Global Employment
Policy Review

Background Paper N°1
November 2019

The effectiveness of interventions to reduce informality in low- and middle income countries

Jonas Jessen

Jochen Kluve

The Effectiveness of Interventions to Reduce Informality in Low- and Middle Income Countries

Abstract

Labor markets in low- and middle income countries are characterized by high levels of informality. A multitude of interventions have been implemented to increase the formalization of firms and workers, including information campaigns, simplified registration procedures, reductions of payroll taxes, and interventions enforcing formalization. We compile a database of 157 impact estimates from 32 academic studies that evaluate empirically one or more of these formalization interventions. The quantitative analysis correlates the impact estimates of the studies — given as a measure of sign and statistical significance, the effect size or percent impacts — with explanatory factors such as intervention type, outcome variable, scope of the intervention, and other covariates. Several key findings emerge: first, the intervention type is not a strong determinant for the effectiveness of formalization interventions. Second, the outcome "worker registration" shows significantly better results than other outcomes. Third, interventions at scale are more effective on average than singular "programs".

JEL: C40 · J08 · J46 · J48

Keywords: Formalization · Labor Registration · Business Registration · Impact Evaluation

1 Introduction

High levels of informality are a central feature of labor markets in low- and middle-income countries. Whereas a precise definition of informality is challenging to give — Maloney (2004) defines it as "[b]roadly speaking, the small-scale, semi-legal, often low-productivity, frequently family-based, perhaps pre-capitalistic enterprise" — the fact that either firms or workers are not registered with the tax or social security system and therefore not formal participants of the labor market is regarded as a key concern. Specifically, there are several reasons why policymakers would worry about informality: first, it reduces the tax base, thus negatively impacting the provision of public goods. Secondly, it may lead to an inefficient allocation of resources, as formal and informal firms compete in the same market but have different marginal costs. Third, informal workers are not covered by any of the institutions — such as pension systems, health insurance, etc. — that protect formal employees.[1]

Figure 1 displays shares of informal employment (as a percentage of non-agricultural employment) in a series of countries in Sub-Saharan Africa (left hand side, blue color), in Latin America (center, grey color), and Asia (right hand side, red color). The figure illustrates that levels of informality are generally pervasive, and differ somewhat between major regions: in Latin America, these shares are in the range of 23.6 (Uruguay, 2017) and 77.3 percent (Bolivia, 2015), and thus lower than in Sub-Saharan Africa – where the majority of countries have shares higher than 70 percent – and Asia. Notwithstanding these regional differences and the general downward trend in informality since the early 2000s (e.g. Maurizio and Vázquez, 2017), the figure shows that overall levels of informal employment have remained high across the developing world.

In order to address this persistent challenge, a multitude of policies and programs have been implemented in many countries with the aim to increase the formalization of firms or workers, or both. Among initiatives to formalize businesses have been, for

[1] In describing and analyzing informal labor markets there have been debates on whether they are best described by workers being in the informal sector voluntarily or involuntarily, and thus whether formal and informal labor markets are segmented or not; or perhaps a mixture of the two. E.g. Khamis (2012) provides a concise overview of these issues. Our starting point is less about what precisely defines informal labor markets and how they are best characterized, but the fact that there is an economic rationale to address informality through a set of interventions.

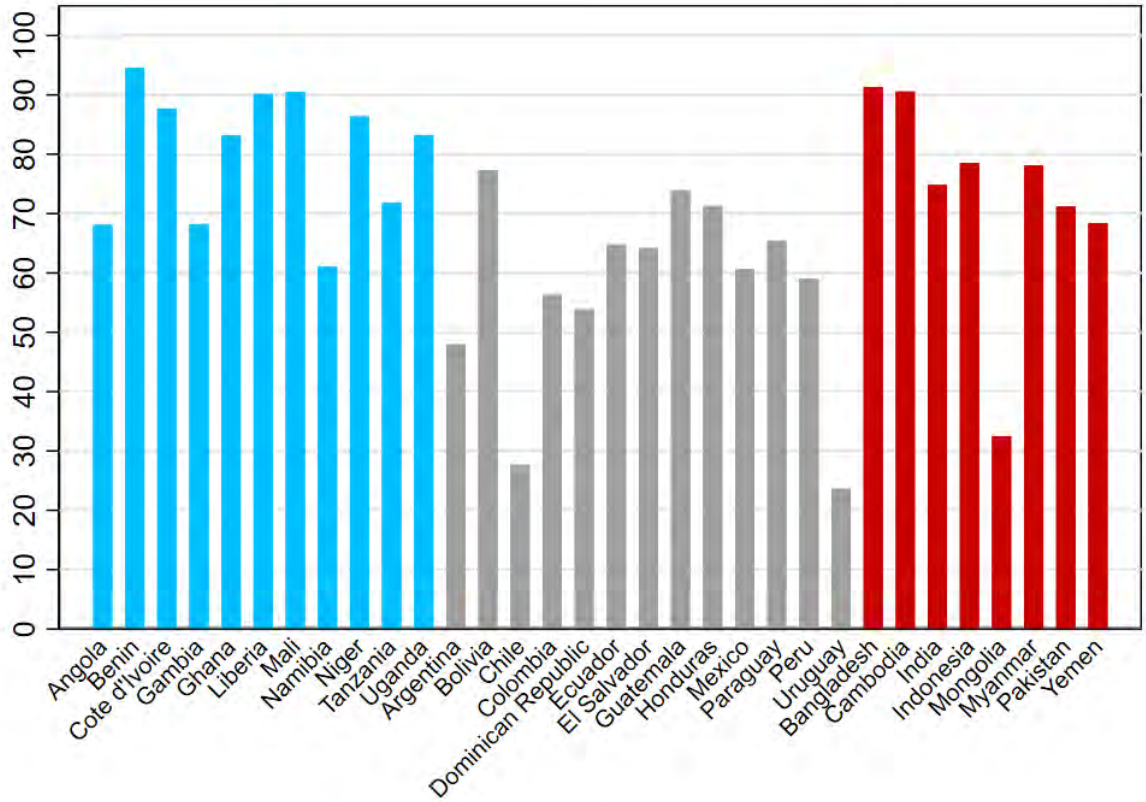


Figure 1: Informal employment as % of non-agricultural employment

Note: World Bank data. Shares reported are from 2017 or latest year available.

instance, the implementation of one-stop shops for business registration and the simplification of payroll taxes and social security contributions (e.g. Bruhn, 2011; Fajnzylber et al., 2011, for Mexico). Other approaches concern information interventions, e.g. information campaigns that explicate the step-by-step procedures and potential benefits of business registration (De Giorgi and Rahman, 2013, for Bangladesh). Also programs that reduce the costs of business registration have been considered and put into practice (Alcázar and Jaramillo, 2016, for Peru), as are financial mechanisms in which a bonus payment is given to firms who are willing to register (de Mel et al., 2013, for Sri Lanka). Finally, a potential policy alternative to incentive-based approaches are interventions that enforce business formalization (e.g. De Giorgi et al., 2018, for Bangladesh).

The second type of approaches targets the formalization of labor, such as the registration of workers. This has included e.g. tax reduction and bureaucracy simplification policies such as SIMPLES in Brazil (Monteiro and Assunção, 2012). Other cost-reducing approaches include reductions in payroll taxes (e.g. Bernal et al., 2017, for Colombia) or the simplification of labor registration (e.g. Ronconi and Colina, 2011, for Argentina). Finally, as in the case of targeting businesses, also when targeting workers the enforcement of formalization legislation is a policy option, given that such legislation is in place (Pignatti, 2017, for Colombia). Clearly, several of the approaches mentioned here and in the above paragraph are potentially combinable into a multi-component approach.

Given the policy relevance of labor market informality and this large spectrum of interventions aiming to increase formality, the empirical evaluation of these interventions is of key interest to policy makers, to learn about the effectiveness of initiatives that intend to reduce informality. Some of the earlier evidence specifically on the formalization of small firms is reviewed in Bruhn and McKenzie (2014) and Khamis (2014), who find, for instance, that approaches focusing on the ease of formalization alone will not induce most informal firms to become formal, while increased enforcement of rules can increase formality. They also find that entry-reform type of interventions result in only a modest increase in the number of formal firms (at best), and, when looking at the "limited existing evidence base" at the time, call for further efforts to evaluate formalization policies and a broader perspective on intervention types, including credit and labor market policies (Bruhn and McKenzie, 2014).

Therefore, several additional policy questions are of key concern: what is known about which type of "formalization" intervention works? Is there a difference between the short-run and long-run effects? Are different outcomes affected differentially? Is it easier to "formalize" firms or labor? Does the scope of the intervention play a role? That is, are interventions more effective when implemented at scale ("policies", typically nationwide and permanent) or when implemented for narrowly defined groups, regions, or sectors ("programs", small-scale and often one-off)?

In this paper, we address these questions. We start with the compilation of a database of available studies worldwide that assess the effects of interventions that reduce informality, in low- and middle income countries. We then provide an empirical analysis of this database of "formalization interventions" in the spirit of a meta analysis (e.g. Stanley and Doucouliagos (2012), Card et al. (2017)) — that is, we quantitatively analyze the patterns of intervention effectiveness by, for instance, intervention type, outcome, time horizon, and scope of the intervention. This approach provides new evidence on the relative, and absolute effectiveness of formalization interventions across low- and middle income countries.

The following Section 2 provides a framework distinguishing five categories of formalization interventions that have been implemented in practice. The section also delineates the compilation of the database. Section 3 presents a descriptive statistical analysis of the data, while Section 4 explains and implements the quantitative approach to investigate correlates of the effectiveness of formalization interventions. Publication bias and precision-weighted estimation are also discussed and implemented. Section 5 concludes.

2 A typology and database of formalization interventions

2.1 Typology

To adequately distinguish between different forms of formalization interventions that have been implemented and evaluated in practice, we identify five main types classifiable as follows:

(i) information interventions

(ii) simplification / registration interventions

(iii) tax incentives / social security reduction

(iv) labor inspection / enforcement interventions

(v) financial incentives

First, information interventions provide informal firms — and/or would-be entrepreneurs — with information regarding (a) the registration process, and (b) the benefits of registration. The latter includes e.g. protection of the business name, (better) access to bank loans, limited liabilities, greater ownership rights, and the enhancement of social status. Information interventions are typically not provided as stand-alone interventions, but are often combined with simplification / registration interventions. The studies in our data (see below) that analyze this intervention type primarily stem from experimental evidence. De Giorgi and Rahman (2013), for instance, use a Randomized Controlled Trial in which informal firms were informed about a recent registration reform in Bangladesh that effectively reduced the duration of the registration of a business from 42 days to one day. The study focuses on randomized exposure to the information about the reform (not on the reform itself) and finds that one year after the treatment the self-reported knowledge about the reform increased among small and medium enterprises, but actual registration did not. The authors conjecture that lack of information may not be the main constraint, but rather higher taxes and regulation, and the (perceived) low benefits of registration.

Second, the purpose of simplification / registration interventions is to simplify business entry regulations, or business registration procedures. These reforms typically lead to a sizable decrease in the number of days required for registration. One example is the study by Bruhn (2011) who analyzes the Rapid Business Opening System (SARE) in Mexico implemented in the years 2002-2006. The reform reduced the average number of days for a business registration from 30.1 to 1.4. The study uses variation in roll-out across time and municipalities to identify the reform effects, and the results indicate that the total number of firms increased by 5 percent in eligible industries. The author identifies former wage workers opening businesses as the main channel for the reform effects.

Third, tax incentive and social security reduction interventions pursue the objective to reduce the "costs of being formal" by reducing the tax burden and/or social security contributions. This is intended to make the registration of firms and workers more attractive. An example is the SIMPLES reform in Brazil 1997 analyzed by Monteiro and Assunção (2012). The reform combined six different federal taxes and social security contributions into one monthly-based rate. The study uses a difference-in-differences approach with sectors affected and not affected by the reform, and finds a statistically significant and large positive effect in one eligible sector (retail), while the estimated effects are insignificant in four other sectors (construction, manufacturing, transportation, services).

The fourth category, labor inspection and enforcement interventions, intends to increase compliance with firm and/or labor registration regulations through enforcement. Such enforcement can take place e.g. through labor inspector visits, or through official letters from the tax authorities. E.g. Pignatti (2017) analyzes the Colombian "Action Plan" of 2011 which doubled the number of labor inspectors (from 400 to 800) within four years. The study uses variation in the roll-out across regions and over time, and finds a small and statistically significant impact on formal employment of workers. Importantly, the effect remains persistent over several years.

Finally, the fifth intervention category provides a financial incentive to enhance the potential effect of (i) information or (b) simplification / registration interventions. That is, it is typically not provided as a stand-alone intervention, but in combination with either of the two. For instance, de Mel et al. (2013) implement an RCT in Sri Lanka with informal firms and four treatment arms T1 through T4: (T1) provides an information intervention plus the reimbursement of the (modest) direct registration costs. (T2) through (T4) each provide the information treatment, too, plus an additional payment of the local equivalent of USD 88, 175, and 350, respectively. The empirical analysis finds that T1 did not show any effect; for T2, 17-22 percent of firms registered, for T3, 48 percent. There was no additional impact for T4.

2.2 Compilation of the database

The objective of the data compilation is to construct a database of the universe of impact evaluations and quantitative assessments of formalization interventions worldwide, focusing on low and middle income countries. This systematic process proceeds in three main steps (e.g. Kluve et al., 2019): 1) The first step is to search for relevant studies that analyze one or more of the intervention types defined in section 2.1; 2) the second step is to verify a set of inclusion criteria to arrive at the final set of relevant studies; and 3) the third step is to systematically extract information from these primary studies and code it into the database.

The first step uses a broad set of search terms (including e.g. terms such as *"formalize"*, *"formalization"*, *"registration simplification"*, *"labor inspection"*, etc.) and applies them to a title and abstract search in a series of websites and research databases in which relevant studies would be contained (such as e.g. the Social Science Research Network SSRN, IDEAS/RePEc, Google Scholar, the 3ie Repository of Published Impact Evaluation studies, etc.). In addition, backward / forward citation search is used to ensure all relevant papers are located. The studies identified through this first step are then given a full-text assessment in the second step, in which the following main inclusion criteria are considered:

— We only include empirical studies with a quantitative assessment of the effect or impact of a formalization program or policy using some version of a selection correction (counterfactual impact assessment, i.e. estimation of a causal treatment effect). In general, this can include studies based on experimental designs (Randomized Controlled Trials) or quasi-experimental methods.

— Distinguishable estimate of the effect or impact of the formalization program or policy, with an indication of the statistical significance of the estimate.

— Distinguishable formalization program or policy that can be categorized into one of the five intervention types.

— Studies that assess the impact on at least one of six relevant outcome variables (specified and discussed in detail below).

— In line with the objectives of the study (see introduction above), and because in general symmetry of effects cannot be assumed, only studies with a *switch-on* type of intervention that target improved formalization outcomes are included. That is, for instance, a study that looks at how tax increases may lead to a reduction in formal employment would not be in-scope.

— Study available in English.

— Search hits of studies focusing on the "hidden economy" or "shadow economy" are not in-scope.

All studies fulfilling these criteria are then used to extract information into a database. In our case, 32 primary studies were identified by the search process that also fulfilled the inclusion criteria. The main information to be extracted from the primary studies concerns, first, the intervention analyzed (as discussed in section 2.1), secondly, the outcome used to measure intervention effectiveness, and thirdly the estimate of the intervention impact.

To measure the effects of interventions, we consider six outcomes; a) firm registration, b) worker registration, c) wages, d) firm profitability, e) tax revenue, and f) investment. More than 70 percent of observations in our data stem from the first two categories. The first outcome looks at the number of registered firms or the probability of a set of firms to register. The second outcome looks at the same outcome at the level of the worker and considers the number of formal jobs, individual registration or the share of formally employed workers in an economy (or local labor market).

When examining the impact on wages we seek to identify whether the formalization interventions have led to increased wages for workers. This could be the case if, e.g. employing registered workers has become cheaper for firms (intervention type (iii) and (v)) which increases workers' bargaining power, or if being registered leads to increased firm performance. The latter point is also directly considered by the outcomes d) and f), which both look at firm outcomes. A reason why formalization interventions have become increasingly popular is that it is widely believed that formality can improve firm performance, e.g. by giving firms access to credit markets and by making it easier for firms to grow.[2] A major goal of governments aiming at increasing formality is to boost tax revenue (outcome (e)). Large informal economies are a main reason why many developing countries have a low tax base. By making it cheaper for firms or workers to become formal, it is hoped that the increase in formality can increase tax revenue despite reducing marginal tax rates.

In addition to intervention type and outcome, and perhaps most importantly, the coding process needs to extract a measure of the intervention effect. This measure can then be correlated in the empirical analysis with other variables from the primary studies, to investigate whether the estimated effectiveness shows systematic features by intervention characteristics. As a first measure, we use a trinomial indicator of sign and statistical significance (convention: at the 5 percent level) of the estimated intervention effect or

[2]Some evidence suggests that firms may stay inefficiently small to avoid being obliged to register and pay taxes (e.g. Bruhn and Loeprick, 2016).

impact: (i) negative and statistically significant, (ii) not statistically different from zero, and (iii) positive and statistically significant.

Ideally, in a second step, one would like to extract a measure of the size of the estimated impact, i.e. the coefficient of the estimated treatment effect (see e.g. Card et al., 2017). Doing this in a comparable way in the given context of formalization interventions is challenging because of the heterogeneity of the outcomes considered in the primary studies. Eventually, however, it was possible to code an effect size measure for more than two thirds quarters of our total sample (68 percent, see below) using estimates evaluating the *percent impact* of the intervention. This measure is defined as

$$\textit{Effect size} = \frac{\hat{\beta}}{\bar{\mu_c}} \tag{1}$$

where $\hat{\beta}$ is the estimated intervention effect coefficient and $\bar{\mu_c}$ refers to the mean outcome of the control group or the pre-intervention value of the treatment group. Our effect size measure deviates somewhat from the commonly used standardized mean difference (Borenstein et al., 2011), as in the quasi-experimental studies $\hat{\beta}$ is often not the mean difference between two groups, and the standard deviation required for the denominator is very rarely reported in the primary studies. Note that whereas this measure cannot be interpreted in units of specific outcomes, the standardization of $\hat{\beta}$ with the baseline value $\bar{\mu_c}$ has the advantage of making the effect size dimensionless and thus comparable across heterogeneous outcomes. This is a desirable property (see Duvendack et al., 2012), which is analogously achieved in more classical meta analysis settings by the conversion from a raw to a standardized mean difference.

Complementing this first effect size measure, we also code the *percentage point impact* of the interventions for our two main outcomes — worker and firm registration. This measure is directly interpretable and of relevance for decision makers: One policy goal, for instance, may be to increase firm registration by ten percentage points. Other related relevant effects for public policy, such as an increase in the tax base, can also be approximated well using this measure. A disadvantage of the percentage point measure is that it cannot say much about the distribution of the dependent variable: The success of an intervention that increased worker registration by one percentage point may be interpreted quite differently depending on whether the baseline value was close to zero (a large relative effect) or close to, say, 80 percent (a small relative effect). This would be captured by the effect size measure explained above.

The coding process, in addition, includes several more variables in three main groups:

a) <u>Study characteristics:</u>

- Country
- Authors; Title; Publication status (year; journal, if applicable)

b) <u>Intervention characteristics:</u>

- Target of the formalization intervention: firms, workers, or both.
- Scope of the intervention: program or policy.
- Year of the policy change or implementation of intervention (if applicable).

c) <u>Empirical analysis:</u>

- Time horizon of the study: start and end
- Unit of observation: (i) firm, (ii) worker, (iii) linked, (iv) other
- Data source and size of the estimation sample
- Time horizon of the outcome measurement: coded in months since reform date / start of the intervention, then categorized as short-run (up to 12 months), medium-run (13-24 months), and long-run (more than 24 months).
- Identification strategy and empirical method

When a study reports estimated impacts for (a) separate interventions, (b) separate outcomes, (c) separate groups of firms or workers, or (d) at separate time horizons, then these estimates are coded separately; that is, one study typically yields more than one observation in the data. Overall, it was thus possible to extract 159 impact estimates from the 32 primary studies. The trinomial measure of sign/significance is available for all estimates, and the effect size is available for 108 estimates. Note that for each of the relevant categories we code only the "best possible" estimate, i.e. either the one explicitly highlighted by the authors as the preferred estimate, or the one interpreted in the paper's findings. That is, we do not code treatment coefficients from slightly varying specifications, or from robustness checks.[3]

	Observations		Studies	
Country	Freq. (1)	Percent (2)	Freq. (3)	Percent (4)
Argentina	6	4	2	6
Bangladesh	3	2	2	6
Benin	18	11	1	3
Brazil	34	21	10	31
Colombia	40	25	6	19
Georgia	10	6	1	3
Indonesia	6	4	1	3
Malawi	12	8	1	3
Mexico	8	5	3	9
Peru	11	7	2	7
Russia	2	1	1	3
Sri Lanka	5	3	1	3
Turkey	4	3	1	3
	159		32	

Table 1: Distribution of countries in data base on formalization interventions

Table 1 presents an overview of the countries in the data. It can be seen that a set of countries with specific reforms, some of which were analyzed in more than one paper, is prominently represented in the data (e.g. Brazil with 34 estimates from 10 studies, and Colombia with 40 estimates from 6 studies). Overall, the majority of analyses of formalization policies and programs originates in countries in Latin America (99 impact estimates = 62%), but still more than one third of estimates (60) are from non-LAC countries. The overall number of countries in the sample (13) is not very large, which indicates that perhaps the use of these interventions — but more likely the quantitative analysis of these interventions — is not very widespread yet. Evidently, at the outset of this research we would have hoped for primary studies from many of the countries featured e.g. in Figure 1.

Figure 2 illustrates how the relevance of these interventions — or, to be precise: the

[3]This meta-analytical approach distinguishes our work e.g. from the paper by Floridi et al. (2019) who code every single estimate (568) available from a set of 18 primary studies analyzing the impact of formalization interventions specifically on firms.

rigorous assessment of their effectiveness — is, in fact, a very recent phenomenon. The figure shows the distribution of the years in which the respective reform or intervention was implemented, and indicates that more than two-thirds of the impact estimates for which this information was available (96 out of 136, i.e. 70%) are from interventions implemented in 2010 or later.

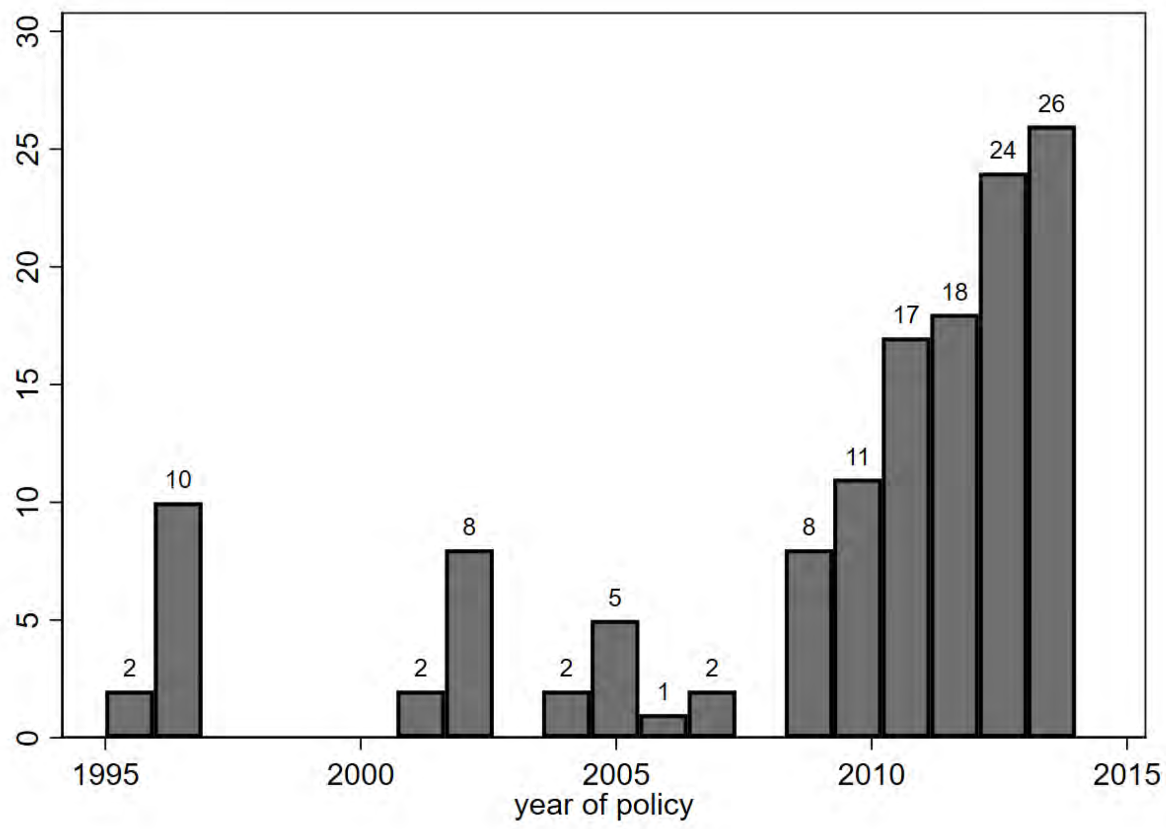


Figure 2: Year in which the formalization intervention was implemented

3 Descriptive analysis

This section presents a descriptive statistical analysis of the main patterns in the database. Table 2 gives an overview of the features of the impact evaluations of formalization interventions. The first two columns give a summary of the full sample, i.e. all estimates. Looking at the "intervention type", estimates of simplification / registration interventions represent the largest share in the data (85 estimates, 54%), followed by tax incentives (40%) and information approaches (30%). Financial incentives also cover almost one quarter in the data, and 19 estimates describe impacts of labor inspection interventions

	Full sample		Experimental		Quasi-exp.	
	Freq. (1)	Perc. (2)	Freq. (3)	Perc. (4)	Freq. (5)	Perc. (6)
Number of estimates	159		56		103	
Number of studies	32		8		24	
Intervention type						
Information intervention	48	30	44	79	4	4
Simplification / registration	85	54	44	79	41	40
Tax incentives	63	40	0	0	63	61
Financial incentives	37	23	28	50	9	9
Labor Inspection	19	12	4	7	15	15
Combination of interventions						
Single interventions	84	53	4	7	80	78
Two combined	57	36	40	71	17	17
Three combined	18	12	21	12	6	6
Formalization target						
Firm	91	57	56	100	35	34
Labor	62	39	0	0	62	60
Both	6	4	0	0	6	6
Formalization scope						
Program	59	37	56	100	3	3
Policy	100	63	0	0	100	97
Time horizon						
Short-term (0-12 months)	79	50	27	48	50	50
Medium-term (13-24 months)	63	40	24	43	39	38
Long-term (>24 months)	17	11	5	9	12	12
Outcome						
Registered firms	58	37	35	63	23	23
Formal jobs	56	35	2	4	52	52
Wages	14	9	2	4	12	12
Firm profitability	11	7	9	16	2	2
Tax revenue	18	11	6	11	12	12
Investment	2	1	2	4	0	0

Table 2: Features of impact evaluations of formalization interventions

Note: Registered firms denote the number of formally registered firms or registration probability. Formal jobs denote number of formal jobs, worker registration or probability to register. Evidently it varies from country to country what precisely "registration" entails. Columns (1) and (2) show an overview for all estimates, columns (3) and (4) show estimates from experimental studies and the last two columns those from quasi-experimental settings. The differentiation between experimental and quasi-experimental studies is almost identical to the differentiation between singular programs and policies (at scale) with one exception where a program is analyzed in a quasi-experimental setting.

(12%).

As the next panel indicates, the majority of estimates originates from single-component intervention and evaluations (84 estimates, or 53%). 36% of estimates are from interventions that combine two different approaches, and 12% of estimates are from interventions that combine three. That is, almost half of the sample covers multi-component interventions.

Looking at the "formalization target" there are slightly more interventions targeting firms than workers (57% and 39%, respectively). A residual 4% of impact estimates is from interventions that target both. In terms of the intervention scope, 100 of the impact estimates (63%) refer to „policy“-type interventions, while 59 (37%) refer to „program“-type interventions. This relation differs slightly when looking at the primary study level (not shown in the table): 10 of the 32 primary studies (31%) analyze singular „programs“, while 22 primary studies (69%) analyze „policies“ at scale. That is, the typical primary study analyzing a program produces on average fewer impact estimates than the typical primary study analyzing a policy.

The large majority of impact estimates are available for the short-term (79, i.e. 50%) and for the medium-term (63, i.e. 40%), while for the long-term time horizon only a limited number of estimates has been produced (17, i.e. 11%).

Finally, and perhaps unsurprisingly, in terms of the outcomes the majority of impact estimates investigate either impacts of formalization interventions on (a) the number of registered firms / firms‘ registration probability, or (b) the number of formal jobs / formal employment / worker registration, with more than a third in the sample each. Tax revenue is analyzed in just 11% of cases, while formalization impacts on wages, firm profitability, and investment remain the exception.

Columns (3) to (6) differentiate by whether the estimates are derived from experimental or quasi-experimental settings. Some substantial differences are apparent; almost all experiments involve some form of information and simplification / registration intervention, and half include a financial incentive. Quasi-experimental estimates on the other hand mostly stem from tax incentives and simplification / registration. Almost all experiments combine several types of interventions, and all are targeted at firms. The majority of quasi-experimental estimates consist of single interventions and most of them target labor registration, which is also reflected in "formal jobs" and "wages" being the predominant outcome measures. These notable differences by methodology evident in columns (3) to (6) therefore suggest that distinguishing the sub-samples of experimental and quasi-experimental estimates will also be important in the subsequent quantitative analysis.

3.1 Sign and statistical significance of estimated intervention effects

In Table 3 we begin investigating patterns of effectiveness by looking at sign and significance of the estimates. The table shows that just below half (73) of impact estimates in the full sample are positive and statistically significant (46%), giving a first measure of the probability with which formalization interventions can be expected to be successful. At the same time, only 9 impact estimates (6%) are negative and statistically significant; this means that close to 48% of impact estimates (77) are not statistically different from zero, which, as a first raw measure, would indicate that about one in two formalization interventions would be expected to show no effect.

Distinguishing between experimental and quasi-experimental designs in columns (3) to (6) of Table 3 shows that the share of positive significant estimates is larger for the quasi-experimental sample (51%, vs. 36% in the experimental sample). At the same time, the share of insignificant estimates is larger in the experimental sample (57%, vs. 44% in the quasi-experimental sample). Since the experimental studies are generally based on (much) smaller sample sizes, this latter pattern suggests that some part of the insignificant estimates may be due to research designs based on low statistical power.

Figure 3 stratifies the distribution of intervention effect estimates that are negative significant, insignificant, or positive significant by intervention type, the outcome and by the time horizon of the outcome measure. The distribution by intervention type in Panel (a) shows that only for "tax incentive" intervention types the share of positive significant

	Full sample		Experimental		Quasi-exp.	
	Freq. (1)	Percent (2)	Freq. (3)	Percent (4)	Freq. (5)	Percent (6)
Negative and statistically significant	9	6	4	7	5	5
Insignificant	77	48	32	57	45	44
Positive and statistically significant	73	46	20	36	53	51

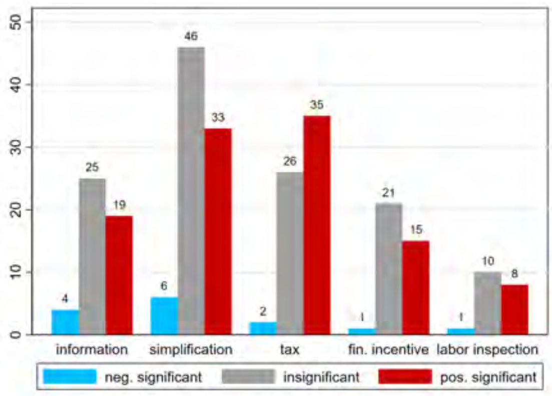
Table 3: Distribution of estimated intervention impacts by sign and significance

Note: Statistical significance is determined at a five per cent significance level.

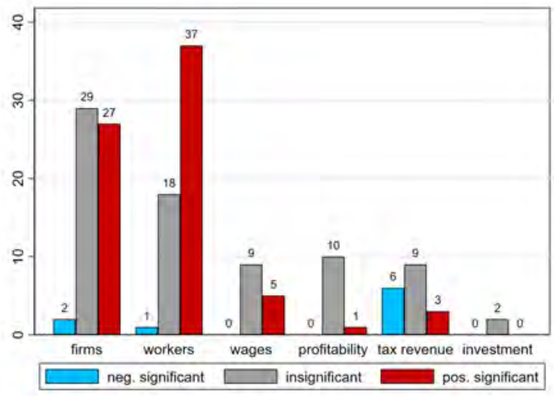
estimates is larger than for insignificant estimates, indicating that these interventions, on average, display more positive results. The second main finding of the graph is that there is no pronounced pattern by intervention type; the relative share of positive significant impacts is highest for tax incentives (56%), followed by labor inspection (42%), financial incentives (41%), information interventions (40%) and simplification/ registration (97%).

The patterns by outcome in Panel (b) on the other hand are rather pronounced. First, the shares of positive significant impacts differ strongly: impacts on the number of formal jobs and worker registration have by far the highest probability of showing positive impacts (66%). For the second main outcome, firm registration, this probability of a positive impact is still 47%. For wages, with much fewer estimates, 36% are significantly positive estimates. All estimates for investment are insignificant, and tax revenue is the only outcome containing a larger number of negative than positive estimates (tax revenue also accounts for two thirds of all negative estimates).

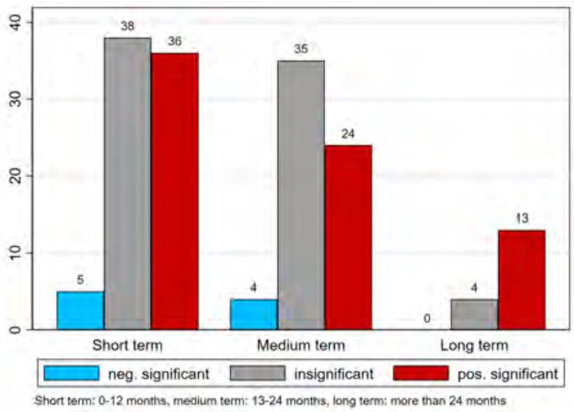
Panel (c) of Figure 3 distinguishes the sign/significance pattern by time horizon. Given the large shares of insignificant estimates at the short-term and medium-term time horizons, no conclusive dynamic pattern can be identified between these two. At the long-term time horizon, statistically significant estimates clearly dominate — given the small number of observations, however, this can at best give a very tentative indication that formalization interventions may have more positive effects in the longer run.



(a) Intervention type



(b) Outcome



(c) Time horizon of outcome measure

Figure 3: Sign and significance of estimates

Note: The number of estimates for the differentiation by intervention type add up to more than 159 estimates as almost half of the estimates combine more than one intervention type (see Table 2). Statistical significance determined at the five percent level.

3.2 Effect sizes

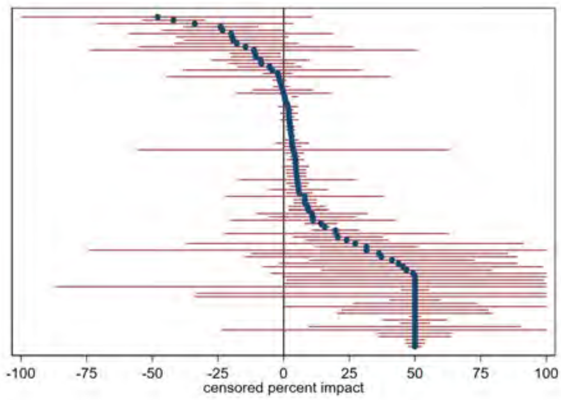
In addition to these main patterns by sign and significance, we can look at more detailed patterns of intervention effectiveness in Figures 4 and 5, which investigate results by the effect sizes — the estimated percent impacts — of the formalization interventions. Note that for presentation purposes we censor the very large percent impacts at 50 percent, and adjust the confidence intervals accordingly: several of the percent impact estimates are very large, attaining values of more than a 50 percent impact (25 out of 108). These originate in the field experiments in which firms are offered/exposed to treatments. Given the design of these studies, the mean value of the outcome in the control group will be relatively low compared to the treatment arms — because the treatment arms are typically offered different formalization incentives, whereas the control group is not — hence generating very large percent impacts (to be clear, this is not a shortcoming of these studies at all, but instead a shortcoming of the percent impact measure, which in turn is the only effect size measure we can use across all primary studies).

To illustrate this, consider the study from de Mel et al. (2013): in the time period of analysis two out of 105 control firms registered (0.019 registration probability), whereas in one of the treatment groups 30 firms did. Controlling for covariates, the treatment on the treated point estimate on the registration probability is 0.471, which implies an estimated percent impact of 2,477 percent (this is the largest in our sample). In order to make the presentation of our findings accessible in the figures, we set all percent impacts larger than 50 percent to 50 percent.

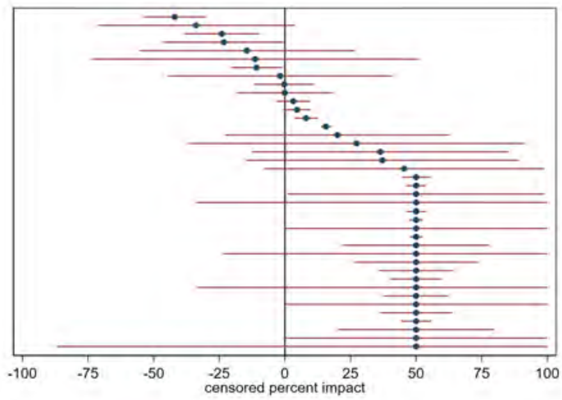
Figure 4 displays the distribution of estimated effect sizes and confidence intervals for all estimates (Panel (a)) and by intervention category (Panels (b) to (f)). The graph for all estimates shows a wide range of effect sizes ranging from sizable negative to very large positive, also with widely varying — and frequently very large — confidence bands. Looking specifically at intervention types, the forest plots look quite different: Information interventions (Panel (b)) show a broad range of effect sizes, with the largest number across interventions of very large point estimates at the censoring value, about half of which are quite precisely estimated, and the other half display broad confidence bands. For simplification / registration interventions (Panel (c)) we observe the largest number of available effect sizes, again covering the full range of effect sizes, and with strongly varying confidence bands.

The effect size estimates for tax incentives (Panel (d)) have a different pattern: There are fewer overall, and a much larger share are small effect sizes estimated with high precision. This likely reflects that often large administrative data sets analyzing policies at scale constitute the origin of these estimates. This pattern is similar for the labor inspection interventions (Panel (e)), only with fewer available effect sizes, while financial incentives interventions (Panel (f)) again display a more steady distribution across a larger range of effect sizes.

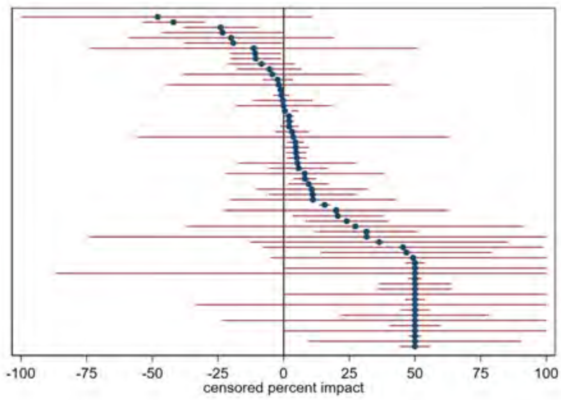
Figure 5 displays additional dimensions of heterogeneity by outcome type, inter-



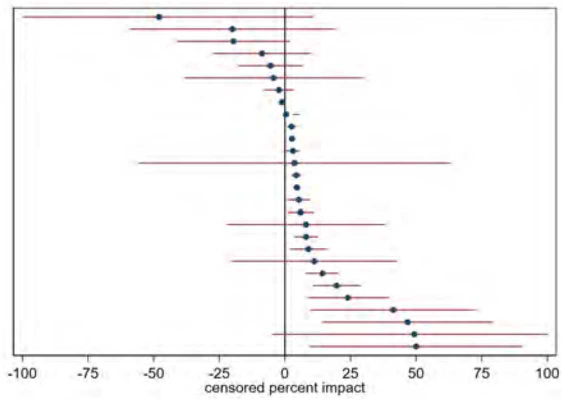
(a) All estimates



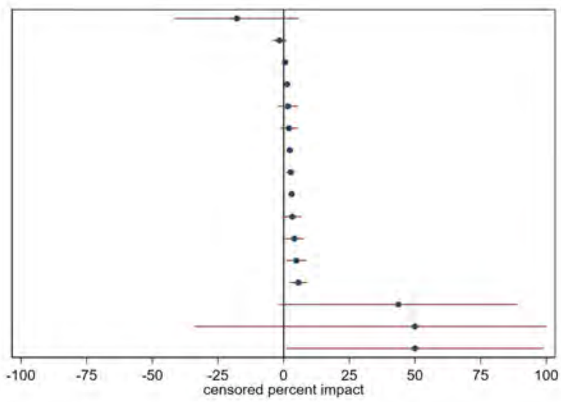
(b) Information intervention



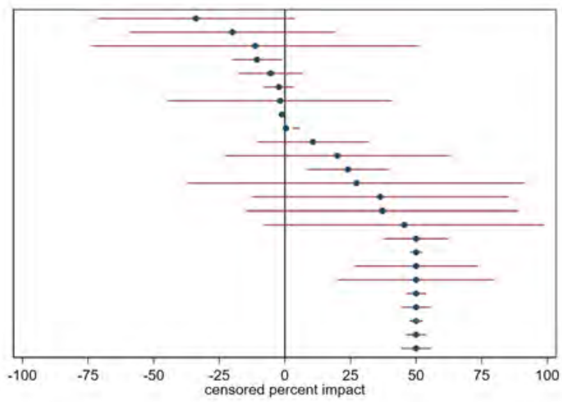
(c) Simplification / registration



(d) Tax incentives



(e) Labor inspections



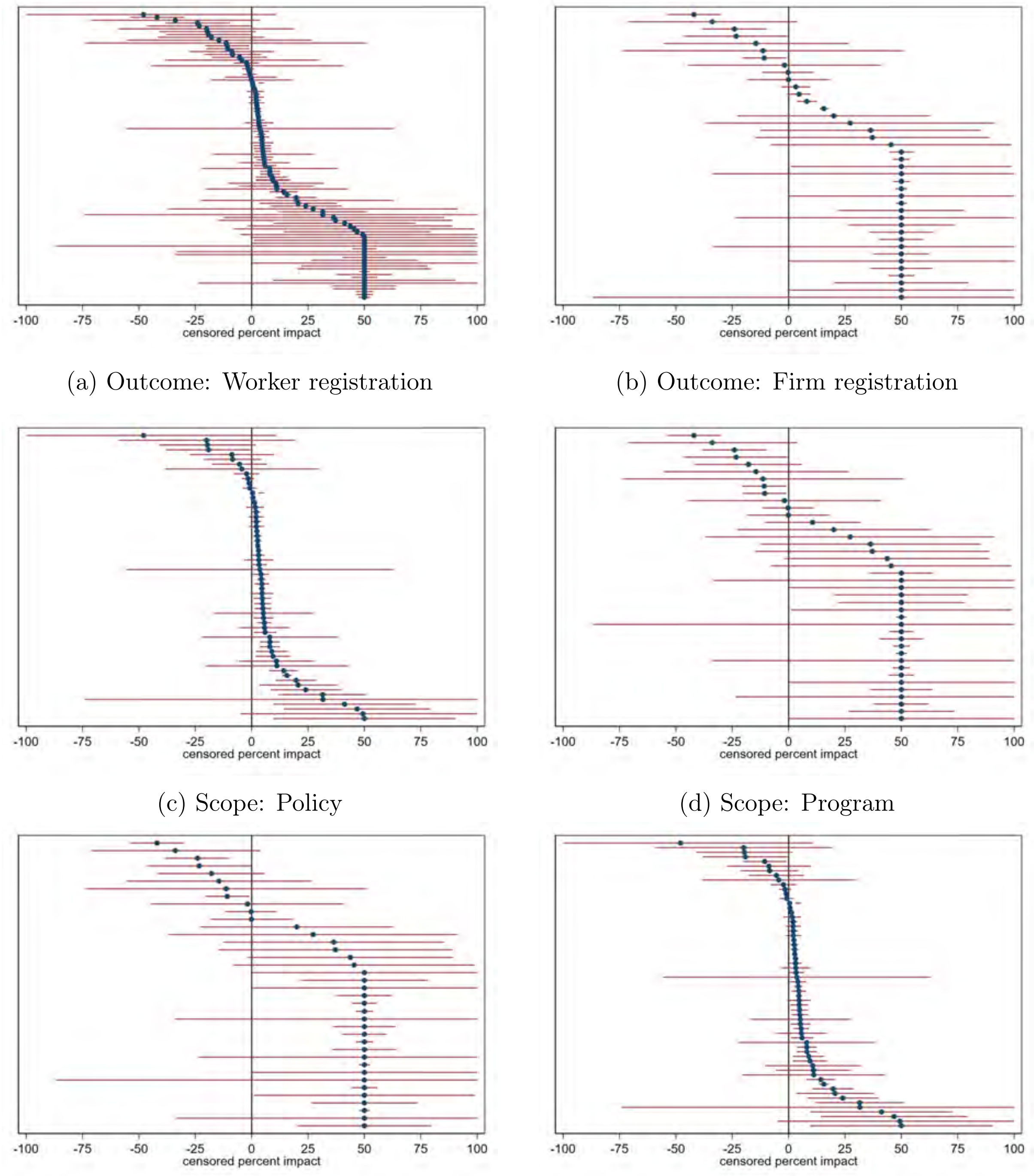
(f) Financial incentive

Figure 4: Forest plots of effect sizes I

Note: Forest plots of percent impacts and corresponding 95 percent confidence intervals for all estimates and by intervention type. For illustrative purposes percent impacts larger than 50 percent are censored at 50 percents and the confidence intervals are scaled accordingly. Similarly, lower and upper bounds of the confidence intervals are censored at -100 and +100 respectively to ensure similar scaling. Confidence intervals are of course, by definition, symmetric. In Tables 6 and 7 the uncensored estimates are used.

vention scope, and research design.[4] Specifically, the top two panels distinguish the

[4]Additional forest plots distinguishing by sign and significance of the estimates and by the time



(a) Outcome: Worker registration

(b) Outcome: Firm registration

(c) Scope: Policy

(d) Scope: Program

(e) Experimental estimates

(f) Quasi-experimental estimates

Figure 5: Forest plots of effect sizes II

Note: Forest plots of percent impacts and 95 percent confidence intervals for two main outcomes, by scope of the intervention and for experimental and quasi-experimental estimates. See Figure 4 for other notes.

outcomes worker registration (Panel (a)) and firm registration (Panel (b)). Whereas the former shows a continuous distribution across the full effect size range, with most weight

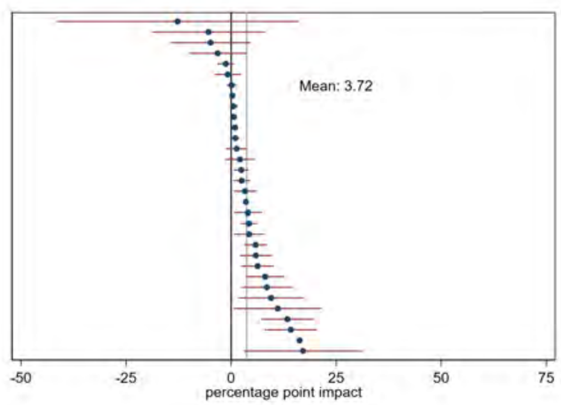
horizon can be found in the appendix.

of the distribution on the small, positive effect sizes (most of which precisely estimated), the latter has most estimated effect sizes at the censoring value, i.e. very large, with strongly varying degrees of precision.

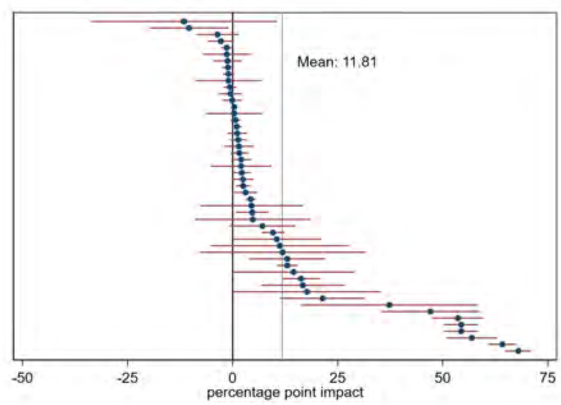
Distinguishing policies (Panel (c)) from programs (Panel (d)) also produces two very diverse patterns: specifically, the estimated effect sizes for policies show a smooth distribution from small negative to sizable positive, with the large majority of effect sizes clustering in the small-positive area. Confidence bands are mostly narrow. Most estimated effect sizes for programs, on the other hand, are very large positive (at the censoring value), and confidence bands are typically wide. Looking at the distinction between experimental and quasi-experimental designs in Panels (e) and (f) virtually mirrors the policy vs. program distinction, as almost all policy estimates (c) are based on quasi-experimental designs (f), while almost all program estimates (d) are based on experiments (e). The latter panel (e) highlights again the conjecture that several experiments with low statistical power contribute to the large number of insignificant intervention effects in the sample.

Overall, the descriptive evidence suggests that tax incentive and labor inspection interventions generally produce positive effects on formalization outcomes — the effects are statistically significant virtually throughout the respective effect size distributions, but typically small in size. The other three intervention types show less conclusive patterns: the effect size varies broadly and covers the full range from sizable negative to very large positive, and frequently the estimated effects are insignificant due to very wide confidence bands.

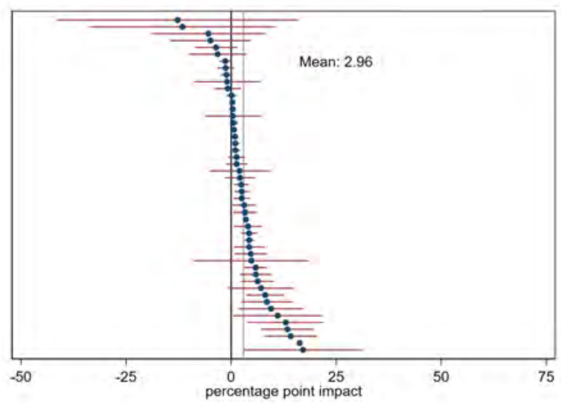
Figure 6 complements the descriptive analysis by plotting the distributions of the percentage point impacts. These distributions in Figure 6 are narrower than those for the percent impacts in the previous two figures, which indicates that the estimated percentage point treatment effects generally relate to small baseline formalization probabilities. As panel (a) shows, the percentage point impacts for worker registration, which primarily stem from interventions at scale, range from -12.7 to 17.1, with a mean of 3.7. For firm registration (panel (b)) the distribution of coefficients is somewhat wider, and the majority of coefficients is in the interval between -11.6 and 21.4 percentage points, with



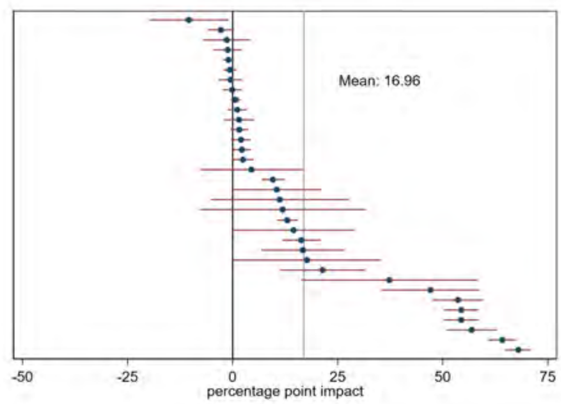
(a) Outcome: Worker registration



(b) Outcome: Firm registration



(c) Scope: Policy



(d) Scope: Program

Figure 6: Percentage point impacts

Note: Forest plots of percentage point impacts and 95 percent confidence intervals. In contrast to Figure 4, coefficients are not standardized by the mean of the control group / baseline mean. Coefficients are only coded for the outcomes worker and firm registration, for which the percentage points increases are straightforward to interpret. The mean value of the coefficients is indicated by the vertical grey line.

a mean of 11.8.[5]

When distinguishing by the scope of the intervention in panels (c) and (d) of Figure 6, one sees a similar pattern as for the percent impacts above, where the *policy* coefficients in (c) are usually quite small and precisely estimated and clustering between -3.6 (10th percentile) and 13 (90th percentile) percentage points, with a mean of 2.9. The *program* estimates (d) cover a notably wider range, with an average percentage point impact of 17.

[5]All estimates larger than 50 percentage points stem from the same study where a large share of firms obtained a business registration certificate in response to the intervention (Campos et al., 2015).

4 Quantitative analysis

4.1 Conceptual framework

The multivariate analysis is based on a conceptual approach used in related research designs, such as e.g. the meta analysis of the effects of active labor market policies (Card et al., 2017). Specifically, consider a formalization intervention that models an outcome y — worker registration, firm registration — observed for members of both a treatment group and a comparison group. Let b represent the estimated impact of the intervention on the outcomes of the treated units from a given evaluation design, and let β represent the probability limit of b (i.e., the estimate that would be obtained if the primary study sample size were infinite). Under standard conditions the estimate b will be approximately normally distributed with mean β and some precision P that depends on both the primary study sample size and the design features of the study. This leads to:

$$b = \beta + P^{-\frac{1}{2}}z \tag{2}$$

where z is a realization from a distribution that will be close to $N(0, 1)$ if the sample size is large enough. The term $P^{-\frac{1}{2}}z$ has the interpretation of the realized sampling error that is incorporated in b. In the next step, assume that the limiting intervention effect associated with a given study (β) can be decomposed as:

$$\beta = X\alpha + \epsilon \tag{3}$$

where α is a vector of coefficients and X captures the observed sources of heterogeneity in β, arising for example from differences in the type of intervention, characteristics of the target group or contextual factors. The term ϵ represents fundamental heterogeneity in the limiting intervention effect arising from the particular way it was implemented, specific features of the intervention or its target group, or the nature of the (labor) market environment. Equations 2 and 3 lead to a model for the observed intervention effect estimates of the form:

$$b = X\alpha + u \tag{4}$$

where the error $u = \epsilon + P^{-\frac{1}{2}}z$ includes both the sampling error in the estimate b and the unobserved determinants of the limiting intervention effect for a given primary study.

Card et al. (2017) propose the use of simple regression models based on Equation 4 to analyze the intervention effects on the relevant outcomes available in the sample of primary studies. In our case these models can be interpreted as providing descriptive summaries of the variation in average intervention effects due to differences in the observed features of a given formalization intervention and target group, and contextual factors (including methodological study features). Given the structure of the error component in Equation 4, Card et al. (2017) prefer OLS estimation, weighting each estimated intervention effect equally, to precision-weighed estimation, which would be efficient under the assumption that $\epsilon = 0$. In the case of formalization interventions the variation in ϵ could be particularly large, reflecting the wide range of factors that can potentially influence a formalization intervention to be more or less successful. This is thus the quantitative approach used for the sample of estimated effect sizes.

In addition, on the basis of having extracted for each estimate information related to whether it was "statistically significant negative", "statistically significant positive", or "not statistically significant from zero", the quantitative analysis estimates an (unweighted) OLS model for this 3-way classification of sign/significance of intervention effects.

4.2 Benchmark estimation results

Table 4 starts with the basic model and reports findings from an OLS regression in which the sign and significance of the estimated impact is correlated with the set of explanatory variables. For this analysis, the full database can be used, as sign and significance was codable for all intervention effect coefficients. While the effect size models are preferable in most aspects, one advantage of the sign and significance model is that it abstracts from the magnitude of the coefficient and more easily incorporates the very large impact estimates. The table step-by-step expands the specification for all estimates (columns

(1) to (5)), then adds results for the quasi-experimental (column (6)) and experimental samples (column (7)) separately.

A first finding of the regression is that there are few significant patterns overall, and in particular no evident correlation between a specific type of formalization intervention and intervention effectiveness. Long-term effects are significantly more likely to be positive than short-term effects, and the main formalization outcomes firm and worker registration are significantly more likely to be positively influenced than the other informality outcomes. In addition, interventions to reduce informality are significantly more likely to have positive effects when the economic context is better, i.e. with lower unemployment, higher GDP growth, and a lower poverty rate. Note that for the experimental sub-sample several covariates are not included because they vary too little, and the contextual variables are not included because the experiments are typically implemented in specific contexts and regions, with narrowly defined treatment and control groups.

Table 5 reports estimates for the meta-analytical effect size model using unweighted OLS as the benchmark model. Given the pattern identified in the descriptive analysis above, that effect sizes and corresponding confidence intervals fundamentally differ between quasi-experimental and experimental studies — recall Figure 5 panels (e) and (f) — we estimate the model separately for the two samples. The specifications for the quasi-experimental sample in columns (1) to (4) only maintain the result that formalization interventions have larger effects during better economic times (GDP growth), but otherwise show no further significant patterns between effect sizes and intervention type or other covariates. The experimental sample produces no conclusive findings, likely due to the underlying wide variation in effect sizes and the small number of studies.

4.3 Publication bias and weighted estimation

In any quantitative empirical exercise that compiles a meta database using a set of primary studies, a potential concern is that of publication bias (e.g. Rothstein et al., 2005) — i.e., the concern that the set of estimated intervention effects in the available literature may contain a systematic positive bias, either because analysts only write up and

	All estimates					Q-exp.	Exp.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Information intervention	-0.028	0.318*	0.156	0.120	0.327	-0.265	-0.158
	(0.170)	(0.164)	(0.214)	(0.244)	(0.404)	(0.696)	(0.262)
Simplification / registration	-0.161	0.084	-0.102	-0.165	-0.124	0.209	0.156
	(0.146)	(0.184)	(0.174)	(0.174)	(0.349)	(0.731)	(0.268)
Tax incentives	0.111	0.374*	0.123	-0.101	0.090	0.407	0.000
	(0.199)	(0.190)	(0.219)	(0.253)	(0.401)	(0.685)	(.)
Financial incentives	0.061	0.193*	-0.010	-0.045	0.122	-0.674***	0.078
	(0.127)	(0.113)	(0.128)	(0.141)	(0.256)	(0.240)	(0.238)
Labor inspection	-0.077	0.091	-0.189	-0.304	-0.051	0.362	-0.278
	(0.260)	(0.238)	(0.240)	(0.260)	(0.415)	(0.736)	(0.411)
Medium-term		-0.072	-0.149	-0.102	-0.143*	-0.026	-0.128
		(0.114)	(0.095)	(0.087)	(0.076)	(0.096)	(0.087)
Long-term		0.460**	0.355**	0.389***	0.360***	0.415***	0.204
		(0.180)	(0.154)	(0.122)	(0.112)	(0.141)	(0.183)
Single intervention		0.322	0.014	-0.258	-0.157	-0.399	
		(0.200)	(0.214)	(0.230)	(0.352)	(0.720)	
Registered firms			0.390*	0.482**	0.535**	-0.004	0.806**
			(0.222)	(0.208)	(0.226)	(0.260)	(0.253)
Formal jobs			0.574***	0.607***	0.567***	0.259	
			(0.177)	(0.190)	(0.177)	(0.182)	
Administrative data				0.237	0.242	0.336*	
				(0.164)	(0.167)	(0.163)	
Program				-0.149	-0.430**	0.000	
				(0.146)	(0.172)	(.)	
Square root of sample / 100				0.023**	0.019*	0.011	
				(0.011)	(0.011)	(0.013)	
Informality share					-0.003	-0.003	
					(0.006)	(0.007)	
Unemployment rate					-0.041*	-0.058**	
					(0.022)	(0.027)	
GDP growth					0.045*	0.063**	
					(0.024)	(0.028)	
Poverty index					0.001	-0.032	
					(0.007)	(0.021)	
Constant	0.448*	-0.133	0.055	0.190	0.342	0.759	-0.199
	(0.229)	(0.310)	(0.302)	(0.354)	(0.888)	(1.442)	(0.398)
R-squared	0.038	0.104	0.224	0.283	0.318	0.397	0.408
Studies	32	32	32	32	32	24	8
Estimates	159	159	159	159	159	103	56

Table 4: Models of sign/significance of estimated program effect

Note: Table entries are coefficients from a linear probability model, in which the dependent variable takes on the values of +1, 0, and –1 for an estimated program effect being positive statistical significant, insignificant, and negative statistical significant, respectively. Standard errors in (parentheses) are clustered at the study level.

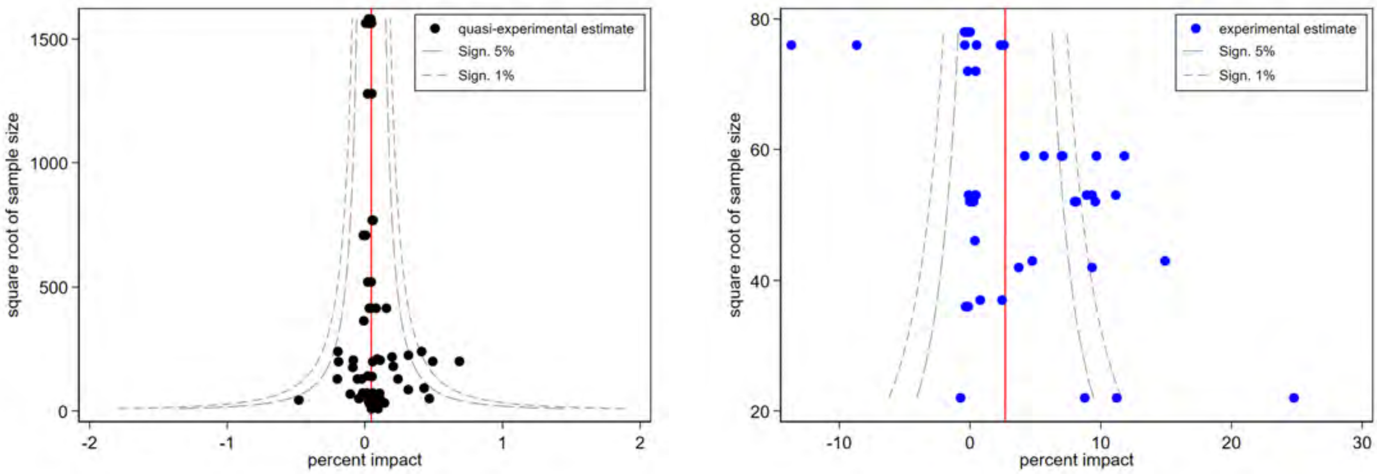
	Quasi-experimental				Exp.
	(1)	(2)	(3)	(4)	(5)
Information intervention	-0.011	-0.311	0.000	0.000	-4.574
	(0.038)	(0.254)	(.)	(.)	(2.887)
Simplification / registration	0.057	-0.148	0.101	0.053	-1.597
	(0.118)	(0.168)	(0.238)	(0.256)	(2.106)
Tax incentives	0.033	-0.152	0.113	0.053	0.000
	(0.112)	(0.166)	(0.227)	(0.234)	(.)
Financial incentives	-0.069	-0.193	-0.157	-0.111	1.969
	(0.109)	(0.154)	(0.167)	(0.160)	(1.657)
Labor inspection	-0.008	-0.207	0.051	-0.005	-4.300
	(0.114)	(0.166)	(0.210)	(0.229)	(2.887)
Medium-term		0.001	-0.009	0.003	-3.190*
		(0.025)	(0.028)	(0.025)	(1.275)
Long-term		0.038	0.034	0.067	0.000
		(0.030)	(0.029)	(0.046)	(.)
Single intervention		-0.289	0.006	-0.076	-6.055*
		(0.234)	(0.068)	(0.103)	(2.887)
Registered firms		-0.079	-0.080	-0.105	
		(0.111)	(0.117)	(0.153)	
Formal jobs		-0.040	-0.063	-0.084	
		(0.064)	(0.073)	(0.098)	
Administrative data			-0.046	-0.046	
			(0.053)	(0.052)	
Program			0.259	0.142	
			(0.261)	(0.257)	
Square root of sample / 100			-0.000	0.002	
			(0.003)	(0.005)	
Informality share				-0.002	
				(0.004)	
Unemployment rate				-0.007	
				(0.009)	
GDP growth				0.018***	
				(0.005)	
Poverty index				0.011	
				(0.009)	
Constant	0.033	0.537	0.025	0.172	10.484
	(0.114)	(0.379)	(0.273)	(0.224)	(5.773)
R-squared	0.034	0.107	0.114	0.163	0.147
Studies	21	21	21	21	6
Estimates	63	63	63	63	45

Table 5: Effect size model for estimated percent impact, unweighted

Note: Table entries are coefficients from a linear probability model, in which the dependent variable is the percent impact of the estimate. Standard errors (in parentheses) are clustered at the study level.

circulate studies that show a positive effect ("file drawer bias") or because they choose specifications that tend to yield positive and significant effects ("p-hacking").

To examine publication bias visually, Figure 7 shows funnel plots of the relationship between the effects sizes and the precision of the estimates (given by the square root of the primary study sample size). The effect sizes of the quasi-experimental estimates (Panel (a)) display an inverted funnel shape in which especially the very precisely estimated impacts center close to zero. This pattern indicates a "well-behaved" sample with no visual evidence for publication bias. For the effect sizes from experimental studies (Panel (b)) a different picture emerges: more precise estimates tend to be smaller or even negative, and large and positive estimates stem from studies with smaller sample sizes.



(a) Quasi-experimental estimates

(b) Experimental estimates

Figure 7: Funnel plots for percent impacts

Note: Figures plot the percent impacts against the square root of the sample size as an indicator of study precision separately for quasi-experimental and experimental estimates. A percent impact of 1 indicates a relative impact of 100 percent. Red vertical line indicates the mean value of percent estimates. Number of estimates in the figures are 63 and 45, respectively.

A more formal test for publication bias is to regress the estimated intervention effect from a given study and specification on the associated sampling error of the estimate and other potential control variables. Using the notation of Section 4.1, the regression model is

$$b = X\alpha + \theta P^{-\frac{1}{2}} + \nu \tag{5}$$

	(1)	(2)	(3)	(4)	(5)	(6)
a) Quasi-experimental						
Unweighted OLS	0.2820	0.1052	0.0731	-0.0417	-0.0273	-0.1285
	(0.4701)	(0.5513)	(0.5993)	(0.6571)	(0.7397)	(0.8631)
Precision weighted	0.3417	0.3786	0.3627	0.3781	0.3106	
	(0.3413)	(0.3501)	(0.3847)	(0.3891)	(0.4387)	(0.4652)
Estimates	62	62	62	62	62	62
b) Experimental						
Unweighted OLS	0.8848***	0.8608***	0.8299***	0.5711***	0.5205***	0.5822***
	(0.1156)	(0.1006)	(0.0703)	(0.0872)	(0.0347)	(0.0723)
Precision weighted	3.7895	2.2569*	2.0887*	1.8822	1.2600	1.2065
	(2.4047)	(1.0667)	(0.9235)	(1.2495)	(1.2009)	(1.2104)
Estimates	45	45	45	45	45	45

Table 6: Test for publication bias

Note: Entries correspond to estimated coefficients of the error term of the percent impact. Precision weighted estimates use the inverse of the variance, winsorized at the 10th and 90th percentiles of the respective samples to stabilize the estimates. Column (1) includes only a constant, the control variables in columns (2) to (6) correspond to the controls introduced in columns (1) to (5) of Table 4. Standard errors (in parentheses) are clustered at the study level.

where ν represents a residual. The estimate of θ is interpreted as a test for asymmetry in the funnel plot relationship between the estimated intervention effects and their precision. If the sample contains more imprecisely estimated large positive effects than large negative effects, θ will be positive.

Estimation results for this model are presented in Table 6. We present estimates from six specifications estimated by unweighted OLS and precision-weighted least squares (Stanley and Doucouliagos, 2012). The results for the quasi-experimental sample (Panel (a)) generally confirm the visual finding from the funnel plot, giving little indication of publication bias. Also for the experimental sample (Panel (b)) the visual results are confirmed, giving evidence of publication bias in all six specifications using unweighted OLS. For the precision-weighted estimates this is less evident, but note that the standard errors become very large.

Based on these findings, Table 7 presents estimation results for a weighted least squares regression using precision weights, and including the standard error of the estimated intervention effect, which is the preferred correction for publication bias as recommended e.g. in Stanley and Doucouliagos (2012). Again, we present results for the quasi-experimental and the experimental subsamples. The results are generally in line

	Quasi-experimental				Exp.
	(1)	(2)	(3)	(4)	(5)
Information intervention	0.096***	0.124	0.000	0.000	3.903***
	(0.006)	(0.121)	(.)	(.)	(0.603)
Simplification / registration	0.070***	0.070***	-0.054	0.092	3.790***
	(0.019)	(0.024)	(0.107)	(0.120)	(0.416)
Tax incentives	0.087***	0.094***	-0.034	0.106	0.000
	(0.019)	(0.025)	(0.107)	(0.124)	(.)
Financial incentives	-0.112***	-0.093	-0.090	-0.174	3.698***
	(0.018)	(0.106)	(0.111)	(0.117)	(0.084)
Labor inspection	0.069***	0.068**	-0.058	0.118	4.080***
	(0.019)	(0.025)	(0.106)	(0.125)	(0.582)
Scaled standard error	0.513	0.534	0.596	0.486	1.223**
	(0.356)	(0.474)	(0.548)	(0.582)	(0.448)
Medium-term		-0.002	-0.000	-0.003	-0.278**
		(0.009)	(0.010)	(0.009)	(0.094)
Long-term		0.014***	0.016***	0.013*	0.000
		(0.005)	(0.005)	(0.007)	(.)
Single intervention		0.015	-0.108***	-0.052**	3.490***
		(0.122)	(0.009)	(0.022)	(0.435)
Registered firms		0.004	0.004	-0.021	
		(0.010)	(0.010)	(0.024)	
Formal jobs		0.011**	0.013	0.007	
		(0.005)	(0.008)	(0.005)	
Administrative data			0.007	0.017	
			(0.011)	(0.017)	
Program			-0.129	-0.004	
			(0.128)	(0.141)	
Square root of sample / 100			0.000	-0.002*	
			(0.001)	(0.001)	
Informality share				0.000	
				(0.000)	
Unemployment rate				-0.005**	
				(0.002)	
GDP growth				-0.001	
				(0.001)	
Poverty index				-0.008***	
				(0.003)	
Constant	-0.052**	-0.081	0.159	0.065	-7.793***
	(0.020)	(0.146)	(0.097)	(0.104)	(0.958)
R-squared	0.525	0.561	0.562	0.602	0.398
Studies	20	20	20	20	6
Estimates	62	62	62	62	45

Table 7: Effect size model for estimated percent impact, precision weighted

Note: Table entries are coefficients from a linear probability model, in which the dependent variable is the percent impact of the estimate. Regressions are weighted by the inverse of the variance, the standard error of the estimates is added as an explanatory variable. Standard errors in (parentheses) are clustered at the study level.

with the findings from the unweighted models in Table 5: as the full specification in column (4) shows, effect sizes show little correlation with the set of covariates, while the pattern that formalization interventions have larger effects in better economic contexts is maintained. Column (5) for the experimental sample again does not identify any conclusive patterns.

The more parsimonious meta regression specifications in columns (1) to (3) for the quasi-experimental sample show that there may be some tentative underlying pattern also in intervention types: the three types of interventions simplification / registration, tax incentives, and labor inspection show significantly larger effects than information interventions and financial incentives, controlling for the precision with which the effects are estimated. These quantitative results from the multivariate analysis thus largely substantiate the results of the descriptive analysis.

5 Conclusion

Against the background of high levels of informality in labor markets in low and middle income countries many interventions have been implemented worldwide in an effort to increase the formalization of workers and firms. As it is of key importance for policy makers to know which intervention works under what circumstances, this paper has analyzed the patterns of the effectiveness of formalization interventions. The analysis has proceeded in a systematic and quantitative way: we first compile a database of primary studies of impact evaluations and quantitative assessments of formalization interventions, and then investigate patterns of the estimated impacts in the data, both descriptively and quantitatively, using meta regression techniques.

Since one main empirical challenge in this exercise is to correlate intervention effectiveness with intervention characteristics, we consider two measures of intervention effectiveness: the sign and significance of the estimated impact (available for all estimates), and the effect size of the estimated impact (for a subset of about three quarters of our database for which this information could be coded). The main explanatory variables considered are: the intervention type — classified into five categories: (i) information intervention, (ii) simplification/registration, (iii) tax incentives, (iv) financial incentives, (v) labor inspection; the outcome variable — comprising, in particular, firm registration and labor registration; the macroeconomic context — as measured by the informality share, unemployment rate, GDP growth, poverty rate; and a series of additional covariates covering intervention and methodological features, of which the intervention scope — policy vs. program — and research design — quasi-experimental vs. experimental designs — are of main relevance.

The database we are able to compile covers 159 impact estimates from 32 studies; these, in turn, originate in 13 countries. Given the pervasiveness of informality in low and middle income countries, one would have hoped for a more comprehensive coverage, but likely the sample reflects the limited number of thorough quantitative evaluations of formalization interventions that are available to date. The fact that the majority of these studies has been produced very recently suggests a surge in interest in this topic, and makes our analysis a timely exercise. Indeed, given the sustained levels of informality in labor markets of low and middle income countries, interventions targeted at reducing informality will doubtlessly remain high on the political agenda of those countries for years to come.

Several key patterns emerge from our analysis. First, a large share of the impact estimates is statistically significant and positive, indicating a generally sizable success probability of formalization interventions, while an equally large share is not statistically different from zero. The latter is partially determined by the experimental studies in the sample, some of which generate relatively imprecisely estimated intervention effects, likely due to low statistical power. Only few estimates are statistically significant and negative. Second, the type of intervention is not a strong predictor of intervention effectiveness, although the descriptive evidence suggests that tax incentives and labor inspection interventions, in particular, are more likely to generate significant impacts. Third, in terms of outcomes we observe that interventions targeting the formalization of workers are more effective than those targeting firms, but that both have a large share of positive and significant estimates. Secondary outcomes — wages, profitability, tax revenue and investments — are rarely positively affected, which could be due to the relatively short time horizon of most studies (few estimates stem from more than 24 months after the interventions), as some of these benefits may take time to materialize. Indeed, fourth, the raw success rate of formalization interventions is largest when looking at outcomes in the long run (>24 months).

The distinction between quasi-experimental and experimental settings is a key feature that emerges in this literature: the former generally analyze policies implemented at-scale, producing precisely estimated — and typically positive, and small in magnitude — effect sizes. The latter generally analyze programs narrowly defined for a specific target group, region, or to test a specific intervention prototype. The designs often yield very large positive effect sizes, frequently with very wide confidence bands. Concerning the intervention scope therefore, these results suggest that policies at scale are more effective than singular programs.

Although the effectiveness of particular interventions to reduce informality will evidently differ depending on the specific context, characteristics of the interventions and the target group, we are able to draw tentative conclusions from our study. Interventions appear to be more effective under favourable economic conditions (lower unemployment and poverty, and higher growth). Attempts to increase the tax base by increasing formalization during an economic downturn may thus not be an effective policy. We also find that targeting the formalization of workers is more effective than targeting firms. While our analysis substantiates this result quantitatively, it may not be too surprising as the stakes of formalizing workers are lower and arguably also more easily reversible. Furthermore, one should take into account that formal firms are often a prerequisite for their workers to be registered, so in countries with few registered firms, this should be the first focus of policy makers.

References

Alcázar, L. and M. Jaramillo (2016): "The Impact of Formality on Microenterprise Performance: A Case Study in Downtown Lima," .

Bernal, R., M. Meléndez, M. Eslava, and A. Pinzon (2017): "Switching from Payroll Taxes to Corporate Income Taxes: Firms' Employment and Wages after the 2012 Colombian Tax Reform," *Economía*, 18, 41–74.

Borenstein, M., L. V. Hedges, J. P. Higgins, and H. R. Rothstein (2011): *Introduction to meta-analysis*, John Wiley & Sons.

Bruhn, M. (2011): "License to sell: the effect of business registration reform on entrepreneurial activity in Mexico," *The Review of Economics and Statistics*, 93, 382–386.

Bruhn, M. and J. Loeprick (2016): "Small business tax policy and informality: evidence from Georgia," *International Tax and Public Finance*, 23, 834–853.

Bruhn, M. and D. McKenzie (2014): "Entry regulation and the formalization of microenterprises in developing countries," *The World Bank Research Observer*, 29, 186–201.

Campos, F., M. Goldstein, and D. McKenzie (2015): *Short-term impacts of formalization assistance and a bank information session on business registration and access to finance in Malawi*, The World Bank.

Card, D., J. Kluve, and A. Weber (2017): "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations," *Journal of the European Economic Association*, 1–38.

De Giorgi, G., M. Ploenzke, and A. Rahman (2018): "Small Firms' Formalisation: The Stick Treatment," *The Journal of Development Studies*, 54, 983–1001.

De Giorgi, G. and A. Rahman (2013): "SME's Registration: Evidence from an RCT in Bangladesh," *Economics Letters*, 120, 573–578.

de Mel, S., D. McKenzie, and C. Woodruff (2013): "The demand for, and consequences of, formalization among informal firms in Sri Lanka," *American Economic Journal: Applied Economics*, 5, 122–50.

Duvendack, M., J. G. Hombrados, R. Palmer-Jones, and H. Waddington (2012): "Assessing 'what works' in international development: meta-analysis for sophisticated dummies," *Journal of development effectiveness*, 4, 456–471.

Fajnzylber, P., W. F. Maloney, and G. V. Montes-Rojas (2011): "Does formality improve micro-firm performance? Evidence from the Brazilian SIMPLES program," *Journal of Development Economics*, 94, 262–276.

Floridi, A., B. Demena, and N. Wagner (2019): "Shedding light on the shadows of informality: A meta-analysis of formalization interventions targeted at informal firms," *ISS Working Paper Series/General Series*, 642, 1–37.

Khamis, M. (2012): "A note on informality in the labour market," *Journal of international development*, 24, 894–908.

——— (2014): "Formalization of jobs and firms in emerging market economies through registration reform," *IZA World of Labor*.

Kluve, J., S. Puerto, D. Robalino, J. M. Romero, F. Rother, J. Stöterau, F. Weidenkaff, and M. Witte (2019): "Do youth employment programs improve labor market outcomes? A quantitative review," *World Development*, 114, 237–253.

Maloney, W. F. (2004): "Informality revisited," *World development*, 32, 1159–1178.

Maurizio, R. and G. Vázquez (2017): "Formal employment generation and transition to formality in developing countries," *mimeo*.

Monteiro, J. C. and J. J. Assunção (2012): "Coming out of the shadows? Estimating the impact of bureaucracy simplification and tax cut on formality in Brazilian microenterprises," *Journal of Development Economics*, 99, 105–115.

Pignatti, C. (2017): "Enforcement and Compliance with Labour Legislation: The Effect of the Colombian "Action Plan"," .

Ronconi, L. and J. Colina (2011): "Simplification of Labor Registration in Argentina: Achievements and Pending Issues," Tech. rep., Inter-American Development Bank.

Rothstein, H. R., A. J. Sutton, and M. Borenstein (2005): "Publication bias in meta-analysis," *Publication bias in meta-analysis: Prevention, assessment and adjustments*, 1–7.

Stanley, T. D. and H. Doucouliagos (2012): *Meta-regression analysis in economics and business*, Routledge.

A Studies in the database

— Alcázar, L. and M. Jaramillo (2016), The Impact of Formality on Microenterprise Performance: A Case Study in Downtown Lima, International Finance Corporation, Washington D.C.: World Bank.

— Almeida, R. and P. Carneiro (2009), Enforcement of Labor Regulation and Firm Size, Journal of Comparative Economics 37, 28-46.

— Almeida, R. and P. Carneiro (2012), Enforcement of Labor Regulation and Informality, American Economic Journal: Applied Economics 4, 64-89.

— Benhassine, N., McKenzie, D., Pouliquen, V., and Santini, M. (2018). Does inducing informal firms to formalize make sense? Experimental evidence from Benin. Journal of Public Economics, 157, 1-14.

— Bernal, R., M. Eslava and M. Meléndez (2015), Taxing Where You Should: Formal Employment and Corporate Income vs Payroll Taxes in the Colombian 2012 Tax Reform, Preliminary.

— Betcherman, G., N. Meltem Daysal and C. Pagés (2010), Do Employment Subsidies Work? Evidence from Regionally Targeted Subsidies in Turkey, Labour Economics 17, 710-722.

— Bosch, M., D. Fernandes and J. M. Villa (2015) Nudging the Self-employed into Contributing to Social Security: Evidence from a Nationwide Quasi Experiment in Brazil. IDB Working Paper Series, WP 633, Washington D.C.: Inter-American Development Bank.

— Bruhn, M. (2011), License to Sell: The Effect of Business Registration Reform on Entrepreneurial Activity in Mexico, The Review of Economics and Statistics, February 2011, 93, 382-386.

— Bruhn, M. (2013), A Tale of Two Species: Revisiting the Effect of Registration Reform on Informal Business Owners in Mexico, Journal of Development Economics 103, 275-283.

— Bruhn, M., and McKenzie, D. (2013). Using administrative data to evaluate municipal reforms: an evaluation of the impact of Minas Fácil Expresso. Journal of Development Effectiveness, 5(3), 319-338.

— Bruhn, M. and J. Loeprick (2016), Small Business Tax policy and Informality: Evidence from Georgia, International Tax and Public Finance 23, 834-853.27

— Campos, F., M. Goldstein and D. McKenzie (2015), Short-Term Impacts of Formalization Assistance and a Bank Information Session on Business registration and Access to Finance in Malawi, Policy Research Working Paper 7183, Washington D.C.: World Bank.

— Cruces, G., Galiani, S., and Kidyba, S. (2010). Payroll taxes, wages and employment: Identification through policy changes. Labour Economics, 17(4), 743-749.

— de Andrade, G. H., M. Bruhn and D. McKenzie (2014), A Helping Hand or the Long Arm of the Law? Experimental Evidence on What Governments Can Do to Formalize Firms, The World Bank Economic Review 30, 24-54.

— Galiani, S., Meléndez, M., and Ahumada, C. N. (2017). On the effect of the costs of operating formally: New experimental evidence. Labour Economics, 45, 143-157.

— De Giorgi, G., and Rahman, A. (2013). SME's Registration: Evidence from an RCT in Bangladesh. Economics Letters, 120(3), 573-578.

— De Giorgi, G., M. Ploenzke and A. Rahman (2017), Small Firms' Formalisation: The Stick Treatment, The Journal of Development Studies 2017.

— de Mel, S., D. McKenzie and C. Woodruff (2013), The Demand for, and Consequence of, Formalization among Informal Firms in Sri Lanka, American Economic Journal: Applied Economics 5, 122-150.

— Fajnzylber, P., W. F. Maloney and G. V. Montes-Rojas (2011) Does Formality Improve Micro-firm Performance? Evidence from the Brazilian SIMPLES Program, Journal of Development Economics 94, 262-276.

— Fernández, C., and Villar, L. (2017). The impact of lowering the payroll tax on informality in Colombia. Economía, 18(1), 125-155.

— Kaplan, D. S., E. Piedra and E. Seira (2011), Entry Regulation and Business Start-ups: Evidence from Mexico, Journal of Public Economics 95, 1501-1515.

— Kugler, A., M. Kugler and L. O. Herrera Prada (2017), Do Payroll Tax Breaks Stimulate Formality? Evidence from Colombia's Reform, NBER Working Paper Series, WP23308.

— Madalozzo, R. and A. Bruscato Bortoluzzo (2011), The Impact of Tax Exemptions on Labor Registration: The Case of Brazilian Domestic Workers, Insper Working Paper 232.

— Monteiro, J. C., and Assunção, J. J. (2012). Coming out of the shadows? Estimating the impact of bureaucracy simplification and tax cut on formality in Brazilian microenterprises. Journal of Development Economics, 99(1), 105-115.

— Morales, L. F., and Medina, C. (2017). Assessing the effect of payroll taxes on formal employment: The case of the 2012 tax reform in Colombia. Economía, 18(1), 75-124.

— Mullainathan, S. and P. Schnabl (2010), Does Less Market Entry Regulation Generate More Entrepreneurs? Evidence from a Regulatory Reform in Peru, International Differences in Entrepreneurship, 159-177, NBER, University of Chicago Press.

— Pignatti, C. (2017), Enforcement and Compliance with Labour Legislation: The Effect of the Colombian "Action Plan", International Labour Organization, mimeo, ILO: Geneva.

— Piza, C. (2018). Out of the Shadows? Revisiting the impact of the Brazilian SIMPLES program on firms' formalization rates. Journal of Development Economics, 134, 125-132.

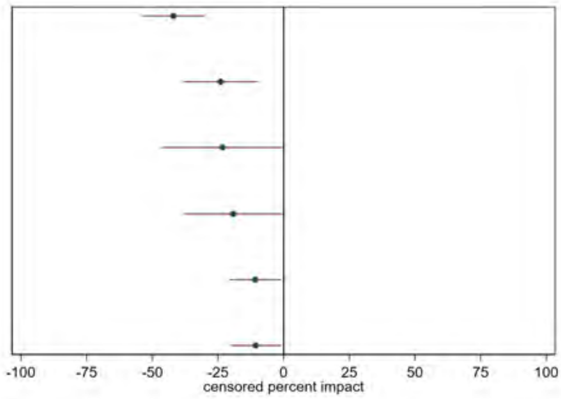
— Rocha, R., Ulyssea, G., and Rachter, L. (2018). Do lower taxes reduce informality? Evidence from Brazil. Journal of Development Economics, 134, 28-49.

— Ronconi, L. and J. Colina (2011), Simplification of Labor Registration in Argentina: Achievements and Pending Issues, IDB Working paper Series No. IDB-WP-277.

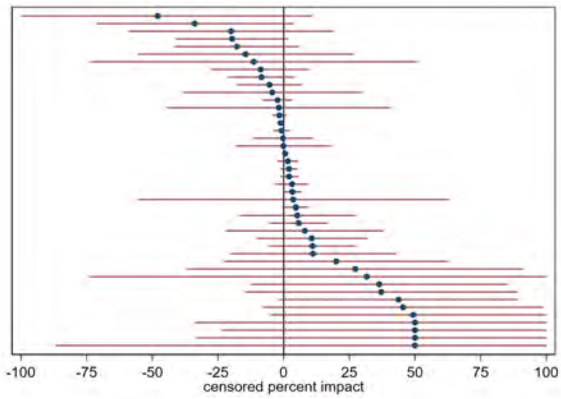
— Rothenberg, A. D., Gaduh, A., Burger, N. E., Chazali, C., Tjandraningsih, I., Radikun, R., Cole, S., and Weilant, S. (2016). Rethinking Indonesia's informal sector. World Development, 80, 96-113.

— Slonimczyk, F. (2012), The Effect of taxation on Informal Employment: Evidence from the Russian Flat Tax Reform, Informal Employment in Emerging and Transition Economies, Chapter 2, 55-99.

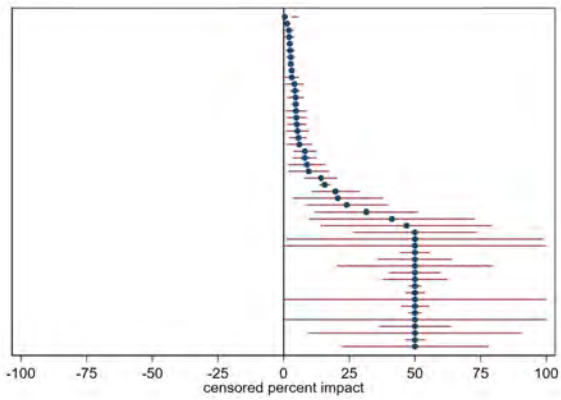
B Figures



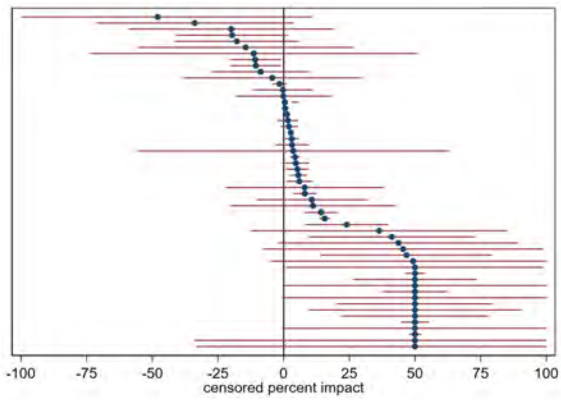
(a) Negative and statistically significant



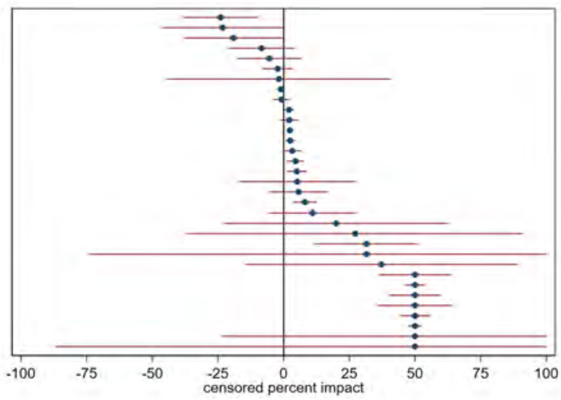
(b) Insignificant



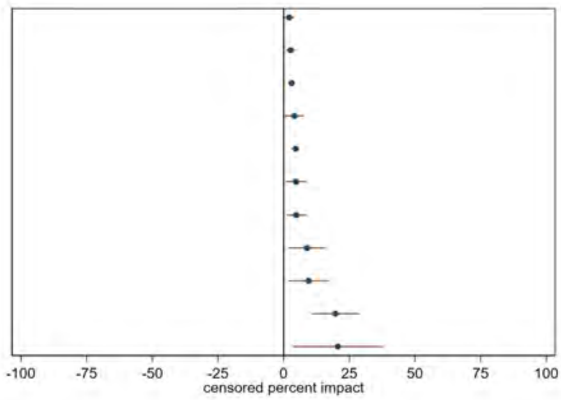
(c) Positive and statistically significant



(d) Time horizon: 0-12 months



(e) Time horizon: 13-24 months



(f) Time horizon: > 24 months

Figure 8: Forest plots of percent impacts

Note: Forest plots of percent impacts and 95 percent confidence intervals by statistical significance (at the 5 percent level) and by time horizon. See Figure 4 for other notes.