

Optimal Screening in Experiments with Partial Compliance*

Christopher Carter[†] Adeline Delavande[‡] Mario Fiorini[§]
Peter Siminski[¶] Patrick Vu^{||}

December 9, 2025

Abstract

This note studies optimal experimental design under partial compliance when experimenters can screen participants prior to randomization. Theoretical results show that retaining all compliers and screening out all non-compliers achieves three complementary aims: (i) the Local Average Treatment Effect is the same as the standard 2SLS estimator with no screening; (ii) median bias is minimized; and (iii) statistical power is maximized. In practice, complier status is unobserved. We therefore discuss feasible screening strategies and propose a simple test for screening efficacy. Future work will conduct an experiment to demonstrate the feasibility and advantages of the optimal screening design.

Keywords: Instrumental variables, Compliance, Experimental Design, Median Bias, Statistical Power

*We thank Peter Hull and seminar participants at the 2025 STEP UP Conference at UNSW for helpful comments.

[†]Department of Economics, University of Technology Sydney. Email: christopher.carter@uts.edu.au

[‡]Department of Economics, University of Technology Sydney. Email: adeline.delavande@uts.edu.au

[§]Department of Economics, University of Technology Sydney. Email: mario.fiorini@uts.edu.au

[¶]Department of Economics, University of Technology Sydney. Email: peter.siminski@uts.edu.au

^{||}Department of Economics, University of New South Wales. Email: patrick_vu@brown.edu

1. INTRODUCTION

In experiments with partial compliance, low take-up rates can make instrument variables ('IV') estimates of the average treatment effect on the treated both biased and imprecise. In response, recent econometrics research has proposed alternative methods aimed at improving precision ([Spiess and Coussens, 2021](#); [Hazard and Löwe, 2025](#)). Such efforts are based on the insight that placing greater weight on complier populations can enhance efficiency. For example, [Spiess and Coussens \(2021\)](#) propose an alternative estimator that weighs each observation according to its estimated compliance; and [Hazard and Löwe \(2025\)](#) exploits covariate information to restrict estimation to a subpopulation with non-zero compliance.

This note studies a related, but distinct, problem. It considers optimal experimental design under partial compliance when experimenters can screen participants prior to randomization. This differs from previous studies, which propose alternative estimators that can be applied to existing IV studies ([Spiess and Coussens, 2021](#); [Hazard and Löwe, 2025](#)). By contrast, this note addresses experimental design.

Our main aim is to establish the conditions under which a screened 2SLS estimator dominates the standard 2SLS estimator in terms of both bias and power. We study this question in the standard setup where both the instrument and treatment are binary. To begin, we consider an idealized setting where the experimenter observes participant types (i.e. compliers, always-takers, and never-takers). In later sections, we discuss how types can be elicited in practice.

A first-order concern is that screening can change the target parameter. This can render comparisons between screened and unscreened estimators ambiguous. Thus, to begin, we establish a sufficient condition which guarantees that the screened 2SLS estimator identifies the same Local Average Treatment Effect ('LATE') as the unscreened 2SLS estimator. Specifically, the sufficient condition states that any screening mechanism which retains all compliers identifies the same LATE. The intuition is simple: because the LATE only identifies the treatment effect for compliers, any screening mechanism retaining all compliers must preserve the target parameter.

Under this sufficient condition, we then show that the optimal screening mechanism is to retain all compliers and screen out all non-compliers (i.e. always-takers and never-takers). This screening mechanism is 'optimal' in the sense that it minimizes bias and maximizes statistical power among the set of screening mechanisms

that retain all compliers. Given that this set also includes no screening, the optimal screened 2SLS estimator therefore dominates the standard, unscreened 2SLS estimator in terms of both bias and power. Intuitively, the 2SLS estimator only identifies the average treatment effect among compliers, so that non-compliers only contribute noise to the estimator. In effect, removing non-compliers therefore ‘strengthens’ the instrument; and this more than offsets decreases in precision due to a smaller sample size. It is noteworthy that the optimal screened estimator does not exhibit a bias-variance trade-off, which can be present in alternative approaches (Spiess and Coussens, 2021; Hazard and Löwe, 2025), as both bias and efficiency improve.

The theoretical results rest on the assumption that the experimenter observes participants’ types, which is not true in practice. Thus, to practically implement the screening estimator, the experimenter must be able to accurately elicit types. To do this, we suggest that experimenters simply ask participants two questions, with the proviso that their answers will have no bearing on whether they will be offered treatment: (i) “*Would you take treatment ($D_i = 1$) if offered ($Z_i = 1$)?*” and (ii) “*Would you take treatment ($D_i = 1$) if not offered ($Z_i = 0$)?*”?¹

If reported compliance matches actual compliance, then types are accurately identified and the optimal screening estimator is feasible. To test whether this is the case, we propose an approach based on results in Borusyak and Hull (2024), which allow one to estimate the probability that stated complier status equals one conditional on actually being a complier.²

This provides a test of the identification assumption that all compliers are retained under screening, and has implications for which estimator should be reported. For example, if the null hypothesis that $\mathbb{P}[\widetilde{CM}_i = 1 | CM_i = 1] = 1$ is rejected, then researchers should use the standard, unscreened estimator because the screened estimator identifies a different LATE. Alternatively, failing to reject the null provides support for the use of the screened estimator, which will identify the same LATE and have better statistical properties as compared with the unscreened estimator in terms of both power and bias.

¹In an RCT with one-sided non-compliance, it is only necessary to ask participants question (i) because treatment is not available to those not offered.

²More formally, let D_i be binary treatment, CM_i be a binary variable for *true* complier type, and \widetilde{CM}_i be a binary variable that equals one for reported compliers. Results in Borusyak and Hull (2024) show that an IV regression (with the full sample) of $\widetilde{CM}_i \times D_i$ on D_i estimates $\mathbb{P}[\widetilde{CM}_i = 1 | CM_i = 1]$.

Finally, it is important to add that we recommend that experimenters actually implement ‘pseudo-screening’, where they ask the screening questions, but retain the full sample for the experiment. This allows them to calculate both the screened and unscreened estimator, estimate the power gain from the screened design, and also perform stronger tests of the identification assumption, as discussed in more detail in the main text.

This note proceeds as follows. Section 2 sets up the problem and presents the main theoretical results. Section 3 shows how to feasibly implement the screened estimator and test for the reliability of stated compliance. Finally, Section 4 concludes with a brief discussion of future work.

2. THEORY

In this section, we describe the set up and present the main theoretical results. Proofs are in Appendix A.

The main aim is to compare the standard *unscreened IV estimator* to an alternative approach where experimenters screen out certain participants and then run the experiment on the remaining subset. We call this the *screened IV estimator*.

The theoretical results are proved under an idealized setting in which participant types are observed by the experimenter (i.e. complier, always-taker, never-taker). We derive an ‘optimal’ screened estimator under this assumption. Since types are not observed in reality, we discuss how to implement this estimator in practice in Section 3.

2.1 Setup

Consider the standard linear IV model with heterogeneous treatment effects:

$$Y_i = \alpha + \beta_i D_i + u_i$$

$$D_i = \phi + \pi Z_i + \eta_i$$

where $\mathbb{E}[D_i u_i] \neq 0$ and we assume, for simplicity, that errors are homoscedastic. Let S_i be a binary variable that equals one if i is screened into the sample, and zero otherwise. We impose the standard assumptions under the LATE framework (Imbens and Angrist, 1994):

Assumption 1 (LATE Identifying Assumptions). *Assume that*

1. *Exclusion restriction:* $Y_i(d, 0) = Y_i(d, 1) = Y_i(d)$ for $d = 0, 1$
2. *Independence:* $(Y_i(D_i(1), 1), Y_i(D_i(0), 0), D_i(1), D_i(0), S_i) \perp Z_i$
3. *First Stage:* $E[D_i(1) - D_i(0)] \neq 0$
4. *Monotonicity:* $D_i(1) - D_i(0) \geq 0$ for all i (or vice versa).

Compared to the standard LATE assumptions, we include one additional assumption, namely, that the instrument is assumed to be independent of screening. This is immediately satisfied in the scenario we consider, where screening occurs prior to randomization.

The main aim is to compare median bias and power of the *screened IV estimator*, $\hat{\beta}_r$, and the *unscreened IV estimator*, $\hat{\beta}_1$, where the subscript denotes the fraction of the original sample that is retained. To make meaningful comparisons, we assume throughout that screening is neither trivial nor degenerate, in that at least one unit is screened out of the experiment and not all are screened out: $r \in (0, 1)$.

A fundamental concern for alternative IV estimators is whether or not they change the target parameter. For example, Spiess and Coussens (2021) and Hazard and Löwe (2025) propose alternative estimators that can provide gains in precision, but, in general, target a different LATE to standard IV estimation. This makes direct comparisons challenging. To avoid this issue, we establish a sufficient condition under which screening does not change the target parameter:

Assumption 2 (Complier Retention). $\mathbb{P}[S_i = 1 | CM_i = 1] = 1$

Under this sufficient condition, the target parameter is unchanged:

Lemma 1. *Under Assumptions 1 and 2,*

$$\frac{\mathbb{E}[Y_i | Z_i = 1, S_i = 1] - \mathbb{E}[Y_i | Z_i = 0, S_i = 1]}{\mathbb{E}[D_i | Z_i = 1, S_i = 1] - \mathbb{E}[D_i | Z_i = 0, S_i = 1]} = \mathbb{E}[Y_i(1) - Y_i(0) | CM_i = 1]$$

In words, this result says that any screening procedure which retains all compliers will target the same LATE as the unscreened estimator (i.e. the average treatment effect on compliers). We maintain this assumption throughout the paper and propose a simple approach to test it in Section 3.

2.2 Optimal Screening Mechanism

This subsection considers the optimal screening mechanism conditional on retaining all compliers (Assumption 2). The mechanism we propose is optimal in two senses: first, that it maximizes statistical power; and second, that it minimizes median bias. This differs from alternative IV estimators in the literature, which typically pose a bias-variance trade-off (Spiess and Coussens, 2021; Hazard and Löwe, 2025).

First, consider statistical power, for which we make the following assumption:

Assumption 3 (Homoscedasticity). $\mathbb{E}[u_i^2] = \sigma_u^2$

The variance of the 2SLS estimator depends on the average residual variance in the estimation sample; under homoscedasticity, this average is unaffected by screening, ensuring that any reduction in the standard error comes solely from the strengthened first stage rather than from changes in error variance.³ With this, we can state the main result:

Proposition 1. *Under Assumptions 1–3, $se(\hat{\beta}_r) = \sqrt{r} \cdot se(\hat{\beta}_1)$ and $se(\hat{\beta}_r)$ is minimized when all non-compliers are screened out.*

This result shows that any (non-trivial) screening mechanism that retains all compliers lowers the standard error, which corresponds to a lower MDE and higher statistical power.⁴ Moreover, the power gain is inversely related to the fraction of the full sample that is retained, r . This implies that the power-maximizing screening mechanism minimizes r , which is achieved by screening out all never-takers and always-takers and retaining only compliers.

The intuition for Proposition 1 is that LATE identifies the effect for compliers only, so removing non-compliers reduces noise and thus increases power by ‘strengthening the instrument’. This general insight is well understood in the existing IV literature (Spiess and Coussens, 2021; Hull and Borusyak, 2025; Hazard and Löwe, 2025). The primary contribution of Proposition 1 relative to existing work is therefore conceptual, in that it shows that this insight can be used to optimize *experimental design*; this

³This can be weakened to allow for heteroscedasticity across units, provided that the average variance is equal across screened and unscreened samples. An even weaker requirement is that the screened sample have weakly lower average residual variance, $\sigma_{u,r}^2 \leq \sigma_{u,1}^2$, which also guarantees improved precision.

⁴Statistical power is increasing in the ratio of the true treatment effect and the standard error. Since the true treatment effect is unchanged under Assumption 2, a lower standard error immediately implies higher statistical power.

differs from the complementary but distinct aim of improving precision in *existing* IV studies.

To gauge the potential practical importance of this insight, consider Tarozzi et al. (2015), which studies the impact of microcredit in Ethiopia using an RCT with partial compliance. The first-stage shows that assignment to access to micro-credit increased borrowing rates by 25 percentage points; note that this also provides an estimate of the share of compliers in the population.⁵

Now suppose experimenters effectively screen out all non-compliers from the estimation sample. Then Lemma 1 implies that the target parameter is unchanged. Moreover, under the assumptions of Proposition 1, the standard error for the screened estimator will be halved relative to the standard, unscreened estimator ($\sqrt{0.25} = 0.5$). Thus, confidence intervals and the MDE would also be halved, which represents an extremely large gain in precision for zero cost.

Next, we turn to bias, where we make the following assumptions:

Assumption 4 (Independent Normal Errors). $u_i \stackrel{iid}{\sim} N(0, \sigma_u^2)$ and $\eta_i \stackrel{iid}{\sim} N(0, \sigma_v^2)$ and $u_i \perp \eta_i$.

Assumption 5 (Sign Screening on Estimated First Stage). *Experimenters screen for IV estimates where the sign of the estimated first-stage matches the sign of the population first-stage.*

Assumption 4 imposes normality of errors for analytical convenience. Following Angrist and Kolesár (2024), Assumption 5 assumes that experimenters screen the estimated first-stage to have the same sign as the population first-stage. In practice, this is a relatively mild assumption.

To illustrate, consider again the example of a microcredit intervention. The population first-stage is generally thought to be positive ($\pi > 0$), since those who are offered microcredit ($Z_i = 1$) should have (weakly) higher take-up rates than those not offered ($Z_i = 0$). Screening the estimated first-stage in this case means that if the estimated take-up rate were *lower* in the group offered microcredit, then the experimenter would discard this quantitative results from the experiment. This is likely to hold in practice because such an outcome would likely invalidate the study since it suggests an extremely weak instrument.

⁵Formally, Assumption 1 implies: $\pi = \mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0] = \mathbb{E}[D_i(1) - D_i(0)] = \mathbb{P}[CM_i = 1]$.

With this, we can state the result:

Proposition 2. *Under Assumptions 1, 2, 4 and 5, the median of the magnitude of bias of the screened 2SLS estimator is strictly lower than that of the unscreened 2SLS estimator and minimised when all non-compliers are screened out.*

This results state that any screening mechanism which retains all compliers lowers median bias – and also that median bias will be lowest when all never-takers and always-takers are excluded. The intuition is similar to Proposition 1: since we only ever identify the impact on compliers, removing always-takers and never-takers effectively increases the strength of the instrument, which in turn lowers bias.

To summarize, this section makes a number of complementary claims. First, that a sufficient condition for maintaining the same target parameter in the screened and unscreened designs is to retain all compliers. Second, that conditional on satisfying this sufficient condition, screening out all always-takers and never-takers will minimize bias and maximize statistical power. Next, we turn to feasible implementation of this screening design.

3. FEASIBLE IMPLEMENTATION

In the preceding section, the theoretical results were derived assuming the experimenter can directly observe participants' types (e.g. complier, always-taker, never-taker) and screen accordingly. In reality, types are unobservable. This section discusses how to implement the optimal screening estimator in practice, and proposes an approach to partially test whether screening effectively identifies types.

3.1 Screening Questions

We recommend experimenters screen for compliers by simply asking them their intended compliance. Thus, for RCTs with two-sided noncompliance, we propose adding two questions to the questionnaire prior to randomisation:

1. Would you take treatment ($D_i = 1$) if offered ($Z_i = 1$)?
2. Would you take treatment ($D_i = 1$) if not offered ($Z_i = 0$)?

Responses to these two questions identify participants' *stated type*. For example, if a participant respond no to (i) and yes to (ii), then their *stated type* is a complier.

The effectiveness of screening depends on whether participants' *stated type* match their *true type*, which we discuss further in the next subsection.

Before that, it is useful to introduce some notation. Let $\widetilde{CM}_i = 1$ if i is a *stated complier* based on the two questions above, and zero otherwise; and let $CM_i = 1$ if i is a *true complier* and zero otherwise. Additionally, we say that *Type-I screening error* occurs when an individual is a stated complier ($\widetilde{CM}_i = 1$) but not a true complier ($CM_i = 0$). Alternatively, *Type-II screening error* occurs when an individual is a stated non-compliers ($\widetilde{CM}_i = 0$) but is in fact a true complier ($CM_i = 1$).

Identification requires that all compliers are retained; in other words, that there is no Type-II error. That, under screening for stated compliers ($S_i = 1 \iff \widetilde{CM}_i = 1$), Assumption 2 becomes $\mathbb{P}[\widetilde{CM}_i = 1 | CM_i = 1] = 1$, which we discuss how to estimate in the next subsection. Note that Type-I error does not, by contrast, harm identification, although it does reduce the power gain because non-compliers included in the sample will increase the variance of the estimator.

Finally, note that in RCTs with one-sided noncompliance only question (1) needs to be added to the questionnaire; this is because the answer to (2) is ‘no’ by design for all participants.

3.2 Pseudo-Screening and Testing Identification

We recommend that experimenters implement what we call *pseudo-screening*. This is where screening questions are asked to elicit stated types, but where the experiment nevertheless is conducted on the full sample. Pseudo-screening means that experimenters can calculate both the IV estimate and standard error for both the screened sample and the full sample, and allows experimenters to measure the power gains from the screened design.

More importantly, however, is that pseudo-screening provides experimenters with greater variation for testing the identification assumption, namely, that screening successfully retains all compliers, which ensure the same LATE as the unscreened estimator (Assumption 2). To see this, consider the case experimenters do not pseudo-screen but actually screened out stated non-compliers ($\widetilde{CM}_i = 0$). Suppose further that a subset of stated non-compliers are in fact compliers ($CM_i = 1$); this represents Type-II error, which violates the identification assumption.

However, under true screening, this type of violation is undetectable because non-compliers are excluded from the experiment, and their behaviour is therefore never

observed. By contrast, under pseudo-screening, a subset of stated non-compliers will be offered treatment, and their choices can be observed to test whether stated non-compliance matches non-compliance in reality.

To see how this can be formally tested, consider first the more general result in Proposition 1 in [Borusyak and Hull \(2024\)](#). This result can be used to show that for any characteristic X_i , the 2SLS coefficient from an IV regression of $X_i \times D_i$ on D_i will estimate the average of that characteristic in the complier subpopulation i.e. $\mathbb{E}[X_i|CM_i = 1]$.

Returning to our setting, let \widetilde{CM}_i be a binary variable that equals one if i reports the intention to comply, and zero otherwise. From the general result above, it follows that an IV regression of $\widetilde{CM}_i \times D_i$ on D_i will estimate the probability that stated compliance equals one conditional on actually being a complier i.e. $\mathbb{P}[\widetilde{CM}_i = 1|CM_i = 1]$. This corresponds exactly the probability in Assumption 2 when the screening mechanism is to retain all stated compliers.

Thus, we propose a one-sided test of the null hypothesis that $\mathbb{P}[\widetilde{CM}_i = 1|CM_i = 1] = 1$, where the alternative hypothesis is $\mathbb{P}[\widetilde{CM}_i = 1|CM_i = 1] < 1$.⁶ Rejecting the null hypothesis suggests that stated compliance is not reliable. This might justify using of the standard, unscreened estimator, which is, of course, calculable under pseudo-screening. On the other hand, failing to reject the null hypothesis would favor using the screened estimator, as this is guaranteed to have better statistical power (Proposition 1) and lower bias (Proposition 2).

3.3 Optimal Screening Estimator

Validating that $\mathbb{P}[\widetilde{CM}_i = 1|CM_i = 1] = 1$ tests for Type-II screening error; and hence it provides a direct test on whether the LATE is preserved and that the screen estimator has better bias and power properties than the unscreened estimator. However, it does not imply that there is no Type-I screening error, and hence whether the optimal screening design has been achieved. That is, some participants may state being compliers ($\widetilde{CM}_i = 1$) but may not in fact be compliers $CM_i = 0$. This outcome might arise, for example, due to Hawthorne effects. More formally, it is possible that $\mathbb{P}[\widetilde{CM}_i = 0|CM_i = 0] \neq 1$, even if $\mathbb{P}[\widetilde{CM}_i = 1|CM_i = 1] = 1$.

⁶An alternative test of the identification assumption would be to estimate whether treatment take-up is zero for the ‘pseudo-screened-out’ group (i.e. stated non-compliers). We opt for the test in the main text because it utilizes the full sample and is therefore more powerful.

Thus, to test whether the optimal screening design was obtained, the experimenter can additionally test the null hypothesis that $\mathbb{P}[\widetilde{CM}_i = 0 | CM_i = 0] = 1$. To calculate this quantity, note that the law of total probability implies the following:

$$\mathbb{P}[\widetilde{CM}_i = 0 | CM_i = 0] = \frac{\mathbb{P}[\widetilde{CM}_i = 0] - \mathbb{P}(CM_i = 1)\mathbb{P}[\widetilde{CM}_i = 0 | CM_i = 1]}{1 - \mathbb{P}[CM_i = 1]}$$

where all quantities on the right-hand side are directly estimatable.

Hence, if both nulls are not rejected, then this provides evidence that stated responses are reliable and that the optimal screening estimator is achieved.

4. CONCLUSION

This note examines optimal experimental design under partial compliance when experimenters can screen participants prior to randomization. It shows that screening for compliers maintains the same LATE as no screening but maximizes statistical power and minimizes bias. Back-of-the-envelope calculations based on actual RCTs suggest that power gains could be quite large e.g. halving confidence intervals.

However, the feasibility of the screened estimator depends on the reliability of eliciting compliance behaviour. Thus, we are pursuing ongoing work to test the feasibility and potential advantages of the optimal screening design over a standard compliance design.

REFERENCES

- Angrist, Joshua D. and Michal Kolesár**, “One Instrument to Rule Them All: The Bias and Coverage of Just-ID IV,” *Journal of Econometrics*, March 2024, 240 (2).
- Borusyak, Kirill and Peter Hull**, “Negative Weights Are No Concern in Design-Based Specifications,” *AEA Papers and Proceedings*, 2024, 114, 597–600.
- Hazard, Yagan and Simon Löwe**, “Improving Late Estimation in Experiments with Imperfect Compliance,” *Working Paper*, 2025.
- Hull, Peter and Kirill Borusyak**, “Optimal Formula Instruments,” *Working Paper*, 2025.
- Imbens, Guido W. and Joshua D. Angrist**, “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 1994, 62 (2), 467–475.
- Spiess, Jann and Stephen Coussens**, “Improving Inference from Simple Instruments through Compliance Estimation,” *Working Paper*, 2021.
- Tarozzi, Alessandro, Jaikishan Desai, and Kristin Johnson**, “The Impacts of Microcredit: Evidence from Ethiopia,” *American Economic Journal: Applied Economics*, 2015, 7 (1), 54–89.

A. PROOFS

Proof of Lemma 1: Observe that

$$\begin{aligned} \frac{\mathbb{E}[Y_i|Z_i = 1, S_i = 1] - \mathbb{E}[Y_i|Z_i = 0, S_i = 1]}{\mathbb{E}[D_i|Z_i = 1, S_i = 1] - \mathbb{E}[D_i|Z_i = 0, S_i = 1]} &= \mathbb{E}[Y_i(1) - Y_i(0)|CM_i = 1, S_i = 1] \\ &= \mathbb{E}[Y_i(1) - Y_i(0)|CM_i = 1] \end{aligned}$$

The first equality follows from the arguments in the standard LATE theorem with Assumption 1 conditional on $S_i = 1$ (Imbens and Angrist, 1994). The second equality follows immediately from Assumption 2 \square

Proof of Proposition 1: The ratio of the standard error under the screened design relative to the standard error under the standard design is, under homoscedasticity,

$$\frac{se(\hat{\beta}_r)}{se(\hat{\beta}_1)} = \sqrt{\frac{\frac{\sigma_{u,r}^2}{N_S \cdot \pi_S^2 \widehat{Var}(Z)}}{\frac{\sigma_{u,1}^2}{N \cdot \pi^2 \widehat{Var}(Z)}}} = \sqrt{\frac{N}{N_S} \cdot \frac{\pi}{\pi_S}} \quad (1)$$

where the final inequality uses $\sigma_{u,S}^2 = \sigma_u^2$ (Assumption 3).

Next, note that Assumption 1 implies $\pi = \mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0] = \mathbb{E}[D_i(1) - D_i(0)] = \mathbb{P}[CM_i = 1]$. By a similar argument, we have

$$\begin{aligned} \pi_S &= \mathbb{P}[CM_i = 1|S_i = 1] = \frac{\mathbb{P}[S_i = 1|CM_i = 1]\mathbb{P}[CM_i = 1]}{\mathbb{P}[S_i = 1]} \\ &= \frac{\mathbb{P}[CM_i = 1]}{r} \end{aligned}$$

where the first equality uses Bayes Rule. It follows that $\pi/\pi^S = r$. Finally, since $N_r = rN$ by definition, it we have that $\sqrt{\frac{N}{N_S}} = \frac{1}{\sqrt{r}}$. Substituting these expressions into equation (1) gives the desired result.

Proof of Proposition 2: Define for convenience $q \equiv \frac{1}{N} \sum_i Z_i$. We begin by deriving an expression for bias of the screened 2SLS estimator.

$$\hat{\beta}_r - \beta = \frac{\sum_{\{i|S_i=1\}} \tilde{Z}_i u_i}{\sum_{\{i|S_i=1\}} \tilde{Z}_i \tilde{D}_i}$$

Note that this expression also captures the unscreened estimator, $\hat{\beta}_1$, when $r = 1$.

Expanding the denominator gives

$$\begin{aligned}
\sum_{\{i|S_i=1\}} \tilde{Z}_i \tilde{D}_i &= \sum_{\{i|S_i=1\}} \tilde{Z}_i D_i - \bar{D} \sum_{\{i|S_i=1\}} \tilde{Z}_i \\
&= \sum_{\{i|S_i=1\}} \tilde{Z}_i D_i = \sum_{\{i|S_i=1\}} \tilde{Z}_i (\phi + \pi_S Z_i + \eta_i) \\
&= \pi_S \sum_{\{i|S_i=1\}} \tilde{Z}_i Z_i + \sum_{\{i|S_i=1\}} \tilde{Z}_i \eta_i \\
&= \pi_S N_S q (1 - q) + \sum_{\{i|S_i=1\}} \tilde{Z}_i \eta_i \\
&= \pi N q (1 - q) + \sum_{\{i|S_i=1\}} \tilde{Z}_i \eta_i
\end{aligned}$$

The second and fourth equalities use the fact that $\sum_{\{i|S_i=1\}} \tilde{Z}_i = \sum_{\{i|S_i=1\}} (Z_i - q) = \sum_{\{i|S_i=1\}} Z_i - rNq = 0$. The final line uses the following simplification.

$$\begin{aligned}
\pi_S &= \mathbb{E}[D_i | Z_i = 1, S_i = 1] - \mathbb{E}[D_i | Z_i = 0, S_i = 1] \\
&= \mathbb{E}[D_i(1) - D_i(0) | S_i = 1] \\
&= \mathbb{E}[CM_i = 1 | S_i = 1] \\
&= \mathbb{P}[CM_i = 1 | S_i = 1]] \\
&= \frac{\mathbb{P}[S_i = 1 | CM_i = 1] \mathbb{P}[CM_i = 1]}{\mathbb{P}[S_i = 1]} = \frac{\pi}{r}
\end{aligned}$$

where the second equality uses independence (Assumption 1); the third equality monotonicity (Assumption 1); the fifth equality Bayes Rule; and the final equality follows due to full complier retention (Assumption 2), the fact that $r \equiv \mathbb{P}[S_i = 1]$, and finally that $\pi = \mathbb{P}[CM_i = 1]$. Since $N_S = rN$, it follows that $\pi_S N_S = \pi N$.

Thus, we have

$$\hat{\beta}_r - \beta = \frac{M_{u,r}}{\mu + M_{\eta,r}}$$

where $\mu \equiv \pi N q (1 - q)$, $M_{\eta,r} \equiv \sum_{\{i|S_i=1\}} \tilde{Z}_i \eta_i$, and $M_{u,r} \equiv \sum_{\{i|S_i=1\}} \tilde{Z}_i u_i$.

Next, observe that $M_{u,r}$ and $M_{\eta,r}$ are linear combinations of independent Gaussian

draws, from which it follows that

$$\begin{aligned}\text{Var}(M_{u,r}) &= \text{Var} \left(\sum_{\{i|S_i=1\}} \tilde{Z}_i u_i \right) = \sum_{\{i|S_i=1\}} \text{Var} (\tilde{Z}_i u_i) \\ &= \sigma_u^2 \sum_{\{i|S_i=1\}} \tilde{Z}_i^2 = r \cdot (\sigma_u^2 N q (1 - q)) \\ &= r \cdot \text{Var}(M_{u,1})\end{aligned}$$

In words, the variance of $M_{u,r}$ is a scaled version of the corresponding unscreened random variable, $M_{u,1}$. By the same argument $\text{Var}(M_{\eta,r}) = r \cdot \text{Var}(M_{\eta,1})$. It follows that

$$\hat{\beta}_r - \beta = \frac{M_{u,r}}{\mu + M_{\eta,r}} \stackrel{d}{=} \frac{\sqrt{r} M_{u,1}}{\mu + \sqrt{r} M_{\eta,1}} \quad (2)$$

Let $F_{Bias|\eta}(\cdot)$ the cdf of bias conditional on a fixed value of $M_{\eta,r}$; and let $\Phi(\cdot)$ denote the normal cdf. Then for any $t > 0$

$$\begin{aligned}F_{Bias|r,\eta}(-t) &= \mathbb{P} \left(\frac{\sqrt{r} M_{u,1}}{\mu + \sqrt{r} \cdot m} < -t \mid M_{\eta,1} = m \right) \\ &= \mathbb{P} \left(M_{u,1} < -\frac{t\mu}{\sqrt{r}} - t \cdot m \mid M_{\eta,1} = m \right) \\ &= \Phi \left(-\frac{t\mu}{\sqrt{r}} - t \cdot m \right)\end{aligned}$$

where the second equality uses the fact that the denominator is positive, which occurs precisely when the estimated first-stage has the same sign as the population first stage (Assumption 5). The final expression is clearly increasing in r .

By similar reasoning, we have that $1 - F_{Bias|r,\eta}(-t) = 1 - \Phi \left(\frac{t\mu}{\sqrt{r}} + t \cdot m \right)$, which is also increasing in r .

Thus, the probability mass of the tails of the *conditional* bias distribution are increasing in r . Next, note that a straightforward application of the dominated convergence theorem implies that the probability mass of the tails of the *unconditional* bias distribution are also increasing in r .

Finally, let $F_{|Bias|,r}(\cdot)$ denote the cdf of the absolute value of bias. From the

preceding arguments, it follows that for any $0 < r_1 < r_2 < 1$, and $t > 0$, that

$$F_{|Bias|,r_2}(t) \leq F_{|Bias|,r_1}(t)$$

In other words, $F_{|Bias|,r_2}(\cdot)$ first-order stochastically dominates $F_{|Bias|,r_1}(\cdot)$, which immediately implies that the median (and in fact all other quantile) under $F_{|Bias|,r_2}(\cdot)$ is larger than under $F_{|Bias|,r_1}(\cdot)$, which is what we wanted to show.

From this, it is immediately clear that the bias minimizing choice for r is to screen out all never-takers and always-takers, such that $r = \mathbb{P}(CM_i = 1)$. \square