Journal of Quantitative Criminology (2021) 37:35–72
https://doi.org/10.1007/s10940-019-09441-w

REVISIONS

Plea Bargaining and the Miscarriage of Justice

Michael Beenstock[1] · **Josh Guetzkow**[2] · **Shir Kamenetsky-Yadan**[3]

Published online: 13 December 2019

Abstract

Objectives We examine whether, on average, plea bargaining encourages guilty pleas among defendants who are factually innocent.

Methods We develop a formal theory of plea bargaining in which defendants take into account the possibility of false convictions or acquittals when making plea-bargain decisions. We use an incidentally truncated bivariate probit model to test the theory, which predicts that if innocent defendants plead guilty, the correlation (ρ) between the unobserved heterogeneity regarding selection into trial and regarding conviction at trial should be sufficiently positive. The method does not require knowledge of whether individual defendants are factually guilty or innocent. Since ρ is also predicted to vary directly with the unobserved toughness of prosecutors, we develop a decomposition theorem to distinguish between the effects of defendants and prosecutors in plea bargain decisions.

Results Using data on 2012 criminal cases decided in Israeli courts from 2010 to 2011, we find that ρ is large and positive. Hence, defendants who did not plea bargain were positively selected in terms of conviction. This means that defendants who accepted plea bargains had smaller counterfactual conviction probabilities than observationally similar defendants who went to trial.

Conclusions The results indicate that, on average, factually innocent defendants in Israel during this period took plea bargains instead of going to trial. This contradicts "innocence effect" theory, which predicts that factually innocent defendants, on average, reject plea bargains. Our findings are important for research on shadow trial theory, since they show that selection into plea bargains cannot be ignored when inferring counterfactual trial outcomes for plea bargainers.

Keywords Plea bargains · Shadow trial · Innocence effect · Selection bias · Bivariate probit

The authors thank the following people for their assistance and feedback: Oren Gazal-Ayal, Liraz Klauzner, Adi Leibovitch, Doron Teichman, Keren Weinshall-Margel and participants of the Criminal Justice Research Center workshop at The Ohio State University.

✉ Josh Guetzkow
joshua.guetzkow@mail.huji.ac.il

[1] Department of Economics, Hebrew University of Jerusalem, Jerusalem, Israel

[2] Department of Sociology and Anthropology, Institute of Criminology, Hebrew University of Jerusalem, Mt. Scopus, 91905 Jerusalem, Israel

[3] Statistician, Bank of Israel, Jerusalem, Israel

Introduction

Plea bargaining is on the rise around the world. According to Fair Trials International, the number of countries practicing plea bargaining increased from 19 in 1990 to 66 in 2017. Most countries have not reached the plea-bargaining rate of the US, where upwards of 95% of convictions are achieved through plea bargains. In Israel, for example, 78% of adjudicated cases are resolved via plea bargaining (calculated from Gazal-Ayal and Weinshall-Margel 2014). Reliance on plea bargaining has raised important concerns about the miscarriage of justice, because plea bargains may encourage innocent defendants to plead guilty (Alschuler 2015; Dervan and Edkins 2013; Rakoff et al. 2014). These concerns have been underscored by the large numbers of exonerated defendants in the US who pled guilty: 397 of the 2167 exonerations (18%) identified from 1989 through 2017 (National Registry of Exonerations 2017). A recent survey of inmates indicates that 6% of respondents believed they were completely innocent of their crimes (Loeffler et al. 2018), and other surveys have reported over one-third of prisoners with mental illnesses reported ever having made a false guilty plea (Redlich et al. 2010; see also Gudjonnson and Sigurdsson 2008; Zottoli et al. 2016).

Ultimately, the extent to which innocents plead guilty is unknowable, since only defendants know if they are factually innocent, and self-reports are unreliable (though see Loeffler et al. 2018). Proponents of plea bargaining argue that no injustice is done if innocents who pled guilty would likely have been convicted otherwise (Church 1979; Easterbrook 1992). As neither innocence nor counterfactual convictions are observable, it would appear at first blush that this issue lies beyond the reach of empirical inquiry. Moreover, the explanatory power of empirical models of plea bargaining and sentencing are rather low because they are influenced by numerous unobservable phenomena. These include: factual guilt or innocence; the strictness of judges; the toughness of prosecutors; errors of judgment; the over-confidence, short-sightedness and loss aversion of defendants; and numerous other phenomena that are inherently unobservable or fraught with measurement error. Indeed, when it comes to plea bargains and sentencing, what we do not observe is arguably more important than what we do. In this paper, we offer a novel approach to the study of plea bargaining that turns this problem to our advantage. What we do not observe in statistical models of plea bargaining and convictions sheds empirical light on the relation between plea bargaining and the miscarriage of justice.

Our approach treats plea bargains and trial outcomes as a problem of incidental truncation or sample selection. We start with the assumption that, on average, defendants who are factually guilty are more likely to be convicted if they stand trial, while the factually innocent are less likely to be convicted. We use a bivariate probit selection model to estimate the correlation, denoted by ρ, between the unobservable phenomena that influence selection into trial (i.e., when no plea agreement is reached) and the unobservable phenomena that influence conviction. If ρ is negative, defendants who select into trial are less likely than observationally similar defendants to be convicted due to unobservables, i.e. they are negatively selected in terms of conviction. This would mean that defendants who go to trial are, on average, more likely to be factually innocent than those who plea bargain. The opposite applies if ρ is positive. In that case, defendants who go to trial are positively selected because they are more likely to be convicted, indicating that they are more likely to be factually guilty compared to observationally similar defendants who plea bargain. We obviously do not know who is factually guilty or innocent, nor does the methodology require such knowledge.

It should be noted that although the incidentally truncated regression model (Heckman 1976) has been applied in a wide variety of areas (see e.g., Maddala 1983; Uggen 1999; Wynand and van Praag 1981), this is the first time it has been used in the study of plea bargaining. Our analysis also offers an important methodological innovation: we decompose ρ into constituent covariation terms in order to assess whether the miscarriage of justice contributes to the empirically estimated value of ρ.

A miscarriage of justice can be said to occur when factually innocent defendants accept plea bargains or are convicted at trial, and when factually guilty defendants are acquitted.[1] We show that this type of miscarriage of justice implies that ρ should be positive for two reasons. First, if ρ is positive it means that the factually guilty are over-represented among defendants who select into trials, and by implication, the number of factually innocent defendants who plea bargain increases. Second, we show that ρ should be positive if factually guilty defendants plead innocent in the hope that they will be acquitted in court. In either case, justice is miscarried. In the first, the factually innocent plead guilty. In the second, the factually guilty rely on judgement error to avoid conviction.

Our method can also be used to inform recent attempts to assess the "shadow trial" model (Abrams 2011; Bushway and Redlich 2012; Ulmer and Bradley 2006) and the "innocence effect" (Gazal-Ayal and Tor 2012). The shadow trial approach seeks to model the plea bargain decision as a rational one based on the perceived probable outcome of the trial. Rational defendants will accept plea bargains when the discounted sentence is smaller than the expected value of the sentence if the case goes to trial, which is a product of the probability of conviction and the likely sentence length. Innocence effect theory argues that, for a variety of reasons discussed below, innocent defendants are more likely to reject plea bargains and insist on going to trial to prove their innocence.

Both streams of research "require heroic assumptions about the comparability of those who go to trial and those who plead guilty" (Bushway et al. 2014, p. 750). Indeed, they assume by default that ρ is zero, i.e. that selectivity is ignorable. The estimation of ρ provides useful information about defendants' comparability, and we show how to correct for sample selection bias if they are not. If ρ is positive (negative), the counterfactual probability of conviction for defendants who accepted plea bargains is smaller (larger) than for observationally equivalent defendants who did not plea bargain. In other words, the estimation of ρ allows us to answer the question: if defendants who accepted plea bargains instead rejected them, how would their conviction probability compare to observationally similar defendants who did go to trial? If, for example, ρ is positive, the factually innocent are more likely to self-select into plea bargains and plead guilty, and the factually guilty are more likely to be tried in court. It follows, therefore, that if plea bargainers had been tried, their conviction probability would have been lower than that of observationally similar defendants who went to trial.

We show that innocence effect theory predicts that ρ should be negative, because on average the factually innocent wish to prove their innocence in court where they expect to be acquitted. We also show that the canonical version of shadow trial theory (Landes 1971; Nagel and Neef 1979) predicts that ρ should be positive, because prosecutors are less likely to offer attractive sentence discounts to defendants who they think are more likely to be convicted in court. Therefore, the estimation of ρ sheds empirical light on these theories.

[1] We do not mean to suggest that these are the only types of injustice that occur in criminal courts, but other forms of injustice fall outside the focus of our research.

We illustrate our approach by estimating ρ for a stratified random sample of criminal cases decided in Israeli courts during 2010 and 2011. To our knowledge, this is the first study of plea bargaining that uses data on jurisdictions outside the US. We find that ρ is large and positive, which is inconsistent with innocence effect theory. Our novel decomposition of ρ also shows that it is too large to be accounted for by shadow trial theory alone. Consequently, our results indicate a miscarriage of justice in Israel during this period.

Recent research on shadow trial theory and the innocence effect is part of a broader renaissance of research on plea bargaining (for reviews see Johnson et al. 2016; Redlich et al. 2017b). Redlich et al. (2017b, p. 465) conclude that "the most pressing need on the theoretical front is for more formal modelling of the predominant theories." Our study responds to this call. However, it also stands apart in important respects. Existing research on plea bargaining primarily focuses on which factors affect the decision-making of defendants, defense attorneys, prosecutors, judges and courtroom workgroups. This research can tell us, for example, what motivates prosecutors to make plea deals or go to trial (Bandyopadhyay and McCannon 2015), how plea-bargain rates vary by the race and ethnicity of defendants (Kutateladze et al. 2014), or what kinds of cognitive biases might affect defendant decision-making (Bibas 2004; Redlich et al. 2017a). This line of research details the push and pull factors that make plea bargains more or less likely but ultimately cannot answer the question we address here about the miscarriage of justice. To address this question, investigators have increasingly turned to experimental methods using mock trials and hypothetical vignettes (Bushway et al. 2014; Dervan and Edkins 2013). Although these studies have important advantages, such as the ability to manipulate case and defendant characteristics, their applicability to the complexity of real-world cases involving actual defendants and prosecutors remains an open question. By contrast, we use empirical data on actual cases decided in criminal courts.

In the next section we introduce the key unobservable phenomena that play a central role in our theory. We also discuss how the observable phenomena that previous researchers have used in their models of sentencing fit into our theory. Then we build a formal selection model to test miscarriage of justice theory, shadow trial theory, and innocence effect theory. Next, we describe our data and methods. After presenting our estimate of ρ, we calculate counterfactual conviction probabilities for different subsets of defendants. We then decompose ρ into constituent covariance components and bound their contribution. We conclude with implications for existing and future research, as well as for policy.

The Miscarriage of Justice: Building Blocks of a Formal Theory

The dominant model of individual plea-bargaining decisions recognizes that they take place in the "shadow of the trial" (Bibas 2004; Bushway and Redlich 2012; Landes 1971; Nagel and Neef 1979; Smith 1986). Broadly speaking, this means that when defendants decide whether to accept or reject plea bargains (and when prosecutors offer plea deals), they take into account the expected outcome of the case if it were to go to trial. We accept this basic proposition. However, in contrast to recent work on shadow trial theory (Abrams 2011; Bushway and Redlich 2012; Bushway et al. 2014; Redlich et al. 2016; Ulmer and Bradley 2006), we are not concerned here with comparing expected versus observed plea discounts[2] or examining what factors influence judge or prosecutorial decision making. We

[2] These studies regress sentences (S) on controls (X) and a plea bargain dummy (PB): $S_i = \alpha + X_i\beta + \theta PB_i + u_i$. If plea bargainers are negatively (or positively) selected, OLS estimates of θ are

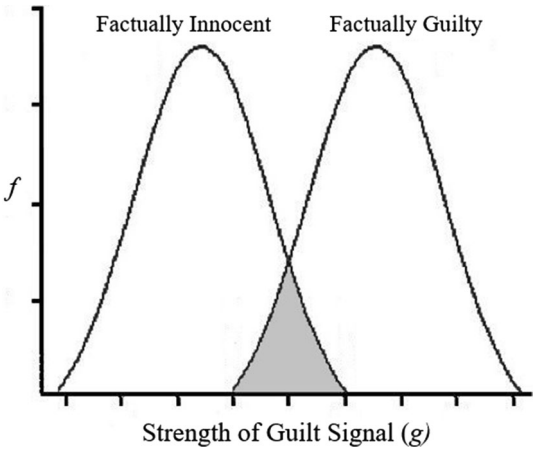

Fig. 1 A bimodal distribution of guilt signals for the factually innocent and guilty

focus instead on identifying the unobservable phenomena that affect plea and conviction outcomes in terms of their relationship to ρ. In the following, we discuss the various considerations that defendants, prosecutors, judges and juries might take into account when making plea and conviction decisions, and we offer hypotheses about how the elements of plea and conviction decisions are expected to co-vary. Only then do we elaborate a formal presentation of our model.

Signals of Guilt and Prosecutors' Toughness

We start with the widely accepted premise that there is a distinction between factual and legal innocence (Hoffman 2007) and that defendants know if they are factually innocent or guilty. We recognize that factual guilt and innocence may not always be clear-cut in every single case, even for defendants. For our purposes, it is enough to define factually innocent defendants as those who are not culpable in any way, while factually guilty defendants are those who are culpable to any degree. Henceforth we will refer to *factual* guilt or innocence simply as guilt or innocence, unless we need to distinguish it from *legal* guilt or innocence. (Legal innocence refers to cases where defendants do have some degree of culpability but a case cannot be made against them for legal reasons, such as lack of evidence, a statute of limitations, or procedural technicalities).

We draw on signaling theory (Spence 1973) to express the relation between factual guilt or innocence, which are unobservable, and what we refer to as a "guilt signal" perceived by judges, denoted by g. The guilt signal includes all aspects of the case upon which judges decide to convict that are ultimately unobservable to researchers.[3] Positive realizations of g indicate guilt, and negative realizations indicate innocence. We assume that the signal is informative: on average, guilty defendants have stronger guilt signals than innocent ones.

Footnote 2 (continued)
biased downward (or upward). The same issue arises in Bushway et al. (2014) where they use experimental data for S. They implicitly assume that the responses are unrelated to the residuals.

[3] While this could refer to juries as well, we refer to judges for simplicity and because Israel does not have juries.

Whereas "factual guilt" is naturally dichotomous, the guilt signal (g) is assumed to have a continuous distribution over a range including positive and negative values. This distribution is expected to be bimodal with some innocents appearing guilty and vice versa, as depicted in the shaded area in Fig. 1.

A key component of g, featured in shadow trial theory, is the strength of evidence: when the evidence is stronger, defendants are more likely to be convicted, and trial discounts reduced (Bushway et al. 2014; Kutateladze et al. 2015). In terms of our theory, g increases with the strength of incriminating evidence and decreases when the evidence favors innocence. Attempts to measure the strength of evidence (Bushway and Redlich 2012; Bushway et al. 2014; LaFree 1985; Smith 1986) are inevitably imperfect. For example, Bushway et al. (2014) found that although three out of four measures for strength of evidence were statistically significant in models predicting conviction probabilities, their explanatory contribution was modest. Furthermore, in models used to estimate counterfactual conviction probabilities for plea bargainers, evidence indicators were either not significant, or had the opposite sign from what was expected. Since our data do not include any direct indicators of evidence, g assumes the entire role of unobservable data on the strength of evidence. If our models did include direct measures for the strength of evidence, g would assume the role of unobservable measurement error.

We distinguish between guilt signals perceived by judges at trial, denoted by g, and guilt signals perceived by defendants at the plea-bargaining stage, denoted by g'. In deciding whether to accept a plea deal, defendants assess their guilt signal, which plays a role in their decision.[4] If g' is less than g, we can say that defendants are over-confident or in denial (Redlich et al. 2017a, b). We assume that defendants' assessments of their guilt signals are positively correlated with judges'. Other things being equal, defendants who believe they have stronger guilt signals are more likely to accept plea bargains in order to avoid trial. Were they to go to trial, they would face a higher probability of conviction compared to defendants with weaker guilt signals.

Prosecutors also assess defendants' guilt signals at the plea-bargaining stage to estimate the probability of a conviction if the case went to trial. We assume that their assessments are positively correlated with g.[5] Following Bushway et al. (2014), we assume that prosecutors are tougher on defendants with stronger guilt signals; they drive harder bargains by offering smaller plea discounts (or none at all), inducing such defendants to reject plea offers. We define the toughness of a prosecutor in terms of the generosity of the plea discounts they offer: other things being equal, prosecutors can be said to be tougher when they offer less generous plea discounts. If prosecutors are tougher on defendants with stronger guilt signals, we would expect to see a positive correlation between their toughness, denoted by d, and the strength of guilt signals. In what follows, we refer to this phenomenon as STT (shadow trial theory), because it is consistent with the shadow trial hypothesis that plea discounts are lower when the probability of conviction is higher (due to stronger guilt signals). Of course, many

[4] Typically, defendants are advised by their attorney about this and other matters, but the distinction between defendants and their attorneys can be ignored for the sake of simplicity, both because defendants ultimately make the final decision and because their perceptions and decisions are almost certainly correlated with their attorney's advice overall.

[5] Of course it is possible that prosecutors' assessments of guilt will sometimes diverge from that of judges. Even so, as long as they agree more often than not, our assumption of positive correlation holds. Ideally we could model this explicitly with fixed effects for judges and prosecutor and interactions between them. However, our data only contain identifying information on judge.

things can contribute to prosecutorial toughness, such as election concerns (Bandyopadhyay and McCannon 2014). However, our concern here is not with what influences prosecutorial toughness in general, but rather on analyzing how (unobservable) toughness influences ρ.

Whereas g is unobservable, sentencing research has pointed to many observable factors that increase the likelihood of conviction, including defendant and case characteristics (see e.g., Spohn 2013; Ulmer and Bradley 2017; Walker et al. 2012). We draw on this research to specify the covariates of the bivariate probit model. We also incorporate courtroom workgroup perspectives (Johnson et al. 2016) by including court fixed effects and courts' caseload pressure: prosecutors are more likely to press for plea bargains in courts with higher caseload pressures (Ulmer et al. 2010; Wooldredge 1989).

Innocence and Defendants' Perception of Judgment Error

Justice is not only miscarried if factually innocent defendants plead guilty; it is also miscarried when factually innocent defendants are falsely convicted at trial and when factually guilty defendants are acquitted. We denote the former probability by θ and the latter by λ. Other things being equal, the more that innocent defendants believe that the judge is likely to convict them in error (i.e., the larger is θ), the more likely they are to accept plea bargains. Also, the more that guilty defendants believe the judge will err in their favor and acquit them (i.e., the larger is λ), the more likely they are to reject plea bargains. Just as g' might be affected by over-confidence, denial and loss-aversion, so might θ and λ be affected (Redlich et al. 2017a, b). For example, λ may be larger among over-confident defendants and θ and λ may be smaller due to loss aversion. We show below that if defendants' perceptions of judgment error are positively correlated with actual error, and if judgment error is correlated with guilt signals, factually innocent defendants might be induced to plead guilty, and factually guilty defendants might be induced to plead innocent.

Gazal-Ayal and Tor (2012) have argued that, ceteris paribus, innocent defendants are more likely to reject plea bargains compared to guilty defendants. Explanations for this "innocence effect" center on the fact that innocent defendants may: (1) reach legitimately different assessments of trial outcomes due to private information; (2) be irrationally optimistic about their prospects (i.e., underestimate θ); and/or (3) maintain their innocence as a matter of principle (Gross 2011; Tor et al. 2010). The first and second explanations can be viewed as cases where innocents perceive their chances of false conviction to be lower, and so are more likely than guilty defendants to reject plea bargains in favor of trials. In other words, innocent defendants are more likely to expect the court to "get it right." This is also consistent with the notion that innocent defendants are more likely to frame potential punishment as a loss, which according to prospect theory would make them more likely to go to trial (Redlich et al. 2017a, b).

If, however, factually innocent defendants with weaker guilt signals fear that judges are more likely to "get it wrong", they will prefer to accept plea bargains. This would be consistent with experimental findings that have produced false guilty plea rates in excess of 50% (Dervan and Edkins 2013). The same applies to the factually guilty. If these defendants with stronger guilt signals think they might go free, they will reject plea bargains that they would otherwise have accepted.

Cost of Going to Trial

Another consideration frequently mentioned in the shadow trial literature is the cost of going to trial. There is a consensus that potential trial costs generally act as a deterrent to all parties against going to trial (Gazal-Ayal and Riza 2009; Gross 2008, 2011; Johnson et al. 2016). Litigation costs should make prosecutors willing to offer more generous deals, but also make defendants more willing to accept less attractive plea bargains. Therefore, the likelihood of going to trial should vary inversely with litigation costs. We assume that the covariation between litigation costs and guilt signals, if any, is positive, since defendants who appear guiltier may need to invest more in their defense.

A Formal Model of Miscarriage of Justice Theory

Sample Selectivity

In this section, we draw on the unobservable phenomena in the preceding discussion to develop a formal sample selection model. We treat plea bargaining as a source of incidental truncation into trial.[6] Ideally, we wish to know the counterfactual trial outcomes for defendants who pled guilty. In "Appendix 1" we provide nonparametric bounds[7] for these counterfactuals, which turn out to be wide in the absence of identifying assumptions. If sample selection into trials is based on observables alone, the method of inverse probability weighting (IPW) would correct for sample selection bias in conviction outcomes (Greene 2012, pp. 776–778), and what is not observed is ignorable.

We propose a theory in which what is not observed in selection into trial is not ignorable, because it is suspected of being correlated with what is not observed in conviction outcomes. This correlation is responsible for inducing hidden confounders in selection into trial. Therefore, IPW is inappropriate a priori because it imposes restrictions that are unnecessarily strong. Instead, we draw on the incidentally truncated regression model (ITM) (Heckman 1976), in which the outcome variable is dichotomous rather than continuous (Wynand and van Praag 1981), and which, in contrast to IPW, does not assume that unobservables in selection are ignorable. Nevertheless, we compare results obtained by IPW, which are sensitive to hidden confounders, with results obtained by our preferred methodology.

Since results estimated by ITM may be sensitive to parametrization, we carry out robustness checks for results obtained using the baseline of the bivariate normal distribution. In this context we use bivariate logit distributions as well as copula methods, which show that our results are insensitive to these parametric alternatives.

Y_2 is a dummy variable, which equals 1 if defendants are convicted at trial, and zero otherwise. Let Y_2^* denote a latent variable such that $Y_2 = 1$ if $Y_2^* > 0$ and $Y_2 = 0$ otherwise:

[6] Truncated outcomes are not observed for reasons unrelated to the outcome of interest, whereas incidental truncation arises when outcomes are not observed for reasons that might be related to the outcome of interest. Since conviction outcomes are unknown for plea-bargainers, and plea-bargain decisions are related to these unknown outcomes, their counterfactual trial outcomes are incidentally truncated. This is also known as endogenous selection.

[7] Nonparametric bounds (Manski 1995) involve no identifying assumptions regarding what motivates selection decisions or how unobserved phenomena are distributed.

$$Y_{2i}^{*} = X_i\beta + y_{2i} \tag{1}$$

where X is a vector of observed covariates hypothesized to determine the probability of conviction, such as the characteristics of defendants, type of indictment, existence and identity of victim, type and weight of evidence, the identity of judges and courts, etc., y_2 is a random variable, which is not observed by researchers, and i labels defendants. We propose the following auxiliary hypothesis for y_2:

$$y_{2i} = g_i + e_i \tag{2a}$$

$$e_i = \pi\theta_i - (1-\pi)\lambda_i \tag{2b}$$

Recall that g denotes guilt signals, and e denotes errors of judgment, which has two components in Eq. (2b), where θ denotes the probability conditional on g that innocents are convicted (false-positive), λ denotes the conditional probability that the guilty are acquitted (false-negative), and π is the unknown proportion of innocents among defendants. Taken together, Eqs. (2a) and (2b) mean that innocents may be convicted when g and θ are large, and the guilty acquitted when g is small and λ is large.

Y_1 is a dummy variable, which equals 1 if no plea bargain is reached (selection into trial) and equals zero if there is a plea bargain. Let Y_1^* denote a latent variable such that $Y_1 = 1$ if $Y_1^* > 0$, and $Y_1 = 0$ otherwise, where:

$$Y_{1i}^{*} = Z_i\gamma + y_{1i} \tag{3a}$$

$$y_{1i} = -g_i' - c_i + d_i - e_i' \tag{3b}$$

Z is a vector of observable covariates hypothesized to determine failure to reach a plea bargain, such as defendant characteristics, court caseloads, etc., y_1 is a random variable not observed by researchers, and g' and e' denote defendants' subjective perceptions of g and e. The auxiliary hypothesis for y_1 (Eq. 3b) has four components: defendants' perceptions of their guilt signals (g'), their litigation costs (c), the prosecutors' toughness (d), and their assessments of judgment error (e'). According to Eq. (3b), defendants with stronger guilt signals, higher litigation costs, and greater expectations of judgment error are less likely to select into trial (in other words, they are more likely to accept plea offers), while defendants with tougher prosecutors are more likely to go to trial. One may add to Eq. (3b) any other unobservable phenomena, such as loss-aversion, base-rate bias, etc., which might influence plea bargain decisions (Redlich et al. 2016; Redlich et al. 2017b). If these phenomena are independent of y_2 they are ignorable. For this reason, we do not include them in Eq. (3b).

In Eqs. (2a) and (3b) y_1 and y_2 depend linearly on signals of guilt if g is positive and innocence if it is negative. When g is close to zero, case quality is naturally more ambiguous than when g is strongly positive or negative. Since the tails of the logistical distribution are fatter than those of the normal distribution, a logit specification would be more sensitive to case quality than the probit. Although our main results are probit based, we show that they are statistically indistinguishable from their logit counterparts, which suggests that case ambiguity as defined is not empirically important in our data. This inference is weak because y_1 and y_2 depend on many other components apart from g, and because, e.g., Pareto and Burr distributions have fatter tails than the logistic, which might be more sensitive to case quality.

The correlation between y_2 and y_1 can be written as:

$$\rho = -\sigma_{gg'} - \sigma_{ee'} - \sigma_{ge'} - \sigma_{g'e} + \sigma_{gd} - \sigma_{gc} + \sigma_{de} - \sigma_{ce} \tag{4}$$

where for example, $\sigma_{gg'}$ denotes the covariance between guilt signals perceived by judges and defendants. If g and g' and e and e' are positively correlated, the first two terms in Eq. (4) are negative.[8] If all other covariances happen to be zero, then ρ must be negative, which implies, on average, that innocents try to prove their innocence in court, while the guilty accept plea bargains. If defendants with stronger guilt signals (larger g and g') are more likely to be convicted because of judgment error (larger e and e') the third and fourth covariances will be negative. If prosecutors are tougher on defendants with stronger guilt signals, then σ_{gd} is positive, because prosecutors' perceptions of g are correlated with g. If litigation costs vary directly with signals of guilt then σ_{gc} is positive. The remaining covariance terms are assumed to be zero, since there is no obvious reason why the toughness of prosecutors and litigation costs should be related to judgment error. The size and sign of ρ are empirical matters, which cannot be determined a priori, even if the signs of the individual covariances were known.

If y_1 and y_2 are bivariate normal random variables with correlation ρ, the probability of conviction conditional on selection into trial is:

$$P_i = \Phi\left[\frac{X_i\beta + \rho Z_i\gamma}{\sqrt{1-\rho^2}}\right] \tag{5a}$$

where $\Phi(z)$ denotes the cumulative standard normal density (Greene 2012, p. 790). Notice that if $\rho = 0$ Eq. (5a) states that the probability of conviction is simply $\Phi(X_i\beta)$, as expected. In this case, consistent estimates of γ may be obtained by estimating Eq. (1) by probit as in Bushway and Redlich (2012), because selection is quasi random and therefore ignorable. If, however, ρ is non-zero, probit estimates of β are biased and inconsistent. Wynand and van Praag (1981) suggested a maximum likelihood estimator for β, γ and ρ under the assumption that y_1 and y_2 are bivariate normal. In this context Greene (2012, p. 933) comments, "For better or for worse, the empirical literature on the subject continues to be dominated by Heckman's original model built around the joint normal distribution." However, we combine joint normality with the use of an instrumental variable to identify β, γ and ρ, as described in more detail in the methods section, and we carry out robustness checks with respect to parametric alternatives.

Counterfactuals

We may use the selection model to estimate the counterfactual probability of conviction had defendants with plea bargains been tried. This counterfactual expresses the conviction probability in the shadow of the trial. The expected value of Y_{2i}^* for the bivariate normal is:

$$E\left(Y_{2i}^*/Y_{1i}\right) = X_i\beta + \rho\lambda_i \tag{5b}$$

[8] Recall, for example, that the correlation between g and g' is $r_{gg'} = \sigma_{gg'}/\sigma_g\sigma_{g'}$, where σ_g denotes the standard deviation of g. Therefore, correlations vary directly with covariances.

$$\lambda_i = \left(2Y_{1i} - 1\right)\frac{\phi\left[\left(2Y_{1i} - 1\right)Z_i\gamma\right]}{\Phi\left[\left(2Y_{1i} - 1\right)Z_i\gamma\right]} \tag{5c}$$

where ϕ(z) denotes the standard normal density. Equations (5b) and (5c) may be used to calculate shadow trial conviction rates for plea bargainers by setting $Y_{1i}=0$. To illustrate we assume that $X_i\beta=1$ and $Z_i\gamma=0.3$. If $Y_{1i}=1$ and $\rho=0.5$, the expected value of $Y_{2i}^{*}=1.171$ and the probability of conviction is 0.879. If $Y_{1i}=0$ (the counterfactual for plea bargainers) the expected value of Y_{2i}^{*} is 0.5 and the probability of conviction is 0.692. The shadow trial conviction rate is less than for observationally similar defendants who were tried because the latter are positively selected in terms of conviction. If instead $\rho=-0.5$, matters are reversed; the shadow trial conviction probability is 0.879 instead of 0.692 because defendants who were tried are negatively selected in terms of conviction.

Quantitative Versus Qualitative Results

These shadow trial conviction probabilities are not directly informative about the proportion of factually innocent plea bargainers because counterfactual judgments may be subject to error. Since ρ depends, *inter alia*, on judgment error (e), factually innocent plea bargainers might be falsely convicted at trial in their counterfactual, and their factually guilty counterparts might be falsely acquitted. It would only be possible to quantify the proportion of factually innocent plea bargainers if plea bargain decisions and convictions depended upon g alone, i.e. c, d and e are zero, or, there are no litigation costs, prosecutors have no role in plea bargaining, and judgments are always correct. The latter means that the guilty are convicted, and the innocent are always acquitted.

Suppose, for example, that there are 100 defendants of which 70 accepted plea bargains and 30 were tried. If the conviction rate among the latter is 2/3, there must be at least 20 guilty defendants and 10 innocent defendants, because judgements are always correct. If ρ is positive, the observed conviction rate (2/3) exceeds its true value because it embodies positive sample selection bias equal to $b = \rho\bar{\lambda}$, i.e. guilty defendants are over-represented in trials, and innocent defendants are over-represented in plea bargains. Since 70% of defendants accepted plea bargains, $\bar{\lambda} = 0.525$. If $\rho=0.317$, $b=1/6$, in which case the true conviction rate is 2/3–1/6 = 1/2. This means that half the defendants are innocent. Since 10 of them were acquitted in court, there must be 40 plea bargainers who are innocent, and 30 who are guilty. Therefore, 57% of plea bargainers are innocent. This proportion varies directly with ρ.

Since ρ does not depend on g alone, but also depends on litigation costs, the toughness of prosecutors, judgment error, and other phenomena such as over-confidence, risk aversion and endowment effects, it is impossible to quantify the miscarriage of justice. However, it is possible to express it qualitatively using a decomposition theorem for ρ (see below under "Decomposing ρ: Does the Miscarriage of Justice Account for the Results? section"). We therefore qualify our results by stating that "on average" plea bargainers are factually innocent.

The Court System in Israel

The Israeli legal system grew out of the system established during the British Mandate over Palestine (1920–1948). It therefore resembles the legal system in many other common law countries. There are, however, key differences between the Israeli and US criminal justice

systems. The most important with respect to plea bargaining is that in Israeli courts, prosecutors must list all evidence at the time of indictment and are not allowed to submit additional evidence later. Consequently, unlike in most US jurisdictions, Israeli defendants know all the evidence against them when they make plea bargain decisions. This is important, because some proposals for plea bargaining reform in the US emphasize that defendants' ignorance of the evidence against them might induce the innocent to plead guilty (Alschuler 1983; Turner and Redlich 2016).

Another difference is that there are no jury trials in Israel. Our method is still applicable to jurisdictions with trials by jury, because juries' perceptions of guilt signals should be positively correlated with that of judges. It may be that juries are more prone to error, but we would still expect their errors to correlate with judges'. The absence of juries in Israel tends to prolong court proceedings. Instead of single trials held over the course of a few days or less so as not to inconvenience the jury, trials in Israel tend to consist of multiple hearings held over weeks or months, with witnesses typically being called as available. Generally speaking, the more witnesses that testify in a case, the more hearings there are. We have confirmed this through discussions with several current and former defense attorneys and prosecutors. This means that the number of hearings may be taken as a rough proxy for the number of witnesses, though of course there can be other reasons for additional hearings.

All courts in Israel belong to a single national system and are centrally administered by the Judicial Authority, which is part of the Ministry of Justice. For criminal cases, there are three levels of courts: magistrate, district and supreme. The Supreme Court acts as a court of appeals, while the magistrate and district courts deal with less and more severe crimes, respectively. Magistrate courts have jurisdiction where the criminal charge carries a potential sentence of up to 7 years imprisonment. District courts deal with cases with higher potential sentences. There are 29 magistrate courts and 5 district courts. In 2017, there were 44,484 new non-traffic related adult criminal cases opened in magistrate courts, and 2794 in district courts (Israeli Courts Administration 2018).

Crimes in Israel are classified as either felonies, misdemeanors, or "contraventions." A felony carries a minimum punishment of more than 3 years in prison; a misdemeanor has a minimum punishment of imprisonment of more than 1 month and up to 3 years; and a contravention has a maximum punishment of imprisonment up to 1 month. Israel does not have uniform sentencing guidelines, nor any guidelines governing plea negotiations. Prosecution of cases is divided between police and state prosecutors. Police prosecutors are responsible for trying misdemeanor and contravention cases, while district attorneys from the State Attorney's Office in the Ministry of Justice are responsible for trying felony cases, which are approximately 10% of all criminal cases (Office of the State Attorney 2018). The Office of the Public Defender was established in 1996. Public defenders are provided for all defendants whose charges could result in prison or have a severe impact on their lives. Since there are only about 150 full-time public defenders, in most cases private attorneys are paid a flat fee by the state to provide legal representation.

Discussions we had with several current and former prosecutors and defense attorneys allow us to paint a portrait of the plea bargain process in Israel as it is actually practiced. At the level of the magistrate courts, where the vast majority of cases are tried, there is an initial meeting between the defense counsel and the prosecuting attorney to discuss their cases. These pre-hearing meetings typically cover many cases and only a brief amount of time is devoted to each individual case. Most plea agreements are typically struck at this stage, although plea agreements can be reached prior to this stage or at any point prior to a final verdict. Typically, the defense attorney proposes a plea deal, but often the prosecutor will seek a deal when they see weaknesses in their case. Gazal-Ayal and Weinshall-Margel (2014) found

that about 30% of plea bargains were for a reduced sentence in exchange for a plea of guilty to all charges, 15% involved a change or reduction in charges, while the remainder involved a combination of these two or an agreement over legal procedures or stipulation of facts.

In many cases, the attorneys fail to reach any agreement. In other cases, defendants later reject the terms that the prosecutor demanded, often hoping that they will be treated more leniently by the judge if they admit their guilt and throw themselves on the mercy of the court. There are no data available on how frequently these two scenarios occur, and the informal estimates given by our informants varied considerably. Our analysis of the data used in this paper (Gazal-Ayal and Weinshall-Margel 2014) shows that a plea bargain was reached in 61% of cases where the data indicated that there had at least been an opportunity for a plea agreement (see discussion of excluded cases in "Data" section). It remains unknown what portion of the remaining 39% were either never presented with or ultimately rejected a plea agreement. It should be noted that the 61% plea-bargain figure is much lower than the 78% figure we cited in the introduction, even though they come from the same data source. The reason is that the 78% figure is calculated using as a denominator all cases where a final judgement occurred, whereas the denominator for the 61% figure includes all cases where there was at least an opportunity in principle to reach a plea agreement, regardless of the subsequent outcome of the case.

Data

Our data come from a study conducted by Gazal-Ayal and Weinshall-Margel (2014), who compiled information on 2012 criminal cases decided in the courts of Israel between May 5, 2010 and May 5, 2011 for a descriptive study of Israeli case processing. These data come from a random sample stratified by court. It includes 3% of criminal cases decided in magistrate courts and 13% of criminal cases decided in district courts. Coding of case information was conducted by 13 law students from the University of Haifa based on the indictments, statements of defense, requests, decisions and protocols of the court hearings in each case. More than 10% of the files were coded by at least two coders with an inter-coder reliability exceeding 90%.

We dropped observations from the analysis if there was no option for a plea bargain. This includes 141 Palestinians from the West Bank who were charged solely for being in Israel without a permit, an offense for which there are never plea bargains. Even if these cases were included, they would be dropped from the model, since they predict selection into trial and conviction perfectly. There were an additional 12 cases where no hearing ever took place, which is also an indication that there was no opportunity for a plea bargain. The remaining 5 cases had missing data on plea bargains. The net sample size after these cases were excluded is 1854.

Dependent Variables

The outcome variable for Y_1 (selection into trial) is whether or not the defendant agreed to a plea bargain. $Y_1 = 0$ when a plea bargain was reached, which was the case in 1137 or about 61% of remaining cases (see Table 1). $Y_1 = 1$ in 717 cases that went to trial.

The outcome variable for Y_2 (conviction) indicates whether defendants who did not reach plea bargains were convicted ($Y_2 = 1$). Of the defendants who did not reach plea agreements, 228 were convicted, which is 32% of all non-plea bargain cases. $Y_2 = 0$ in 213

Table 1 Case dispositions

Case disposition	N
Plea bargains	1137
Trials	717
Total convictions	228
Total acquittals	213
De jure acquittals	8
De facto acquittals	205
Prosecution withdrew charges	142
Prosecution cancelled indictment	63
Total censored cases	276
Defendant not located	145
Stay of proceedings	15
Suspension of proceedings	28
Defendant deemed incompetent to stand trial	49
Other or N/A	39
Subtotal	1854
Excluded from the analysis	158
Total N	2012

cases, which is nearly 30% of all non-plea bargain cases. This category includes both *de jure* and *de facto* acquittals. There are only 8 cases where defendants were formally acquitted at trial. We counted withdrawn charges (n = 142) and canceled indictments (n = 63) as *de facto* acquittals. The reason is that, from a legal perspective, such outcomes are equivalent to *de jure* acquittals, apart from the issue of double jeopardy. They are also a meaningful indication of the strength of guilt signals, since prosecutors typically cancel indictments and withdraw charges when they feel the case is not strong enough to go to trial. They do this mainly to save their resources for their strongest cases and limit the embarrassment of *de jure* acquittals.[9] From the defendants' perspective, these dispositions clearly represent meaningful forms of non-conviction, since one of the key issues in the debate over plea bargaining is whether or not defendants who accepted plea bargains would have been convicted or not. Certainly, if defendants knew ahead of time that prosecutors were going to drop or cancel the charges against them, they would have rejected any plea bargain. In any event, it is simply not feasible methodologically to distinguish between different types of non-convictions or different types of plea bargains, since this would require the estimation of multinomial selection models. Grouping these different forms of non-convictions together is a simplification, but the key to deciding whether this simplification is problematic is whether it affects the estimation of ρ. If defendants are primarily concerned with not being convicted, and regard the reasons for not being convicted as of secondary importance, the estimate of ρ should not be sensitive. A similar logic applies to aggregating different types of plea bargains, such as charge bargains and sentence reductions: if defendants are chiefly concerned with reducing their sentences and both types of bargains lead to reduced sentences, then ρ should not be sensitive to the type of bargain.

[9] Our understanding of these issues was confirmed by several lawyers consulted for this project who have extensive experience with criminal cases in Israel.

Censoring

There is a natural censoring problem in the data because trial verdicts take place after plea bargains have been reached. For example, some defendants died before trial. In their case, there are data for Y_1 but not for Y_2. Y_1 is observed for plea bargains reached before the cut-off date of May 5, 2011, but Y_2 is not observed for cases disposed of after this date. These include 43 cases that were stayed or suspended, and 145 defendants who could not be located for trial. In 49 cases, defendants were deemed incompetent to stand trial,[10] and in 39 cases dispositions were coded as "other" or "not available."

If only uncensored observations are used, information would be discarded on plea bargains simply because conviction outcomes were still unavailable at the cut-off on May 5, 2011. To avoid this loss of information, we use all the data on Y_1 by including an indicator for censored observations in the Y_2 model. The total number of censored observations is 276. These observations do not directly affect the estimation of ρ because by definition the model for Y_2 explains these observations perfectly. However, they do so indirectly because they contribute to the parameter estimates of the selection model (Y_1). We are implicitly assuming that the selection process governing censoring is independent of the unobservable phenomena such as g, d, c, etc, because censor status has nothing to do with judgment error, the toughness of prosecutors, litigation costs and guilt signals. Consequently, had these cases been decided rather than censored, they would not have affected the estimate of ρ.

Independent Variables

In addition to case disposition, the data also contain information on the criminal charges, maximum punishments, identities of judges and courts, dates of indictment, pre-trial detention status, legal representation and limited information on defendant characteristics. Table 2 presents the distribution of these variables overall and broken down by plea bargain and conviction status.

The plea bargain rate was largest for first offenders and smallest for defendants whose criminal history is unknown. Plea bargain rates also vary by citizenship status. It was largest for foreigners and smallest for Palestinians from the West Bank and East Jerusalem. Less than 8% of defendants were women, and women were less likely to plead guilty than men. The majority of defendants (66%) were not detained prior to trial, with only 22% held until sentencing. The plea-bargaining rate was lowest for the former and greatest for defendants who had other restrictions imposed. Most defendants had legal representation, but defendants with no legal representation were considerably less likely to plead guilty. Note that the Public Defender's office contracts much of their work out to private lawyers. The data therefore do not allow us to distinguish between private lawyers working under contract and those hired directly by defendants, so we cannot differentiate between public and private representation. The plea-bargain rate was similar for most offense types except for white-collar crimes, which was much lower. This is not surprising given that defendants accused of white-collar crimes are more likely to have the resources to take their cases to trial. We wish to stress that Table 2 describes the data; it is not intended to suggest causal

[10] We did not treat "incompetent to stand trial" as a form of *de facto* acquittal since the legal implications were not as clear cut as they were for the other dispositions treated as acquittals. However, we tested models that treated this disposition as a form of acquittal, and the estimate of ρ was nearly identical.

Table 2 Sample characteristics

	Proportion from total	Proportion w/ plea bargain	Proportion w/o plea bargain	N	Proportion acquitted	Proportion convicted	N
Criminal history							
Yes	0.44	0.82	0.18	820	0.22	0.78	149
No	0.26	0.76	0.24	487	0.28	0.72	115
Unknown	0.29	0.17	0.83	540	0.94	0.61	443
Missing	0.00	0.14	0.86	7	0.67	0.33	3
Citizenship							
Israel	0.96	0.61	0.39	1782	0.69	0.31	685
Foreign	0.01	0.69	0.31	26	0.50	0.50	8
West Bank	0.01	0.94	0.06	17	0.00	1.00	1
East Jerusalem	0.01	0.68	0.32	19	0.00	1.00	6
Missing	0.01	0.00	1.00	10	0.70	0.30	10
Gender							
Female	0.08	0.51	0.49	140	0.72	0.28	68
Male	0.91	0.63	0.37	1693	0.68	0.32	626
Missing	0.01	0.24	0.76	21	0.44	0.56	16
Pre-trial detention							
None	0.66	0.53	0.47	1223	0.73	0.27	566
Other restrictions	0.13	0.80	0.20	242	0.49	0.51	49
Arrested, but not up until sentencing	0.04	0.73	0.27	66	0.56	0.44	18
Arrested until sentencing	0.16	0.82	0.18	294	0.21	0.79	52
Hospitalization in psychiatric facility	0.01	0.91	0.91	22	0.95	0.05	20
Missing	0.00	0.29	0.71	7	0.80	0.20	5
Legal representation							
Public defender	0.31	0.71	0.29	567	0.53	0.47	165
Lawyer (private or public)	0.50	0.72	0.28	926	0.55	0.45	257
No legal representation	0.19	0.19	0.81	360	0.88	0.12	287
Moste severe criminal charge							
Drugs	0.13	0.66	0.34	243	0.73	0.27	82
Property	0.17	0.60	0.40	316	0.66	0.34	126
Public order	0.13	0.63	0.37	234	0.70	0.30	86
White collar	0.07	0.38	0.62	138	0.54	0.46	84
Violent	0.42	0.63	0.37	779	0.74	0.26	287
Other	0.07	0.68	0.32	139	0.48	0.52	44
Missing	0.00	0.20	0.80	5	1.00	0.00	1.00
Court							
District: Nazareth	0.03	0.83	0.17	48	0.13	0.88	8
District: Haifa	0.03	0.77	0.23	56	0.15	0.85	13
District: Jerusalem	0.02	0.65	0.35	34	0.25	0.75	12
District: Central	0.02	0.86	0.14	35	0.40	0.60	5
District: Tel Aviv	0.03	0.87	0.13	63	0.13	0.88	8

Table 2 (continued)

	Proportion from total	Proportion w/ plea bargain	Proportion w/o plea bargain	N	Proportion acquitted	Proportion convicted	N
District: Beer Sheva	0.04	0.84	0.16	83	0.62	0.38	13
Magistrate: North	0.07	0.66	0.34	127	0.55	0.45	42
Magistrate: Haifa	0.10	0.51	0.49	186	0.56	0.44	91
Magistrate: Jerusalem	0.09	0.58	0.42	159	0.47	0.53	62
Magistrate: Central	0.22	0.61	0.39	416	0.83	0.17	164
Magistrate: Tel Aviv	0.16	0.40	0.60	293	0.78	0.22	177
Magistrate: South	0.19	0.67	0.33	354	0.77	0.23	115

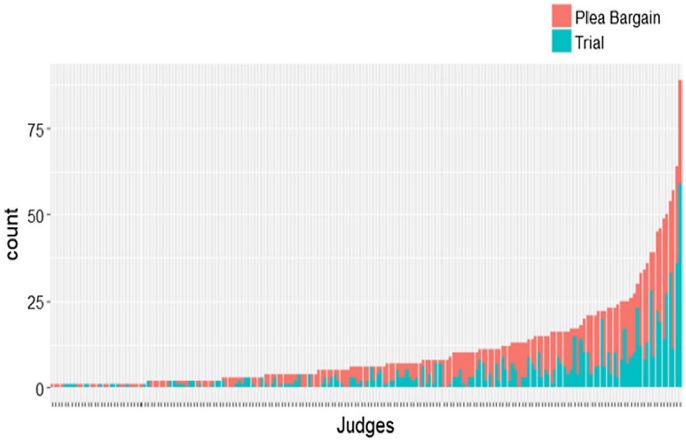

Fig. 2 Repeat observations by judges

relationships. This is especially relevant to arrest type and legal representation, which are endogenous variables. We did not include all of the variables listed in Table 2 in our model, for reasons discussed below.

Plea-bargain rates also vary by court and by judge. It was highest in the district court of Tel Aviv and lowest in the magistrate court in Tel Aviv. There is a rich literature detailing the institutional and organizational features of courts and courtroom workgroups that can be expected to affect the likelihood of plea bargaining and convictions (see Johnson et al. 2016 for review). What this implies for our framework is that the process governing selection into trial, and hence the value of ρ, can be expected to vary across judges, across courtroom work groups, and across jurisdictions. For example, if there is a norm in a particular court to offer minimal or no plea discounts for defendants with very strong guilt signals, then factually guilty defendants can be expected to select into trial more frequently. Or, if courts are particularly disorganized, this might lead to a higher-than-normal rates of error. In principle, if organizational features could be measured, then the value of ρ could be calculated separately for each organizational unit and compared to determine which features are correlated with a larger miscarriage of justice. Lacking such measures, we simply use fixed effects to control for variation between judges and courts.

More than a hundred judges were involved in the cases in our data. Most judges were involved in more than one case, and many judges were involved in more than 10 cases (Fig. 2). One judge was involved in more than 80 cases. We use these data to estimate judge fixed effects, as described in the "Findings" section below. In plea bargains the identity of judges is less important than in trials, because judges almost always accept plea bargains, whereas the role of judges is more salient in cases that go to trial, especially in Israel where there is no jury system. Nevertheless, it is noteworthy that for some judges, all of the cases assigned to them eventually plea bargained.

Table 2 does not include all the data used to specify the covariates for Y_1 and Y_2. For example, the data record whether the offense involved a victim. For some offenses this is irrelevant, but for others, such as grievous bodily harm, it means there is a victim who may give evidence. Unfortunately, the data do not record the number of witnesses, but they do record the number of hearings, which, as noted above, may be correlated with the number of witnesses. Another variable is court pressure, which serves as an instrumental variable in the Y_1 model, as discussed in the next section.

Methodology

Estimation Procedure

Since bivariate probit models are complex and nonlinear, we set initial parameter values by estimating β and γ assuming $\rho = 0$, i.e. by ignoring selectivity. We use the general-to-specific (GTS) procedure (Hendry 1995) for specifying the variables in these models. GTS combines backward stepwise procedures with checks for path dependence in which variables eliminated at an earlier stage are subsequently respecified. These initial model specifications are then re-estimated by maximum likelihood together with ρ. Finally, GTS is used once again to obtain final estimates of β, γ and ρ. This was implemented using the heckprob command in Stata. In "Appendix 2" we discuss parametric alternatives to the normal and a non-nested test to distinguish between these alternatives.

Model Specification

The estimation procedure can fail to converge if there are too many parameters in the model. We have therefore prioritized parsimonious specifications in which variables with z-statistics of less than one were omitted. Our model is also limited to the information available in the dataset. The conviction model (Y_2) includes two offense types (drugs and white collar) for the most serious charge, the number of hearings (which is likely correlated with the number of witnesses), whether there is a victim (who can in principle be called as a witness), and court and judge fixed effects. The plea bargain model (Y_1) includes citizenship, criminal history, offense types, maximum sentence, an indicator of whether the person was held in a psychiatric facility prior to sentencing, whether there is a victim, an indicator of court pressure (described in the next section), and court and judge fixed effects. We expect that these variables may play a role in determining plea-bargain decisions, and that they are exogenous with respect to defendants' true guilt. For example, according to STT, defendants take into account the length of the maximum sentence that they may be expected to serve if found guilty.

Criminal history should not matter for conviction, and is therefore omitted from the model for Y_2. However, prosecutors are likely to take it into consideration in deciding what kind of plea discount—if any—to offer, and so it may affect plea-bargaining outcomes. The existence of a victim is an indicator of evidence against the defendant since victims can act as witnesses.

Since judges are assigned to cases prior to plea bargaining, defendants and prosecutors know the identity of judges at the plea bargain stage. Since this knowledge might influence plea bargains, we included judge fixed effects in both the conviction (Y_2) and plea bargaining (Y_1) models. For example, a defendant assigned to a "hanging judge" might be more inclined to accept a plea offer than one assigned to a judge who is known for being lenient. Since the number of judges is large, it was not feasible to estimate individual judge fixed effects. We grouped them as follows. We formed a reference category that includes all judges with less than 10 cases. The remaining 65 judges were formed into 7 groups. The compositions of these groups were selected on the basis of goodness-of-fit in probit models for Y_2. Finally, we estimated the bivariate probit model with 7 judge group fixed effects in the models for Y_2, but only 6 fixed effects in Y_1 because the 7th group did not satisfy the inclusion criterion of having a z-statistic of greater than one.

For pre-trial detention status, we only include an indicator for defendants held in psychiatric facilities. Although pre-trial detention status is thought to have a major influence on defendants' readiness to accept plea bargains (Wooldredge et al. 2015), we do not specify other forms of detention status in the Y_1 model. As explained in "Appendix 3", if defendants are detained for reasons related to their conviction outcomes, the causal effect of detention on plea bargaining is not identified. We therefore leave this matter for subsequent research, which involves finding instrumental variables for detention status and trivariate selection.

Similar considerations apply to the causal effect of legal representation on the probability of conviction as well as the probability of accepting plea bargains. We have already noted that ρ decreases if defendants with stronger signals of guilt invest more in legal representation. However, we refrain from specifying legal representation in the models for Y_1 and Y_2 because, in addition to data limitations described above, just as the causal effect of detention is not identified, neither is the causal effect of legal representation (for more details see "Appendix 3"). Matters would be different if instrumental variables were available for identifying the causal effects of legal representation. Here, too, the selection problem becomes trivariate. We defer this issue to future research as well.

Identification

In view of parametric assumptions about the joint distribution of y_1 and y_2, identification should also be based on instrumental variables, which are included among the covariates hypothesized to affect selection into trial (Z) but excluded from the covariates hypothesized to affect conviction (X). There is no reason why Z and X should be the same since there are variables that affect plea bargains, which do not affect convictions, and there are variables that affect convictions, which do not affect plea bargains. For example, defendants cannot know in advance what will transpire in court. Prosecutors play a central role in plea bargains while judges determine convictions. Hence, in principle, information on prosecutors would be specified in Z but not X. We use caseload pressure on courts as an instrumental variable, represented by the total number of criminal cases opened, pending and closed to the total number of judges in a particular

court for the year in which indictments were filed. These annual data come from the statistical reports published by the Israeli court system (Israeli Courts Administration 2003–2011). We hypothesize that greater pressure induces prosecutors to reach plea bargains in order to relieve pressure on the courts, but there is no reason to suspect that busier courts are more likely to convict. We wish to stress, therefore, that the parameters are identified by instrumental variables and not just by assumptions of bivariate normality.

Findings

Our main priority is the estimation of ρ. However, its estimate is conditional on the specification of Y_1 and Y_2. Our central concern is not with the specification of Y_1 and Y_2; we do not seek to make strong claims about the covariates X and Z, which have causal effects on trial outcomes and plea bargains. Rather, the specification of Y_1 and Y_2 serve to provide estimates of ρ. Nevertheless, we check the robustness of ρ with respect to alternative specifications of X and Z, and alternative parametric assumptions. The observable determinants of plea bargains and trial outcomes are of secondary priority.

Table 3 Results from bivariate probit model of conviction and selection into trial

	Coef.	SE
Conviction model (Y2)		
Victim—yes	−.393***	.122
Number of hearings	.071***	.019
Crime category—drugs	−.140	.191
Crime category—white collar	.438*	.172
Constant	−1.213***	.140
Selection into trial/no plea bargain model (Y1)		
Citizenship—foreign	−.447	.272
Citizenship—West Bank	−1.385**	.527
Criminal history—none	.258***	.078
Criminal history—not known	1.995***	.090
Crime category—drugs	−.301*	.133
Crime category—property	.136	.097
Crime category—public order	−.110	.107
Crime category—white collar	.486***	.147
Maximum statutory sentence across counts	−.001	.001
Pre-trial detention—hospitalization in psych facility	1.813***	.402
Victim—yes	−.239**	.089
Court pressure	−.001*	.000
Constant	−.833***	.123
ρ (rho)	**.983*** **	**.016**

N = 1757 (1083 in Y2); log likelihood = −874.4. Court and judge fixed effects reported in Table 4

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Table 4 Judge and court fixed effects from bivariate probit models of conviction and selection into trial

	Coef.	SE
Conviction model (Y2)		
District court—Beer Sheva	−1.793*	.903
Magistrate court—Jerusalem	.548*	.230
Magistrate court—Tel Aviv	.836***	.179
Judge group 1	.422*	.187
Judge group 2	1.205**	.455
Judge group 3	−.648**	.231
Judge group 4	−.436	.234
Judge group 5	−.355	.345
Judge group 6	.468	.375
Judge group 7	−.814*	.322
Selection into trial/no plea bargain model (Y1)		
District court—Jerusalem	.251	.218
District court—Beer Sheva	−1.030**	.349
Magistrate court—Jerusalem	.249*	.127
Magistrate court—Tel Aviv	.625***	.116
Judge group 1	.373**	.140
Judge group 2	.406	.254
Judge group 3	−.306	.162
Judge group 4	−.214	.190
Judge group 5	1.208**	.466
Judge group 6	.366*	.150

N = 1757 (1083 in Y2); log likelihood = −874.4

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Main Results

Results using the model specification procedure described above for the bivariate normal specification for Y_1 and Y_2 are presented in Table 3. For convenience, fixed effects for court and judge groups are reported in Table 4. Because of missing data for certain covariates, this model is estimated with 1757 observations instead of 1854, which slightly increases the representation of plea bargains. The top panel of Tables 3 and 4 present the β parameters of the X covariates in the Y_2 conviction model from Eq. (1). These parameters are estimated using the 441 observations remaining after censoring and dropping cases with missing values. The bottom panel refers to the γ parameters and Z covariates in the plea bargain Eq. (3a), where the dependent variable is $Y_1 = 1$ if there was no plea bargain. These parameters are estimated using all 1757 observations, with an indicator variable (not shown) for the observations that were censored in y_2 as described above in the "Data" section.

The most important parameter in Table 3 is the estimate of ρ, which is positive and close to 1. We wish to stress that this estimate is not a technical consequence of the fact that ρ has an upper limit of 1. If it had been, the iteration procedure would have aborted. Instead, the estimate of ρ converges from below on 0.98. Refining the sensitivity of tolerance, iteration criteria and starting values make no difference to the result, including starting values

Table 5 Goodness-of-fit statistics for models in Tables 3 and 4

Model	Pseudo R^2	Kay-Little R^2	Kendall tau
Selection	0.517	0.726	0.492
Conviction	0.091	0.606	0.459

of $\rho = 0.99$. Moreover, as shown below, the estimate of ρ is robust to alternative parametric assumptions to bivariate normality. We therefore conclude that the estimate of ρ is genuine. As ρ approaches 1, small differences in ρ may be statistically significant. A Fisher test for the difference between the estimate of ρ in Table 3 and the null of $\rho = 0.99$ has a z-value of 8.44, which overwhelmingly rejects the null. This rejection applies a fortiori when the null is $\rho = 1$.

A brief discussion of the results from Tables 3 and 4 is in order. We begin with the bottom (Y_1) panel in Table 3 where a positive sign means that the variables concerned increase the probability of going to trial and reduce the probability of plea bargains. Defendants from the West Bank were less likely to go to trial compared to the reference category, which was comprised of Israeli citizens and East Jerusalem residents. Defendants with no criminal history, or whose criminal history was unknown, were more likely to reject plea bargains compared to defendants with criminal histories. Crime category does not matter for plea bargaining except for white-collar crimes where plea bargains were more likely to be rejected, and drug-related crimes where going to trial was less likely. There is some indication that higher maximum sentences induce more plea bargaining, but this variable is not statistically significant at conventional levels. When there is a victim, defendants are more likely to plead guilty. Plea bargains are less likely in the large magistrate courts and in the district courts in Jerusalem and Haifa. As expected, court pressure increases the incidence of plea bargaining.

The Y_2 panel of Table 3 shows that conviction probabilities vary directly with the number of court hearings, but are smaller if there is a victim. It also shows that conviction rates are higher for white collar crimes. Table 4 shows that conviction is most likely at the magistrate court in Tel Aviv and least likely at the district court in Beer Sheva. Plea bargaining is most likely at the district court in Beer Sheva and least likely at the magistrate court in Tel Aviv. Judges in group 2 were most likely to convict, whereas judges in group 7 were least likely to convict.

Goodness-of-fit statistics for the plea bargain and conviction models in Tables 3 and 4 are reported in Table 5. These statistics are bounded between 0 and 1. It is well known that (unlike logit models) probit models do not replicate aggregate selection and conviction probabilities in the data. Instead, they maximize the likelihood of observing the data given the model. Therefore, in contrast to linear regression, R^2 is not a criterion for model specification. Nor does it embody diagnostic information. The various measures reported in Table 5 reflect the absence of a consensus on how to measure goodness-of-fit for nonlinear estimators. Nevertheless, they compare favorably with the R^2 reported by Bushway and Redlich (2012) for their linear estimator of conviction rates.

Robustness Checks

On the whole, omitting variables with relatively large *p*-values such as "district court Beer Sheva" from panel A in Table 3 or "crime category property" from panel B make

Table 6 Robustness tests for alternative parametric assumptions

Marginal model	Copula	ρ	log likelihood
Probit	Normal	0.981	−874.4
Logit	Normal	0.980	−873.4
Logit	t (df = 3)	0.995	−874.2
Probit	Frank	0.978	−875.1
Logit	Frank	0.978	−875.4
Bivariate probit (Table 3)	n/a	0.983	−874.4

no difference to the parameter estimates, including ρ. The same applies to when we omit observations for which previous criminal history is unknown.[11] Indeed, the insensitivity of ρ to these omissions reinforces our judgment that the estimate of ρ in Table 3 is not merely technical. The same applies to adding variables such as gender, time to indictment and other defendant characteristics to panel A.

In Table 6 we report robustness tests with respect to parametric alternatives to bivariate normality, described in more detail in "Appendix 2". We use various copulas (Smith 2003), estimated with R, to represent the cumulative distribution between y_1 and y_2, which do not restrict the estimate of ρ. With the exception of the t copula, which uses 3 degrees of freedom, the other Archimedean copulas use only one degree of freedom. Relative to the alternatives, the Frank copula assumes that large negative values of y_1 and y_2 are weakly related, while values near the mode are strongly related. Also, tail dependence is more symmetric. We focus on the estimate of ρ because it is likely to be more sensitive to parametric alternatives. The bivariate probit result is the baseline case from Table 3 reported at the bottom of Table 6 for convenience. Since Pearson correlations are undefined, the estimates of ρ in Table 6 refer to their Spearman counterparts.

The estimates of ρ in Table 6 are strikingly similar. Also, they have almost identical likelihoods. These tests establish that the results are robust with respect to parametric alternatives, both in terms of the marginal distributions and the cumulative densities for y_1 and y_2.

We also estimated the conviction model using the method of inverse probability weighting (IPW) where the selection probabilities (trimmed for outliers) are determined by the plea bargain model in Table 3. Although the parameter estimates in Table 3 and their IPW counterparts (not shown) are broadly similar in terms of their estimates and p-values, the court and judge fixed effects (Table 4) are different. The likelihood of the IPW model is considerably smaller than the likelihood contribution of the conviction model in Table 3. IPW assumes that selection into plea bargains depends entirely upon observables, which are represented by Z in Eq. (3a), implying that y_{1i} is zero for all defendants. This would mean that unobservables such as guilt signals (g), the toughness of prosecutors (d), etc. do not matter for plea-bargaining.

Proponents of IPW use sensitivity analysis to check whether IPW results are sensitive to hidden confounders as follows: First, they check whether defendants who share common Z covariates have different selection probabilities, which they should not under the null. Rosenbaum (2002, p. 107) refers to such hidden confounders by u. Second, they assume that

[11] Defendants' criminal histories marked as "unknown" in the data are essentially missing from the data even though the prosecutors and defense attorneys involved in the case do know their histories.

u might also influence the outcome (conviction) probabilities positively or negatively to various degrees, denoted by Γ. Γ = 4 means that one defendant matched for Z is 4 times more likely to plea-bargain than the other because of the hidden confounder. Third, upper and lower bounds are calculated for conviction probabilities using Γ (Rosenbaum 2002, p. 113).

Suppose, for example, the IPW conviction probability conditional on Z is 0.4, and when Γ = 4 this probability may be as low as 0.39 and as high as 0.42 so that these bounds are not significantly different from 0.4. It might reasonably be concluded that the IPW estimate of 0.4 is insensitive to hidden confounders. Matters would be different if either or both of the bounds are significantly different from 0.4 when Γ is 1.3, i.e. when one matched defendant is slightly more likely than the other to plea bargain.

We reconcile IPW with the selection methodology in Table 3 in the following manner. The hidden confounder (u) is represented by y_1. Unlike IPW, we hypothesize the existence of y_2 which is correlated with y_1 through ρ. For a given Γ, the sensitivity analysis implicitly assumes that y_2 is perfectly correlated with y_1. Whereas sensitivity analysis experiments with different values of Γ to determine indirectly whether hidden confounders matter, the incidentally truncated methodology in Table 3 determines this issue directly by estimating ρ. If $\rho = 0$, this is equivalent to insensitivity for IPW when Γ is large. Whereas hidden confounders are hypothesized *ab initio* in Table 3, the opposite applies with IPW. Sensitivity analysis is inherently a judgment call since critical values for Γ are unavailable. By contrast, hypotheses concerning ρ may be tested.

We use a marginal sensitivity procedure in R proposed by Zhao, Small and Bhattacharya (2019) to conduct sensitivity analysis on our IPW results. When Γ = equals 1 (no sensitivity to hidden confounders) the probability of conviction is 0.5845. The conviction probability bounds are sensitive to Γ. For example, when Γ = 1.25 the lower bound is 0.5196 and the upper bound is 0.6458. The lower bound becomes statistically significant ($p = 0.05$) when Γ is only 1.16.

The sensitivity of IPW to hidden confounders come as no surprise in view of the fact that the estimate of ρ in Table 3 is large and positive. Proponents of IPW criticize the bivariate probit methodology for its parametric assumptions. Are the assumptions of IPW weaker or stronger? We resolve this dilemma by undertaking robustness checks as reported in Table 6, indicating that the results of ITM in this case are not sensitive to a wide range of parametric assumptions.

Calculating Counterfactual Conviction Probabilities

We use the results in Table 3 to parameterize Eqs. (5b) and (5c) and calculate counterfactual conviction probabilities for defendants with similar observed characteristics. That is, for defendants who pled guilty, we estimate their counterfactual expected probability of conviction had they gone to trial, and compare it to the probability of conviction for similar defendants who were tried. Table 7 provides examples of observed and counterfactual conviction probabilities for 9 defendant profiles. The baseline profile (Example 1) refers to defendants charged with a drug-related crime, who are Israeli citizens with no criminal history, had one hearing, are in the reference 'other' category for pre-trial detention and court group, had a judge in group 2 and a mean court pressure value of 71.7. These hypothetical defendants are charged with crimes that did not come with a potential prison sentences, so the maximum sentence in months is zero.

The bottom section of Table 7 under "Conviction Estimates" compares the conviction probabilities for defendants who went to trial with the counterfactual conviction

Table 7 Counterfactual probability of conviction for different defendant profiles

Profile	1	2	3	4	5	6	7	8	9
Crime category	Drugs	Drugs	Drugs	Drugs	**White collar**	Drugs	Drugs	Drugs	Drugs
Citizenship	Israel	Israel	Israel	Israel	Israel	**Foreign**	Israel	Israel	Israel
Number of hearings	1	**10**	**10**	1	1	1	**8**	1	1
Court group	Other	Other	**Tel Aviv—Magistrate**	**Tel Aviv—Magistrate**	Other	Other	Other	Other	Other
Judge group	2	2	2	2	2	2	2	2	2
Criminal history	None	None	None	None	None	None	None	None	None
Max sentence (months)	0	0	0	0	0	0	0	**50**	0
Pre-trial detention	Other	Other	Other	Other	Other	Other	Other	Other	Other
Victim	None	None	None	None	None	None	None	None	None
Court pressure	71.7	71.7	71.7	71.7	71.7	71.7	71.7	71.7	**160**
Conviction estimates									
Conviction likelihood for defendants w/o plea bargain: $E(Y_2^*\|Y_1=1)$									
z-score	1.08	1.71	2.13	1.49	1.14	1.41	1.57	1.11	1.14
Probability	0.86	0.96	0.98	0.93	0.87	0.92	0.94	0.86	0.87
Counterfactual conviction likelihood for defendants w/plea bargain: $E(Y_2^*\|Y_1=0)$									
z-score	−0.56	0.08	0.56	−0.08	−0.44	−0.36	−0.06	−0.53	−0.51
Probability	0.29	0.53	0.71	0.47	0.33	0.36	0.48	0.30	0.31
Difference between z-scores	1.63	1.63	1.57	1.57	1.58	1.78	1.63	1.64	1.65

Differences from the baseline model (1) are in bold

probabilities for defendants with the same profile who plea bargained. The profiles shown were chosen for illustration only. The first statistic reports the expected value of Y_2^* for defendants who did not plea bargain. Recall from Eq. (1) that Y_2^* denotes a latent variable where $Y_2 = 1$ if $Y_2^* > 0$. The distribution of Y_2^* is standard normal, and so the conviction likelihood can be calculated as a z-score on the standard normal distribution, which allows for easy comparability across profiles. In the next row, we report the conviction probabilities generated by z-scores from the standard normal distribution. The next two rows report the estimated z-scores and probabilities for the counterfactuals where defendants with similar profiles plea bargained instead of going to court. The bottom row shows the difference between the estimated z-scores for defendants who went to trial and those who plea bargained.

Table 7 shows that across all profiles, the counterfactual probabilities of conviction for plea bargainers are much smaller than the probabilities of conviction for observationally similar defendants who went to trial. For example, defendants who fit our baseline profile in column 1 and went to trial have an expected conviction probability of 86%, compared to a 29% counterfactual probability of conviction for observationally similar defendants who accepted plea bargains. Profile 3 shows a less pronounced difference, where the counterfactual conviction rate drops from 98% to 71%. The only difference between profiles 1 and 3 is that the latter involved 10 hearings and were tried in the Tel Aviv magistrate court, which, as shown in Table 4, had the second-highest coefficient for conviction likelihood. Most of this difference is due to the additional hearings (compare profile 3 and 4). Since more hearings likely indicate the presence of more witnesses, this suggests that more or better evidence leads to outcomes that are more just—or at least to fewer people pleading guilty when they would not have been convicted. The counterfactuals reported in Table 7 are unbiased estimates, or expected values. They obviously have confidence intervals, which we have not calculated for technical reasons. Nevertheless, the differences between the counterfactuals and their comparators are almost surely statistically significant.

It is important to note that that even if defendants would not have been convicted had they gone to trial, this does not imply they must have been factually innocent. They could be found legally innocent due to, say, a reasonable doubt even though they are factually guilty. We cannot distinguish between legal and factual innocence, nor can we determine what proportion of defendants belong to each of those categories. We simply assume that people who are not convicted are more likely to be factually innocent than those who are convicted. Furthermore, the information in Table 7 does not tell us to what extent the differences between observed and counterfactual conviction probabilities is due to a miscarriage of justice. We now turn to address this issue.

Decomposing ρ: Does the Miscarriage of Justice Account for the Results?

Having established that the value of ρ is large and positive resulting in large differences between observed and counterfactual conviction probabilities, we address the question: to what extent does miscarriage of justice theory (MJT) account for ρ compared to shadow trial theory (STT)? Recall that STT predicts a positive value for ρ since defendants who appear guilty (and so are more likely on average to be factually guilty) will be offered less attractive plea offers, especially by tough prosecutors, and so they should be more likely to reject plea bargains. In fact, if their guilt signals are especially strong or prosecutors are really tough, they may not be offered plea bargains at all. How can we be sure, then, that STT does not entirely account for the large positive value of ρ?

To address this, we return to Eq. (4), which is reproduced here for convenience:

$$\rho = -\sigma_{gg'} - \sigma_{ee'} - \sigma_{ge'} - \sigma_{g'e} + \sigma_{gd} - \sigma_{gc} + \sigma_{de} - \sigma_{ce} \tag{4}$$

We use Eq. (4) to interpret the estimate of ρ, which is composed of 8 covariance terms. If the last 6 terms are zero, ρ should be negative, assuming, as seems reasonable, that g and g' are positively correlated, as are e and e'. This result would have been consistent with innocence effect theory (Gazal-Ayal and Tor 2012), which contends that on average defendants who do not have plea bargains are negatively selected in terms of conviction. According to Eq. (4), ρ would be positive if the combined effect of the remaining 6 covariance terms were sufficiently positive. Unfortunately, it is impossible to identify these covariance terms individually. It is difficult to conceive of an a priori reason why σ_{ce} and σ_{de} should be other than zero since litigation costs (c) and prosecutors' toughness (d) should be independent of judgment error (e). If defendants with stronger signals of guilt incur higher litigation costs ($\sigma_{gc} > 0$), as seems reasonable, this would imply that ρ should be yet more negative.

Therefore, for ρ to be positive, either σ_{gd} must be sufficiently positive and/or $\sigma_{g'e}$ and $\sigma_{ge'}$ must be sufficiently negative. The former would be consistent with shadow trial theory (STT), which predicts that prosecutors are tougher on defendants with stronger guilt signals. The latter would be consistent with miscarriage of justice theory (MJT), because it means that defendants with weaker guilt signals worry that they are more likely to be convicted, and/or defendants with stronger guilt signals hope that they are more likely to be found innocent. In summary, $\sigma_{g'e}$ and $\sigma_{ge'}$ are negative if innocent defendants, on average, expect to be falsely convicted and/or guilty defendants expect judgment error to be in their favor and therefore to be falsely acquitted.

Although the individual components of ρ in Eq. (4) cannot be identified, we can nevertheless bound the contribution of STT to ρ under the assumption that MJT is false. If these bounds turn out to be inconsistent with the estimate of ρ, then by implication MJT is not false. In other words, if STT cannot account for the large, positive value of ρ, then MJT must account for some portion. This procedure prioritizes STT over MJT because we see MJT as the rival hypothesis.

In Table 3 the variances of y_1 and y_2 equal 1 due to bivariate normality. Hence from Eqs. (2a) and (3b):

$$\sigma_{y2}^2 = \sigma_g^2 + \sigma_e^2 + 2\sigma_{ge} = 1 \tag{6a}$$

$$\sigma_{y1}^2 = \sigma_{g'}^2 + \sigma_{e'}^2 + \sigma_d^2 + \sigma_c^2 + 2(\sigma_{g'e'} + \sigma_{g'c} - \sigma_{g'd} - \sigma_{dc} + \sigma_{e'c} - \sigma_{e'd}) = 1 \tag{6b}$$

If MJT is false, judgment error is independent of guilt signals. Hence, $\sigma_{ge} = \sigma_{g'e} = \sigma_{g'e'} = 0$. We continue to assume $\sigma_{dc} = \sigma_{de} = \sigma_{ce'} = 0$. Under these restrictions, Eq. (6a) simplifies to:

$$\sigma_g^2 + \sigma_e^2 = 1 \tag{6c}$$

in which case σ_g^2 and σ_e^2 are less than 1, and their average is 0.5. We assume that g, and g' have the same variances, as do e and e', and that $\sigma_{gd} = \sigma_{g'd}$ and $\sigma_{gc} = \sigma_{g'c}$.[12] Substituting these assumptions into Eq. (4) implies:

[12] Formally, this implies e.g. $r_{gd} / r_{g'd} = \sigma_{g'd} / \sigma_{gd}$.

$$1+\rho=\frac{1}{2}\left(\sigma_d^2+\sigma_c^2\right) \tag{6d}$$

Since the estimate of ρ is 0.98, Eq. (6d) implies that the variances of c and d would have to be large relative to the variances of g and e for our results to be consistent with STT alone. If these variances were almost twice as large as the variances of g and e (0.99), Eq. (6d) would hold exactly. If, as seems reasonable, all variances are roughly similar (0.5), the right hand side of Eq. (6d) would be 1, which is less than $1+\rho$ (1.98). Therefore, STT alone cannot account for why ρ is positive, and by implication the role of MJT is salient and large. If the variances are equal, therefore, the contribution of STT to $1+\rho$ would be 1 and the contribution of MJT would be 0.98. However, for reasons given in the section on "Quantitative Versus Qualitative Results," we are unable to quantify the proportion of plea bargainers who are factually innocent.

Having established a role for MJT, we decompose it into innocent defendants who fear that they would be convicted if they plead innocent, and guilty defendants who plead innocent in the hope that they might be acquitted. We denote the contribution of MJT to ρ by ρ_{mj}, which is positive. In Eq. (4) $\rho_{mj}=-(\sigma_{g'e}+\sigma_{ge'})$, which implies that the covariance terms in brackets are negative, i.e. defendants with stronger guilt signals are less likely to be convicted in error. According to Eq. (2b) ρ_{mj} may be decomposed into the contribution of innocent defendants who pled guilty in plea bargains (ρ_{mji}), as defined in Eq. (7c), and the contribution of guilty defendants who reject plea bargains in the hope that they will not be convicted (ρ_{mjg}), as defined in Eq. (7b):

$$\rho_{mj}=\rho_{mji}+\rho_{mjg} \tag{7a}$$

$$\rho_{mjg}=(1-\pi)(\sigma_{g'\lambda}+\sigma_{g\lambda'}) \tag{7b}$$

$$\rho_{mji}=-\pi(\sigma_{g'\theta}+\sigma_{g\theta'}) \tag{7c}$$

Recall that π denotes the unknown proportion of innocents among defendants. Since the individual covariance terms are not identified, the decomposition of ρ_{mj} is not feasible. The guilty may reject plea bargains in the hope of being acquitted. If this tendency is more pronounced among the guilty with weaker guilt signals (who appear less guilty), the covariance between g and λ would be negative in Eq. (7b), in which case ρ_{mjg} would be negative, and ρ_{mji} even more positive to account for the large, positive value of ρ_{mj}. In this case, the miscarriage of justice due to innocents pleading guilty would be even larger. If the covariance between g and λ is positive, the opposite would be true. However, this would imply, unreasonably in our view, that the factually guilty who look guiltier expect more judgement error and prefer to take their chances in court.

If the covariance for the factually guilty is negative, as seems reasonable, Eq. (7c) would imply that the covariance between g and θ is negative, since ρ is positive. This would mean that the factually innocent with weaker guilt signals (who appear more innocent) prefer plea bargains compared to the factually innocent with stronger guilt signals. Is this reasonable? It would be, if the lack of confidence in the courts among the factually innocent is greater for those who look more innocent. Getting arrested and accused of a crime even though you are not only innocent but also look innocent (have weaker guilt signals), could undermine defendants' trust in the judicial process. In our view, this interpretation is more plausible than the alternative that factually guilty defendants with stronger guilt signals take their chances in court.

In summary, a conservative interpretation of ρ implies, on average, that innocent defendants plead guilty and guilty defendants plead innocent. However, the main component in the miscarriage of justice theory comprises the innocents who plead guilty.

Alternative Explanations

Other unobservable phenomena may have been omitted from our analysis. For example, conditional on their expectations of conviction in court, more risk averse defendants are inclined to accept plea bargains; they prefer the certainty of plea bargains to taking their chances in court. Second, guilty defendants may feel remorse; they wish to admit their guilt and are inclined to accept plea bargains. Third, in addition to the toughness of prosecutors, defendants may take into consideration the severity of their judges; defendants who are allocated to severer judges are more inclined to accept plea bargains. These omissions all affect ρ negatively. For example, adding unobserved judge severity in Eq. (1) and subtracting it from Eq. (2b), or subtracting unobserved remorse from Eq. (2b) decreases ρ. Therefore, these omissions serve to strengthen the case in favor of MJT because they would increase the inability of STT alone to explain why ρ is positive. Subtracting unobserved risk aversion from Eq. (2b) would increase ρ if defendants with stronger guilt signals are less risk averse. In this respect, risk aversion is similar to defendants' estimation of judgment error and works in a similar way. Although risk aversion and estimates of error are conceptually distinct, in practical and empirical terms it is difficult if not impossible to disentangle them. Therefore, the omission of these other unobservable phenomena either strengthens our conclusions, or it makes no difference.

Another source of unmeasured variation comes from defendants' minority status, which is missing due to data limitations. Given evidence of judicial bias against minorities in Israel (Fishman and Rattner 1997; Fishman et al. 2006; Gazal-Ayal and Sulitzeanu-Kenan 2010), minority status would tend to increase defendants' conviction likelihood. To the extent that minority defendants are aware of this bias, they should be more willing to accept a plea deal. Indeed, evidence from the U.S. supports the notion that minorities are more likely to enter false guilty pleas (Redlich et al. 2010). In this regard, minority status would operate similarly to risk aversion and would thus tend to make ρ negative.

Conclusion

There are two important empirical results in this paper concerning the miscarriage of justice and plea bargaining. The first confirms fears voiced by critics that plea bargaining induces innocent defendants to plead guilty (Alschuler 2015; Dervan and Edkins 2013; Rakoff et al. 2014). The second, which follows from the first, is that shadow trial probabilities of conviction are smaller for plea bargainers than for observationally similar defendants who were tried in court. These results should be of major interest to policymakers and jurists, as well as to criminologists concerned with estimating probabilities of conviction for shadow trials (Abrams 2011; Bushway and Redlich 2012; Bushway et al. 2014; Redlich et al. 2016; Ulmer and Bradley 2006). The demonstrated importance of selection bias in the estimation of probabilistic models of conviction and the need to take selection bias into account has important implications for sentencing research more broadly, and especially in view of the "life course" approach to the study of sentencing and criminal cases (see e.g., Johnson 2015).

We have demonstrated a new approach to the study of plea bargains that enables researchers to shed empirical light on whether plea bargains result in a miscarriage of justice. It is consistent with existing research on wrongful convictions (Huff et al. 1996; Loeffler et al. 2018) and false guilty pleas (Gudjonnson and Sigurdsson 2008; Redlich et al. 2010; Zottoli et al. 2016) that rely on offender self-reports or expert evaluations. Defenders of plea bargaining argue that even if innocent people plead guilty, this does no injustice since they probably would have been convicted had they gone to trial. However, because defendants who plea bargain do not go to trial, it has been impossible to infer what would have happened to plea bargainers if they went to court. Drawing upon an established bivariate probit methodology seldom used in criminology (though see Uggen 1999), we estimated the correlation (ρ) between the residuals from probabilistic models of selection into trial and conviction in court. The estimate of ρ is positive and large, implying that defendants who did not have plea bargains are positively selected in terms of conviction. This result is inconsistent with innocence effect theory, which predicts that factually innocent defendants are more likely to reject plea bargains on average (Gazal-Ayal and Tor 2012), in which case ρ should be negative.

Our conclusion is based on a novel decomposition of ρ into three components: the first component is induced by the toughness of prosecutors; the second is induced by factually guilty defendants who plead innocent; and the third is induced by factually innocent defendants who nevertheless accept plea bargains. We showed that the first two components are insufficiently large to account for the estimate of ρ.

Of course, our study is not without its limitations. To begin with, the results are based on a single jurisdiction, Israel, over the span of a single year. Second, the data include a relatively limited number of covariates. We would especially have wanted to include a measure for criminal history with fewer missing values and better evidentiary indicators beyond the potential existence of a witness (the victim) or the proxy indicator of number of court hearings. In addition, we would have ideally been able to include prosecutor fixed effects and judge-prosecutor interactions. Finally, the bivariate probit methodology may be sensitive to its parametric assumptions. And although our robustness checks with respect to a variety of parametric alternatives suggest our results are not sensitive, one can never be certain that results are not driven by parametric assumptions.

The incidentally truncated regression model we employ has seldom been used to examine selection bias in criminology, which is surprising since it holds great promise for studying many topics of concern to criminologists, such as the effects of different sentencing practices or the impact of rehabilitation or re-entry programs. We hope to see the method used more widely in the discipline. It is well-suited to the study of plea bargaining, because it exploits phenomena that are either unobservable or difficult to measure in order to shed light on hypotheses concerning convictions and sentencing, apart from plea bargaining.

Furthermore, the limitations of our study point to promising avenues for future research: for example, conducting similar research in different jurisdictions. Future research may examine what distinguishes jurisdictions where ρ is lower from those where it is higher. For example, is ρ lower in jurisdictions where limits are placed on sentence discounts that prosecutors are allowed to offer defendants? If so, this could indicate a promising way to mitigate the miscarriage of justice in plea bargaining regimes (Gazal-Ayal 2006). More generally, ρ can be estimated for different courts and then correlated with the various organizational and institutional features of courtroom work groups that have historically been central to criminological and sociological accounts of plea bargaining. Such studies would identify the factors that exacerbate or mitigate the miscarriage of justice. Also, it is unclear how the estimation of ρ would be affected by improved controls for strength of evidence; this is something to examine in future research with data that include such

information. Another avenue for future research involves a more complex methodology to examine the causal effect of legal representation and/or pre-trial detention, discussed in "Appendix 3".

Another question is whether enabling defendants to review the evidence against them before they accept a plea bargain will make it harder for prosecutors to pressure innocent defendants into plea agreements, as suggested by Alschuler (1983). Our results indicate that such a policy change is unlikely to make a difference, because we found a large, positive value for ρ in a jurisdiction (Israel) where prosecutors must disclose all evidence that could be presented in court at the time of indictment. However, since Israel is different from the U.S. in many other ways, further research using U.S. data is called for.

Our results refer to a jurisdiction where the plea bargaining rates, although high, are still lower than in the U.S. We expect that as the incidence of plea bargaining increases, the miscarriage of justice will intensify, because high rates of plea bargaining are a sign that courtroom actors are exerting a great deal of pressure on defendants to settle out of court in order to reduce workloads. Indeed, we find that given everything else, the probability of plea bargains varies directly with the workload of courts. We also suspect a vicious cycle in which plea bargains are perceived to economize on the need for more judges, which in turn increases the workload on incumbent judges, which further increases the incidence of plea bargaining.

There was a brief period in the late 1970's and early 1980's when the dominance of plea bargaining in the United States fell under intense academic scrutiny and criticism. That burst of attention withered on the vine, and plea bargaining became even more widely used and deeply entrenched. Perhaps this was due in part to the absence of sound empirical research demonstrating convincingly that plea bargaining promotes the miscarriage of justice. We believe that our approach can provide such evidence. If plea bargaining causes a miscarriage of justice, it is in the public interest to reduce or eliminate it. The judiciary should be expanded or made more efficient so that innocent defendants are not pressured to plead guilty, and the guilty do not benefit from sentence discounts. Using plea bargains to reduce pressure on the courts creates injustice and may undermine confidence in the judiciary. Justice is not a matter to be decided in the plea-bargaining bazaar.

Funding The authors are grateful for financial support from the Barak Center for Interdisciplinary Legal Studies.

Appendix 1: Bounding the Effect of Plea Bargaining on Conviction Probabilities

Let Y_1 denote a dichotomous variable, which equals 0 if defendants plead guilty (plea bargain) and equals 1 if defendants are tried in court. Let Y_2 denote a dichotomous variable, which equals 0 if defendants are acquitted and equals 1 if convicted. Y_2 is observed when $Y_1 = 1$. Since both prosecutor and defendant must agree to a plea bargain, $Y_1 = 0$ when there is mutual agreement. In practice, however, the data do not reveal why $Y_1 = 1$. This eventuality arises when either or both did not plea bargain.

Let $P(Y_2 = 1/\ Y_1 = 1,\ X)$ denote the conditional probability of conviction where X is a set of observed exogenous controls. This conditional probability is revealed by the data. Let $P(Y_1 = 1/\ X)$ denote the conditional probability of no plea bargain, which is also revealed by the data. Let $P(Y_2 = 1/\ X) = P^*$ denote the unconditional probability of conviction, which is the counterfactual conviction probability in the absence of plea bargaining, and which is not revealed by the data. Applying Bayes' theorem this counterfactual probability is:

$$P^* = P(Y_2 = 1/Y_1 = 1, X)P(Y_1 = 1/X) + P(Y_2 = 1/Y_1 = 0, X)P(Y_1 = 0/X) \tag{8}$$

Equation (8) shows that the counterfactual status of this probability depends on the counterfactual probability of conviction had plea bargainers been tried, $P(Y_2 = 1/Y_1 = 0,\ X) = C$. According to shadow trial theory, this counterfactual is the shadow outcome, which motivates plea bargains. Since C is naturally bounded between 0 and 1, the solution to Eq. (8) is bounded between:

$$P_L^* = P(Y_2 = 1/Y_1 = 1, X)P(Y_1 = 1/X) \leq P^* \leq P_L^* + P(Y_1 = 0/X) = P_H^* \tag{9}$$

For example, if 40% of defendants are plea bargainers and 70% of those who faced trial are convicted, $P_L^* = 0.42$ and $P_H^* = 0.82$. These bounds mean that in the absence of plea bargaining, the probability of conviction could not have been less than 0.42, i.e. when all plea bargainers are acquitted, and it could not be greater than 0.82, i.e. when all plea bargainers are convicted. The data are informative about what cannot be. In the absence of empirical data, the bound of ignorance, or ambiguity, concerning the probability of conviction is 1 since in theory all defendants might have been convicted and all might have been acquitted. In this hypothetical case, the data reduce this bound to 0.82–0.42 = 0.4; so the cup is (slightly) more than half-full. Note that since plea bargainers plead guilty, 82% of defendants are guilty under plea bargaining. Since this rate equals P_H it must be the case that in the absence of plea bargaining the number of convictions would have been less, and could be as low as 42%.

The bounds discussed above are entirely non-parametric in the sense of Manski (1995). This minimalistic approach simply uses the data without making assumptions. The bounds may be narrowed by making various a priori or parametric assumptions about self-selection and its statistical distribution. For example, if defendants are rational, the counterfactual that all plea bargainers would be acquitted is unreasonable. Defendants who were tried rejected plea bargains despite a conviction rate of 0.7. Since these defendants are observationally similar to plea bargainers, this suggests that the lower bound for C is 0.7 rather than 0. On the other hand, there is no reason why C cannot be 1. Substituting C = 0.7 into Eq. (9) increases the lower bound to 0.7 from 0.42 while the upper bound remains at 0.82, and the ambiguity is reduced from 0.4 to 0.12. Making untestable assumptions, such as

rationality, reduces ambiguity. Indeed, the assumption that plea bargainers randomly select themselves would eliminate the ambiguity entirely, in which case P^* is 0.7 in this hypothetical case.

Appendix 2: Parametric Alternatives

In the numerical illustration of Eq. (7) it was assumed that β, γ and ρ are known. They are, of course, unknown and dependent. Our objective is to obtain consistent estimates of β, γ and ρ. Having obtained these estimates, we may use Eq. (7) to calculate conviction probabilities for individual defendants that allow for self-selection in plea bargaining.

If y_1 and y_2 have a bivariate normal distribution with correlation ρ their cumulative density function is:

$$F(y_1, y_2, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \exp\left[-\frac{1}{2(1-\rho^2)}\left(y_1^2 + y_2^2 - \rho y_1 y_2\right)\right] dy_1 dy_2 \quad (10a)$$

If instead y_1 and y_2 have a bivariate logistic distribution their cumulative density function does not involve integrals (Dubin and Rivers 1989):

$$F(y_1, y_2, \rho) = \frac{1}{1 + \left[\exp\left(-y_1/k\right) + \exp\left(-y_2/\kappa\right)\right]^{\kappa}} \quad (10b)$$

$$\rho = 1 - \frac{\kappa^2}{2} \qquad 0 < \kappa \leq \sqrt{2} \quad (10c)$$

The logistic distribution has fatter tails than the normal distribution, and is therefore more likely to capture extreme behavior. If, for example, innocents feel strongly that they wish to prove their innocence in court, or the guilty strongly resist plea bargains, their behavior might be more appropriately modelled by the logistic instead of the normal distribution. Also, the contours of the bivariate logistic distribution, unlike the contours for the bivariate normal, are asymmetric due to skewness. The claim that in practice probit and logit models produce similar results might be applicable to mundane choices regarding mortgages, cars, etc., but matters might be different regarding deviant behavior such as murder, rape and insistence on criminal trial. Other possible parametric assumptions include the bivariate Burr (type II) and Gumbel distributions, which add skewness as well different tail properties than in the normal (Kotz et al. 2000). However, an advantage of the bivariate normal is that ρ is unrestricted, whereas Eq. (10c) implies that ρ must be positive. Kotz et al. (2000) show that ρ is bounded between 0 and −0.40365 for the Gumbel I case; it is bounded between −1/4 and 1/4 in the Gumbel II case; and it is positive in the bivariate gamma case.

Dubin and Rivers (1989) suggested a ML estimator based on Eq. (10b), i.e. the bivariate logistic distribution. An obvious limitation of the bivariate logistic model is that according to Eq. (10c) ρ cannot be negative. A further limitation is that despite the simplicity of Eq. (10b) relative to Eq. (10a), the first order conditions of the likelihood function are more nonlinear than their bivariate normal counterparts. Perhaps this explains why the bivariate normal specification has been more popular than its logit counterpart, especially in the social sciences. However, Kotz et al. (2000) note that bivariate exponential distributions are popular in the

natural sciences. It is well known that results might not be robust with respect to what are arbitrary parametric assumptions (Eyal and Beenstock 2008; Heckman and Singer 1984). We therefore report results for alternative parametric specifications in the interest of robustness.

More recently, sample selectivity has been based on copulas (Smith 2003), which parametrize the cumulative distribution of y_1 and y_2 independently of the probabilistic models for sample selection and outcomes. For example, we use probit models for plea bargains and convictions without assuming that y_1 and y_2 are bivariate normal. We assume instead, for example, that their cumulative density is specified in terms of a Frank copula:

$$F(y_1 y_2) = -\frac{1}{\theta} ln \left[1 + \frac{\left(e^{-\theta y_1} - 1\right)\left(e^{-\theta y_2} - 1\right)}{e^{-y_1}} \right]$$

The advantage of this copula over its rivals (e.g. Clayton and Gumbel) is that it does not restrict the sign of the correlation between y_1 and y_2, nor does it restrict its range. However, in common with other copulas, only the Spearman correlation is defined.

Since alternative parametric assumptions are generally non-nested, a non-nested test (Santos Silva 2001) may be used to distinguish between rival hypotheses. Let $\hat{P}_{0i}$ denote the predicted probability for individual i according to the null hypothesis (model 0), and $\hat{P}_{1i}$ denote the predicted probability from a rival hypothesis (model 1). If these hypotheses are nested, a likelihood ratio test may be used to distinguish between them (as in the case of Eq. 10b and Burr II). In the non-nested case, matters are different because neither hypothesis is a special case of its rival. The following generated regressor is specified as a covariate in model 0 with coefficient δ_0:

$$\psi_{0i} = \frac{\hat{P}_{0i}(1 - \hat{P}_{0i})}{\eta \partial \hat{P}_{0i} / \partial x_{0i}} \left[\left(\frac{\hat{P}_{1i}}{\hat{P}_{0i}} \right)^{\eta} - \left(\frac{1 - \hat{P}_{1i}}{1 - \hat{P}_{0i}} \right)^{\eta} \right] \tag{11}$$

where $0 \leq \eta \leq 1$ and x_0 is vector of covariates in the null. Note that Eq. (11) simplifies to:

$$\psi_{0i} = (\hat{P}_{1i} - \hat{P}_{0i}) \div \frac{\partial \hat{P}_{0i}}{\partial x_{0i}} \tag{12}$$

Next, the roles of null and rival are exchanged in Eq. (11) with model 1 as the null and model 0 as its rival, and ψ_{1i} is specified alongside x_{1i} in model 1 with coefficient δ_1. There are four possible results:

(i) δ_0 is not statistically significant and δ_1 is statistically significant, i.e. model 0 rejects model 1 but model 1 does not reject model 0. Therefore, model 0 is preferred to model 1.
(ii) δ_1 is not statistically significant and δ_0 is statistically significant, i.e. model 1 rejects model 0 but model 0 does not reject model 1. Therefore, model 1 is preferred to model 0.
(iii) δ_0 and δ_1 are not statistically significant, i.e. model 0 does not reject model 1 and model 1 does not reject model 0. Therefore, neither model is preferred.
(iv) δ_0 and δ_1 are statistically significant, i.e. model 0 rejects model 1 but so does model 1 reject model 0. Therefore, neither model is preferred. However, a mixture model of models 0 and 1 is preferred to models 0 and 1.
(v) The partial derivatives from Eq. (5) are:

$$\frac{\partial P_i}{\partial X_{ki}} = \frac{\beta_k}{\sqrt{1-\rho^2}}\phi(z) \tag{13a}$$

$$\frac{\partial P_i}{\partial Z_{ki}} = \frac{\rho\gamma_k}{\sqrt{1-\rho^2}}\phi(z) \tag{13b}$$

$$\frac{\partial P_i}{\partial \rho} = \frac{(1-\rho^2)(1+\rho)Z_i\gamma + X_i\beta}{(1-\rho^2)^{3/2}}\phi(z) \tag{13c}$$

.

(vi) where ϕ(z) denotes the density of the standard normal distribution. Note that since $\phi'(z) > 0$ when $z<0$ and is negative when $z>0$, the signs of 2nd order derivatives depend on z. Notice that whereas the signs of Eq. (13a) and (13c) do not depend on the sign of ρ, the sign of Eq. (13b) depends on the sign of ρ. Equation (13a) states that if a variable increases the probability of conviction, its marginal effect on the probability of conviction is positive, as expected. Equation (13b) states that if a variable increases the probability of no plea bargain, its marginal effect on the probability of conviction is positive if ρ is positive, and negative if ρ is negative. The intuition for the former is that defendants are more likely to be convicted at trial because their g is larger. Finally, Eq. (13c) states that given what is observed about defendants (Z and X), their probability of conviction varies directly with ρ. The intuition is that on average their g is larger.

Appendix 3: Pre-trial Detention and Other Endogenous Variables

The specification of Z and X parameters in Y_1 and Y_2, respectively, refers to exogenous variables and excludes variables over which defendants or prosecutors exercise control. For example, although detention and bail status may influence defendants' decisions on plea bargaining (Euvrard and Leclerc 2017; Kellough and Wortley 2002), such variables are just as endogenous as plea bargaining itself. It would only be legitimate to include detention status in Z if its latent variable is independent of y_2, i.e. if it is unrelated to guilt.

Let D = 1 if defendants are detained and is zero otherwise. The latent variable for D is hypothesized as:

$$D_i^* = K_i\kappa + f_i \tag{14a}$$

$$f_i = g_i + d_i \tag{14b}$$

where K denotes covariates hypothesized to influence detention status, which also depends on defendants' guilt signals, and the toughness of their DAs. Suppose detention status is specified in Eq. (3a):

$$Y_{1i}^* = Z_i\gamma + \psi D_i^* + v_i \tag{15a}$$

$$v_i = -g_i' - e_i' - c_i + d_i \tag{15b}$$

where $\psi < 0$ is hypothesized to be negative. For mathematical convenience D^* is specified in Eq. (15a) rather than D. The covariance between v and f is:

$$\sigma_{vf} = -\sigma_{gg'} - \sigma_{g'd} - \sigma_{e'e} - \sigma_{e'e} - \sigma_{gc} + \sigma_{dg} + \sigma_d^2 \tag{16}$$

If $\sigma_{vf} > 0$, the estimate of ψ is inconsistent and is biased upwards. Therefore, an estimate of $\psi = 0$ does not mean that detention does not affect plea bargaining. If $\sigma_{vf} < 0$, the estimate of ψ is biased downwards. Therefore, if ψ is estimated to be negative it does not mean that detention encourages plea bargaining. The problem is that the causal effect of detention on plea bargains (ψ) is not identified in Eq. (15a). Identification requires that Z excludes variables that are specified in K.

Substituting Eq. (14a) into Eq. (15a) generates:

$$Y_{1i}^* = Z_i\gamma + K_i\psi\kappa + v_i + \psi f_i \tag{17}$$

which is how we interpret the selection in Eq. (8), i.e. variables that might affect detention status are specified in the selection model for plea bargaining. Similar arguments apply to legal representation (Alschuler 1975), which would be endogenous in the y_2 model as well as in the y_1 model.

References

Abrams DS (2011) Is pleading really a bargain. J Empir Legal Stud 8(S1):22

Alschuler AW (1975) The defense Attorney's role in plea bargaining. Yale Law J 84(6):1179–1314. https://doi.org/10.2307/795498

Alschuler AW (1983) Implementing the criminal defendant's right to trial: alternatives to the plea bargaining system. Univ Chicago Law Rev 50(3):120

Alschuler AW (2015) A nearly perfect system for convicting the innocent. Albany Law Rev. 79:919

Bandyopadhyay S, McCannon BC (2014) The effect of the election of prosecutors on criminal trials. Public Choice 161(1–2):16

Bandyopadhyay S, McCannon BC (2015) Prosecutorial retention: signaling by trial. J Public Econ Theory 17(2):37

Bibas S (2004) Plea bargaining outside the shadow of trial. Harv Law Rev 117(8):2463–2547. https://doi.org/10.2307/4093404

Bushway SD, Redlich AD (2012) Is plea bargaining in the "shadow of the trial" a mirage? J Quant Criminol 28(3):437–454

Bushway SD, Redlich AD, Norris RJ (2014) An explicit test of plea bargaining in the "shadow of the trial". Criminology 52(4):32

Church TW (1979) In defense of "bargain justice". Law Soc Rev 13(2):509–525. https://doi.org/10.2307/3053266

Dervan LE, Edkins VA (2013) The innocent defendant's dilemma: an innovative empirical study of plea bargaining's innocence problem. J Crim Law Criminol 103(1):48

Dubin JA, Rivers D (1989) Selection bias in linear regression, logit and probit models. Sociol Methods Res 18(2–3):360–390

Easterbrook FH (1992) Plea bargaining as compromise. Yale Law J 101(8):10

Euvrard E, Leclerc C (2017) Pre-trial detention and guilty pleas: inducement or coercion? Punishm Soc 19(5):525–542

Eyal Y, Beenstock M (2008) Sign reversal in LIVE treatment effect estimates: the effect of vocational training on unemployment duration. Labour Econ 15(5):1102–1125

Fishman G, Rattner A (1997) The Israeli criminal justice system in action-is justice administered differentially? J Quant Criminol 13(1):22

Fishman G, Rattner A, Turjeman H (2006) Sentencing outcomes in a multinational society: when judges, defendants and victims can be either arabs or jews. Eur J Criminol 3(1):16

Gazal-Ayal O (2006) Partial ban on plea bargains. Cardozo Law Rev 27(5):56

Gazal-Ayal O, Riza L (2009) Plea-bargaining and prosecution. In: Garoupa N (ed) Criminal law and economics. Edward Elgar, Cheltenham

Gazal-Ayal O, Sulitzeanu-Kenan R (2010) Let my people go: ethnic in-group bias in judicial decisions—evidence from a randomized natural experiment. J Empir Legal Stud 7(3):26
Gazal-Ayal O, Tor A (2012) The innocence effect. Duke Law J 62:339
Gazal-Ayal O, Weinshall-Margel (2014) The power of the prosecution in criminal proceedings—an empirical study. Mishpatim 44(3):55
Greene WH (2012) Econometric analysis, 7th edn. Pearson, London
Gross SR (2008) Convicting the innocent. Ann Rev Law Soc Sci 4:173–192
Gross SR (2011) Pretrial incentives, post-conviction review, and sorting criminal prosecutions by guilt or innocence. NY L Sch L Rev 56(3):1009–1030
Gudjonnson GH, Sigurdsson JF (2008) How frequently do false confessions occur? An empirical study among prison inmates. Psychol Crime Law 1(1):6
Heckman J (1976) The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimation for such models. Ann Econ Soc Meas 5(4):17
Heckman J, Singer B (1984) A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52:271–320
Hendry DF (1995) Dynamic econometrics. Oxford University Press, Oxford
Hoffman M (2007) The myth of factual innocence. Chicago-Kent Law Rev 82(2):27
Huff R, Rattner A, Sagarin E (1996) Convicted but innocent: wrongful conviction and public policy. SAGE, Thousand Oaks
Israeli Courts Administration (2003–2011) Data and statistical report. https://www.gov.il/he/Departments/publications/. Accessed 1 Mar 2017
Israeli Courts Administration (2018) Annual data and statistical report. https://www.gov.il/he/Departments/publications/. Accessed 15 May 2019
Johnson BD (2015) Examining the "life course" of criminal cases: a new frontier in sentencing research. Criminol Public Policy 14(2):4
Johnson BD, King RD, Spohn C (2016) Sociolegal approaches to the study of guilty pleas and prosecution. Ann Rev Law Soc Sci 12:479–495
Kellough G, Wortley S (2002) Remand for plea Bail decisions and plea bargaining as commensurate decisions. Br J Criminol 42(1):186–210
Kotz S, Balakrishnan N, Johnson NL (2000) Bivariate and trivariate normal distributions. In: Continuous multivariate distributions: models and applications, 2nd ed, vol 1. Wiley & Sons, New York
Kutateladze BL, Andiloro NR, Johnson BD, Spohn CC (2014) Cumulative disadvantage: examining racial and ethnic disparity in prosecution and sentencing. Criminology 52(3):514–551
Kutateladze BL, Lawson VZ, Andiloro NR (2015) Does evidence really matter? An exploratory analysis of the role of evidence in plea bargaining in felony drug cases. Law Hum Behav 39(5):431
LaFree GD (1985) Adversarial and nonadversarial justice: a comparison of guilty pleas and trials. Criminology 23(2):289–312
Landes WM (1971) An economic analysis of the courts. J Law Econ 14(1):61–107
Loeffler C, Hyatt J, Ridgeway G (2018) Measuring self-reported wrongful convictions among prisoners. J Quant Criminol. https://doi.org/10.1007/s10940-018-9381-1
Maddala GS (1983) Limited-dependent and qualitative variables in econometrics. Cambridge University Press, Cambridge
Manski C (1995) Identification problems in the social sciences. Harvard University Press, Cambridge
Nagel SS, Neef M (1979) Decision theory and the legal process. Lexington Books, Lexington
National Registry of Exonerations (2017) Spreadsheet downloaded from National Registry of Exonerations website on 12/31/17. Retrieved from https://www.law.umich.edu/special/exoneration/Pages/about.aspx
Office of the State Attorney (2018) Summary Report. Retrieved from https://www.gov.il/he/Departments/publications/
Rakoff JS, Daumier H, Case AC (2014) Why innocent people plead guilty. NY Rev Books 61(18):1–12
Redlich AD, Summers A, Hoover S (2010) Self-reported false confessions and false guilty pleas among offenders with mental illness. Law Hum Behav 34(1):12
Redlich AD, Bushway SD, Norris RJ (2016) Plea decision-making by attorneys and judges. J Exp Criminol 12:25
Redlich AD, Bibas S, Edkins VA, Madon S (2017a) The psychology of defendant plea decision making. Am Psychol 72(4):14
Redlich AD, Wilford MM, Bushway SD (2017b) Understanding guilty pleas through the lens of social science. Psychol Public Policy Law 23(4):14
Rosenbaum PR (2002) Observational studies. Springer-Verlag, New York
Santos Silva JMC (2001) A score test for non-nested hypotheses with applications to discrete data models. J Appl Econom 16(5):577–597

Smith DA (1986) The plea bargaining controversy. J Crim Law Criminol 77(3):949–968
Smith MD (2003) Modelling sample selection using archimedean copulas. Econom J 6(1):24
Spence M (1973) Job market signaling. Q J Econ 87(3):355–374
Spohn C (2013) Racial disparities in prosecution, sentencing, and punishment. The Oxford handbook of ethnicity, crime, and immigration, 166–193
Tor A, Gazal-Ayal O, Garcia SM (2010) Fairness and the willingness to accept plea bargain offers. J Empir Legal Stud 7(1):97–116
Turner JI, Redlich AD (2016) Two models of pre-plea discovery in criminal cases: an empirical comparison. Wash Lee Law Rev 73(1):124
Uggen C (1999) Ex-offenders and the conformist alternative: a job quality model of work and crime. Soc Probl 46(1):25
Ulmer JT, Bradley MS (2006) Variation in trial penalties among serious violent offenses. Criminology 44(3):40
Ulmer JT, Bradley MS (2017) Handbook on punishment decisions: locations of disparity. Routledge, New York
Ulmer J, Eisenstein J, Johnson BD (2010) Trial penalties in federal sentencing: extra-guidelines factors and district variation. Justice Q 27(4):32
Walker S, Spohn C, DeLone M (2012) The color of justice: race, ethnicity, and crime in America. Cengage Learning, Wadsworth
Wooldredge J (1989) An aggregate-level examination of the caseload pressure hypothesis. J Quant Criminol 5(3):24
Wooldredge J, Frank J, Goulette N, Travis L (2015) Is the impact of cumulative disadvantage on sentencing greater for black defendants? Criminol Public Policy 14(2):37
Wynand PVdV, van Praag BMS (1981) The demand for deductibles in private health insurance: a probit model with sample selection. J Econom 17(2):23
Zhao Q, Small D, Bhattacharya B (2019) Sensitivity analysis for inverse probability weighting estimators via the percentile bootstrap. J R Stat Soc Ser B. https://doi.org/10.1111/rssb.12327
Zottoli TM, Daftary-Kapur T, Winters GM, Hogan C (2016) Plea discounts, time pressures, and false-guilty pleas in youth and adults who pleaded guilty to felonies in New York City. Psychol Public Policy Law 22(3):9