

ECON2390 Fall 2024

Applied Econometrics I

Final

Adrien Foutelet

December 18, 2024

- 1 This question is based on the article by Agan, Doleac, and Harvey (2023, QJE), “Misdemeanor Prosecution” [1] posted in the exam folder of the Canvas course page.
 - 1.1 The authors use a residualized leave-out ADA leniency as an instrument for the nonprosecution decision. Explain how the instrument is constructed, what its value represents (e.g., how to interpret a leniency value of 0.05), and why it satisfies the Random Assignment assumption.

Agan, Doleac, and Harvey focus on the effect of nonprosecution decisions on defendants' post arraignment outcomes. To address causality, they carry out an IV analysis. For that, they use the residualized leave-out mean measure of Assistant District Attorney (ADA) leniency as an instrument for nonprosecution decisions.

Consider individual i , case c , and year t . The structural model is:

$$Y_{ict} = \beta_1 \cdot NotProsecuted_{ict} + \beta_2 X_{it} + \gamma_{ct} + \epsilon_{ict}$$

Consider court r . Consider the following projection:

$$NotProsecuted_{icrt} = \gamma'_{rt} + \delta_{irt}$$

Define the residualised nonprosecution decision:

$$NotProsecuted_{icrt}^* = NotProsecuted_{icrt} - \gamma'_{rt}$$

This variable solves the issue of confoundedness by selection of nonprosecution decisions. Indeed, nonviolent misdemeanor case types may vary by court, month of the year, and day of the week.

Consider ADA a , number of nonviolent misdemeanor cases arraigned by a n_a , and number of nonviolent misdeaminor cases invovling defendant i arraigned by ADA a $n_{i,a}$. The instrument is:

$$Z_{cta} = \frac{1}{n_a - n_{i,a}} \left(\sum_{k=0}^{n_a} NotProsecuted_{ict}^* - \sum_{k=0}^{n_{i,a}} NotProsecuted_{ict}^* \right)$$

The second stage of the 2SLS estimation is then:

$$NotProsecuted_{ict} = \alpha_1 Z_{cta} + \alpha_2 X_{it} + \gamma''_{ct} + u_{ict}$$

The value of the instrument corresponds to how more likeley to not prosecute a nonviolent misdemeanor case an ADA is relative other ADAs holding case type fixed (i.e, holding court, month of the year and day of the week fixed).

A value of 0.05 means that, for the current defendant, the ADA is five percentage points more likeley than the average ADA to not prosecute a nonviolent misdemeanor.

The Random Assignment assumption consists in the fact that the instrument variable is randomly assigned independent of any observed heterogeneity determining one's outcome and/or treatment selection. The reasons for which the instrument satisfies the random assignment assumption are:

- ADA leniency is a leave out mean measure of arraigning ADA non-prosecution decisions.

This ensures that the instrument be not correlated with non-prosecution through ϵ_{ict} (see next answer for more detail).

- As-if random assignment of misdemeanor cases to arraigning ADAs:

- By construction, since the cases are assigned to ADAs based on availability and court schedules, since ADAs are assigned to arraignment courtrooms on a rotating scheduel (weekly basis), and since there is no possibility for defendants and ADAs to choose each other, defendants are not sorted systematically to ADAs.

- Empirically, observable defendant characteristics are balanced across ADA leniency scores.

Thanks to this, the only possible confounders in the relationship between ADA leniency and non-prosecution are court, month of the year, and day of the week.

- Residualisation of the non-prosecution decision in the computatation of ADA leniency.

The residualisation includes court, month of the year, and day of the week fixed effects, which leaves no possible counfounder in the relation ship between ADA leniency and non-prosecution.

- 1.2** On page 1467, the authors state: “As is standard to avoid the small-sample correlation between the ADA decision in this case and her average leniency, we then construct the leave-out mean measure of ADA nonprosecution (leniency).” Explain why constructing leniency by leaving-out is important. How might failing to account for this issue affect the instrument exogeneity?

The naive OLS estimation of the structural model obviously suffers from endogeneity, i.e., non zero covariance between the defendant’s non-prosecution and the error term. For example, arraigning ADAs are more likely to prosecute misdemeanor defendants who have prior criminal convictions and defendants with prior convictions are more likely to have subsequent criminal justice contacts.

Then if, for each defendant, ADA leniency is computed as a mean including the defendant’s non-prosecution, mechanically the instrument is itself correlated with the second stage error term, causing endogeneity in the second stage (breaking the exclusion restriction) and biasing the estimate (self-inclusion bias). What is more, the smaller the sample, the more ”driven” by the defendant’s non prosecution ADA leniency is.

The leave-out mean measure addresses that issue, removing the defendant’s cases from the sample from which ADA leniency is computed (no overlap between the instrument and the dependent variable).

- 1.3** A potential caveat of the paper is that the authors do not directly observe nonprosecution and instead proxy for it by the absence of further case events. Could the use of a proxy introduce bias in the estimates? If so, explain in which direction and why. If not, provide an argument to support the validity of this approach.

Doing so, they would:

- Count prosecuted defendants with only one case event (e.g., dying past arraignment and prior to further case events; case not closed; specific prosecutions needing no further case event) as not prosecuted (Type 1 error, False positives).
- Count non-prosecuted defendants with multiple case events (e.g., prosecution decision split in several events due to adjournment for exceptional circumstances) as prosecuted (Type 2 error, False negatives).

The summary statistics provided in Table 1 only provide the average number of case event for prosecuted and non prosecuted defendants which makes difficult to validate the possibility of the former case. It validates the possibility of the latter case as the average number of case events for non-prosecuted defendants is different from 1 (1.002).

To discuss the bias, consider the fact that prosecuted defendants have higher rates of recidivism, as seen in Table 1:

- In case of type I error, there will be an upward bias of the rates of recidivism of units proxied as non-prosecuted. Therefore, there will be a downward bias of the estimate of the effect of non-prosecution on recidivism.

- In case of type II error, there will be an downward bias of the rates of recidivism of units proxied as prosecuted. Therefore, there will be an upward bias of the estimate of the effect of non-prosecution on recidivism.

1.4 Is the monotonicity assumption plausible in this setting? Explain. If it is plausible, discuss how the tests they do in the paper support your conclusion. If it does not seem plausible, how would your interpretation of the results change? What would the instrument be identifying instead?

The monotonicity assumption, under heterogeneous treatment effects, is plausible in this setting. It consists in the fact that the effect of ADA assignment on the probability of nonprosecution is monotonic across defendants. It implies that a defendant not prosecuted by a strict ADA would not be prosecuted by a lenient ADA; that a defendant prosecuted by a lenient ADA would also be prosecuted by a strict ADA.

The authors walk in the footsteps of Frandsen, Lefgren, and Leslie (2023) [2], who show that the assumption can be relaxed to an average monotonicity assumption. The authors show that ADA leniency is positively correlated with nonprosecution in a large variety of subsamples. This means that prosecutors who are more lenient overall are more likely to not prosecute people in any observable subgroup. This positively tests that the covariance between a defendant's prosecutor-specific treatment and the prosecutor's overall propensity to not prosecute is weakly positive, which in turn positively tests the average monotonicity assumption.

1.5 Is the exclusion assumption plausible in this context? The authors assess this assumption by investigating whether leniency predicts subsequent case outcomes for prosecuted defendants. However, on page 1473, they concede that "it is also not clear that if we did find associations between ADA leniency and case outcomes that this would represent a violation of the exclusion restriction." What alternative approach could they use to evaluate not only the exclusion restriction but also the overall validity of the instrument? In your answer, giving the main idea of your approach suffices.

The exclusion assumption consists in the fact that ADA leniency only affects post-arrangement outcomes through the prosecution decision. It seems plausible since the ADA making the non-prosecution decision has (practically) no other role in the defendant's case past the arraignment. Notably, most bail, pre-trial motions and case outcomes are handled by different ADAs unrelated to the initial one. The second argument, the validity of which they are not certain about, lies in the fact that the authors find no correlation between leniency and post-arrangement outcomes for prosecuted defendants.

An alternative approach to evaluate the exclusion restriction and the overall validity of the instrument would be to use a placebo test assessing how ADA leniency predicts defendants' pre-treatment characteristics. The tested hypothesis

would be:

$$H_0 : \text{"The instrument is correlated with pre-treatment characteristics"}$$

If H_0 cannot be rejected (a significant correlation is estimated), we would cast light on the non-random distribution of ADA leniency across defendants, which would invalidate the random assignment assumption. Further, this would give the possibility of alternative channels to prosecution decisions for ADA leniency to effect post-arrainment outcomes. For example, it could be that richer defendants pay lawyers who, because of their deep knowledge of courts or corruption, could make sure that their clients will be facing the most lenient ADAs. It is not unrealistic that, if such lawyers existed, they would assess ADA leniency by computing the same leave-out mean of an ADA's prosecution decision. Then there would be selection of the richest defendants both amongst the non-prosecuted and the ones facing lenient ADAs, such that ADA leniency would affect post-arrainment outcomes both through the prosecution decision and the affluence of the defendants, invalidating the exclusion restriction.

Our placebo test would therefore evaluate both the plausibility of the random assignment assumption and the exclusion restriction at once, hence would test the overall validity of the instrument.

1.6 Consider the results presented in Table III. Interpret the findings, focusing on what the coefficients imply about the impact of nonprosecution on subsequent criminal complaints. Given that the analysis uses a continuous, constructed instrument (ADA estimated leniency), who are the compliers in this setting?

Given that the analysis uses a continuous instrument, the compliers are the marginal defendants: those whose prosecution decision is on the razor edge and could fall on one side or the other depending on leniency. A marginal increase in ADA leniency shifts the marginal defendant to the non-prosecuted side.

Table III presents the OLS, 2SLS, and reduced form estimates for two post-arrainment outcomes: likelihood of subsequent criminal complaints within two years and number of subsequent criminal complaints within two years. The OLS and 2SLS estimates are presented with court x time fixed effects and with and without a set of case and defendant covariates.

Estimates are significant at the 1 percent level, except the ones in the reduced form model and the one for the number of subsequent criminal complaints within two years in the 2SLS model with the set of covariates. Magnitudes are large. We can see that all the OLS and 2SLS estimates loose magnitude when adding the set of covariates but stay at the same order.

We can see (OLS with controls) that non-prosecution reduces the probability of subsequent criminal complaints within two years by 10 percentage points and the number of these complaints by 0.51.

We can see (2SLS with controls) that non-prosecution of marginal defendants reduces the probability of subsequent criminal complaints within two years by 29 percentage points and the number of these complaints by 1.71.

We can see (reduced form) that ADA leniency reduces the probability of subsequent criminal complaints within two years by 16 percentage points and the number of these complaints by 0.93.

These effects have to be considered with a mean probability of subsequent complaints within two years and a mean number of these complaints for prosecuted defendants of 0.37 and 1.64 respectively.

The reduced form show that, even if the exclusion restriction were violated, we would still have causality between ADA leniency and on post-arraigning outcomes. Indeed, it captures the total effect of ADA leniency, without decomposing into the channels through which it operates: the direct effect (via non-prosecution decision) and the indirect effect (breaking the exclusion restriction).

1.7 Seeing the 2SLS estimate from the paper, policymakers suggest a new policy eliminating the possibility of prosecuting misdemeanors for all defendants. What additional steps and/or assumptions do you need to estimate what would be the effect of this policy on future criminal complaints within two years? Discuss.

To estimate the effect of this policy on future criminal complaints within two years, we would need to move from a Local Average Treatment Effect (LATE) estimation to an Average Treatment Effect. The LATE estimation, which is what we did with 2SLS, focuses on the marginal compliers, defendants whose non-prosecution changed because of ADA leniency. By contrast, the ATE focuses on the whole population, compliers and non-compliers, hence encompasses the effect on always takers (those who would always be non-prosecuted irrespective of ADA leniency) and the never takers (those who would always be prosecuted irrespective of ADA leniency).

That switch from LATE to ATE can be made by satisfying several assumptions.

- The representativeness of compliers assumption. It means that the set of compliers is a representative sample of the broad defendant population, in light of individual and case characteristics. In that respect, if a balance test failed, recomputing the LATE with inverse probability weighting would solve the issue.
- The homogeneous treatment effect assumption. It means that the causal effect of being not prosecuted on post-arraigning outcomes for compliers is the same as that for non-compliers. This assumption seems plausible as we can assume that the compliers in our case are not too different from the non-compliers cognitively and psychologically and, because of that, the non-compliers would not react differently.
- No selection on unobservables. This assumption seems plausible for the same reason as the previous point.
- Stable Unit Treatment Value Assumption (SUTVA). It means that there is no interference (the treatment assigned to compliers does not affect the treatment of other never-takers and always-takers) and that there is no variations in treatment. This assumption is more questionable, since non-prosecution could affect police behavior (disappointment, triggering more zeal in the dealing with the next cases) or future ADA decisions (a "self-defined" quota may guide

the decision of an ADA: after deciding to not prosecute a certain number of people during one session, the ADA may feel a sort of duty to prosecute some comparable cases not to appear too lenient).

- 1.8 A researcher points out that ADAs not only have very different prosecution rates but also claim conflicts of interest at very different rates. Furthermore, they claim that anecdotal evidence suggests that ADAs with the highest prosecution rates are also those with the lowest rates of claiming conflicts of interest. If this concern is empirically valid, which assumption, if any, would it violate? If an assumption is violated, would you expect the main estimate of the paper to be biased upward or downward? Discuss.

If this concern were empirically valid, it would challenge the exclusion restriction: leniency would affect the post-arrainment outcomes through non-prosecution and another channel, the diligence of the ADA to see the defendants in certain cases prosecuted. It would also cause selection bias on unobservables: since more lenient ADAs would be more likely to wave defendants than less lenient ADAs, then defendants subject to systematic conflicts of interest (unobservable) would be more likely to eventually face a less lenient ADA than a lenient ADA, hence would be more likely to be prosecuted.

If the defendants systematically subject to conflicts of interest with ADA were also the most severe case in terms of offence, and hence in terms of likeliness to repeat offense, then we would get an upward biased effect of non-prosecution on post-arraining outcomes.

What is more, if the data were built such that, in case of waving, the recorded arraigning ADA is the one initially assigned, then the first stage estimate (regression of non-prosecution on leniency) would be of smaller magnitude than it should (attenuation bias due to first stage contamination), leading to a larger second stage estimate as it should. We would therefore get an even stronger upward bias of our effect.

2 This question is based on the analysis shown in “Social Media and Mental Health” by Braghieri, Levy, and Makarin (2022, AER) [3], as posted on Canvas.

Consider the following potential outcome framework. Let i denote students, c colleges, and t time periods. Let $D_{i,c,t}$ and $(Y_{i,c,t}(0), Y_{i,c,t}(1))$ respectively denote the treatment status and the potential outcomes without and with treatment of student i in college c at period t . Furthermore, let $G_{i,c} = \min\{t : D_{i,c,t} = 1\}$ denote the first time period that the students i in college c are exposed to the treatment.

2.1 Formally state the no anticipation and parallel trends assumption using the potential outcome framework and interpret them in the context of the paper

Consider student i in college c at period t .

- The no anticipation assumption:

$$\forall t < G_{i,c}, Y_{i,c,t}(1) = Y_{i,c,t}(0)$$

This means that a student's potential outcome (mental health) at any time before the rollout of the treatment (Facebook) in their college is not influenced by their future exposure to treatment.

- The parallel trend assumption:

$$\forall t > t', c \neq c', E[Y_{i,c,t}(0) - Y_{i,c,t'}(0)] = E[Y_{i,c',t}(0) - Y_{i,c',t'}(0)]$$

This means that the untreated potential outcome (mental health) would have followed the same trend across treated (those who would have gotten on Facebook) and untreated students (those who would not have gotten on Facebook) in the absence of treatment (no Facebook rollout on campus).

2.2 On page 3664, the authors claim that “our estimates capture the general equilibrium effects of introducing social media in an entire community. Such general equilibrium effects are arguably particularly important for technologies like social media that exhibit strong network externalities.” Explain why and how this claim might imply the violation of the main identifying assumptions in the Difference-in-Differences framework.

They describe spillover effects and network externalities, which could violate both assumptions.

- The no anticipation assumption: The individuals on each campus may anticipate the future rollout of Facebook, having heard from the strong network externalities associated with it. This can consist in a lot of excitement, time spent documenting themselves about the phenomenon, listing the people they would add as friends, worrying that their significant others may not think about adding them, hesitating whether they should choose to get treated or not, etc. These anticipatory behaviors affect their potential mental health, so the no anticipation assumption breaks.
- The parallel trend assumption: The individuals on each campus all belong to networks, in addition to the campus networks. Those in the treated group are likely to have a network of people equally interested in Facebook scattered on multiple campuses. This implies that, before campus rollout, those in the treatment group are likely to have someone in their network who started using Facebook. Since using Facebook has strong externalities, the individuals in question are likely to be affected by these externalities before campus rollout. These externalities could consist in the consequences on their own mental health of the mental health effect of Facebook on the person from their network. With treatment and control group mental health diverging pre-treatment, the parallel trend assumption breaks.

- 2.3 The NCHA data that the authors used to construct the outcome variable is a repeated cross-section of students instead of a panel (i.e., we do not observe the same students over time). What additional assumption is required to identify the treatment effect in this setup? Express this assumption using the potential outcome framework above.

The parallel trend assumption needs to hold in a cross sectional setting. This acknowledges the fact that we do not track the same individuals over time but the group of individuals specific to each college and time pair:

$$\forall t > t', c \neq c', i \neq i', E[Y_{i,c,t}(0) - Y_{i,c,t'}(0)] = E[Y_{i',c',t}(0) - Y_{i',c',t'}(0)]$$

This means that the untreated potential outcome (mental health) would have followed the same trend across the students in the treated group at each period (those who would have gotten on Facebook) and the students in the untreated group at each period (those who would not have gotten on Facebook) in the absence of treatment (no Facebook rollout on campus). Behind this assumption is the idea that the groups of students filling each college and time pair are comparable with one another in terms of observable and unobservable characteristics.

- 2.4 Consider the two-way fixed effects (TWFE) specification (1) on page 3671. Now, assume a simplified two-period setup $t = 1, 2$ with two colleges $c = 1, 2$. College $c = 1$ gets treated in the first period (i.e., $G_{i,1} = 1$) and college $c = 2$ gets treated in the second period (i.e., $G_{i,2} = 2$). What does the coefficient β identify in this setup? Derive the expression for β , explain it, and discuss whether you observe anything unusual about it.

In this setup, β is the Two Way Fixed Effect (TWFE) of the introduction of Facebook on student mental health. It estimates the difference in means of $Y_{i,1,t} - Y_{i,2,t}$ between "comparison" period $t = 2$ and "control" period $t = 1$. We know from the notes that:

$$\beta = \frac{1}{2} \left(\left(E[Y_{i,c,t} | c = 1, t = 1] - E[Y_{i,c,t} | c = 1, t = 2] \right) - \left(E[Y_{i,c,t} | c = 2, t = 1] - E[Y_{i,c,t} | c = 2, t = 2] \right) \right)$$

We have:

$$Facebook_{i,1,1} = 1$$

$$Facebook_{i,1,2} = 1$$

Hence:

$$Y_{i,1,1} = Y_{i,1,1}(1)$$

$$Y_{i,1,2} = Y_{i,1,2}(1)$$

Hence:

$$E[Y_{i,1,1} | c = 1, t = 1] = Y_{1,1}(1)$$

$$E[Y_{i,1,2} \mid c = 1, t = 2] = Y_{1,2}(1)$$

Hence:

$$E[Y_{i,c,t} \mid c = 1, t = 2] - E[Y_{i,c,t} \mid c = 1, t = 1] = Y_{1,2}(1) - Y_{1,1}(1)$$

Also, we have:

$$Facebook_{i,2,1} = 0$$

$$Facebook_{i,2,2} = 1$$

Hence:

$$Y_{i,2,1} = Y_{i,2,1}(0)$$

$$Y_{i,2,2} = Y_{i,2,2}(1)$$

Hence:

$$E[Y_{i,2,1} \mid c = 2, t = 1] = Y_{2,1}(0)$$

$$E[Y_{i,2,2} \mid c = 2, t = 2] = Y_{2,2}(1)$$

Hence:

$$E[Y_{i,c,t} \mid c = 2, t = 2] - E[Y_{i,c,t} \mid c = 2, t = 1] = Y_{2,2}(1) - Y_{2,1}(0)$$

Therefore

$$\beta = \frac{1}{2} \left(\left(Y_{1,2}(1) - Y_{1,1}(1) \right) - \left(Y_{2,2}(1) - Y_{2,1}(0) \right) \right)$$

What is unusual about this estimate is that college 1 is treated in both periods, but is used as a control for college 2. Instead of comparing college 2 to a truly untreated college, college 1 is used as a control, while it was already treated.

If there is homogeneity of the treatment across colleges and time (constant $SATT_{c,t}$), this is not an issue and the TWFE is equal to the ATT. Indeed, this ensures that the outcome variable in college 1 does not evolve differently across pre-treatment periods from how it would evolve there across post-treatment periods. This homogeneity assumption is a strong one.

2.5 More generally (i.e., with many time periods and many groups being treated at different times), what issues arise with specification (1) on page 3671 when treatment effects are heterogeneous across either time or units? What is the source of this problem?

When treatment effects are heterogeneous across either time or units, the TWFE is not equal to the ATT but is equal to the weighted average of the $SATT_{c,t}$ with weights equal to:

$$w_{c,t} = \frac{N_c \epsilon_{c,t}}{\sum_{c,t:G_{i,c}=c} N_c \epsilon_{c,t}}$$

with $\epsilon_{c,t}$ the residuals from regressing the treatment variable on the college and time fixed effects

(de Chaisemartin and D'Haultfoeuille 2020 [4]).

The issue comes from the fact that some weights can be negative, which implies that the overall estimate can be biased in an unpredictable fashion.

As we said earlier, the source of the problem is that the TWFE estimator implicitly uses earlier-treated units as controls for later-treated units, even though those earlier-treated units were already affected by the treatment. When treatment effects are heterogeneous across time or units, this leads to invalid comparisons and these negative weights, causing the TWFE estimator to produce a biased and inconsistent estimate of the true treatment effect.

3 A statistical decision problem

3. A statistical decision problem.

Consider:

$i \in [1, n]$: individual i in a population of size n .
 $y = (y_i)_{i \in [1, n]} \in \mathbb{R}^n$: the net income vector of the population.

$\tau = (\tau_i)_{i \in [1, n]} \in \mathbb{R}^n$: the net income adjustment vector of the population

$L(\tau, y) = \sum_{i=1}^n (z - y_i - \tau_i)^2$: the government's income transfer policy performance - evaluating function.

$\hat{y} = (\hat{y}_i)_{i \in [1, n]}$: the net income noisy estimate vector observed by the government.

Assume $\hat{y}_i \sim N(y_i, 1)$.

(c) The key elements of the statistical decision problem:

- The state space (where y lives): $\mathbb{H} = \mathbb{R}^n$
- The observations (where random variable \hat{y} lives): $\mathcal{Y} = \mathbb{R}^n$,
with draws distributed along $N(y, I_n)$
- The action space (where τ lives): $\mathcal{A} = \mathbb{R}^n$
- The decision function (the function corresponding to the government mapping observation to action):

$$\begin{aligned} s: \mathcal{Y} &\longrightarrow \mathcal{A} \\ \hat{y} &\longmapsto \underset{\tau}{\operatorname{argmin}} \left\{ d(z, \hat{y}) \right\} \\ &= \underset{\tau}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (z_i - \hat{y}_i - \tau_i)^2 \right\} \end{aligned}$$

- The loss function (the loss that the government has to "pay" in reference to the true state of nature when actions are taken):

$$L: \mathcal{A} \times \mathbb{H} \longrightarrow \mathbb{R}$$

$$\begin{aligned} (z, y) &\longmapsto L(z, y) \\ &= \sum_{i=1}^n (z_i - y_i - \tau_i)^2 \end{aligned}$$

- The risk (the average loss that the government has to "pay" in reference to the true state of nature when a decision function is chosen):

$$\begin{aligned}
 R: \mathcal{G}(y, \theta) \times \Theta &\longrightarrow \mathbb{R} \\
 (S \times y) &\longmapsto \mathbb{E}_{\hat{g}} \left[L(S(\hat{g}), y) \mid y \right] \\
 &= \int_{\mathcal{Y}} L(S(\hat{g}), y) dP(y \mid y)
 \end{aligned}$$

(b) Assume the decision rule is:

$$\hat{\tau}_c^{(c)}(y) = c(z - \hat{g}_i), \quad \text{with } c \in \mathbb{R}.$$

The risk then is:

$$\begin{aligned}
 R(\hat{\tau}^{(c)}, y) &= \mathbb{E}_{\hat{g}} \left[L \left[\sum_{i \in [1, n]} \hat{\tau}_i^{(c)}(y) \right], y \mid y \right] \\
 &= \mathbb{E}_{\hat{g}} \left[\sum_{i=1}^n (z - g_i - \hat{\tau}_i^{(c)}(y))^2 \mid y \right]
 \end{aligned}$$

$$= E_{\hat{y}} \left[\sum_{i=1}^n (z - y_i - c(\hat{z} - \hat{y}_i))^2 \mid y \right]$$

$$= \sum_{i=1}^n E_{\hat{y}} \left[(z - y_i - c(\hat{z} - \hat{y}_i))^2 \mid y \right]$$

$$= \sum_{i=1}^n E_{\hat{y}} \left[(z - y_i - cz + c\hat{y}_i)^2 \mid y \right]$$

$$= \sum_{i=1}^n E_{\hat{y}} \left[(z - y_i - cz)^2 + (c\hat{y}_i)^2 + 2(z - y_i - cz)c\hat{y}_i \mid y \right]$$

$$= \sum_{i=1}^n \left[(z - y_i - cz)^2 + c^2 E_{\hat{y}}(\hat{y}_i^2 \mid y) + 2(z - y_i - cz)c E_{\hat{y}}(\hat{y}_i \mid y) \right]$$

$$= \sum_{i=1}^n \left[(z - y_i - cz)^2 + c^2 \left(V_{\hat{y}}(\hat{y}_i \mid y) + [E_{\hat{y}}(\hat{y}_i \mid y)]^2 \right) + 2(z - y_i - cz)c y_i \right]$$

$$= \sum_{i=1}^n \left[(z - y_i - cz)^2 + c^2 (1 + y_i^2) + 2(z - y_i - cz)c y_i \right]$$

i.e. $R(\tau^c, y)$

$$= \sum_{i=1}^n \left[(z - y_i - cz)^2 + (z - y_i - cz) 2c y_i + c^2 (1 + y_i^2) \right]$$

Since we focus on a single individual i , we focus on one term of this sum only and call it the risk or from know on.

We have:

$$r: \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$(c, y_i) \mapsto (z - y_i - cz)^2 + (z - y_i - cz) 2c y_i + c^2 (1 + y_i^2)$$

$$\begin{aligned} (c)r_1(c, y_i) &= -2(z - y_i - cz) + 2c(1 + y_i^2) + 2zy_i - 2y_i^2 - 4zy_i c \\ &= -2z^2 - 2y_i^2 + 4zy_i + 2c[z^2 + 1 + y_i^2 - 2zy_i] \\ &= -2[z^2 + y_i^2 - 2zy_i] + 2c[1 + z^2 + y_i^2 - 2zy_i] \\ &= -2(z - y_i)^2 + 2c(z - y_i)^2 + 2c \end{aligned}$$

$$\text{i.e } r_1(c, y_i) = 2(z - y_i)^2 [c - 1] + 2c$$

Therefore $\forall y_i \in \mathbb{R}, \forall c \geq 1, r_1(c, y_i) > 0$.

This implies that the risk is strictly increasing in c on $[1, +\infty]$. And since the risk is continuous in c on the interval:

$$\forall y_i \in \mathbb{R}, \forall c \in [1, +\infty], \exists \delta > 0 : r(c - \delta, y_i) < r(c, y_i)$$

Therefore, for $c > 1$, the decision rule $\pi_i^{(c)}$ is inadmissible.

(d) Note that:

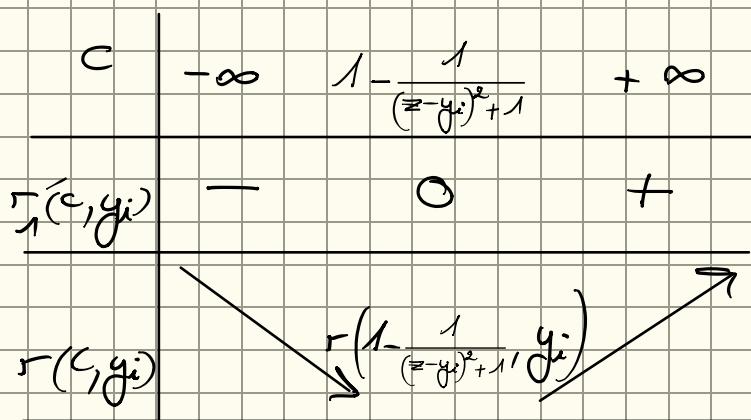
$$\bullet r'_1(c, y_i) = 0 \Leftrightarrow (z - y_i)^2 = c \left[(z - y_i)^2 + 1 \right]$$

$$\Leftrightarrow c = 1 - \frac{1}{(z - y_i)^2 + 1}$$

$$\bullet r'_1(c, y_i) > 0 \Leftrightarrow c > 1 - \frac{1}{(z - y_i)^2 + 1}$$

By continuity, we infer the following notations for the risk:

$$\forall y_i \in \mathbb{R},$$



Now the worst case risk is equal to

$$\max_{y_i \in \mathbb{R}} \{ r(c, y_i) \}$$

Recall that:

$$r(c, y_i) = (z - y_i - cz)^2 + (z - y_i - cz) 2c y_i + c^2 (1 + y_i^2)$$

Hence:

$$\begin{aligned}
r'_2(c, y_i) &= -2(z - y_i - cz) + 2cz - 2c^2z - 4cy_i + 2c^2y_i \\
&= -2z + 2cz + 2cz - 2c^2z + 2y_i - 4cy_i + 2c^2y_i \\
&= -2z \left[1 - 2c + c^2 \right] + 2y_i \left[1 - 2c + c^2 \right]
\end{aligned}$$

Note that:

- If $c \in \mathbb{R} \setminus \{1\}$:

$$\bullet r'_2(c, y_i) = 0 \Rightarrow 2y_i \left[1 - 2c + c^2 \right] = 2z \left[1 - 2c + c^2 \right]$$

$$\Leftrightarrow y_i = z$$

$$\bullet r'_2(c, y_i) > 0 \Leftrightarrow y_i > z$$

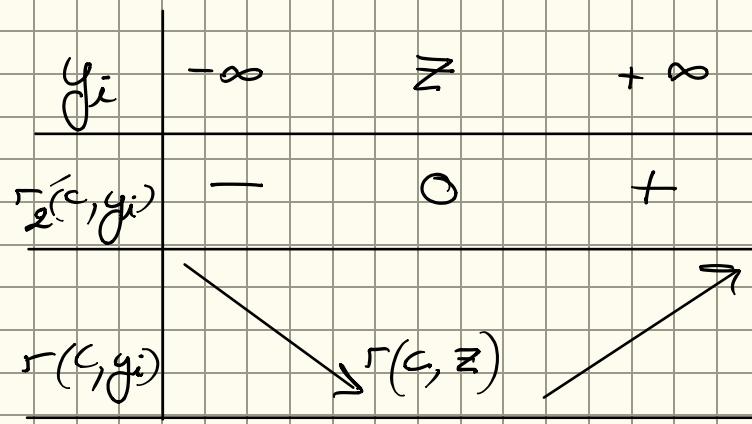
- If $c = 1$:

$$\bullet \forall y_i \in \mathbb{R}, r'_2(c, y_i) = 0$$

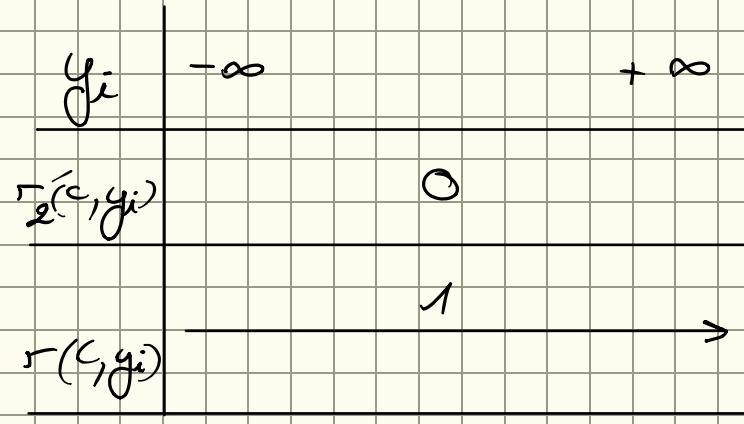
$$\begin{aligned}
\bullet r(c, y_i) &= (z - y_i - z)^2 + (z - y_i - z)2y_i + 1 + y_i^2 \\
&= (z - y_i - z + y_i)^2 + 1 \\
&= 1
\end{aligned}$$

By continuity, we infer the following notations for the risk:

$\forall c \in \mathbb{R} \setminus \{1\}$:



If $c = 1$:



Therefore, the worst case risk is:

$$\max_{y_i \in \mathbb{R}} \left\{ r(c, y_i) \right\} = \begin{cases} +\infty \text{ (unbounded)} & \text{if } c \in \mathbb{R} \setminus \{1\} \\ 1 & \text{if } c = 1. \end{cases}$$

Therefore the value of c that minimizes the worst-case risk is:

$$\underset{c \in \mathbb{R}}{\operatorname{argmin}} \left[\max_{y_i \in \mathbb{R}} \left\{ r(c, y_i) \right\} \right] = 1.$$

(e) Call π the prior distribution for y_i over \mathbb{R} , such that $y_i \sim N(0, \gamma^2)$.

The Bayes risk is:

$$R_B(\pi, c) = \int_{\mathbb{R}} R(c, y_i) d\pi(y_i)$$

$$= E_{y_i} \left[(z - y_i - cz)^2 + (z - y_i - cz) 2c y_i + c^2 (1 + y_i^2) \mid c \right]$$

$$= E_{y_i} \left[(z - cz)^2 + y_i^2 - 2(z - cz)y_i + 2cz y_i - 2c^2 z y_i + c^2 + c^2 y_i^2 \mid c \right]$$

$$= (z - cz)^2 + E_{y_i} (y_i^2 \mid c) - 2(z - cz) E_{y_i} (y_i \mid c) + 2cz E_{y_i} (y_i \mid c)$$

$$- 2c E_{y_i} (y_i^2 \mid c) - 2c^2 z E_{y_i} (y_i \mid c) + c^2 + c^2 E_{y_i} (y_i^2 \mid c)$$

$$= (z - cz)^2 + \gamma^2 - 2c \gamma^2 + c^2 + c^2 \gamma^2$$

$$= (z - cz)^2 + c^2 + \gamma^2 [1 - 2c + c^2]$$

i.e. $R_B(\pi, c) = z^2 (1 - c)^2 + c^2 + \gamma^2 [1 - 2c + c^2]$

We can now find c that minimizes the Bayes risk. The Bayes risk is a polynomial of degree two in c with a positive coefficient for its second-degree monomial. Therefore its minimum is reached where its first order derivative in c is null and this point is unique. Therefore:

$$c_{\text{Bayes}} \equiv \underset{c \in \mathbb{R}}{\operatorname{argmin}} \left\{ r_B(\pi, c) \right\}$$

$$= \left\{ c \in \mathbb{R} : r'_B(\pi, c) = 0 \right\}$$

$$= \left\{ c \in \mathbb{R} : -2z^2(1-c) + 2c - 2\gamma^2 + 2\gamma^2 c = 0 \right\}$$

$$= \left\{ c \in \mathbb{R} : -z^2 - \gamma^2 + z^2 c + c + \gamma^2 c = 0 \right\}$$

$$= \left\{ c \in \mathbb{R} : c [1 + z^2 + \gamma^2] = z^2 + \gamma^2 \right\}$$

$$\equiv \frac{z^2 + \gamma^2}{1 + z^2 + \gamma^2}$$

$$= 1 - \frac{1}{1 + z^2 + \gamma^2}$$

Therefore Box's rule corresponds to

$$c_{\text{Boxe}s} = 1 - \frac{1}{1 + z^2 + \delta^2}.$$

Now we switch back to the original setting -

(f) Recall question (b). We had found:

$$R(z^{(c)}, y) \\ = \sum_{i=1}^n r(c, y_i)$$

$$\begin{aligned} &\text{ie } R(z^{(c)}, y) \\ &= \sum_{i=1}^n \left[(z - y_i - cz)^2 + (z - y_i - cz) 2c y_i + c^2 (1 + y_i^2) \right] \end{aligned}$$

(g) Recalling the computations from question (b) again, we immediately see that we have:

$$\frac{\partial R}{\partial c}(z^{(c)}, y) = 0 \Leftrightarrow \sum_{i=1}^n \left[2(z - y_i)^2 [c^* - 1] + 2c^* \right] = 0$$

$$\Leftrightarrow [c^* - 1] \sum_{i=1}^n (z - y_i)^2 + nc^* = 0$$

$$\Leftrightarrow c^* \left[n + \sum_{i=1}^n (z - y_i)^2 \right] = \sum_{i=1}^n (z - y_i)^2$$

$$\Leftrightarrow c^* = 1 - \frac{n}{\sum_{i=1}^n (z - y_i)^2 + n}$$

Therefore the total risk across all individuals is minimized for :

$$c^* = 1 - \frac{n}{\sum_{i=1}^n (z - y_i)^2 + n} \neq 1.$$

Therefore, minimizing the total risk across all individuals is not compatible with minimizing the individual worst case risk (unless all people already earn the target income level, a case for which the whole programme is pointless).

Recall question (d) where we showed that minimizing

the individual risk yielded an optimum:

$$c_{\text{individual}}^* = 1 - \frac{1}{(z-y_i)^2 + 1}$$

Therefore, minimizing total risk and individual risk are not necessarily compatible.

Note that we have:

$$\begin{aligned} c^* &= 1 - \frac{n}{\sum_{i=1}^n (z-y_i)^2 + n} \\ &= 1 - \frac{1}{\frac{\sum_{i=1}^n (z-y_i)^2}{n} + 1} \end{aligned}$$

Therefore, minimizing total risk minimizes individual risk only for those whose squared difference with the target income is equal to the average one in the whole population. There may be no individual for whom this is the case.

(h) I propose the following estimator :

$$\hat{c}^* = 1 - \frac{1}{\frac{\sum_{i=1}^n (z - \hat{y}_i)^2}{n} + 1}$$

which gives :

$$\tau_i(\hat{c}^*)(\hat{y}) = \left[1 - \frac{1}{\frac{\sum_{i=1}^n (z - \hat{y}_i)^2}{n} + 1} \right] (z - \hat{y}_i)$$

Since $\hat{y}_i \sim N(y_i, 1)$, $\hat{y}_{i_{SS}} = \left[1 - \frac{n-2}{\sum_{i=1}^n \hat{y}_i^2} \right] \hat{y}_i$

Then :

$$\tau_i(\hat{c}^*)(\hat{y}_{i_{SS}}) = \left[1 - \frac{1}{\frac{\sum_{i=1}^n (z - \hat{y}_{i_{SS}})^2}{n} + 1} \right] (z - \hat{y}_{i_{SS}})$$

$$\text{ie } \bar{x}_i \hat{(c^*)} (\hat{y}_{\text{IS}})$$

$$= \left[1 - \frac{1}{\sum_{i=1}^n \left[1 - \left[1 - \frac{n-2}{\sum_{i=1}^n y_i} \right] \hat{y}_i \right]^2 + 1} \right]$$

$$= \left[1 - \frac{n-2}{\sum_{i=1}^n y_i} \right] \hat{y}_i$$

I'm not sure I fully understand what you meant here but it seems that the two estimate don't coincide. I don't understand what value you want me to estimate with
 James-Stein.

References

- [1] Amanda Agan, Jennifer L Doleac, Anna Harvey, Misdemeanor Prosecution, *The Quarterly Journal of Economics*, Volume 138, Issue 3, August 2023, Pages 1453–1505
- [2] Frandsen, Brigham, Lars Lefgren, and Emily Leslie. 2023. "Judging Judge Fixed Effects." *American Economic Review*, 113 (1): 253–77.
- [3] Braghieri, Luca, Ro'ee Levy, and Alexey Makarin. 2022. "Social Media and Mental Health." *American Economic Review*, 112 (11): 3660–93.
- [4] de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110 (9): 2964–96.