Pre-Analysis Plan: Delays, Corruption and Monitoring in government service provision

Martin Mattsson

August 13, 2018

Abstract

Countries with high levels of corruption also have slower government service delivery but it is not known if corruption causes delays, delays causes corruption or if the two are causally related at all. In this project I propose a model for how bureaucrats choose processing times and bribe demands to government service applicants. The models shows how, under certain information settings, the possibility of demanding bribes creates incentives for bureaucrats to create inefficiently long processing times for some applications.

I will test the predictions of this model using an experiment in the context of a particular government service in a particular setting, namely changes to land records in Bangladesh. The experiment will first test a management information system in the form of a monthly performance scorecard, making it visible to bureaucrats' managers if there are delays in the processing of applications for land record changes. The first question the experiment will answer is if the scorecard actually reduce delays in this government service. The second question is if the scorecard, having created an incentive to reduce the number of delayed applications, also reduces the amount of bribes paid by applicants. My model generates different predictions for how the bribe payments change under different information settings. Hence the experiment will not only test the model but also test the information setting under which bureaucrats operate.

In this pre-analysis plan I will describe how I will analyze the data from the experiment evaluating the effect of Performance Scorecards. I will motivate the analysis by first describing the theory that generates the predictions that the RCT will test.

Contents

1 Introduction

The stereotypical view of government service provision in low-income countries is that they are fraught with delays and that “speed money” is often required to get any services at all.[1] While this is not true for all services in all low-income countries, the prevalence of corruption and the number of days it takes to receive government services are positively correlated in cross country data. However, it has not been established if corruption causes slow government service delivery.

The experiment described in this pre-analysis plan is designed to test a specific model of how corrupt government institutions may cause delays in service delivery by providing an incentive to bureaucrats to intentionally slow down applications from applicants who do not pay bribes. The intervention evaluated in the proposed experiment is a monthly Performance Scorecard sent to bureaucrats and their superiors. The scorecard is designed to improve the superiors monitoring ability as well as "nudging" the bureaucrats to perform better and thereby creating an incentive for bureaucrats to reduce delays in service delivery to citizens. The prediction of the model is that this intervention will not only speed up service delivery but also change who pays what bribe payments.

The context in which we are conducting this study are local offices of the Ministry of Land in Bangladesh, Upazila Land Offices (ULOs). These offices perform the important service of changing the official land records every time a plot of land is sold and then provide the new owner with a Record of Rights (RoR) of the plot of land. This process is called a land record “mutation” and it is possible to implement the monitoring system due to the recent introduction of a digital system known as the “eMutation system”. The eMutation system has been developed and is being implemented across all of Bangladesh by a2i (www.a2i.pmo.gov.bd), an agency situated within the Prime Minister’s Office that is the main implementation partner for the research project alongside the Land Reforms Board of Bangladesh who oversees the ULOs.

2 Theory

2.1 Introduction to the model

This is a static model where a large number of applicants apply for the same service to one government bureaucrat at the same time. The bureaucrat can ask for different bribe payments from the applicants and can offer the service with different processing times or refuse to provide the service. Once a processing time and bribe payment is agreed upon the applicant pays the bribe and the bureaucrat must honor the agreement. Two types of agents in this model, bureaucrats and applicants, are described in detail below. The model is adopted from an asymmetric information model of price discrimination under monopoly described in Bolton et al. (2005).

The model presented here makes use of several simplifying assumptions while still generating the main predictions that will be tested in the experiment. In particular I will solve the model for two types of agents and a linear disutility of labor for the government bureaucrat. A more general model where these assumptions are relaxed will be presented in the paper reporting the results of this study.

The performance scorecard intervention is modeled as an additional benefit to the bureaucrat for each application that is processed within the time limit. The predictions of the model with regards to the experiment will come from moving this benefit from zero to a positive value.

[1] This is also the starting point of Banerjee (1997).

2.1.1 Bureaucrat

The bureaucrat maximizes her utility, $U^G(L, B)$, which is a function of total bribe income B and the total amount of labor L. When the scorecards are being sent out, i.e. in the treatment group after the start of the intervention, the bureaucrat also takes into account the future career benefits from processing applications within a government mandated time limit.

To make the model tractable I will make several simplifying assumption about this utility function. First I will give it a specific functional form:

$$U^G(L, B) = B - dL + \chi \sum_{i=1}^{A} \mathbf{1}(t_i \leq t_{limit})$$

Where A is the number of potential applicants and each applicant is indexed by i. b_i is the bribe paid by applicant i and B is total bribe income $B = \sum_{i=1}^{A} b_i$. l_i is the labor the bureaucrat spent on processing applicant i's application and L is the total amount of labor $L = \sum_{i=1}^{A} l_i$. t_i is the applicant i's processing time and $\mathbf{1}(t_i \leq t_{limit})$ is an indicator for if t_i is below or equal to t_{limit}. χ is the future career benefit the bureaucrat receives for processing an application within the time limit t_{limit} when the scorecard intervention is active. We can think of χ as the net present value of an improvement of the bureaucrat's career prospects as a result of the scorecard showing one more application processed within the time limit. For convenience I will assume that χ is paid for by a lump-sum tax on all the applicants so that setting χ higher or lower only affects total welfare through how it affects the allocation of bureaucrat labor.

There are a few specific assumptions made in this functional form. First, the bureaucrat is risk neutral. Second, the bribes enter linearly without a coefficient, in other words one unit of money generates on unit of utility. Third, the bureaucrat can costlessly decline the application without a valid reason or simply never process it. Forth, it is risk free and cost less for the bureaucrat to ask for bribes. Fifth, the bureaucrat does not have an outside option, a more realistic way of thinking about this is that holding the government position is so attractive to the bureaucrat that she would never want to give it up. The model can be set up so that the fifth assumption always holds, without any changes to the results, by having the applicant's pay a lump-sum tax that pays for a fixed wage to the bureaucrat that always makes the government position more attractive than the second best alternative.

The government service production function The time it takes for applicant i to get the service is $t_i = G(l_i)$ where l_i is the labor spent on processing the application from an applicant i. $G'(l) < 0$ so more labor mean shorter processing times. Furthermore $G''(l) \geq 0$ so that there are "weakly diminishing returns" to bureaucrat labor in making the processing time shorter. It does not take any effort for the bureaucrat to not process the application.

2.1.2 Applicants

Applicants derive utility from having their application processed, this value is discounted by the time it takes to process the application. Similar to the bureaucrat's case above, the bribe payment enters the applicant's utility function linearly and with a negative coefficient equal to one so that one unit of money equals one unit of utility. Applicants only differ in terms of their valuation of the government service.

The applicants' utility function takes the following functional form:

$$U^A(\theta_i, b_i, t_i) = -b_i + v(t_i)\,\theta_i$$

Where θ_i is the valuation of the service, b_i is the bribe paid and t_i is the time it takes to get the service.θ takes the value θ_H for a proportion β of the population, the "high types". While it takes the value θ_L, which is lower than θ_H, for the rest of the population, the "low types". Here I omit the lump-sum tax that is used to raise funds for the bureaucrats future career benefits for simplicity.

Regarding the discounting function $v(t_i)$ I will assume that $v'(t_i) < 0$, i.e. the quicker the applicant gets the service the less discounted it is. Furthermore, I will assume that $G''(l_i^*) > -\frac{v''(G(l_i^*))}{v'(G(l_i^*))}$, this is because if $v''(G(l_i^*))$ is positive and too large we will end up with a situation where higher valuation applicants should receive less bureaucrat labor in the Pareto optimal case, this goes against the situation I want the model to describe.

Spence-Mirrlees condition I will assume that the Spence-Mirrlees single crossing condition holds which in this model implies that:

$$\frac{\partial\left[\frac{\frac{\partial U^A(\theta_i,b_i,t_i)}{\partial t_i}}{\frac{\partial U^A(\theta_i,b_i,t_i)}{\partial b_i}}\right]}{\partial\theta_i} < 0$$

Since $\frac{\partial U^A(\theta_i,b_i,t_i)}{\partial b_i} = 1$ it is sufficient that $\frac{\partial^2 U^A(\theta_i,b_i,t_i)}{\partial t_i \partial\theta_i} < 0$. In words, as the valuation increases the marginal utility of getting the service quicker also increases (the marginal cost of longer processing time increases).

2.2 Condition for Pareto optimality

Given that the functional forms of the utility functions, one dollar in the hand of an applicant is valued the same as one dollar in the hand of the bureaucrat, hence all Pareto optimal solutions maximize the total surplus in the economy, i.e. the value generated by all the applications. I will start by simplifying the utility functions by removing the terms that are pure transfers since these cancel out when calculating the total surplus.

$\chi \sum_{i=1}^{A} \mathbf{1}(t_i \leq t_{limit})$ is simply a transfer funded by the lump-sum tax on applicants and hence I drop it from the welfare calculation. Bribes are also simply transfers, I drop them from the utility functions and I am left with $U^G(L)$ and $U^A(\theta, t(\theta))$. Finally, I will assume that all applicants of the same type receive the same service so that there are only two service levels t_H and t_L.

All Pareto optimal solutions will maximize the total surplus generated by the application processing. Formally:

$$max\left\{U^G(L,B) + \beta U^A(\theta_H, t_H) + (1-\beta)\,U^A(\theta_L, t_L)\right\} \tag{1}$$

Using the functional form assumptions, specifically the assumption that the disutility of labor is linear we can rewrite this at the individual applicant level as:

$$max\left\{-dl_i + v(G(l_i))\,\theta_i\right\}$$

Using unconstrained maximization for the individual case we get that for each applicant i, FOC (w.r.t. l_i):

$$v'\left(G\left(l_i^*\right)\right) G'\left(l_i^*\right) \theta_i = d \tag{2}$$

This can be rearranged as:

$$v'\left(G\left(l_i^*\right)\right) = \frac{d}{G'\left(l_i^*\right) \theta_i} \tag{3}$$

Equation 2 shows that the marginal willingness to pay for bureaucrat labor by the applicant has to be equal to the marginal willingness to accept bribes to work more by the bureaucrat, divided by the marginal improvement in terms of time one unit of additional bureaucrat labor provides. We can derive the comparative static:

$$\frac{\partial l_i}{\partial \theta_i} = -\frac{\left(v'\left(G\left(l_i^*\right)\right)\right)^2}{d\left[G''\left(l_i^*\right) v'\left(G\left(l_i^*\right)\right) + v''\left(G\left(l_i^*\right)\right)\right]} > 0 \tag{4}$$

Where the inequality comes from the assumption that $G''\left(l_i^*\right) > -\frac{v''\left(G\left(l_i^*\right)\right)}{v'\left(G\left(l_i^*\right)\right)}$. Equation 4 shows that types that value the service more will receive more bureaucrat labor in the Pareto optimal allocation.

Equation 2 characterize the unique Pareto optimal allocation of bureaucrat labor for the individuals who receive the government service in the Pareto optimal allocation, i.e. for all applicants for whom $l_i^* > 0$. However, only applicants for whom $v\left(G\left(l_i^*\right)\right)\theta_i - l_i^* d > 0$ receives any service in the Pareto optimal solution. The cutoff for when an applicant would receive any service in the Pareto optimal case can be written as:

$$\theta^{cutoff} = \frac{dl^*}{v\left(G\left(l^*\right)\right)}$$

I assume that $\theta^{cutoff} < \theta_L < \theta_H$ so that all applicants receive the service in the Pareto optimal allocation.

2.2.1 Graphical exposition of model

In the Figure 1 below I show a set-up graphical exposition of the model designed to aid intuition, I will use this type of graph throughout the paper.

Figure 2 below shows the Pareto optimal allocation for L-types and H-types graphically.

2.3 Full information and asymmetric information equilibria without performance scorecards

2.3.1 Full information equilibria

One potential equilibria in this model is one where the bureaucrat can perfectly observe the applicants willingness to pay. Using this information the bureaucrat can offer different bribe-processing time combinations for each applicant that maximize the overall surplus and ensures that the bureaucrat captures all of this surplus. This is the analogue of a perfectly price discriminating monopolist that captures all of the consumer surplus.

One way to reach this equilibrium is for the bureaucrat to sets an individual bribe-processing time schedule of $b_i = v\left(G\left(l_i\right)\right)\theta_i$ for each applicant. The applicant will chose the bribe-processing time combination that

Figure 1: Graph set-up

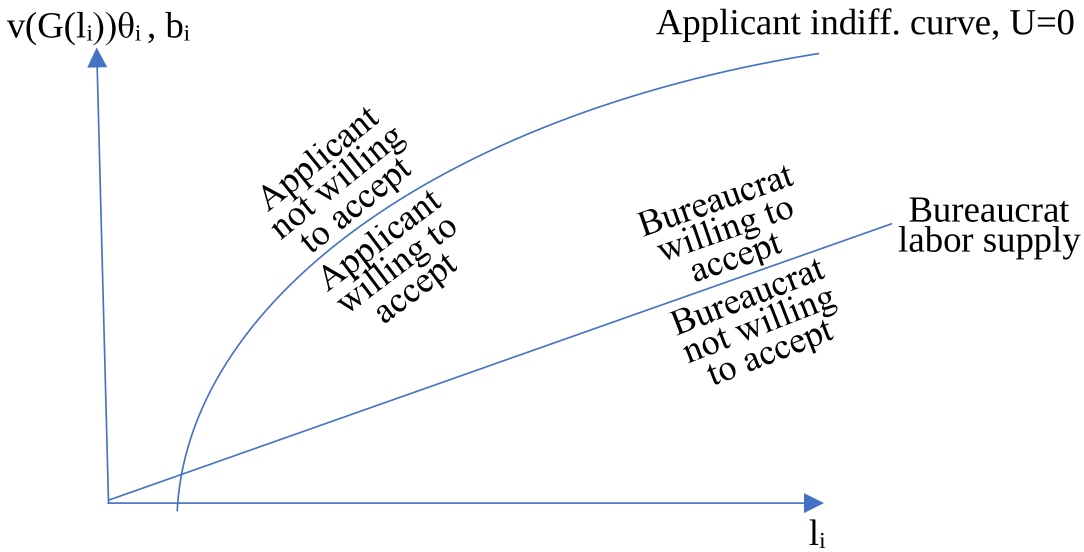


Figure 2: Pareto optimal allocation

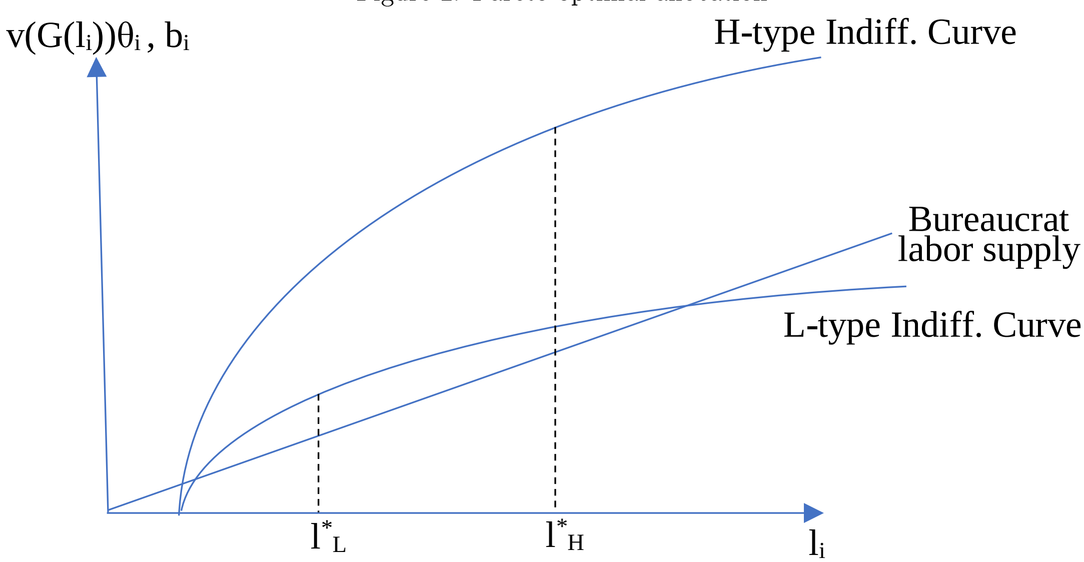
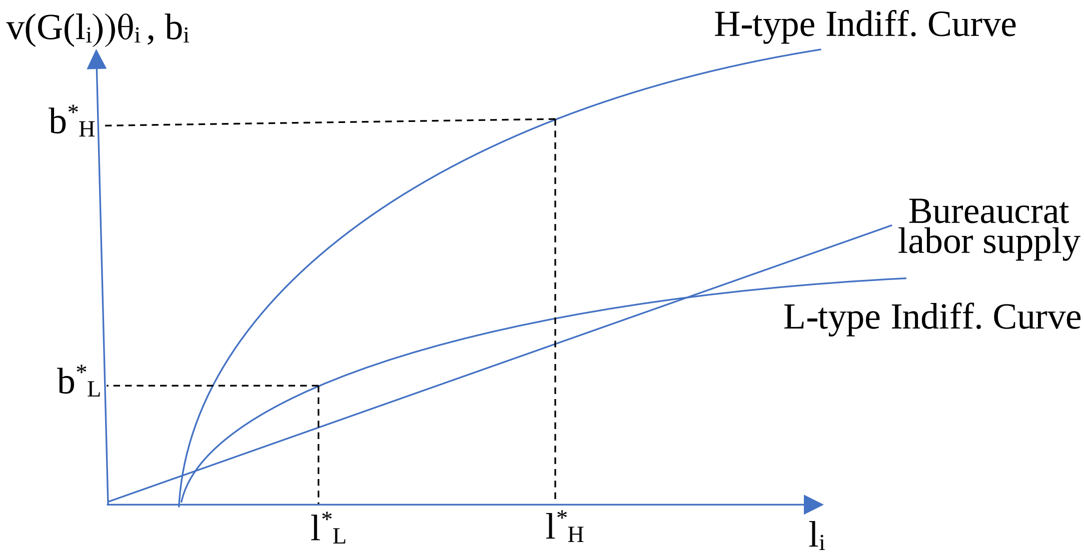


Figure 3: Full information equilibrium



maximizes his utility but since the schedule is set to capture all the surplus the applicant will end up with zero utility. The Pareto optimal allocation of bureaucrat labor would be achieved since this is the level of labor maximizing the surplus that the bureaucrat can obtain.

2.3.2 Asymmetric information

If the government does not know what type each individual applicant is, but does know the distribution of applicants she can offer the same "menu" of bribe-processing time combinations to all applicants. Government bureaucrats can set a non-linear price schedule $b(t)$ but applying the revelation principle we can simplify the bureaucrats problem to setting two sets of price-bribe combinations: $\{(t_H, b_H);(t_L, b_L)\}$

$$max_{\{t_H,b_H\},\{t_L,b_L\}}\left\{\beta b_H + (1-\beta)\, b_L - D\left(\beta G^{-1}(t_H) + (1-\beta)\, G^{-1}(t_L)\right)\right\}$$

Subject to:

- IR_H: Individual Rationality constraint H-types: $IR_H : -b_H + v(t_H)\,\theta_H \geq 0$
- IR_L: Individual Rationality constraint L-types: $IR_L : -b_L + v(t_L)\,\theta_L \geq 0$
- IC_H: Incentive Compatibility constraint H-types: $IC_H : -b_H + v(t_H)\,\theta_H \geq -b_L + v(t_L)\,\theta_H$
- IC_L: Incentive Compatibility constraint L-types: $IR_L : -b_L + v(t_L)\,\theta_L \geq -b_H + v(t_H)\,\theta_L$

Solution Assume that only IR_L and IC_H binds (this can be confirmed later):

$$IR_L : -b_L + v(t_L)\,\theta_L = 0 \tag{5}$$

$$IC_H : -b_H + v(t_H)\,\theta_H = -b_L + v(t_L)\,\theta_H \tag{6}$$

Sub IR_L into IC_H to get:

$$-b_H + v(t_H)\theta_H = -v(t_L)\theta_L + v(t_L)\theta_H \Rightarrow b_H = v(t_H)\theta_H - v(t_L)(\theta_H - \theta_L)$$

Eliminate b_H and b_L and rewrite the maximization problem:

$$max_{t_H,t_L}\left\{\beta\left[v(t_H)\theta_H - v(t_L)(\theta_H-\theta_L)\right] + (1-\beta)\left[v(t_L)\theta_L\right] - d\left(\beta G^{-1}(t_H) + (1-\beta)G^{-1}(t_L)\right)\right\}$$

This maximization problem generates the following first order conditions:

FOC 1 (w.r.t. t_H):

$$\beta\left[v'\left(t_H^{assym}\right)\theta_H\right] - \beta\frac{dG^{-1}(t_H)}{dt_H}d = 0$$

$$\Rightarrow v'\left(t_H^{assym}\right) = \frac{d}{G'\left(l_H^{assym}\right)\theta_H} \tag{7}$$

FOC 2 (w.r.t. t_L):

$$\beta\left[-v'\left(t_L^{assym}\right)(\theta_H-\theta_L)\right] + (1-\beta)\left[v'\left(t_L^{assym}\right)\theta_L\right] - (1-\beta)\frac{dG^{-1}(t_L)}{dt_L}d = 0$$

$$\Rightarrow v'\left(t_L^{assym}\right) = \frac{d}{G'\left(l_L^{assym}\right)\theta_L\left[1-\left(\frac{\beta}{1-\beta}\right)\left(\frac{\theta_H-\theta_L}{\theta_L}\right)\right]} \tag{8}$$

As can be seen in Equation 8 above the denominator can be negative if β is close to 1 or if θ_L is very small compared to θ_H. In this case L-types does not get any service. For the purpose of connecting this model to my research I will assume that this is not the case and that L-types value the service sufficiently and that the group of L-types if sufficiently large so that the bureaucrat provides them some service.

In terms of the bribes, IR_L implies that the L-types pay a bribe equal to their willingness to pay:

$$b_L^{assym} = v\left(t_L^{assym}\right)\theta_L \tag{9}$$

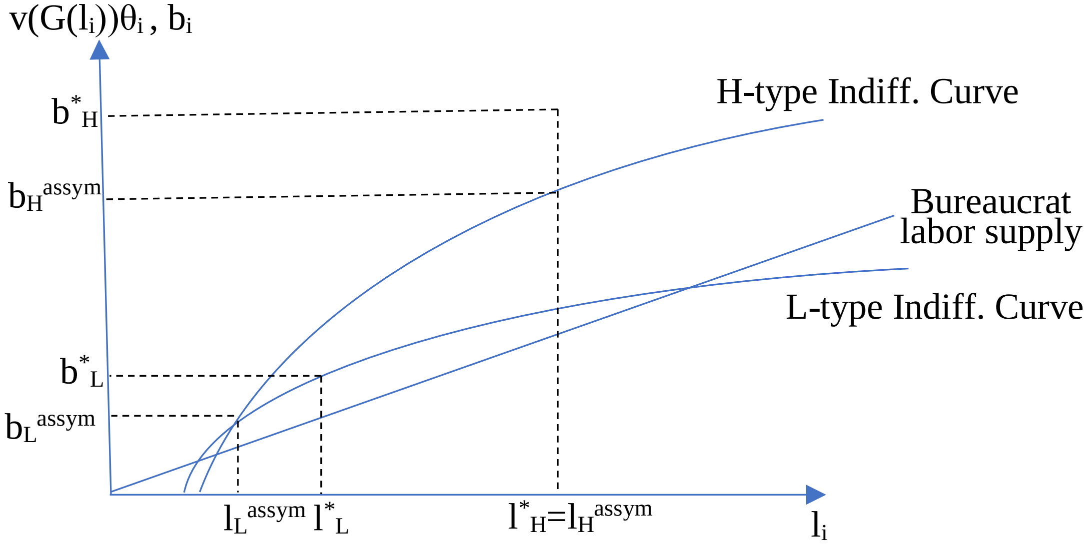
Using IC_H and Equation we can see that H-types pay a bribe below their willingness to pay and get a positive utility from the so called "information rent":

$$b_H^{assym} = \underbrace{v\left(t_H^{assym}\right)\theta_H}_{\text{Willingness to pay}} - \underbrace{v\left(t_L^{assym}\right)(\theta_H-\theta_L)}_{\text{Information rent}} \tag{10}$$

Comparing the asymmetric information equilibrium allocation to the Pareto optimal allocation
Comparing Equation 7 with the Pareto optimal solution in Equation 2 we can immediately see that they are the same. This is the classic "no distortion at the top" result from (Mirrlees, 1971). However, as described above the bribe paid by the H-types is lower than in the full information case.

The equality between the Pareto optimal labor allocation and the allocation reached under the asymmetric information equilibrium does not hold when comparing Equation 2 and Equation 8. Since $\left[1-\left(\frac{\beta}{1-\beta}\right)\left(\frac{\theta_H-\theta_L}{\theta_L}\right)\right] <$

Figure 4: Asymmetric information equilibrium



1, L-types will be allocated less than the Pareto optimal level of labor. Since L-types are allocated less labor their willingness to pay for the service decreases and they hence pay a lower bribes will still ending up with zero utility.

Figure 4 displays this equilibrium graphically.

2.4 Effects of the Performance Scorecards

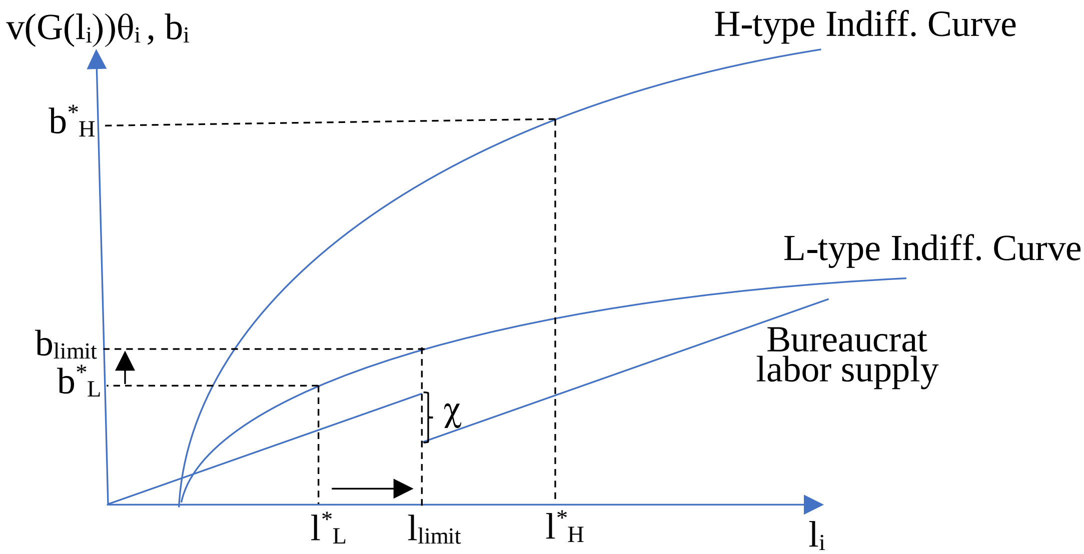
2.4.1 Full information with Performance Scorecards

When the Performance Scorecards are implemented (i.e. when $\chi > 0$) bureaucrats receive additional utility, a "performance bonus" when an application is processed within a specific time limit t_{limit}. I define $l_{limit} = G^{-1}(t_{limit})$ so that the bureaucrat can be said to receive this bonus is $l_i > l_{limit}$. Under full information this does not affect the bureaucrats behavior much. The bureaucrat simply continues to maximize the surplus between herself and the applicant, including the performance payment χ, and continues to extracts all of the surplus through a bribe. There are three ranges of θ for which applicants are affected differently from the increase in χ. If the applicant's valuation θ is high enough she will not be affected because the application would have been processed within the time limit even without the performance scorecard. If the valuation is low enough (below some lower cutoff θ_{LC}) the applicant is not affected because the application will be processed slower than the government mandated time limit even with the Performance Scorecard. Finally, there is a range of valuations such that the performance bonus encourages the bureaucrat to process the application just within the time limit.

To assist the intuition behind these three cases, in the formal definition I will be using l_i^* instead of θ_i, but since l_i^* is directly determined by θ_i through Equation 3 the conditions below could just as well been described in terms of ranges for θ_i.

1. If $l_i^* \geq l_{limit}$ (or equivalently $t_i^* \leq t_{limit}$) applicants are not affected by χ since they receive the service within the time limit even when $\chi = 0$

Figure 5: Performance Scorecard under Full Information



2. If $l_i^* < l_{LC}$ applicants are not affected by χ since even with χ there is no gain to the bureaucrat from increasing l_i from l_i^* to l_{limit}

3. If $l_{LC} < l_i^* < l_{limit}$ applicants will be affected by χ and they will now receive the service exactly at the limit, i.e. $l_i = l_{limit}$

The lower cutoff l_{LC}, or equivalently θ_{LC}which is the θ_i such that $l_i^* = l_{LC}$ is determined by the following equation:

$$v\left(G\left(l_{LC}\right)\right)\theta_{LC} - dl_{LC} = v\left(G\left(l_{limit}\right)\right)\theta_{LC} - dl_{limit} + \chi$$

$$\left(v\left(G\left(l_{LC}^*\right)\right) - v\left(G\left(l_{limit}\right)\right)\right)\theta_{LC} + d\left(l_{limit} - l_i^*\right) = \chi$$

We can now derive a comparative static for how changes in χ affects the size of the group getting the service at the time limit:

$$\frac{d\chi}{d\theta_{LC}} = \underbrace{v\left(G\left(l_{LC}^*\right)\right) - v\left(G\left(l_{limit}\right)\right)}_{<0} + \frac{dl_{LC}^*}{d\theta_{LC}}\underbrace{\left(G'\left(l_{LC}^*\right)v'\left(G\left(l_{LC}^*\right)\right)\theta_{LC} - d\right)}_{=0} < 0$$

Where the second term is equal zero due to Equation 3.

$$\frac{d\chi}{d\theta_{LC}} < 0 \Longleftrightarrow \frac{d\theta_{LC}}{d\chi} < 0$$

So that if χ is increased, the number of individuals getting the services at exactly t_{limit} increases since the lower cutoff θ_{LC} decreases.

2.4.2 Model predictions under full information

Processing times

The model predicts that under full information, the performance scorecard will increase the processing speed for a group of individuals who otherwise would have had processing time just below the time limit.

Bribes

Under full information bribes will only be affected for those applicants with a θ_i such that l_i^* lies between l_{LC} and l_{limit}. All other applicants will continue to pay their full willingness to pay and since their processing times are unchanged their willingness to pay will also be unchanged.

For the group that have their processing times increased from t_i^* to t_{limit} by the Performance Scorecard bribes will also increase. Since the bureaucrats extract the full willingness to pay from the applicants the bribe payments will increase from $v\left(G\left(l_i^*\right)\right)\theta_i$ to $v\left(G\left(l_{limit}\right)\right)\theta_i$ for this group.

An additional prediction is that the dispersion of bribe payment at t_{limit} will increase. Before the Performance scorecard anyone who got their application processed at exactly the time limit would have the same valuation θ_{limit} and hence would have paid the same bribe $v\left(G\left(l_{limit}\right)\right)\theta_{limit}$. However, with the scorecard a larger number of applicants with valuations ranging from θ_{LC} to θ_{limit} will have their applications processed at the time limit and hence bribe payments between $v\left(G\left(l_{limit}\right)\right)\theta_{LC}$ and $v\left(G\left(l_{limit}\right)\right)\theta_{limit}$ should be observed.

2.4.3 Asymmetric information and Performance Scorecards

In the asymmetric information equilibrium there are also different values of θ_L and θ_H for which the introduction of the Performance Scorecard have different implications. Remember that when $\chi = 0$ Equation 7 and 8 characterize the interior solution for the optimal level of labor effort from the bureaucrat. To study the interesting case that this model yields, I will assume that the time limit is set such that $l_H^{assym} > l_{limit}$ and $l_L^{assym} < l_{limit}$. There are two potential solutions for the bureaucrat, one is to keep the interior solution $l_L = l_L^{assym}$, the other is the corner solution, i.e. to set $l_L = l_{limit}$. The corner solution will result in a higher bureaucrat utility when:

$$\beta\left[v\left(t_H^{assym}\right)\theta_H - v\left(t_L^{assym}\right)\left(\theta_H - \theta_L\right)\right] + (1-\beta)\left[v\left(t_L^{assym}\right)\theta_L\right] - d\left(1-\beta\right)G^{-1}\left(t_L^{assym}\right) < \beta\left[v\left(t_H^{assym}\right)\theta_H - v\left(t_{limit}\right)\left(\theta_H - \theta_L\right.\right.$$

As we can see from the equation above, for all θ_L and θ_H such that $l_H^{assym} > l_{limit}$ and $l_L^{assym} < l_{limit}$, there exist a level of performance bonus χ such that the corner solution is optimal for the bureaucrat.

The corner solution $l_L = l_{limit}$ does not only change the labor allocation and bribes for the L-types but, as opposed to the full information case, also reduce the bribe payments for the H-types since they have to be made better off in order for them not to choose the menu option meant for the L-types.

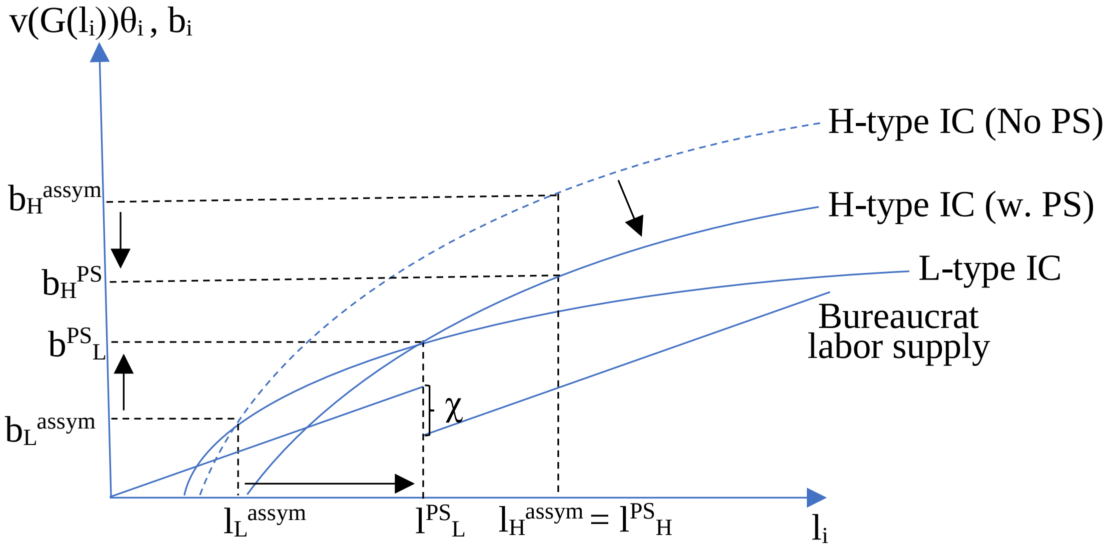
Below I denote the corner solution allocation under the Performance Scorecard with the superscript PS:

$$l_L^{PS} = l_{limit}$$

L-types will receive the service exactly at the time limit.

$$b_L^{PS} = v\left(G\left(l_{limit}\right)\right)\theta_L$$

Figure 6: Performance Scorecard under Asymmetric Information



Since L-types are still paying their willingness to pay they will pay exactly as much as they are willing to at the new faster service delivery.

$$l_H^{PS} = l_H^*$$

H-type still receive the Pareto optimal labor allocation.

$$b_H^{PS} = v\left(G\left(l_H^*\right)\right)\theta_H - v\left(G\left(l_{limit}\right)\right)\left(\theta_H - \theta_L\right)$$

The bribe payment for the H-types is still determined by Equation 10 but now the information rent has increased since the L-types are receiving their service faster.

2.4.4 Model predictions under asymmetric information

Processing times The L-types processing times decrease as:

$$l_L^{limit} < t_L^*$$

The H-types processing times are unaffected:

$$t_H^{limit} = t_H^*$$

Bribes The L-types bribes increase as their willingness to pay for the faster service increase: $b_L^{limit} = v\left(t_L^{limit}\right)\theta_L > b_L^* = v\left(t_L^*\right)\theta_L$

The H-types bribes decrease since their information rents has to increase to keep them from choosing the menu option meant for the L-types:
$b_H^{limit} = v\left(t_L^{limit}\right)\left(\theta_L - \theta_H\right) + v\left(t_H^{limit}\right)\theta_H < b_H^* = v\left(t_L^*\right)\left(\theta_L - \theta_H\right) + v\left(t_H^*\right)\theta_H$

2.5 Model predictions for experiment

This model has two sets of predictions depending under what information setting the experiment takes place, some of the predictions are common to both information setting while some of them differ. Furthermore, all of these predictions assume that the Performance Scorecard indeed create an additional incentive for bureaucrats to process applications within the time limit, i.e. it causes χ to become positive. I will group the predictions into two groups, main predictions and secondary prediction. The main predictions are those where the effects should be large enough for me to see them in my experiment. I do not expect to see the secondary predictions in my experiment but I do include them here for completeness and will test for them in the data.

Main predictions:

- Prediction 1 (FI and AI): More applications will be processed just within the time limit and fewer applications will be processed just outside of the time limit
- Prediction 2-FI: Bribe payments will not be affected for those who had their applications processed well above the time limit
- Prediction 2-AI: Bribe payments will decrease for those who had their applications processed well above the time limit

Secondary prediction:

- Prediction 3 (FI and AI): Bribe payments will increase marginally for those with applications around the time limit in the processing time distribution
- Prediction 4 (FI and AI): The dispersion of bribe payments will increase marginally for those with applications just above the time limit in the processing time distribution

3 Experiment

3.1 Context: Mutations applications and the eMutation system

Applications for land record mutations in Bangladesh are sent to the Upazila Land Office (ULO)[2]. There are 491 ULOs in Bangladesh and each ULO is headed by an Assistant Commissioner (Land) (ACL), which is an entry-level position in the Bangladesh Civil Service. Each ULO covers 5-15 Union Parishads, the lowest tier of government in Bangladesh. The ACL is directly supervised by the Upazila Nirbahi Officer (UNO), the top bureaucrat at the Upazila level and a mid-level officer in the Bangladesh Civil Service. After a land record mutation application has been submitted to the ULO, the Union Parishad Land Office (UPLO) inspects the documents submitted as well as the land in question. If the UPLO finds the application to be correct the UPLO prepares a "proposal for mutation" and submits this back to the ULO. The ULO then calls both the parties of the land transaction for a hearing with the ACL and the ACL can then accept the mutation and issue a "Record of Right" (RoR or *Khatian* in Bengali) to the new owner. The whole process from submission

[2] In urban areas these offices are simply called Assistant Commissioner (Land) Office, since there are no Upazilas in urban areas, but for the purpose of this paper I will refer to all land offices as ULOs.

to acceptance of the mutation should not take more than 45 working days [3]. However, applications are often delayed well beyond the set time and ACLs and the staff working in the ULO and UPLO are notorious for requiring bribes. In a survey conducted by Transparency International, 58.6% of households that had done a mutation had also stated they had paid a bribe with this mutation. The average bribe amount was BDT 10,014 (approximately USD 130) .

The eMutation system allows applications for land records to be made online or at a computer in the ULO with the assistance of a clerk. The system helps the ULO to communicate digitally with the UPLO and generate and print out the required documents faster. In qualitative interviews ACLs using the eMutation system claim that it increases the control they have over each application since they can track it in the system via a "dashboard", reduces the number of times the applicant has to visit the ULO in person and reduces the risk of forgery of the Record of Rights. Another feature of the eMutation system is that it saves the information regarding each mutation digitally on a server managed by a2i. This generation of digital data in theory enables UNOs and the Ministry of Land to see how many mutations that has been processed in each ULO and how long it took for each application to be processed. However, to date this data has not been used by the UNOs or the Ministry or Land to assess the performance of the ACLs.

The roll-out of the eMutation system is under way and currently 114 ULOs have been connected to the system. It is in these ULOs our experiment will take place. For some of the ULOs (currently 18) all the applications for mutations are processed through the eMutation system but for most of the ULOs there are some applications that are still processed manually. One reason for this is that not all UPLOs in these Upazilas are connected to the eMutation system. During a transition period, applications from some Union Parishads within the Upazilas will be done using the eMutation system while others will be processed using the traditional paper based system. As described below, we will use this partial implementation to measure spillovers of our experiment on applications that does not affect the scorecard.

3.2 Experimental intervention

The intervention of our experiment is to generate a monthly “Performance scorecard” for each ACL and share this with the ACL and the UNO as well as with the relevant bureaucrats at the District level i.e. the Additional Deputy Commissioner (revenue) and the Deputy Commissioner. The scorecard will contain information on the number of applications processed within 45 working days in the past month as well as the number of applications pending for more than 45 working days at the end of each month. In addition to these figures, a percentile ranking among the ACLs connected to the eMutation system will be generated so anyone receiving the scorecard can assess the ACL relative to other ACLs in the country. The purpose of the scorecard is to allow the UNO and other superiors of the ACL to observe the ACL’s performance in order to incentivize the ACL to provide a timely service to applicants. The UNO is the “Report Initiating Officer” for the ACL, meaning that it is the UNO that assesses the ACL in the annual review that forms the basis for the ACLs career progression. We therefore expect the ACL to be responsive to this improved monitoring by the UNO by making sure that the scorecards reflect positively on the ACL’s work.

[3] Whenever "days" is mentioned in this paper I am referring to working days.

3.3 Randomization

The main identification strategy of the study will be the randomization of which offices the scorecard system is implemented in. Out of 114 ULOs currently connected to the eMutation system, 57 will have scorecards generated for them and sent to the ACLs themselves, their respective UNOs and District Commissioners. There are currently 18 ULOs that have fully implemented the eMutation system so that all applications are processed using the digital system. The randomization will be done separately for the group with the 100% implementation and for the group with partial implementation. After these two groups have been separated the randomization will be stratified using the following strata:

- Having processed above/below the median number of applications within 45 working days in the months of June and July, 2018 in full implementation group
- Having processed within the first, second or third tertile number of applications within 45 working days in the months of June and July, 2018 in partial implementation group
- Having above/below the median number of applications pending for more than 45 working days in full implementation group
- Having a number within the first, second or third tertile of applications pending for more than 45 working days in partial implementation group

This gives me 13 strata. Within each strata half of the ULOs are assigned to treatment. [4]

3.4 Data sources

This project will have three main data sources, digital administrative data, analog administrative data and applicant survey data. I will describe the administrative and survey data separately.

3.4.1 Administrative data

This is the data that comes from each application's administrative paper work or in the case of eMutations data digitally entered into the eMutation system. For applications that were processed through the eMutation system this data is stored on a government server and will be transferred to me every month while the experiment is still running. For applications that were processed manually I will visit each ULO and digitize the key fields from a random sample of applications approximately 5 months after the intervention. The key variables coming out of the administrative data for my empirical tests are the following:

- Processing time of application (used to generate number of applications processed within the time limit)
- Number of applications pending
- Contact information of applicants

[4] If there is an odd number of ULOs in a strata the last ULO is grouped together with other such "misfits" in their implementation group and half of the misfits are randomly assigned treatment. Again if there are misfits these are grouped together with potential misfits from the other implementation group and half of those are assigned treatment. Finally if there is still one misfit left this ULO is given the treatment with 50% probability.

3.4.2 Survey data

I will use the contact information of the applicants to survey a random sample of applicants, ideally through phone surveys but if that is not possible I will use in-person surveys. The choice between phone surveys and in-person surveys will be done after a pilot of the response rate to phone surveys. The key variables coming out of the surveys for my empirical tests are the following:

- Bribes paid (broken down by to whom the bribe was paid)
- Willingness to pay for faster service

I will also use the survey data to cross-check that the information in the administrative data is correct.

4 Pre-Analysis Plan for data from experiment

4.1 Pre-analysis plan strategy

Following Olken (2015) I will pre-specify what one could call the "first level" of analysis, in other words the analysis I had in mind when designing the experiment. In general I will not attempt to design a decision tree and make statements such as "if I find result X I will conduct analysis Y" instead I will simply state how I will conduct the test for result X and once I have seen the result I will decide on what further analysis would be interesting to conduct. This strategy will be explained in the publications following this experiment and I will clearly state what parts of the analysis was pre-specified and what parts arose as a result of the first level results.

Despite my best efforts to write a comprehensive pre-analysis plan there might arise "first-level" interesting questions and/or econometric specifications and tests that were not included in the pre-analysis plan. We may report such finding but with the caveat that they are exploratory findings that needs to be replicated independently in order for the p-values to have their standard interpretation.

I have organized the pre-analysis plan according to the domains of outcome variables that I will estimate the Performance Scorecards effect on. I have aggregated the outcome variables into these domains in order to create one index per domain that I will test a hypothesis on using a statistical test. Aggregating outcomes into a smaller number of indexes reduces the number of hypotheses I will test and therefore reduces the need for multiple hypothesis adjustment of the p-values of these tests, see Section 5.2 for details about how I will adjust for multiple hypothesis testing. The domains I have aggregated our outcome variables into are the following:

- Domain 1: Effect of Performance Scorecard on bureaucrat performance as measured by the scorecards (Section 4.3)
- Domain 2: Benefits of Performance Scorecard to applicants (Section 4.4)
- Domain 3: Testing predictions of the model outlined in Section 2 (Section 4.5)
- Domain 4: Spillover effect on applications not entering the scorecards (Section 4.6)

For all the analysis below I will allow for 1 month of the intervention before I start measuring the outcomes. I.e. I will ignore the first month of data after the intervention has started and will measure the effect of the intervention starting from the second month.

For all the continuous variables described below I will winsorize them at the 99th percentile if their lowest possible value is zero and at both the 1st and 99th percentile if they can take on negative values.

4.2 Balance of Randomization

The first analysis I will conduct is to test for the equivalence of the Upazilas and applications in the pre-intervention period. In specific, I will restrict the pre-intervention period to the month before the first Scorecard was sent out. First I will test the treatment-control balance between the Upazilas using t-tests for the averages of the following variables from the digital administrative data:

1. The number of eMutation applications received
2. The Performance Scorecard rankings of the Upazila (two observations for each Upazila)
3. Number of applications processed within 45 working days
4. Number of cases pending for more than 45 working days

At the application level I will test for balance on the following variables for all applications that were submitted in the year leading up to the intervention:

1. A dummy variable for if the application was processed within 45 working days
    (a) A dummy variable for if the application was processed within 30 to 45 working days
    (b) A dummy variable for if the application was processed within 46 to 65 working days
2. Working days the applications took to process among processed applications
3. A dummy variable for if a bribe was paid or gift was given (either a direct bribe or a payment to an agent)
4. The bribe amount (among those who paid a bribe)
5. Total cost of the application: estimated monetary value of time spent and monetary costs
6. A dummy variable for if the application was approved

Finally I will check for balance on demographic variables among the applicants:

1. Age
2. Gender
3. Address in rural area (outside Upazila, District or Division capital)

In addition to the individual t-tests I will also perform three F-tests (one at the Upazila level, one for the variables available for all applications and one for the variables available for processed applications) for the joint significance of all these variables predictive power on treatment.

4.3 Bureaucrat performance

4.3.1 Upazila-Month level analysis

The first and most obvious effect the Performance Scorecards may have is that on the variables directly targeted by the Scorecards themselves, namely the number of applications processed within 45 working days for each month and the number of applications pending for more than 45 working days at the end of each month. I will test the effect on these outcomes using the digital administrative data using the following econometric specification:

$$Outcome_{it} = \alpha + \beta_1 Treatment_i + \beta_2 Outcome_{i,-1} + Strata_i + \gamma_t + \varepsilon_{it} \tag{11}$$

Where the $Outcome_{it}$ is one of three outcome variables for Upazila i in month t, where $t = -1$ is the baseline month, $t = 0$ is the first month of the intervention and $t > 0$ are the months for which the effect of the intervention will be measured. The four outcome variables are:

1. The log [5] of the number of eMutation applications received. Since I do not have a strong prior for this analysis it will not enter into the index or multiple hypothesis testing adjustment described in Section 5.2 below.
2. The rankings of the Upazila (two observations for each Upazila)
3. The log of the number of applications processed within 45 working days
4. The log of cases pending for more than 45 working days[6] $Treatment_i$ is the treatment status of ULO i. $Outcome_{i,-1}$ is the value of the outcome variable in the baseline period, the baseline period is defined as the month preceding the start of the intervention. $Strata_i$ is a strata fixed effects and γ_t are month fixed effects.

4.3.2 Short-term vs. medium term effects

To separate out the short-term and medium term effects of the scorecard I will estimate 5 different β_1 using month 2-6 after the implementation of the scorecard. The econometric specification of this analysis will be the following:

$$Outcome_i = \alpha + \beta_1 Treatment_i + \beta_2 Outcome_{i,t-1} + Strata_i + \varepsilon_i$$

This regression will be run 5 times, one for each of the months in the period from month 2 to 6 after the experiment. Since I do not have a strong prior for this analysis it will not enter into the index or multiple hypothesis testing adjustment described in Section 5.2 below.

[5] Throughout this Pre-analysis plan I will refer to the natural logarithm as the "log".

[6] I do not expect to have any zero values. However, if I were to have a zero values in any of the offices I will use the inverse hyperbolic sine transformation instead of the natural logarithm Burbidge et al. (1988).

4.3.3 Heterogeneous effects between high vs. low performers

It is possible to imagine that the Performance Scorecard has different effects depending on if they are positive or negative. [7]

To test for heterogeneous effects between those receiving a positive first scorecard vs those who received a negative first scorecard I use the following econometric specification:

$$Outcome_{it} = \alpha + \beta_1 Treatment_i + \beta_2 Treatment_i \times Baseline > 50pctil_i + \beta_3 Baseline > 50pctil_i + Strata_i + \gamma_t + \varepsilon_{ti}$$

Where the outcomes are defined as above and $Baseline > 50pctil_i$ is a dummy for if the sum of the two percentile rankings is above or below the 50th percentile of this sum. Since I do not have a strong prior for which way the heterogeneous effects will go these will not enter into the index or multiple hypothesis testing adjustment described in Section 5.2 below.

4.4 Benefits of Performance Scorecard to applicants

While the analysis in Section 4.3 above answers the question, "Does Performance Scorecards increase performance as measured by the scorecards?" Another interesting and policy relevant question is "What is the benefit of the performance scorecards for the applicant?" To measure this I will estimate the effect of the intervention on individual applications using the following specification:

$$Outcome_{ait} = \alpha + \beta Treatment_i + Strata_i + \gamma_t + \varepsilon_{ait} \tag{12}$$

Where $Outcome_{ait}$ is an outcome for application a in ULO i. γ is a dummy variable for the month the application was created in (not when it was disposed by the bureaucrat). For this analysis I will use a large number of outcomes:

1. A dummy variable for if the application was processed within 45 working days
2. The log of days the applications took to process where unprocessed applications will be given a predicted processing time equal to the average processing time from the second month of the intervention for that ULO
3. Number of trips related to application
4. Total amount of hours spent on application (trips and preparation)
5. Monetary costs: travel costs, fees, payments to agents and direct bribes
6. A dummy variable for if a bribe was paid or gift was given (either a direct bribe or a payment to an agent)
7. The bribe amount (zero for those who paid no bribe)

[7] Evidence of such heterogeneous effects are described in Ashraf (2018) although this paper empathize the importance of actual performance compared to expected performance in the report cards, something that I will not measure in my study.

8. Total cost of the application: estimated monetary value of time spent and monetary costs (Since this is a sum of variables tested individually it will not enter into the index or multiple hypothesis testing adjustment described in Section 5.2.)

9. A dummy variable for if the application was approved (I do not have a strong prior for this outcome and hence it will not enter into the index or multiple hypothesis testing adjustment described in Section 5.2.)

Note that outcomes number 3 through 8 are coming from the survey data and will have substantially smaller sample size than outcomes 1, 2 and 9.

4.5 Testing the predictions of the model

As described above the model has different predictions depending on the information setting but the econometric specification will be the same when testing the model under both information settings.

To test Prediction 1, "*More applications will be processed just within the time limit and fewer applications will be processed just outside of the time limit*", I will use the following specifications:

$$Between\_30\_45_{ait} = \alpha + \beta_{just\_within} Treatment_i + Strata_i + \gamma_t + \varepsilon_{ait} \tag{13}$$

$$Between\_46\_65_{ait} = \alpha + \beta_{just\_outside} Treatment_i + Strata_i + \gamma_t + \varepsilon_{ait} \tag{14}$$

Where a $Between\_30\_45_{ait}$ is a dummy for if an application was processed in 30 to 45 working days and $Between\_46\_65_{ait}$ is a dummy for if an application was processed in 46 to 65 working days. The model's prediction is that $\beta_{just\_within} > 0$ and that $\beta_{just\_outside} < 0$.

Prediction 2 is an important prediction because it will allow us to differentiate between the full information setting and the asymmetric information setting and hence understand if the changes caused by the experiment are Pareto improvements or changes away from a Pareto optimal equilibrium. I will test Prediction 2 using the following specification for applications processed within 30 working days.

$$Bribe_{ait} = \alpha + \beta Treatment_i + Strata_i + \gamma_t + \varepsilon_{ait} \tag{15}$$

Where $Bribe_{ait}$ is the self reported payments beyond the official fees to government officials and agents. As described in the Section 2, in a full information setting the model predicts$\beta = 0$ while in an asymmetric information setting the model predicts $\beta < 0$. I will interpret a negative and statistically significant β as evidence in favor of an asymmetric information setting.

Since the test of Prediction 2 does not use the full data set, but only data for those applicants who were processed within 30 working days, I will conduct the following test to confirm that there is no selection effects on what type of applications are within the treatment vs. control group.

$$Within\_30_{ait} = \alpha + \beta Treatment_i + Strata_i + \gamma_t + \varepsilon_{ait}$$

Where $Within\_30_{ait}$ is a dummy for if the application was processed within 30 working days. If $\beta = 0$ I will conclude that the treatment does not affect what applications are processed within 30 working days.

However, if $\beta \neq 0$ and statistically significant at the 10% level, I will adjust the sample used in Equation 15 to include the same fraction of applications in the control group as the fraction who's application was processed within 30 working days in the treatment group. So, for example, if 25% of applications in the control group were processed within 30 working days I will include the fastest 25% of processing times in the treatment group. As long as the treatment does not change the order of which application was processed first there should be no issue of selection using this procedure.

4.5.1 Secondary predictions

As described above, I do not expect that my experiment has sufficient power to find evidence for these secondary effects. Hence, I will not include them in the index or multiple hypothesis testing adjustment described in Section 5.2 but I will still include them in my analysis for completeness.

To test Prediction 3 "*Bribe payments will increase marginally for those with applications around the time limit in the processing time distribution*", I will use the following specification for applicants to had their application processed between 30 and 65 working days:

$$Bribe_{ait} = \alpha + \beta Treatment_i + Strata_i + \gamma_t + \varepsilon_{ait} \tag{16}$$

The prediction is that $\beta > 0$.

To test prediction 4 I will use the absolute value of the errors from Equation 16 to run the following regression:

$$abs\left(\varepsilon_{ait}\right) = \beta Treatment_i + v_{ait} \tag{17}$$

The prediction is that $\beta > 0$.

4.6 Spillovers

I will estimate spillovers of the intervention by estimating the effect on mutation applications that were done manually and therefore did not enter the Performance Scorecard. There are two types of applications that were done manually. First, those that were done manually since they were because the come from a Union Parishad where the UPLO is not connected to the eMutation system and second those that were done manually despite coming from a Union Parishad connected to the eMutation system. I will start by testing if the intervention affected the number of applications received from these two groups:

$$ln\_applications_{it} = \alpha + \beta Treatment_i + Strata_i + \gamma_t + \varepsilon_{it}$$

Where $ln\_applications_{it}$ is the number of manual applications coming from Union Parishads with/without the eMutation system. Since I do not have a strong prior for this analysis it will not enter into the index or multiple hypothesis testing adjustment described in Section 5.2 below.

After having analyzed if the treatment affected the pool of applications in the manual system I will test for if bureaucrat labor is diverted away from these application to the applications that affects the performance scorecards using a specification very similar to that in Equation 11:

$$Outcome_{it} = \alpha + \beta_1 Treatment_i + \beta_2 Outcome_{i,-1} + Strata_i + \gamma_t + \varepsilon_{it}$$

Where the outcomes will be:

1. The log of the number of applications processed within 45 working days
2. The log of cases pending for more than 45 days[8]

I will also perform an application level analysis similar to that in Equation 12:

$$Outcome_{ait} = \alpha + \beta Treatment_i + Strata_i + \gamma_t + \varepsilon_{ait} \tag{18}$$

For this analysis I will use three different outcomes:

1. A dummy variable for if the application was processed within 45 working days
2. The log of working days the applications took to process where unprocessed applications will be given a predicted processing time equal to the average processing time from the second month of the intervention for that ULO
3. A dummy variable for if the application was approved (I do not have a strong prior for this outcome and hence it will not enter into the index or multiple hypothesis testing adjustment described in Section 5.2.)

This analysis has an important implication for the model. If there are spillovers this suggests that approximating the Bureaucrat's disutility of labor as linear in labor effort is not correct. If the analysis does suggest that there are spillovers I will incorporate this into my model.

5 Standard errors, adjustments for multiple hypothesis testing and unbalanced attrition

5.1 Clustering of standard errors

For each econometric specification I will report heteroskedasticity robust standard errors clustered at the ULO level. I will present two set of p-values, one which are standard p-values based on the standard errors and one which are adjusted for multiple hypothesis testing as per the description in Section 5.2.

5.2 Multiple hypothesis testing

As is evident from the many analyses described above, there is a high risk of false positives is all estimates significantly different from zero are automatically treated as "the effect of the intervention". To minimize the risk for spurious results due to "chance" and for our p-values to have the standard interpretation I will define domains of related outcome variables for which I will create summary indexes following Anderson (2008). Within each domain I also analyze each component of the domain while adjusting the p-values for multiple hypothesis-testing. Furthermore, I will analyze the domains in a pre-determined sequential way to avoid having to adjust the p-values of the standardized treatment effects for multiple hypothesis testing.

[8]If I were to have a zero values in any of the offices, I will use the inverse hyperbolic sine transformation instead of the natural logarithm Burbidge et al. (1988).

5.2.1 Creating summary indices

Within each domain with multiple outcome variables (all except Section) I will calculate a standardized treatment effect by creating an index. I will do this by following the 4 steps below:[9]

1. I will change signs of the outcome variables so that a higher number is always the direction of our prior. Variables for which I do not have a prior I will not enter into the indices.

2. Convert all outcome into standard deviations. I will denote the vector of standardized outcomes in domain j $\tilde{y}_j$.

3. Create a vector of weights for each outcome within a domain equal to the sum of the row entries of the inverted covariance matrix for all variables within that domain, normalized by dividing it by the sum of all weights. Formally, $w_j = \left[\left(1'\hat{\Sigma}_j^{-1} \right) \left(1'\hat{\Sigma}_j^{-1} 1 \right)^{-1} \right]'$ where $\hat{\Sigma}_j$ is the covariance matrix of all variables in domain j and **1** is a column vector of ones.

4. Create an index that is equal to the weighted average of the standard deviations of the outcome variables. Formally $Index_{ij} = w_j^{'} \tilde{y}_{ij}$

After having generated these indices we will run the following regression to test the hypothesis that the treatment has no effect on any of the variables:

$$Index_i = \alpha + \beta Treat_i + Strata_i + \varepsilon_i \tag{19}$$

This procedure has the advantage of reducing the number of hypothesis tests within a domain into one test and thereby avoiding the need for multiple hypothesis testing adjustments within a domain. It also increases our statistical power if my hypothesis is correct, since several insignificant results can add up to one significant result if many of the insignificant estimates are in the same directions as my hypothesis.

5.2.2 Family-wise error rate adjusted p-values within domain

Within each domain I will calculate the Family-Wise Error Rate (FWER) adjusted p-values based on 10,000 iterations of Anderson (2008) version of the free step-down re-sampling method of Westfall et al. (1993). The number of tests, M that the p-values will be adjusted for are all of the outcomes described in the pre-analysis plan above except those for which I have explicitly stated that they will not be part of the multiple hypothesis testing. The algorithm has the following 7 steps for a family of M outcomes:

1. Sort outcomes in order of increasing p-values of the rejection of the null hypothesis, $y_1, y_2, ..., y_M$ such that $p_1 < p_2 < ... < p_M$

2. Simulate a data set where the null hypothesis is true by drawing simulated treatment assignments from the actual distribution of the treatment assignment without replacement.

3. Calculate a set of simulated p-values $(p_1^*, p_2^*, ..., p_M^*)$ for the rejection of the null for each outcome using the simulated treatment assignments. These p-values will not have the same monotonicity as $p_1 < p_2 < ... < p_M$.

[9]Adopted from Anderson (2008).

4. Enforce the original monotonicity by setting $p_r^{**} = min\left\{p_r^*, p_{r+1}^*, ..., p_M^*\right\}$ for all r, where r represents the number in the original ordering. Now, by construction, $p_1^{**} \leq p_2^{**} \leq ... \leq p_M^{**}$.

5. Perform 10,000 replications of steps 2-4 and count S_r, the number of times that $p_r^{**} < p_r$, for each r.

6. Generate $p_r^{FWER*} = \frac{S_r}{10{,}000}$

7. The Family-Wise Error Rate adjusted p-values will be $p_r^{FWER} = min\left\{p_r^{FWER*}, p_{r+1}^{FWER*}, ..., p_M^{FWER*}\right\}$

For each outcome I will report the significance level in terms of Family-Wise Error Rate adjusted p-values and unadjusted p-values as well as the standard errors.

5.3 Attrition

I will report "attrition" as the difference between the number of applications randomly selected from the digital administrative data and the number of successful interviews with those applicants. I will test for if attrition was affected by the treatment using the following regression:

$$Attrition_i = \alpha + \beta Treatment_i + Strata_i + \gamma_t + \varepsilon_{ait} \tag{20}$$

Furthermore, since attrition can be of a different nature in the treatment and control group even if it is on average the same, we will restrict the sample to the attriting applications and run the analysis as we did for the baseline balance check (see Section 4.2). If either the β in regression equation 20 is significantly different from zero or if the balance check of the attriting applications rejects the statistical tests described in Section 4.2 we will use a Heckman selection correction Heckman (1979) as well as Lee bounds Lee (2009) to assess the extent to which the attrition may effect our estimates of the effect of the Performance Scorecard.

References

Michael L Anderson. 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, 2008.

Anik Ashraf. Do performance ranks increase productivity? evidence from a field experiment. 2018.

Abhijit V Banerjee. A theory of misgovernance. *The Quarterly Journal of Economics*, 112(4):1289–1332, 1997.

Patrick Bolton, Mathias Dewatripont, et al. *Contract theory*. MIT press, 2005.

John B Burbidge, Lonnie Magee, and A Leslie Robb. Alternative transformations to handle extreme values of the dependent variable. *Journal of the American Statistical Association*, 83(401):123–127, 1988.

J Heckman. Sample selection bias as a specification error. *Econometrica, 47 (1)*, 1979.

David S Lee. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3):1071–1102, 2009.

James A Mirrlees. An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208, 1971.

Benjamin A Olken. Promises and perils of pre-analysis plans. *Journal of Economic Perspectives*, 29(3): 61–80, 2015.

Peter H Westfall, S Stanley Young, and S Paul Wright. On adjusting p-values for multiplicity. *Biometrics*, 49(3):941–945, 1993.