1. a.
```
cap program drop montecarlo // program define montecarlo, rclass // clear
// set obs 20 // gen e = rnormal(0, sqrt(3)) //
gen z = rnormal() // gen u = rnormal() // gen x = z + e + u //
gen y = x + e// reg x z // return scalar b1 = _b[z] // return scalar rss
= `e(rss)'// test _b[z] = 0 // return scalar F = `r(F)' //
ivreg y (x = z) // return scalar b2 = _b[x]
// gen xtilde = z // reg y xtilde // return scalar b3 = _b[xtilde]
// gen ess = (e+u)^2// sum ess // return scalar ess = `r(N)' * `r(mean)'
// drop e z u x y xtilde// end //simulate rss=r(rss) ess=r(ess) //
F=r(F) b1=r(b1) b2=r(b2) b3=r(b3), seed(42) reps(10000): montecarlo
*beta_2SLS with estimated first stage coefficient
sum b2, d // mean = .4375127
*beta_2SLS with true first stage coefficient
sum b3, d // mean = 1.008521
```

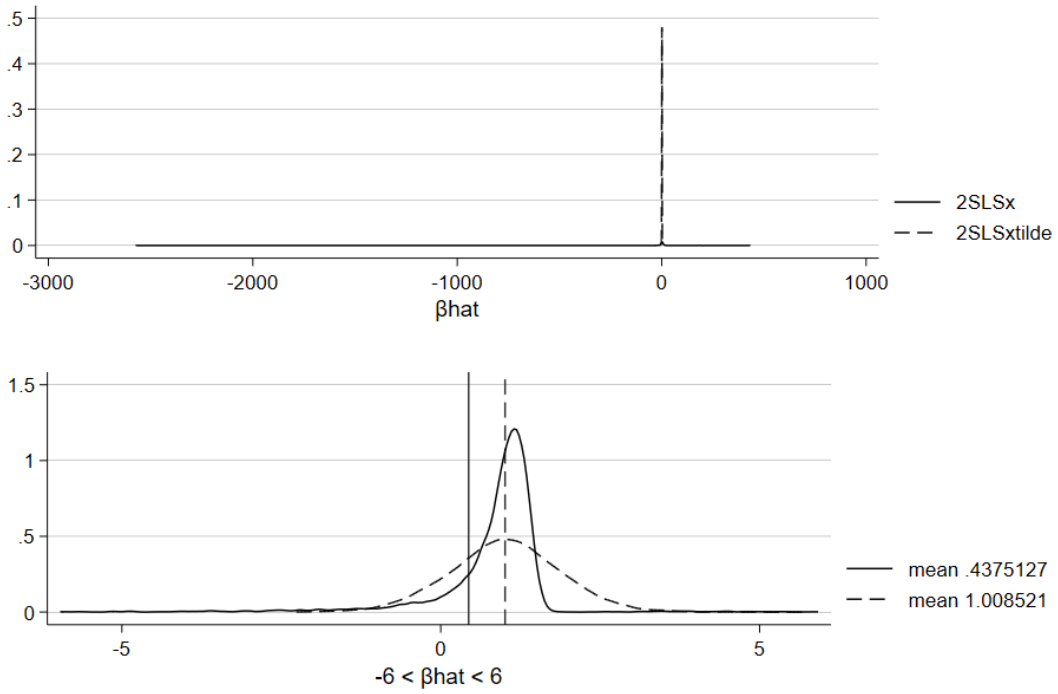

Figure 1: $\hat{\beta}_{2SLS}$ kernel density and mean

b.
```
*absolute second stage bias
gen absbias = abs(b2 - 1)
*in case that beta_FS estimate < beta_FS
sum absbias if b1 > 1 // on average .2243988
*in case that beta_FS estimate > beta_FS
sum absbias if b1 < 1 // on average 2.838942
```

Figure 2 demonstrates that when the first-stage F-statistic exceeds 10, the second-stage bias does not decrease; instead, it increases

[1]mach5689@student.su.se

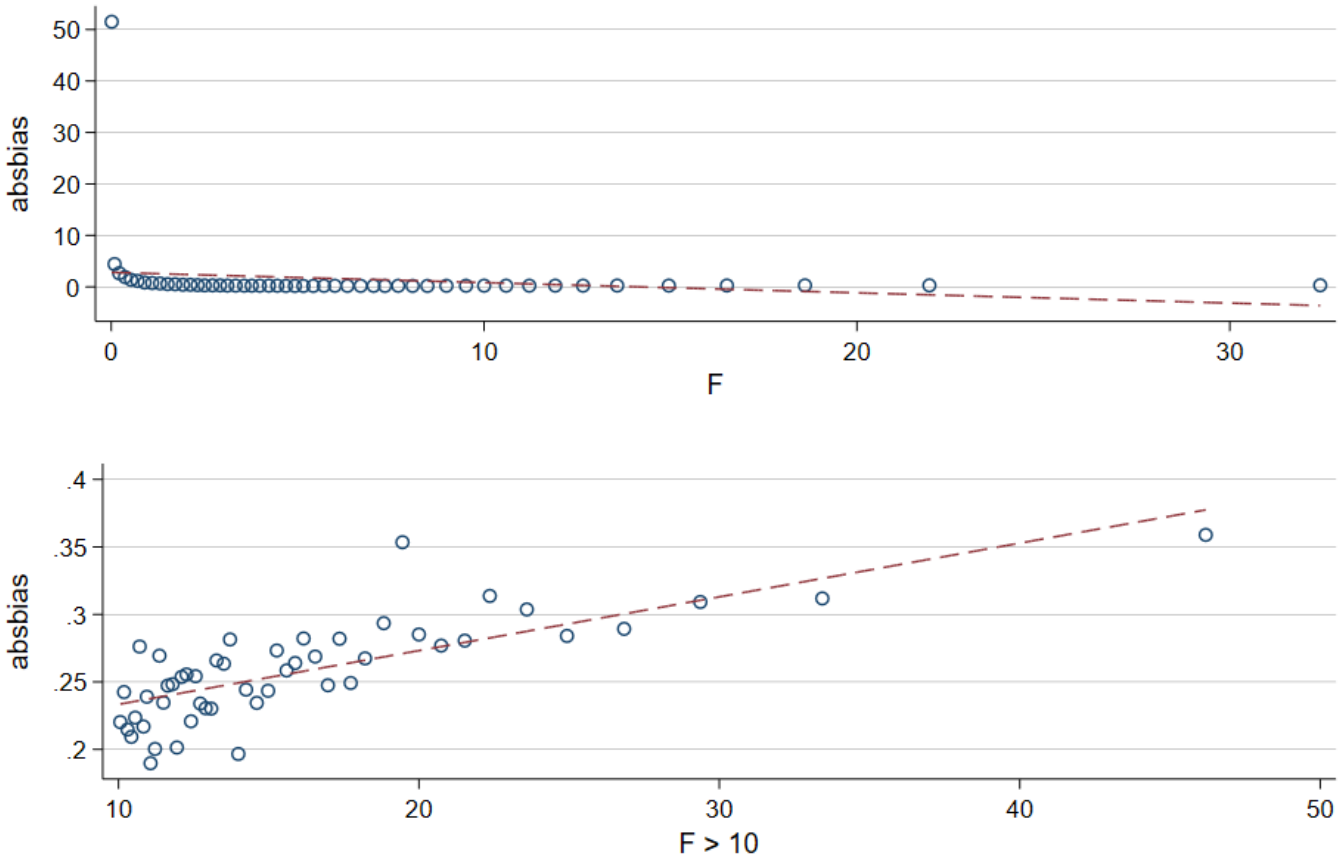

Figure 2: $|\hat{\beta}_{2SLS} - \beta_x|$ over F

c.
```
*ratio of residual sum of squares over error sum of squares
gen ssratio = rss/ess
sum ssratio // the highest across all simulations is .999998
```

d.
```
*the largest absolute second stage bias
gen absbiasdesc = -absbias
sort absbiasdesc
list b1 in 1/5 // for beta_FS = .000352
*the smallest absolute second stage bias
sort absbias
list b1 in 1/5 // for beta_FS = .7315149
```

Figure 4 illustrates that the first-stage F-statistic does not appear to be a useful diagnostic for detecting bias in cases where $\hat{\beta}_{FS} > 0.5$. This is because F linearly increases as a function of $\hat{\beta}_{FS}$, while $|\hat{\beta}_{2SLS} - \beta_x|$ is minimized at $\hat{\beta}_{FS} = \beta_{FS} = 1$.

e. Figure 1 shows that the bias in $\hat{\beta}_{2SLS}$ arises from estimating the first stage. Figure 3 highlights that the source of bias is overfitting: the RSS does not vary with $\hat{\beta}_{FS}$ and is consistently smaller than the ESS, which is minimized at $\hat{\beta}_{FS} = \beta_{FS} = 1$ so that the ratio between the two decreases as $\hat{\beta}_{FS}$ deviates from its true value. OLS mechanics adjust for both exogenous and endogenous variation, with the latter causing bias in the second stage by making the first-stage fitted values correlated with the error term. Further, a critical issue occurs a when the first stage is weak, as seen by the largest absolute bias occuring in cases when $\hat{\beta}_{FS}$ is near 0 (2SLS is a ratio where any slight bias in the reduced form is effectively multiplied by weak first stage). Figures 2 and 4, as noted above, highlight that small values of the sample F statistic are alarming for severe bias (weak first stage), while the same does not hold if their values are large enough (overfitting).

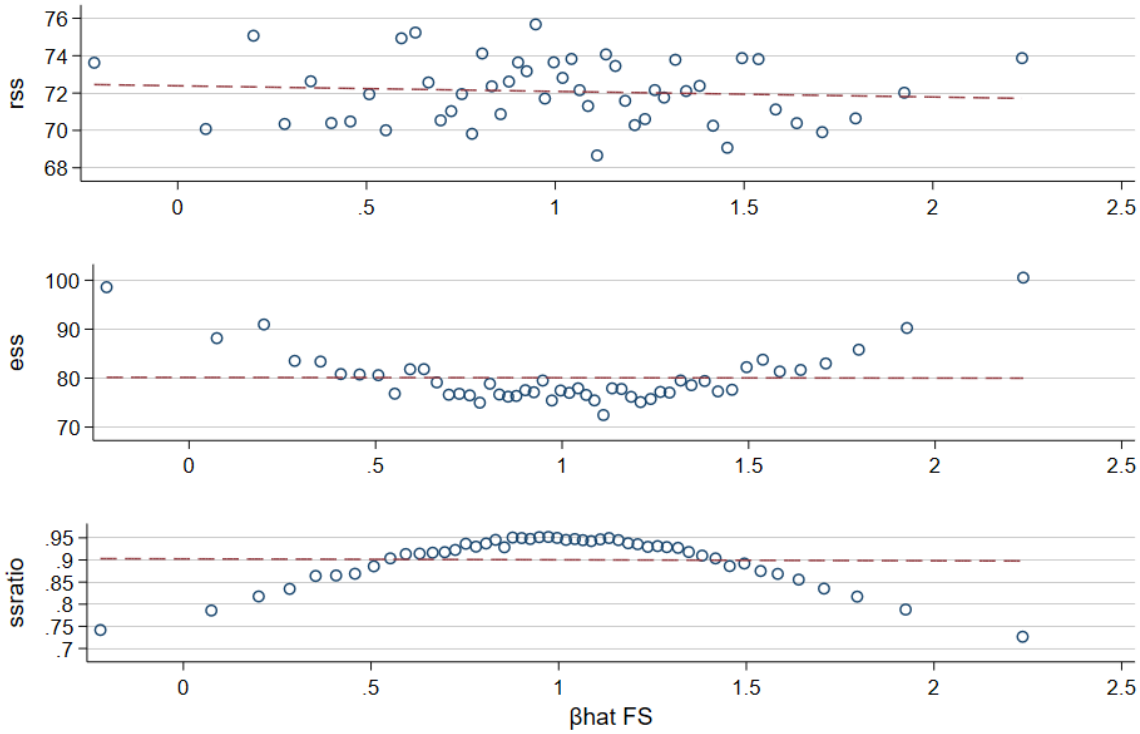

Figure 3: sums of squares over $\hat{\beta}_{FS}$

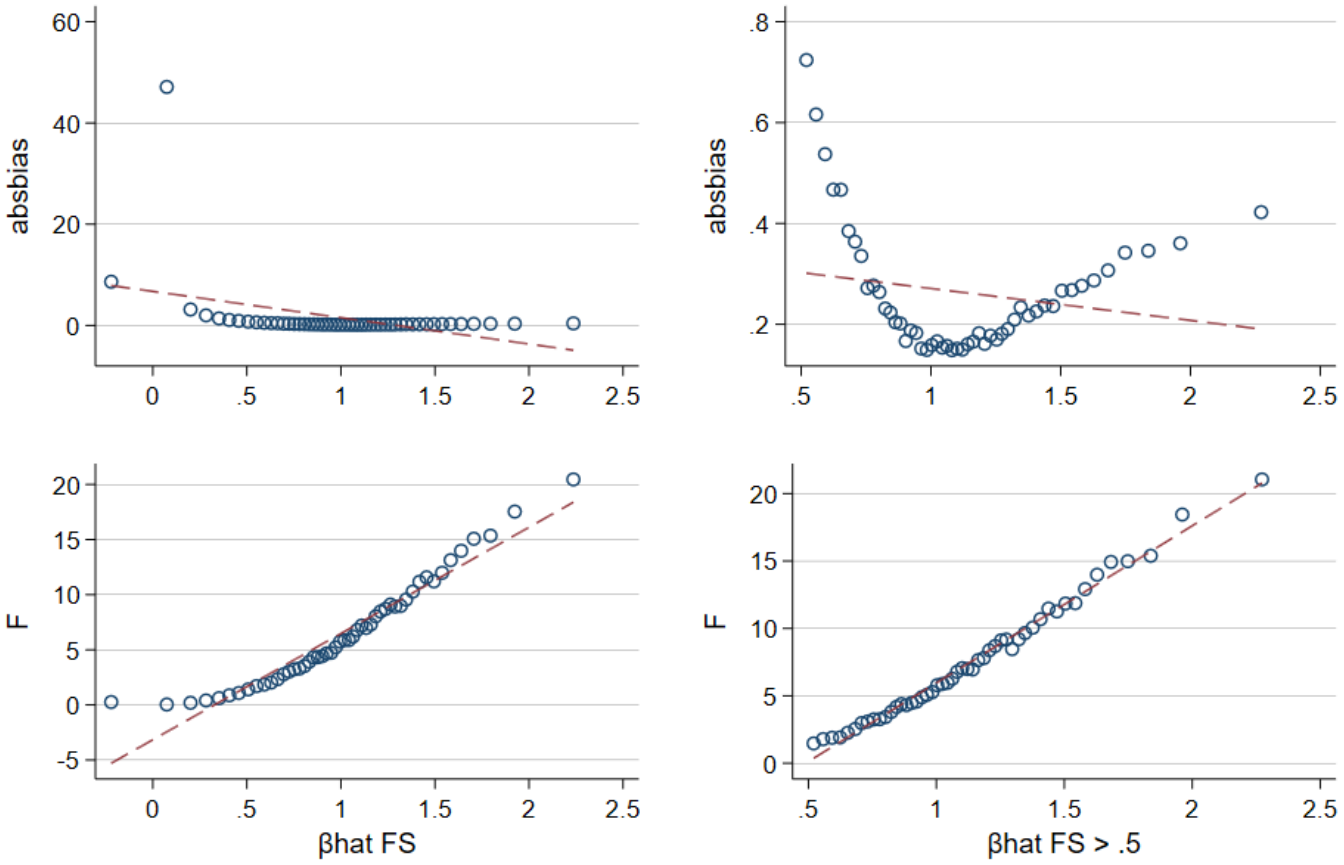

Figure 4: $|\hat{\beta}_{2SLS} - \beta_x|$ and F over $\hat{\beta}_{FS}$

2. 1. The compliance type is unobserved and thus cannot be individually identified because we can't see both $D_i(1)$ and $D_i(0)$ for each individual. We need to assume monotonicity to rule out defiers, and independence of the instrument to give a causal interpretation of the reduced form. Then the shares of always-takers, never-takers, and compliers: π_j, $j = a, n, c$, respectively, are identified as $\pi_a = \mathbb{E}(D_i = 1 \mid Z_i = 0)$, $\pi_n = \mathbb{E}(D_i = 0 \mid Z_i = 1)$, $\pi_c = \mathbb{E}(D_i = 1 \mid Z_i = 1) - \mathbb{E}(D_i = 1 \mid Z_i = 0)$.
   2. Denote j_i, $j = a; n; c$ the individuals with $D_i(1) = D_i(0) = 1$; $D_i(1) = D_i(0) = 0$; $D_i(1) = 1$, $D_i(0) = 0$ respectively, and $P(j_i)$ the sample fraction: $P(a_i) = 3/10, P(n_i) = 3/10, P(c_i) = 4/10$. Further, $P(Z_i = 1) = P(Z_i = 0) = 5/10$. $P(D_i = 0, Z_i = 1) = P(D_i = 1, Z_i = 0) = 2/10$, $P(D_i = 0, Z_i = 0) = P(D_i = 1, Z_i = 1) = 3/10$. Since $P(n_i, Z_i = 0) = P(n_i) - P(D_i = 0, Z_i = 1) = 1/10$ the fraction of never-takers amongst those who were not assigned to treatment $P(n_i|Z_i = 0) = \frac{P(n_i, Z_i=0)}{P(Z_i=0)} = 2/10$ is not the same amongst those who were assigned to treatment $P(n_i|Z_i = 1) = \frac{P(n_i, Z_i=1)}{P(Z_i=1)} = 4/10$ where $P(n_i, Z_i = 1) = P(D_i = 0, Z_i = 1)$ by the absence of defiers, thus randomization was not successful. Note $P(a_i|Z_i = 0) = 2/10, P(a_i|Z_i = 0) = 4/10$ by symmetry so that $P(c_i|Z_i = 0) = P(c_i|Z_i = 1) = 4/10$.
   3. $P(D_i = 1|Z_i = 1) - P(D_i = 1|Z_i = 0) = \frac{P(D_i=1, Z_i=1)}{P(Z_i=1)} - \frac{P(D_i=1, Z_i=0)}{P(Z_i=0)} = \frac{6-4}{10} = 2/10 \neq 4/10 = P(c_i)$ The coefficient estimate of the first stage regression of a dummy indicating treatment status on a dummy indicating treatment assignment and a constant is not identical to the true fraction of compliers in the sample. Assignment of Z_i determines the relationship between the two numbers. Since $P(D_i = 1|Z_i = 1) - P(D_i = 1|Z_i = 0) = P(c_i|Z_i = 1) + P(a_i|Z_i = 1) - P(a_i|Z_i = 0)$, they are equal only under successful randomization, while in this setting the former is $2/10$ less than the latter since $P(a_i|Z_i = 1) - P(a_i|Z_i = 0) = -2/10$.
   4. The second stage coefficient estimate on the dummy indicating treatment status is the reduced form estimate divided by the first stage estimate $\frac{P(Y_i=1|Y_i=1) - P(D_i=1|Z_i=0)}{P(D_i=1|Z_i=1) - P(D_i=1|Z_i=0)}$. The reduced form estimate is the difference in Y_i means between the Z_i groups. Averaging over always-takers, never-takers, and compliers, respectively, the mean outcomes for the treated $5 * 2/10 + 1 * 4/10 + (2 + 2) * 4/10$ and for the untreated $5 * 4/10 + 1 * 2/10 + (2 + 0) * 4/10$ are both equal to 3. The heterogeneity of outcomes across compliance groups, together with the failure of randomization, biases the estimate for the true sample's 'LATE' by -2 to 0, as always-takers with higher outcomes are relatively over-represented in the untreated group while the contrary holds for never-takers.

3. Given monotonicity (no-defiers) and independence (conditioning on Z_i), we have

$$1 \Pr[D_i(1) - D_i(0) = 1] + 0 + 0 = E[D_i(1) - D_i(0)] = E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0] > 0,$$

where always/never-takers drop out of the expcation. From the definition of conditional probability

$$\begin{aligned}\Pr[X_i = x|D_i(1) - D_i(0) = 1] &= \frac{\Pr[D_i(1) - D_i(0) = 1|X_i = x] \Pr[X_i = x]}{\Pr[D_i(1) - D_i(0) = 1]} \\ &= \frac{\Pr(X_i = x)(E[D_i \mid Z_i = 1, X_i = x] - E[D_i \mid Z_i = 0, , X_i = x])}{\Pr[D_i(1) > D_i(0)]}\end{aligned}$$

where the treatment status being completely determined by Z_i for compliers and independence Z_i is used. Finally, scaling the right hand side due to conditioning on $X_i = x$, we have

$$E[X_i|D_i(1) - D_i(0) = 1] = \frac{E[X_i D_i \mid Z_i = 1] - E[X_i D_i \mid Z_i = 0]}{\Pr[D_i(1) - D_i(0) = 1]}$$

so that substituting the first result for the denominator gives the desired rewriting. Alternatively, working from the hint, using monotonicity, independence and applying LIE over compliance types

$E[X_i D_i \mid Z_i = 1] - E[X_i D_i \mid Z_i = 0] = 1E[X_i|(D_i(1) - D_i(0) = 1]Pr[D_i(1) - D_i(0) = 1] +$
$0E[X_i|(D_i(1) - D_i(0) = 0]Pr[D_i(1) - D_i(0) = 0] - 1E[X_i|(D_i(1) - D_i(0) = -1]Pr[D_i(1) - D_i(0) = -1]]$
$= E[X_i|(D_i(1) - D_i(0) = 1]1Pr[D_i(1) - D_i(0) = 1]]$

combined with the derived expression for the first stage shows the given Wald estimator can be rewritten as $E[X_i|D_i(1) - D_i(0) = 1]$.