THE UNIVERSITY OF CHICAGO

# DETECTION OF IRREGULAR ASSIGNMENTS OF CASES TO JUDGES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY
FRANCISCO DE ASIS DEL VILLAR ORTIZ MENA

CHICAGO, ILLINOIS
JUNE 2022

# Dedication

A mi papá, por ser diferente. A mi mamá, por su comprensión. A mi papá y a mi mamá, por su apoyo que, para fines prácticos, nunca ha tenido límites.

# Contents

# List of Figures

# List of Tables

# Acknowledgements

Special thanks goes to my advisor, Ali Hortaçsu. He welcomed me to participate in his research endeavors in the early days of my Ph.D., and vouched for me later on. His sober and practical approach to economics research has greatly influenced mine. For their large and incisive volume of comments on this and many previous research projects that did not see the light of day, I thank my committee members Stéphane Bonhomme and Alexander Torgovitsky. I have been lucky to have Stéphane's benevolent razor-sharp attention in numerous occasions. His drive to engage with whatever idea I put on presentation slides was a decisive thrust for my work. Finally, my greatest single intellectual debt is due to Alex. He opened the door for me to join his econometrics advising group and, in so doing, engineered the hybrid applied economist and econometrician that I strive to become.

# Abstract

I develop tools to detect irregular assignments of cases to judges and apply them to Ecuador's judicial system. I derive the sharp bounds on the overall, court-specific, and judge-specific probabilities that a case's assignment is inconsistent with existing regulations. The bounds rely on administrative case assignment data and one, or both, of the following assumptions: (i) that certain observed case characteristics do not influence which judge a case should be assigned to, and (ii) that the probability distribution over the judges that each case should be assigned to is known (e.g. uniform, random assignment). I construct a database of all publicly-available case assignments in Ecuador's district courts, with over two million assignments from 2016 to 2020 and I find 6.6% of judges to be involved in irregular assignments and 2.6% of assignments, or 58 thousand assignments, to be irregular overall.

# 1 Introduction

Regulations that govern the assignment of judicial cases to judges are ubiquitous and can be traced back to the 4th century BCE in Athens.[1] They aim to abolish the market for judges, where judicial decisions tend to favor the party with a higher willingness to pay for a favorable decision, the party that files the case and, hence, has a first-mover advantage, or the party that knows more about the set of available judges (Egan, Matvos, and Seru 2021). Indeed, there is widespread advocacy for clear rules to govern case assignments (Transparency International 2007). In practice, however, their successful implementation requires enforcement resources. When enforcement is low, a non-trivial amount of actual assignments could be *irregular*, or inconsistent with the regulations.

From an enforcement policy perspective, this raises the following questions. How many case assignments are irregular? To what extent are irregular assignments made in particular courts or to specific judges? In this paper, I develop tools to address these questions in any given judiciary. Then, I apply them in Ecuador, where multiple irregular assignment scandals have surfaced in 2021, one of which involves the recently deposed mayor of Quito, Ecuador's capital city. I construct a database that contains over 2 million case assignments made in district courts between March 2016 and February 2020, and I detect irregular assignments that are highly localized. I find 6.6% of judges to be involved in irregular assignments and at least 2.6% of assignments, or 58 thousand assignments, to be irregular overall.

At the heart of these measurements lie the case assignment regulations. They imply that various judicial case characteristics do not influence which judge the case should be

[1] During that century, Athenian jurors were randomly selected to participate in a given trial using a random, multi-stage selection process that involved an allotment machine called a *kleroterion*. See Dow (1937) p.198.

assigned to. Examples include the plaintiff's friendly ties with government officials, the amount of money claimed in a payment dispute, or the amount of paperwork that the plaintiff or prosecutor submits when she files the case. If they are sufficiently specific, then they also imply a probability distribution over the judges that each case should be assigned to (e.g. uniform, random assignment). Each implication yields an identification assumption that is informative for the probability that a case's assignment is irregular, given data on actual assignments. However, the second implication yields a more informative assumption that allows me to measure another parameter of interest: the probability that a case's assignment is irregular, conditional on a given judge. These parameters give a granular view on the structure of irregular assignments and are valuable inputs to guide the allocation of regulatory enforcement resources.

I begin by relating the judge that a case is assigned to with the judge that the case would have been assigned to, had its assignment been irregular, and the judge that the case would have been assigned to, had its assignment been *regulatory*, or in accordance with existing regulations. The counterfactual assignment that is observed will depend on whether the case's assignment is irregular or not. Unlike program evaluation models, where the treatment status is observed, we do not observe any case's irregular assignment status. Indeed, the distribution of the case's irregular assignment status is the object of interest. Thus, our model involves a discrete, two-component mixture.

I then study identification under two assumptions. First, the researcher observes a case characteristic that is statistically independent of counterfactual regulatory assignments only. I call such an instrumental variable a *one-sided instrument*. One-sided instruments are weaker than traditional instrumental variables (e.g. Imbens and Angrist 1994), which require exclusion from both counterfactual outcomes, and their identification power has not been studied in the context of mixture models. They differ from the instrumental variables studied by Henry, Kitamura, and Salanié (2014), which are excluded from observed outcomes, conditional on the unobserved state — the case's irregular assignment status.[2] The second assumption involves a stronger interpretation of the regulations: the probability mass function of counterfactual regulatory assignments is known.

Under each assumption, I obtain analytical solutions for the sharp bounds on the probability that a case's assignment is irregular, as well as the corresponding probabilities conditional on covariates (e.g. the case's court and time of assignment) and conditional on the judge that the case is assigned to. This is a two-step process. First, I show that the parameters of interest are examples of linear and scalar parameters whose identified sets are obtained by solving two linear programming problems. Related characterizations include Balke and Pearl (1997), Torgovitsky (2019), Tebaldi, Torgovitsky, and Yang (2019) and Laffërs (2019b). Then, I reformulate the linear programs that define the lower bounds for our parameters of interest in terms of optimal transport problems with convenient graphical interpretations and obtain their closed-form solutions.

The closed-form solutions for the bounds are simple. If the distribution of counterfactual regulatory assignments is known, the lower bound on the probability of an irregular assignment equals the absolute ($\ell_1$, taxicab, Manhattan) distance between this distribution and that of actual assignments. With a binary one-sided instrument, this lower bound equals the absolute distance between the two conditional distributions of actual assignments, weighted by the mass of the smaller group, defined in terms of the instrument values. Moreover, knowledge of the distribution of counterfactual regulatory assignments implies that the probability that a case's assignment is irregular, conditional on a given judge, is greater than the rate

[2]They also differ from the mismeasured counterparts used in the literature on data misclassification, which satisfy the same exclusion restriction as the instrumental variables of Henry, Kitamura, and Salanié (2014) (see Bollinger 1996, Mahajan 2006, Y. Hu 2008 and DiTraglia and García-Jimeno 2019).

of cases that the judge received in excess of what she should have received.

I then apply these findings in Ecuador, a country with a GDP per capita that is roughly ten times smaller than that of the U.S. (2019, World Bank), where multiple scandals involving manipulated assignments of cases to judges have surfaced in recent months. According to Ecuadorian regulations, the set of judges that a case can be assigned to depends on the case's location and field of law (i.e. whether the case pertains to criminal law, family law, administrative law, etc. . . ). Within a court, judges are selected on the basis of a lottery system that is not fully specified.

My primary data source is the public, plain text version of the government's database of judicial cases, available on a government website that facilitates individual case searches. I collected and structured the case assignment information for all cases that are contained in this website, to assemble a database with over two million case assignments, performed between March 2016 and February 2020 in Ecuador's 331 district courts.

The one-sided instrument that I consider is the amount of paperwork that the plaintiff or prosecutor submits when she files the case. Its exogeneity stems from the fact that cases with large or small amounts of plaintiff or prosecutor paperwork should not be assigned to judges differently. This instrument alone singles out a criminal court in the commercial hub of Guayaquil, with over 15% of irregular assignments.

A more specific interpretation of the regulations is that cases should be assigned to each of the competent judges in any given court with equal probabilities. I assume that a judge is competent for a given case if she is in an active spell when the case is assigned, and if she has a small enough case workload, compared with her peers. A scalar parameter governs each criterion and I select the parameters so as to obtain a conservative lower bound on the probability that a case's assignment is irregular. This exercise implies that at least 2.6% of case assignments are irregular. Moreover, it reveals that, out of the 1568 judges that are present in the data, 111 were involved in irregular assignments. In particular, the judge that I find received the largest amount of irregular assignments faced corruption charges that involved his participation in illegal sales of sentence reductions for inmates, only a few weeks after the end of my sample period.

The primary contribution of this paper is to develop tools to quantify the extent of irregular assignments of cases to judges on the basis of existing regulations and observed assignments. My setting is close to that of Daljord, Pouliot, Xiao, and M. Hu (2021), who measure the extent of black market trade of Beijing license plates under a local government rationing policy. When the distribution of counterfactual regulatory assignments is known, the lower bound on the probability that a case's assignment is irregular equals their optimal transport estimator of the lower bound on the probability that a license plate is traded in the black market. I build on their analysis by introducing one-sided instruments as a means to estimate the same parameter without imposing knowledge of the distribution of a counterfactual outcome. Second, I show that their optimal transport estimator equals the sharp lower bound on the parameter of interest when knowledge of the distribution of one counterfactual outcome is imposed.

My application to Ecuador's judiciary showcases the practical value of these tools to quantify behavior that is typically hard to measure. Indeed, early studies of government corruption (Reinikka and Svensson 2004, Fisman and Wei 2004, Olken 2006) rely on access to the joint distribution of actual outcomes and a potentially noisy measure of the outcomes that would have been observed, had there been no corruption.

From an econometric point of view, this paper introduces one-sided instruments to study non-parametric identification of mixture models (e.g. Hall and Zhou (2003), Henry, Kitamura, and Salanié (2014), Compiani and Kitamura (2016), Kitamura and Laage (2018)). My linear programming formulation of the identified set for the parameters of interest can be seen as an application of Lafférs (2019b), is inspired by Tebaldi, Torgovitsky, and Yang (2019), and is related with Balke and Pearl (1997), Lafférs (2013), Demuynck (2015), Lafférs (2019a) and Torgovitsky (2019). In my setting, I do not observe a proxy variable for the cases' irregular assignment statuses, a common feature in the data misclassification literature (e.g. Bollinger 1996, Mahajan 2006, Molinari 2008, Y. Hu 2008 and DiTraglia and García-Jimeno 2019), I do not have access to the irregular assignment status for a subset of cases, as in Molinari (2010), nor can I credibly set an upper bound on the probability that a case's assignment is irregular, as in Horowitz and Manski (1995).

I organize the paper as follows. Section 2 introduces the econometric framework in a stylized environment and presents the identification results. Section 3 develops the theory of identification that underlies the results presented in Section 2. Section 4 discusses the Ecuadorian context and the available data. Section 5 adapts the econometric framework to the Ecuadorian context and discusses estimation. Section 6 presents the estimation results and section 7 concludes.

# 2 Illustrative Framework

This section illustrates my identification results in a stylized econometric framework. The framework forms the basis for the empirical model that I use to measure irregular assignments in Ecuador.

## 2.1 Setting

Consider a stylized setting where a number of judicial cases, indexed by $i$, are assigned to one of $n_Y$ judges who worked in a given court during a specified time period (e.g. a quarter). Let $Y_i$ denote the judge that case $i$ is assigned to. Label judges from 1 to $n_Y$, so that $Y_i$ is an observed random variable that takes values in $\{1, \ldots, n_Y\}$.

In this setting, there exist regulations that specify how cases should be assigned to these judges. For example, regulations could mandate simple random assignment, or simple random assignment among a subset of judges. In practice, however, case $i$'s assignment may be *irregular*, or inconsistent with the regulations. Let $S_i$ indicate if $i$'s assignment is irregular or not. This is a latent, binary random variable.

Irregular assignments can arise for various reasons, which I will not attempt to distinguish at this stage. Some reasons, such as administrative errors, do not necessarily involve illegal behavior; others, such as transactions in the black market for judges, do; and some may involve behavior whose legal status is unclear, as with judge shopping, the practice of filing and withdrawing the same case multiple times until the case is assigned to the desired judge.

Consider two counterfactual assignments for any given case. The first counterfactual assignment is the judge that case $i$ would have been assigned to, had its assignment been irregular, $S_i = 1$. The second one is the judge that case $i$ would have been assigned to, had its assignment been *regulatory*, or consistent with the regulations, $S_i = 0$. We denote counterfactual irregular assignments with variable $Y_i(1)$ and counterfactual regulatory assignments with variable $Y_i(0)$. They relate to actual assignments $Y_i$ according to the potential outcomes equation:

$$Y_i \quad = \quad S_i Y_i(1) \; + \; (1 - S_i) Y_i(0). \tag{1}$$

That is, the judge that case $i$ is assigned to equals $Y_i(1)$ if $i$'s assignment is irregular ($S_i = 1$), and equals $Y_i(0)$ otherwise.

For the sake of illustration, I implicitly condition on case $i$'s covariates. I am interested in two parameters: the rate of irregular assignments, and the judge-specific rates of irregular assignments: $\Pr(S_i = 1)$ and $\Pr(S_i = 1 \,|\, Y_i = y^*)$ for each $y^* \in \{1, \ldots, n_Y\}$, respectively. In this setting, these parameters offer a detailed view of the extent and structure of irregular assignments. They are policy-relevant, since they inform the allocation of regulatory enforcement resources.

To measure these quantities, we need assumptions. No component on the right-hand side of (1) is observed. Thus, it is possible that $\Pr\big(Y_i = Y_i(1)\big) = 1$ and $\Pr(S_i = 1) = \Pr(S_i = 1 \,|\, Y_i = y^*) = 1$, for all $y^* \in \{1, \ldots, n_Y\}$. Similarly, it is possible that $\Pr\big(Y_i = Y_i(0)\big) = 1$ and $\Pr(S_i = 1) = \Pr(S_i = 1 \,|\, Y_i = y^*) = 0$, for all $y^* \in \{1, \ldots, n_Y\}$.

I consider two assumptions: that the distribution of regulatory assignments is known, and that the researcher observes a case characteristic $Z_i$ with finite support $\mathcal{Z}$ that does not influence the judge that the case would have been assigned to, had the case's assignment been regulatory.

**Assumption PMF.** *The probability mass function of* $Y_i(0)$ *is known.*

**Assumption IV.** $Z_i$ *is statistically independent of* $Y_i(0)$*.*

I now introduce my identification results for the parameters of interest under each assumption, in turn. At a conceptual level, the discussion of the identification results under Assumption PMF lays the groundwork to introduce the results under Assumption IV.

## 2.2 Identification Results under Assumption PMF

Assumption PMF states that the distribution of regulatory assignments is known. In the context of my application, I interpret Ecuadorian assignment regulations to mean that, within the court where case $i$ is assigned, $i$'s judge is drawn from a uniform distribution defined over the set of judges that are competent for case $i$ at the time of $i$'s assignment.

Consider the task of measuring the rate of irregular assignments, $\Pr(S_i = 1)$. According to (1), if case $i$'s actual and regulatory assignments differ ($Y_i \neq Y_i(0)$), then $i$'s assignment must be irregular ($S_i = 1$). This means that

$$\Pr(S_i = 1) \;\geq\; \Pr(Y_i \neq Y_i(0)).$$

$\Pr(Y_i \neq Y_i(0))$ is not identified, since the joint distribution of $(Y_i, Y_i(0))$ is unknown. However, we know the marginal distribution of $Y_i$, since actual assignments are observable, as well as the marginal distribution of $Y_i(0)$, under Assumption PMF. The unobserved joint distribution of $(Y_i, Y_i(0))$ must be consistent with the known marginal distributions. Therefore, a lower bound on the rate of irregular assignments is the minimum probability that case $i$'s actual and regulatory assignments differ that can be obtained from a joint distribution of $(Y_i, Y_i(0))$ that is consistent with the marginal distributions of $Y_i$ and $Y_i(0)$.

Let $\Gamma$ be the set of probability mass functions defined over $\{1, \ldots, n_Y\} \times \{1, \ldots, n_Y\}$. To summarize our discussion:

$$\begin{aligned}
\Pr(S_i = 1) \quad &\geq \quad \Pr(Y_i \neq Y_i(0)) \\
&\geq \quad \min_{\gamma \in \Gamma} \quad \sum_{y=1}^{n_Y} \sum_{y_0=1}^{n_Y} 1\{y \neq y_0\}\, \gamma(y, y_0) \qquad \text{subject to:} \\
&\qquad (i) \sum_{y_0=1}^{n_Y} \gamma(y, y_0) \; = \; \Pr(Y_i = y) \quad \text{for all } y \in \mathcal{Y} \\
&\qquad (ii) \sum_{y=1}^{n_Y} \gamma(y, y_0) \; = \; \Pr(Y_i(0) = y_0) \quad \text{for all } y_0 \in \mathcal{Y}.
\end{aligned} \tag{2}$$

Problem (2) is a discrete optimal transport problem (see Galichon (2016)). Because of its binary cost function, which assigns a cost of one if actual and regulatory assignments differ ($y \neq y_0$) and a cost of zero if they do not, problem 2 is particularly tractable. Indeed, its closed-form solution equals half of the absolute ($\ell_1$, Taxicab, Manhattan) distance between the marginal distributions of $Y_i$ and $Y_i(0)$ (see Propositions 4.2 and 4.7 of Levin and Peres (2017) for a textbook treatment):

$$\frac{1}{2} \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y) - \Pr(Y_i(0) = y) \Big|.$$

A similar reasoning produces a lower bound for the judge-specific rates of irregular assignments. Consider judge $y^* \in \{1, \ldots, n_Y\}$ and parameter $\Pr(S_i = 1 \,|\, Y_i = y^*)$. Model (1) implies that:

$$\Pr(S_i = 1, Y_i = y^*) \;\geq\; \Pr(Y_i \neq Y_i(0), Y_i = y^*).$$

That is, if case $i$'s actual and regulatory assignments differ, then $i$'s assignment is irregular, irrespective of the judge that the case was actually assigned to. $\Pr(Y_i \neq Y_i(0), Y_i = y^*)$ is not identified, since we do not observe the joint distribution of $(Y_i, Y_i(0))$. However, the data on actual assignments and Assumption PMF allow us to place a lower bound on this quantity with the minimum probability that $Y_i \neq Y_i(0)$ and $Y_i = y^*$ that can be obtained from a joint distribution of $(Y_i, Y_i(0))$ that is consistent with the marginal distributions of $Y_i$ and $Y_i(0)$:

$$\begin{aligned}
\Pr(S_i = 1, Y_i = y^*) \quad &\geq \quad \Pr(Y_i \neq Y_i(0), Y_i = y^*) \\
&\geq \quad \min_{\gamma \in \Gamma} \sum_{y_0=1}^{n_Y} 1\{y^* \neq y_0\}\, \gamma(y^*, y_0) \qquad \text{subject to:} \qquad (3) \\
&\qquad (i) \sum_{y_0=1}^{n_Y} \gamma(y, y_0) \;=\; \Pr(Y_i = y) \quad \text{for all } y \in \mathcal{Y} \\
&\qquad (ii) \sum_{y=1}^{n_Y} \gamma(y, y_0) \;=\; \Pr(Y_i(0) = y_0) \quad \text{for all } y_0 \in \mathcal{Y}.
\end{aligned}$$

Once again, (3) is an optimal transport problem. Its cost function assigns a cost of one if actual and regulatory assignments differ *and* the actual assignment is judge $y^*$, and assigns zero cost otherwise. In this case, the closed-form solution to this problem is simply the amount of cases assigned to judge $y^*$, beyond the amount of cases that should have been assigned to judge $y^*$:

$$\max\Big\{0, \quad \Pr(Y_i = y^*) - \Pr(Y_i(0) = y^*\Big\}.$$

Thus, Assumption PMF will place informative lower bounds on the rates of irregular assignments, to the extent that the distributions of actual and regulatory assignments differ. In contrast, Assumption PMF does not place an informative upper bound on these quantities. Because Assumption PMF does not place any restrictions on the distribution of irregular assignments, $Y_i(1)$, it is consistent with the possibility that assignments coincide with irregular assignments: $\Pr(Y_i = Y_i(1)) = 1$. In this case, every case assignment can be irregular: $\Pr(S_i = 1) = \Pr(S_i = 1 \,|\, Y_i = y^*) = 1$ for all $y^* \in \{1, \ldots, n_Y\}$. Proposition 1 summarizes the discussion.

**Proposition 1.** *If Assumption PMF holds, then*

1. $\frac{1}{2}\sum_{y=1}^{n_Y}\Big|\operatorname{Pr}(Y_i = y) \; - \; \operatorname{Pr}(Y_i(0) = y)\Big| \quad \leq \quad \operatorname{Pr}(S_i = 1) \quad \leq \quad 1.$

2. *For all* $y^* \in \{1, \ldots, n_Y\}$,

$$\max\left\{0,\ \frac{\operatorname{Pr}(Y_i = y^*) - \operatorname{Pr}(Y_i(0) = y^*)}{\operatorname{Pr}(Y_i = y^*)}\right\} \quad \leq \quad \operatorname{Pr}(S_i = 1 \,|\, Y_i = y^*) \quad \leq \quad 1.$$

*These bounds are sharp.*

Proposition 1, proven in Appendix C, asserts that the lower bounds that I have introduced are sharp. This means that, for each bound, there exists a joint distribution of the data, $(Y_i(0), Y_i(1), S_i)$, that is consistent with the known distribution of $Y_i(0)$ and with the distribution of actual assignments under equation (1), and generates the given bound. Informally, this means that more information on the parameters of interest cannot be obtained without more data or further assumptions.

Daljord, Pouliot, Xiao, and M. Hu (2021) first proposed the solution to problem (2) as a lower bound on the quantity of black market transactions of license plates in China, following the introduction of a lottery-based rationing system. In their setting, the observed outcome is the price of the car associated with license plate $i$, and they use the fact that license plates were supposed to be allocated by a lottery to obtain the distribution of car prices in the absence of a black market. Proposition 1 shows that this lower bound is sharp and, hence, promotes the estimand they propose.

## 2.3 Identification Results under Assumption IV

Under Assumption IV, one observes case characteristics $Z_i$ which, in conjunction with (1), may generate variation in actual assignments $Y_i$ through $S_i$ and/or $Y_i(1)$ exclusively. Because $Z_i$ is excluded only from $Y_i(0)$, I call it a *one-sided instrument*. It differs from the traditional exclusion restriction (e.g. Imbens and Angrist 1994), whereby the instrument generates variation in assignments through $S_i$ only (i.e. statistical independence holds with respect to $(Y_i(0), Y_i(1))$). It also differs from the exclusion restriction proposed by Henry, Kitamura, and Salanié (2014), which requires that $Z_i$ be independent of $Y_i$, conditional on $S_i$.[3]

Case characteristics $Z_i$ are discrete. In my application, $Z_i$ is a binary measure of the amount of paperwork submitted by the plaintiff/prosecutor when she files the case. In support of Assumption IV, I argue that Ecuadorian regulations do not contain specific assignment procedures for cases that differ along this dimension, and that this case characteristic is independent of the case characteristics that determine regulatory assignments. Notice that the traditional exclusion restriction, $Z_i \perp\!\!\!\perp (Y_i(0), Y_i(1))$, is unlikely to hold for $Z_i$. Indeed, $Z_i$ is presumably correlated with irregular assignments, $Y_i(1)$: plaintiffs that file cases with larger amounts of paperwork may value judge attributes differently from others. Plaintiffs with different preferences over judges would select different judges if they were given the chance to do so.

Consider the task of measuring the rate of irregular assignments, $\Pr(S_i = 1)$, under Assumption IV. I proceed as in the discussion of identification under Assumption PMF. Conditional on instrument realization $z \in \mathcal{Z}$, where $\mathcal{Z}$ is the finite support of $Z_i$, a case's assignment is irregular if its actual assignment differs from its regulatory assignment, by

[3] An equivalent formulation of this statement in terms of potential outcomes is that $Z_i$ is independent of $Y_i(1)$ within the subpopulation with $S_i = 1$ and that $Z_i$ is independent of $Y_i(0)$ within the subpopulation with $S_i = 0$.

model (1). This observation yields a lower bound on the rate of irregular assignments conditional on $Z_i = z$:

$$\Pr(S_i = 1 \,|\, Z_i = z) \;\geq\; \Pr(Y_i \neq Y_i(0) \,|\, Z_i = z).$$

A further lower bound on $\Pr(S_i = 1 \,|\, Z_i = z)$ is given by the minimum probability that case $i$'s actual and regulatory assignments differ, conditional on $Z_i = z$, that can be obtained from a joint distribution of $(Y_i, Y_i(0)) \,|\, Z_i = z$ that is consistent with the marginal distributions of $Y_i \,|\, Z_i = z$ and $Y_i(0) \,|\, Z_i = z$:

$$\begin{aligned}
\Pr(S_i = 1 \,|\, Z_i = z) \quad &\geq \quad \Pr(Y_i \neq Y_i(0) \,|\, Z_i = z) \\
&\geq \quad \min_{\gamma \in \Gamma} \quad \sum_{y=1}^{n_Y} \sum_{y_0=1}^{n_Y} 1\{y \neq y_0\}\, \gamma(y, y_0) \qquad \text{subject to:} \qquad (4) \\
&\qquad (i) \sum_{y_0=1}^{n_Y} \gamma(y, y_0) \;=\; \Pr(Y_i = y \,|\, Z_i = z) \quad \text{for all } y \in \mathcal{Y} \\
&\qquad (ii) \sum_{y=1}^{n_Y} \gamma(y, y_0) \;=\; \Pr(Y_i(0) = y_0 \,|\, Z_i = z) \quad \text{for all } y_0 \in \mathcal{Y}, \\
&= \quad \frac{1}{2} \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \,|\, Z_i = z) \;-\; \Pr(Y_i(0) = y \,|\, Z_i = z) \Big|
\end{aligned}$$

where $\Gamma$ is the set of probability mass functions defined over $\{1, \ldots, n_Y\} \times \{1, \ldots, n_Y\}$, as before, and the last equality follows from the binary cost structure of optimal transport problem (4). It follows that

$$\begin{aligned}
\Pr(S_i = 1) &= \sum_{z \in \mathcal{Z}} \Pr(Z_i = z) \Pr(S_i = 1 \mid Z_i = z) \\
&\geq \frac{1}{2} \sum_{z \in \mathcal{Z}} \Pr(Z_i = z) \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \mid Z_i = z) \; - \; \Pr(Y_i(0) = y \mid Z_i = z) \Big| \\
&= \frac{1}{2} \sum_{z \in \mathcal{Z}} \Pr(Z_i = z) \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \mid Z_i = z) \; - \; \Pr(Y_i(0) = y) \Big|,
\end{aligned}$$

where the last equality follows Assumption IV. This lower bound is not identified, however, since the marginal distribution of $Y_i(0)$ is unknown. A further lower bound that is observable is the lowest possible lower bound that is implied by a distribution of $Y_i(0)$. Let $\Phi$ be the set of probability mass functions defined over $\{1, \ldots, n_Y\}$. It follows that

$$\Pr(S_i = 1) \quad \geq \quad \min_{\phi \in \Phi} \frac{1}{2} \sum_{z \in \mathcal{Z}} \Pr(Z_i = z) \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \mid Z_i = z) \; - \; \phi(y) \Big|. \qquad (5)$$

When $Z_i$ is binary, so that $\mathcal{Z} = \{0, 1\}$, this problem has a closed-form solution. To see this, let $p_{\min} \equiv \min\{\Pr(Z_i = 0), \Pr(Z_i = 1)\}$. For any $\phi \in \Phi$, it follows that

$$\begin{aligned}
& \frac{1}{2} \sum_{z \in \mathcal{Z}} \Pr(Z_i = z) \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \mid Z_i = z) \; - \; \phi(y) \Big| \\
= \; & \frac{1}{2} \Pr(Z_i = 0) \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \mid Z_i = 0) \; - \; \phi(y) \Big| \\
& \qquad + \; \frac{1}{2} \Pr(Z_i = 1) \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \mid Z_i = 1) \; - \; \phi(y) \Big| \\
\geq \; & \frac{p_{\min}}{2} \left( \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \mid Z_i = 0) \; - \; \phi(y) \Big| + \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \mid Z_i = 1) \; - \; \phi(y) \Big| \right) \\
\geq \; & \frac{p_{\min}}{2} \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \mid Z_i = 0) \; - \; \Pr(Y_i = y \mid Z_i = 1) \Big|,
\end{aligned}$$

where the last inequality follows from the triangle inequality. Moreover, this quantity is achieved by $\phi^*$, where

$$\phi^*(y) = \begin{cases} \Pr(Y_i = y \,|\, Z_i = 0) & \text{if} \quad \Pr(Z_i = 1) \leq \Pr(Z_i = 0) \\ \Pr(Y_i = y \,|\, Z_i = 1) & \text{otherwise.} \end{cases}$$

The following proposition summarizes.

**Proposition 2.** *If Assumption IV holds, then*

1. $\min_{\phi \in \Phi} \ \frac{1}{2} \sum_{z \in \mathcal{Z}} \Pr(Z_i = z) \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \,|\, Z_i = z) \ - \ \phi(y) \Big| \ \leq \ \Pr(S_i = 1) \ \leq \ 1$

2. *For all* $y^* \in \{1, \ldots, n_Y\}$, $\quad 0 \ \leq \ \Pr(S_i = 1 \,|\, Y_i = y) \ \leq \ 1.$

*In addition, if* $Z_i$ *is binary, so that* $\mathcal{Z} = \{0, 1\}$*, then*

$$\frac{p_{min}}{2} \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \,|\, Z_i = 0) \ - \ \Pr(Y_i = y \,|\, Z_i = 1) \Big| \ \leq \ \Pr(S_i = 1) \ \leq \ 1,$$

*where* $p_{min} \equiv \min\{\Pr(Z_i = 0), \Pr(Z_i = 1)\}$*. Each of these bounds is sharp.*

Proposition 2, proven in Appendix C, shows that binary instruments that yield informative lower bounds on $\Pr(S_i = 1)$ will satisfy two conditions. First, a relevance condition: $Z_i$ must induce variation in assignments for the absolute distance between the conditional assignment distributions to be positive. Second, $Z_i$ must be relatively balanced. This is intuitive: if the mass of cases with $Z_i = 0$ is small, the lower bound on $\Pr(S_i = 1)$ arises when regulatory assignments $Y_i(0)$ are distributed according to $Y_i \,|\, Z_i = 1$, in which case only a small fraction of cases' assignments may be irregular.

Proposition 2 also shows that Assumption IV does not yield informative bounds on

$\Pr(S_i = 1 \mid Y_i = y)$. Intuitively, this follows because, if judge $y$ is assigned a substantial amount of cases with $Z_i = 0$ but no cases with $Z_i = 1$, she could either have several cases with irregular assignments if $\Pr(Y_i(0) = y) = \Pr(Y_i = y \mid Z_i = 1)$, or no cases with irregular assignments if $\Pr(Y_i(0) = y) = \Pr(Y_i = y \mid Z_i = 0)$, and Assumption IV cannot distinguish between these possibilities. Thus, my measurements of judge-specific rates of irregular assignments in Ecuador will necessarily involve Assumption PMF.

Furthermore, Proposition 2 shows that Assumption PMF does not place an informative upper bound on the parameters of interest. The intuition for this result is that Assumption IV, like Assumption PMF, does not restrict the distribution of irregular assignments, $Y_i(1)$, so that it is consistent with $\Pr(Y_i = Y_i(1)) = 1$, and $\Pr(S_i = 1) = \Pr(S_i = 1 \mid Y_i = y^*) = 1$ for all $y^* \in \{1, \ldots, n_Y\}$.

Finally, the bounds that Proposition 2 presents are sharp. This means that, for each bound, there exists a joint distribution of the latent and observed data, $(Y_i(0), Y_i(1), S_i, Z_i)$, that satisfies Assumption IV, is consistent with the observed joint distribution of $(Y_i, Z_i)$ under model (1), and generates the given bound. Informally, this means that more information on the parameters of interest cannot be obtained without more data or further assumptions.

# 3 Identification Analysis

This section develops the theory of identification that underlies propositions 2 and 1, in a setting with explicit covariates.

In addition to the discrete random variables $Y_i$ and $Z_i$, the researcher observes case characteristics $X_i$. $X_i$ is a random vector that takes values in finite set $\mathcal{X}$. In the empirical framework of section 5, these characteristics will be the case's court, field of law and time period of assignment. I reformulate Assumptions IV and PMF as:

**Assumption IVx.** *$Z_i$ is statistically independent of $Y_i(0)$, conditional on $X_i$.*

**Assumption PMFx.** *The probability mass function of $Y_i(0) \mid X_i = x$ is known, for all $x \in \mathcal{X}$.*

Note that if $X_i$ is degenerate, then Assumptions IVx and PMFx are identical to Assumptions IV and PMF.

The joint distribution of $\big(Y_i(0), Y_i(1), S_i, Z_i\big)$, conditional on covariates $X_i$, is the cornerstone of the identification analysis, for three reasons. First, the available data, i.e. the probability mass function of $(Y_i, Z_i, X_i)$, constitute restrictions on this distribution, under equation (1). Second, assumptions IVx and PMFx can be reformulated as restrictions on this distribution. Finally, any feature of the joint distribution of the data that we do not observe, any parameter, can be seen as a function of this distribution. Let $\mathcal{F}$ denote the set of probability mass functions of $\big(Y_i(0), Y_i(1), S_i, Z_i\big)$ conditional on $X_i$. $f$ denotes a typical element of $\mathcal{F}$, and $f(y_0,\, y_1,\, s,\, z \mid x)$ denotes a typical value of $f$.

I proceed in two steps. First, I obtain the restrictions imposed by our data and assumptions on the primitive conditional distribution, $f$, to define its identified set. Then, I define the identified sets for the parameters of interest and characterize them.

## 3.1 Identified set for $f$

The identified set for $f$ is the set of all distributions in $\mathcal{F}$ that are observationally equivalent under model (1), and are consistent with assumptions IVx and PMFx. $f \in \mathcal{F}$ satisfies observational equivalence under model (1) if:

$$\sum_{y_0,y_1,s} 1\big\{sy_1 + (1-s)y_0 = y\big\}\, f\big(y_0,\, y_1,\, s,\, z \,|\, x\big) = \Pr(Y_i = y,\, Z_i = z \,|\, X_i = x) \qquad \forall\, y, z, x. \tag{$R_{OE}$}$$

In other words, $f$ is observationally equivalent whenever its implied distribution of $(Y_i, Z_i) \,|\, X_i$ under model (1) matches that which is observed. Next, any $f$ that is observationally equivalent is consistent with Assumption IVx if:

$$\sum_{y_1,s} f(y_0,\, y_1,\, s,\, z \,|\, x) \;=\; \Pr(Z_i = z \,|\, X_i = x) \sum_{y_1,s,\tilde{z}} f(y_0, y_1, s, \tilde{z} \,|\, x). \qquad \forall\, y_0, z, x. \tag{$R_{IV}$}$$

That is, $f$ is consistent with Assumption IVx if its implied distribution of $(Y_i(0), Z_i) \,|\, X_i = x$ equals the product of the implied marginal distributions. Notice that the implied distribution of $Z_i \,|\, X_i$ equals the observed distribution by observational equivalence. Finally, $f \in \mathcal{F}$ is consistent with Assumption PMFx if:

$$\sum_{y_1,s,z} f(y_0,\, y_1,\, s,\, z \,|\, x) \;=\; \Pr(Y_i(0) = y_0 \,|\, X_i = x) \qquad \forall\, y_0, x, \tag{$R_{PMF}$}$$

where $\Pr(Y_i(0) = y_0 \,|\, X_i = x)$ is known, for all $y_0 \in \{1, \ldots, n_Y\}$ and $x \in \mathcal{X}$.

Assumptions IVx and PMFx are associated with identified sets $\mathcal{F}^*_{\text{IV}}$ and $\mathcal{F}^*_{\text{PMF}}$, respectively, where

$$\begin{aligned}
\mathcal{F}^*_{\text{IV}} \quad &\equiv \quad \big\{f \in \mathcal{F} \;:\; f \text{ satisfies restrictions } (R_{OE}) \text{ and } (R_{IV})\big\} \quad \text{and} \\
\mathcal{F}^*_{\text{PMF}} \quad &\equiv \quad \big\{f \in \mathcal{F} \;:\; f \text{ satisfies restrictions } (R_{OE}) \text{ and } (R_{PMF})\big\}.
\end{aligned}$$

The case where both Assumptions IVx and PMFx are imposed need not be treated separately. Under both assumptions, the distribution of $Y_i(0) \,|\, X_i, Z_i$ equals that of $Y_i(0) \,|\, X_i$, which is known. Hence, both assumptions can be cast as Assumption PMFx with covariates $\widetilde{X}_i = (X_i, Z_i)$.

### 3.2 Identified sets for parameters of interest

I cast parameters as linear functions of distributions in $\mathcal{F}$, $\theta : \mathcal{F} \mapsto \mathbb{R}^{d_\theta}$, where $d_\theta$ is the dimensionality of parameter $\theta$. Each parameter $\theta = (\theta_1, \ldots, \theta_{d_\theta})$ that I consider is associated with $d_\theta$ vectors of known non-negative coefficients $c = (c_1, \ldots, c_{d_\theta})$, so that

$$\theta(f;\, c) \quad \equiv \quad \begin{pmatrix} \sum_{y_0, y_1, s, z, x} c_1(y_0, y_1,\, s,\, z,\, x)\; f(y_0, y_1,\, s,\, z \,|\, x) \\ \vdots \\ \sum_{y_0, y_1, s, z, x} c_{d_\theta}(y_0, y_1,\, s,\, z,\, x)\; f(y_0, y_1,\, s,\, z \,|\, x) \end{pmatrix}.$$

When $\theta$ is scalar, $\theta(f;\, c) \;\equiv\; \sum_{y_0, y_1, s, z, x} c(y_0, y_1,\, s,\, z,\, x)\; f(y_0, y_1,\, s,\, z \,|\, x)$. The identified set for parameter $\theta(\cdot\,;\, c)$ is the set of parameter values that are associated with distributions that belong to the identified set for $f$:

$$\begin{aligned} \Theta^*_{\text{IV}}(c) \quad &\equiv \quad \{\theta(f\,;\, c) : f \in \mathcal{F}^*_{\text{IV}}\} \quad \text{and} \\ \Theta^*_{\text{PMF}}(c) \quad &\equiv \quad \{\theta(f\,;\, c) : f \in \mathcal{F}^*_{\text{PMF}}\}. \end{aligned}$$

Table 1 shows that all of our parameters of interest are linear and presents the associated vectors of coefficients.[4]

[4]In fact, linear parameters are widespread. See, e.g. Mogstad, Santos, and Torgovitsky (2018). For example, the expectation of counterfactual outcome $Y_i(1)$ is the linear parameter associated with coefficients $c^1$, where $c^1(y_0, y_1, s, z, x) = y_1$; the "Average Treatment Effect" — the average difference between $Y_i(1)$ and $Y_i(0)$ — is the linear parameter associated with $c^{ATE}$, where $c^{ATE}(y_0, y_1, s, z, x) = y_1 - y_0$; the probability that $Y_i(1)$ (or $Y_i(0)$) equals a given $y \in \mathcal{Y}$ is also a linear parameter. Moreover, the versions of these parameters that condition on $X_i = x$ or $Y_i = y$ are also linear.

| Description | Parameter of Interest | $c(y_0, y_1, s, z, x)$ |
|---|---|---|
| Rate of Irregular Assignments | $\Pr(S_i = 1)$ | $\Pr(X_i = x) \cdot 1\{s = 1\}$ |
| Judge $y^*$'s Rate of Irregular Assignments | $\Pr(S_i = 1 \mid Y_i = y^*)$ | $\frac{\Pr(X_i = x)}{\Pr(Y_i = y^*)} \cdot 1\{s = 1\} \cdot 1\{y_1 = y^*\}$ |
| Rate of Irregular Assignments, given $X_i = x_0$ | $\Pr(S_i = 1 \mid X_i = x_0)$ | $1\{x = x_0\} \cdot 1\{s = 1\}$ |
| Rate of Irregular Assignments, given $X_i \in \mathcal{X}_0$ | $\Pr(S_i = 1 \mid X_i \in \mathcal{X}_0 \subseteq \mathcal{X})$ | $\frac{\Pr(X_i = x)}{\Pr(X_i \in \mathcal{X}_0)} \cdot 1\{x \in \mathcal{X}_0\} \cdot 1\{s = 1\}$ |

Table 1: Coefficients of the Linear Parameters of Interest

I now turn to the characterization, or computation, of identified sets. Notice first that $\mathcal{F}^*_{\text{IV}}$ and $\mathcal{F}^*_{\text{PMF}}$ are convex sets: the convex combination of any two elements of $\mathcal{F}^*_{\text{IV}}$ (or $\mathcal{F}^*_{\text{PMF}}$) is a well-defined probability mass function that also satisfies restrictions ($\text{R}_{\text{OE}}$) and ($\text{R}_{\text{IV}}$) (or ($\text{R}_{\text{PMF}}$)). It is well defined because $\mathcal{F}$, the set of probability mass functions of $\big(Y_i(0), Y_i(1), S_i, Z_i\big)$ conditional on $X_i$, is convex. It satisfies these restrictions because the solution set to ($\text{R}_{\text{OE}}$) and ($\text{R}_{\text{IV}}$) (or ($\text{R}_{\text{PMF}}$)) is convex, which follows from the fact that these restrictions are linear equations in $f$.

Now, fix non-negative coefficients $c$ and consider parameter $\theta(\cdot\,;\, c)$. Its identified sets, $\Theta^*_{\text{IV}}(c)$ and $\Theta^*_{\text{PMF}}(c)$, are also convex. In particular, let $f_1, f_2 \in \mathcal{F}^*_{\text{IV}}$. For a given $\lambda \in [0, 1]$, $\lambda f_1 + (1 - \lambda) f_2 \in \mathcal{F}^*_{\text{IV}}$ and

$$\lambda \underbrace{\theta(f_1;\, c)}_{\in \Theta^*_{\text{IV}}(c)} + (1 - \lambda) \underbrace{\theta(f_2;\, c)}_{\in \Theta^*_{\text{IV}}(c)} \quad = \quad \theta\big(\lambda f_1 + (1 - \lambda) f_2;\ c\big) \ \in \ \Theta^*_{\text{IV}}(c).$$

Thus, the identified set for a scalar and linear parameter under either Assumption IVx or PMFx equals an interval in $\mathbb{R}^+$. What is left to determine are the two extreme points of this interval, also known as the *sharp bounds*. But this is straightforward: since the parameter and the restrictions are linear, the extreme points of this interval equal the solution to two linear programming problems that minimize/maximize the parameter value subject to restrictions ($R_{OE}$), ($R_{IV}$) and ($R_{PMF}$). That is, given a vector of non-negative coefficients $c$, $\Theta^*_{IV}(c) = \left[\underline{\theta}_{IV}(c), \overline{\theta}_{IV}(c)\right]$, where

$$\begin{aligned}
\underline{\theta}_{IV}(c) &= \min_{f \in \mathcal{F}} \quad \theta(f;\, c) \quad \text{subject to } (R_{OE}) \text{ and } (R_{IV}) \\
\overline{\theta}_{IV}(c) &= \max_{f \in \mathcal{F}} \quad \theta(f;\, c) \quad \text{subject to } (R_{OE}) \text{ and } (R_{IV}),
\end{aligned}$$

and $\underline{\theta}_{PMF}(c)$ and $\overline{\theta}_{PMF}(c)$ are defined analogously.

For our parameters of interest, listed in Table 1, these linear programs either have closed-form solutions or simpler formulations. Table 2 lists the results for parameters $\Pr(S_i = 1 \,|\, X_i = x)$ and $\Pr(S_i = 1 \,|\, Y_i = y^*, X_i = x)$ and Appendix C proves them. Sharp lower bounds for more aggregate parameters such as $\Pr(S_i = 1)$ or $\Pr(S_i = 1 \,|\, Y_i = y^*)$ can be obtained from the lower bounds listed in Table 2 through appropriate aggregation. For example, the lower bound for $\Pr(S_i = 1)$ under Assumption PMFx is:

$$\sum_{x \in \mathcal{X}} \Pr(X_i = x) \left( \frac{1}{2} \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \,|\, X_i = x) \; - \; \Pr(Y_i(0) = y \,|\, X_i = x) \Big| \right).$$

# 4 Context and Data

This section gives an overview of Ecuador's judicial system, discusses the existing regulations on the assignment of cases to judges, and presents the assignment data.

| Assumption | Parameter of Interest | $\underline{\theta}(c)$ | $\overline{\theta}(c)$ |
|---|---|---|---|
| IVx | $\Pr(S_i = 1 \mid X_i = x)$ | $\min_{\phi \in \Phi} \sum_{z,y} \frac{1}{2} \Pr(Z_i = z \mid x) \Big\lvert \Pr(Y_i = y \mid x, z) \; - \; \phi(y \mid x) \Big\rvert,$ where $\Phi$ is the set of p.m.f.s of $Y_i(0) \mid X_i$. | 1 |
| IVx | $\Pr(S_i = 1 \mid Y_i = y^*, X_i = x)$ | 0 | 1 |
| PMFx | $\Pr(S_i = 1 \mid X_i = x)$ | $\frac{1}{2} \sum_y \Big\lvert \Pr(Y_i = y \mid x) \; - \; \Pr(Y_i(0) = y \mid x) \Big\rvert$ | 1 |
| PMFx | $\Pr(S_i = 1 \mid Y_i = y^*, X_i = x)$ | $\max \left\{ 0, \; \frac{\Pr(Y_i = y^* \mid x) - \Pr(Y_i(0) = y^* \mid x)}{\Pr(Y_i = y^* \mid x)} \right\}$ | 1 |

Table 2: Sharp Bounds on the Parameters of Interest, conditional on $X_i = x$.

## 4.1 Context

Unlike federal states, such as Brazil, Mexico, or the United States, Ecuadorian law is homogeneous across its administrative divisions. Ecuador's judiciary has a 3-tiered judiciary, composed of 331 district courts, 24 provincial courts, the National Court of Justice, and a governing body called the Judicial Council. In this paper, I focus on case assignments to judges in the country's district courts.

Table 3 presents the key institutional components that govern the assignment of cases to judges. Lottery offices deployed throughout the country perform assignments. Personnel attached to these offices use a dedicated computer program to draw assignments. In the event of a power outage or any other circumstance where the computer program is not accessible, the personnel draw cases that await assignment sequentially at random and assign them to available judges, who have been arranged in a pre-defined order.[5] Ecuadorian regulations leave the precise implementation of the computer program to the Judicial Council. In a recent interview, however, the president of the Judicial Council briefly explains the implementation: the computer program assigns cases at random among available judges who have a relatively low case workload at the time of assignment. Finally, judges who are available for a given case must work in courts that have competence over the case's field of law[6] and location.[7]

Ecuador offers an ideal setting to study irregular assignments of cases to judges, for two reasons. First, this topic is salient and raises concerns among public officials in the Judicial Council, and among the general public. In recent months, several case assignment scandals have surfaced which involve judges in courts across the country as well as high profile individuals, such as the mayor of Quito, the country's capital, who was recently removed from office.[8]

Second, large scale access to case-level assignment information across the country's courts is possible for non-confidential cases, and this information is regularly updated by the Judicial

[5]This procedure dates from 2004, when assignments were still being performed manually in some Ecuadorian provinces. Since 2013, all case assignments are computer-based by default.

[6]Each district court has competence over cases that belong to a subset of the following fields of law: criminal law (e.g. a homicide), civil law (e.g. a payment dispute that involves a bank and a credit card debtor), administrative law (e.g. a dispute related with a government contract), tax law (e.g. a tax payment dispute), juvenile law (e.g. a robbery conducted by someone under 18 years of age), transit law (e.g. drunk driving), family violence law (e.g. a case of household violence), family law (e.g. a divorce), labor law (e.g. wrongful termination of an employee) and landlord-tenant law.

[7]The Judicial Council specifies the territory associated with each court. In general, the location of criminal cases is the location where the alleged crime was committed and the location of other cases is the address of the defendant. See article 404 of Código Orgánico Integral Penal 2014, which contains further rules to obtain the jurisdiction if the location of the crime is unknown, and articles 9-15 of Código Orgánico General de Procesos 2015.

[8]See the media coverage here, here, here, and here.

Council.[9] In Latin America, this is exceptional: case-level assignment data is scattered across different judiciaries in federal states such as Mexico or Brazil, and large scale access to case assignment information is effectively denied to the general public in countries such as Argentina, Chile, Colombia, Mexico or Peru.

## 4.2 Data

My data is a collection of lottery certificates that record individual judicial case assignments to judges. I source this data from `http://consultas.funcionjudicial.gob.ec/informacionjudicial/public/informacion.jsf`, a website that is maintained by the judicial regulator of Ecuador, the *Consejo de la Judicatura*. This website makes available to the public the government's unique database of judicial cases, called *Sistema Automático de Trámite Judicial Ecuatoriano*.[10]

Every judicial case in the country is given a unique identifier at the time of filing that consists of a two digit number that is associated with each of Ecuador's twenty-four provinces, a three digit number used by the Judicial Council for internal purposes, the four-digit year, and a consecutive number. My data collection exercise requested the information on file for every possible judicial case unique identifier over a period of two months in 2021. For each successful request, I obtained a plain text *.html* file that contained the case's lottery certificate. I then extracted the lottery certificate from this file. For each certificate, I

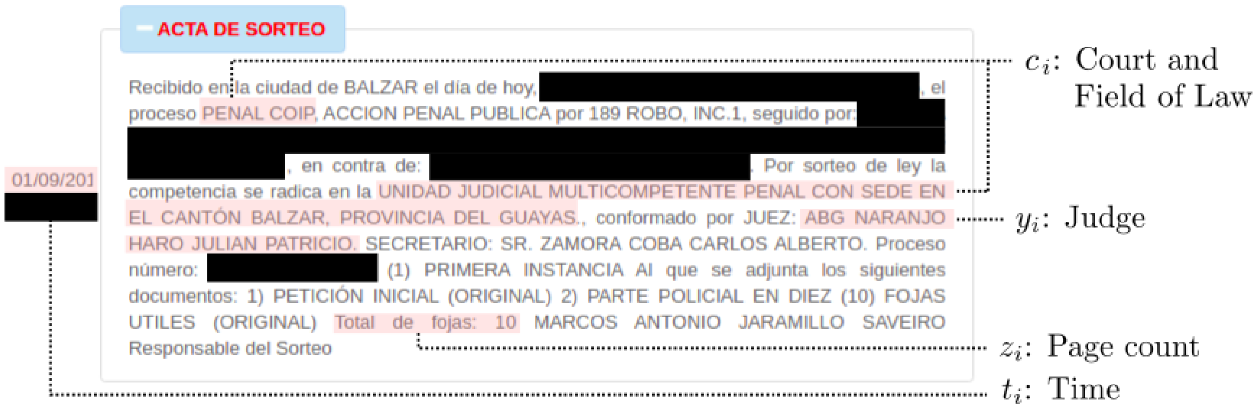
[9]Confidential cases are those that involve sexual crimes, family violence, and crimes against the state. See article 562 of Código Orgánico Integral Penal (2014). Crimes against the state are listed in arts. 336-365. They include rebellion, insubordination of military and police personnel, sabotage, treason, espionage, non-authorized possession of firearms and arms dealing.

[10]See Machasilla, Mejía, and Torres Feraud (2020) for a description of this database, and articles $118-119$ in Código Orgánico General de Procesos (2015) and $578-579$ in Código Orgánico Integral Penal (2014) for the legal content requirements of this database.

| Regulation | Original text | Source |
|---|---|---|
| The use of the automatic system for case lotteries is compulsory in all districts that have the technological facilities and the system installed. | En los distritos que cuentan con las facilidades tecnológicas y se encuentre instalado el sistema automático de sorteo de causas para primera y segunda instancia, su uso será obligatorio. | Article 9, Reglamento de Sorteo de Juicios (2004) |
| Districts that do not have the system installed will perform lotteries as follows: after numbering the cases, one ticket for each case is inserted in a container. Tickets are then randomly drawn and determine the judge that the case must be assigned to. | En los distritos o lugares carentes del sistema informático para el sorteo éste tendrá el procedimiento siguiente: Numeradas las demandas o expedientes con arreglo en un recipiente apropiado se colocarán tantas fichas cuantas sean aquellos. Estas fichas se sacarán por la suerte y determinarán a los jueces que deben conocer de las causas. | Article 11, Reglamento de Sorteo de Juicios (2004) |
| The algorithm of the system assigns cases to judges randomly, according to the judges' case workloads. That is, if we have five judges and each judge has a case workload of ten, then (the system) assigns randomly. But if one of them has a case workload of one hundred, then the system skips that judge, because she has too high a workload | El algoritmo del sistema asigna de manera aleatoria las causas segun la carga procesal que tenga un juzgador. Es decir, si tenemos cinco juzgadores, los cinco tienen carga procesal de diez, entonces va asignando aleatoriamente. Pero si a uno de ellos se le pone una carga procesal de cien causas en trámite, entonces el sistema se salta ese juzgador porque tiene muchas causas en trámite | Minutes 5:48 – 6:30 of an interview with the President of the Judicial Council, available here. |

Table 3: Ecuadorian Case Assignment Regulations

*Note:* English translations are my own.


ACTA DE SORTEO

01/09/201 [redacted]

Recibido en la ciudad de BALZAR el día de hoy, [redacted], el proceso PENAL COIP, ACCION PENAL PUBLICA por 189 ROBO, INC.1, seguido por: [redacted], en contra de: [redacted]. Por sorteo de ley la competencia se radica en la UNIDAD JUDICIAL MULTICOMPETENTE PENAL CON SEDE EN EL CANTÓN BALZAR, PROVINCIA DEL GUAYAS., conformado por JUEZ: ABG NARANJO HARO JULIAN PATRICIO. SECRETARIO: SR. ZAMORA COBA CARLOS ALBERTO. Proceso número: [redacted] (1) PRIMERA INSTANCIA Al que se adjunta los siguientes documentos: 1) PETICIÓN INICIAL (ORIGINAL) 2) PARTE POLICIAL EN DIEZ (10) FOJAS UTILES (ORIGINAL) Total de fojas: 10 MARCOS ANTONIO JARAMILLO SAVEIRO Responsable del Sorteo



Figure 1: An annotated lottery certificate

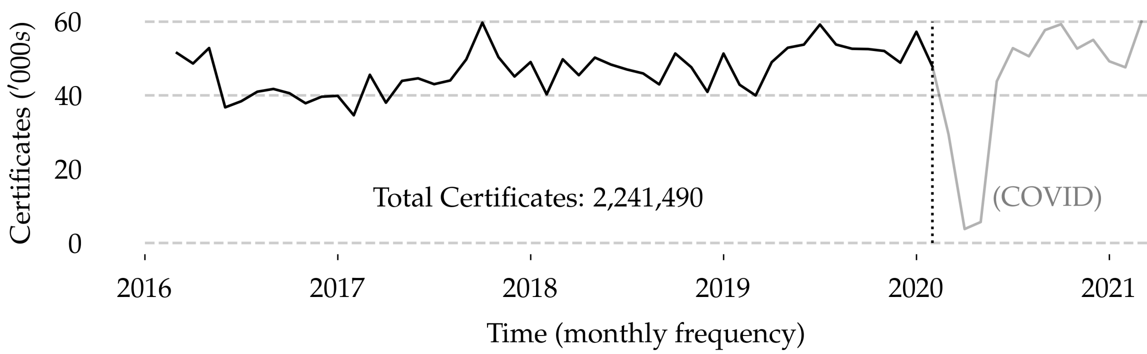



Figure 2: Lottery certificates over time

extracted and processed the date when the assignment is produced, the case's reported field of law, the court where the case is assigned, the judge that the case is assigned to, and the amount of paperwork that the plaintiff or prosecutor submits when the case is filed. Figure 1 depicts a lottery certificate, as it appears in the government's webpage.[11]

My data collection exercise returns over 2 million lottery certificates that record assignments made in district courts between March, 2016 and February, 2020. I chose the beginning of my sample period for practical reasons: before this time, lottery certificates come in a vast array of formats, which makes the construction of accurate text processing programs

[11]The black regions conceal the case's identifiable information.

| | Percentiles | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 5 | 10 | 25 | 50 | 75 | 90 | 95 | observations |
| Certificates per Judge | 10 | 72 | 582 | 1213 | 1886 | 2937 | 3827 | 1568 |
| Certificates per Court | 17 | 152 | 1408 | 3118 | 6025 | 16348 | 31815 | 331 |
| Plaintiff Paperwork (number of pages) | 0 | 0 | 1 | 5 | 12 | 26 | 44 | $2,023,010$ |

Table 4: Summary Statistics

a daunting task. I chose to end my sample period before the onset of the SARS-CoV-2 pandemic, which had a sizeable impact on the judiciary's activities, as Figure 2 shows.

# 5 Empirical Framework

This section extends the within-court model of case assignments of section 2.1 to include assignments made in every Ecuadorian district court, between March, 2016 and February, 2020 and discusses statistical inference.

## 5.1 Setting

I index lottery certificates – loosely referred to as judicial cases – with $i \in \{1, \ldots, n\}$. For each case, I observe the case's field of law and the court where it is assigned, $C_i$, as well as the date and time of assignment, $T_i$. The judge that the case is assigned to is $Y_i$, a discrete random variable that takes values in $\mathcal{Y} = \{1, \ldots, n_Y\}$, the set of all judges who I observe received at least one assignment in district courts between March of 2016 and February of 2020. Finally, I observe the page count of paperwork reported by the case's plaintiff or prosecutor when she filed the case and before the case was assigned to a judge. $Z_i$ measures this characteristic and takes values in finite set $\mathcal{Z}$. Hence, I view the available collection of lottery certificates as a realization of the collection of random vectors, $\{(Y_1, Z_1, C_1, T_1), \ldots, (Y_n, Z_n, C_n, T_n)\}$, which I assume to be identically distributed.

The recent case assignment scandals and my conversations with Judicial Council officials reveal a variety of ways that give rise to irregular assignments. Personnel who work in the case assignment offices may manipulate the Judicial Council's computer program that generates assignments in order to direct assignments. Third parties may infiltrate the Judicial Council's computer system and direct assignments. Judges may call in sick in order to avoid being assigned to specific cases. The cases' locations can be manipulated so as to target certain courts and the cases' fields of law can be misrepresented so as to target certain judges.

Thus, irregular assignments involve manipulations of the judge that is selected in the case's court and time of assignment, or they involve manipulations of the case's court and time of assignment. Because my analysis is made conditional on the cases' court and time of assignment, I exclusively detect the former class of manipulations. I therefore interpret $S_i$ to indicate if the assignment of case $i$ involved manipulations of the judge that was selected, within its court and time of assignment. This distinction is implicit in the econometric model of section 2, which is a model of case assignments in a given court. As in section 2, $Y_i(0)$ denotes the case's counterfactual regulatory assignment, and $Y_i(1)$ denotes its counterfactual irregular assignment. They relate with actual assignments according to the potential outcomes model (1).

## 5.2 Assumptions

I consider two distinct identification assumptions that can be imposed separately: that regulatory assignments are uniformly distributed among the set of competent judges, and that case characteristic $Z_i$ is a one-sided instrument.

At any given time and court, a judge is competent for case assignments if she *should* be available for assignments. Appointed judges need not be competent. At times, they may be on a legitimate medical leave or on vacation, for example. Competent judges need not be available: a judge that takes a medical leave so as to avoid a specific case should have been available when the case was assigned. Let $\mathcal{J}_{ct}$ be the set of competent judges in court and field of law $c$ at time $t$. Judges belong to $\mathcal{J}_{ct}$ if they receive cases in court $c$ during time period $t$ and if they have a relatively low workload compared with their peers during this time period. In practice, I select the criteria to measure $\mathcal{J}_{ct}$ so as to obtain conservative estimates of irregular assignments. Appendix section A discusses how I measure this set in more detail. Define $\mathcal{T}_i^J$ as the largest time interval that contains case $i$'s time of assignment, $T_i$, and features the same set of competent judges in $C_i$, i.e. the joint judge spell when case $i$ was assigned. Concretely, this spell is $\mathcal{T}_i^J = \mathcal{T}^J(C_i, T_i)$, where $\mathcal{T}^J(c,t) = \left[\underline{t},\ \bar{t}\right]$ and $\underline{t}, \bar{t}$ are such that (i) $t \in \left[\underline{t},\ \bar{t}\right]$ and (ii) $\bar{t} - \underline{t} \ = \ \sup\left\{\bar{\tau} - \underline{\tau} :\ t \in \left[\underline{\tau},\ \bar{\tau}\right] \text{ and } \mathcal{J}_{ct} = \mathcal{J}_{c\tau} \text{ for all } \tau \in \left[\underline{\tau},\ \bar{\tau}\right]\right\}$.[12] Assumption PMFe states that cases' regulatory assignments are uniformly distributed, conditional on the case's court and joint judge spell.

**Assumption PMFe.** $Y_i(0) \,|\, C_i = c, \mathcal{T}_i^J = \mathbf{t} \ \sim \ Unif(\mathcal{J}_{ct})$ *where* $t \in \mathbf{t}$*, for all* $c$ *and* $\mathbf{t}$*.*

Notice that Assumption PMFe corresponds with Assumption PMFx with covariates $C_i$ and $\mathcal{T}_i^J$. Two features of the institutional setting, listed in Table 3, motivate Assumption PMFe. First, the case assignment procedure that was in place before the implementation

[12] $\underline{t} = \underline{t}(c,t)$ and $\bar{t} = \bar{t}(c,t)$. I omit these arguments for conciseness. Moreover, notice that $\underline{t}$ and $\bar{t}$ are uniquely defined. Let $\underline{t}_1, \bar{t}_1$ and $\underline{t}_2, \bar{t}_2$ satisfy (i) and (ii). Since they satisfy (i), $\left[\underline{t}_1, \bar{t}_1\right] \cap \left[\underline{t}_2, \bar{t}_2\right] \neq \varnothing$. Define $\underline{t}^* = \min\{\underline{t}_1, \underline{t}_2\}$ and $\bar{t}^* = \max\{\bar{t}_1, \bar{t}_2\}$. Because $\left[\underline{t}^*, \bar{t}^*\right] = \left[\underline{t}_1, \bar{t}_1\right] \cup \left[\underline{t}_2, \bar{t}_2\right]$,

$$\sup\left\{ \bar{\tau} - \underline{\tau} \ :\quad t \in \left[\underline{\tau},\ \bar{\tau}\right] \quad \text{and} \quad \mathcal{J}_{ct} = \mathcal{J}_{c\tau} \text{ for all } \tau \in \left[\underline{\tau},\ \bar{\tau}\right]\right\} \ = \ \bar{t}^* - \underline{t}^* \ \geq \ \max\left\{\bar{t}_1 - \underline{t}_1,\ \bar{t}_2 - \underline{t}_2\right\}.$$

Thus, $\underline{t}_1 = \underline{t}_2 = \underline{t}^*$ and $\bar{t}_1 = \bar{t}_2 = \bar{t}^*$.

of the Judicial Council's computer system, and is still in place when the computer system is out of order implies that cases be assigned to competent judges with equal probabilities. Second, the President of the Judicial Council confirmed in a recent public interview that the computer system draws assignments randomly.

The second assumption that I consider, Assumption IVe, views the case's amount of plaintiff or prosecutor paperwork as a one-sided instrument, conditional on the case's court and time period of assignment. Let $\mathcal{T}_i$ denote the quarter-year when case $i$ is assigned.

**Assumption IVe.** *$Z_i$ is statistically independent of $Y_i(0)$, conditional on $(C_i, \mathcal{T}_i)$.*

Assumption IVe corresponds with Assumption IVx with covariates $C_i$ and $\mathcal{T}_i$. The motivation for this assumption is that, in Ecuador's district courts, cases with different amounts of plaintiff or prosecutor paperwork should not be assigned to judges differently. The threat to Assumption IVe are fluctuations in the composition of plaintiff or prosecutor paperwork within court, field of law and quarter-year, whereby judges who are active at times when a large share of incoming cases have extensive plaintiff paperwork would receive cases with more plaintiff paperwork than others. Assumption IVe rules out such compositional variation.

## 5.3 Statistical Analysis

I now discuss the measurement of the sharp bounds on the rates of irregular assignments in the Ecuadorian context with the finite amount of available case assignment information, $\{(Y_1, Z_1, C_1, T_1), \ldots, (Y_n, Z_n, C_n, T_n)\}$.

### 5.3.1 Detecting irregular assignments under Assumption PMFe

Under Assumption PMFe, I obtain the bounds on the rates of irregular assignments from Table 2. The rate of irregular assignments within the set of courts $\mathcal{C}_0$ and time periods $\mathcal{T}_0$ satisfies $\Pr(S_i = 1 \,|\, C_i \in \mathcal{C}_0, \mathcal{T}_i \in \mathcal{T}_0) \in [\mathrm{LB}_{\mathrm{PMF}}(\mathcal{C}_0, \mathcal{T}_0), 1]$ , where

$$\begin{aligned}
\mathrm{LB}_{\mathrm{PMF}}(\mathcal{C}_0, \mathcal{T}_0) \quad &\equiv \quad \sum_{c,\mathbf{t}} \Pr(C_i = c, \mathcal{T}_i^J = \mathbf{t} \,|\, C_i \in \mathcal{C}_0, \mathcal{T}_i^J \in \mathcal{T}_0) \;\cdot\; \mathrm{LB}_{\mathrm{PMF}}(c, \mathbf{t}) \\
\mathrm{LB}_{\mathrm{PMF}}(c, \mathbf{t}) \quad &\equiv \quad \frac{1}{2} \sum_{y \in \mathcal{Y}} \left| \Pr(Y_i = y \,|\, C_i = c, \mathcal{T}_i^J = \mathbf{t}) \;-\; \frac{1\{y \in \mathcal{J}_{c\mathbf{t}}\}}{\#\mathcal{J}_{c\mathbf{t}}} \right|,
\end{aligned}$$

and judge $y^*$'s rate of irregular assignments ($y^* \in \mathcal{Y}$) satisfies $\Pr(S_i = 1 \,|\, Y_i = y) \in [\mathrm{LB}_y, 1]$, where

$$\begin{aligned}
\mathrm{LB}_y \quad &\equiv \quad \sum_{c,\mathbf{t}} \frac{\Pr(C_i = c, \mathcal{T}_i^J = \mathbf{t} \,|\, Y_i = y)}{\Pr(Y_i = y \,|\, C_i = c, \mathcal{T}_i^J = \mathbf{t})} \cdot \mathrm{LB}_y(c, \mathbf{t}) \\
\mathrm{LB}_y(c, \mathbf{t}) \quad &\equiv \quad \max \left\{ 0, \quad \Pr(Y_i = y \,|\, C_i = c, \mathcal{T}_i^J = \mathbf{t}) - \frac{1\{y \in \mathcal{J}_{c\mathbf{t}}\}}{\#\mathcal{J}_{c\mathbf{t}}} \right\}
\end{aligned}$$

**Measurement of $\mathbf{LB_{PMF}}(\mathcal{C}_0, \mathcal{T}_0)$.** Consider the empirical counterpart of $\mathrm{LB}_{\mathrm{PMF}}(c, \mathbf{t})$:

$$\widehat{\mathrm{LB}}_{\mathrm{PMF}}(c, \mathbf{t}) \quad \equiv \quad \frac{1}{2} \sum_{y \in \mathcal{Y}} \left| \widehat{\Pr}(Y_i = y \,|\, C_i = c, \mathcal{T}_i^J = \mathbf{t}) \;-\; \frac{1\{y \in \mathcal{J}_{c\mathbf{t}}\}}{\#\mathcal{J}_{c\mathbf{t}}} \right|,$$

where $\widehat{\Pr}(Y_i = y \,|\, C_i = c, \mathcal{T}_i^J = \mathbf{t}) \equiv \frac{\sum_{i=1}^n 1\{Y_i = y, C_i = c, \mathcal{T}_i^J = \mathbf{t}\}}{\sum_{i=1}^n 1\{C_i = c, \mathcal{T}_i^J = \mathbf{t}\}}$. Unfortunately, $\widehat{\mathrm{LB}}_{\mathrm{PMF}}(c, \mathbf{t})$ is not an unbiased estimator of $\mathrm{LB}_{\mathrm{PMF}}(c, \mathbf{t})$. Indeed, $\widehat{\mathrm{LB}}_{\mathrm{PMF}}(c, \mathbf{t}) > 0$ if $Y_i = Y_i(0)$ (almost surely). To obtain a consistent and mean-conservative estimator of $\mathrm{LB}_{\mathrm{PMF}}(c, \mathbf{t})$ and a conservative confidence interval, I mirror Pouliot et al. (2022) in their derivations of statistical properties of the distributionally-robust optimal transport estimator for Wasserstein distances.

The central element in the construction of these quantities is the triangle inequality.

Concretely, let $\Delta(c,\mathbf{t}) \equiv \frac{1}{2} \sum_{y\in\mathcal{Y}} \left| \widehat{\Pr}(Y_i = y \,|\, c, \mathbf{t}) - \Pr(Y_i = y \,|\, c, \mathbf{t}) \right|$. By the triangle inequality for $\ell_1$ distances,

$$\widehat{\mathrm{LB}}_{\mathrm{PMF}}(c,\mathbf{t}) \;\le\; \Delta(c,\mathbf{t}) + \mathrm{LB}_{\mathrm{PMF}}(c,\mathbf{t}) \qquad \text{almost surely.} \tag{6}$$

Thus, $\mathbb{E}\big[\widehat{\mathrm{LB}}_{\mathrm{PMF}}(c,\mathbf{t}) - \Delta(c,\mathbf{t})\big] \le \mathrm{LB}_{\mathrm{PMF}}(c,\mathbf{t})$ and, for all $\alpha \in (0,1)$,

$$\Pr\left(\widehat{\mathrm{LB}}_{\mathrm{PMF}}(c,\mathbf{t}) - q_{1-\alpha} \le \mathrm{LB}_{\mathrm{PMF}}(c,\mathbf{t})\right) \;\ge\; \Pr\left(\Delta(c,\mathbf{t}) \le q_{1-\alpha}\right) \;=\; 1-\alpha,$$

where $q_{1-\alpha}$ is the $1-\alpha$-quantile of the distribution of $\Delta(c,\mathbf{t})$. A mean-conservative estimator of $\mathrm{LB}_{\mathrm{PMF}}(c,\mathbf{t})$ is therefore

$$\widehat{\mathrm{LB}}^{*}_{\mathrm{PMF}}(c,\mathbf{t}) \quad\equiv\quad \widehat{\mathrm{LB}}_{\mathrm{PMF}}(c,\mathbf{t}) - \mathbb{E}\big[\Delta(c,\mathbf{t})\big],$$

since $\mathbb{E}\big[\widehat{\mathrm{LB}}^{*}_{\mathrm{PMF}}(c,\mathbf{t})\big] = \mathbb{E}\big[\widehat{\mathrm{LB}}_{\mathrm{PMF}}(c,\mathbf{t}) - \Delta(c,\mathbf{t})\big] \le \mathrm{LB}_{\mathrm{PMF}}(c,\mathbf{t})$, and a conservative confidence interval of level $\alpha \in (0,1)$ is $\big[\widehat{\mathrm{LB}}_{\mathrm{PMF}}(c,\mathbf{t}) - q_{1-\alpha},\ 1\big]$.

To determine if $\widehat{\mathrm{LB}}^{*}_{\mathrm{PMF}}(c,\mathbf{t})$ is consistent and to measure the distribution of $\Delta(c,\mathbf{t})$, suppose that $\big\{(Y_1,C_1,T_1),\ldots,(Y_n,C_n,T_n)\big\}$ is mutually independent. In this case, the law of large numbers and the continuous mapping theorem guarantee consistency of $\widehat{\mathrm{LB}}^{*}_{\mathrm{PMF}}(c,\mathbf{t})$, and the central limit theorem and the delta method imply that:

$$\sqrt{n}\Big(\widehat{\Pr}(Y_i = y \,|\, c,\mathbf{t}) - \Pr(Y_i = y \,|\, c,\mathbf{t})\Big)_{y\in\mathcal{J}_{c\mathbf{t}}} \;\xrightarrow{d}\; (V_y(c,\mathbf{t}))_{y\in\mathcal{J}_{c\mathbf{t}}} \;\sim\; N(0,\Sigma),$$

where, for all $j, k \in \mathcal{Y}$ such that $j \neq k$:

$$\begin{aligned} \Sigma_{jj} &= \frac{\Pr(Y_i = j \,|\, c, \mathbf{t}) \cdot \big[1 - \Pr(Y_i = j \,|\, c, \mathbf{t})\big]}{\Pr(C_i = c, \mathcal{T}_i^J = \mathbf{t})}, \quad \text{and} \\ \Sigma_{jk} &= \frac{-\Pr(Y_i = j \,|\, c, \mathbf{t}) \cdot \Pr(Y_i = k \,|\, c, \mathbf{t})}{\Pr(C_i = c, \mathcal{T}_i^J = \mathbf{t})}. \end{aligned}$$

Hence, by the continuous mapping theorem, $\sqrt{n}\Delta(c, \mathbf{t}) \xrightarrow{d} \frac{1}{2}\sum_{y \in \mathcal{J}_{c\mathbf{t}}} |V_y(c, \mathbf{t})|$. I obtain the asymptotic distribution of $\Delta(c, \mathbf{t})$ through simulation from $\big(\widehat{V}_y\big)_{y \in \mathcal{J}_{c\mathbf{t}}} \sim N\big(0, \widehat{\Sigma}\big)$, where $\widehat{\Sigma}$ is the sample counterpart of $\Sigma$.

Given the sets of courts $\mathcal{C}_0$ and quarters $\mathcal{T}_0$, I then estimate $\text{LB}_{\text{PMF}}(\mathcal{C}_0, \mathcal{T}_0)$ with

$$\widehat{\text{LB}}^*_{\text{PMF}}(\mathcal{C}_0, \mathcal{T}_0) \quad \equiv \quad \widehat{\text{LB}}_{\text{PMF}}(\mathcal{C}_0, \mathcal{T}_0) - \mathbb{E}\big[\Delta(\mathcal{C}_0, \mathcal{T}_0)\big],$$

where

$$\begin{aligned} \widehat{\text{LB}}_{\text{PMF}}(\mathcal{C}_0, \mathcal{T}_0) &\equiv \sum_{c, \mathbf{t}} \widehat{\Pr}(C_i = c, \mathcal{T}_i^J = \mathbf{t} \,|\, C_i \in \mathcal{C}_0, \mathcal{T}_i^J \in \mathcal{T}_0) \cdot \widehat{\text{LB}}_{\text{PMF}}(c, \mathbf{t}) \\ \Delta(\mathcal{C}_0, \mathcal{T}_0) &\equiv \sum_{c, \mathbf{t}} \widehat{\Pr}(C_i = c, \mathcal{T}_i^J = \mathbf{t} \,|\, C_i \in \mathcal{C}_0, \mathcal{T}_i^J \in \mathcal{T}_0) \cdot \Delta(c, \mathbf{t}). \end{aligned}$$

Since (6) holds for every court $c$ and quarter $\mathbf{t}$,

$$\widehat{\text{LB}}_{\text{PMF}}(\mathcal{C}_0, \mathcal{T}_0) \leq \Delta(\mathcal{C}_0, \mathcal{T}_0) + \text{LB}_{\text{PMF}}(\mathcal{C}_0, \mathcal{T}_0) \qquad \text{almost surely.}$$

Thus, $\mathbb{E}\big[\widehat{\text{LB}}^*_{\text{PMF}}(\mathcal{C}_0, \mathcal{T}_0)\big] \leq \mathbb{E}\big[\text{LB}_{\text{PMF}}(\mathcal{C}_0, \mathcal{T}_0)\big]$ and a confidence interval with level (weakly) greater than $\alpha \in (0, 1)$ is $\big[\widehat{\text{LB}}_{\text{PMF}}(\mathcal{C}_0, \mathcal{T}_0) - q_{1-\alpha},\ 1\big]$, where $q_{1-\alpha}$ is the $1 - \alpha$ quantile of the distribution of $\Delta(\mathcal{C}_0, \mathcal{T}_0)$ and

$$\sqrt{n}\Delta(\mathcal{C}_0, \mathcal{T}_0) \xrightarrow{d} \sum_{c,\mathbf{t}} \left( \Pr(C_i = c, \mathcal{T}_i^J = \mathbf{t} \,|\, C_i \in \mathcal{C}_0, \mathcal{T}_i^J \in \mathcal{T}_0) \;\cdot\; \frac{1}{2} \sum_{y \in \mathcal{J}_{c\mathbf{t}}} |V_y(c, \mathbf{t})| \right).$$

**Measurement of $\mathbf{LB}_y$.** The empirical counterpart of $\text{LB}_y(c, \mathbf{t})$,

$$\widehat{\text{LB}}_y(c, \mathbf{t}) \;\equiv\; \max\left\{0, \quad \widehat{\Pr}(Y_i = y \,|\, c, \mathbf{t}) - \frac{1\{y \in \mathcal{J}_{c\mathbf{t}}\}}{\#\mathcal{J}_{c\mathbf{t}}}\right\}$$

is likely to be a biased estimator of $\text{LB}_y$: $\max\left\{0, \quad \widehat{\Pr}(Y_i = y \,|\, c, \mathbf{t}) - \frac{1\{y \in \mathcal{J}_{c\mathbf{t}}\}}{\#\mathcal{J}_{c\mathbf{t}}}\right\}$ is positive for any finite sample, even if, for example, $Y_i = Y_i(0)$ (almost surely). To construct a point estimator of $\text{LB}_y$, define

$$\Delta_y(c, \mathbf{t}) \;\equiv\; \max\left\{0,\; \widehat{\Pr}(Y_i = y \,|\, c, \mathbf{t}) - \Pr(Y_i = y \,|\, c, \mathbf{t})\right\}.$$

Since $x \mapsto \max\{0, x\}$ is a subadditive function, it follows that:

$$\widehat{\text{LB}}_y(c, \mathbf{t}) - \Delta_y(c, \mathbf{t}) \le \text{LB}_y(c, \mathbf{t}) \quad \text{a.s.}$$

A mean-conservative estimator of $\text{LB}_y(c, \mathbf{t})$ is therefore $\widehat{\text{LB}}_y(c, \mathbf{t}) - \mathbb{E}\left[\Delta_y(c, \mathbf{t})\right]$. Thus, I estimate $\text{LB}_y$ with:

$$\widehat{\text{LB}}_y^* \quad = \quad \sum_{c,\mathbf{t}} \frac{\widehat{\Pr}(C_i = c, \mathcal{T}_i^J = \mathbf{t})}{\widehat{\Pr}(Y_i = y)} \cdot \widehat{\text{LB}}_y(c, \mathbf{t}) - \mathbb{E}\left[\sum_{c,\mathbf{t}} \frac{\widehat{\Pr}(C_i = c, \mathcal{T}_i^J = \mathbf{t})}{\widehat{\Pr}(Y_i = y)} \cdot \Delta_y(c, \mathbf{t})\right].$$

I obtain the bias correction term under the assumption that $\{(Y_1, C_1, T_1), \ldots, (Y_n, C_n, T_n)\}$ is mutually independent, so that, for every court $c$ and time period $\mathbf{t}$,

$$\sqrt{n(c, \mathbf{t})}\Delta_y(c, \mathbf{t}) \quad \xrightarrow{d} \quad \max\{0,\; V_y(c, \mathbf{t})\},$$

where $V_y(c,\mathbf{t}) \sim N\left(0,\ \frac{\Pr(Y_i=y \mid c,\mathbf{t})[1-\Pr(Y_i=y \mid c,\mathbf{t})]}{\Pr(C_i=c,\mathcal{T}_i^J=\mathbf{t})}\right)$. Hence,

$$\sqrt{n}\sum_{c,\mathbf{t}} \frac{\widehat{\Pr}(C_i=c,\mathcal{T}_i^J=\mathbf{t})}{\widehat{\Pr}(Y_i=y)} \cdot \Delta_y(c,\mathbf{t}) \quad \xrightarrow{d} \quad \sum_{c,\mathbf{t}} \frac{\Pr(C_i=c,\mathcal{T}_i^J=\mathbf{t})}{\Pr(Y_i=y)} \max\{0,\ V_y(c,\mathbf{t})\}.$$

To test if judge $y$ received irregular assignments or not, consider the following hypotheses:

$$H_0:\ \Pr(S_i=1 \mid Y_i=y)=0 \qquad \text{vs.} \qquad H_1:\ \Pr(S_i=1 \mid Y_i=y)>0. \tag{7}$$

I decide between these two hypotheses for every judge in two steps. First, given judge $y$, I propose a test statistic for (7) and obtain its asymptotic distribution. Then, I correct the p-values for the tests associated with each judge to account for multiple hypothesis testing, using the procedure from Romano and Wolf (2016).

Consider the test statistic $T_y \equiv \widehat{\Pr}(Y_i=y) - \widehat{\Pr}(Y_i(0)=y)$, obtained by replacing the unknown population probabilities in $\Pr(Y_i=y) - \Pr(Y_i(0)=y)$ with their empirical counterparts. Given independent and identically distributed case assignment information, $\left((Y_i, C_i, \mathcal{T}_i^J)\right)_{i=1}^n$, the classical central limit theorem implies that:

$$\sqrt{n}\left[\left(\widehat{\Pr}(Y_i=y), \left(\widehat{\Pr}(C_i=c,\mathcal{T}_i^J=\mathbf{t})\right)_{c,\mathbf{t}}\right) - \left(\Pr(Y_i=y), \left(\Pr(C_i=c,\mathcal{T}_i^J=\mathbf{t})\right)_{c,\mathbf{t}}\right)\right]$$
$$\xrightarrow{d} \quad N(0,\Sigma_y),$$

where $\Sigma_y$ is the variance-covariance matrix of $\left(1\{Y_i=y\}, \left(1\{C_i=c,\mathcal{T}_i^J=\mathbf{t}\}\right)_{c,\mathbf{t}}\right)$. By the continuous mapping theorem,

$$
\begin{aligned}
&\sqrt{n}\left(\widehat{\Pr}(Y_i = y) - \widehat{\Pr}(Y_i(0) = y) - [\Pr(Y_i = y) - \Pr(Y_i(0) = y)]\right) \\
&= h_y'\sqrt{n}\left[\left(\widehat{\Pr}(Y_i = y), \left(\widehat{\Pr}(C_i = c, \mathcal{T}_i^J = \mathbf{t})\right)_{c,\mathbf{t}}\right) - \left(\Pr(Y_i = y), \left(\Pr(C_i = c, \mathcal{T}_i^J = \mathbf{t})\right)_{c,\mathbf{t}}\right)\right] \\
&\xrightarrow{d} \quad N(0, h_y'\Sigma_y h_y),
\end{aligned}
$$

where $h_y = \Big(1,\ \big(-1\{y \in \mathcal{J}_{c\mathbf{t}}\}/\#\mathcal{J}_{c\mathbf{t}}\big)_{c,\mathbf{t}}\Big)$.

The null hypothesis in (7) implies that $\text{LB}_y = 0$, in which case $\Pr(Y_i = y) - \Pr(Y_i(0) = y) \leq 0$. Thus, a conservative test of size $\alpha \in (0,1)$ is obtained if we decide in favor of $H_1$ in (7) whenever the standardized test statistic exceeds $\Phi_{1-\alpha}$, the $1-\alpha$ quantile of a standard normal distribution:

$$
\begin{aligned}
\alpha &\approx \Pr\left(T_y - [\Pr(Y_i = y) - \Pr(Y_i(0) = y)] > \Phi_{1-\alpha}\sqrt{h_y'\Sigma_y h_y} \;\middle|\; H_0\right) \\
&\geq \Pr\left(T_y > \Phi_{1-\alpha}\sqrt{h_y'\Sigma_y h_y} \;\middle|\; H_0\right),
\end{aligned}
$$

and the $p$-value is given by $1 - \Phi^{-1}\big(T_y/\sqrt{h_y'\Sigma_y h_y}\big)$. In practice, I obtain the variance of the test statistic with the empirical counterparts of $h_y$ and $\Sigma_y$.

I conduct such a test for every judge. To adjust the $p$-values for multiple testing, I implement the Romano-Wolf correction developed by Romano and Wolf (2016) (see also Clarke, Romano, and Wolf (2020) for a detailed overview).

### 5.3.2 Detecting irregular assignments under Assumption IVe

Table 2 shows the lower bounds on the rates of irregular assignments under Assumption IVe. Within the set of courts $\mathcal{C}_0$ and quarters $\mathcal{T}_0$, the rate of irregular assignments satisfies

$\Pr(S_i = 1 \,|\, C_i \in \mathcal{C}_0, \mathcal{T}_i \in \mathcal{T}_0) \in [\mathrm{LB}_{\mathrm{IV}}(\mathcal{C}_0, \mathcal{T}_0), 1]$, where

$$\begin{aligned} \mathrm{LB}_{\mathrm{IV}}(\mathcal{C}_0, \mathcal{T}_0) &\equiv \sum_{c,\mathbf{t}} \Pr(C_i = c, \mathcal{T}_i = \mathbf{t} \,|\, C_i \in \mathcal{C}_0, \mathcal{T}_i \in \mathcal{T}_0) \cdot \mathrm{LB}_{\mathrm{IV}}(c, \mathbf{t}) \\ \mathrm{LB}_{\mathrm{IV}}(c, \mathbf{t}) &\equiv \min_{\phi \in \Phi} \sum_{z \in \mathcal{Z}, y \in \mathcal{Y}} \frac{\Pr(Z_i = z \,|\, c, \mathbf{t})}{2} \cdot \Big| \Pr(Y_i = y \,|\, Z_i = z, c, \mathbf{t}) - \phi(y) \Big| \end{aligned}$$

and $\Phi$ is the set of probability mass functions defined over $\mathcal{Y}$. In particular, when the one-sided instrument $Z_i$ is binary, Proposition 2 shows that $\mathrm{LB}_{\mathrm{IV}}(c, \mathbf{t})$ can be cast in terms of the $\ell_1$ distance between the two conditional distributions of assignments:

$$\mathrm{LB}_{\mathrm{IV}}(c, \mathbf{t}) = \frac{p_{\min}(c, \mathbf{t})}{2} \cdot \sum_{y \in \mathcal{Y}} \Big| \Pr(Y_i = y \,|\, Z_i = 1, c, \mathbf{t}) - \Pr(Y_i = y \,|\, Z_i = 0, c, \mathbf{t}) \Big|.$$

where $p_{\min}(c, \mathbf{t}) = \min\big\{ \Pr(Z_i = 0 \,|\, c, \mathbf{t}), \Pr(Z_i = 1 \,|\, c, \mathbf{t}) \big\}$.

**Measurement of $\mathbf{LB_{IV}}(\mathcal{C}_0, \mathcal{T}_0)$ when $Z_i$ is binary.** The derivation of a consistent and mean conservative estimator of $\mathrm{LB}_{\mathrm{IV}}(c, \mathbf{t})$ is similar to that of $\mathrm{LB}_{\mathrm{PMF}}(c, \mathbf{t})$. It is ultimately based on the triangle inequality for $\ell_1$ distances (see Pouliot et al. 2022). Let $\widehat{p}_{\min}(c, \mathbf{t})$ and $\widehat{\mathrm{LB}}_{\mathrm{IV}}(c, \mathbf{t})$ be the empirical counterparts of $p_{\min}(c, \mathbf{t})$ and $\mathrm{LB}_{\mathrm{IV}}(c, \mathbf{t})$. For $z \in \{0, 1\}$, define

$$\Delta^z(c, \mathbf{t}) \equiv \frac{\widehat{p}_{\min}(c, \mathbf{t})}{2} \sum_{y \in \mathcal{Y}} \Big| \widehat{\Pr}(Y_i = y \,|\, Z_i = z, c, \mathbf{t}) - \Pr(Y_i = y \,|\, Z_i = z, c, \mathbf{t}) \Big|.$$

Two applications of the triangle inequality for $\ell_1$ distances imply that

$$\widehat{\mathrm{LB}}_{\mathrm{IV}}(c, \mathbf{t}) \le \Delta^0(c, \mathbf{t}) + \frac{\widehat{p}_{\min}(c, \mathbf{t})}{p_{\min}(c, \mathbf{t})} \mathrm{LB}_{\mathrm{IV}}(c, \mathbf{t}) + \Delta^1(c, \mathbf{t}) \qquad \text{almost surely,} \tag{8}$$

and Jensen's inequality yields $\mathbb{E}[\widehat{p}_{\min}(c,\mathbf{t})] < p_{\min}(c,\mathbf{t})$.[13] The estimator

$$\widehat{\mathrm{LB}}^*_{\mathrm{IV}}(c,\mathbf{t}) \;\equiv\; \widehat{\mathrm{LB}}_{\mathrm{IV}}(c,\mathbf{t}) \;-\; \mathbb{E}\big[\Delta^0(c,\mathbf{t}) - \Delta^1(c,\mathbf{t})\big]$$

is therefore mean-conservative:

$$\mathbb{E}\big[\widehat{\mathrm{LB}}^*_{\mathrm{IV}}(c,\mathbf{t})\big] \;\leq\; \mathbb{E}\left[\frac{\widehat{p}_{\min}(c,\mathbf{t})}{p_{\min}(c,\mathbf{t})}\mathrm{LB}_{\mathrm{IV}}(c,\mathbf{t})\right] \;\leq\; \mathrm{LB}_{\mathrm{IV}}(c,\mathbf{t}).$$

On the other hand, (8) implies that, for all $\alpha \in (0,1)$,

$$\begin{aligned}\Pr\left(\frac{p_{\min}(c,\mathbf{t})}{\widehat{p}_{\min}(c,\mathbf{t})}\big[\widehat{\mathrm{LB}}_{\mathrm{IV}}(c,\mathbf{t}) - q_{1-\alpha}\big] < \mathrm{LB}_{\mathrm{IV}}(c,\mathbf{t})\right) &\geq \Pr\Big(\Delta^0(c,\mathbf{t}) \;+\; \Delta^1(c,\mathbf{t}) < q_{1-\alpha}\Big)\\ &= 1-\alpha,\end{aligned}$$

where $q_{1-\alpha}$ is the $1-\alpha$-quantile of the distribution of $\Delta^0(c,\mathbf{t}) + \Delta^1(c,\mathbf{t})$. Determining whether $\big[\widehat{\mathrm{LB}}_{\mathrm{IV}}(c,\mathbf{t}) - q_{1-\alpha},\, 1\big]$ is a valid confidence interval of size $\alpha$ or not is work in progress.

I obtain an asymptotic approximation to the distribution of $\Delta^0(c,\mathbf{t}) + \Delta^1(c,\mathbf{t})$ under the assumption that $\big\{(Y_1,Z_1,C_1,T_1),\ldots,(Y_n,Z_n,C_n,T_n)\big\}$ is mutually independent. In this case,

$$\sqrt{n}\Big(\Delta^0(c,\mathbf{t}) + \Delta^1(c,\mathbf{t})\Big) \;\xrightarrow{d}\; \frac{p_{\min}(c,\mathbf{t})}{2}\sum_{y\in\mathcal{Y}}\big|V^0_y(c,\mathbf{t})\big| + \big|V^1_y(c,\mathbf{t})\big|\,,$$

[13] $g : [0,1] \mapsto \mathbb{R}$, where $g(x) = \min\{x,\, 1-x\}$, is a concave function, and $\mathbb{E}\big[\widehat{\Pr}(Z_i = 0\,|\,c,\mathbf{t})\big] = \Pr(Z_i = 0\,|\,c,\mathbf{t})$, so that

$$\begin{aligned}\mathbb{E}[\widehat{p}_{\min}(c,\mathbf{t})] &= \mathbb{E}\big[\min\big\{\widehat{\Pr}(Z_i = 0\,|\,c,\mathbf{t}),\, 1-\widehat{\Pr}(Z_i = 0\,|\,c,\mathbf{t})\big\}\big]\\ &\leq \min\big\{\mathbb{E}\big[\widehat{\Pr}(Z_i = 0\,|\,c,\mathbf{t})\big],\, 1-\mathbb{E}\big[\widehat{\Pr}(Z_i = 0\,|\,c,\mathbf{t})\big]\big\} \;=\; p_{\min}(c,\mathbf{t}).\end{aligned}$$

where, for each $z \in \{0,1\}$, $\big(V_y^z\big)_{y\in\mathcal{Y}} \sim N\big(0, \Sigma^z\big)$, and for all $j, k \in \mathcal{Y}$ such that $j \neq k$,

$$\begin{aligned}\Sigma^z_{jj} &= \frac{\Pr(Y_i = j \mid Z_i = z, c, \mathbf{t})\big[1 - \Pr(Y_i = j \mid Z_i = z, c, \mathbf{t})\big]}{\Pr(Z_i = z, C_i = c, \mathcal{T}_i = \mathbf{t})} \quad \text{and} \\ \Sigma^z_{jk} &= -\frac{\Pr(Y_i = j \mid Z_i = z, c, \mathbf{t}) \Pr(Y_i = k \mid Z_i = z, c, \mathbf{t})}{\Pr(Z_i = z, C_i = c, \mathcal{T}_i = \mathbf{t})}.\end{aligned}$$

I compute the asymptotic distribution of $\Delta^0(c,\mathbf{t})+\Delta^1(c,\mathbf{t})$ by simulation, using the empirical counterparts of $p_{\min}(c,\mathbf{t})$, $\Sigma^0$ and $\Sigma^1$.

Given the sets of courts $\mathcal{C}_0$ and quarters $\mathcal{T}_0$, I estimate $\text{LB}_{\text{IV}}(\mathcal{C}_0, \mathcal{T}_0)$ with

$$\widehat{\text{LB}}^*_{\text{IV}}(\mathcal{C}_0, \mathcal{T}_0) \equiv \widehat{\text{LB}}_{\text{IV}}(\mathcal{C}_0, \mathcal{T}_0) - \mathbb{E}\big[\Delta^0(\mathcal{C}_0, \mathcal{T}_0) + \Delta^1(\mathcal{C}_0, \mathcal{T}_0)\big],$$

where

$$\begin{aligned}\widehat{\text{LB}}_{\text{IV}}(\mathcal{C}_0, \mathcal{T}_0) &\equiv \sum_{c,\mathbf{t}} \widehat{\Pr}(C_i = c, \mathcal{T}_i^J = \mathbf{t} \mid C_i \in \mathcal{C}_0, \mathcal{T}_i^J \in \mathcal{T}_0) \cdot \widehat{\text{LB}}_{\text{IV}}(c,\mathbf{t}) \\ \Delta^z(\mathcal{C}_0, \mathcal{T}_0) &\equiv \sum_{c,\mathbf{t}} \widehat{\Pr}(C_i = c, \mathcal{T}_i^J = \mathbf{t} \mid C_i \in \mathcal{C}_0, \mathcal{T}_i^J \in \mathcal{T}_0) \cdot \Delta^z(c,\mathbf{t}), \quad z \in \{0,1\}.\end{aligned}$$

Since (8) holds for every court $c$ and quarter $\mathbf{t}$,

$$\begin{aligned}&\widehat{\text{LB}}_{\text{IV}}(\mathcal{C}_0, \mathcal{T}_0) - \big[\Delta^0(\mathcal{C}_0, \mathcal{T}_0) + \Delta^1(\mathcal{C}_0, \mathcal{T}_0)\big] \\ &\quad \leq \sum_{c,\mathbf{t}} \widehat{\Pr}(C_i = c, \mathcal{T}_i^J = \mathbf{t} \mid C_i \in \mathcal{C}_0, \mathcal{T}_i^J \in \mathcal{T}_0) \frac{\widehat{p}_{\min}(c,\mathbf{t})}{p_{\min}(c,\mathbf{t})} \text{LB}_{\text{IV}}(c,\mathbf{t}) \qquad \text{almost surely.} \quad (9)\end{aligned}$$

Together with the law of iterated expectations and Jensen's inequality, (9) implies that $\mathbb{E}\big[\widehat{\text{LB}}^*_{\text{IV}}(\mathcal{C}_0, \mathcal{T}_0)\big] \leq \text{LB}_{\text{IV}}(\mathcal{C}_0, \mathcal{T}_0)$. Determining conservative confidence intervals of level $\alpha$ on the basis of (9) is work in progress. Finally, by the continuous mapping theorem,

$$\sqrt{n}\Big(\Delta^0(\mathcal{C}_0,\mathcal{T}_0)\;+\;\Delta^1(\mathcal{C}_0,\mathcal{T}_0)\Big)$$
$$\xrightarrow{d}\sum_{c,\mathbf{t}}\Bigg(\Pr(C_i=c,\mathcal{T}_i^J=\mathbf{t}\,|\,C_i\in\mathcal{C}_0,\mathcal{T}_i^J\in\mathcal{T}_0)\;\cdot\;\frac{p_{\min}(c,\mathbf{t})}{2}\sum_{y\in\mathcal{J}_{c\mathbf{t}}}\big|V_y^0(c,\mathbf{t})\big|+\big|V_y^1(c,\mathbf{t})\big|\Bigg).$$

**Measurement of $\mathbf{LB_{IV}}(\mathcal{C}_0,\mathcal{T}_0)$ when $Z_i$ is categorical.** To test if irregular assignments took place in courts $\mathcal{C}_0$ and quarters $\mathcal{T}_0$, consider the following hypotheses:

$$H_0:\ \Pr(S_i=1\,|\,\mathcal{C}_0,\mathcal{T}_0)=0\qquad\text{vs.}\qquad H_1:\ \Pr(S_i=1\,|\,\mathcal{C}_0,\mathcal{T}_0)>0. \tag{10}$$

Under the null hypothesis, $\mathrm{LB}_{\mathrm{IV}}(\mathcal{C}_0,\mathcal{T}_0)=0$. In addition, if the one-sided instrument $Z_i$ has full support within each court and quarter contained in $(\mathcal{C}_0,\mathcal{T}_0)$, then $\mathrm{LB}_{\mathrm{IV}}(\mathcal{C}_0,\mathcal{T}_0)=0$ implies that actual assignments $Y_i$ are statistically independent of $Z_i$, conditional on $\{C_i=c,\mathcal{T}_i=\mathbf{t}\}$, for all $c\in\mathcal{C}_0$ and $\mathbf{t}\in\mathcal{T}_0$. I propose to use this implication to justify the randomization hypothesis that underpins a permutation test of (10).

To introduce the test, let $W\equiv\big((Y_1,Z_1,C_1,\mathcal{T}_1),\ldots,(Y_n,Z_n,C_n,\mathcal{T}_n)\big)$ be the random vector that takes values in set $\mathcal{W}$ and whose realization gives rise to the observed data. Given the set of courts $\mathcal{C}_0$ and quarters $\mathcal{T}_0$, consider a map $g:\mathcal{W}\mapsto\mathcal{W}$ such that:

$$\begin{aligned}&g\big((y_1,z_1,c_1,\mathbf{t}_1),(y_2,z_2,c_2,\mathbf{t}_2),\ldots(y_n,z_n,c_n,\mathbf{t}_n)\big)\\=\;&\big((y_1,z_{\gamma(1)},c_1,\mathbf{t}_1),(y_2,z_{\gamma(2)},c_2,\mathbf{t}_2),\ldots(y_n,z_{\gamma(n)},c_n,\mathbf{t}_n)\big),\end{aligned}$$

where

$$\gamma(i)\;=\;\begin{cases}\pi(i) & \text{if } x_i\in\mathcal{C}_0 \text{ and } \mathbf{t}_i\in\mathcal{T}_0\\ i & \text{otherwise}\end{cases}$$

and $\pi$ is a permutation of $\{i: \ x_i \in \mathcal{C}_0 \text{ and } \mathbf{t}_i \in \mathcal{T}_0\}$. Let $\mathcal{G}$ be the set of all such maps.[14]

**Proposition 3.** *Fix a set of courts $\mathcal{C}_0$ and quarters $\mathcal{T}_0$ and consider the hypothesis $H_0$ : $\Pr(S_i = 1 \,|\, C_i \in \mathcal{C}_0, \mathcal{T}_i \in \mathcal{T}_0) = 0$. If (i) Assumption IVe holds, (ii) $\big((Y_i, Z_i, C_i, \mathcal{T}_i)\big)_{i=1}^n$ is independent and identically distributed under $H_0$, and (iii) $\Pr(Z_i = z \,|\, C_i = c, \mathcal{T}_i = \mathbf{t}) > 0$ for all $z \in \mathcal{Z}$ and for every $c \in \mathcal{C}_0$ and $\mathbf{t} \in \mathcal{T}_0$ under $H_0$, then $W$ and $g(W)$ are identically distributed for all $g \in \mathcal{G}$ under $H_0$.*

*Proof.* Fix $w = \big((y_1, z_1, c_1, \mathbf{t}_1), \ldots, (y_n, z_n, c_n, \mathbf{t}_n)\big) \in \mathcal{W}$ and $g \in \mathcal{G}$. Let $\mathcal{I}_w \equiv \{i \in \{1, \ldots, n\} : \ x_i \in \mathcal{C}_0, \mathbf{t}_i \in \mathcal{T}_0\}$ and denote by $\pi$ the permutation of $\mathcal{I}_w$ associated with $g$. Under $H_0$, it follows that

$$
\begin{aligned}
&\Pr(W = w) \\
&= \Pr\big((Y_i, Z_i, C_i, \mathcal{T}_i) = (y_i, z_i, c_i, \mathbf{t}_i) \ \forall i \notin \mathcal{I}_w\big) \\
&\quad \times \prod_{i \in \mathcal{I}_w} \Pr\big((C_i, \mathcal{T}_i) = (c_i, \mathbf{t}_i)\big) \Pr\big(Y_i = y_i \,|\, c_i, \mathbf{t}_i\big) \Pr\big(Z_i = z_i \,|\, c_i, \mathbf{t}_i\big) \\
&= \Pr\big((Y_i, Z_i, C_i, \mathcal{T}_i) = (y_i, z_i, c_i, \mathbf{t}_i) \ \forall i \notin \mathcal{I}_w\big) \\
&\quad \times \prod_{i \in \mathcal{I}_w} \Pr\big((C_i, \mathcal{T}_i) = (c_i, \mathbf{t}_i)\big) \Pr\big(Y_i = y_i \,|\, c_i, \mathbf{t}_i\big) \Pr\big(Z_i = z_{\pi^{-1}(i)} \,|\, c_i, \mathbf{t}_i\big) \\
&= \Pr\big(W = g^{-1}(w)\big) \\
&= \Pr\big(g(W) = w\big),
\end{aligned}
$$

where the first equality uses the fact that $H_0$ and our full support assumption on $Z_i$ imply that $Y_i \perp\!\!\!\perp Z_i \,|\, C_i = x, \mathcal{T}_i = \mathbf{t}$ for all $c \in \mathcal{C}_0$ and $\mathbf{t} \in \mathcal{T}_0$. ■

Proposition 3 states the conditions under which the randomization hypothesis holds, i.e.

[14] I omit the dependence of $\mathcal{G}$ on $(\mathcal{C}_0, \mathcal{T}_0)$ for notational brevity.

| Assumption | **PMFe** | **IVe** | **PMFe & IVe** |
|---|---|---|---|
| Estimate of l.b. on $\Pr(S_i = 1)$ | 2.59 | −0.11 | 2.08 |
| [95% confidence interval] | [2.43, 100] | | [2.05, 100] |
| Expected upper bound of bias | 3.59 | 4.06 | 4.82 |
| Empirical counterpart of l.b. | 6.05 | 3.94 | 6.90 |
| [95% rej. region, $H_0$ : l.b. = 0] | [3.28, 100] | [3.24, 100] | [4.00, 100] |
| # Groups | 12,413 | 19,953 | 18,117 |
| # Case Assignments | 2,241,490 | 2,023,010 | 2,023,010 |
| # Assignments/Group: p50 | 7 | 15 | 6 |
| p75 | 49 | 59 | 42 |
| p90 | 340 | 210 | 218 |
| p95 | 830 | 456 | 502 |
| p99 | 3,140 | 1,605 | 1,875 |

Table 5: Aggregate Lower Bounds on Irregular Assignments

*Notes:* Each column shows the estimation results for the sharp lower bound on the aggregate rate of irregular assignments, $\Pr(S_i = 1)$, across Ecuador's district courts between March 2016 and February 2020. The first column shows the results under assumption PMFe, the second column shows the results under assumption IVe, and the third column shows the results under both assumptions PMFe and IVe. The first row shows the mean-conservative estimates and conservative confidence intervals. The second row shows the estimated upper bound on the expected bias of the empirical counterpart of the population lower bound on $\Pr(S_i = 1)$. The third row shows this empirical lower bound and the rejection region to decide between $H_0 : \Pr(S_i = 1) = 0$ and $H_1 : \Pr(S_i = 1) > 0$ at a type-I error level lower than 5% using the empirical lower bound as the test statistic. The remaining rows show the number of court-time period combinations used in estimation, the number of case assignments in the data, and percentiles of the number of case assignments per court-time period combination, respectively.

the distribution of our data is invariant to transformations in $\mathcal{G}$ under the null hypothesis. To test (10), consider the test statistic $\widehat{\mathrm{LB}}_{\mathrm{IV}}(\mathcal{C}_0, \mathcal{T}_0)$, obtained by replacing the unknown probabilities in $\mathrm{LB}_{\mathrm{IV}}(\mathcal{C}_0, \mathcal{T}_0)$ with their empirical counterparts. Under the randomization hypothesis, the quantiles of the set of test statistic values obtained across possible transformations of the data $g \in \mathcal{G}$ serve as the quantiles of the distribution of the test statistic under the null hypothesis to test (10) at a given size $\alpha$. For a textbook treatment of permutation tests, see section 15.2 in Lehmann and Romano (2005).

# 6 Results

This section describes the estimates on the rates of irregular assignments in Ecuador. Table 5 reports the lower bound on the extent of irregular assignments across district courts, between March 2016 and February 2020. When regulatory case assignmnents are uniformly distributed, I find that at least 2.59% of case assignments are irregular. This estimate is mean-conservative, in the sense that its estimand has a lower expected value than the population lower bound on $\Pr(S_i = 1)$. The empirical counterpart of this lower bound is considerably higher at 6.05%, with a positive expected bias of at most 3.59%. Exclusive reliance on my measure of case paperwork as a one-sided instrument reveals a negative mean-conservative estimate of the lower bound on the overall rate of irregular assignments. This estimate is too conservative, however: a permutation test rejects the null hypothesis of no irregular assignments at type-I error levels of at most 1%. On the other hand, the empirical counterpart of the lower bound on $\Pr(S_i = 1)$ when I impose both assumptions PMFe and IVe equals 6.9% and does exceed the one that I find exclusively on the basis of assumption PMFe. However, it is also associated with a larger bias correction. After this correction, assumptions PMFe and IVe show a smaller lower bound on $\Pr(S_i = 1)$ of 2.08%.

Few courts display considerable amounts of irregular assignments. Figure 3 shows that assumption IVe singles out a criminal court in the commercial hub of Guayaquil with about 15% of irregular assignments, whereas assumption PMFe finds fifteen courts with rates of irregular assignments that exceed 10%. The court that is singled out under assumption IVe had at least 12% of irregular assignments under assumption PMFe.

Few judges are found to be involved in irregular assignments. Figure 4 shows that 6.6% of judges in the data were involved in irregular assignments, with 5 judges receiving over 2 irregular assignments for every 5 assignments received. A ranking of the judges who were found to be involved in irregular assignments by the total amount of irregular assignments they received shows four judges who stand out from the rest. In particular, the judge who I find received the largest amount of irregular assignments, $1,500$ throughout the sample period, faced corruption charges in March of 2020 because of his involvement in the sale of sentence reductions to inmates in exchange for money or alcohol.

# 7 Conclusion

In this paper, I developed a method to evaluate a basic aspect of judicial activity: the assignment of cases to judges. The method yielded measurements on the extent to which actual assignments violate the regulations that govern them. In particular, it provided the most informative bounds on the extent of violations that can be achieved with individual case assignment data and knowledge of the existing regulations. Such data is available to the public in Ecuador, but is routinely collected by judiciaries around the world.

In Ecuador, a weak interpretation of the regulations suggested an instrumental variable that detected irregular assignments in a handful of courts. A stronger interpretation of the regulations implied that 6.6% of judges who worked in district courts between March, 2016 and February, 2020 were involved in such assignments, and that 2.6% of case assignments violated the regulations. In either case, the irregular assignments that I detected are highly localized. These findings suggest that the method is a useful tool to direct regulatory enforcement resources.

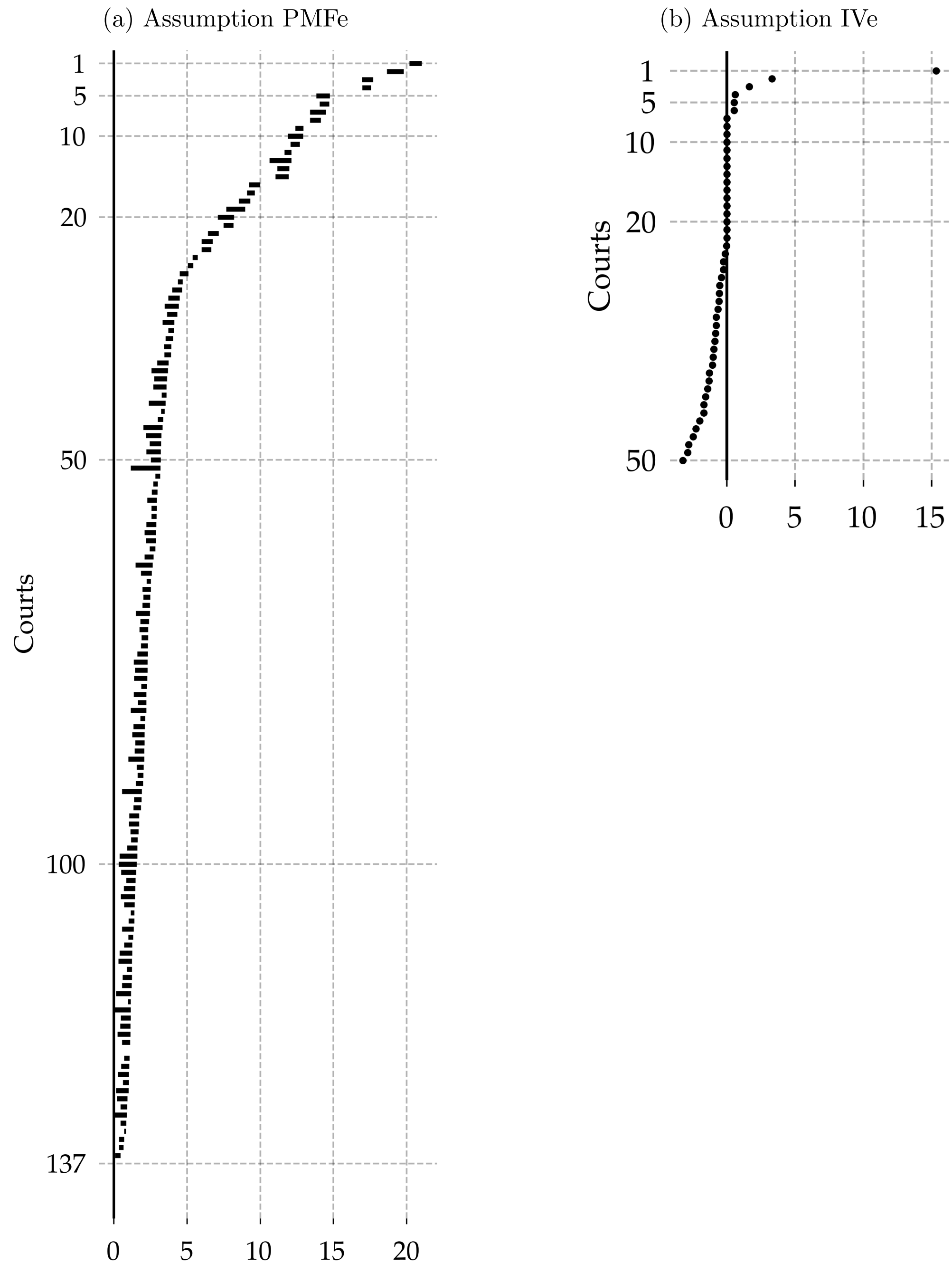



Figure 3: Rates of Irregular Assignments by Court under Assumptions PMFe and IVe

*Notes:* Panel (a) shows the point estimate and 99% confidence interval for the lower bound on the rate of irregular assignments under Assumption PMFe for all courts where the confidence interval exceeds zero. The result for each court is depicted as a line whose left and right endpoints equal the lower end of the confidence interval and the point estimate, respectively. Panel (b) shows the point estimate for the lower bound on the rate of irregular assignments for all courts that I find to be involved in irregular assignments under Assumption IVe using a permutation test at type-I error level of 1%. The result for each court is depicted as a dot.

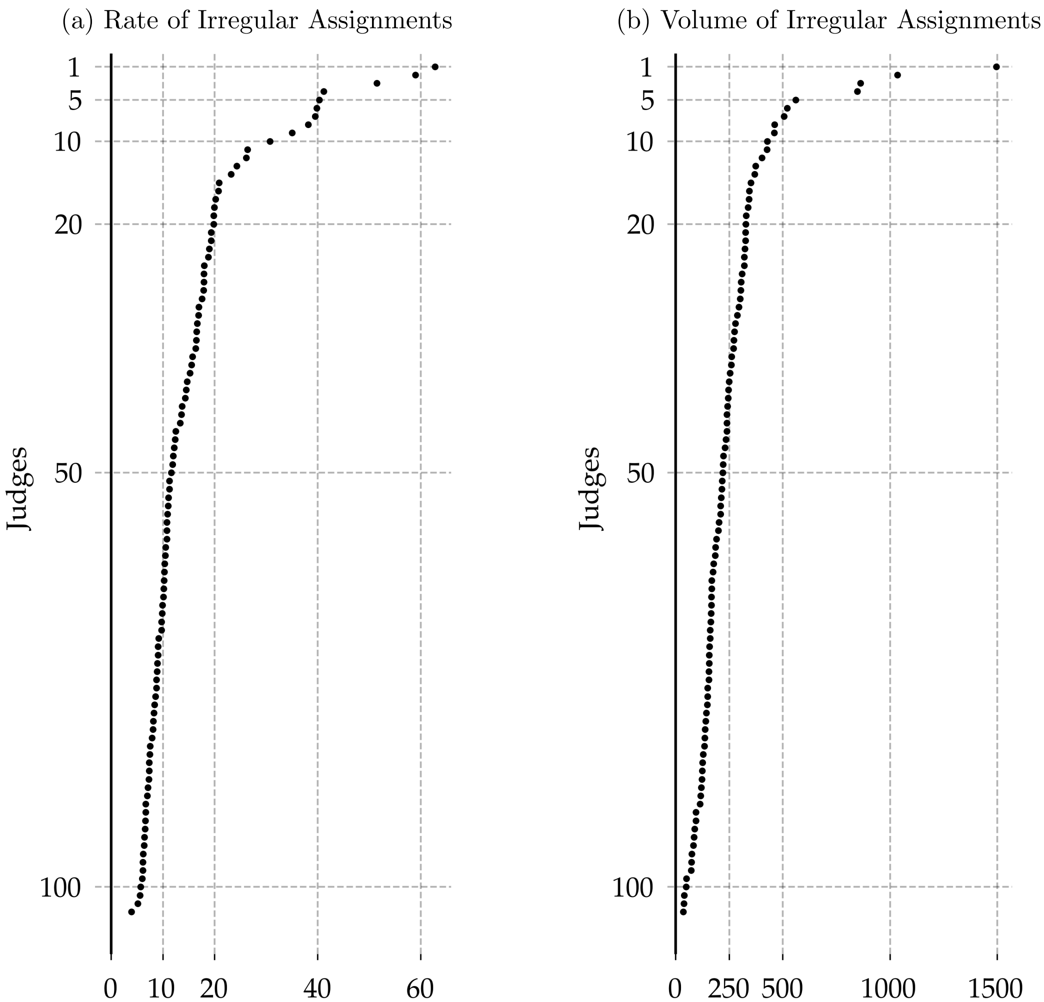



Figure 4: Rates of Irregular Assignments by Judge

*Notes:* Panel (a) shows the point estimate for the lower bound on the rate of irregular assignments for all judges that I find to be involved in irregular assignments at a family-wise error rate of at most 5%, using the stepdown resampling method to construct $p$-values adjusted for multiple hypothesis testing of Romano and Wolf (2016). Panel (b) shows the volume of irregular assignments received by each of the judges that appear in Panel (a), but sorted by volume.

# References


Balke, Alexander and Judea Pearl (1997). “Bounds on Treatment Effects from Studies with Imperfect Compliance”. In: *Journal of the American Statistical Association* 92.439, pp. 1171–1176. DOI: 10.1080/01621459.1997.10474074. eprint: https://doi.org/10.1080/01621459.1997.10474074. URL: https://doi.org/10.1080/01621459.1997.10474074.

Bollinger, Christopher R (1996). “Bounding mean regressions when a binary regressor is mismeasured”. In: *Journal of Econometrics* 73.2, pp. 387–399. ISSN: 0304-4076. DOI: https://doi.org/10.1016/S0304-4076(95)01730-5. URL: https://www.sciencedirect.com/science/article/pii/S0304407695017305.

Clarke, Damian, Joseph P. Romano, and Michael Wolf (2020). “The Romano–Wolf multiple-hypothesis correction in Stata”. In: *The Stata Journal* 20.4, pp. 812–843. DOI: 10.1177/1536867X20976314. eprint: https://doi.org/10.1177/1536867X20976314. URL: https://doi.org/10.1177/1536867X20976314.

Código Orgánico General de Procesos (May 2015). *Registro Oficial, Suplemento 506*. Asamblea Nacional del Ecuador. Última modificación: 2016-12-09.

Código Orgánico Integral Penal (Feb. 2014). *Registro Oficial, Suplemento 180*. Asamblea Nacional del Ecuador. Última modificación: 2018-02-05.

Compiani, Giovanni and Yuichi Kitamura (Oct. 2016). “Using Mixtures in Econometric Models: A Brief Review and Some New Results”. In: *The Econometrics Journal* 19.3, pp. C95–C127. ISSN: 1368-4221. DOI: 10.1111/ectj.12068. eprint: https://academic.oup.com/ectj/article-pdf/19/3/C95/27684243/ectj0c95.pdf. URL: https://doi.org/10.1111/ectj.12068.

Daljord, Oystein, Guillaume Pouliot, Junji Xiao, and Mandy Hu (May 2021). *The Black Market for Beijing License Plates*. Papers 2105.00517. arXiv.org. URL: https://ideas.repec.org/p/arx/papers/2105.00517.html.

Demuynck, Thomas (2015). “Bounding average treatment effects: A linear programming approach”. In: *Economics Letters* 137, pp. 75–77. ISSN: 0165-1765. DOI: https://doi.org/10.1016/j.econlet.2015.09.042. URL: https://www.sciencedirect.com/science/article/pii/S0165176515004012.

DiTraglia, Francis J. and Camilo García-Jimeno (2019). “Identifying the effect of a mis-classified, binary, endogenous regressor”. In: *Journal of Econometrics* 209.2, pp. 376–390. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2019.01.007. URL: https://www.sciencedirect.com/science/article/pii/S0304407619300181.

Dow, Sterling (1937). “Prytaneis: a study of the inscriptions honoring the Athenian councillors”. In: *Hesperia Supplements* 1, pp. 1–258. URL: https://www.jstor.org/stable/1370763.

Egan, Mark, Gregor Matvos, and Amit Seru (May 2021). *Arbitration with Uninformed Consumers*. Tech. rep. 19-046. DOI: http://dx.doi.org/10.2139/ssrn.3260442. URL: https://ssrn.com/abstract=3260442.

Fisman, Raymond and Shang-Jin Wei (2004). “Tax Rates and Tax Evasion: Evidence from “Missing Imports” in China”. In: *Journal of Political Economy* 112.2, pp. 471–496. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/10.1086/381476.

Galichon, Alfred (2016). *Optimal Transport Methods in Economics*. Princeton University Press. URL: http://www.jstor.org/stable/j.ctt1q1xs9h.

Hall, Peter and Xiao-Hua Zhou (2003). "Nonparametric estimation of component distributions in a multivariate mixture". In: *The Annals of Statistics* 31.1, pp. 201–224. DOI: 10.1214/aos/1046294462. URL: https://doi.org/10.1214/aos/1046294462.

Henry, Marc, Yuichi Kitamura, and Bernard Salanié (2014). "Partial identification of finite mixtures in econometric models". In: *Quantitative Economics* 5.1, pp. 123–144. DOI: https://doi.org/10.3982/QE170. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/QE170. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/QE170.

Horowitz, Joel L. and Charles F. Manski (1995). "Identification and Robustness with Contaminated and Corrupted Data". In: *Econometrica* 63.2, pp. 281–302. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2951627.

Hu, Yingyao (2008). "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution". In: *Journal of Econometrics* 144.1, pp. 27–61. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2007.12.001. URL: https://www.sciencedirect.com/science/article/pii/S0304407607002436.

Imbens, Guido W. and Joshua D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects". In: *Econometrica* 62.2, pp. 467–475. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2951620.

Kitamura, Yuichi and Louise Laage (2018). *Nonparametric Analysis of Finite Mixtures*. arXiv: 1811.02727 [econ.EM].

Lafférs, Lukáš (2013). "Essays in Partial Identification". In: *Ph.D Thesis, Norwegian School of Economics*. URL: https://static1.squarespace.com/static/52e69d46e4b05a145935f24d/t/5334b151e4b0a31591863425/1395962193849/Dissertation+Laffers.pdf.

— (2019a). "Bounding average treatment effects using linear programming". In: *Empirical Economics* 57, pp. 727–767. ISSN: 1435-8921. DOI: 10.1007/s00181-018-1474-z. URL: https://doi.org/10.1007/s00181-018-1474-z.

— (Feb. 2019b). "Identification in Models with Discrete Variables". In: *Computational Economics* 53.2, pp. 657–696. ISSN: 1572-9974. DOI: 10.1007/s10614-017-9758-5. URL: https://doi.org/10.1007/s10614-017-9758-5.

Lehmann, Erich L. and Joseph P. Romano (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, New York, NY. DOI: https://doi.org/10.1007/0-387-27605-X.

Levin, David A and Yuval Peres (2017). *Markov Chains and Mixing Times: Second Edition*. Vol. 107. American Mathematical Society Non-Series Monographs. DOI: https://doi.org/10.1090/mbk/107.

Machasilla, Érika, Verónica Mejía, and Douglas Torres Feraud (June 2020). *Estadísticas de Producción Judicial, Metodología*. Consejo de la Judicatura. Última modificación: 2015-05-22.

Mahajan, Aprajit (2006). "Identification and Estimation of Regression Models with Misclassification". In: *Econometrica* 74.3, pp. 631–665. DOI: https://doi.org/10.1111/j.1468-0262.2006.00677.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-

0262.2006.00677.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0262.2006.00677.x.

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky (2018). "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters". In: *Econometrica* 86.5, pp. 1589–1619. DOI: https://doi.org/10.3982/ECTA15463. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA15463. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA15463.

Molinari, Francesca (2008). "Partial identification of probability distributions with misclassified data". In: *Journal of Econometrics* 144.1, pp. 81–117. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2007.12.003. URL: https://www.sciencedirect.com/science/article/pii/S0304407607002564.

— (2010). "Missing Treatments". In: *Journal of Business & Economic Statistics* 28.1, pp. 82–95. ISSN: 07350015. URL: http://www.jstor.org/stable/27799116.

Olken, Benjamin A. (2006). "Corruption and the costs of redistribution: Micro evidence from Indonesia". In: *Journal of Public Economics* 90.4, pp. 853–870. ISSN: 0047-2727. DOI: https://doi.org/10.1016/j.jpubeco.2005.05.004. URL: https://www.sciencedirect.com/science/article/pii/S0047272705000678.

Reglamento de Sorteo de Juicios (Sept. 2004). *Registro Oficial 421*. Consejo de la Judicatura. Última modificación: 2013-08-23.

Reinikka, Ritva and Jakob Svensson (2004). "Local Capture: Evidence from a Central Government Transfer Program in Uganda". In: *The Quarterly Journal of Economics* 119.2, pp. 679–705. ISSN: 00335533, 15314650. URL: http://www.jstor.org/stable/25098697.

Romano, Joseph P. and Michael Wolf (2016). "Efficient computation of adjusted p-values for resampling-based stepdown multiple testing". In: *Statistics & Probability Letters* 113, pp. 38–40. ISSN: 0167-7152. DOI: https://doi.org/10.1016/j.spl.2016.02.012. URL: https://www.sciencedirect.com/science/article/pii/S0167715216000389.

Tebaldi, Pietro, Alexander Torgovitsky, and Hanbin Yang (May 2019). *Nonparametric Estimates of Demand in the California Health Insurance Exchange*. Working Paper 25827. National Bureau of Economic Research. DOI: 10.3386/w25827. URL: http://www.nber.org/papers/w25827.

Torgovitsky, Alexander (2019). "Nonparametric Inference on State Dependence in Unemployment". In: *Econometrica* 87.5, pp. 1475–1505. DOI: https://doi.org/10.3982/ECTA14138. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA14138. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA14138.

Transparency International (2007). *Global Corruption Report 2007*. Cambridge University Press. ISBN: 978-0-521-70070-2. URL: www.cambridge.org/9780521700702.

# A Measurement of the Set of Competent Judges

Assumption PMFe raises a challenge: which judges are competent in a given court and field of law $x$, and point in time $t$? I consider a judge to be competent in $(x, t)$ whenever she is active and has a low relative workload.

Competent judges must be in an active spell. Judge $y$ is in an active spell in court and field of law $x$ at time $t$ if $t$ falls within a window of less than $\alpha$ days between the time when she was last assigned a case in $x$, and the time when she will be next assigned a case in $x$. Thus, lower values of $\alpha$ require judges to receive cases at a higher frequency to be considered active. Any choice of $\alpha$ entails two errors. A judge could be considered active at $(x, t)$ when she actually was not; and a judge could be considered inactive at $(x, t)$ when she actually was active. High values of $\alpha$ produce the former errors, whereas low values of $\alpha$ produce the latter errors.

Competent judges must have a case workload that is less than $\beta$ times the workload of their peer with the lowest case workload at $(x, t)$. I include this criterion because it is listed as such by the President of the Judicial Council (see Table 3).

I select parameters $\alpha$ and $\beta$ so as to obtain conservative lower bounds on the overall rate of irregular assignments, $\Pr(S_i = 1)$. Figure A.1 shows that I achieve a conservative lower bound on this parameter when I ignore judges' case workloads ($\beta = \infty$), and when judges must receive cases at a frequency of at least $\alpha = 30$ days to be considered active between assignments.

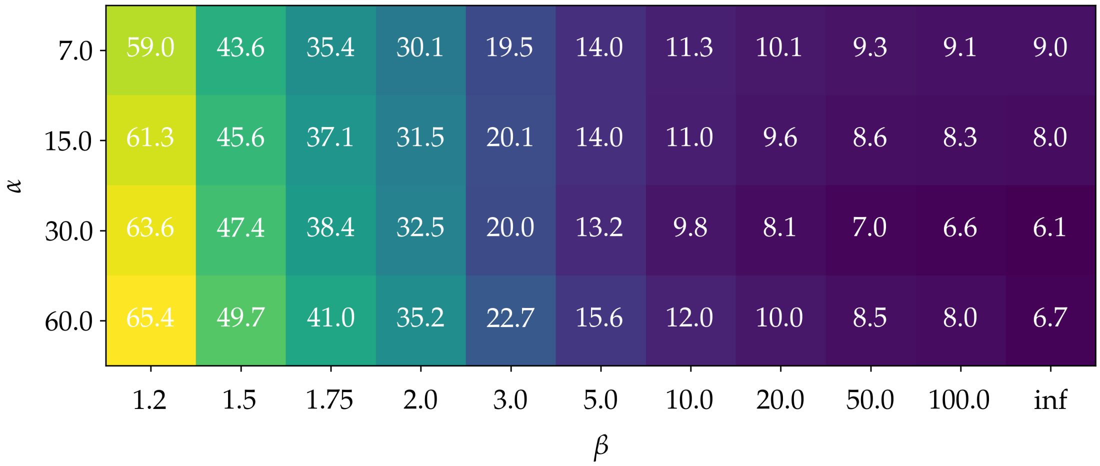



Figure A.1: Lower Bounds on Overall Irregular Assignments across Tuning Parameters

Each cell shows the empirical counterpart of the population lower bound on the overall rate of irregular assignments, $\Pr(S_i = 1)$, under Assumption PMFe for a given value of $\alpha$ and $\beta$. $\alpha$ is measured in days and equals the minimum frequency of cases assignments made to a judge for her to be in an active spell. $\beta$ is the maximum case workload of a judge (relative to the peer with the lowest workload) for her to be available for case assignments. When $\beta = \infty$, judges' workloads do not determine if they are available for case assignments or not.

# B Missing Data

In this section, I consider interpretations of the findings in Table 2 that are conscious of missing judicial case assignment information.

In section 3 of the main text, I assume that the unconditional distribution case assignment information, $(Y_i, Z_i, X_i)$, is known or estimable. In the application, however, the data includes the assignment information of publicly-available judicial cases that I retrieved from a website of the Ecuadorian government. Three kinds of judicial cases are therefore missing in my data: non-confidential cases that are unavailable in the government's website, non-confidential cases that are available in the government's website, but were not retrieved by my data collection exercise, and confidential cases.

Let $M_i$ indicate if case $i$ belongs to one of these three mutually-exclusive categories. Given the data-collection procedure, only the distribution of case assignment information among non-missing judicial cases, $(Y_i, Z_i, X_i) \,|\, M_i = 0$, can be known or estimable. The identification results listed in Table 2 are easily adapted to this setting however, provided the following assumptions are made.

**Assumption PMFm.** *$M_i$ is statistically independent of $Y_i(0)$, conditional on $X_i$.*

**Assumption IVm.** *$Z_i$ is statistically independent of $Y_i(0)$, conditional on $X_i$ and $M_i = 0$.*

Assumptions PMFx and PMFm amount to taking the distribution of regulatory assignments, $Y_i(0)$, to be known (conditional on $X_i$). Assumption PMFm requires that $M_i$ satisfy the one-sided instrument exclusion restriction. In support of this assumption, I note that Ecuadorian regulations do not specify distinct case assignment procedures for confidential cases. Under Assumptions PMFx and PMFm, the identification results in Table 2 hold, conditional on $M_i = 0$: for all $x \in \mathcal{X}$ and $y^* \in \mathcal{Y}$,

$$\begin{aligned} &\Pr(S_i = 1 \,|\, X_i = x, M_i = 0) \\ &\qquad \in \left[ \frac{1}{2} \sum_y \Big| \Pr(Y_i = y \,|\, x, M_i = 0) \; - \; \Pr(Y_i(0) = y \,|\, x) \Big|, \quad 1 \right] \end{aligned} \tag{A.1}$$

and

$$\begin{aligned} &\Pr(S_i = 1 \,|\, Y_i = y^*, X_i = x, M_i = 0) \\ &\qquad \in \left[ \max\left\{ 0, \; \frac{\Pr(Y_i = y^* \,|\, x, M_i = 0) - \Pr(Y_i(0) = y^* \,|\, x)}{\Pr(Y_i = y^* \,|\, x, M_i = 0)} \right\}, \quad 1 \right]. \end{aligned} \tag{A.2}$$

On the other hand, Assumption IVm posits that $Z_i$ is a one-sided instrument, conditional on the judicial case not being missing. Under Assumption IVm, the identification results under one-sided instruments in Table 2 hold conditional on $M_i = 0$: for every $x \in \mathcal{X}$,

$$\begin{aligned} &\Pr(S_i = 1 \,|\, X_i = x, M_i = 0) &\text{(A.3)} \\ &\in \left[ \min_{\phi \in \Phi} \sum_{z,y} \frac{1}{2} \Pr(Z_i = z \,|\, x, M_i = 0) \Big| \Pr(Y_i = y \,|\, x, z, M_i = 0) \;-\; \phi(y \,|\, x) \Big|, \quad 1 \right]. \end{aligned}$$

I note that the bounds presented in (A.1), (A.2) and (A.3) are valid for the rates of irregular assignments among all non-confidential cases, not just among non-missing cases, under two additional assumptions: that my data-collection exercise fails to retrieve the information of available judicial cases at random, and that the assignment of a non-confidential judicial case that is not available in the government's website is more likely to be irregular than that the assignment of a non-confidential case that is available in the government's website.

# C Proofs

Lemmas 1 – 5 construct the sharp bounds presented in Propositions 1 and 2 on the basis of the theory of identification from Section 3. Notice that Propositions 1 and 2 do not involve case covariates, $X_i$, for the sake of illustration. This means that the support of $X_i$, $\mathcal{X}$, is implicitly assumed to be a singleton.

*Proof of Proposition 1.* By Lemma 1, the sharp upper bound for $\Pr(S_i = 1)$ and the sharp upper bounds for $\Pr(S_i = 1 \,|\, Y_i = y^*)$ for every $y^* \in \{1, \ldots, n_Y\}$ all equal one. By Lemma 5, the sharp lower bound for $\Pr(S_i = 1)$ is:

$$\sum_{y \in \mathcal{Y}} \frac{1}{2} \Big| \Pr(Y_i = y) \;-\; \Pr(Y_i(0) = y) \Big|.$$

whereas, for any $y^* \in \{1, \ldots, n_Y\}$, the sharp lower bound for $\Pr(S_i = 1 \,|\, Y_i = y^*)$ is:

$$\max\left\{0, \quad \frac{\Pr(Y_i = y^*) \; - \; \Pr(Y_i(0) = y^*)}{\Pr(Y_i = y^*)}\right\}.$$

■

*Proof of Proposition 2.* By Lemma 1, the sharp upper bound for $\Pr(S_i = 1)$ and the sharp upper bounds for $\Pr(S_i = 1 \,|\, Y_i = y^*)$ for every $y^* \in \{1, \ldots, n_Y\}$ all equal one. By Lemma 3, the sharp lower bound for $\Pr(S_i = 1)$ is:

$$\sum_{y \in \mathcal{Y}, z \in \mathcal{Z}} \frac{p_{\min}}{2} \Big| \Pr(Y_i = y \,|\, Z_i = 0) \; - \; \Pr(Y_i = y \,|\, Z_i = 1) \Big|$$

whereas, for any $y^* \in \{1, \ldots, n_Y\}$, the sharp lower bound for $\Pr(S_i = 1 \,|\, Y_i = y^*)$ equals zero. ■

**Lemma 1.** Fix $x^* \in \mathcal{X}$ and $y^* \in \{1, \dots, n_Y\}$. Consider the parameter associated with $\Pr(S_i = 1 \mid X_i = x^*)$, $\theta(f; c)$, where $c(y_0, y_1, s, z, x) \equiv 1\{x = x^*\} \cdot 1\{s = 1\}$, and the parameter associated with $\Pr(S_i = 1 \mid Y_i = y^*, X_i = x^*)$, $\theta(f; c_y)$, where $c_y(y_0, y_1, s, z, x) \equiv \frac{1\{x=x^*, y_1=y^*\}}{\Pr(Y_i=y^* \mid X_i=x^*)} \cdot 1\{s = 1\}$. Then,

$$\overline{\theta}_{\text{PMF}}(c) = \overline{\theta}_{\text{PMF}}(c_y) = \overline{\theta}_{\text{IV}}(c_y) = \overline{\theta}_{\text{IV}}(c) = 1.$$

*Proof.* Let Assumption PMFx hold and define the data-generating process $f_{\text{PMF}}$ as:

$$f_{\text{PMF}}(y_0, y_1, s, z \mid x) = \begin{cases} 0 & \text{if } s = 0 \\ \Pr(Y_i(0) = y_0 \mid X_i = x) \Pr(Y_i = y_1, Z_i = z \mid X_i = x) & \text{if } s = 1. \end{cases}$$

$f_{\text{PMF}}$ is well-defined, since it is weakly positive and adds up to one for each $x \in \mathcal{X}$. Now, $f_{\text{PMF}} \in \mathcal{F}^*_{\text{PMF}}$, since it satisfies restrictions ($\text{R}_{\text{OE}}$) and ($\text{R}_{\text{PMF}}$). Moreover,

$$\begin{aligned}
\theta\left(f_{\text{PMF}};\ c\right) &= \sum_{y_0, y_1, s, z, x} c(y_0, y_1, s, z, x) f_{\text{PMF}}(y_0, y_1, s, z \mid x) \\
&= \sum_{y_0, y_1, z} f_{\text{PMF}}(y_0, y_1, 1, z \mid x^*) \\
&= \sum_{y_0, y_1, z} \Pr(Y_i(0) = y_0 \mid X_i = x^*) \Pr(Y_i = y_1, Z_i = z \mid X_i = x^*) \\
&= \sum_{y_1, z} \Pr(Y_i = y_1, Z_i = z \mid X_i = x^*) \\
&= 1.
\end{aligned}$$

and

$$
\begin{aligned}
\theta\left(f_{\mathrm{PMF}};\ c_y\right) &= \sum_{y_0,y_1,s,z,x} c_y(y_0,y_1,s,z,x) f_{\mathrm{PMF}}(y_0,y_1,s,z\,|\,x) \\
&= \sum_{y_0,z} \frac{f_{\mathrm{PMF}}(y_0,y^*,1,z\,|\,x^*)}{\Pr(Y_i=y^*\,|\,X_i=x^*)} \\
&= \sum_{y_0,z} \Pr(Y_i(0)=y_0\,|\,X_i=x^*) \frac{\Pr(Y_i=y^*,Z_i=z\,|\,X_i=x^*)}{\Pr(Y_i=y^*\,|\,X_i=x^*)} \\
&= 1.
\end{aligned}
$$

Since $\theta(f;c) \leq 1$ and $\theta(f;c_y) \leq 1$ for all $f \in \mathcal{F}$, it follows that $\overline{\theta}_{\mathrm{PMF}}(c) = \overline{\theta}_{\mathrm{PMF}}(c_y) = 1$.

Now let Assumption IVx hold and, for a given probability mass function $\phi$ defined over $\{1,\ldots,n_Y\}$ for each $x \in \mathcal{X}$, define the data-generating process $f_{\mathrm{IV}}$ as:

$$
f_{\mathrm{IV}}(y_0,y_1,s,z\,|\,x) = \begin{cases} 0 & \text{if } s=0 \\ \phi(y_0\,|\,x)\Pr(Y_i=y_1,Z_i=z\,|\,X_i=x) & \text{if } s=1. \end{cases}
$$

Since $f_{\mathrm{IV}}$ is weakly positive, adds up to one for every $x \in \mathcal{X}$ and satisfies restrictions ($\mathrm{R_{OE}}$) and ($\mathrm{R_{IV}}$), $f_{\mathrm{IV}} \in \mathcal{F}^*_{\mathrm{IV}}$. Moreover, $\theta\left(f_{\mathrm{IV}};c\right) = \theta\left(f_{\mathrm{IV}};c_y\right) = 1$, so that $\overline{\theta}_{\mathrm{IV}}(c) = \overline{\theta}_{\mathrm{IV}}(c_y) = 1$. ∎

**Lemma 2.** Define $\mathcal{Y} \equiv \{1,\ldots,n_Y\}$ and consider the linear parameter $\theta(f;c^S)$ associated with scalar coefficients $c^S$, such that $c^S(y_0,y_1,s,z,x) \equiv \omega(y_1,x)1\{s=1\}$ and $\omega : \mathcal{Y}\times\mathcal{X} \mapsto \mathbb{R}$. Let $\Gamma$ be the set of probability mass functions defined over $\mathcal{Y}\times\mathcal{Y}$ and $\Phi$ be the set of probability mass functions defined over $\mathcal{Y}$ for each $x \in \mathcal{X}$. Then

$$\begin{aligned}
\underline{\theta}_{\text{IV}}(c^S) &\equiv \min_{f \in \mathcal{F}^*_{\text{IV}}} \quad \theta(f;\, c^S) \\
&= \min_{\phi \in \Phi} \sum_{z \in \mathcal{Z}, x \in \mathcal{X}} \Pr(Z_i = z \,|\, X_i = x)\ \lambda(\phi, z, x)
\end{aligned}$$

where

$$\begin{aligned}
\lambda(\phi, z, x) \quad &= \quad \min_{\gamma \in \Gamma} \sum_{y_0, y \in \mathcal{Y}} \omega(y, x)\ 1\{y_0 \neq y\}\ \gamma(y_0, y) \qquad \text{s.t.} \\
&\qquad (i) \quad \sum_{y_0 \in \mathcal{Y}} \gamma(y_0, y) \ = \ \Pr(Y_i = y \,|\, Z_i = z, X_i = x) \qquad \forall\, y \\
&\qquad (ii) \quad \sum_{y \in \mathcal{Y}} \gamma(y_0, y) \ = \ \phi(y_0 \,|\, x) \qquad \forall\, y_0
\end{aligned}$$

*Proof.* The proof has two parts. In Part I, I show that $\underline{\theta}_{\text{IV}}(c^S)$ equals the solution to a convenient linear program:

$$\begin{aligned}
\underline{\theta}_{\text{IV}}(c^S) = \quad & \min_{\xi \in \Xi} \sum_{y_0, y \in \mathcal{Y}, z \in \mathcal{Z}, x \in \mathcal{X}} \omega(y, x)\ 1\{y_0 \neq y\} \xi(y_0, y, z \,|\, x) \qquad \text{subject to:} \\
(i) \quad & \sum_{y_0 \in \mathcal{Y}} \xi(y_0, y, z \,|\, x) \ = \ \Pr(Y_i = y, Z_i = z \,|\, X_i = x) \qquad \forall\, y, z, x \\
(ii) \quad & \sum_{y \in \mathcal{Y}} \xi(y_0, y, z \,|\, x) \ = \ \Pr(Z_i = z \,|\, X_i = x) \sum_{y \in \mathcal{Y}, z' \in \mathcal{Z}} \xi(y_0, y, z' \,|\, x) \qquad \forall\, y_0, z, x,
\end{aligned}$$

where $\Xi$ is the set of probability mass functions defined over $\mathcal{Y} \times \mathcal{Y} \times \mathcal{Z}$ for each $x \in \mathcal{X}$. In Part II, I reformulate this linear program.

**Part I**

For a given $\xi \in \Xi$, let $\mathcal{F}^*_\xi$ denote the subset of $\mathcal{F}^*_{\text{IV}}$ that is consistent with $\xi$:

$$\mathcal{F}^*_\xi \equiv \Bigg\{ f \in \mathcal{F}^*_{\text{IV}} \; : \sum_{y_1 \in \mathcal{Y}, s \in \{0,1\}} 1\{sy_1 + (1-s)y_0 = y\} f(y_0, y_1, s, z \,|\, x) \; = \; \xi(y_0, y, z \,|\, x) \quad \forall y_0, y, z, x \Bigg\}.$$

Clearly, $\mathcal{F}^*_{\text{IV}} \; = \; \big\{ f \in \mathcal{F}^*_\xi \; : \; \xi \in \Xi \big\} \; = \; \big\{ f \in \mathcal{F}^*_\xi \; : \; \xi \in \Xi \text{ and } \mathcal{F}^*_\xi \neq \varnothing \big\}$, so that

$$\underline{\theta}_{\text{IV}}(c^S) \quad = \quad \min_{\xi \in \Xi} \left( \min_{f \in \mathcal{F}^*_\xi} \theta(f; c^S) \right) \quad \text{subject to} \quad \mathcal{F}^*_\xi \neq \varnothing.$$

I will now show that:

$$\mathcal{F}^*_\xi \neq \varnothing \quad \Longleftrightarrow \quad \begin{cases} \sum_{y_0 \in \mathcal{Y}} \xi(y_0, y, z|x) \; = \; \Pr(Y_i = y, Z_i = z \,|\, X_i = x) \quad \forall y \in \mathcal{Y}, z \in \mathcal{Z}, x \in \mathcal{X} \\ \sum_{y} \xi(y_0, y, z|x) \; = \; \Pr(Z_i = z \,|\, X_i = x) \sum_{y, z'} \xi(y_0, y, z'|x) \quad \forall y_0 \in \mathcal{Y}, z, x. \end{cases} \tag{A.4}$$

To prove sufficiency, suppose that $\mathcal{F}^*_{\hat{\xi}}$ is non-empty for some $\hat{\xi} \in \Xi$, and let $\hat{f} \in \mathcal{F}^*_{\hat{\xi}}$. It follows that, for all $y \in \mathcal{Y}, z \in \mathcal{Z}$,

$$\begin{aligned} & \sum_{y_0 \in \mathcal{Y}} \hat{\xi}(y_0, y, z|x) \\ = & \sum_{y_0, y_1 \in \mathcal{Y}, s \in \{0,1\}} 1\{sy_1 + (1-s)y_0 = y\} \hat{f}(y_0, y_1, s, z \,|\, x) \\ = & \Pr(Y_i = y, Z_i = z \,|\, X_i = x), \end{aligned} \tag{A.5}$$

where the first equality holds by the definition of $\mathcal{F}^*_\xi$ and the fact that $\hat{f} \in \mathcal{F}^*_\xi$, and the second equality holds because $\hat{f} \in \mathcal{F}^*_{\text{IV}}$, so that $\hat{f}$ satisfies restriction ($\text{R}_{\text{OE}}$). On the other hand, for every $y_0 \in \mathcal{Y}$, $z \in \mathcal{Z}$,

$$\begin{aligned}
\sum_{y\in\mathcal{Y}} \hat{\xi}(y_0, y, z|x) &= \sum_{y_1\in\mathcal{Y}, s\in\{0,1\}} \sum_{y\in\mathcal{Y}} 1\{sy_1 + (1-s)y_0 = y\}\hat{f}(y_0, y_1, s, z \,|\, x) \\
&= \sum_{y_1\in\mathcal{Y}, s\in\{0,1\}} \hat{f}(y_0, y_1, s, z \,|\, x) \\
&= \Pr(Z_i = z \,|\, X_i = x) \sum_{y_1\in\mathcal{Y}, s\in\{0,1\}, z'\in\mathcal{Z}} \hat{f}(y_0, y_1, s, z' \,|\, x) \\
&= \Pr(Z_i = z \,|\, X_i = x) \sum_{y\in\mathcal{Y}, z'\in\mathcal{Z}} \hat{\xi}(y_0, y, z'|x), \qquad \text{(A.6)}
\end{aligned}$$

where I used the definition of $\mathcal{F}^*_{\hat{\xi}}$ and the fact that $\hat{f} \in \mathcal{F}^*_{\xi}$ in the first and last equalities, and the third equality holds because $\hat{f} \in \mathcal{F}^*_{\text{IV}}$, so that $\hat{f}$ satisfies restriction ($\text{R}_{\text{IV}}$).

To prove necessity, suppose that a given $\xi \in \Xi$ satisfies the restrictions in the right-hand-side of (A.16) and define $f_\xi$ as:

$$f_\xi(y_0, y_1, s, z \,|\, x) = \begin{cases} 1\{y_0 \neq y_1\}\xi(y_0, y_1, z \,|\, x) & \text{if} \quad s = 1, \\ 1\{y_0 = y_1\}\xi(y_0, y_1, z \,|\, x) & \text{if} \quad s = 0. \end{cases} \qquad \text{(A.7)}$$

I will show that $f_\xi \in \mathcal{F}^*_\xi$. Notice first that $f_\xi \in \mathcal{F}$, since $\xi \in \Xi$. In addition, $f_\xi$ satisfies ($\text{R}_{\text{OE}}$): for every $y \in \mathcal{Y}$, $z \in \mathcal{Z}$ and $x \in \mathcal{X}$,

$$\begin{aligned}
&\sum_{y_0,y_1,s} 1\big\{sy_1 + (1-s)y_0 = y\big\} \; f_\xi(y_0, y_1, s, z \,|\, x) \\
&= \sum_{y_0} \sum_{y_1} \Big(1\big\{y_1 = y\big\} \; 1\{y_0 \neq y_1\} \; \xi\big(y_0, y_1, z|x\big) + 1\big\{y_0 = y\big\} \; 1\{y_0 = y_1\} \; \xi\big(y_0, y_1, z|x\big)\Big) \\
&= \sum_{y_0} \xi(y_0, y, z|x) \\
&= \Pr(Y_i = y, \; Z_i = z \,|\, X_i = x),
\end{aligned}$$

where the last equality is shown in (A.5). Finally, $f_\xi$ satisfies ($\mathrm{R_{IV}}$): for all $y_0 \in \mathcal{Y}$ and $z \in \mathcal{Z}$ and $x \in \mathcal{X}$,

$$
\begin{aligned}
\sum_{y\in\mathcal{Y},s\in\{0,1\}} f_\xi(y_0,y,s,z\,|\,x) &= \sum_{y_1\in\mathcal{Y},s\in\{0,1\}}\sum_{y\in\mathcal{Y}} 1\{sy_1+(1-s)y_0=y\}f_\xi(y_0,y_1,s,z\,|\,x)\\
&= \sum_{y\in\mathcal{Y}} \xi(y_0,y,z\,|\,x)\\
&= \Pr(Z_i=z\,|\,X_i=x)\sum_{y\in\mathcal{Y},z'\in\mathcal{Z}} \xi(y_0,y,z'\,|\,x)\\
&= \Pr(Z_i=z\,|\,X_i=x)\sum_{y_1\in\mathcal{Y},s\in\{0,1\},z'\in\mathcal{Z}} f_\xi(y_0,y_1,s,z'\,|\,x),
\end{aligned}
$$

where the second and last equalities follow from the definition of $\mathcal{F}^*_\xi$, and the third equality is obtained from the steps shown in (A.6).

I now determine $\min_{f\in\mathcal{F}^*_\xi}\theta(f;c^S)$ for all non-empty $\mathcal{F}^*_\xi$. Consider a given $\xi$ such that $\mathcal{F}^*_\xi \neq \varnothing$ and define $f_\xi$ as in (A.7). I will show that $\theta(f_\xi;c^S) = \min_{f\in\mathcal{F}^*_\xi}\theta(f;c^S)$. The previous discussion showed that $f_\xi \in \mathcal{F}^*_\xi$. For any $f \in \mathcal{F}^*_\xi$, it follows that:

$$
\begin{aligned}
\theta(f_\xi; c^S) &= \sum_{y_0,y,s,z,x} \omega(y,x)\ 1\{s=1\} f_\xi(y_0,y,s,z\,|\,x) \\
&= \sum_{y_0,y,z,x} \omega(y,x)\ f_\xi(y_0,y,1,z\,|\,x) \\
&= \sum_{y_0,y,z,x} \omega(y,x)\ 1\{y_0 \neq y\} \xi(y_0,y,z\,|\,x) \\
&= \sum_{y_0,y,z,x} \omega(y,x)\ 1\{y_0 \neq y\} \sum_{y_1,s} 1\{s y_1 + (1-s) y_0 = y\} f(y_0,y_1,s,z\,|\,x) \\
&= \sum_{y_0,y,z,x} \omega(y,x)\ 1\{y_0 \neq y\} \sum_{y_1} 1\{y_1 = y\} f(y_0,y_1,1,z\,|\,x) \\
&= \sum_{y_0,y,z,x} \omega(y,x)\ 1\{y_0 \neq y\} f(y_0,y,1,z\,|\,x) \\
&\leq \sum_{y_0,y,z,x} \omega(y,x)\ f(y_0,y,1,z\,|\,x) \\
&= \sum_{y_0,y,s,z,x} \omega(y,x)\ 1\{s=1\} f(y_0,y,1,z\,|\,x) \\
&= \theta(f; c^S),
\end{aligned}
$$

where the fourth equality follows from the fact that $f \in \mathcal{F}_\xi^*$. Hence,

$$
\begin{aligned}
\underline{\theta}_{\mathrm{IV}}(c^S) \;&\equiv\; \min_{f \in \mathcal{F}_{\mathrm{IV}}^*} \quad \theta(f;\, c^S) \\
&=\; \min_{\xi \in \Xi} \left( \min_{f \in \mathcal{F}_\xi^*} \theta(f; c^S) \right) \quad \text{s.t.} \quad \mathcal{F}_\xi^* \neq \varnothing \\
&=\; \min_{\xi \in \Xi} \sum_{y_0, y \in \mathcal{Y}, z \in \mathcal{Z}, x \in \mathcal{X}} \omega(y,x)\; 1\{y_0 \neq y\}\; \xi(y_0, y, z \,|\, x) \qquad \text{s.t.} \\
&(i) \quad \sum_{y_0 \in \mathcal{Y}} \xi(y_0, y, z \,|\, x) \;=\; \Pr(Y_i = y, Z_i = z \,|\, X_i = x) \qquad \forall\, y, z, x \\
&(ii) \quad \sum_{y \in \mathcal{Y}} \xi(y_0, y, z \,|\, x) \;=\; \Pr(Z_i = z \,|\, X_i = x) \sum_{y \in \mathcal{Y}, z' \in \mathcal{Z}} \xi(y_0, y, z' \,|\, x) \qquad \forall\, y_0, z, x
\end{aligned}
\tag{A.8}
$$

**Part II**

An equivalent formulation of (A.8) is:

$$
\begin{aligned}
&\min_{\phi \in \Phi, \psi \in \Psi} \sum_{z \in \mathcal{Z}, x \in \mathcal{X}} \Pr(Z_i = z \,|\, X_i = x) \sum_{y_0, y \in \mathcal{Y}} \omega(y,x)\; 1\{y_0 \neq y\}\; \psi(y_0, y \,|\, z, x) \qquad \text{s.t.} \\
&\quad (i) \quad \sum_{y_0 \in \mathcal{Y}} \psi(y_0, y \,|\, z, x) \;=\; \Pr(Y_i = y \,|\, Z_i = z, X_i = x) \qquad \forall\, y, z, x \\
&\quad (ii) \quad \sum_{y \in \mathcal{Y}} \psi(y_0, y \,|\, z, x) \;=\; \phi(y_0 \,|\, x) \qquad \forall\, y_0, z, x
\end{aligned}
\tag{A.9}
$$

where $\Phi$ is the set of probability mass functions defined over $\mathcal{Y}$ for each $x \in \mathcal{X}$ and $\Psi$ is the set of probability mass functions defined over $\mathcal{Y} \times \mathcal{Y}$ for each $z \in \mathcal{Z}$ and $x \in \mathcal{X}$. To see that

(A.8) and (A.9) are equal, let $\xi^*$ solve problem (A.8) and define

$$\begin{aligned} \phi_{\xi^*}(y_0 \,|\, x) \quad &\equiv \quad \sum_{y \in \mathcal{Y}, z' \in \mathcal{Z}} \xi^*(y_0, y, z' \,|\, x) \qquad \forall y_0 \in \mathcal{Y},\ x \in \mathcal{X} \\ \psi_{\xi^*}(y_0, y \,|\, z, x) \quad &\equiv \quad \frac{\xi^*(y_0, y, z \,|\, x)}{\Pr(Z_i = z \,|\, X_i = x)} \qquad \forall y, y_0 \in \mathcal{Y},\ x \in \mathcal{X},\ z \in \mathcal{Z}. \end{aligned}$$

$\phi_{\xi^*}$ and $\psi_{\xi^*}$ are well-defined conditional probability mass functions and they clearly yield the same objective. Moreover, $\phi_{\xi^*}$ and $\psi_{\xi^*}$ are feasible in problem (A.9). Hence, (A.9) must be weakly smaller than (A.8). Conversely, let $\phi^*$ and $\psi^*$ solve problem (A.9) and define

$$\xi_{\psi^*}(y_0, y, z \,|\, x) \quad \equiv \quad \Pr(Z_i = z \,|\, X_i = x) \psi^*(y_0, y \,|\, z, x) \qquad \forall\, y, y_0 \in \mathcal{Y},\ x \in \mathcal{X},\ z \in \mathcal{Z}.$$

$\xi_{\psi^*}$ yields the same objective value and is well-defined: it is a probability mass function over $\mathcal{Y} \times \mathcal{Y} \times \mathcal{Z}$ for each $x \in \mathcal{X}$. Moreover, $\xi_{\psi^*}$ is feasible in problem (A.8). Hence, (A.8) must be weakly smaller than (A.9). It follows that (A.8) and (A.9) are equal.

Finally, problem (A.9) equals

$$\min_{\phi \in \Phi} \sum_{z \in \mathcal{Z}, x \in \mathcal{X}} \Pr(Z_i = z \,|\, X_i = x)\ \lambda(\phi, z, x) \tag{A.10}$$

where

$$\begin{aligned} \lambda(\phi, z, x) \quad = \quad & \min_{\gamma \in \Gamma} \sum_{y_0, y \in \mathcal{Y}} \omega(y, x)\ 1\{y_0 \neq y\}\ \gamma(y_0, y) \qquad \text{s.t.} \\ & (i) \quad \sum_{y_0 \in \mathcal{Y}} \gamma(y_0, y) \ = \ \Pr(Y_i = y \,|\, Z_i = z, X_i = x) \qquad \forall\, y \\ & (ii) \quad \sum_{y \in \mathcal{Y}} \gamma(y_0, y) \ = \ \phi(y_0 \,|\, x) \qquad \forall\, y_0 \end{aligned} \tag{A.11}$$

To see this, fix some $\phi \in \Phi$, and let $\gamma^*_{z,x}$ solve (A.11) for each $z \in \mathcal{Z}$ and $x \in \mathcal{X}$, given $\phi$. Because $\psi_{\gamma^*}$, defined as $\psi(y_0, y \,|\, z, x) \equiv \gamma^*_{z,x}(y_0, y)$, is feasible in (A.9) and yields the same objective given $\phi$, (A.9) is weakly smaller than (A.10). Similarly, let $\psi^*$ solve the (A.9) given $\phi$, and define $\gamma_{z,x}$ for all $z \in \mathcal{Z}$ and $x \in \mathcal{X}$ as $\gamma_{z,x}(y_0, y) \equiv \psi^*(y_0, y \,|\, z, x)$. For each $(z, x)$, $\gamma_{z,x}$ is feasible in (A.11) and $\big\{\gamma_{z,x} : z \in \mathcal{Z}, x \in \mathcal{X}\big\}$ yield the same objective value. Therefore, (A.10) is weakly smaller than (A.9). It follows that problems (A.9) and (A.10) are equal. ■

**Lemma 3.** Let Assumption IVx hold and let $\Phi$ be the set of probability mass functions defined over $\{1, \ldots, n_Y\}$ for each $x \in \mathcal{X}$. Then, given $x^* \in \mathcal{X}$, $\Pr(S_i = 1 \,|\, X_i = x^*) \geq \text{LB}(x^*)$ where

$$\text{LB}(x^*) \;=\; \min_{\phi \in \Phi} \sum_{y \in \mathcal{Y}, z \in \mathcal{Z}} \frac{1}{2} \Pr(Z_i = z \,|\, X_i = x^*) \Big| \Pr(Y_i = y \,|\, X_i = x^*, Z_i = z) \;-\; \phi(y \,|\, x^*) \Big|,$$

and $\Pr(S_i = 1 \,|\, Y_i = y, X_i = x^*) \geq 0$. These lower bounds are sharp.

In addition, suppose that $Z_i$ is binary, so that $\mathcal{Z} = \{0, 1\}$. Let $p_{\min} \equiv \min\big\{\Pr(Z_i = 0 \,|\, X_i = x^*), \Pr(Z_i = 1 \,|\, X_i = x^*)\big\}$. Then

$$\text{LB}(x^*) \;=\; \sum_{y \in \mathcal{Y}, z \in \mathcal{Z}} \frac{p_{\min}}{2} \Big| \Pr(Y_i = y \,|\, X_i = x^*, Z_i = 0) \;-\; \Pr(Y_i = y \,|\, X_i = x^*, Z_i = 1) \Big|$$

*Proof.* Fix $x^* \in \mathcal{X}$. Note that, under Assumption IVx, $\text{LB}(x^*) = \underline{\theta}_{\text{IV}}(c)$, where

$$c(y_0, y_1, s, z, x) = 1\{x = x^*\} \cdot 1\{s = 1\}.$$

By Lemma 2,

$$\underline{\theta}_{\text{IV}}(c) \quad = \quad \min_{\phi \in \Phi} \sum_{z \in \mathcal{Z}} \Pr(Z_i = z \,|\, X_i = x^*) \; \lambda(\phi, z, x^*) \tag{A.12}$$

where

$$\lambda(\phi, z, x^*) \;=\; \min_{\gamma\in\Gamma} \sum_{y_0, y\in\mathcal{Y}} 1\{y_0 \neq y\}\, \gamma(y_0, y) \qquad \text{s.t.} \tag{A.13}$$

$$(i) \quad \sum_{y_0\in\mathcal{Y}} \gamma(y_0, y) \;=\; \Pr(Y_i = y \,|\, Z_i = z, X_i = x^*) \qquad \forall\, y$$

$$(ii) \quad \sum_{y\in\mathcal{Y}} \gamma(y_0, y) \;=\; \phi(y_0 \,|\, x^*) \qquad \forall\, y_0$$

and $\Gamma$ is the set of probability mass functions defined over $\mathcal{Y} \times \mathcal{Y}$. Problem (A.13) is a Monge-Kantorovich transportation (optimal transport) problem with binary costs. With this particular cost structure, it admits a closed-form solution, given by half of the absolute distance between the marginal distributions (see Propositions 4.7 and 4.2 of Levin and Peres (2017) for a recent textbook treatment), so that:

$$\lambda(\phi, z, x^*) \;=\; \frac{1}{2}\sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \,|\, Z_i = z, X_i = x^*) - \phi(y \,|\, x^*) \Big|$$

In conjunction with (A.12), this result gives the desired lower bound for $\Pr(S_i = 1 \,|\, X_i = x^*)$, $\text{LB}(x^*)$.

If $Z_i$ is binary and $\mathcal{Z} = \{0, 1\}$, then for any $\phi \in \Phi$,

$$
\begin{aligned}
& \sum_{z \in \mathcal{Z}} \Pr(Z_i = z \,|\, X_i = x^*)\ \lambda(\phi, z, x^*) \\
= \quad & \Pr(Z_i = 0 \,|\, X_i = x^*)\ \lambda(\phi, 0, x^*)\ +\ \Pr(Z_i = 1 \,|\, X_i = x^*)\ \lambda(\phi, 1, x^*) \\
\geq \quad & p_{\min}\ \lambda(\phi, 0, x^*)\ +\ p_{\min}\ \lambda(\phi, 1, x^*) \\
= \quad & \frac{p_{\min}}{2} \cdot \Bigg( \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \,|\, Z_i = 0, X_i = x^*) - \phi(y \,|\, x^*) \Big| \\
& \qquad + \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \,|\, Z_i = 1, X_i = x^*) - \phi(y \,|\, x^*) \Big| \Bigg) \\
\geq \quad & \frac{p_{\min}}{2} \cdot \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \,|\, Z_i = 0, X_i = x^*) - \Pr(Y_i = y \,|\, Z_i = 1, X_i = x^*) \Big|,
\end{aligned}
$$

where $p_{\min} \equiv \min\big\{ \Pr(Z_i = 0 \,|\, X_i = x^*), \Pr(Z_i = 1 \,|\, X_i = x^*) \big\}$ and the last inequality follows from the triangle inequality, a property of the absolute ($\ell_1$, Taxicab, Manhattan) distance. But this lower bound is achieved by $\phi^*(\cdot, x^*)$, defined as:

$$
\phi^*(y \,|\, x^*) \ = \begin{cases} \Pr(Y_i = y \,|\, Z_i = 0, X_i = x^*) & \text{if} \quad p = \Pr(Z_i = 1 \,|\, X_i = x^*) \\ \Pr(Y_i = y \,|\, Z_i = 1, X_i = x^*) & \text{if} \quad p = \Pr(Z_i = 0 \,|\, X_i = x^*) \end{cases}
$$

Therefore,

$$
\mathrm{LB}(x^*) \quad = \quad \frac{p_{\min}}{2} \cdot \sum_{y=1}^{n_Y} \Big| \Pr(Y_i = y \,|\, Z_i = 0, X_i = x^*) - \Pr(Y_i = y \,|\, Z_i = 1, X_i = x^*) \Big|.
$$

On the other hand, given $y^* \in \mathcal{Y}$ and $x^* \in \mathcal{X}$, the sharp lower bound for $\Pr(S_i = 1 \,|\, Y_i = y^*, X_i = x^*)$ is $\underline{\theta}_{\mathrm{IV}}(c_y)$, where

$$
c_y(y_0, y_1, s, z, x) = \frac{1\{x = x^*, y_1 = y^*\}}{\Pr(Y_i = y^* \,|\, X_i = x^*)} \cdot 1\{s = 1\}.
$$

By Lemma 2,

$$\underline{\theta}_{\mathrm{IV}}(c_y) \quad = \quad \min_{\phi\in\Phi} \sum_{z\in\mathcal{Z}} \Pr(Z_i = z \,|\, X_i = x^*)\; \lambda(\phi, z, x^*) \tag{A.14}$$

where

$$\lambda(\phi, z, x^*) \;=\; \min_{\gamma\in\Gamma} \sum_{y_0\in\mathcal{Y}} \frac{1\{y_0 \neq y^*\}}{\Pr(Y_i = y^* | X_i = x^*)}\, \gamma(y_0, y^*) \qquad \text{s.t.} \tag{A.15}$$
$$(i) \quad \sum_{y_0\in\mathcal{Y}} \gamma(y_0, y) \;=\; \Pr(Y_i = y \,|\, Z_i = z, X_i = x^*) \qquad \forall\, y$$
$$(ii) \quad \sum_{y\in\mathcal{Y}} \gamma(y_0, y) \;=\; \phi(y_0 \,|\, x^*) \qquad \forall\, y_0$$

The constraints in problem (A.15) imply that

$$\sum_{y_0\in\mathcal{Y}} 1\{y_0 \neq y^*\}\gamma(y_0, y^*) \;-\; \sum_{y\in\mathcal{Y}} 1\{y \neq y^*\}\gamma(y^*, y) \;=\; \Pr(Y_i = y^* \,|\, Z_i = z, X_i = x^*) \;-\; \phi(y^* \,|\, x^*),$$

so that the optimal solution to problem A.15 is given by any feasible $\gamma^*$ such that:

$$\sum_{y_0\in\mathcal{Y}} 1\{y_0 \neq y^*\}\gamma^*(y_0, y^*) \;=\; \max\Big\{0, \quad \Pr(Y_i = y^* \,|\, Z_i = z, X_i = x^*) \;-\; \phi(y^* \,|\, x^*)\Big\},$$

so that

$$\lambda(\phi, z, x^*) \;=\; \max\left\{0, \quad \frac{\Pr(Y_i = y^* \,|\, Z_i = z, X_i = x^*) \;-\; \phi(y^* \,|\, x^*)}{\Pr(Y_i = y^* \,|\, X_i = x^*)}\right\}.$$

Now, define $\phi^*(y \,|\, x^*) = \Pr(Y_i = y \,|\, Z_i = z^{\max}, X_i = x^*)$, where $z^{\max} = \arg\max_z \Pr(Y_i = y^* \,|\, Z_i = z, X_i = x^*)$. For all $z \in \mathcal{Z}$, it follows that

$$\Pr(Y_i = y^* \,|\, Z_i = z, X_i = x^*) \;-\; \phi(y^* \,|\, x^*) \leq 0.$$

Thus, $\lambda(\phi^*, z, x^*) = 0$ for all $z$, and $\underline{\theta}_{\text{IV}}(c_y) = 0$. ∎

**Lemma 4.** Suppose that $Z_i$ is degenerate and define $\mathcal{Y} \equiv \{1, \ldots, n_Y\}$. Consider the linear parameter $\theta(f; c^S)$ associated with scalar coefficients $c^S$, such that $c^S(y_0, y_1, s, z, x) \equiv \omega(y_1, x)1\{s = 1\}$ and $\omega : \mathcal{Y} \times \mathcal{X} \mapsto \mathbb{R}$. Then

$$\begin{aligned}
\underline{\theta}_{\text{PMF}}(c^S) \quad &\equiv \quad \min_{f \in \mathcal{F}^*_{\text{PMF}}} \quad \theta(f;\, c^S) \\
&= \quad \min_{\gamma \in \Gamma} \sum_{y_0, y \in \mathcal{Y}, x \in \mathcal{X}} \omega(y, x)\; 1\{y_0 \neq y\}\gamma(y_0, y \,|\, x) \qquad \text{subject to:} \\
(i) \quad & \sum_{y_0 \in \mathcal{Y}} \gamma(y_0, y \,|\, x) \;=\; \Pr(Y_i = y \,|\, X_i = x) \qquad \forall\, y, x \\
(ii) \quad & \sum_{y \in \mathcal{Y}} \gamma(y_0, y \,|\, x) \;=\; \Pr(Y_i(0) = y \,|\, X_i = x) \qquad \forall\, y_0, x
\end{aligned}$$

*Proof.* The proof is analogous to that of Lemma 2. In this proof, I simply show the main steps to avoid repetition.

For a given $\gamma \in \Gamma$, let $\mathcal{F}^*_\gamma$ denote the subset of $\mathcal{F}^*_{\text{PMF}}$ that is consistent with $\gamma$:

$$\mathcal{F}^*_\gamma \equiv \Bigg\{ f \in \mathcal{F}^*_{\text{PMF}} \;: \sum_{y_1 \in \mathcal{Y}, s \in \{0,1\}} 1\{s y_1 + (1 - s) y_0 = y\} f(y_0, y_1, s \,|\, x) \;=\; \gamma(y_0, y \,|\, x) \quad \forall y_0, y, x \Bigg\}.$$

Thus,

$$\underline{\theta}_{\text{PMF}}(c^S) \quad = \quad \min_{\gamma \in \Gamma} \left( \min_{f \in \mathcal{F}^*_\gamma} \theta(f; c^S) \right) \quad \text{subject to} \quad \mathcal{F}^*_\gamma \neq \varnothing.$$

Moreover,

$$\mathcal{F}^*_\gamma \neq \varnothing \iff \begin{cases} \sum\limits_{y_0 \in \mathcal{Y}} \gamma(y_0, y|x) \;=\; \Pr(Y_i = y \,|\, X_i = x) \quad \forall y \in \mathcal{Y}, x \in \mathcal{X} \\ \\ \sum\limits_{y} \gamma(y_0, y|x) \;=\; \Pr(Y_i(0) = y \,|\, X_i = x) \quad \forall y_0 \in \mathcal{Y}, x \in \mathcal{X}. \end{cases} \tag{A.16}$$

On the other hand,

$$\theta(f_\gamma;\ c^S) \quad = \quad \min_{f \in \mathcal{F}^*_\gamma} \theta(f; c^S),$$

where

$$f_\gamma(y_0, y_1, s \,|\, x) \quad = \quad \begin{cases} 1\{y_0 \neq y_1\}\gamma(y_0, y_1 \,|\, x) \quad \text{if} \quad s = 1, \\ \\ 1\{y_0 = y_1\}\gamma(y_0, y_1 \,|\, x) \quad \text{if} \quad s = 0. \end{cases} \tag{A.17}$$

We therefore conclude that:

$$\begin{aligned} \underline{\theta}_{\text{PMF}}(c^S) \quad &= \quad \min_{\gamma \in \Gamma} \sum_{y_0, y \in \mathcal{Y}, x \in \mathcal{X}} \omega(y, x)\; 1\{y_0 \neq y\}\gamma(y_0, y \,|\, x) \qquad \text{subject to:} \\ &(i) \quad \sum_{y_0 \in \mathcal{Y}} \gamma(y_0, y \,|\, x) \;=\; \Pr(Y_i = y \,|\, X_i = x) \qquad \forall\; y, x \\ &(ii) \quad \sum_{y \in \mathcal{Y}} \gamma(y_0, y \,|\, x) \;=\; \Pr(Y_i(0) = y \,|\, X_i = x) \qquad \forall\; y_0, x. \end{aligned}$$

■

**Lemma 5.** Let Assumption PMFx hold and suppose that $Z_i$ is degenerate. Then, given $x^* \in \mathcal{X}$, the sharp lower bound for $\Pr(S_i = 1 \,|\, X_i = x^*)$ equals

$$\sum_{y \in \mathcal{Y}} \frac{1}{2} \Big| \Pr(Y_i = y \,|\, X_i = x^*) \;-\; \Pr(Y_i(0) = y \,|\, X_i = x^*) \Big|.$$

On the other hand, the sharp lower bound for $\Pr(S_i = 1 \,|\, Y_i = y, X_i = x^*)$ equals

$$\max\left\{0, \quad \frac{\Pr(Y_i = y^* \,|\, X_i = x^*) \;-\; \Pr(Y_i(0) = y^* \,|\, X_i = x^*)}{\Pr(Y_i = y^* \,|\, X_i = x^*)}\right\}.$$

*Proof.* Fix $x^* \in \mathcal{X}$. Note that, under Assumption PMFx, the sharp lower bound for $\Pr(S_i = 1 \,|\, X_i = x^*)$ is $\underline{\theta}_{\text{PMF}}(c)$, where $c(y_0, y_1, s, z, x) = 1\{x = x^*\} \cdot 1\{s = 1\}$. By Lemma 4,

$$\begin{aligned} \underline{\theta}_{\text{PMF}}(c) \quad &= \quad \min_{\gamma \in \Gamma} \sum_{y_0, y \in \mathcal{Y}} 1\{y_0 \neq y\}\, \gamma(y_0, y) \qquad \text{s.t.} \\ &\qquad (i) \quad \sum_{y_0 \in \mathcal{Y}} \gamma(y_0, y) \;=\; \Pr(Y_i = y \,|\, X_i = x^*) \qquad \forall\, y \\ &\qquad (ii) \quad \sum_{y \in \mathcal{Y}} \gamma(y_0, y) \;=\; \Pr(Y_i(0) = y_0 \,|\, X_i = x^*) \qquad \forall\, y_0 \end{aligned} \tag{A.18}$$

where $\Gamma$ is the set of probability mass functions defined over $\mathcal{Y} \times \mathcal{Y}$. Like problem (A.13) in Lemma 3, Problem (A.18) is an optimal transport problem with binary costs and a well-known solution:

$$\underline{\theta}_{\text{PMF}}(c) \quad = \quad \sum_{y \in \mathcal{Y}} \frac{1}{2}\Big|\Pr(Y_i = y \,|\, X_i = x^*) \;-\; \Pr(Y_i(0) = y \,|\, X_i = x^*)\Big|.$$

On the other hand, given $y^* \in \mathcal{Y}$ and $x^* \in \mathcal{X}$, the sharp lower bound for $\Pr(S_i = 1 \,|\, Y_i = y^*, X_i = x^*)$ is $\underline{\theta}_{\text{PMF}}(c_y)$, where

$$c_y(y_0, y_1, s, z, x) = \frac{1\{x = x^*, y_1 = y^*\}}{\Pr(Y_i = y^* \,|\, X_i = x^*)} \cdot 1\{s = 1\}.$$

By Lemma 4,

$$\underline{\theta}_{\text{PMF}}(c_y) \quad = \quad \min_{\gamma \in \Gamma} \sum_{y_0 \in \mathcal{Y}} \frac{1\{y_0 \neq y^*\}}{\Pr(Y_i = y^* | X_i = x^*)} \, \gamma(y_0, y^*) \qquad \text{s.t.}$$

$$(i) \quad \sum_{y_0 \in \mathcal{Y}} \gamma(y_0, y) \; = \; \Pr(Y_i = y \,|\, X_i = x^*) \qquad \forall \, y$$

$$(ii) \quad \sum_{y \in \mathcal{Y}} \gamma(y_0, y) \; = \; \Pr(Y_i(0) = y_0 \,|\, X_i = x^*) \qquad \forall \, y_0.$$

Now, notice that the constraints in this problem imply that

$$\sum_{y_0 \in \mathcal{Y}} 1\{y_0 \neq y^*\} \gamma(y_0, y^*) \; - \; \sum_{y \in \mathcal{Y}} 1\{y \neq y^*\} \gamma(y^*, y)$$
$$= \; \Pr(Y_i = y^* \,|\, X_i = x^*) \; - \; \Pr(Y_i(0) = y^* \,|\, X_i = x^*).$$

Thus, the optimal objective is attained by any feasible $\gamma^*$ that satisfies:

$$\sum_{y_0 \in \mathcal{Y}} 1\{y_0 \neq y^*\} \gamma^*(y_0, y^*) \; = \; \max\Big\{0, \quad \Pr(Y_i = y^* \,|\, X_i = x^*) \; - \; \Pr(Y_i(0) = y^* \,|\, X_i = x^*)\Big\},$$

so that

$$\underline{\theta}_{\text{PMF}}(c_y) \; = \; \max\left\{0, \quad \frac{\Pr(Y_i = y^* \,|\, X_i = x^*) - \Pr(Y_i(0) = y^* \,|\, X_i = x^*)}{\Pr(Y_i = y^* \,|\, X_i = x^*)}\right\}.$$

∎