Regulation by Reputation? Intermediaries, Labor Abuses, and International Migration

A. Nilesh Fernando Niharika Singh*

Migrant workers and employers rely on intermediaries to facilitate labor market placements. If information frictions obscure their reputation, however, intermediaries may under-invest in placement quality. Using data on over 1.5 million Sri Lankan migrants to the Gulf region, we examine the effects of a government-run intermediary rating program that publicly revealed ratings two years after it was announced. Prior to the ratings' public release, under-performing intermediaries invest in the rating criteria and place migrants in less abusive and more remunerative employment. Our results suggest the government's threat of quality revelation induced intermediaries to prospectively screen employers and improve placement quality.

JEL Classification Numbers: O12, J63, O15

Keywords: Labor Intermediaries, Placement Quality, International Migration

---

* Fernando (nilesh.fernando@nd.edu) and Singh (nsingh6@nd.edu), University of Notre Dame. The Sri Lanka Bureau of Foreign Employment, the Ministry of Foreign Employment, Verité Research, Azam Bakeer Markar, Tamra Ranasinghe, and the Hon. Min. Thalatha Atukorale were essential to facilitating this research. We are grateful to the Ford Program in Human Development Studies and Solidarity and the Institute for Scholarship in the Liberal Arts at the University of Notre Dame for providing funding. Alison Lodermeier, Zigzmund Fernandopulle, Layton Hall, Mika Inoue, Thashiya Nauki, L.K Ruhunage, Paul Shaloka, Mehak Siddiquei, Joe Tatarka, and Renee Yaseen provided excellent research assistance. We thank numerous seminar participants and colleagues for detailed comments.

Prospective migrants often rely on intermediaries to facilitate access to labor market opportunities abroad. Across two of the world's largest migration corridors—the Mexico-US and Gulf-South Asia migration corridor—estimates place the share of intermediated migration as high as 80%.[1] The Gulf region,[2] the setting of our study, hosts nearly a quarter of the world's 170 million migrant workers and the vast majority of migrant workers from South Asia (International Labor Organization, 2019; Shah, 2013). Prior work in this context has shown that migrants routinely experience abuse and these concerns have led a number of governments to institute restrictions on Gulf migration at a potentially great welfare cost to prospective migrants and their families (Fernando and Lodermeier, 2022; Shivakoti et al., 2021).

Labor market intermediaries—be they migration intermediaries, temporary help agencies or internet platforms—arise because they have a comparative advantage in addressing market failures (Autor, 2008). Migration intermediaries are tasked with screening migrants and employers and coordinating mutually desirable matches. However, where information frictions obscure the observability of investments in screening, intermediaries may under-invest in placement quality. Further, given migrants' limited recourse and restricted labor mobility in the Gulf region (Naidu et al., 2016), intermediaries may instead seek to exploit migrants through unscrupulous recruitment practices.[3] These factors raise the stakes of a migrant's initial placement and the importance of policy that can rein in errant intermediaries.

We examine whether a policy that rates intermediaries and publicly reveals their quality can relax information frictions and improve migrant placement quality. Our study focuses on Sri Lankan migration intermediaries (hereafter, recruitment agencies) contracted by employers and *foreign* recruitment agencies acting on behalf of employers, whom we collectively refer to

[1] See Demaret (2006) for the Gulf region and Orrenius (2001) for the US southern border.

[2] 'Gulf region' refers to the countries in the Gulf Cooperation Council (GCC).

[3] E.g. UNODC (2015) and Naidu et al. (2023) discuss how intermediaries engage in 'bait and switch' practices and charge exorbitant fees inducing debt bondage.

as *foreign employers*.[4] The regulator first announced the criteria on which agencies would be rated and, two years later, revealed a 'star rating' to the public. The eligibility criteria for the program motivates a difference-in-difference model, which we estimate using administrative data on over 1.5 million Sri Lankan migrants over a decade (2005-2015), linked to recruitment agencies and foreign employers. To assess placement quality, we merge this dataset with data on contractual provisions and worker complaints—ranging from physical abuse to breach of contract—made to consulates across the Gulf region.

Our key finding is that the policy induced eligible agencies—particularly, previously underperforming agencies—to screen foreign employers and place migrants with less abusive employers *prior* to the revelation of the star ratings. To guide the interpretation of our empirical results, we develop a model of agency reputations and show that the ratings program improves the information environment by providing the market with a signal of an agency's investment in placement quality. Agencies invest in placement quality to influence an employer's belief about their type (i.e. their reputation) and, consequently, future demand for their services. Employers reward agencies for investing in screening because it improves match quality and limits employee turnover. In contrast, migrants have limited information about agency reputations, both because an agency's identity is obscured by a network of sub-intermediaries and because the ratings were largely inaccessible to migrants.[5]

The rating program created incentives for recruitment agencies to improve their placement quality both across job types and sectors and within them, by placing migrants with higher quality employers. The government assessed placement quality using an aggregate measure intended to capture both of these margins: the *Quantitative Score*.[6] The program also encouraged eligible agencies to invest in recommended management practices that were subsequently

[4]Our data do not allow us to distinguish between these groups.

[5]In a survey of migrants, we found that less than half the respondents surveyed could remember the name of their recruitment agency (see Section 4.1), similar to Bazzi et al. (2021).

[6]The Quantitative Score is based on total recruitment, the skill intensity of jobs matched, and an agency's performance in resolving complaints, see Appendix A.2.

audited by third-party inspectors. Compliance with these recommendations was assessed using another metric: the *Audit Score*. The government combined these two scores to produce a continuous score which maps to a star rating. The government announced the rating criteria in 2010 and then publicly revealed a star rating in 2012.

Eligibility for the program is based on criteria that preceded the announcement of the program: agencies were eligible if they recruited at least 100 migrants in 2009. Consequently, agencies could not select into the program and, as per the program rules, they could not opt out of it. These facts motivate a difference-in-difference design that compares eligible agencies who recruited just above the 100 migrant threshold (100-155) to a set of comparison agencies who recruited just below it (45-99).[7] Our identifying assumption is that eligibility is not systematically correlated with agency trends but, rather, reflects stochastic differences in recruitment; an assumption supported by F-tests of the lagged differences of our main outcomes.[8]

We estimate the effects of the program for the period after the rating criteria were announced but prior to the revelation of ratings (2010-2012)—which we refer to as the *Reputational Incentives* (RI) phase—and the period succeeding the public release of the star ratings, which we refer to as the *Quality Revelation* (QR) phase (after 2012). In addition, to estimate heterogeneous effects by (pre-program) agency quality, we use the ratings formula to compute a *Predicted Rating* for both eligible and comparison agencies.[9] We denote agencies that have

[7]To determine this window, we use an algorithm developed by Calonico et al. (2014), which yields an optimal sample containing 207 agencies who recruited 45-155 migrants in 2009.

[8]Though our empirical strategy resembles a 'difference-in-discontinuity' design (Grembi et al., 2016; Eggers et al., 2018), these designs use a pre-treatment period to isolate the effect of a specific treatment when it is bundled with confounding treatments which vary at the same discontinuity. In contrast, our design—as in a standard difference-in-difference—uses temporal variation to examine pre-trends and how treatments effects vary over time.

[9]To compute the Predicted Rating we use the fact that the rating calculated by the government is a function of administrative data collected *prior* to the announcement of the program. As such, we use this data to predict the continuous rating and use the fitted values to construct

a below median Predicted Rating as pre-program low-quality agencies. Importantly, outcomes during the RI-phase can *only* be influenced by agencies as the star ratings are yet to be made public.[10] In contrast, the QR-phase combines the effects of revealing a rating—to which both employers and migrants can respond—and the effects of reputational incentives.

We have three main findings. *First, eligible pre-program low-quality agencies respond to reputational incentives, secure higher demand, and are less likely to exit the market relative to low-quality comparison agencies.* To assess investments in the rating criteria, we compute a monthly analog of the Quantitative Score, which aggregates outcomes relating to the composition of recruitment and complaint management (see Appendix A.2). Low-quality eligible agencies invest in the rating criteria and their Quantitative Score is 11% higher relative to low-quality comparison agencies prior to quality revelation. Eligible low-quality agencies also receive substantially more job orders (65%), a measure of employer demand. Overall, while eligible agencies are approximately 44% less likely to exit the market, this result is driven by low-quality agencies who are 75% less likely to exit the market relative to similar comparison agencies;[11] in contrast, exit by high-quality agencies is unaffected.

Second, the program induces eligible agencies to improve placement quality and this too is largely driven by pre-program low-quality agencies. We find that migrants placed by eligible low-quality agencies report monthly salaries that are $20 (12%) higher than the comparison mean. Further, these agencies are 29% more likely to place migrants with a *Good Employer*: i.e. one with a below median pre-program complaint rate. Using an index that aggregates multiple measures of contract quality, we find that low-quality eligible agencies increase contract quality by 0.17 standard deviations.[12] Perhaps as a result of better placement quality, migrants placed by low-quality eligible agencies are 19% more likely to renew their contract with an employer

---

a counterfactual Predicted Rating for comparison agencies.

[10] In Section C.10 we find little evidence in support of indirect employer responses.

[11] Exit is defined as zero migrants recruited by an agency for at least 12 consecutive months.

[12] Contract quality is a weighted normalized index consisting of whether a contract includes health insurance, accommodation, return airfare, and the salary range (Anderson, 2008).

relative to the comparison mean. These investments narrow the gap in average quality between low and high-quality agencies by 71% during the RI-phase.[13]

Finally, improvements in placement quality occur prior to quality revelation, suggesting the program induced low-quality agencies to prospectively screen employers. Increases in average salaries and job orders from Good Employers begin during the RI-phase. As the star ratings are yet to be made public, these effects are *necessarily* a consequence of agency responses to the program announcement and largely preclude responses by foreign employers and migrants. During this period, we find little difference between migrants placed by eligible and comparison agencies. In addition, eligible agencies are no more likely to place migrants with *new employers* with whom they had no prior business dealings. Rather, we find that eligible agencies are more likely to place migrants with previously known Good Employers. This result suggests agencies respond to the program by scrutinizing employers more carefully; a survey we conducted with approximately one-fifth of recruitment agencies supports this view and shows that agencies routinely report *rejecting* job orders and cite safety concerns as a reason for doing so.

While we do not find that the program increased recruitment by eligible agencies, this may in part result from spillover effects to comparison agencies that are in the immediate vicinity of an eligible agency. For these comparison agencies, we find that a nearby eligible agency induces *positive* spillovers, increasing migrant recruitment and job orders, though the latter estimate is imprecise. Spillover effects may also occur if the program induces Bad Employers—those with an above average pre-program complaint rate—to seek out lowly rated or ineligible agencies who they view as unencumbered by reputational incentives, but we do not find evidence supporting this hypothesis either before or after the ratings were released.

Rather, we find that the program led employers to reward highly rated agencies with more job orders. Using the exact ratings formula to construct a regression discontinuity design, we compare agencies who are just above and below star-specific thresholds. Pooling across the thresholds, we find that an additional star leads to a 67% increase in job orders. In contrast, we

[13]The average gap in the (yearly) Quantitative Score between high and low-quality agencies reduces from 51.1% in 2009 to 15.0% in 2011.

do not find that migrants responded to the star ratings. Higher rated agencies do not recruit more migrants in general nor migrants who are especially likely to care about the ratings. This lack of migrant response is unsurprising as the ratings were largely made available on the internet, which less that 12% of the Sri Lankan population had access to at the end point of our dataset.

Our results suggest that the program induced eligible agencies to invest in the rating criteria prior to the QR-phase. Eligible low-quality agencies invest more in the *Quantitative Score*, relative to high-quality eligible agencies, but they also invest in management practices assessed by the *Audit Score*—especially in marketing and record-keeping—which account for the majority of their star rating. Perhaps as a consequence, low-quality eligible agencies are less likely to exit the market relative to low-quality comparison agencies. Overall, low-quality eligible agencies drive many of the observed effects of the program; while these arise in the RI-phase, they persist thereafter suggesting the program likely had a *certification* effect that improved their visibility, regardless of the number of stars they ultimately received.

Assessing whether the program improved market efficiency is complicated by the fact that the market was in steady decline over this period. However, we use a proxy for match surplus—*Total Salaries:* the product of migrants recruited and the salary they are contractually owed—and find that the program increased this proxy by approximately 70% for eligible agencies. Using this result, we estimate that the program generated nearly USD 2 million in increased remittance income per year, net of cost. We note, however, that these gains to Sri Lankan migrants may come at a cost to non-Sri Lankan migrants who are placed with Bad Employers. Additionally, we lack data on commissions and the entry and exit decisions of employers and are therefore unable to definitively assess the effects of the program on their welfare.

Recent empirical research on intermediaries has typically focused on how product market intermediaries address market failures and influence competition.[14] In contrast, labor market intermediaries have received limited attention and are often functionally different from product market intermediaries.[15] The paper closest to our own, Bazzi et al. (2021), uses an experiment

[14] E.g. Mitra et al. (2018); Bergquist and Dinerstein (2020); Atkin and Donaldson (2015).

[15] An exception is the work of Lee (2007) on public employment agencies in the US.

to estimate how *migrants* respond to information on intermediary quality. The policy we study, in contrast, influences the actions of employers and intermediaries who we view as an especially relevant margin because intermediaries often wield considerable market power inhibiting the scope for migrant choice and delegate recruitment to sub-intermediaries, thereby obscuring their own identity. Our paper is unique in its ability to capture the response of migration intermediaries to changes in regulatory policy.

Several studies have quantified the wage gains from international migration (Clemens, 2019; Naidu et al., 2023; Mobarak et al., 2023). Comparatively less attention has been paid to the *quality* of migrant placements, likely owing to the difficulty and political sensitivity of collecting reliable data. However, studies have examined how migrants respond to information about risks (Shrestha, 2017; Batista and McKenzie, 2021). We contribute to the evidence-base for migration policy by using a combination of consular reports of abuse and data on salaries and amenities to examine how domestic regulatory policy can improve placement quality.

Finally, our paper explores the effects of quality disclosure and reputational mechanisms in markets. In development economics, research has focused on the importance of relational contracting in buyer-seller relationships where contract enforcement is costly (Macchiavello and Morjaria, 2015, 2021) and how quality disclosure can influence market-level outcomes (Andrabi et al., 2017); see Dranove and Jin (2010) for a review of the literature. Our contribution is to show how quality revelation in the high-stakes world of international migration acts as a reputational mechanism that induces intermediaries to improve placement quality.

2 International Labor Migration from Sri Lanka

An estimated 1.5 million Sri Lankan workers, roughly 17% of the labor force, were employed abroad in the Gulf region in 2015.[16] Each year, roughly 200,000 migrants—evenly split between men and women—seek employment abroad primarily as construction workers, vehicle

[16]This estimate was calculated using publicly available data from the International Labor Organization and a report from the Foreign Ministry of Sri Lanka (Aryasinha, 2020).

operators, and domestic workers. Migrant remittances are an important source of foreign currency, accounting for 8.5% of Sri Lanka's GDP in 2015. However, there are persistent concerns about labor abuses experienced by migrants in the Gulf region, as documented by human rights groups (Human Rights Watch, 2007; Amnesty, 2020) and ethnographic studies[17] that are consistent with descriptive evidence from this setting (Fernando and Lodermeier, 2022). In late 2013, renewed concerns about migrant abuse led the Sri Lankan government to impose restrictions on women intending to migrate to the Gulf region (Peru, 2023). As a consequence, female domestic worker migration saw a large reduction after 2014.

2.1 Regulatory Framework and Market Structure

Concerns about abuse following the rapid expansion of migrant labor demand in the 1970's, led to the creation of a regulatory body to oversee migration in 1985: the Sri Lanka Bureau of Foreign Employment (SLBFE). The SLBFE was given legal authority to enter into agreements with foreign countries and regulate local recruitment agencies through the creation and enforcement of legal standards. All recruitment agencies are required to apply for a license with the SLBFE and migrant registration with the SLBFE has been compulsory since 1996. This is in large part enforced by monitoring at the airport; prospective migrants are denied boarding if they do not have a SLBFE registration stamp on their passport.

While consular assistance is available to all Sri Lankans, only registered migrants have access to support from the SLBFE. This support includes a dedicated labor section at consulates that registers and addresses complaints, and arranges repatriation and access to safe houses in response to severe cases of abuse (International Labor Organization, 2013). Migrant workers or their family members are able to register complaints at Sri Lankan consulates (see Appendix A.5 for an example of back and forth between the SLBFE and an agency regarding a complaint).

Turning to the structure of recruitment, local agencies recruit migrants on behalf of foreign employers and foreign agencies. Our data do not allow us to distinguish between these

[17] See Abu-Habib (1998); Frantz (2008); Jureidini and Moukarbel (2004).

two parties, so we collectively refer to them as 'foreign employers'.[18] Alternatively, migrants previously placed at an employer may renew their contract directly without the need for a recruitment agency (see Appendix A.7). The process for recruiting a migrant worker involves obtaining a number of approvals from the SLBFE and can take anywhere from two to five months to complete (see Appendix A.8 for details).

Agencies earn commissions in return for successfully recruiting migrant workers. We do not observe these commissions in our administrative data, but a survey of agencies (n=109) we conducted suggests foreign employers pay agencies fixed commissions per migrant recruited (see Section 4.1).[19] In our dataset spanning 2005-2015, there are 800 active recruitment agencies each year (median: 736), that recruit approximately 198 migrants per year on average (median: 66); the top 20 agencies account for approximately a quarter of migrants recruited via agencies during this period.[20]

The agency survey suggests that job orders are plentiful—50% of agencies report rejecting a job order in the prior year— but less than 20% of agencies report being able to fulfill a job order. Two-thirds of agencies state the main reason they were unable to do so is because they can't find workers. The characteristics of job orders vary considerably and are consistent with prior studies demonstrating a high degree of employer monopsony power in this setting (Naidu et al., 2016). We find that the least abusive employers pay higher salaries and offer more amenities both within and across sectors (see Appendix A.16).[21]

[18] Foreign agencies are themselves intermediaries and place migrants at their ultimate employer, e.g. they recruit domestic workers and place them at households.

[19] In the agency survey, 22 agencies specializing in the recruitment of domestic workers reported earning per worker commissions between $342-$382 across several countries.

[20] Agencies account for roughly 160,000 migrants per year, but approximately 60,000 migrants each year directly renew their contract with an employer.

[21] The lack of a compensating differential is suggestive of a dual market consisting employers that offer identical workers jobs with very different characteristics (Bulow and Summers, 1986).

2.2 Agency Ratings Program

Due to domestic political pressures and the increased salience of migrant abuse, the SLBFE devised a program to rate local recruitment agencies towards improving professionalism in the industry and the quality of job opportunities available to Sri Lankan migrants.[22] The program's stated goals include assisting foreign employers to identify suitable local agencies, motivating agencies to improve their own standards, and providing information to migrants on reputable agencies. In the booklet provided to agencies (see Appendix A.6 for a summarized version), the criteria listed for evaluation reveal two key concerns that the program intends to address:

- **Management Practices:** the program encourages the standardization of record-keeping practices towards minimizing fraud and maintaining orderly premises.
- **Placement Quality:** the program encourages agencies to develop new markets and high-skill opportunities for Sri Lankan workers. The goal here in particular is to reward placements in sectors and destinations that may have less abusive conditions. In addition the program encourages agencies to resolve complaints made by migrants and be more vigilant about issues migrants face at their workplace.

2.2.1 Eligibility for Program, Timing, & Program Phases

Agencies that recruited more than 100 migrants during 2009 and held a valid license for more than a year were eligible for the program.[23] The program consisted of two key phases: (1) Reputational Incentives (RI) Phase which lasted from 2010-2012 and (2) the Quality Revelation (QR) Phase which began after 2012 (see Appendix A.1 for a timeline). During the RI-phase,

[22] The case of Rizana Nafeek, a Sri Lankan domestic worker who was sentenced to death in 2009 and subsequently executed in Saudi Arabia was a pivotal moment (Silva, 2013).

[23] Technically, eligibility also required that agencies had not previously been engaged in underage recruitment, were not convicted of submitting false documents, and did not have more than 10 pending legal cases. These conditions are virtually guaranteed by having a valid license.

eligible agencies were informed about the program and given the rating criteria. Agencies were made aware of the program in January, 2010 through an online system managed by the SLBFE and provided with a booklet that described eligibility for the program, the program goals, and the criteria used in the assessment (see Appendix A.6). Thereafter, between July 2010-May 2011, third-party audits were conducted to assess compliance with recommended management practices.[24] During the QR-phase, the star ratings were announced in a public award ceremony in February, 2012. In June, 2012, the ratings were released on the SLBFE website.

2.2.2 Ratings Formula and Distribution of Star Ratings

Agencies were awarded a *star rating* between zero to five stars based on thresholds corresponding to a continuous score (hereafter, the *Combined Score*) that aggregates the rating criteria.[25] The Combined Score is itself a weighted combination of two constituent scores: the Quantitative Score and the Audit Score. The Quantitative Score uses administrative data to assess placement quality and its criteria include the volume of recruitment, skill and market diversification, and the successful resolution of complaints (9 criteria in total, see Appendix A.2). The Audit Score assesses compliance with recommended management practices including record-keeping and marketing (24 criteria in total, see Appendix A.2). We explain how the ratings were calculated in more detail in Appendix A.3, but on average the Audit Score accounted for 85% of the Combined Score and the Quantitative Score accounted for the balance 15%.

As the third-party audits occurred after the program announcement, agencies were able to influence their Audit Score. However, the government used *pre-program* data from 2009 to compute the Quantitative Score that was used for the star ratings. To reiterate, in spite of informing agencies that their star rating would be based on an assessment of their investments

[24] The audits were conducted by the Sri Lanka Standards Institute (SLSI) and the Sri Lanka Institute of Marketing (SLIM).

[25] Where c is the Combined Score, $c = F(Quantitative\ Score, Audit\ Score)$ and c is then mapped to a star rating as follows—0 stars: $0 \leq c \leq 19.9$, 1 star: $20 \leq c \leq 34.9$, 2 stars: $35 \leq c \leq 49.9$, 3 stars: $50 \leq c \leq 59.9$, 4 stars: $60 \leq c \leq 79.9$, and 5 stars: $c \geq 80$.

in the rating criteria during the RI-phase, part of the star rating was a function of data that, unknown to agencies, could *not* be influenced by their actions.[26] In summary, agencies were made aware of their eligibility for the program, the criteria for the rating, and *when* they would be released to the public. Agencies are *not* told the precise formula which maps the criteria to a star rating, but were told that their rating would be assessed during the RI-phase (2010-2012).

Based on the eligibility criteria, 307 out of 653 operational agencies were eligible for the program. Out of these eligible agencies, 34 agencies (11%) did not receive a star, 58 agencies (19%) received 1 star, 155 agencies (51%) received 2 stars, 52 agencies (16%) received 3 stars, and 8 agencies (3%) received 4 stars. No agency received 5 stars (see Appendix A.9). Consequently, while the star ratings certainly have a cardinal interpretation—a four star agency is certainly better than a two star agency—the more salient effect might have been a *certification effect* (i.e. whether or not they received a star) particularly for less well-known agencies.

3 Migration Intermediaries and Information Frictions

Descriptive evidence from our context suggests that recruitment agencies play a key role in influencing placement quality (Fernando and Lodermeier, 2022). Beyond securing passports and visas, agencies also play an important role in screening both employers and migrants in order to produce desirable matches. Whereas migrants care about avoiding abusive employers, employers instead wish to limit costly turnover. Yet, as screening effort may be difficult to observe and contract upon, whether agencies are willing to invest in screening may depend on whether they are rewarded for this effort. In particular, if information frictions obscure market learning about an agency's prior effort, they may have little incentive to invest in placement quality.[27] Under these conditions, a ratings program that provides a credible public signal of an

[26] We compute that if the government instead used the 2011 Quantitative Score, 5/77 eligible agencies in our main sample would receive a different number of stars.

[27] Consistent with the presence of such information frictions, we find little evidence of matching between more (less) reputable employers—defined as those with low (high) pre-program

agency's placement quality may induce agencies to invest in screening.

3.1 A Two-Period Bayesian Model of Agency Reputations

In this section, we outline a framework describing the interaction between agencies, employers, and migrants to provide insight into the contracting environment and guide the interpretation of our empirical results (see Appendix A.20 for details). In our framework, information frictions obscure an agency's type (adverse selection) and investment in screening (moral hazard). However, agencies can influence the market's beliefs about their type (i.e their reputation) if there is a mechanism that provides a signal of their effort (Cabral, 2005). The model consists of two periods that mimic the structure of the ratings program: the first period is the Reputational Incentives (RI) phase and the second period is the Quality Revelation (QR) phase. The central insight of our framework is that the credible threat of quality revelation in the second period can induce lower-quality agencies to invest in screening in the first period; they do so because employers favorably update their beliefs and reward them with increased demand.[28]

In this framework, there are agencies, employers, and migrants.[29] Agencies live for both periods, but employers and workers live for one period. We assume agencies vary in quality, *High* (H) and *Low* (L), and facilitate matches between employers and migrants in both periods. H-type agencies always produce 'good' placements, defined as one where a contract is not prematurely terminated. Meanwhile, L-type agencies produce good placements with probability $p(e)$ at cost $c(e)$, which are both functions of screening effort, e; the share of H-type agencies, θ, is common knowledge. Agencies maximize expected demand for their services, which is a product of employers' and migrants' demand, over the two periods.

---

complaint rates—and high (low) quality agencies prior to the ratings program (Appendix A.15).

[28] In support of the underlying structure of this model, we show that agencies with higher ratings receive more job orders (see section 5.5.1).

[29] Migrant labor markets resemble a two-sided platform, where an agency acts as the platform matching employers and migrants (Armstrong, 2006; Rysman, 2009).

Employers also consist of two types, *Good* or *Bad*, and live for one period. *Good* employers care about placement quality because turnover is costly. Their demand subsequently depends on the probability of good placements in each period. *Bad* employers, however, are not concerned with placement quality and, consequently, their demand is fixed in each period.[30]

Migrants' would like to be placed with a *Good* employer, so their demand depends on this probability. However, because migrants engage with agencies through a network of sub-intermediaries and did not have access to the ratings, we assume that their demand is fixed. With this setup, our primary object of interest is whether an *L*-type agency invests screening effort in the first period under different information environments.[31] The equilibrium concept is the perfect Bayesian equilibrium where the choice of first period effort by an *L*-type is a mutual best response given *Good* employers' belief about their effort.

In the absence of a mechanism to publicly reveal placement quality in the first period, the information environment in the second period looks exactly the same as the first period. Similar to *Bad* employers and migrants, *Good* employers will have the same demand in each period, since their beliefs about agency types and effort do not change. Therefore, in equilibrium, to maximize total expected demand, an *L*-type agency will choose to not invest in screening effort in the first period, because it (correctly) expects no return to this effort.

With a ratings program, the regulator credibly commits to revealing an agency's first period placement quality in the second period. Consequently, if placement quality in the first period is revealed to be bad, an agency is revealed as an *L*-type. *Good* employers would rationally avoid this agency in the second period knowing that it will not exert effort in the second period. The *L*-type agency would thus lose demand from *Good* employers in the second period. On the other hand, if placement quality is revealed as good, then *Good* employers update their belief that an agency is an *H*-type agency using Bayes' Rule.

An *L*-type agency faces a trade-off between the effort cost of screening and the benefit from

[30] See Section C.10 and 7.2 for further discussion and supporting empirical evidence.

[31] As the game ends after two periods, an *L*-type agency never exerts effort in the second period; *Good* employers anticipate this and avoid contracting with *L*-types.

increased second period demand. It therefore maximizes expected total demand to equate the marginal benefit from increased second period demand from *Good* employers with its marginal cost of first period effort. Under standard assumptions about $p(e)$ and $c(e)$, we show that there exists an interior solution whereby *L*-type agencies increase screening: e. The key result here is that *L*-type agencies increase their investment in placement quality in response to the program; trivially, this also implies a differential investment since the actions of *H*-types are unchanged.

This framework shows how easing information frictions can encourage an *L*-type agency to *prospectively* invest in placement quality to increase future demand from *Good* employers. In the subsequent sections, we show empirically that low-quality agencies indeed improve placement quality in the RI-phase. While our framework focuses on placement quality, we note that in practice, the ratings program also rewarded management practices via the Audit Score. Investment in the management practices—e.g. better record-keeping and marketing—may also raise demand for an agency's services and improve their outcomes (see Section 5.4).

4 Data Sources and Empirical Strategy

4.1 Administrative Data and Primary Data Collection

The primary data we use are administrative data collected by the SLBFE. We complement this data with a detailed survey of recruitment agencies, the 'agency survey' (n=109), and a detailed survey of migrants, the 'migrant survey' (n=240), both of which were conducted in 2009.[32] The specific administrative data we observe consists of data on: (1) *Job Orders* initiated by an employer and then approved by the SLBFE (data is missing for 2005 and 2014). These data include

[32]For the agency survey, we used administrative data to sample an even split between agencies based in Colombo or Kurenegala that recruited above and below 100 migrants. For the migrant survey, we surveyed 116 prospective migrants and 124 returned migrants; we sampled workers in areas where they are likely to congregate (e.g. temples, schools) across two major migrant-sending cities: Colombo and Kurunegala.

information on the number of workers requested by an employer (i.e. vacancies) and the salary, and amenities being offered, (2) *Migrant Trips* through 1510 recruitment agencies—1,543,441 trips, 1,066,825 unique individuals—to foreign destinations. This data includes agency identifiers and a limited set of migrant characteristics, and (3) *Complaints* made by migrants or their family members to a Sri Lankan consulate. There are 127,238 complaints from 112,000 unique individuals (see Appendix A.4 for the composition of complaints).

4.2 Identifying Variation

As agencies were informed in early 2010 about the program, their eligibility is retrospective because it depends on the number of migrants recruited in 2009. Consequently, agencies can neither select into the program, nor select out of it since they cannot refuse a rating per the program rules. As such, a natural counterfactual are the set of agencies that have been licensed for at least a year but recruited fewer than 100 migrants in 2009. At first glance, the eligibility threshold may suggest the appropriateness of a regression discontinuity design. However, a key concern is whether agencies that were narrowly eligible or ineligible were on different trends.

Our ideal counterfactual would be the set of agencies who, on average, follow similar trends in recruitment but recruited fewer migrants in 2009 due to the stochastic nature of recruitment. A difference-in-difference design allows us to transparently assess pre-trends for our outcomes of interest and control for theoretically important sources of agency-specific, time-varying heterogeneity. In addition, given multiple phases of the program, the event study graphs also allow us to assess *when* program effects occur—a key issue for interpreting our empirical results as we discuss in the conceptual framework in Section 3.

To pick a non-arbitrary set of agencies located about the 100 migrant threshold, we use an algorithm developed to conduct balance tests on pre-treatment variables by Calonico et al. (2014). The algorithm conducts balance tests for pre-specified pre-treatment characteristics using a progressively larger window across the threshold until the test statistic suggests an imbalance.[33] We specified variables that capture the fundamentals of agency operations: migrant

[33] The test constructs a sequence of nested windows around the 100-migrant eligibility cutoff;

recruitment, job orders, and complaint intensity between 2005-2008. We find that the optimal sample contains agencies who recruited between 45-155 migrants in 2009.[34] This sample restriction yields 207 agencies, accounting for 32% of all active agencies. If the government instead used the 2008 migrant total to determine eligibility, 35% of our eligible agencies would have been ineligible and, conversely, 26% of our comparison agencies would have been eligible. This fact, and a visualization of pre-program recruitment trends (see Appendix A.10), provide suggestive evidence that program eligibility is unrelated to underlying trends.

4.3 Difference-in-Differences Specification

For agency i in month-year t, our main specification is:

$$Y_{it} = \delta_t + \alpha_i + \beta_1 Elig * \mathbb{1}[2010 \leq t < 2012]_{it} + \beta_2 Elig * \mathbb{1}[t \geq 2012]_{it} + \tau X_{it} + \varepsilon_{it} \quad (1)$$

In the specification above, δ_t are month-year fixed effects, α_i is an agency fixed effect, $Elig$ is an indicator variable for an eligible agency, $\mathbb{1}[2010 \leq t < 2012]$ is an indicator variable for the period between January 2010 and May 2012, $\mathbb{1}[t \geq 2012]$ is an indicator variable for the period after (and including) June, 2012, and X_{it} are two sets of time varying controls: the interaction of month-year fixed effects with average migrant recruitment between 2005-08, and the interaction of month-year fixed effects with the average Quantitative Score between 2005-08.[35] We cluster standard errors at the agency level. Under the assumption of parallel trends,

---

we specify a minimum window of 20 migrants to ensure a large enough sample for inference. The algorithm successively increases the window size by 2.5 migrants, conducts balance tests for our given covariates. The algorithm then chooses the largest window around the cutoff such that the p-value of the balance test is larger than a pre-specified level.

[34] Our results are robust to using alternative samples (see Appendix C.11).

[35] Since agencies of varying size and quality may be subject to different shocks, we allow for trends that vary with pre-program migrant recruitment (size) and the Quantitative Score (quality). We show in Appendix C.7 that our results are robust to excluding these trend controls.

β_1 is the causal effect of the program during the Reputational Incentives phase, while β_2 is the casual effect of the program in the period after the Quality Revelation phase, which includes the effects of the RI-phase that precedes it.

Next, we estimate how the program effect varies by (pre-program) agency quality: the key source of heterogeneity in our conceptual framework. Ideally, we would compare an eligible agency to an ineligible one of comparable quality that would have received a similar rating had they been eligible. To approximate this ideal, we compute a *Predicted Rating* for all agencies, using the fact that an eligible agency's 2012 star rating is partly determined by the pre-treatment (i.e. 2009) value of their Quantitative Score, which we are also able to construct for the comparison group (see Appendix A.13).[36] To do so, we interact equation (1) with an indicator variable for whether an agency has an above (*High*) or below (*Low*) median *Predicted Rating*:[37]

$$\begin{aligned} Y_{it} = \delta_t + \alpha_i + \gamma_1 High * Post_{it} + \beta_1 High * Elig * Post_{it} + \\ \beta_2 Low * Elig * Post_{it} + \tau X_{it} + \varepsilon_{it} \end{aligned} \tag{2}$$

The terms are as defined above, however, $Post_{it}$ refers to month-year observations after the introduction of the program. The $High * Post_{it}$ term captures the trend for pre-program high-quality agencies in the comparison group.[38] In the model above, β_1 is the causal effect of the program for pre-program high-quality agencies, while β_2 is the analogous effect for pre-program low-quality agencies. We note that since eligible agencies are compared to comparison agencies of the same quality, our estimates cannot be driven by mean reversion (see Section 6.2). Finally, we estimate equation (2) for each program phase, to understand how agencies of varying quality

[36] We regress an eligible agency's Combined Score—the continuous value that determines star ratings—on their 2009 Quantitative Score (t-stat = 15.05). We then use the estimated coefficients to construct a predicted Combined Score for the comparison group.

[37] The *Predicted Rating* varies considerably within each group of agencies, allowing us to estimate heterogeneous effects by pre-program agency quality.

[38] When we estimate heterogeneous effects by agency quality, we omit the interaction between month-year fixed effects and the average Quantitative Score between 2005-08.

respond to the program in each phase.

Our identifying assumption is that eligible agencies and comparison agencies would have followed a parallel trend in the absence of the program. To assess the plausibility of this assumption, we first plot the raw quarterly average of migrants recruited by eligible agencies and comparison agencies (Panel A in Appendix B.5) and see that the trends track each other closely prior to the introduction of the program (vertical blue line). Analogous charts plotting job orders (Panel B) and the complaint rate (Panel C) yield a similar conclusion.[39] To investigate pre-trends more formally, we test whether we can reject that the lagged effects for a series of outcomes are jointly zero; both for the sample of all agencies and the sub-samples of low and high quality agencies.[40] In Appendix C.1, across several key outcomes we are unable to reject the joint null for the lagged effects with p-values ranging from 0.20 to 0.82; we provide plots of these lagged effects in Appendix C.2.

4.4 Summary Statistics

Table 1 provides descriptive statistics for our sample at the agency-month-year level using data from 2005 to 2015. The average firm in our sample recruits 8.23 migrants per month. A majority of migrants are female, engage in domestic work, and nearly a quarter sought work in Saudi Arabia: the most common destination for Sri Lankan migrants. The median monthly salary for a migrant worker during this period is $178 (2015 US dollars). In terms of demand from foreign employers, an agency receives 0.48 job orders per month, each of which has 68 vacancies on average.[41] On average 8% of migrants placed by an agency made a formal complaint to a

[39] The complaint rate is the number of complaints divided by the number of migrants.

[40] Owing to power concerns and variation in outcomes, failure to reject the joint null need not preclude pre-trends nor would rejecting the null necessarily imply a violation of the assumption (Roth, 2022). A strength of our design that reduces concerns about power is that the F-tests we report are based on at least 15 lagged periods.

[41] Job order data is missing for 2005 and 2014, resulting in a smaller sample size.

consulate. In Appendix A.4 we categorize these complaints; the three most frequent complaints are: (1) non-receipt of wages (17%), (2) breach of employment contract (13%), and (3) harassment (11%). Agencies resolve 88% of complaints filed, taking 4.6 months on average to reach a resolution.

5 Results

In this section, we report difference-in-difference estimates of the program. We first assess whether eligible agencies invested in the rating criteria and then estimate how the program influenced agency operations (Table 2) and placement quality (Table 3).

5.1 Investment in Rating Criteria

To investigate whether agencies responded to the rating criteria, we constructed a *monthly* version of the Quantitative Score that captures investment in a subset of the rating criteria that we observe for eligible and comparison agencies (see Section 2.2.2). To recap, the Quantitative Score is a function of a number of criteria including the number of migrants recruited, the skill-level of jobs, and the complaint solve rate (see Appendix A.13 for details). The Audit Score matters even more for the star ratings, but we only observe this score for eligible agencies.

In Panel A of Table 2, we find that eligible agencies, on average, have a higher monthly Quantitative Score during the RI-phase (i.e. the coefficient on $Elig * [2010-12]$), but this estimate is somewhat imprecise (t-stat = 1.64). However, in Panel B, we find that pre-program low-quality agencies (i.e. below median *Predicted Rating*), are significantly more likely to invest in the rating criteria: their Quantitative Score increases by 0.66 points, which represents an increase of 11% relative to the pre-program comparison mean. In Panel C, we estimate program effects by agency quality and program phase and find that the increase in the Quantitative Score for low-quality agencies (largely) occurs during the RI-phase.

In examining the components of the Quantitative Score, we do not find that a specific criterion drives it's observed increase during the RI-phase (See Panel A of Appendix B.6).

While low-quality agencies are more likely to send skilled migrants abroad (Panel B), this effect arises *after* the RI-phase—suggesting it may be driven by employers and/or migrants responding to the ratings—and, therefore, cannot explain the higher monthly Quantitative Score *during* the RI-phase. One interpretation of these results is that in contrast to examining the effects on a *specific* criterion, the monthly Quantitative Score yields a more powerful test because it combines the criteria and weights them in a way that reflects the government's priorities.

The Audit Score assesses agencies across 24 criteria relating to recommended management practices including investments in record-keeping, marketing, and the implementation of awareness programs. Although we're unable to assess the causal effect of the program on the Audit Score, in Appendix A.18 we show how the Audit Score and it's components vary by agency quality. Overall, the Audit Score for low-quality eligible agencies is 4% higher than the score for high-quality eligible agencies, though this difference is not statistically significant; the score is driven by investments in marketing (44%) and record-keeping (40%).

In sum, investments in the rating criteria substantially narrowed the gap in quality among eligible agencies. The average gap in the (yearly) Quantitative Score between low and high-quality agencies was 51.1% in 2009 and 15.0% in 2011, in the second year of the RI-phase: a reduction of 70.7%. We note again that the government chose to use the 2009 Quantitative Score when computing the star ratings, effectively handicapping low-quality agencies. In spite of this, and likely due to their investments in the Audit Score, low-quality agencies ultimately earn 1.8 stars on average, relative to 1.9 stars for high-quality agencies.

5.2 Market Outcomes

In column 2 of Table 2, we find that the program did not significantly affect migrant recruitment by eligible agencies on average (Panel A), nor by pre-program agency quality (Panels B & C); we note, however, that the estimates are generally positive. In column 3, we see that job orders for eligible agencies increase by 31% during the RI-phase. After quality revelation, job orders for eligible agencies are 42% higher relative to the pre-program comparison mean. As with investments in the rating criteria examined in the prior section, we find that the effect on

job orders is driven by low-quality agencies. Job orders for these agencies increase by 65% relative to the comparison mean. Panel A in Figure 1 (all agencies) and Figure 2 (low quality agencies) use event study plots to demonstrate these effects graphically and, additionally, show the absence of pre-trends which is considered in further detail in Panel D of Appendix C.2.

As eligible low-quality agencies have more job orders in both phases of the program, it is perhaps surprising that they do not recruit more migrants. While the imprecision of our estimates for migrant recruitment may suggest we are under-powered, they may also reflect a trade-off between the quantity and quality of placements.[42] An increase in job orders may arise either because employers have reached out to agencies or vice versa. In Panel C, we find that the increase in job orders for low-quality agencies begins *prior* to the QR-phase. The timing of this result suggests that the increase in job orders is initiated by agencies because prior to the release of the star ratings, employers and migrants do not have any new information to which to respond. While increases in job orders do not influence an agency's rating, these results suggest that eligible agencies intent on improving placement quality may instead reach out to more reputable employers. We discuss evidence supporting this interpretation in Section 5.5.

In column 4 of Table 2, we estimate whether the program influenced the survival of eligible agencies. To do so, we construct an 'Agency Exit' variable that records whether an agency had 12 consecutive months of zero migrant recruitment.[43] We see that even during the RI-phase, eligible agencies are 44% less likely to exit the market. The estimated effect is driven by low-quality agencies, who are 75% less likely to exit the market relative to similar comparison agencies; low-quality agencies are both less likely to exit prior to quality revelation and after it (column 4, Panel C). In contrast, we do not find that the program affects the likelihood of high-quality agencies exiting the market.

The agency exit effects in the QR-phase are consistent with the theoretical predictions

[42] Assuming the errors are asymptotically normal, power = 0.80, and alpha = 0.05, a two-sided test would yield a minimum detectable effect (MDE) for migrant recruitment of 4.438 $\sim$ 60% (82% for low-quality agencies) of the comparison mean during the RI-phase.

[43] These results are robust to using a 6 month and 24 month definition. See Appendix C.3.

of the model in Section 3; i.e. *L*-type agencies invest in the rating criteria so as to secure more job orders once quality is revealed. However, low-quality eligible agencies are also less likely to exit during the RI-phase. This result may instead be a consequence of the adoption of recommended management practices, which we discuss in Section 5.4.

5.3 Placement Quality

In Table 3, we assess whether the program influenced a series of proxies for placement quality. We adjust all outcomes for agency exit and, therefore, our sample sizes are smaller than in Table 2.[44] In column 1, we find that the program increased the average salary of migrants placed by eligible agencies in both phases of the program, but these estimates are imprecisely estimated (t-stat $\sim$ 1.3). Turning to effect heterogeneity by agency quality (Panel B), we find that low-quality agencies place migrants in jobs with significantly higher salaries ($20 or 12%).[45]

An alternative measure of placement quality is to assess whether migrants placed by agencies decide to renew their contracts with the same employer, as turnover is an important proxy of match quality. In column 2, we compute the program effect on 'Contract Renewal' which reports the share of migrants placed by an agency who renew their contract with an employer. While we do not find that the program had an average effect on contract renewal (Panel A), we find that migrants placed by low-quality eligible agencies are significantly more likely to return to the same employer after the program; the effect size is substantial and reflects an increase of 15% relative to the pre-program comparison mean (Panel B). Panel C shows that the contract renewal effect largely arises *after* quality revelation, likely due to the fact that contracts of migrants placed during the RI-phase would only be renewed two to three years later.

[44] We code an outcome as missing if an agency had 12 consecutive months of zero recruitment; similarly adjusting the estimated effects in the prior section reduces their precision but not their substantive interpretation (see Appendix C.6).

[45] This increase in average salaries is driven by migrants placed by low-quality agencies during the RI-phase (Panel C), consistent with event study graph in Panel B of Figure 2.

In column 3, we turn our attention to the quality of foreign employers at which migrants were placed. We use pre-program information on these employers to identify job orders originating from Good Employers (i.e. those with below median pre-program complaint rates) and find that the program resulted in a 31% increase in such job orders (Good Orders) for eligible agencies prior to quality revelation. While both high and low-quality eligible agencies have significantly more Good Orders (Panel B), the increase in Good Orders for low-quality eligible agencies is more precisely estimated and occurs during the RI-phase, suggesting the program induced these agencies to initiate contact with employers to improve their placement quality.

Finally, we use an index of contract quality that aggregates the provision of amenities such as health insurance, accommodation, and return airfare to assess the quality of placements. We find that the program increases contract quality by 0.17 standard deviations for low-quality eligible agencies (Panel B) and this increase occurs during the RI-phase (Panel C) as demonstrated graphically in Panel C of Figure 2.[46] Yet another way to assess placement quality is to estimate the effect of the program on the complaint rate. In Appendix C.5, we do not find that the program influenced the complaint rate in either phase of the program or by agency quality. Our design is likely under-powered to detect changes in the incidence of complaints owing to considerable variation even *within* an employer-agency match owing to idiosyncratic factors.[47]

Our results suggest the program induced eligible agencies—and, especially, pre-program low-quality agencies—to improve their placements *prior* to quality revelation. We again note that this finding is consistent with predictions of our model for *L*-type agencies in Section 3. In

[46] Appendix B.3 computes program effects for each of the constituent variables of the contract quality index. We again find that low-quality agencies drive the results and place migrants at employers who are more likely to provide them with higher salaries and better amenities.

[47] Using a migrant-year level dataset, agency x employer fixed effects explain five times as much variation in whether a migrant makes a complaint as country, sector, and time fixed effects, but they jointly yield an adjusted r-squared of just 0.25 (Fernando and Lodermeier, 2022). Furthermore, assuming asymptotically normal errors the errors, power = 0.80, and alpha = 0.05, a two-sided test would yield a MDE of 0.02 $\sim$ 25% of the comparison mean.

the model, the mechanism by which *L*-type agencies improve placement quality is by screening employers, which we discuss in the following section.

5.4 Discussion: Ratings Investment and Agency Exit

Our results show that low-quality eligible agencies are especially responsive to the program and drive many of the observed outcomes. In Appendix A.19, we compute agency exit before (2005-09) and after (2010-15) the program by agency quality. High-quality agencies are less likely to exit the market in the pre-period, but this trend is *reversed* after the introduction of the program: 18% of low-quality agencies exit the market, relative to 23% of high-quality agencies. This reversal, however, obscures important heterogeneity by eligibility for the program: 6% of low-quality *eligible* agencies exit, relative to 23% of low-quality *comparison* agencies.

While we lack cash-flow data on agency operations that would permit a more detailed accounting of the improved survival of low-quality eligible agencies, we note that low-quality agencies invest more in the Quantitative Score than their high-quality counterparts and have a higher Audit Score on average (though this difference is not significant). The realized Audit Scores are largely composed of practices relating to marketing (40%) and record-keeping (44%) which may well account for the improved survival of low-quality eligible agencies (see Appendix A.18). de Oliveira (2023) finds a similar result on the effect of management practices on firm survival. Although we are unable to estimate whether the adoption of these practices causes less exit, we note that a higher Audit Score is correlated with being less likely to exit the market for low-quality eligible agencies (t-stat = -2.69).[48]

5.5 Discussion: Screening and Job Order Composition

Why doesn't the increase in job orders for eligible agencies result in additional migrant recruitment? Job orders contain multiple vacancies, many of which remain unfilled over time. Consequently, additional job orders need not imply an increase in recruitment. An agency in-

[48] We do not find that these effects vary *among* low-quality agencies (see Appendix B.9).

tent on prioritizing placement quality may restrict the quantity of migrants placed by selectively filling vacancies for job orders from more reputable employers.

Figure 3 shows how matches between agencies and Good Employers—those with a below median pre-program complaint rate—evolve with the introduction of the program. Each panel plots a series of distributions showing the density of Good Employers matching with agencies of varying quality, as measured by their Combined Score/2012 agency rating (x-axis). Although we plot all density functions with respect to the 2012 Combined Score, the agency ratings only became public after 2012, i.e. the density functions with solid lines in Panel C. Consequently, the density functions in Panel A and B visualize matches between Good Employers and agencies in an information environment where the ratings are as yet unknown.

In Section 3, we argue that information frictions obscure agency quality prior to quality revelation. The dashed gray line in Panel A of Figure 3 supports this assumption by showing that Good Employers do not systematically match with higher quality agencies *prior* to the introduction of the program. However, Panel B shows that the darker line (triangles) denoting the second year of the RI-phase shifts its mass towards agencies with a higher rating.[49] Two points are of note here: first, the only actors in the market that can influence this shift are agencies as the star ratings are only revealed in the QR-phase. Second, this shift is driven by agencies who achieved a Combined Score exceeding 50 points, i.e. three stars or better.

This graphical evidence is consistent with an increase in Good Orders prior to quality revelation for low-quality agencies (Panel C of Table 3). In Appendix B.2, we examine the characteristics of employers linked to job orders. We find that low-quality eligible agencies increase the number of *unique* employers with whom they place migrants during the RI-phase (column 1). However, these employers were previously known to the agencies and not *new employers* with whom the agencies had no prior interaction (column 2). Consequently, low-quality agencies improve placement quality by favoring less-abusive (Good), previously known employers, resulting in an increase in the number and share of Good Orders during the RI-phase,

[49] In Appendix B.11, we use tests of the equality of distributions to show that the density functions shift significantly after the introduction of the program.

but there is no change in either the share or number of Bad Orders. In contrast, we do not find evidence suggesting that agencies screened migrants.[50]

Collectively, these results suggest eligible agencies screen job orders to improve placement quality prior to the revelation of the star ratings. In support of the plausibility of this mechanism, the agency survey shows that approximately half (55) of the 109 agencies surveyed report rejecting a job order in the last year; 52% of these agencies cite the reason for rejection as being safety concerns related to the reputation of the employer (see Panel B of Appendix B.1).

5.5.1 Employer Response to the Star Ratings

Foreign employers are the most likely to have access to the star ratings. The manner in which ratings were made available—through a searchable database on the SLBFE website (see Appendix A.11)—makes them far more accessible to employers relative to migrants who largely lack access to the internet. As we discuss in Section 2.1, employers are required to fill out several forms and enter details on the SLBFE's online information management system, so they have good reason to access this website. In addition, interviews with consular officials suggest that foreign missions sometimes provided prospective employers with lists of recruitment agencies and their star ratings, though this was not a uniform practice.

In Panel C of Figure 3, we find that in QR-phase there is a noticeable intensification in the likelihood of a match between Good Employers and highly rated agencies; Appendix A.17 shows that we do *not* find similar evidence showing Bad Employers responded to the ratings. To more rigorously test if employers respond to the star ratings, we use a regression discontinuity design that estimates the effect of an agency receiving an additional star. In particular, we use discontinuities in the star-specific cut-offs that map the Combined Score to a star rating. We compare agencies just above and below the cut-offs and pool the thresholds to arrive at an estimate of the causal effect of receiving an additional star.[51] In Table 4, we find that an

[50] In Appendix B.6 and Appendix B.4, we find no evidence to suggest there were compositional changes in the migrants recruited by eligible agencies prior to quality revelation.

[51] We detail this specification in Appendix B.7 and probe its robustness in Appendix C.14.

additional star increases job orders by 67% (significant at the 5% level). We interpret this result as providing evidence of employer response to the star ratings and lending support to the structural assumptions of the conceptual framework in Section 3.

5.5.2 Migrant Response to Star Ratings

In theory, migrants may have responded to the star ratings and influenced placement quality. However, we believe this is unlikely because the primary manner in which the SLBFE chose to disseminate the ratings was through their website. In 2015 (the final year of our dataset), just 12.1% of the Sri Lankan population had access to the internet, making it very unlikely that prospective migrants had access to the ratings (World Bank, 2021). Nevertheless, using the same RDD discussed in the previous section, we test whether migrants respond to higher ratings. In Table 4, we find that migrant recruitment (column 2) and the share of domestic workers (column 3) do not respond to an increase in the star rating of an agency. Migrants who made complaints prior to the ratings program may be especially sensitive to the ratings, but we find that agencies receiving higher ratings are no more likely to attract such migrants.

6 Robustness

6.1 Spillover Effects: Geographic Proximity & Assortative Matching

Agencies are often clustered in the same area as they are not allowed to have multiple physical premises by law. Panel A of Appendix A.14 shows that many agencies are located in Colombo and Kurunegala, while Panel B shows clusters of agencies in Colombo District. The results in the preceding sections suggest that eligible agencies secure more job orders and are less likely to exit the market. These effects may result from eligible agencies poaching job orders and/or migrants available to comparison agencies. To test this hypothesis, in Table 5 we estimate whether

While the estimates are fairly stable, we have limited power and the data are quite variable (see Appendix C.13), so we report bias-corrected estimates (Cattaneo et al., 2019).

the presence of an eligible agency in a 100 foot radius of a comparison agency influences outcomes. We use this radius because agencies tend to be clustered in the same building, but our results are robust to alternative radii (see C.9). The specification includes a fixed effect of the local market interacted with year fixed effects and we cluster standard errors at the local market level (see Panel B of Appendix A.14).

Rather than eligible agencies poaching job orders or migrants from their neighbors in the comparison group, we find that the presence of an eligible agency *increases* migrant recruitment (column 1) and job orders (column 3) for comparison agencies, though the latter estimate is marginally imprecise (t-stat $= 1.64$). Being in close proximity to an eligible agency also does not reduce the job order quality (i.e. Good Orders) of comparison agencies (column 4). These results suggest that the positive market outcomes that eligible agencies experience do not come at the expense of nearby comparison agencies, who are no more likely to exit the market (column 2) when they are located near an eligible agency as opposed to when they are not. If anything, the positive spillovers—perhaps owing to high fixed costs in securing new job orders which are then shared with neighboring agencies—on migrant recruitment and job orders suggest that we are underestimating the effects of the program.[52]

6.2 Agency Quality & Mean Reversion

Measurement error in our classification of pre-program agency quality may imply that the heterogeneous effects we report result from mean reversion. That is, the program doesn't improve outcomes for low-quality agencies, but rather that these were high-quality agencies that were improperly classified and trended back to their mean. The key argument against mean reversion is that high and low-quality eligible agencies are always being compared to high and low-quality comparison agencies. As such, any measurement error would apply similarly to both sets of agencies and the heterogeneous treatment effects we report, therefore, capture the *differential* effect net of any measurement error. We also assess the robustness of our main results to two alternative measures of agency quality: the 2009 Quantitative Score and the average Quanti-

[52] In Appendix C.10 we find limited evidence in support of alternative spillover mechanisms.

tative Score for 2007-09. Across these measures of agency quality our estimates have similar magnitudes, albeit with varying precision (see Appendix C.8).

7 Discussion and Conclusion

In the presence of information frictions, adversely selected labor market intermediaries may exploit their informational advantage to the detriment of both workers and employers (Autor, 2008). A policy that relaxes these frictions may induce intermediaries to improve their match quality and result in the exit of low-quality intermediaries from the market as in Lee (2007). In assessing the effect of the program on market efficiency, we first focus on the quantity and quality of migrant placements and then discuss the entry and exit of agencies and employers.

7.1 Extent of Migration, Placement Quality, and Match Surplus

Inference about how the program influences the extent of migration is complicated by the fact that the overall trend in the market is one of steady decline—see Panel A of Appendix B.5. This decline is in part due to the partial ban on migration discussed in Section 2.1.[53] However, our estimates suggest that the program *may* have increased the amount of migrants recruited by eligible agencies, though these estimates are imprecise. If migrants are scarce, an increase in recruitment by eligible agencies may come at the expense of comparison agencies. Yet, when we explore this possibility in Section 6.1, we find that comparison agencies located near eligible agencies *increased* the amount of migrants they recruited. Overall, our results suggest that the program may have, if anything, increased the extent of migration.

Though the effect of the program on the extent of migration is inconclusive, our findings suggest the program was successful in improving placement quality. The results in Table 3 suggest eligible agencies place migrants with employers that pay higher salaries, provide more amenities, and are less abusive. Unsurprisingly, this results in more contract renewals. More-

[53] We note that the reform has little effect on our key estimates (see Appendix C.12).

over, we do not find evidence to suggest that these improvements came at the cost of some migrants. Eligible agencies do not alter the composition of migrants placed abroad (Appendix B.4) and, likewise, whether or not a comparison agency is exposed to nearby a eligible agency does not change whether they place migrants with Good Employers (Table 5).

Finally, in assessing the effect of the program on match surplus we must combine our previous analysis of how the program affected both the extent of migration and placement quality. To create a proxy for overall match surplus that combines these margins, we note again that a consequence of a dual-market characterized by monopsony power is that less abusive employers tend to offer better contracts both within and across industries (see Section 2.1). Consequently, a proxy for the match surplus generated by an agency is the sum of the salaries of migrants they have placed, which we term 'total salaries'.

In Appendix C.4, we find that total salaries for eligible agencies increase by 69% prior to the QR-phase and stay 71% higher thereafter (column 1); the effect is larger for low-quality eligible agencies (109%) and positive, but imprecise, for high-quality agencies. We do not find evidence for negative spillover effects using this proxy. Instead, total salaries are higher for comparison agencies in close proximity to eligible agencies, though this estimate is imprecise (t-stat = 1.26). Using the estimated effect on total salaries, we compute the program generates approximately USD 2 million in increased remittance income per year (see Appendix A.21).

7.2 Effects on Agencies and Employers

Our results suggest the program left migrants better off with higher quality and, likely, more placements. We weigh these gains in match surplus against entry and exit for agencies and employers. First, we find that eligible agencies are less likely to exit the market in both phases of the program in Panel A of Table 2. Low-quality eligible agencies drive this result: by 2015, 23% of comparison agencies (both high and low-quality) and high-quality eligible agencies had exited the market. In contrast, just 6% of low-quality eligible agencies exit the market (Appendix A.19). Although eligible agencies are less likely to exit the market, this does not appear to come at the cost of nearby comparison agencies (Panel B, column 2 of Table 5). These

results suggest the program greatly improved the viability of low-quality eligible agencies.

Turning to employers, we first note that we do not observe their exit and entry decisions or the commissions they pay agencies. Consequently, our ability to assess the effects of the program on employers is limited. However, we note that Good Employers are more likely to match with highly rated agencies after quality revelation (Panel C of Figure 3). While this result suggests the program may leave them better off, if competition intensifies for the services of highly rated agencies, they may also have to increase the commissions paid to these agencies. In contrast, we think demand from Bad Employers is more elastic; these employers likely value match quality less and may recruit non-Sri Lankan migrants in response to the ratings program. In support of this view, we find limited evidence to suggest that Bad Employers responded to quality revelation (See Appendices B.8 and A.17). In sum, while our results suggest the program benefited migrants and agencies, we are unable to definitively assess its effects on employers.

7.3 Implications for Policy and Future Research

Labor abuse experienced by international migrants is a phenomenon of great human concern, yet it is understudied in the economics literature. This perhaps owes to the politically sensitive nature of the topic and the difficulty of collecting reliable data on these outcomes. By linking complaints made by migrants to local intermediaries and foreign employers, we are able to assess the effects of domestic regulatory policy on migrant placement outcomes abroad.

Migrant-sending countries typically have limited power to influence labor laws in destination countries and must instead resort to pre-departure training or banning migration to curb abuse (Peru, 2023). Our results and those from Bazzi et al. (2021) showing how an information intervention can influence migrant intermediary choice, instead show that domestic regulatory policy can improve migrant placement by influencing the actions of migrants, intermediaries, and foreign employers. While intermediaries may understandably be keen to comply with the regulator, why do foreign employers respond to the program? Our results suggest the ratings act as certification: rather than conveying a precise signal of quality, agencies that carry any star rating—often included in their own marketing, see Appendix A.12—likely appear more

reputable to prospective employers.

A limitation of the program is that it does not influence the incidence of employer abuse so much as the burden on whom it falls. While placement quality for Sri Lankan migrants improves, this may come at the expense of non-Sri Lankan migrants placed at Bad Employers. Alternatively, if other sending countries invest in similar programs this may either discipline Bad Employers or drive them out of the market; estimating these spillover effects is an important direction for future research. In addition, future research should aim to understand how the market power of intermediaries may limit the welfare gains from migration and assess which policies can promote migrant choice and welfare. As labor intermediaries do not engage in arbitrage, the lack of prices through a supply chain limits the use of 'pass-through' as a way to deduce market structure. Future work may benefit from imposing structure on the process of matching so as to estimate the importance of intermediary market power.

References

Abu-Habib, Lina (1998). The use and abuse of female domestic workers from Sri Lanka in Lebanon. *Gender & Development*, 6(1):52–56.

Amnesty (2020). Reality check 2020: Countdown to the 2022 World Cup migrant workers' rights in Qatar.

Anderson, Michael L (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Andrabi, Tahir, Das, Jishnu, and Khwaja, Asim Ijaz (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review*, 107(6):1535–63.

Armstrong, Mark (2006). Competition in two-sided markets. *The RAND journal of economics*, 37(3):668–691.

Aryasinha, Ravinatha (2020). The role of foreign employment in the economic revival of Sri Lanka; challenges and opportunities. *Ministry of Foreign Relations*.

Atkin, David and Donaldson, Dave (2015). Who's getting globalized? the size and implications of intra-national trade costs. *Mimeo*.

Autor, David (2008). The economics of labor market intermediation: an analytic framework.

Batista, Catia and McKenzie, David J (2021). Testing classic theories of migration in the lab.

Bazzi, Samuel, Cameron, Lisa, Schaner, Simone, and Witoelar, Firman (2021). Information, intermediaries, and international migration.

Bergquist, Lauren Falcao and Dinerstein, Michael (2020). Competition and entry in agricultural markets: Experimental evidence from Kenya. *American Economic Review*, 110(12).

Bulow, Jeremy I and Summers, Lawrence H (1986). A theory of dual labor markets with application to industrial policy, discrimination, and Keynesian unemployment. *Journal of labor Economics*, 4(3, Part 1):376–414.

Cabral, Luís MB (2005). The economics of trust and reputation: A primer. *New York University and CEPR*, 2005.

Calonico, Sebastian, Cattaneo, Matias D, and Titiunik, Rocio (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Cattaneo, Matias D, Idrobo, Nicolás, and Titiunik, Rocío (2019). *A practical introduction to regression discontinuity designs: Foundations*. Cambridge University Press.

Clemens, Michael A (2019). Measuring the spatial misallocation of labor: The returns to India-Gulf guest work in a natural experiment. *IZA Discussion Paper*.

de Oliveira, Priscilla (2023). Why businesses fail: Underadoption of improved practices by Brazilian micro-enterprises. *Mimeo*.

Demaret, Luc (2006). Private employment agencies: The challenges ahead from the workers' perspective. *Merchants of Labour*, page 159.

Dranove, David and Jin, Ginger Zhe (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4):935–963.

Eggers, Andrew C, Freier, Ronny, Grembi, Veronica, and Nannicini, Tommaso (2018). Regression discontinuity designs based on population thresholds: Pitfalls and solutions. *American Journal of Political Science*, 62(1):210–229.

Fernando, A. Nilesh and Lodermeier, Alison (2022). Understanding adverse outcomes in Gulf migration: Evidence from administrative data from Sri Lanka. *International Migration Review*.

Frantz, Elizabeth (2008). Of maids and madams: Sri Lankan domestic workers and their employers in Jordan. *Critical Asian Studies*, 40(4):609–638.

Grembi, Veronica, Nannicini, Tommaso, and Troiano, Ugo (2016). Do fiscal rules matter? *American Economic Journal: Applied Economics*, pages 1–30.

Human Rights Watch (2007). Exported and exposed: Abuses against Sri Lankan domestic workers in Saudi Arabia, Kuwait, Lebanon, and the United Arab Emirates.

International Labor Organization (2013). Operational manual for labour sections of Sri Lankan diplomatic missions in labour receiving countries.

International Labor Organization (2019). ILO global estimates on international migrant workers.

Jureidini, Ray and Moukarbel, Nayla (2004). Female Sri Lankan domestic labour in Lebanon: contractual, slavery-like practices and conditions. *Journal of Ethnic and Migration Studies*, 30(4):581–607.

Lee, Woong (2007). Private deception and the rise of public employment offices in the United States, 1890-1930. Technical report, National Bureau of Economic Research.

Macchiavello, Rocco and Morjaria, Ameet (2015). The value of relationships: evidence from a supply shock to Kenyan rose exports. *American Economic Review*, 105(9):2911–45.

Macchiavello, Rocco and Morjaria, Ameet (2021). Competition and relational contracts in the Rwanda coffee chain. *The Quarterly Journal of Economics*, 136(2):1089–1143.

Mitra, Sandip, Mookherjee, Dilip, Torero, Maximo, and Visaria, Sujata (2018). Asymmetric information and middleman margins: An experiment with Indian potato farmers. *Review of Economics and Statistics*, 100(1):1–13.

Mobarak, Ahmed Mushfiq, Sharif, Iffath, and Shrestha, Maheshwor (2023). Returns to international migration: Evidence from a Bangladesh-Malaysia visa lottery. *American Economic Journal: Applied Economics*, 15(4):353–388.

Naidu, Suresh, Nyarko, Yaw, and Wang, Shing-Yi (2016). Monopsony power in migrant labor markets: evidence from the United Arab Emirates. *Journal of Political Economy*, 124(6):1735–1792.

Naidu, Suresh, Nyarko, Yaw, and Wang, Shing-Yi (2023). The benefits and costs of guest worker programs: Experimental evidence from the India-UAE migration corridor.

Orrenius, Pia M (2001). Illegal immigration and enforcement along the US-Mexico border: An overview. *Economic and Financial Review-Federal Reserve Bank of Dallas*, (1):2–11.

Peru, Dhanushka Arunasiri (2023). Women's labor market opportunities and fertility decisions: Evidence from Sri Lanka. *Mimeo*.

Roth, Jonathan (2022). Pre-test with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*.

Rysman, Marc (2009). The economics of two-sided markets. *Journal of economic perspectives*, 23(3):125–143.

Seshan, Ganesh and Yang, Dean (2014). Motivating migrants: A field experiment on financial decision-making in transnational households. *Journal of Development Economics*, 108:119–127.

Shah, Nasra M (2013). Labour migration from Asian to GCC countries: Trends, patterns and policies. *Middle East Law and Governance*, 5(1-2):36–70.

Shivakoti, Richa, Henderson, Sophie, and Withers, Matt (2021). The migration ban policy cycle: a comparative analysis of restrictions on the emigration of women domestic workers. *Comparative Migration Studies*, 9(1):1–18.

Shrestha, Maheshwor (2017). Get rich or die tryin': perceived earnings, perceived mortality rate and the value of a statistical life of potential work-migrants from Nepal.

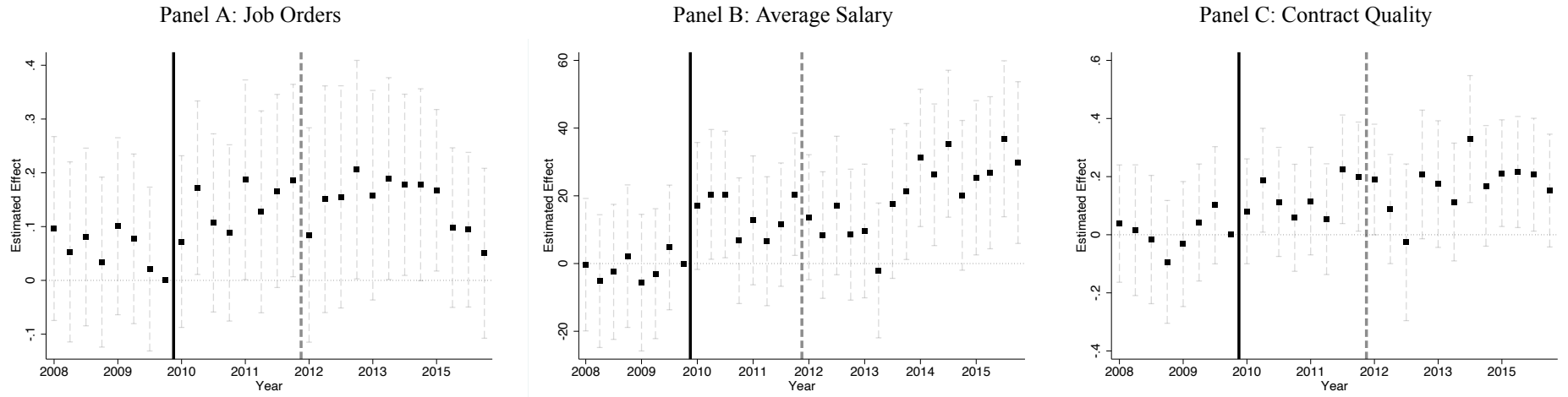
Silva, D (2013). Rizana Nafeek executed. *Daily Mirror*.

UNODC (2015). The role of recruitment fees and abusive and fraudulent recruitment practices of recruitment agencies in trafficking in persons.

World Bank (2021). *World Development Indicators 2021*. The World Bank.

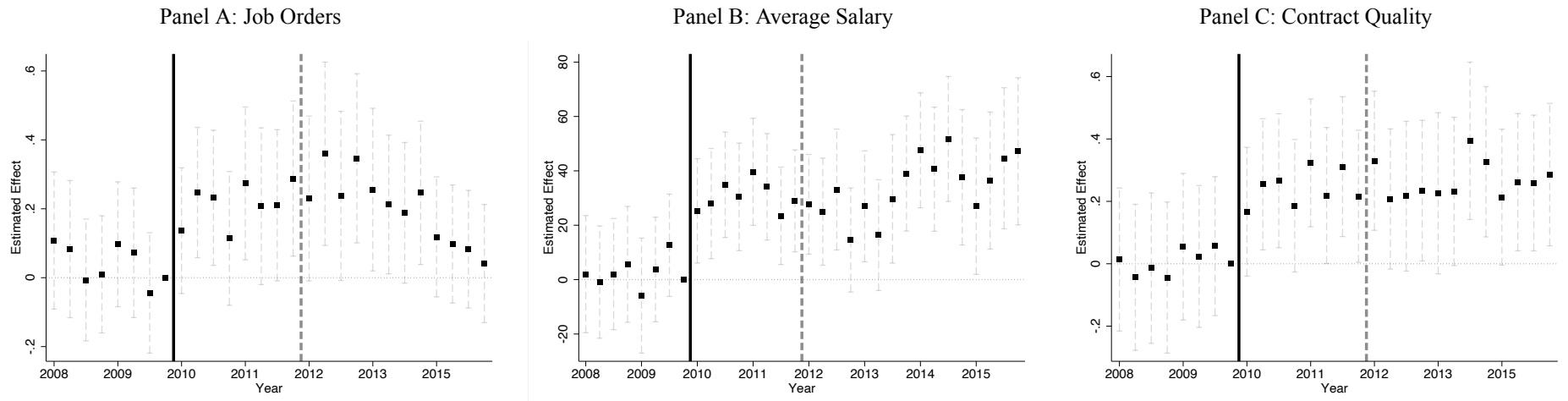
Yang, Dean (2011). Migrant remittances. *Journal of Economic perspectives*, 25(3):129–152.

Figure 1: Event Study Graphs - All Agencies



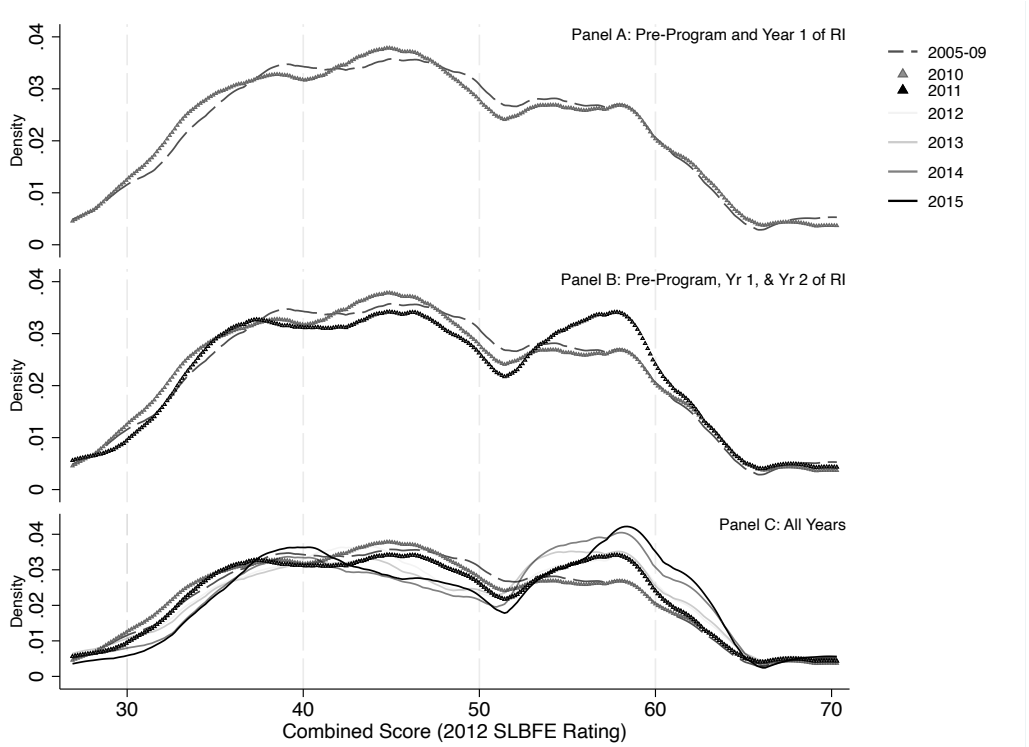
Notes: These figures show event study graphs for all agencies across three outcomes: job orders, average salaries, and a contract quality index (see Appendix A.13 for details). Data for 2014 in missing for Panel A & C. The figures plot an indicator for an eligible firm interacted with a series of quarterly indicators (i.e. the black squares) for 2008-15 (see Appendix C.1 for the details). The light gray dashed lines are the 95% confidence intervals. The reference group is the final quarter of 2009. The solid black vertical line indicates the program announcement, while the dashed gray vertical line indicates the ratings' release; both lines are shifted half a quarter to the left.

Figure 2: Event Study Graphs - Low Quality Agencies



Notes: These figures show event study graphs for low quality agencies across three outcomes: job orders, average salaries, and a contract quality index (see Appendix A.13 for details). Data for 2014 in missing for Panel A & C. The figures plot an indicator for an eligible firm interacted with a series of quarterly indicators (i.e. the black squares) for 2008-15 (see Appendix C.1 for the details). The light gray dashed lines are the 95% confidence intervals. The reference group is the final quarter of 2009. The solid black vertical line indicates the program announcement, while the dashed gray vertical line indicates the ratings' release; both lines are shifted half a quarter to the left.

Figure 3: The Distribution of Matches to Good Employers by Agency Quality



Notes: This figure plots probability density functions describing matches between agencies and 'Good Employers': employers with a below median pre-program ($<$ 2010) complaint rate. The domain is the continuous Combined Score/2012 agency rating which is only known to the public after 2012. Panel A plots the density function for pre-program matches (dashed gray line) and those during 2010, i.e. the first year of the RI-phase (lighter triangles). Panel B adds 2011, year two of the RI-phase (darker triangles). Panel C plots the remaining years (2012-2015) corresponding to the QR-phase (solid lines, darker lines refer to later years).

Table 1: Agency-Month Level Summary Statistics, 2005-2015

	Mean	SD	Median	N
	(1)	(2)	(3)	(4)
# Migrants	8.23	20.79	4.00	27324
# Female Migrants	4.95	8.84	2.00	27324
# Domestic Worker Migrants	5.02	9.74	2.00	27324
# Migrants to Saudi Arabia	1.81	8.00	0.00	27324
Salary (USD), 2015 dollars	178.80	118.38	176.92	23777
# Job Orders	0.48	0.87	0.00	22356
# Vacancies	68.30	177.21	0.00	22356
# Complaints	0.66	1.62	0.00	27324
Complaint rate	0.08	0.15	0.00	19659
Complaint solved rate	0.88	0.31	1.00	6765
Complaint solved time	4.62	5.09	3.00	6765

Notes: This table reports summary statistics at the agency-month level. Salaries are converted to 2015 US dollars. Job Orders are the number of orders that an agency received in a specific month from employers abroad (i.e. demand), while vacancies reports the average number of vacant positions contained in a job order. Note, data for 2005 and 2014 is missing for the job order data. Complaints reports the number of complaints made to Sri Lankan consulates by migrants who were placed by agencies in the relevant month. The complaint rate measures the average number of complaints by migrants placed by an agency in a given month conditional on the agency placing non-zero migrants. The complaint solved rate and time further condition on the set of agencies who had non-zero complaints in a given month.

Table 2: Investment in Rating Criteria and Market Outcomes

	Quant Score (1)	Migrants (2)	Job Orders (3)	Agency Exit (4)
Panel A: Effects by Program Phase				
Elig*2010-12	0.343	0.199	0.107*	-0.070***
	(0.209)	(1.585)	(0.062)	(0.026)
Elig*2012-15	0.254	3.378	0.145*	-0.083*
	(0.195)	(3.266)	(0.078)	(0.044)
Panel B: Effects by Agency Quality				
Elig*High*Post-2010	-0.342	2.183	-0.011	0.007
	(0.265)	(5.408)	(0.114)	(0.074)
Elig*Low*Post-2010	0.657***	1.860	0.220***	-0.126***
	(0.200)	(2.093)	(0.083)	(0.039)
Panel C: Effects by Agency Quality and Program Phase				
Elig*High*2010-12	-0.072	-4.428	-0.062	-0.023
	(0.346)	(3.392)	(0.098)	(0.053)
Elig*High*2012-15	0.398	5.919	0.029	0.024
	(0.335)	(7.755)	(0.145)	(0.093)
Elig*Low*2010-12	0.672***	2.393	0.215***	-0.092***
	(0.240)	(2.009)	(0.078)	(0.031)
Elig*Low*2012-15	0.358*	1.558	0.224**	-0.144***
	(0.212)	(2.224)	(0.094)	(0.046)
N	27324	27324	22356	27324
Comparison Mean	5.80	7.61	0.34	0.16

Notes: *Quant Score* summarizes investment in the rating criteria, *Migrants* reports monthly recruitment, *Job Orders* are contracts received from employers (data from 2005 and 2015 are missing), and *Agency Exit* indicates a 12 month sequence of zero recruitment (see Appendix A.13). All specifications cluster at the agency level, include agency FE, and month-year FE x avg. migrant recruitment (05-08). Panel A also contains month-year FE x avg. Quantitative Score (05-08).

Table 3: Placement Quality

	Avg. Salary	Contract Renewal	Good Orders	Contract Quality
	(1)	(2)	(3)	(4)
Panel A: Effects by Program Phase				
Elig*2010-12	11.488	0.009	0.053**	0.050
	(8.859)	(0.011)	(0.025)	(0.069)
Elig*2012-15	13.415	0.014	0.035	0.068
	(10.329)	(0.011)	(0.025)	(0.087)
Panel B: Effects by Agency Quality				
Elig*High*Post-2010	8.731	0.001	0.076*	-0.063
	(19.052)	(0.013)	(0.045)	(0.151)
Elig*Low*Post-2010	19.754**	0.023**	0.050**	0.167**
	(9.570)	(0.011)	(0.025)	(0.078)
Panel C: Effects by Agency Quality and Program Phase				
Elig*High*2010-12	-5.647	0.009	0.061	-0.140
	(18.394)	(0.024)	(0.044)	(0.133)
Elig*High*2012-15	18.926	-0.001	0.090*	0.009
	(22.909)	(0.014)	(0.052)	(0.194)
Elig*Low*2010-12	24.924***	0.013	0.068**	0.191***
	(9.123)	(0.013)	(0.029)	(0.073)
Elig*Low*2012-15	16.642	0.028**	0.035	0.148
	(11.781)	(0.013)	(0.026)	(0.099)
N	21865	21865	18365	18365
Comparison Mean	164.90	0.15	0.17	0.31

Notes: *Avg. Salary* refers to average migrant salaries (constant in 2015 USD). *Contract Renewal* is the share of migrants who renewed their contracts, *Good Orders* are job orders from employers with a below median pre-program complaint rate, and *Contract Quality* is a normalized index of non-wage amenities (see Appendix A.13). All specifications cluster at the agency level, include agency FE, and month-year FE x avg. migrant recruitment (05-08). Panel A also contains month-year FE x avg. Quantitative Score (05-08).

Table 4: The Effect of Receiving a Higher Rating

	Job Orders	Migrants	Domestic Share	Pre-Complaints
	(1)	(2)	(3)	(4)
Additional Star	1.006**	6.093	-0.009	0.164
	(0.464)	(8.669)	(0.108)	(0.228)
N	20714	28060	28058	28060
Depvar. Mean	1.50	25.43	0.25	0.56

Notes: This table estimates the effect of receiving an additional star on the government's star rating program using a regression discontinuity design. The sample consists of *all* agencies eligible for the rating program. We report bias-corrected estimates which pool across the star specific thresholds (see Appendix B.7 for details), use a uniform kernel, and use the MSE-optimal bandwidth using the rdrobust package (see Appendix C.14 for conventional and robust estimates). All specifications include year fixed effects and standard errors are clustered at the agency-level. Job Orders are the number of orders that an agency received in a specific month from employers abroad (i.e. demand). Migrants are the number of migrants placed by an agency in a given month. Domestic Share is the share of migrants who are domestic workers placed by the agency in a given month. Pre-Complaints is a count variable for the number of migrants who made a complaint prior to the program and subsequently migrated again after the program.

Table 5: Spillover Effects

	Migrants	Agency Exit	Job Orders	Good Orders
	(1)	(2)	(3)	(4)
Any Elig*Post-2010	4.217*	0.002	0.242	0.082
	(2.275)	(0.059)	(0.147)	(0.061)
N	16104	16104	13176	10650
Comparison Mean	5.93	0.20	0.32	0.17

Notes: This table estimates spillover effects of the program and restricts attention to the sample of comparison agencies. 'Comparison Mean' reports the pre-program (2005-2009) mean of the dependent variable for comparison agencies that do not have an eligible agency within 100 ft. 'Agency Exit' is a proxy for an agency closing down. It is coded as an indicator variable that switches on at the beginning of a 12 month sequence of zero recruitment after the program was introduced. 'Job Orders' are the number of orders that an agency received in a specific month from employers abroad (i.e. demand). Note, data for 2005 and 2014 is missing for the job order data. 'Good Orders' are job orders corresponding to employers with a below median pre-program complaint rate. 'Any Elig' is a dummy variable coded as 1 if there is an eligible agency within 100 ft of a comparison agency. The estimates are clustered at the local market level and the specification includes agency fixed effects, month-year fixed effects interacted with average 2005-2008 migrant recruitment, month-year fixed effects interacted with average Quantitative Score 2005-2008 fixed effects, and local market x month-year fixed effects.

Appendices - Table of Contents

For Online Publication

A Program Information and Context **49**

A.1 Program Timeline 49

A.2 Criteria used for Agency Rating Program 50

A.3 Calculation of Star Ratings 52

A.4 Complaint Types 53

A.5 Example of Complaint Mechanism 54

A.6 Booklet Provided to Agencies 55

A.7 Migration Channels 58

A.8 Job Order Process 59

A.9 Distribution of Ratings 60

A.10 Quarterly Variation in Recruitment 61

A.11 Agency Rating Online Database 62

A.12 Online Advertisements 63

A.13 Data Appendix 64

A.14 Agency Locations 66

A.15 Pre-Program Employer to Agency Matches 67

A.16 Wages and Amenities by Employer Complaint Rates 68

A.17 Matches to Bad Employers by Agency Quality 69

A.18 Investment in the Audit Score 70

A.19 Agency Exit Over Time 71

A.20 A Model of Agency Reputations 72

A.21 Returns to the Program 75

B Additional Results **76**

B.1 Results from Agency Survey 76

B.2 Characteristics of Job Orders 77

B.3 Components of Contract Quality 78

B.4 Composition of Migrants Recruited 79

B.5 Raw Trends 80

B.6 Components of Quantitative Score 81

B.7 The Effect of Receiving a Higher Rating 82

B.8 Salaries Paid by Bad Employers 83

B.9 Heterogeneous Effects for Low Quality Agencies 84

B.10 Spillovers and Ethnic Networks 85

B.11 K-S Tests for Employer-Agency Matches 86

C Robustness Tests **87**

C.1 Pre-Trends 87

C.2 Plots of Lagged Effects 88

C.3 Agency Exit 89

C.4 Total Salaries 90

C.5 Complaint Rate 91

C.6 Robustness to Agency Exit 92

C.7 Robustness to Trend Controls 93

C.8 Mean Reversion and Measures of Agency Quality 94

C.9 Robustness of Spillover Effects 96

C.10 Do Bad Employers match with Comparison Agencies? 97

C.11 Program Effects with Alternative Samples 98

C.12 Robustness of Estimates to FBR Reform 99

C.13 RD Plots for Job Orders 100

C.14 Robustness of RD Estimates 101

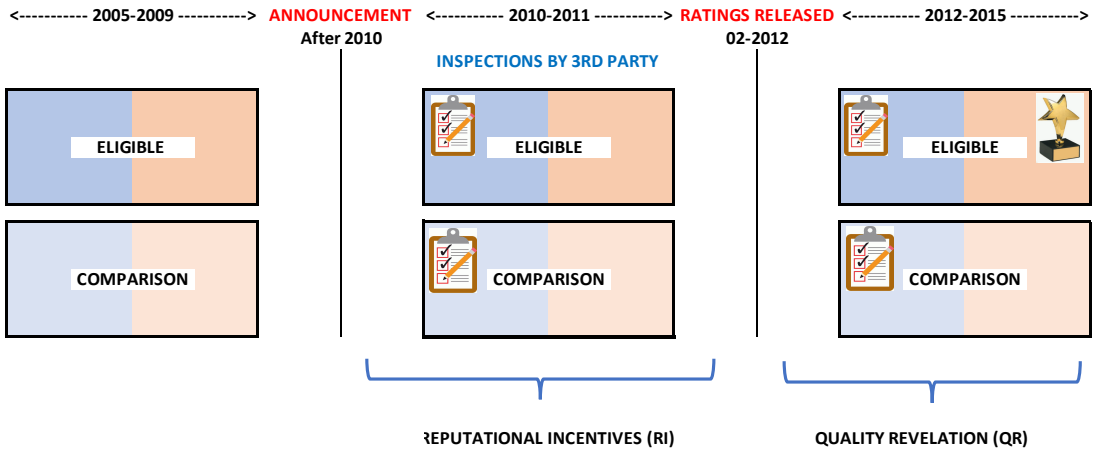
Appendix A Program Information and Context

Appendix A.1 Program Timeline

Panel A: Program Phases

- Reputational Incentives (RI) Phase (2010-2012):
  - *January 2010:* Program announced to agencies and booklet with grading criteria is sent through SLBFE management system
  - *July 2010- May 2011:* Third-party inspections/audits
- Quality Revelation (QR) Phase (2012-2015):
  - *February, 2012:* Ratings announced at public ceremony
  - *June, 2012:* Ratings made available on online database

Panel B: Timeline with Potential Effects



Notes: Panel A shows a timeline for the agency ratings program and lists the potential effects below it. In Panel B, the timeline with potential effects illustrates the difference-in-difference design. We compare a set of eligible agencies to comparison agencies below the 100 migrant recruitment threshold in 2009. After the announcement, the rating criteria (signified by the notebook icon) are announced to all agencies, and third-party inspections take place shortly thereafter. The ratings are released to the public in February, 2012. We refer to the period between the program announcement and the public release of ratings as the 'Reputational Incentives' phase of the program. We refer to the period after the public release of the ratings as the 'Quality Revelation' period. This period, of course, also includes the effects of RI. The blue and red shading denotes agencies of varying pre-program quality.

Appendix A.2 Criteria used for Agency Rating Program

A2.1: Criteria Used for Quantitative Score

- Total Numbers of Persons Recruited
- Skill Level of Jobs
  - Professional jobs
  - Middle-level Employment
  - Skilled workers
  - Semi-skilled workers
  - Unskilled workers
  - Domestic Workers
- Recruitment to New Destinations
- Dispute Settlement Rate
- Diversification of Job Market
- Cess payment

A2.2: Criteria Used for Audit Score

- Record-Keeping
  - Timely submission of license renewal application
  - Timely submission of bank guarantee for license application
  - Book A: Register containing details of migrants
  - Book B: Register of job offers
  - Book C: Register on remittance reimbursed from the Bureau
  - Book D: Register containing details of Foreign Employers or Commissions
  - Maintaining Receipts
  - Register containing details of passports
- Marketing
  - Proper Maintenance of Office
    - Location of agency
    - Display of name boards
    - Attractive appearance
  - Interior Layout

        - Division of office area
        - Process Management
        - Good equipment
    - Sustainability of business (continuity of operation)
    - Outlook of Office Staff
        - Uniform
        - ID Card
    - Implementing Awareness Programs
        - Registry of bio-data of employers
        - Registry of bio-data of recruits and details of their family background
        - Maintenance of a notice board
    - Implementing pre-migration awareness programs which educate them on:
        - Destination country
        - Company or employer
        - Nature of employment
        - Material data on country
        - Common customs to be followed
        - Do away with social stigmas
        - Actions to be taken if problems crop up

- Staff
    - Assessment of Human Resources
        - Graduate and diploma holders
        - A-Level holders
        - O-Level holders

- Other(Bonus Points)
    - Recruiting persons without charging fees
    - Sending the employee direct to employer
    - License not temporarily discontinued during license year
    - Not receiving any letters of warnings over non-addressing of complaints
    - Not receiving any other letters of warning issued by the Bureau

Appendix A.3 Calculation of Star Ratings

Agencies were awarded a *star rating* between zero to five stars based on thresholds corresponding to a continuous score: the *Combined Score*). The Combined Score is itself a weighted combination of two constituent scores: the Quantitative Score and the Audit Score. The Quantitative Score uses administrative data to assess placement quality and its criteria include the volume of recruitment, skill and market diversification, and the successful resolution of complaints (9 criteria in total). The Audit Score uses data collected through third-party audits to assess compliance with recommended management practices such as record-keeping, the maintenance of office premises, and the qualifications of agency staff (24 criteria in total). For a full list of the criteria, see Appendix A.2.

Where c is the Combined Score, $c = (Quantitative\ Score + Audit\ Score) * (4/7)$ and c is then mapped to a star rating as follows—0 stars: $0 \leq c \leq 19.9$, 1 star: $20 \leq c \leq 34.9$, 2 stars: $35 \leq c \leq 49.9$, 3 stars: $50 \leq c \leq 59.9$, 4 stars: $60 \leq c \leq 79.9$, and 5 stars: $c \geq 80$.

The Quantitative Score is computed by assigning letter grades to continuous values for 9 criteria. For each variable, the value of the variable was converted into a letter grade (A=5, B=4, C=3, D=2, E=1, Z=0). Letter grades were assigned in the following way: the range of the variable is divided into 5 equal bins; then letter grades are assigned based on these bins where the highest-scoring bin receives letter grade "A" and the lowest receives "Z." If a score could not be calculated for the criterion, a letter grade "Z" was assigned. A total score was then calculated by summing the numeric values associated with all letter grades.

For each of the 24 criteria relating to the Audit Score, agencies were similarly awarded letter grades which converted into the same number of points as with the Quantitative Score. In this case, the grades awarded relate to compliance with the recommended practices, though the precise manner through which these were awarded are unclear.

Appendix A.4 Complaint Types

	Complaints	Percentage of Total Complaints
Non Receipt of Wages	19741	16.62
Breach of Employment Contract	15201	12.80
Harassment	13407	11.29
Sickness	12522	10.54
Others	11775	9.92
Lack of Communication	11322	9.53
Not Sent Back After Completion of Contract Period	9509	8.01
Problem at Employees Home (Sri Lanka)	6207	5.23
No Correspondence to Complaints	3866	3.26
Being Stranded Without Employment	2849	2.40

Notes: The table above shows the decomposition of complaints over the span of the dataset 2005-2015.

Appendix A.5 Example of Complaint Mechanism

Licensee,
[redacted] 2018.10.30

Complaint no: [redacted]
Migrant worker: [redacted]
Passport no: [redacted] DEPARTURE DATE: 2016.[redacted]

The conciliation division received a complaint on the 2018.06.08 regarding the migrant worker mentioned above. This is regarding the complaint and the investigation on the 2018.06.20.

It was mentioned in her contract that her salary would be 1125 SR but she was actually paid 900 SR. Every month she was not paid 225 SR. Thus the total outstanding salary for 24 months is (9225x 24) for 24 months is 5400 SR. The letter dated 2018.06.20 states that the balance salary should be paid and make arrangements get her back to Sri Lanka. However, She came back on the 2018. 10. 15 but the balance salary was not paid.

Therefore, make sure that that the balance salary of 5400SR is paid before the 2018. 11.30 and handover the documents related to that to our office. We kindly inform you that If you fail to do pay the balance, we will take legal action.

Manager (Conciliation)
SLBFE

Licensee,
[redacted] 2018.12.24

Complaint no: [redacted]
Migrant worker: [redacted]
Passport no: [redacted] DEPARTURE DATE: 2016.[redacted]

The conciliation division received a complaint on the 2018.06.08 regarding the migrant worker mentioned above. This is regarding the complaint and the investigation on the 2018.06.20.

An order was given to you to settle the balance payment of the salary (225x24) 5400 SR. This was informed to you on2018.06.20 and 2018.10.30. However, you have failed to settle this matter.

Therefore, we have forwarded your case to the legal division for legal action.

Manager (Conciliation)
SLBFE

Appendix A.6 Booklet Provided to Agencies

A6.1: Importance of the grading methodology

- To bring all the agencies into a well-recognized standard.
- The objectives are to recognize, encourage and appreciate the agencies maintaining high standards.
- Establishing an impartial, justifiable and scientific methodology to grade agencies.
- The Act requires SLBFE to guide agencies towards their growth, development and standardizing their levels.

A6.2: Policies, qualifying/disqualifying and prospects

- Agencies will be qualified for this grading system only if the agency holds a valid license for at least a year and have sent more than 100 people for employment during the assessment year.
- There are 11 indicators selected for this grading methodology which are listed out below:

1. (i) The total number of recruitment (ii) Recruitment according to manpower groups (Professional, middle level, clerks and classical related jobs, skilled labour, semi-skilled labour, unskilled labour and house maids).
2. Settlement of complaints is assessed in two ways: (i) Settlement rate of the conciliation division looks at the annual rate of the resolved complaints by an agency in that division
   (ii) Investigation and complaints settlement rate looks at the annual rate of resolved complaints in the special investigation division The final rate is calculated by taking the average of the two divisions.
3. Recruitment to new destination has two categories, (i) Entering into agreements with new countries (countries that have not received a SL employees through SLBFE) and should have at least sent one person for employment (ii) Entering into agreements with new agencies and companies in the countries where SL has already started business.
4. Ratio difference between the number of vacancies approved and the number of vacancies fulfilled (the number of people recruited).
5. The payment of cess out the commission they receive from foreign countries for providing employees
6. Evaluating human recourses- points if employees are granted with employee's provident fund (EPF) and points according to the educational qualifications such as graduate or a diploma holder, A/L and O/L qualified.
7. Prosecution- If complaints are forwarded to legal division as it has not been solved, law will be enforced. If 10 prosecutions are against an agency, then they will be disqualified from the grading system.

8. Agencies will receive minus points if there are complaints against agencies lodged by the migrant worker, relatives, family members, foreign agencies and foreign employers.
9. Minus points are given to agencies for receiving the approval of a job order but fail to provide employment within the valid period of approval.

A6.3: Disqualification:

- Any agency will not be considered if the following conditions can be applied:

1. If the average complaint ratio is 100%
2. Have sent less than 100 people
3. If they are punished or warned by a court of law for producing forged documents, recruiting underage people, recruited disabled people and has been prosecuted by SLBFE for more than 10 court cases.

A6.4: Proposals and Suggestions

1. Updating the records and documents that are maintained at present such as data sheets of Sri Lankan employees and approved job orders.
2. Records of statutory refunds made by SLBFE, passport details, employers and commissions received.
3. Maintain a log book for government authorized officers to lodge their recommendations.
4. Classifying the letters received from SLBFE according to licensing division, conciliation division and general letters
5. Provide a valid letter of appointment for everyone serving in the agency, listing out the duties and mark attendance.
6. Maintain records of migrant workers who have returned before the contract period expired and sent back to Sri Lanka before the contract period and provide it to SLBFE which has to be authorised by bureau officers.
7. Make the agency staff thorough with laws, regulations, functions of the agency, the agency grading methodology and pay them well and on time.
8. Get the staff to be members of EPF, provide them with uniforms, ID cards and train and educate them on modern technology, the importance of public relations.
9. Submit the application for extending the license and bank guarantee on time.
10. Be vigilant of the authenticity of certificates forwarded by applicants.
11. Maintain a record of the family members of the migrant workers and take maximum action for the welfare of their family members.
12. Be vigilant of the workers being paid regularly and extend insurance if necessary.

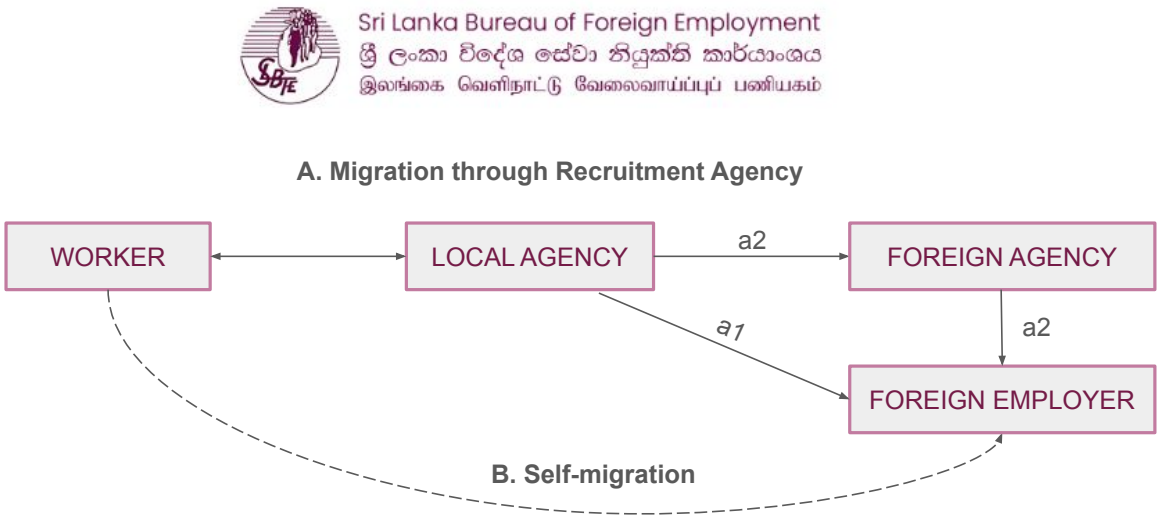
13. Pay attention towards the benefits and salary of the migrant worker as mentioned in the agreement throughout the contract period and persuade the foreign employer if necessary.

A6.5: Benefits

Period of Supervision is from 1st of January to 31st of December every year

1. They will be invited to participate in foreign employment promotion organized by SLBFE.
2. Issue special identity card for the officer in charge of the agency and provide special service counters.
3. Their names will be advertised in SLBFE publications.
4. Granting approval to open web page in the SLBFE website.
5. Discussing to provide all the benefits of government, through the Ministry of Foreign Employment.
6. Standardization of agencies.

Appendix A.7 Migration Channels

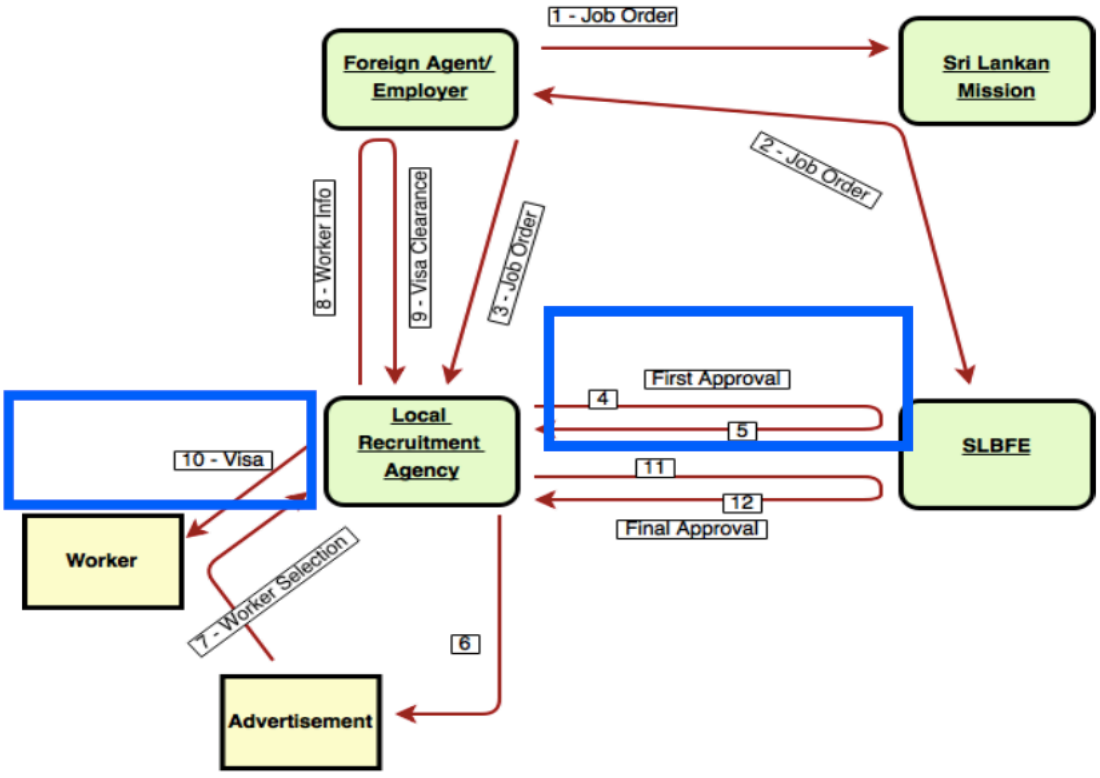


Notes: This figure provides a stylized depiction of the channels through which a migrant is ultimately placed at an employer abroad. In Option A, a migrant is recruited by a local agency (i.e. one based in Sri Lanka) on behalf of either a foreign agency (e.g. one recruiting many domestic workers) or a foreign employer (e.g. a household or a construction company). In path a1, the migrant is placed directly at the ultimate employer. In path a2, the foreign agency ultimately places the migrant at their ultimate employer. In Option B, termed 'Self-migration', a migrant foregoes a recruitment agency and contracts directly with an employer; this is often the case when a migrant is initially placed at an employer through Option A and subsequently renews a contract with this employer. In both cases, the regulator, the Sri Lanka Bureau of Foreign Employment (SLBFE), provides oversight of the migration process.

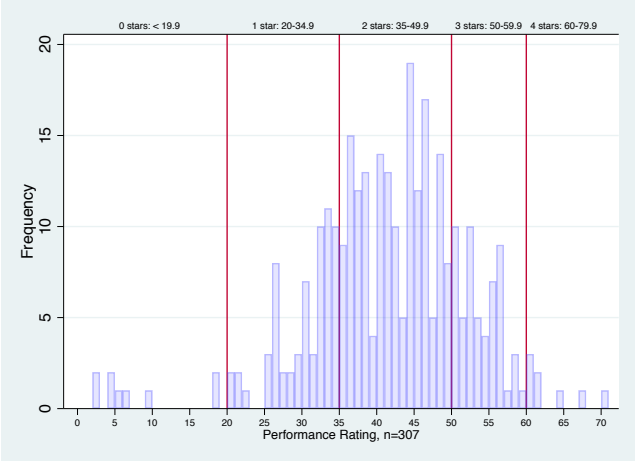
Appendix A.8 Job Order Process

The figure below details a 12-step process from the point at which contact is initiated by a foreign employer to a consulate to to the final approval for a migrant worker's departure. In the first step, the foreign employer creates a job order that describes the number of workers required, their salary, and their benefits. These benefits include whether an employer provides food, accommodation, air travel, and/or health insurance. The Foreign Mission will authenticate the job order and then the local agency will receive the job order. The foreign employer may send the same job order to up to five local agencies. The local agency must then complete the paperwork and submit the job order to the SLBFE for a First Approval.

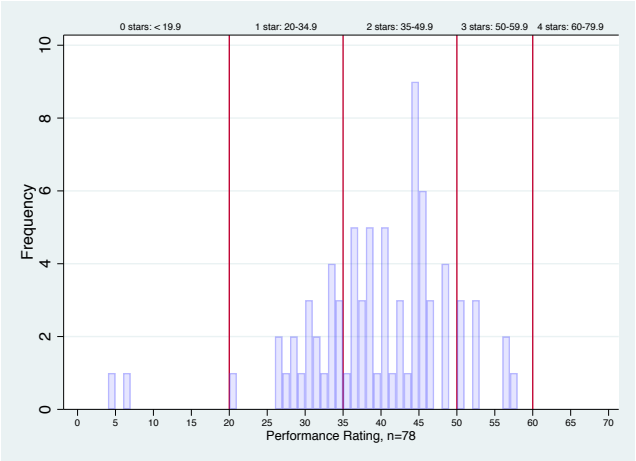
If the job order is approved by the SLBFE, the local agency may advertise in the SLBFE job bank or through local media. Potential migrants are then interviewed and selected. Selected individuals must pass a medical examination, a background check by the Sri Lankan police, and, in some cases, undergo a basic training. The local agency sends prospective worker information to the foreign employer who then obtains visa clearance for the migrant and the local agency obtains visa and air tickets. If the prospective migrant clears these requirements, the job order is sent for final approval to the SLBFE. In the final step, the SLBFE grants final approval and stamps the migrant worker's passport.



Appendix A.9 Distribution of Ratings



(a) Panel A: All Eligible Agencies



(b) Panel B: Recruited 100-155 Migrants in 2009

Notes: This figure displays a frequency distribution of the Combined Score that was calculated for agencies in the agency grading program. The star cutoffs are as follows: 0 stars: 0-19.9, 1 star: 20-34.9, 2 stars: 35-49.9, 3 stars: 50-59.9, 4 stars: 60-79.9, 5 stars: 80-100. Panel A show the distribution for all eligible agencies while Panel B restricts the sample to those who recruited between 100 and 155 migrants in 2009.

Appendix A.10 Quarterly Variation in Recruitment

Notes: This figure shows quarterly recruitment at the agency level for the period determining eligibility (four quarters in 2009) and two quarters prior to 2009. Attention is restricted here to agencies that recruited between 45 and 155 migrants in 2009, where eligible agencies recruited at least 100 migrants.

Appendix A.11 Agency Rating Online Database

Notes: The image above is a screenshot of the agency rating online database from the SLBFE website.

Appendix A.12 Online Advertisements

(a) Panel A

(b) Panel B

Notes: The figures above show advertisements placed by agency owners on their websites.

Appendix A.13 Data Appendix

We describe the main variables used in our analyses below. All variables are measured at the agency-month level, unless stated otherwise.

Average salary (USD): Mean salary of migrants departing per month, deflated to 2015 US dollars, winsorized at the 5% level.

Contract renewal: Share of migrants in a given month who renew their contract with the same employer or self-migrate to the same country and sector in a subsequent trip. We are able to match migrant trips by the same individual using unique national identifiers available in our data.

Complaint rate: Fraction of migrants departing in a given month who file a complaint at anytime. For example, if agency A sent 10 migrants in March 2013 and 3 of these migrants complain at a later date, the complaint rate for the agency in March 2013 is 0.3.

Solve time for complaints: Average number of months taken to resolve complaints filed in a given month

Job orders: Number of job orders received by an agency in a month, after approval by the regulator. Winsorized at the 5% level. We do not have data for years 2005 and 2014.

Good Orders: Number of job orders from foreign employers whose pre-program (prior to 2010) complaint rate is below the median. This value is only generated for employers who recruit at least 50 migrants during the pre-period (accounting for 93% of all migrants). We do not have data for years 2005 and 2014.

Migrants: Number of migrants recruited per month, winsorized at the 5% level.

Construction of Agency Quality Measures

We use the formula used by the SLBFE to compute a Quantitative Score. This is computed both at the yearly level and at the monthly level as we detail below. First, we detail how the government computes the Quantitative Score:

How the Quantitative Score is Computed:

The Quantitative Score is computed by assigning letter grades to continuous values for 9 assessment variables listed below. For each variable, the value of the variable was converted into a letter grade (A=5, B=4, C=3, D=2, E=1, Z=0). Letter grades were assigned in the following way: the range of the variable is divided into 5 equal bins; then letter grades are assigned based on these bins where the highest-scoring bin receives letter grade “A” and the lowest receives “Z.” If a score could not be calculated for the criterion, a letter grade “Z” was assigned. A total score was then calculated by summing the numeric values associated with all

letter grades.

We use this same formula for 9 assessment criteria observed in our administrative data to construct a Quantitative Score. These criteria used are:

- Total migrant recruitment
- Professional recruitment
- Middle level and clerical recruitment
- Skilled recruitment
- Semi-skilled recruitment
- Unskilled recruitment
- Female domestic worker recruitment
- Dispute settlement ratio
- Business sustainability (new markets entered into)

Quantitative Score 2005-08 * Month-Year FE: In order to control for differential agency quality trends, we compute the average yearly Quantitative Score for an agency between 2005 and 2008 and then interact this average with month-year fixed effects.

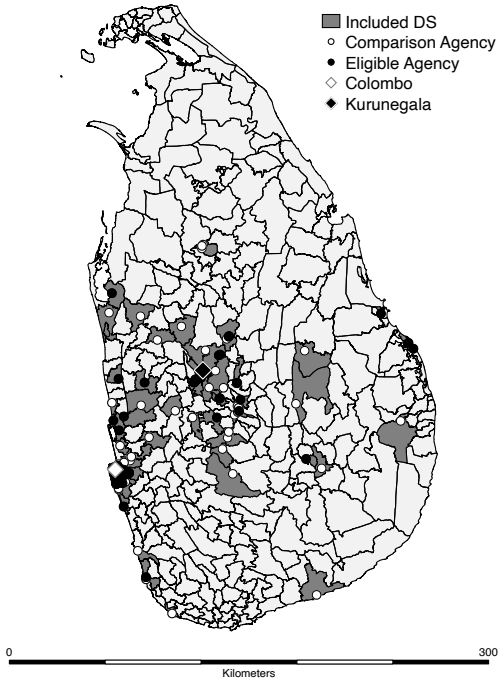
Quantitative Score 2009: In order to look at heterogeneity by pre-program agency quality, we instead use the Quantitative Score computed for 2009. Note, this is also the Quantitative Score used by the government to compute the final rating in 2012.

Predicted Rating: We construct a Predicted Rating using the fact that the government based the actual 2012 ratings on a Quantitative Score that only uses (pre-program) data from 2009 and third-party audit data collected after the program was introduced. We regress an (eligible) agency's actual continuous rating (i.e. their Combined Score) on their 2009 Quantitative Score. The fitted values from this regression are the 'Predicted Rating' for eligible agencies. In addition, we use the estimated coefficient from this regression to compute a 'Predicted Rating' for comparison agencies as well.

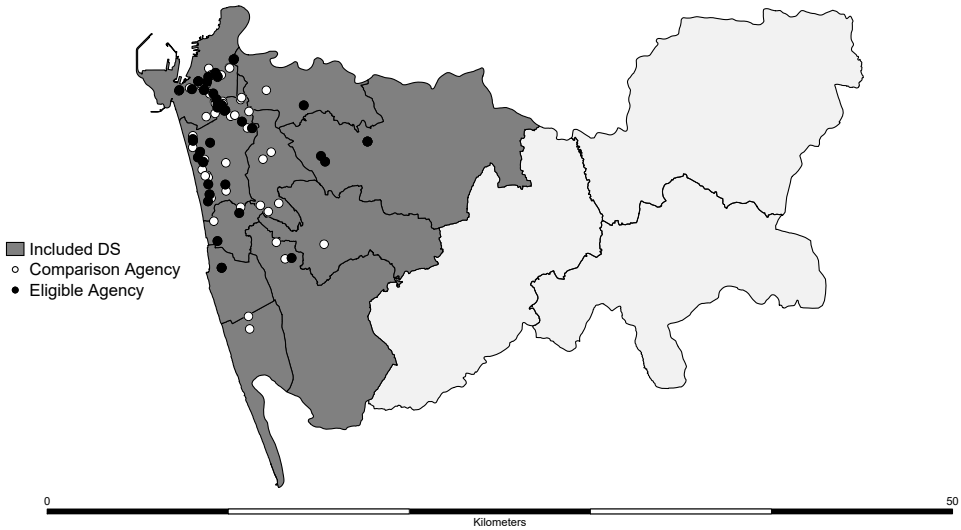
Monthly Quantitative Score: In Table 2, for example, we also use a 'Quant Score' variable to summarize investment in the rating criteria. We do so by computing the *monthly* Quantitative Score for each agency-month. We use the same formula described above, except at a month-level, to compute this score.

Appendix A.14 Agency Locations

Panel A: Agency Locations Across Sri Lanka

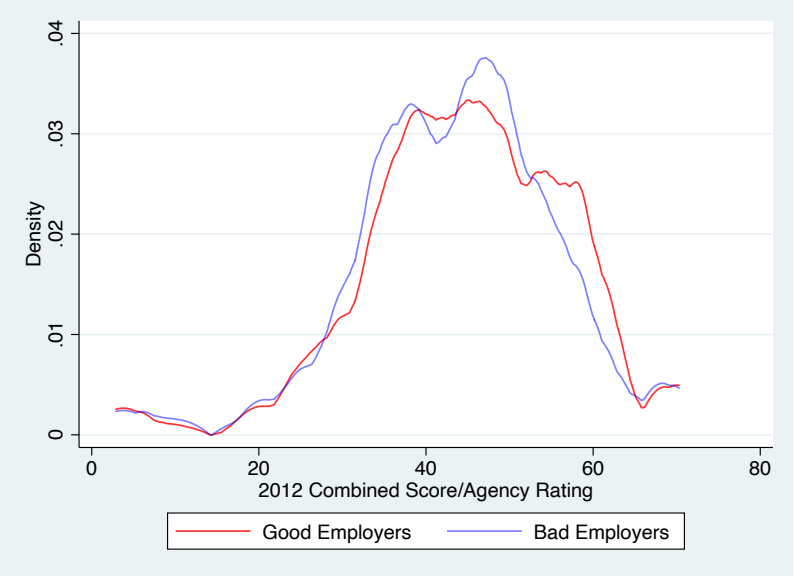


Panel B: Agency Locations Across Colombo District



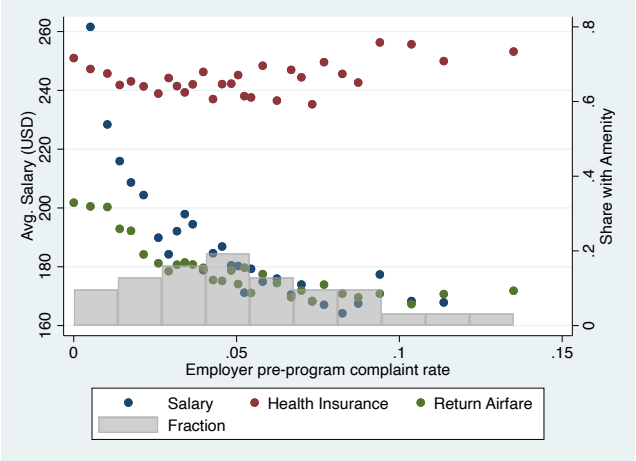
Notes: Panel A shows the location of eligible and comparison agencies included in the study across Sri Lanka. Panel B shows the location of agencies in Colombo District only. 'DS' refers to 'Divisional Secretariat', an administrative unit comparable to a US county.

Appendix A.15 Pre-Program Employer to Agency Matches

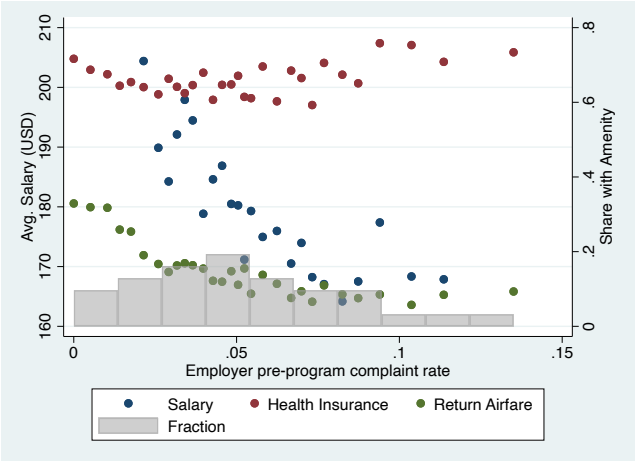


Notes: This figure plots two probability density functions that describe the pre-program matches between recruitment agencies, 'Good Employers' and 'Bad Employers': defined as employers who are below or above the median pre-program (< 2010) complaint rate, respectively. The domain is the continuous Combined Score/2012 agency rating which is only known to the public after 2012.

Appendix A.16 Wages and Amenities by Employer Complaint Rates



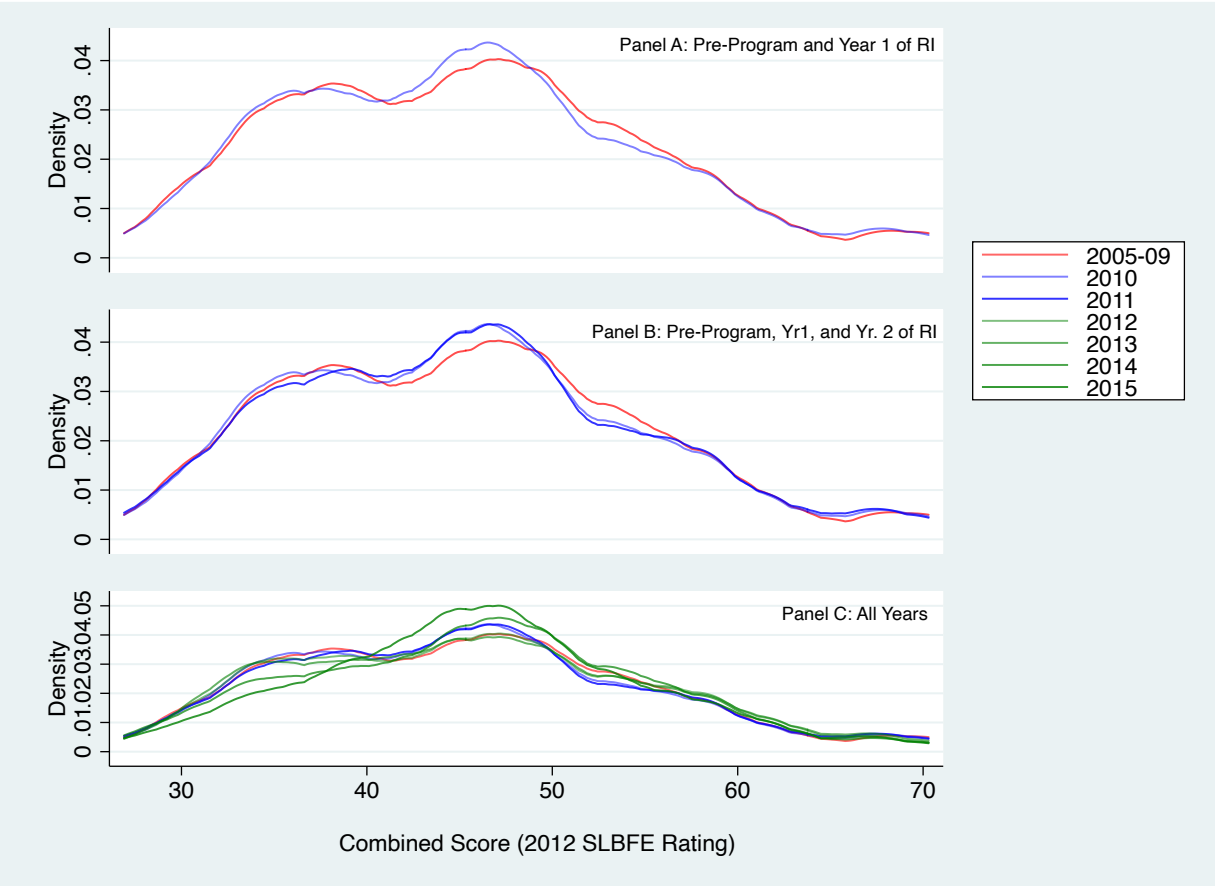
(a) Panel A: All Employers



(b) Panel B: Employers & Agencies Specializing in Domestic Work

Notes: This figure displays average salaries (2015 USD) and the share of amenities by pre-program employer complaint rates. The histogram shows the fraction of employers that fall into each complaint rate bin. Panel A restricts attention to employers who recruited at least 50 migrants prior to the program, which accounts for 94% of all pre-program recruitment. Panel B additionally focuses restricts attention to employers (and/or foreign agencies) for whom at least 75% of recruited migrants were domestic workers.

Appendix A.17 Matches to Bad Employers by Agency Quality



Notes: This figure plots a series of probability density functions that describe which recruitment agencies are matched to 'Bad Employers': defined as employers who have an above median pre-program ($<$ 2010) complaint rate. A match here is defined as whether an agency sends a migrant to a 'Good Employer' in a specific year. The domain is the continuous Combined Score/2012 agency rating which is only known to the public after 2012. Panel A plots the density function for matches occurring prior to the program (red line) and during the first year of the program (2010), i.e. the Reputational Incentives (RI) phase (the blue line). Panel B adds 2011, the second year of RI. Panel C plots the data for the remaining years (2012-2015) of the Quality Revelation (QR) period (the green lines).

Appendix A.18 Investment in the Audit Score

	Low Quality Agencies by 2009 Quantitative Score						High	Low
	4	5	6	7	8	9		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Quantitative Score (2011)	7	8.6	8.3	8.2	8.6	9	9.9	8.6
Audit Score	52	61.6	57.7	46.7	67.2	65.7	59.5	61.8
Record-Keeping	22	24.6	27	20.8	30.6	28.7	26.6	27.4
Marketing	22.5	26.8	20.2	18.7	26.6	27.4	23.6	24.9
Staff	0	1.2	0.2	0.5	1.1	1.5	0.9	1
Other	3	3	3.7	2.5	3.1	2.5	3.2	2.9
Combined Score	33.4	38.4	36.5	30.6	42.5	42.3	39.7	39.4
Stars	1	1.6	1.5	1.2	1.9	2.1	1.9	1.8
N(Agency Exit)	0	0	0	2	0	1	7	3
N	2	5	6	6	14	15	30	48

Notes: This table reports the mean for a series of variables for low-quality eligible agencies by their 2009 Quantitative score (columns 1-6), all high-quality eligible agencies (column 7) and all low-quality eligible agencies (column 8). 'Quantitative Score (2011)' is the computed quantitative score for 2011. The Audit score aggregates outcomes relating to management practices which were audited by third-party inspectors. We grouped outcomes into four sub-categories relating to Record-Keeping, Marketing, Staff, and Other. See Appendix A.2 for the outcomes included in each sub-category. The 'Combined Score' is a weighted combination of the 2009 Quantitative Score and the Audit Score, which was used to determine the star ratings; 'Stars' reports the average number of stars earned for each group of agencies. N(Agency Exit) reports the number of agencies which had exited the market (i.e. 12 consecutive months of zero migrant recruitment) at the end point our data (December, 2015).

Appendix A.19 Agency Exit Over Time

	Pre(2005-09)	Post(2010-15)	Post(2010-15) *Eligible*	Post(2010-15) *Comparison*
	(1)	(2)	(3)	(4)
High	22.9%	23.3%	23.3%	23.3%
Low	19.4%	17.7%	6.2%	23.2%
N	137	207	78	129

Notes: This table reports average agency exit rates prior to (2005-2009) and after (2010-2015) the introduction of the ratings program. Agency exit is defined as 12 consecutive months of zero migrant recruitment. In the pre-period, we define high and low-quality agencies as whether they are above or below the median Quantitative Score computed for 2005. In the post-period, we use the definition of agency quality used in the paper, i.e. above/below median Predicted Rating. The sample is restricted to agencies that recruited between 45 and 155 migrants in 2005 (col. 1) and 2009 (cols. 2-4). Column 3 further restricts attention to eligible agencies, while column 4 does likewise for comparison agencies.

Appendix A.20 A Model of Agency Reputations

We adapt a framework developed by Cabral (2005) to model how the announcement of the ratings program shapes the incentives and actions of recruitment agencies and their interactions with employers and migrants. The model has two periods, intended to mimic the structure of the SLBFE agency ratings program: the first period is the Reputational Incentives (RI) phase, while the second period is the Quality Revelation (QR) phase. In the model, low-quality agencies invest in placement quality in the first period because it positively influences employer beliefs about their type (i.e. their reputation) and raises demand for their services (i.e. job orders) in the future.

The model focuses on agency responses during the RI-phase for two reasons. First, our key results arise prior to quality revelation implying that behavioral responses to the announcement of the program are critical to interpreting our estimates. Second, agencies are the only actors who can *directly* respond to the program during the RI-phase as the star ratings are yet to be publicly revealed. While, in theory, employers may have general equilibrium responses to the program in the RI-phase, we do not find evidence to support such responses in Section C.10; likely owing to Sri Lanka's limited market share in international migration.

Setup

Assume there are *High* (H) type and *Low* (L) type agencies that facilitate matches between employers and migrants in *each* period in order to fulfill a job order (i.e. recruit a fixed number of migrants). While agency type is not observable, it is common knowledge that the share of H-type agencies in the market is θ, where $0 < \theta < 1$. Agencies differ in their ability to produce 'good' matches (or placements), which we define as those where a contract is not prematurely terminated by either party. Further, assume that H-type agencies always produce good placements, but L-type agencies only produce good placements with probability $p(e)$ and cost $c(e)$, which depend on their screening effort, e. We assume that $p(e)$ is decreasing and concave in effort and $p(0) > 0$, while $c(e)$ is increasing and convex in effort and $c(0) = 0$. We first define the demand for agencies, foreign employers, and migrants, and then specify the objective function of agencies.

Foreign Employers: The market consists of many foreign employers, who are of two types, *Good* (G) and *Bad* (B). These employers exist in each period, but only live for one period. We make this assumption in order to keep the model simple. While repeated interactions would allow a specific employer to learn about an agency's type, other employers would not benefit from this information. They would like to enlist the services of agencies to recruit migrant workers.

Good employers care about the quality of the agency they are contracting with, because turnover is costly for them. As such, they value good placement quality. However, since an agency's type and effort is private information, *Good* employers' demand for an agency's services, the number of job orders (q_G), is based on their prior about the likelihood of good placement quality $(Pr(g))$. It is given as follows:

$$q_G = Pr(g) = Pr(g|H)Pr(H) + Pr(g|L)Pr(L) = \theta + (1-\theta)p(e) \tag{3}$$

On the other hand, *Bad* employers do not care about placement quality as we assume turnover is not costly for them. Consequently, they also do not care about agency quality and offer a fixed level of demand, denoted by $\bar{q}_B$.

Migrants: Migrant workers would like foreign job placements with *Good* employers, but due to high search costs cannot coordinate with them directly. Instead, they must engage with local agencies. Migrants know that H-type agencies will always place them with *Good* employers, whereas L-type agencies will only place them with *Good* employers with fixed probability s, where $0 < s < 1$. We further assume that since migrants' relationship with agencies is mediated through an additional layer of sub-intermediaries, migrants never learn about agency type. This suggests that migrants' demand for agencies is always fixed and depends on their prior about the probability of being placed with a *Good* employer, which we denote as $\bar{s}$ and define as follows:

$$\bar{s} = Pr(G|H)Pr(H) + Pr(G|L)Pr(L) = \theta + (1-\theta)s \tag{4}$$

Agencies: Agencies live for both periods and, in each period, they produce matches between employers and migrants. The objective function for an agency is to maximize its total expected demand from employers and migrants across the two periods. This demand in each period is the sum of employer demands ($q_G + \bar{q}_B$) multiplied by worker demand ($\bar{s}$).

While there are two types of agencies (H and L), our setup is designed to focus on the actions and incentives of the L-type agency. In particular, in each period, an L-type agency must decide whether to invest effort in screening to improve placement quality, g. However, as the game ends after two periods, we note that L-type agencies will never invest effort to screen in the second period. Rational employers realize this and would, therefore, like to avoid contracting the services of L-type agencies in the second period.

Thus, we now consider whether an L-type agency will invest effort in the first period under two information environments: (1) where *Good* employers do not learn about first period placement quality because of information frictions and (2) when *Good* employers learn about first period placement quality in the second period via a ratings program. We note that, by assumption, quality revelation only matters for *Good* employers, as *Bad* employers do not care about agency quality and migrants do not learn about quality. The equilibrium concept we use is a perfect Bayesian equilibrium, where the choice of first period effort by an L-type is a mutual best response given other agents' beliefs about their effort.

Equilibrium without the Ratings Program

In the presence of information frictions, first period placement quality remains private information between transacting parties. Consequently, in equilibrium, an L-type agency will not invest effort in the first period ($e^* = 0$), because it cannot influence *Good* employers beliefs about their type and, hence, second period demand. *Good* employers realize this and simply offer the agency a fixed level of demand ($\bar{q}_G$) in each period equal to the probability of good placement quality, as defined by equation 3.

As the demand from *Bad* employers ($\bar{q}_B$) and migrants ($\bar{s}$) is fixed, the total expected de-

mand, denoted by Q, for an L-type agency's services is therefore:

$$Q = 2\bar{s}(\bar{q}_G + \bar{q}_B) \tag{5}$$

Equilibrium with the Ratings Program

Now, the regulator commits to publicly revealing an agency's first period placement quality in the second period. Moreover, agencies are informed about this credible threat in the first period, as in the RI-phase of the ratings program. To derive the equilibrium under this information environment, we begin by assuming that $\bar{e}$ represents *Good* employers' beliefs about an L-type agency's first period effort. Consequently, first period demand from *Good* employers is q_{G1} job orders in the first period, where $q_{G1} = \theta + (1-\theta)p(\bar{e})$.

In the second period, however, employers can observe placement quality from the first period. If first period placement quality is bad $[1-p(e)]$, then the agency is revealed to be an L-type with certainty. It will therefore not receive demand in the second period from *Good* employers (as with certainty L-type agencies do not exert in the second period).

On the other hand, if *Good* employers observe good placement quality $[p(e)]$ in the first period, then they update their beliefs about the agency's type using Bayes' rule—since H-type agencies never shirk on effort, *Good* employers always prefer their services in the second period—where θ is the proportion of H-types:

$$Pr(H|g) = \frac{Pr(g|H)Pr(H)}{Pr(g)} = \frac{\theta}{\theta + (1-\theta)p(\bar{e})} \tag{6}$$

Then, the expected second period demand from *Good* employers is therefore given as follows:

$$q_{G2} = 1 \cdot p(e)Pr(H|g) + 0 \cdot [1-p(e)] = p(e)Pr(H|g) \tag{7}$$

Putting together the two periods, the L-type agency's total expected demand is therefore:

$$\begin{aligned} Q &= \bar{s}(q_{G1} + \bar{q}_B) + \bar{s}(q_{G2} + \bar{q}_B) - c(e) \\ &= 2\bar{s}\bar{q}_B + \bar{s}(q_{G1} + q_{G2}) - c(e) \end{aligned} \tag{8}$$

The L-type agency maximizes equation 8 with respect to e. It chooses first period screening effort, e, so as to equate its marginal benefit (increased second period demand from *Good* employers) with its marginal cost, $c'(e)$, yielding the following first order condition:

$$\left[\frac{\bar{s}\theta}{\theta + (1-\theta)p(\bar{e})}\right] = \frac{c'(e)}{p'(e)} \tag{9}$$

Given the assumptions on $c(e)$ and $p(e)$, the LHS of equation 9 is decreasing in $\bar{e}$ while the RHS is increasing in e. Consequently, there is a unique interior solution $e^* = \bar{e}$ such that an L-type agency's first period screening effort (e) is set equal to employers beliefs about its effort $(\bar{e})$.

Appendix A.21 Returns to the Program

We use the proxy for surplus (Total Salaries: i.e. the sum of all salaries for migrants placed by an agency in a month) to estimate the additional remittance income induced by the program. We estimate that the program increased Total Salaries by approximately 70% on average in both phases of the program (see Col. 1, Appendix Table C.4). Using the average of Total Salaries for eligible agencies prior to the program (USD 1,886 per month), this implies an annual increase in salaries amounting to USD 1,886 x 12 months x 0.7 = USD 15,842 per agency, per year, or USD 79,212 over the five years of the program (measured in 2015 USD). +

The causal effects we estimate are necessarily related to the agencies in our sample, which includes 78/307 eligible agencies. However, the remaining 229 eligible agencies are, by definition, much larger and the program may well have induced an even larger increase in salaries and remittances (especially because we do not find evidence to suggest there were negative spillovers). To be conservative, we assume the program induced an identical increase in Total Salaries (i.e. USD 79,212) for all eligible agencies, implying an increase of USD 24,318,084 for all eligible agencies. Estimates of the share of earnings sent home as remittances vary greatly (Yang, 2011) but perhaps the most relevant estimate for our purposes comes from a study of Keralan male migrants in Qatar, which estimates that migrants remit 42% of their income (Seshan and Yang, 2014). This yields a marginal change in remittance income induced by the program amounting to USD 10,213,595.

This does not take into account the costs of the program which are not publicly available. The key cost components of the program were: (1) having a full-time program manager, (2) the cost of the third-party audits of eligible agencies, (3) the cost of developing the website, (4) the prize giving ceremony. We use the costs of the agency survey we conducted to approximate these costs. We estimate it would cost USD 15 per agency (for 307 eligible agencies, this yields USD 4,605) to conduct the third-party audits, a further USD 5,000 for website development, USD 12,291 for an award ceremony for 500 guests, and a further USD 25,063 to pay wages for a full-time program manager for five years. Assuming 25% overhead yields a total cost of USD 46,959, which we round off to USD 50,000. Consequently, we estimate the program induced additional remittances (net of costs) amounting USD 10,163,595 over the five years of the program or approximately USD 2 million per year.

Appendix B Additional Results

Appendix B.1 Results from Agency Survey

Panel A: Learning about Employer Reputation

Source	N (1)	Percent (2)
Check if Blacklisted	83	76%
Consult Other Agencies	65	60%
Consult Returned Migrants	73	67%
Contact SL Consulate	48	44%
Contact SLBFE	46	42%

Panel B: Reasons for Rejecting Job Order

Source	N (1)	Percent (2)
Job Wasn't Safe	29	53%
Employer Reputation	17	31%
Foreign Agent Reputation	18	33%
Too Many Amenities	11	20%
Low Commission	7	13%
Hard to find Workers	19	35%

Notes: Panel A asks agencies how they learn about the reputation of an employer. Panel B reports the reasons an agency states for rejecting a job order in the past year—55 out of 109 agencies reported rejecting a job order. Both tables use data collected through the agency survey in 2019 (n=109).

Appendix B.2 Characteristics of Job Orders

	# Employers (1)	# New Employers (2)	# Good Orders (3)	Good Orders (%) (4)	# Bad Orders (5)	Bad Orders (%) (6)
Panel A: Effects by Program Phase						
Elig*2010-12	0.5900	-0.0236	0.0526**	0.0270*	0.0021	-0.0051
	(0.500)	(0.422)	(0.025)	(0.015)	(0.022)	(0.015)
Elig*2012-15	0.7266	0.4437	0.0351	0.0185	-0.0037	-0.0116
	(0.718)	(0.706)	(0.025)	(0.016)	(0.022)	(0.015)
Panel B: Effects by Agency Quality						
Elig*High*Post-2010	0.1205	-0.2765	0.0764*	0.0459*	-0.0377	-0.0287
	(1.087)	(1.182)	(0.045)	(0.027)	(0.031)	(0.019)
Elig*Low*Post-2010	1.0869*	0.2546	0.0498**	0.0245	0.0193	0.0016
	(0.652)	(0.527)	(0.025)	(0.016)	(0.025)	(0.018)
Panel C: Effects by Agency Quality and Program Phase						
Elig*High*2010-12	-0.0810	-0.4967	0.0614	0.0392	-0.0362	-0.0250
	(0.846)	(0.828)	(0.044)	(0.028)	(0.035)	(0.020)
Elig*High*2012-15	0.3159	-0.0660	0.0900*	0.0519*	-0.0388	-0.0321
	(1.512)	(1.606)	(0.052)	(0.030)	(0.031)	(0.021)
Elig*Low*2010-12	1.1489*	0.0386	0.0678**	0.0329*	0.0279	0.0077
	(0.592)	(0.417)	(0.029)	(0.019)	(0.027)	(0.020)
Elig*Low*2012-15	1.0393	0.4361	0.0348	0.0175	0.0124	-0.0033
	(0.776)	(0.668)	(0.026)	(0.018)	(0.028)	(0.021)
N	18365	18365	18365	18365	18365	18365
Comparison Mean	5.48	2.53	0.17	0.13	0.17	0.13

Notes: This table reports program effects on the characteristics of job orders. 'Comparison Mean' reports the pre-program (2005-2009) mean of the dependent variable for the comparison group. Panel A compares eligible agencies to comparison agencies for the period where agencies were solely exposed to Reputational Incentives (2010-2012) and the period after Quality Revelation (2012-2015). Panel B computes heterogeneous effects by pre-program agency quality; 'High' denotes above median pre-program quality, while 'Low' is below median pre-program quality, where quality is based on the 'Predicted Rating' variable. Panel C reports effects by program phase and agency quality. '# Employers' reports the number of unique employers with whom an agency did business, '# New Employers' refers to an employer with whom the agency had not done business prior to the introduction of the program. 'Good Orders' ('Bad Orders') are those from an employer with a below median (above median) pre-program complaint rate. All specifications cluster at the agency level and include agency FE. Panels B & C contain month-year FE interacted with avg. migrant recruitment (2005-08). Panel A contains the same controls as Panel B and month-year FE interacted with the avg. yearly quantitative score (2005-08).

Appendix B.3 Components of Contract Quality

	Return Ticket (1)	Health Insurance (2)	Accommodation (3)	Food (4)	Avg. Salary (USD) (5)
			Panel A: Effects by Program Phase		
Elig*2010-12	0.0009	0.0244	0.0207	0.0233	9.9086
	(0.017)	(0.024)	(0.024)	(0.026)	(9.104)
Elig*2012-15	0.0087	0.0226	0.0186	0.0125	9.7274
	(0.021)	(0.027)	(0.027)	(0.028)	(12.714)
			Panel B: Effects by Agency Quality		
Elig*High*Post-2010	-0.0291	-0.0195	-0.0143	-0.0026	1.4770
	(0.036)	(0.044)	(0.044)	(0.042)	(23.784)
Elig*Low*Post-2010	0.0289	0.0560**	0.0487*	0.0474	17.1335*
	(0.019)	(0.026)	(0.026)	(0.030)	(9.597)
N	18365	18365	18365	18365	18365
Comparison Mean	0.10	0.31	0.32	0.30	100.70

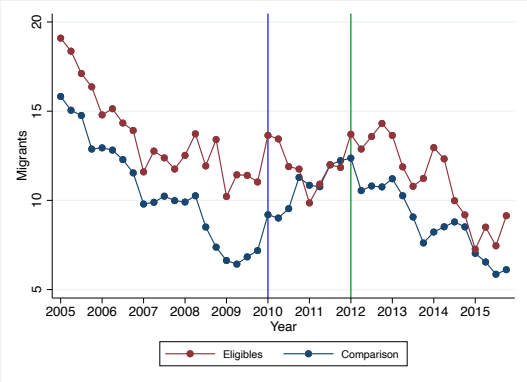
Notes: This table reports difference-in-difference estimates of effect of the program on individual components of the contract quality index. Panel A compares eligible agencies to comparison agencies. Panel B computes heterogeneous effects by pre-program agency quality; 'High' denotes above median pre-program quality, while 'Low' is below median pre-program quality, where quality is based on the 'Predicted Rating' variable. The outcome variables are binary variables that record whether a job order included the provision of return airfare (col. 1), health insurance (col. 2), accommodation (col. 3) and/or food (col. 4). The 'Avg. Salary' variable is either the midpoint of a salary range or the advertised salary (in 2015 USD) depending on what the job order included. All specifications cluster at the agency level and include agency fixed effects, month-year fixed effects interacted with average 2005-2008 migrant recruitment and month-year fixed effects interacted with the average quantitative score between 2005-2008.

Appendix B.4 Composition of Migrants Recruited

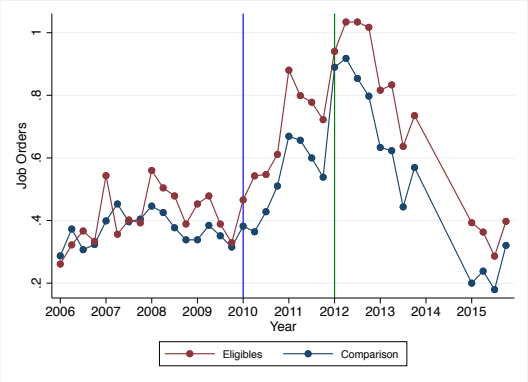
	% Female (1)	% Pre-Complaints (2)	% Under 30 (3)	% Over 50 (4)	% Saudi (5)
Panel A: Effects by Program Phase					
Elig*2010-12	0.0263	0.0054	0.0010	0.0115	-0.0238
	(0.031)	(0.004)	(0.011)	(0.013)	(0.025)
Elig*2012-15	-0.0160	0.0015	0.0050	0.0034	-0.0136
	(0.036)	(0.004)	(0.014)	(0.013)	(0.032)
Panel B: Effects by Agency Quality					
Elig*High*Post-2010	-0.0082	-0.0073	-0.0137	0.0103	-0.0403
	(0.061)	(0.006)	(0.019)	(0.018)	(0.044)
Elig*Low*Post-2010	0.0079	0.0057	0.0032	0.0065	0.0037
	(0.035)	(0.005)	(0.014)	(0.015)	(0.029)
N	21865	21865	21865	21865	21865
Comparison Mean	0.69	0.05	0.06	0.20	0.13

Notes: This table reports difference-in-difference estimates on the composition of migrants recruited by agencies. 'Comparison Mean' reports the pre-program (2005-2009) mean of the dependent variable for the comparison group. Panel A compares eligible agencies to comparison agencies both for the period where agencies were solely exposed to Reputational Incentives (2010-2012) and the period after Quality Revelation (2012-2015). Panel B computes heterogeneous effects by pre-program agency quality; 'High' denotes above median pre-program quality, while 'Low' is below median pre-program quality, where quality is based on the 'Predicted Rating' variable (See Appendix A20 for details). All variables condition on whether non-zero migrants were recruited by an agency in a given month. Column 1 reports the share of migrants who are women, column 2 reports the share who had previously (i.e. pre-program) made complaints, column 3 reports the share of migrants under the age of 30, column 4 reports the share over 50, and, finally, column 5 reports the share whose destination was Saudi Arabia. All specifications cluster at the agency level and include agency fixed effects. Panel B contains month-year fixed effects interacted with average migrant recruitment (2005-08). Panel A contains the same controls as Panel B and month-year fixed effects interacted with the average yearly quantitative score (2005-08).

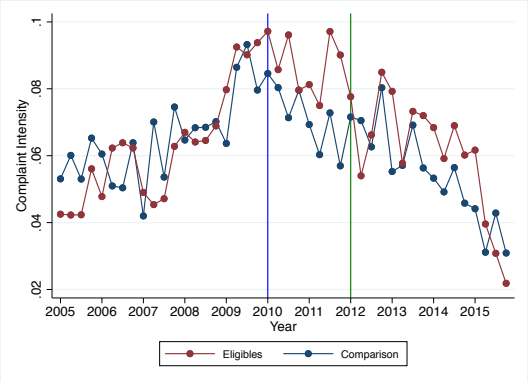
Appendix B.5 Raw Trends



Panel A: Migrant Recruitment



Panel B: Job Orders



Panel C: Complaint Intensity

Notes: These figures show quarterly averages in migrants recruited (Panel A), job orders received (Panel B), and the complaint rate (Panel C). The sample is restricted to agencies who recruit between 45 and 155 migrants in 2009 and who had a license that was valid for at least 1 year in 2009. The vertical blue line is the point prior at which agencies became aware of the program. The vertical green line shows when ratings were released to the public.

Appendix B.6 Components of Quantitative Score

	Professional (1)	Gen. Labor (2)	Middle (3)	Skilled (4)	Semi-Skilled (5)	Operators (6)	Domestic (7)	Solved Rate (8)
Panel A: Effects by Program Phase								
Elig*2010-12	-0.0076	-0.0821	-0.0951	-0.0769	-0.0289	0.6417	0.3538	0.0096
	(0.015)	(0.062)	(0.102)	(0.723)	(0.051)	(0.757)	(0.994)	(0.007)
Elig*2012-15	0.0133	0.6275	0.5709	0.8724	0.0496	1.8421	0.2613	-0.0184
	(0.015)	(0.669)	(0.546)	(0.666)	(0.044)	(1.210)	(0.901)	(0.019)
Panel B: Effects by Pre-Program Agency Quality								
Elig*High*Post-2010	0.0046	0.9968	1.1373	-0.5032	0.0463	2.5717	-0.5258	-0.0040
	(0.042)	(1.044)	(1.276)	(1.251)	(0.088)	(1.792)	(1.766)	(0.017)
Elig*Low*Post-2010	0.0057	-0.0225	-0.0236	1.1760**	0.0215	0.7765	0.9329	-0.0099
	(0.005)	(0.032)	(0.040)	(0.567)	(0.047)	(1.158)	(1.030)	(0.013)
Panel C: Effects by Agency Quality and Program Phase								
Elig*High*2010-12	-0.0199	-0.2919	-0.1554	-2.1400	-0.0847	-0.0286	-0.6396	0.0055
	(0.048)	(0.334)	(0.169)	(1.586)	(0.111)	(0.627)	(1.963)	(0.010)
Elig*High*2012-15	0.0184	1.7253	1.8680	0.4219	0.1204	4.0414	-0.4615	-0.0137
	(0.055)	(1.603)	(1.965)	(1.486)	(0.097)	(2.683)	(1.708)	(0.037)
Elig*Low*2010-12	0.0003	-0.0235	-0.0395	1.0342	-0.0050	0.8787	1.2434	0.0086
	(0.008)	(0.033)	(0.044)	(0.646)	(0.055)	(1.055)	(1.125)	(0.009)
Elig*Low*2012-15	0.0087*	-0.0219	-0.0145	1.2561**	0.0365	0.7188	0.7574	-0.0241
	(0.005)	(0.037)	(0.044)	(0.567)	(0.046)	(1.245)	(1.053)	(0.023)
N	27324	27324	27324	27324	27324	27324	27324	6759
Comparison Mean	0.03	0.08	0.12	1.62	0.11	1.73	3.94	1.00

Notes: This table reports difference-in-difference estimates on the components of the 'Quantitative Score' measure. 'Comparison Mean' reports the pre-program (2005-2009) mean of the dependent variable for the comparison group. Panel A compares eligible agencies to comparison agencies both for the period where agencies were solely exposed to Reputational Incentives (2010-2012) and the period after Quality Revelation (2012-2015). Panel B computes heterogeneous effects by pre-program agency quality; 'High' denotes above median pre-program quality, while 'Low' is below median pre-program quality, where quality is based on the 'Predicted Rating' variable. Columns 1-7 are count variables for migrants placed in varying skill categories. Column 8 reports estimated effects on the complaint solved rate (or dispute settlement rate) which conditions on the set of agencies who had non-zero complaints in a given month. All specifications cluster at the agency level and include agency fixed effects, month-year fixed effects interacted with average 2005-2008 migrant recruitment and month-year fixed effects interacted with average quantitative score 2005-2008.

Appendix B.7 The Effect of Receiving a Higher Rating

In order to separately identify the effect of receiving a higher rating, we instead use the sample of all eligible agencies and compare agencies that are just above and below the star-specific cutoffs. Appendix A.9 shows the distribution of ratings. The domain reports the Combined Score which is the continuous score that the *Quantitative Score* and *Audit Score* on which agencies were rated. In order to estimate the separate effect of receiving a higher rating, we argue that agencies on either side of the star-cutoff are similar aside from the fact that they are arbitrarily assigned to receive either more or less stars.

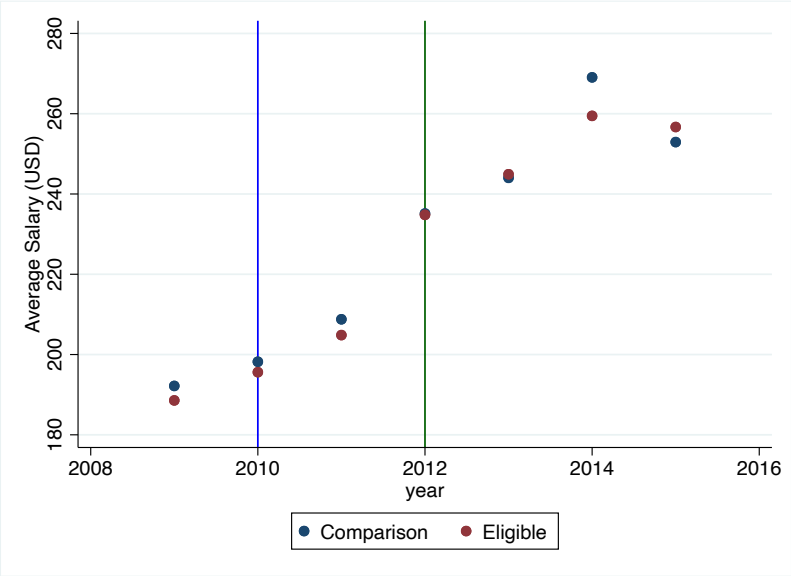
Empirical Strategy: Regression Discontinuity Design

To operationalize this identification strategy we use a regression discontinuity design. To maximize power, we center the treatment variable at the thresholds and assign a value of 1 if the score is above the threshold and 0 if it is below the threshold. For agency i in month m and year t:

$$Y_{imt} = \delta_t + \beta_1 Combined\ Score_i + \beta_2 Treat_i + \beta_3 Treat * Combined\ Score_i + \varepsilon_{imt}$$

In this regression, Y_{imt} is our outcome variable measured for agency i in month m and year t. $Treat_i$ is a dummy that takes a value of 1 when the score is above the star cutoff. $Combined\ Score_i$ is the centered forcing variable and $Treat * Combined\ Score_i$ is an interaction between the treatment and forcing variable. The interaction term allows the effect from the rating to vary on either side of the cutoff. We also include year fixed effects, δ_t, and cluster standard errors at the agency level. We use a uniform kernel and the mean-square error optimizing bandwidth. Note, in Table 4 we refer to $Treat_i$ as $Additional\ Star_i$.

Appendix B.8 Salaries Paid by Bad Employers



Notes: This figure displays average salaries (2015 USD) paid by 'Bad Employers'—those with an above median pre-program (< 2010) complaint rate—to eligible and comparison agencies. The vertical blue line is the point prior at which agencies became aware of the program. The vertical green line shows when ratings were released to the public.

Appendix B.9 Heterogeneous Effects for Low Quality Agencies

	Quant Score (1)	Migrants (2)	Job Orders (3)	Agency Exit (4)
	Panel A: Experience			
Elig*Post-2010	0.4823*	3.8052*	0.2929**	-0.1701**
	(0.283)	(1.993)	(0.126)	(0.066)
Exp*Elig*Post	-0.2565	-4.8573	-0.1695	0.1007
	(0.366)	(4.289)	(0.158)	(0.078)
	Panel B: Nearby Eligible Agency			
Elig*Post-2010	0.5141*	-1.6053	0.1346	-0.1297**
	(0.290)	(2.742)	(0.115)	(0.060)
Near*Elig*Post-2010	-0.1498	7.5156*	0.1689	0.0436
	(0.428)	(4.400)	(0.154)	(0.070)
	Panel C: Good Employers			
Elig*Post-2010	0.5600**	2.7201	0.1939*	-0.1807***
	(0.239)	(2.180)	(0.102)	(0.047)
Good*Elig*Post	-0.1305	-1.1438	0.0603	0.0935
	(0.314)	(3.274)	(0.162)	(0.061)
N	19404	19404	15876	19404

Notes: This table restricts attention to low quality agencies and examine treatment effect heterogeneity with respect to three characteristics: (1) Experience as measured by the number of years an agency has been operational (Panel A), (2) The presence of a nearby eligible agency within a 100 ft radius (Panel B), and (3) the number of (unique) Good Employers with whom the agency has worked (Panel C). In each case, we compute a pre-program measure of the characteristic and interact a standard difference-in-difference specification with an indicator variable that takes a value of one for an above median value of the characteristic. 'Quant Score' summarizes an agency's investment in the rating criteria introduced by the government (See Appendix A.13). 'Migrants' reports the total number of migrants recruited by an agency in a given month. 'Job Orders' are the number of orders that an agency received in a specific month from employers abroad (i.e. demand). Note, data for 2005 and 2014 is missing for the job order data. 'Agency Exit' is a proxy for an agency closing down. It is coded as an indicator variable that switches on at the beginning of a 12 month sequence of zero recruitment after the program was introduced. All specifications cluster at the agency level and include agency fixed effects.

Appendix B.10 Spillovers and Ethnic Networks

	(1) Migrants	(2) Agency Exit	(3) Job orders	(4) Good orders
		Panel A: 100 ft radius		
Any Elig*Post-2010	-1.9575	0.0084	0.2760	0.1070
	(7.679)	(0.059)	(0.251)	(0.070)
Co-Ethnic*Post-2010	11.8527	0.0095	0.0022	-0.0394
	(10.209)	(0.103)	(0.262)	(0.043)
		Panel B: 500 ft radius		
Any Elig*Post-2010	-0.2446	-0.0901	0.2261	0.0006
	(5.806)	(0.110)	(0.234)	(0.081)
Co-Ethnic*Post-2010	9.6858	0.0087	-0.0476	-0.0002
	(9.686)	(0.099)	(0.238)	(0.059)
		Panel C: 0.5 km radius		
Any Elig*Post-2010	0.7878	0.0516	-0.0320	-0.0046
	(3.868)	(0.193)	(0.186)	(0.055)
Co-Ethnic*Post-2010	1.0461	-0.1504	0.0518	-0.0091
	(3.661)	(0.109)	(0.202)	(0.047)
N	14916	14916	12204	10026

Notes: This table estimates spillover effects of the program and restricts attention to the sample of comparison agencies. Panel A reports estimates use a 100 foot radius, Panel B does likewise using a 500 foot radius, and Panel C uses a 0.5 km (0.3 miles) radius. For each of these radii, 'Any Elig' is a dummy variable coded as 1 if there is an eligible agency within the specified radius from a comparison agency. 'Co-Ethnic' is a dummy variable coded as 1 if a nearby eligible agency is owned by a co-ethnic. 'Agency Exit' is a proxy for an agency closing down. It is coded as an indicator variable that switches on at the beginning of a 12 month sequence of zero recruitment after the program was introduced. 'Job Orders' are the number of orders that an agency received in a specific month from employers abroad (i.e. demand). Note, data for 2005 and 2014 is missing for the job order data. 'Good Orders' are job orders corresponding to employers with a below median pre-program complaint rate. The estimates are clustered at the local market level and the specification includes agency fixed effects, month-year fixed effects interacted with average 2005-2008 migrant recruitment, month-year fixed effects interacted with average Quantitative Score 2005-2008 fixed effects, and local market x month-year fixed effects.

Appendix B.11 K-S Tests for Employer-Agency Matches

Year	Good Employers		Bad Employers	
	D	*p*-value	D	*p*-value
	(1)	(2)	(3)	(4)
2010	0.0230	0.952	0.0281	0.593
2011	0.0600	0.111	0.0306	0.583
2012	0.0735	0.060	0.0357	0.463
2013	0.0878	0.018	0.0491	0.231
2014	0.1319	0.001	0.0658	0.047
2015	0.1392	0.001	0.0985	0.003

Notes: The table above reports the test-statistic (column 1 & 3) and associated *p*-value (column 2 & 4) for a series of two-sample Kolmogorov-Smirnov tests that examine the equality of distributions. The goal of the tests is to assess whether the distribution of matches between Good Employers and recruitment agencies (columns 1 & 2) and Bad Employers and recruitment agencies (columns 3 & 4) changed in response to the ratings program. To do so, we compare density functions describing the likelihood of Good and Bad Employers matching with agencies by their 2012 Combined Score (i.e. the support). For each post-program year (i.e. 2010-2015) we test whether the density function for a specific year comes from the same distribution as the density function describing all pre-program years (i.e. < 2010), i.e. the density function which describes how Good and Bad Employers match to agencies prior to the introduction of the program.

Appendix C Robustness Tests

Appendix C.1 Pre-Trends

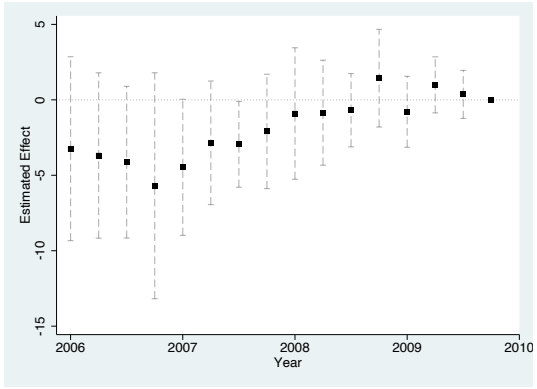
Variable	Sample (1)	F-Stat (2)	p-value (3)
Migrant	All	1.1246	0.3360
	Low	1.4589	0.1281
	High	0.8616	0.6081
Job Orders	All	1.1708	0.2968
	Low	1.2599	0.2350
	High	0.5676	0.8877
Complaint Rate	All	1.3032	0.2023
	Low	1.0045	0.4538
	High	1.3235	0.2178
Avg. Salary	All	0.8942	0.5715
	Low	1.2384	0.2498
	High	0.8743	0.5948
Contract Renewal	All	0.6557	0.8255
	Low	1.2464	0.2442
	High	1.2452	0.2661
Good Orders	All	1.1556	0.3093
	Low	1.9270	0.0248
	High	0.6672	0.8052
Contract Quality	All	0.9034	0.5611
	Low	0.6124	0.8613
	High	1.3662	0.1946

Notes: This table reports joint F-tests of j lags for several key outcomes analyzed in the paper. For agency i in month-year t, we run the specification:

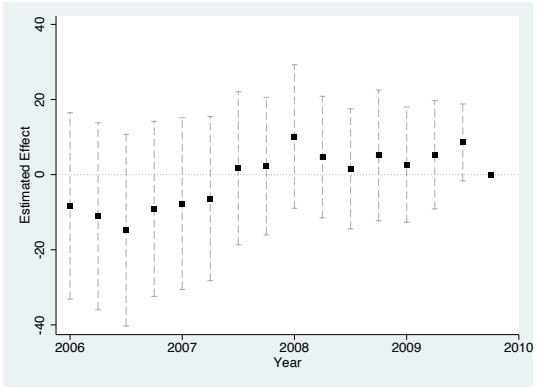
$$Y_{it} = \delta_t + \alpha_i + \sum_{k=1}^{j} \beta_k (Elig * \mathbb{1}[t = k])_{it} + \varepsilon_{it}$$

For each outcome, we separately report the F-test for all agencies, only pre-program low-quality agencies, and only pre-program high-quality agencies. In each case, we test the joint null for 15 lags (i.e. the sum of the β_k's, where each a period corresponds to a quarter) for the period 2006-2009 where the reference period is always the quarter immediately preceding Jan 2010. Standard errors are clustered at the agency level.

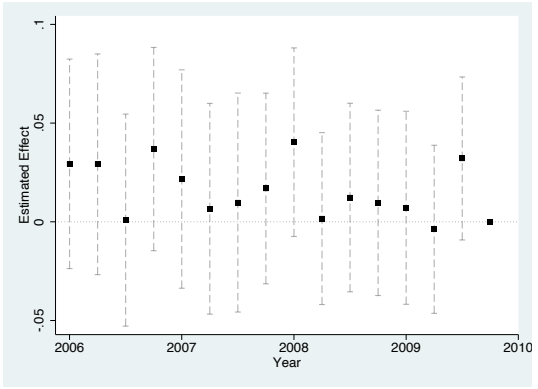
Appendix C.2 Plots of Lagged Effects



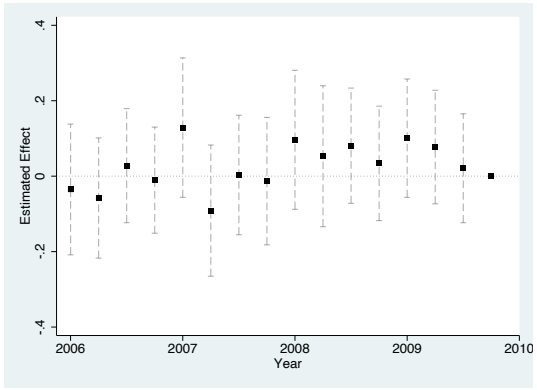
Panel A: Migrants Recruited



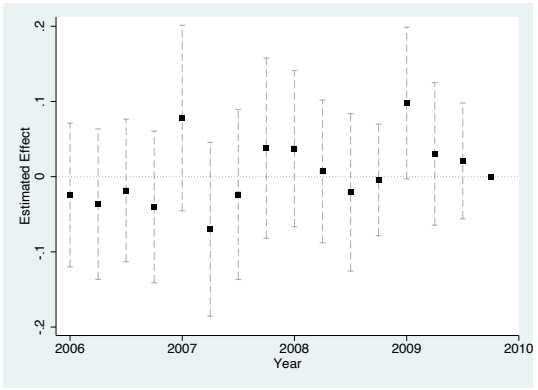
Panel B: Avg. Salary



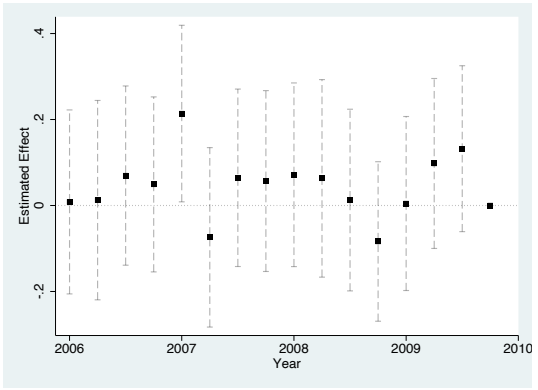
Panel C: Contract Renewal



Panel D: Job Orders



Panel E: Good Orders



Panel F: Contract Quality

Notes: These figures show event study graphs for lagged outcomes. They plot the interaction coefficients—an indicator for an eligible firm interacted with a series of quarterly indicators for 2006-2010. The capped dotted lines are the 95% confidence intervals. In each case the reference is the final quarter of 2009. Standard errors are clustered at the agency level.

Appendix C.3 Agency Exit

	Agency Exit (6 mo) (1)	Agency Exit (1 yr) (2)	Agency Exit (2 yr) (3)
Panel A: Effects by Program Phase			
Elig*2010-12	-0.0666***	-0.0701***	-0.0666***
	(0.026)	(0.026)	(0.026)
Elig*2012-15	-0.0791*	-0.0826*	-0.0741*
	(0.044)	(0.044)	(0.043)
Panel B: Effects by Pre-Program Agency Quality			
Elig*High*Post-2010	0.0013	0.0067	0.0133
	(0.072)	(0.074)	(0.073)
Elig*Low*Post-2010	-0.1196***	-0.1256***	-0.1173***
	(0.038)	(0.039)	(0.038)
N	27324	27324	27324
Comparison Mean	0.17	0.16	0.15

Notes: This table reports difference-in-difference estimates for agency exit. 'Comparison Mean' reports the post-program (2010-2015) mean of the dependent variable for the comparison group. Panel A compares eligible agencies to comparison agencies both for the period where agencies were solely exposed to Reputational Incentives (2010-2012) and the period after Quality Revelation (2012-2015). Panel B computes heterogeneous effects by pre-program agency quality; 'High' denotes above median pre-program quality, while 'Low' is below median pre-program quality, where quality is based on the 'Predicted Rating' variable. 'Agency Exit' is a proxy for an agency closing down. It is coded as an indicator variable that switches on at the beginning of a 6 month (column 1), 12 month (column 2) and 24 month (column 3) sequence of zero recruitment after the program was introduced. All specifications cluster at the agency level and include agency fixed effects. Panel B contains month-year fixed effects interacted with average migrant recruitment (2005-08). Panel A contains the same controls as Panel B and month-year fixed effects interacted with the average yearly quantitative score (2005-08).

Appendix C.4 Total Salaries

	Total Salaries (1)	Total Salaries (2)	Total Salaries (3)
Elig*2010-12	0.6958**		
	(0.289)		
Elig*2012-15	0.7053**		
	(0.341)		
Elig*High*Post-2010		0.7056	
		(0.537)	
Elig*Low*Post-2010		1.0911***	
		(0.396)	
Any Elig*Post-2010			0.9067
			(0.720)
N	25412	25412	14674
Comparison Mean	5.33	5.33	5.33

Notes: This table reports program effects on total salaries, a proxy for match surplus. Col. 3 restricts the sample to comparison agencies. 'Comparison Mean' reports the pre-program (2005-2009) mean of the dependent variable for the comparison group. 'Total Salaries' is computed by summing all salaries of migrants placed by an agency in the relevant month, and this variable is transformed using the inverse hyperbolic sine. The specification in col. 2 contains month-year fixed effects interacted with avg. migrant recruitment (2005-08), clusters at the agency level, and includes agency fixed effects. Col. 1 contains the same controls and month-year fixed effects interacted with the average yearly quantitative score (2005-08). 'Any Elig' is a dummy variable coded as 1 if there is an eligible agency within 100 ft of a comparison agency. The estimate in col. 3 includes the same controls as col. 1 but is clustered at the local market level and includes local market x month-year fixed effects.

Appendix C.5 Complaint Rate

	Complaint Rate (1)	% Breach (2)	% Harassment (3)
Panel A: Effects by Program Phase			
Elig*2010-12	0.006	-0.005	0.012
	(0.007)	(0.023)	(0.018)
Elig*2012-15	-0.002	0.051**	-0.020
	(0.006)	(0.024)	(0.019)
Panel B: Effects by Agency Quality			
Elig*High*Post-2010	-0.003	0.013	-0.005
	(0.008)	(0.040)	(0.035)
Elig*Low*Post-2010	0.004	0.025	-0.004
	(0.007)	(0.022)	(0.018)
Panel C: Effects by Agency Quality and Program Phase			
Elig*High*2010-12	0.003	-0.034	0.054
	(0.009)	(0.046)	(0.036)
Elig*High*2012-15	-0.008	0.060	-0.065
	(0.011)	(0.053)	(0.039)
Elig*Low*2010-12	0.008	-0.000	0.002
	(0.009)	(0.025)	(0.020)
Elig*Low*2012-15	0.001	0.048*	-0.009
	(0.007)	(0.026)	(0.022)
N	19659	8470	8470
Comparison Mean	0.08	0.40	0.14

Notes: This table reports difference-in-difference estimates on complaints. Panel A compares eligible agencies to comparison agencies. Panel B computes heterogeneous effects by pre-program agency quality; 'High' denotes above median pre-program quality, while 'Low' is below median pre-program quality, where quality is based on the 'Predicted Rating' variable. Panel C reports estimates by agency quality and program phase. The complaint rate measures the average number of complaints by migrants placed by an agency in a given month conditional on the agency placing non-zero migrants. % Breach reports the fraction of complaints that are due to breach of contract (including non-payment of wages) while % Harassment refers to the fraction of complaints that are due to physical harassment. Column 1 conditions on non-zero migrants recruited, while cols. 2 & 3 additionally condition on non-zero complaints. All specifications cluster at the agency level and include agency fixed effects. The specification reported in Panel B & C contain month-year fixed effects interacted with average migrant recruitment (2005-08). Panel A contains the same controls as Panel B and month-year fixed effects interacted with the average yearly quantitative score (2005-08).

Appendix C.6 Robustness to Agency Exit

	Quant Score (1)	Migrants (2)	Job Orders (3)
Panel A: Effects by Program Phase			
Elig*2010-12	0.2681	-0.2830	0.0760
	(0.211)	(1.633)	(0.061)
Elig*2012-15	0.2854	3.6344	0.1123
	(0.202)	(3.540)	(0.081)
Panel B: Effects by Pre-Program Agency Quality			
Elig*High*Post-2010	-0.2284	2.6620	-0.0296
	(0.279)	(5.906)	(0.119)
Elig*Low*Post-2010	0.5720***	1.6925	0.1793**
	(0.207)	(2.186)	(0.082)
N	25412	25412	20869
Comparison Mean	6.25	6.92	0.34

Notes: This table reports difference-in-difference estimates for main outcomes that are adjusted for agency exit. 'Quant Score' summarizes an agency's investment in the rating criteria introduced by the government (See Appendix A.13). 'Migrants' reports the total number of migrants recruited by an agency in a given month. 'Job Orders' are the number of orders that an agency received in a specific month from employers abroad (i.e. demand). Note, data for 2005 and 2014 is missing for the job order data. All specifications cluster at the agency level and include agency fixed effects. The specification reported in Panel B contains month-year fixed effects interacted with average migrant recruitment (2005-08). Panel A contains the same controls as Panel B and month-year fixed effects interacted with the average yearly quantitative score (2005-08).

Appendix C.7 Robustness to Trend Controls

	Quant Score (1)	Avg. Salary (USD) (2)	Job Orders (3)	Agency Exit (4)
		Panel A: Quality Trend Only		
Elig*2010-12	0.2357	12.5657	0.1074*	-0.0803***
	(0.154)	(8.765)	(0.061)	(0.028)
Elig*2012-15	0.0805	16.7593	0.1330*	-0.0895**
	(0.127)	(10.394)	(0.078)	(0.045)
		Panel B: Recruitment Trend Only		
Elig*2010-12	0.3426	12.1132	0.1121*	-0.0703***
	(0.209)	(8.810)	(0.063)	(0.026)
Elig*2012-15	0.2537	15.2904	0.1508*	-0.0846*
	(0.195)	(10.434)	(0.079)	(0.044)
		Panel C: No Controls		
Elig*2010-12	0.1254	12.3014	0.1020*	-0.0821***
	(0.222)	(8.732)	(0.061)	(0.029)
Elig*2012-15	-0.0585	16.6366	0.1247	-0.0889*
	(0.235)	(10.253)	(0.079)	(0.046)
N	27324	21865	22356	27324
Comparison Mean	5.80	178.18	0.34	0.17

Notes: This table assesses the robustness of the difference-in-difference estimates reported in the paper to varying combinations of trend controls. 'Comparison Mean' reports the pre-program (2005-2009) mean of the dependent variable for the comparison group. Panel A reports the estimates including only month-year fixed effects interacted with the average yearly quantitative score (2005-08). Panel B uses month-year fixed effects interacted with average migrant recruitment (2005-08). Panel C only includes month-year fixed effects. 'Quant Score' summarizes an agency's investment in the rating criteria introduced by the government (See Appendix A.13). 'Job Orders' are the number of orders that an agency received in a specific month from employers abroad (i.e. demand). Note, data for 2005 and 2014 is missing for the job order data. 'Agency Exit' is a proxy for an agency closing down. It is coded as an indicator variable that switches on at the beginning of a 12 month sequence of zero recruitment after the program was introduced. All specifications cluster at the agency level and include agency fixed effects.

Appendix C.8 Mean Reversion and Measures of Agency Quality

One concern with our measure of pre-program agency quality might be that it captures something transient as a result of measurement error and that agency quality may exhibit mean reversion thereafter. Put otherwise, it is not that the program improves outcomes for low-quality agencies, but that these were simply high-quality agencies that were improperly classified and who then trended back to their mean. We believe this concern is unlikely to explain our estimates for a number of reasons.

First, it is important to note that high and low-quality eligible agencies are always being compared to high and low-quality comparison agencies. There are 207 agencies in the sample, 78 eligible agencies (30 high-quality, 48 low-quality) and 129 comparison agencies (30 high-quality, 99 low-quality). As such, if there is measurement error in how we specify quality and, for example, a subset of low-quality eligible agencies are trending upward because of measurement error, this should also be the case for low-quality comparison agencies. The estimated heterogeneous treatment effects are the differential effect for high/low eligible agencies relative to high/low comparison agencies, whose trend is computed by the High*Post term in equation (2). As such, our estimates 'net out' any potential measurement error.

If the measurement error were itself classified with treatment assignment then it would bias our estimates; however this amounts to a test of pre-trends for pre-program agency quality. Using five years worth of (monthly) pre-program data, we find that we cannot reject the null for the joint F-test of the lags for the Quantitative Score variable (p-value $= 0.4807$). Though not a formal test, it is also useful to note that the point estimates in Table 2 & 3 do not suggest a systematic trend in the estimates for high and low agencies; in some cases the estimated effect is positive for low-quality agencies and negative for high-quality agencies, in others it is the opposite and in still others the effects are in the same direction. It is, therefore, surprising why some outcomes would revert to the mean but not others unless the measurement error were, for some reason, weighted towards specific outcomes.

Nevertheless, we take concerns about measurement error in pre-program agency quality seriously and in Appendix C.8 we assess the robustness of our main results to alternative measures of pre-program agency quality. In Panel A, we report our preferred measure, the Predicted Rating, while in Panel B we report heterogeneous treatment effect estimates by pre-program agency quality defined using the Quantitative Score from 2009. In Panel C we do likewise with the average yearly Quantitative Score between 2007 and 2009. The point estimates are similar in both magnitude and direction across the three measures of pre-program agency quality. We do, however, note that the precision of the estimates is somewhat lower when pre-program agency quality is measured using the average Quantitative Score between 2007-2009 (Panel C).

Our view here is that the Predicted Rating and the 2009 Quantitative Score yield agency quality measures that are more salient from the perspective of an eligible agency. As the program is announced in early 2010 and lays out the criteria for a rating that is forward looking, an agency that was high-quality three years ago but is low-quality in 2009 can't expect to receive a good rating in the future. In addition, we believe the Quantitative Score only captures a part of the regulator's ultimate rating which also depends on the 'Audit Score' which is in turn a function of third-party inspections. By using the Predicted Rating, we weight the Quantitative and Audit scores by how important they are for the actual rating, yielding a measure of quality that is consistent with the regulator's priorities.

Robustness of Estimates to Varying Measures of Pre-Program Agency Quality

	Quant Score (1)	Migrants (2)	Job Orders (3)	Agency Exit (4)	Good Orders (5)	Contract Quality (6)
			Panel A: Predicted Rating			
Elig*High*Post-2010	-0.3422	2.1827	-0.0105	0.0067	0.0764*	-0.0629
	(0.265)	(5.407)	(0.114)	(0.074)	(0.045)	(0.151)
Elig*Low*Post-2010	0.6574***	1.8596	0.2204***	-0.1256***	0.0498**	0.1669**
	(0.200)	(2.093)	(0.083)	(0.039)	(0.025)	(0.078)
			Panel B: Quantitative Score 2009			
Elig*High*Post-2010	0.0603	1.9859	0.1353	-0.0524	0.0686**	0.1168
	(0.261)	(3.951)	(0.099)	(0.056)	(0.034)	(0.109)
Elig*Low*Post-2010	0.6837***	2.4841	0.1627*	-0.1136**	0.0470	0.0810
	(0.220)	(1.941)	(0.087)	(0.052)	(0.030)	(0.093)
			Panel C: Avg. Quantitative Score 2007-2009			
Elig*High*Post-2010	0.2727	-2.0280	0.0929	-0.0618	0.0773**	0.1303
	(0.206)	(2.111)	(0.106)	(0.055)	(0.034)	(0.105)
Elig*Low*Post-2010	0.6494***	7.5747	0.1887**	-0.1052**	0.0281	0.0427
	(0.207)	(4.638)	(0.082)	(0.051)	(0.027)	(0.091)
N	27324	27324	22356	27324	18365	18365
Comparison Mean	5.80	7.61	0.34	0.16	0.17	0.31

Notes: This table reports heterogeneous difference-in-difference estimates using alternative definitions of pre-program agency quality. Panel A uses our preferred specification, i.e. the 'Predicted Rating'. Panel B uses the Quantitative Score from 2009, while Panel C uses the average Quantitative Score between 2007 and 2009 (see Appendix A.13 for details). 'Comparison Mean' reports the pre-program (2005-2009) mean of the dependent variable for the comparison group. 'Quant Score' summarizes an agency's investment in the rating criteria introduced by the government (See Appendix A.13). 'Migrants' reports the total number of migrants recruited by an agency in a given month. 'Job Orders' are the number of orders that an agency received in a specific month from employers abroad (i.e. demand). Note, data for 2005 and 2014 is missing for the job order data. 'Agency Exit' is a proxy for an agency closing down. It is coded as an indicator variable that switches on at the beginning of a 12 month sequence of zero recruitment after the program was introduced. 'Good Orders' are job orders corresponding to employers with a below median pre-program complaint rate. 'Contract Quality' is a normalized index incorporating whether or not a job order specifies that health insurance, accommodation, and return airfare are provided as well as the salary range paid to workers. Index components are weighted by the inverse of the covariance matrix as in Anderson (2008). All specifications cluster at the agency level and include agency fixed effects and month-year fixed effects interacted with average 2005-2008 migrant recruitment.

Appendix C.9 Robustness of Spillover Effects

	Migrants (1)	Agency Exit (2)	Job orders (3)	Good orders (4)
	Panel A: 100 ft radius			
Any Elig*Post-2010	4.2168*	0.0016	0.2418	0.0821
	(2.275)	(0.059)	(0.147)	(0.061)
	Panel B: 500 ft radius			
Any Elig*Post-2010	4.6292*	-0.1151*	0.1911	-0.0031
	(2.335)	(0.068)	(0.156)	(0.052)
	Panel C: 0.5 km radius			
Any Elig*Post-2010	1.6217	-0.0596	0.0046	-0.0240
	(1.821)	(0.129)	(0.068)	(0.031)
N	16104	16104	13176	10650

Notes: This table estimates spillover effects of the program and restricts attention to the sample of comparison agencies. Panel A reports estimates use a 100 foot radius, Panel B does likewise using a 500 foot radius, and Panel C uses a 0.5 km (0.3 miles) radius. For each of these radii, 'Any Elig' is a dummy variable coded as 1 if there is an eligible agency within the specified radius from a comparison agency. 'Agency Exit' is a proxy for an agency closing down. It is coded as an indicator variable that switches on at the beginning of a 12 month sequence of zero recruitment after the program was introduced. 'Job Orders' are the number of orders that an agency received in a specific month from employers abroad (i.e. demand). Note, data for 2005 and 2014 is missing for the job order data. 'Good Orders' are job orders corresponding to employers with a below median pre-program complaint rate. The estimates are clustered at the local market level and the specification includes agency fixed effects, month-year fixed effects interacted with average 2005-2008 migrant recruitment, month-year fixed effects interacted with average Quantitative Score 2005-2008 fixed effects, and local market x month-year fixed effects.

Appendix C.10 Do Bad Employers match with Comparison Agencies?

The spillover effects discussed in Section 6.1 are limited to those occurring when agencies are in close proximity to each other. In this section we consider two other types of spillover effects. First, if eligible agencies are reluctant to work with Bad Employers during the RI-phase, these employers may switch to working with comparison agencies. Second, after the QR-phase, Bad Employers may prefer either low rated agencies or those ineligible for the program if they believe these agencies are less concerned with reputational incentives. In addition, we note that spillovers may instead propagate through ethnicity-based or kin networks. However agencies clustered together also tend to be owned by co-ethnics, limiting variation that would allow for a test of this mechanism (see Appendix B.10). Overall, we concede that our ability to assess the effects of the program on employers is limited because we neither observe their exit and entry decisions nor the commissions they pay agencies.

Nevertheless, we offer three pieces of evidence that collectively serve to limit concern about such spillovers. First, we estimate the effect of the program on the characteristics of job orders in Appendix B.2. In Panel A, we do not find evidence suggesting that the program affected either the number (column 5) or share (column 6) of Bad Orders: i.e. those from employers with an above median pre-program complaint rate. in either phase of the program. Rather, the point estimates are an order of magnitude smaller than the analogous estimates for Good Orders (columns 3-4) and insignificant. Second, in Appendix B.8, we show that the average salaries paid by Bad Employers do not vary by eligible and comparison agencies either prior to or after quality revelation, which we view as supporting the hypothesis that Bad Employers have relatively elastic demand.

Finally, we plot density functions showing how the distribution of agencies matching with Good Employers varies by program phase. We find that Good Employers are more likely to match with highly rated agencies both prior to and after quality revelation (Panel B and C of Figure 3, respectively). When we construct analogous density functions for Bad Employers, we find little evidence to suggest they were more likely to match with lower rated agencies either prior to (Panel B) or after quality revelation (Panel C) in Appendix A.17. We compute two-sample tests for equality of distributions in Appendix B.11, which support the view that there is limited movement in matches between Bad Employers and agencies.

Appendix C.11 Program Effects with Alternative Samples

	Agency Exit (1)	Job orders (2)	Migrants (3)
Panel A: Recruited 45-155 Migrants in 2009			
Elig*2010-12	-0.0701***	0.1067*	0.1991
	(0.026)	(0.062)	(1.585)
Elig*2012-15	-0.0826*	0.1451*	3.3784
	(0.044)	(0.078)	(3.266)
N	27324	22356	27324
Panel B: Recruited 25-175 Migrants in 2009			
Elig*2010-12	-0.0521**	0.1453***	0.5659
	(0.022)	(0.055)	(1.306)
Elig*2012-15	-0.0856**	0.1897***	1.7608
	(0.039)	(0.069)	(2.787)
N	39204	32076	39204
Panel C: All Agencies			
Elig*2010-12	-0.0657***	0.2951***	2.1962
	(0.019)	(0.046)	(1.651)
Elig*2012-15	-0.1354***	0.3889***	-1.0664
	(0.030)	(0.061)	(2.242)
N	86064	70416	86064

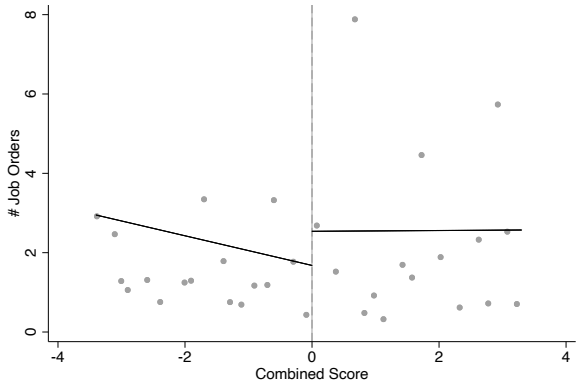
Notes: This table reports difference-in-difference estimates of the overall program effect using varying intervals about the 100 migrant program threshold. Panel A reports the 'optimal bandwidth' used in the paper (45-155 migrants recruited in 2009), Panel B restricts the sample to 25-175 migrants recruited in 2009 and Panel C places no restrictions on the sample. 'Agency Exit' is a proxy for an agency closing down. It is coded as an indicator variable that switches on at the beginning of a 12 month sequence of zero recruitment after the program was introduced. 'Job Orders' are the number of orders that an agency received in a specific month from employers abroad (i.e. demand). Note, data for 2005 and 2014 is missing for the job order data. All specifications cluster at the agency level and include agency fixed effects, month-year fixed effects interacted with average migrant recruitment (2005-08) and month-year fixed effects interacted with average yearly quantitative score (2005-08).

Appendix C.12 Robustness of Estimates to FBR Reform

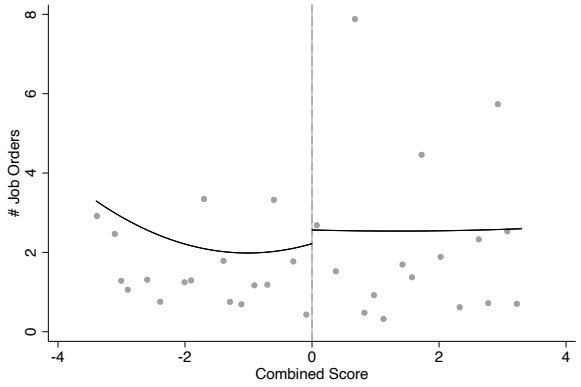
	Agency Exit (1)	Job orders (2)	Migrants (3)
	Panel A: Main Specification		
Elig*2010-12	-0.0701***	0.1067*	0.6913
	(0.026)	(0.062)	(1.133)
Elig*2012-15	-0.0826*	0.1451*	1.0862
	(0.044)	(0.078)	(1.445)
N	27324	22356	27324
	Panel B: Pre-July, 2013 Data		
Elig*2010-12	-0.0701***	0.1067*	0.6913
	(0.026)	(0.062)	(1.133)
Elig*2012-15	-0.0697*	0.1641	1.4457
	(0.041)	(0.114)	(1.990)
N	21114	18630	21114
	Panel C: Domestic Workers Trend		
Elig*2010-12	-0.0658**	0.1142*	0.4154
	(0.025)	(0.064)	(1.295)
Elig*2012-15	-0.0783*	0.1552*	0.9762
	(0.043)	(0.080)	(1.593)
N	27324	22356	27324

Notes: This table assesses whether difference-in-difference estimates of the overall program effect are affected by the Family Background Report (FBR) reform in July, 2013. Panel A reports our preferred specification using all data and Panel B drops data after July, 2013. Since the FBR reform targeted female domestic workers, in Panel C we assess whether agencies recruiting a larger share of domestic workers prior to the agency rating program were differentially affected. We do so by including a trend that interacts the pre-program share of migrants recruited who are domestic workers with month-year fixed effects. 'Agency Exit' is a proxy for an agency closing down. It is coded as an indicator variable that switches on at the beginning of a 12 month sequence of zero recruitment after the program was introduced. 'Job Orders' are the number of orders that an agency received in a specific month from employers abroad (i.e. demand). Note, data for 2005 and 2014 is missing for the job order data. All specifications cluster at the agency level and include agency fixed effects, month-year fixed effects interacted with average migrant recruitment (2005-08) and month-year fixed effects interacted with average yearly quantitative score (2005-08).

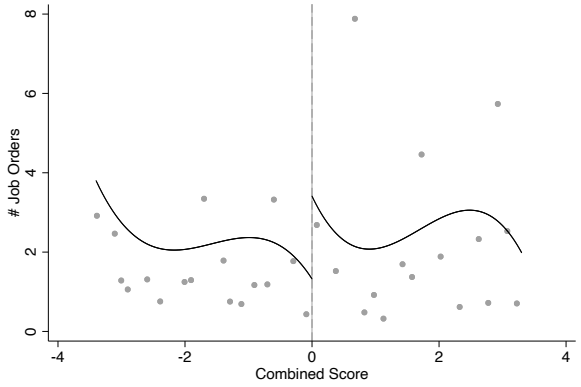
Appendix C.13 RD Plots for Job Orders



Panel A: First-Order Polynomial



Panel B: Second-Order Polynomial



Panel C: Third-Order Polynomial

Notes: This appendix shows regression discontinuity plots for the number of job orders. In each panel, the running variable is the centered Combined Score, across the star thresholds. We use a first-order polynomial in Panel A, a second-order polynomial in Panel B, and a third-order polynomial in Panel C. The sample consists of *all* agencies eligible for the rating program. The estimates pool across the star specific thresholds (see Appendix B.7 for details), use a uniform kernel, and use the MSE-optimal bandwidth using the rdrobust package. All specifications include year fixed effects and standard errors are clustered at the agency-level.

Appendix C.14 Robustness of RD Estimates

	Job Orders (1)	Migrants (2)	Domestic Share (3)	Pre-Complaints (4)
	Panel A: First-Order Polynomial			
Conventional	0.861*	3.788	-0.018	0.113
	(0.464)	(8.669)	(0.108)	(0.228)
Bias-corrected	1.006**	6.093	-0.009	0.164
	(0.464)	(8.669)	(0.108)	(0.228)
Robust	1.006*	6.093	-0.009	0.164
	(0.528)	(9.892)	(0.129)	(0.293)
	Panel B: Second-Order Polynomial			
Conventional	0.808	12.762	-0.031	0.026
	(0.545)	(11.023)	(0.122)	(0.325)
Bias-corrected	0.925*	14.512	-0.006	0.116
	(0.545)	(11.023)	(0.122)	(0.325)
Robust	0.925	14.512	-0.006	0.116
	(0.600)	(12.558)	(0.139)	(0.402)
	Panel C: Third-Order Polynomial			
Conventional	1.088*	9.951	-0.045	0.307
	(0.606)	(13.637)	(0.141)	(0.406)
Bias-corrected	1.257**	13.705	-0.054	0.522
	(0.606)	(13.637)	(0.141)	(0.406)
Robust	1.257*	13.705	-0.054	0.522
	(0.655)	(15.067)	(0.157)	(0.488)
N	20714	28060	28058	28060
Depvar. Mean	1.50	25.43	0.25	0.56

Notes: This table reports conventional, bias-corrected, and robust estimates the effect of receiving an additional star on the government's star rating program using a regression discontinuity design. We report estimates using a first-order polynomial in Panel A, a second-order polynomial in Panel B, and a third-order polynomial in Panel C. The sample consists of *all* agencies eligible for the rating program. The estimates pool across the star specific thresholds (see Appendix B.7 for details), use a uniform kernel, and use the MSE-optimal bandwidth using the rdrobust package. All specifications include year fixed effects and standard errors are clustered at the agency-level. Job Orders are the number of orders that an agency received in a specific month from employers abroad (i.e. demand). Migrants are the number of migrants placed by an agency in a given month. Domestic Share is the share of migrants who are domestic workers placed by the agency in a given month. Pre-Complaints is a count variable for the number of migrants who made a complaint prior to the program and subsequently migrated again after the program.