begin with the realization that such benefits do not automatically result from increased private investment or improved efficiency. And as long as private partners are mostly responsible for developing work plans, as part of the bid process and later during early stages of implementation, there is no reason to expect effective measures for benefit distribution. This is not a particularly new problem. PPP contract designs and tariff structures that fail to benefit—or even disadvantage—major constituencies of stakeholders have led to renegotiations and project collapses from the early days of infrastructure concessions in the early 1990s. Many recommendations have been made about the need to undertake stakeholder analysis by modeling the distribution of project benefits before a contract has been awarded (e.g., van den Berg 2000). But again, this kind of preparation requires financial resources and expertise that must be brought into play before PPP contracts are signed.

**Governance Issues**

Governance refers to the various ways in which governments’ institutional capacity or lack thereof, intentionally or unintentionally affects the performance of PPPs. Political leadership is a rare but highly important way in which governments can facilitate significant infrastructure investment via PPI arrangements (Jones, Jammal, and Gokgur 2002; Eberhard, Kolker, and Leigland 2014). But generally, strong political support for PPPs in low-income countries is rare. A country’s investment climate can also have a major positive impact on the development of PPP projects (IMF 2006), but almost no low-income countries have investment grade credit ratings that would reflect strong investment climates.

Engel, Fischer, and Galetovic (2014) argued that the governance structure under which most PPPs operate is “usually defective.” This is largely because PPPs share many of the same defects as standard public works, including improper design, poor procurement, and frequent renegotiations that Engel and his co-authors conclude are often opportunistic attempts by private contractors to increase profits. A recent study by the World Bank’s Public Private Infrastructure Advisory Facility (PPIAF) found
that in the interests of getting projects to completion, many developing countries use limited competition and direct negotiation to procure private sponsors and operators (PPIAF 2014). But PPIAF also found that projects procured in this manner tend to be more expensive and subject to more problems in implementation.

Eberhard and Gratwick (2010) found that few African countries have established the key legal and regulatory elements needed to guide private participation in their power sectors—a clear policy statement and supporting legislation. Some countries have passed laws permitting independent power producer (IPP) projects, but almost none have addressed the relationship between IPPs and state-owned energy providers. In a more recent study, Eberhard et al. (2016) conclude that in general, Africa’s power sector continues to reflect a partial approach to reforms that facilitate more private investment, with little achieved over the last 20 years in terms of unbundling, privatization, or the introduction of competition.

After “PPP units” became prominent in countries like the United Kingdom and Australia, they have become the preferred institutional tool to regulate the process of PPP project development in many developing countries. A variety of international organizations have recommended establishing PPP units, including the OECD (2010), the UN (2007), and APEC (2008). The rapid growth in the number of such units seems to have resulted from a widespread perception that well-functioning PPP units are important for the success of a country’s PPP program. However, a recent literature review by Lemma (2013) concluded that “there is very little quantitative evidence of the value of centralized PPP coordination units...”. There seems to be consensus that a large and growing number of such units exist, but they differ so widely by country and sector that it is difficult to generalize about their value, or even agree on how to measure their worth (Burger and Hawkesworth 2011; EPEC 2014).

The contribution of PPP units to achieving success may not be entirely clear, but the reasons for their ineffectiveness are more so. One of the earliest studies of these units, sponsored by PPIAF (2007), noted that they were often established due to frustration on the parts of governments with the inability of existing institutions to move PPP programs forward. But the study found that in such situations the governments often overlooked the fact that the very shortcomings that made the existing institutions underperform may well undermine the new PPP units as well (PPIAF 2007). The PPIAF study concluded that in countries where top politicians do not support the PPP program, where procurement of infrastructure and capital works is not transparent and competitive, and where the machinery of government is chronically uncoordinated, PPP units are unlikely to be effective. The study concluded that there is little chance that PPP units can work well where such problems persist.

One problem with PPP units that seems to have limited their usefulness is their disappointing performance in accelerating PPP deal flow. Government officials in developing countries often expect such units to speed up the process of getting PPPs to completion. But such units often produce exactly the opposite result—the flow of
deals slows down because of the perceived need to carefully apply standardized procedures in project development. Better-prepared projects should be more sustainable, and that should off-set lengthy preparation periods to some degree. But this logic rarely satisfies government officials interested in getting projects done as quickly as possible.

South Africa’s PPP Unit, which was established in mid-2000 in the National Treasury, and which boasted procedures and guidebooks that quickly became good practice examples used by other national agencies in Africa, is a case in point (Lemma 2013). The guidelines called for over two dozen relatively elaborate preparation steps, as well as a series of National Treasury approvals of the process recommended by the Treasury’s PPP Unit. The entire process (including the use of “public sector comparators”) was quickly judged by the private sector to be too lengthy, too complex, very expensive, highly over-regulated, and unfortunately responsible for slowing down PPP deal flow (Castalia Strategic Advisors 2007). The PPP Unit itself admitted that its required VFM studies at the municipal level averaged about 30 months (Levinsohn and Reardon 2007). As a result, under this regulatory regime, PPP deal flow remained stagnant in South Africa.

In 2012, the PPP Unit lost its status as a separate department within National Treasury when it was absorbed by the Government Technical Advisory Centre (GTAC), and its role changed from process regulator to the provider of PPP transaction advisory services (GTAC 2015). The experience of South Africa’s PPP Unit seems consistent with the findings of several researchers, confirming that complicated frameworks for regulating the development of PPP projects can significantly slow the process of getting projects to financial close (Zhang and Kumaraswamy 2001; Chan et al. 2010).

PPIAF’s long experience with the creation of PPP units, particularly in Africa, suggests two conditions under which these entities can add value. First, if units have no authority to require compliance with their recommendations, they tend to be quickly relegated to minor advisory roles without much influence. This is because in poor countries governments cannot afford to employ the sort of experts whose advice is of significant value to public or private sector project proponents. Sometimes experts can be seconded to PPP units from private sector firms, or recruited for limited periods with the help of donor funding. But these kinds of experts typically cannot be retained long enough to transfer significant levels of professional skills to regular staff (and if they do, regular staff members tend to leave for higher-paying jobs). To enforce government rules regarding how PPPs should be developed and structured, as opposed to just providing advice, units can function reasonably well with a small central core of expertise. The rest of the staff component can be made up of competent project managers with more general kinds of professional skills.

Second, the rules that PPP units enforce should minimize direct pre-approval of actions by senior government officials based on vague decision-making guidance,
in favor of routine, post-audit reviews of actions based on clear, specific procedural guidelines or regulations. Senior officials can become involved after the fact, if the post-audit reviews reveal that proper procedures were not followed. The pre-approval approach encourages arbitrary decision making and can slow implementation to a standstill. But the post-audit approach is not easy to implement: it requires clear and complete procedural guidelines or regulations, along with consistent, reasonably high-quality auditing skills and procedures. PPP units will not work well in countries already plagued with severe governance defects affecting standard public works procurement and contracting. This is one of the reasons why PPPs are not ideal tools for every developing country.

Transformational PPPs

In light of the preceding discussion, what is to be made of the G20’s enthusiasm for regional or “transformational” PPP projects? Indeed, why would such projects be any more successful than smaller and simpler PPP projects in meeting expectations regarding costs, profits, finance, investment, equity, efficiency, and governance? Although the term is widely used, there is no precise, universally-accepted definition of “transformation project”, but as noted previously, the term is normally used to refer to large, regional, or cross-border infrastructure projects involving private participation. Various prioritized lists of such projects have been compiled by several interlocutors, including the World Bank (2011), the G20 MDB Working Group (2011) and the African Union Commission (2012).

Such regional projects are extremely rare. SSA’s actual experience with regional infrastructure PPI projects is almost non-existent. The World Bank’s PPI Database has recorded only seven regional infrastructure projects on the African continent since 2000 (all in SSA). Five have been transport projects, while; two have been natural gas transmission projects (World Bank, PPI Project Database). But it would be misleading to imply that large regional projects are easier to do in other regions. Since 1992, the PPI Project Database has recorded only 15 cross-border projects globally, in any infrastructure sector. The disadvantage of so little experience with large regional projects is clear. Koppenjan (2008) notes that the learning curve associated with structuring mega-projects as PPPs is extremely steep and such projects seem to be most successful in countries and sectors with a considerable amount of PPP experience at a variety of levels.

Those who defend these large PPP projects point to the advantages of economies-of-scale and of scope. This “bigger is better” approach is also consistent with popular theories of economic development, which argue that big development challenges like poverty alleviation, energy scarcity, and urbanization can only be solved by “big push” solutions involving large projects and large amounts of aid (Sachs 2006). However, these “transformational” projects frequently possess the size and complexity that
would make them subject to what Flyvbjerg would call the “iron law of megaprojects.” According to Flyvbjerg (2014), 90 percent of projects costing over one billion dollars, end up “over budget, over time, over and over again.” This is a problem for all large projects, not just PPPs, but Flyvbjerg points out that “Private capital is no panacea for the ills in megaproject management, to be sure; in some cases, private capital may even make things worse...” Flyvbjerg (2014). Part of the reason for this is that the complexity of large PPPs—with multiple players, complicated and lengthy contract negotiations, high transaction costs, forecasting uncertainties, etc.—leads to what Ansar et al. 2016 call a low “investment fragility threshold”—a high vulnerability of the investments to become unrecoverable due to the impact of random events.

With large, transformational projects, economies-of-scale are cited as a factor that should help reduce the amount of preparation costs. But using the cost metrics discussed earlier, and even assuming some economies-of-scale, the costs of preparing transformational PPP projects are staggering. The average total cost of the 14 regional power sector projects in the Priority Action Plan of the Program for Infrastructure Development for Africa (PIDA) is $2.9 billion (AUC 2012). Preparing projects of that size could require as much as $430 million in total development costs, with roughly one-third of that ($140 million) being the responsibility of the government and its development partners. These development costs tend to be more expensive than costs incurred for developing conventional publicly procured projects, because in addition to more complicated contracts and contact negotiations, they involve much more up-stream preparation. Many low-income countries do not have a specific PPP legislative framework conducive to large PPPs or a single set of procedures for the award of a suitable PPP contract. Without such a legal-regulatory framework in place, project sponsors need to obtain multiple licenses, permits, and authorizations from various governmental and local authorities in order to implement the project. All of this increases the political risk for private investors, in case one or more key permissions cannot be obtained, and makes it more difficult to have these licenses, permits, and authorizations assigned to the project financiers as a form of security under the financing arrangements.

What role would private finance play in these large, regional PPPs? Inga III, a regional power project which is probably on more lists of “transformational” infrastructure projects than any other in SSA, suggests some answers to this question. When it came under serious consideration in the early 2000s, Inga III was intended to be the third and largest hydropower project in the Inga basin of the Congo River in the Democratic Republic of the Congo (DRC). Private sector participation has always been described as a requirement for the project rather than an option because the combination of the government plus donors and MDBs did not have the kind of capital required for project investment (World Bank 2014). But as the project developed, it
became clear that private partners would not necessarily be responsible for all or even most of the finance required for the project.

In a “preliminary financing feasibility study” commissioned by the World Bank (BNP Paribas 2009), the consultants estimated the total cost of the project at about $9 billion for a facility that would generate 4,500 MW. The debt share of total costs ($6.7 billion) would be provided by four MDBs, six DFIs, several export credit agencies, and Chinese policy banks. Commercial banks were expected to provide no more than about 10 percent of the total, assuming that MDB credit enhancements could be accessed from the World Bank’s Partial Risk Guarantee (PRG) facility and the Multilateral Investment Guarantee Agency (MIGA). The study assumed that one-third of the equity needed for the project (about $2.2 billion) would come from the government, sourced as loans from MDBs. This government equity contribution (about $730 million) would be in addition to project preparation costs that could be as much as $400 million for a project of this size, using the metrics discussed above.

At about $1 billion, the government contribution needed to make Inga III viable as a PPP seems untenable. In mid-2010, the government reached the completion point under the Enhanced Heavily Indebted Poor Countries (HIPC) initiative after benefitting from assistance under the Multilateral Debt Relief Initiative (MDRI). By 2013, the DRC’s total outstanding central government debt owed to bi- and multilateral creditors was $2.9 billion (IMF 2015). This total reflected a marked improvement over the 2009 level. But the DRC’s debt sustainability rating was still assessed as “moderate” by the IMF, meaning that any sharp increase in external debt, even to MDBs or donors, risked the return of debt distress. The wisdom of a precipitous 30 percent increase in such debt to support Inga III would no doubt be questioned. The conclusions regarding Inga III seem clear: it would not be predominantly financed by the private sector, leaving the government with a financing burden that it could not afford. The alternative to a transformational PPP in this case is not a project done using traditional public procurement. Rather, it is a collection of smaller projects (public or private) that are more easily financed and managed.

Is there any evidence to suggest that benefits resulting from efficiencies or investment generated by large transformational PPPs would be distributed equitably across the country’s population? In terms of equity arguments, MDBs have not been consistent in their messaging. On the one hand, the World Bank (2011) argued in its infrastructure strategy update that “Large infrastructure projects have often been successful in making project affected people the beneficiaries of the project displacing them, as well as achieving development objectives, like the benefit sharing arrangements in hydropower.”

But the same document (World Bank 2011) acknowledged that the Bank’s strategy of pursuing economic growth through such infrastructure projects, with the expectation that benefits will “trickle down” to the poor, had not been borne out by the facts, and that the results of this approach have often been “slow”. Poverty impacts
have proven “complex to achieve and demonstrate,” and “Learning from past experiences, the Group will do more to enhance the delivery of infrastructure services to the poor,” (World Bank 2011).

At least in the power sector, views on the socio-economic value of large regional projects may be changing. The International Energy Agency (IEA) has acknowledged that 70 percent of rural areas in developing countries are best electrified with mini-grid and other off-grid solutions (mostly wind and solar), and that these kinds of solutions would need to receive most of the new investments recommended for the power sector (International Energy Agency 2011). The value of this small-scale approach seems to be supported even in the DRC, where the government’s Inga III project director recently acknowledged that the project will not generate many benefits for rural areas, and that off-grid solutions involving wind, solar, and small hydropower will be needed to adequately serve populations in those areas (Worldfolio 2016).

The larger the PPP, the more important it is to have knowledgeable and well-capacitated partners, including governments who must play a principal role in project preparation, procurement, and post-closure performance monitoring. In the poorest countries, that kind of capacity does not exist, so donors and MDBs try to create it, sometimes specifically for transformational projects.

But how effective is this kind of just-in-time institutional development? Again, Inga III serves as an instructive example. As a condition of receiving an IDA grant for project preparation studies in 2014, the DRC Government was asked by the World Bank to adopt a key element of improved project governance. By 2015, the government agreed to create a ring-fenced development authority (Agence pour le Développement et la Promotion d’Inga – ADEPI) to manage Inga’s development, and mobilize private participation and public financing. ADEPI was supposed to be created by law as an autonomous entity reporting to the Prime Minister’s office. It was to have a Board of Directors representing various Inga development stakeholders. All ADEPI staff members, including the Director, were supposed to be recruited competitively. In other words, ADEPI was expected to act as a kind of independent PPP unit specifically for the Inga project, combining financial resources and expertise to ensure that responsible best practice was followed in the development of the project.

The grant was made, but the government evidently did not feel compelled to meet these obligations. On September 12, 2016, the World Bank (2016b) cancelled its grant support to Inga III. The activity had already been suspended in July due to the government’s decision to “take the project in a different strategic direction” to that agreed with the Bank in 2014. At the time of suspension, only 6 percent of total preparation financing had been disbursed; most of the studies had not been completed, and many had not been started. ADEPI had not been established as an autonomous entity (World Bank 2016a). In late 2015, President Kabila moved ADEPI into his cabinet and took personal control of the project. But this meant that the independent PPP unit functions would be lost and some of the key environmental and
social studies probably would not be completed. The lesson here also seems clear: governments without a recent history of legal, regulatory, and institutional reforms are unlikely to make rapid, significant progress in these areas during the course of a single large project.

Are there ways of increasing the chances that transformational PPPs will be successful? There are very few examples of successful PPPs of transformational scale, but Ansar and his colleagues (Ansar et al. 2016) make recommendations regarding ways of avoiding problems with mega-projects, which apply equally well to large PPPs: (i) be skeptical of the numbers presented at appraisal; (ii) seek to de-bias estimates of time to task completion, costs, and benefits by demanding more extreme stress tests; and (iii) look for evidence of meaningful cost-benefit analysis as part of appraisal.

A second way to improve the chances of mega-PPP success is to involve multiple countries in the project and ensure that one or more of the project partners has some PPP capacity and can take a leadership role in the project. The N4 Toll Road Concession, awarded in 1997, linked South Africa’s most industrialized, but effectively land-locked northern and eastern regions (Gauteng and Mpumalanga provinces) to the Mozambican port of Maputo. When the project reached closure, it was recognized as a pioneering accomplishment—it was the first toll road concession signed in SSA. It was the first cross-border transport PPP, and only the second regional PPP in any sector. The project was successful because of high-level support that came directly from the presidents of the two countries, but also because of South Africa’s technical and managerial expertise, and their already considerable experience with toll roads. South Africa led the project and built capacity among Mozambican counterparts along the way (Thomas 2009).

A third way to avoid problems with transformational PPPs is to do smaller PPPs instead. This is particularly the case when transformational PPPs involve massive capital investments based on razor-thin profit margins. When possible, projects of this kind need to be broken down into smaller projects. In the case of projects like Inga III, this suggestion is more practical than it may sound. Since the last wave of work on Inga III began in the late 2000s, a change in thinking about investment in renewable power generation has occurred in Africa, thanks to South Africa’s Renewable Energy IPP Program (REIPPPP) and the Scaling Solar Renewable Energy IPP Program in Zambia. Over four rounds of bidding between 2012 and 2015, South Africa’s program attracted $19 billion in private investment in 92 on-grid IPPs, totaling 6,327 MW, with wind and solar PV prices among the lowest in the world (Eberhard, Kolker, and Leigland 2014). Instead of being primarily about climate change mitigation, or off-grid service in rural areas, REIPPPP has shown how quickly renewable energy can add generating capacity to power grids, as compared with large hydropower or coal-fired plants. Construction times for these solar projects are typically under 12 months per facility. In other words, by 2018, South Africa will have 32 percent more generating capacity in place than Inga III is expected to provide after ten years of...
The Scaling Solar program, designed by the IFC for the Zambian Government, is a much smaller tender program, but has reduced power prices even more than in South Africa (IEA 2017). REIPPPP and Scaling Solar both involve extensive pre-bid project design, government and third-party guarantees, donor and MDB financing (of preparation as well as capital investment), and non-negotiable project documentation.

Conclusions: The Emergence of Hybrid PPPs

PPPs have not, in general, met expectations. Based on the findings of the survey by Burger and Hawkesworth (2011), it seems realistic to assume that PPPs may eventually account for 10 to 15 percent of public sector infrastructure investment in OECD countries and up to 25 percent in developing countries. These are of course not insignificant percentages, but they are nothing like what was expected of PPPs in the 1990s. In particular, the evidence shows that PPPs have never successfully gained a foothold in the poorest countries.

The studies reviewed in this paper suggest that governments and their development partners have not done a very effective job of anticipating PPP costs and profits, so assessments of value for money involving tools like public sector comparators have been frequently off the mark. This is true of all kinds of infrastructure projects, in developed as well as developing countries, but especially so in the case of PPPs. In the developing world, particularly in the poorest countries, PPP project costs have been higher than expected, and profits lower. This of course has contributed to the limited use of PPPs in these countries.

PPP project preparation costs are much higher than were expected in the 1990s, and higher than the costs of preparing projects for traditional public procurement. This is partly because of the upstream legal and regulatory preparation required in many developing countries, as well as the complexity and higher costs of PPP procurement compared with traditional government construction contracting (value-for-money assessments involving tools like public sector comparators also tend to be more rigorously applied to projects involving private participation, but most experts argue that such tools should be used for both PPP procurement and traditional government contracting). High preparation costs present governments and their development partners with a difficult problem: who pays for preparation costs that cannot always be recovered in some way from the project?

New data compiled by the World Bank in 2016 (see table 2), suggests another surprising conclusion about infrastructure PPPs: in many situations, private finance is not the predominant source of investment funding for these projects. Governments, or their state-owned banks, along with donors and MDBs, account for more than half of this funding in many cases, even in middle-income regions like Latin America. This
kind of public funding seems essential to attract private finance for many kinds of projects. But this reality challenges some of the traditional definitions of PPPs, especially with regard to the role of the private partner.

Perhaps one of the least surprising findings about infrastructure PPPs is the fact that developmental benefits in terms of poverty reduction are far from automatic. This has been a special focus of former World Bank economist, Antonio Estache, and his critical assessments have found their way into some of the most powerful civil society critiques of PPPs. Estache and others point out that positive poverty impacts must be purposely designed into projects, and then monitored by regulators (Estache and Rossi 2004; Andrés, Schwartz, and Guasch 2013). Private partners do not typically pursue these benefits without substantial incentivization by governments and donors.

Governance problems can have significant negative impacts on PPP project development and oversight. Governments must play a strong role in PPPs, but often the ultimate justification for doing a project as a PPP is the weakness of host government institutions and their lack of skilled staff. PPP units can help, but cannot be expected to work more effectively than other existing governmental agencies. It is not always an exaggeration to say that a government capable of fully playing its role in designing, developing, implementing, and regulating PPPs is probably better off using traditional public procurement to achieve the same objectives. This supports Klein’s point that well-run public companies match the performance of private firms in regulated sectors.

The bigger the PPP, the more prominent are all of these problems—“bigger is better” does not seem to apply to these kinds of projects. Part of the problem is simply the size of the project, but doing a project as a PPP accentuates the problems created by size. This is the reason why transformational PPPs are so rare, and existing projects like Inga III have encountered so many problems. But the best alternative to a transformational PPP is not necessarily a project done via traditional public procurement. It is scaling them down to smaller projects, or a collection of smaller projects, done as either public or private projects. Smaller projects reduce the overall investment fragility of these undertakings, making them more easily financeable. What still seems unclear is why the G20 and other international groups continue to promote mega-PPPs as solutions to infrastructure problems in poor countries. Civil society groups seem to have taken the lead in criticizing the G20’s approach (Alexander 2013). But despite the fact that these groups now routinely cite studies making up the emerging evidence-based critique of PPPs, their opinions have had no noticeable impact on G20 discussions.

If PPPs continue to account for 15 to 25 percent of infrastructure investment, this suggests that governments and their development partners may be finding some situations in which conditions are right for PPPs to provide value for money. The emerging evidence-based critique of PPPs suggests the type of conditions...
required, but these are not the same ones envisioned in the 1990s. It is no longer assumed that self-sustaining PPPs that fully cover all operating and capital costs are workable and desirable in a large variety of infrastructure service delivery situations in poor countries. The World Bank long ago stopped advising governments that were granting concessions to insist on full cost-recovery tariffs, as it did in the 1990s (Kerf et al. 1998).

The evidence-based critique of PPPs suggests that compared with the traditional PPP model promoted in the 1990s, successful PPPs today tend to be hybrids, combining some elements of traditional PPPs with those of consulting or EPC contracts. These hybrids involve less of a role, and lower risks, for private sector operators and financiers, and more of a role (and more risks) for governments and their development partners. This shift in roles is evident across the spectrum of criticisms discussed in this paper: (i) “blended” or subsidized finance now plays a much stronger role in ensuring private sector profits from PPPs; (ii) project preparation is no longer left solely to private partners, but is increasingly paid for and managed by governments and donors; (iii) significant amounts of private finance are available only if governments and MDBs provide guarantees and other risk-mitigation mechanisms and share the financing burden using public or concessional money; (iv) developmental benefits are achieved when projects are designed to include them, and this design work is no longer left to the private sector; and (v) governments must play a strong role in the PPP project development and oversight, but improving governance in poor countries typically requires considerable amounts of funding and pressure from donors and MDBs. In situations where even hybrid projects are difficult to carry out, mainly in the poorest countries or with mega “transformational” scale projects, PPPs have not been very successful, and are no longer frequently attempted.

Notes

1. This database defines “PPI projects” to include management contracts, utility privatizations, and merchant projects, as well as projects more widely thought of as PPPs.

2. In a 1998 publication that asked authors to speculate on the future of MDBs like the World Bank and EIB, Klein “looked back” from a 2044 vantage point and noted that “...the private provision of most economic and social services rendered the funding and guarantee functions of the World Bank group largely superfluous.” (Klein, 1998).

3. Examples are Africa50, established by the African Development Bank, and the World Bank Group’s Global Infrastructure Facility (GIF), both of which emphasize project preparation as well as financing.

4. Wind and solar power are much more variable than hydropower, and cannot provide the base load capacity that most developing countries need. Thus, this comparison between large hydropower and wind/solar is somewhat unfair. But when combined with small hydropower plants or gas-based generation, wind and solar could dramatically reduce the need for a massive hydropower investment in a country like DRC, especially if the primary objective is increasing energy access for the largely rural population.
References


European PPP Expertise Centre. 2014. Establishing and Reforming PPP Units: Analysis of EPEC Member PPP Units and Lessons Learnt. Luxembourg: European Investment Bank.


