

Documentation for replication files for Belloni, Chernozhukov, and Hansen “Inference on Treatment Effects after Selection amongst High-Dimensional Controls”

1. Files for Simulations:

We performed all simulations in Matlab and have included all files called in obtaining the simulation results below.

1. ***RunLassoRedFormSims.m***: Wrapper to run simulations for different designs. Used to obtain baseline single-selection and double-selection results. This file runs the main simulation designs, specifically Designs 22, 44, and 666 which are reported in the main paper. The remaining designs are reported in the supplementary appendix. This function calls the files
MC_TE_Script_Lasso_RedForm_Design1001.m,
MC_TE_Script_Lasso_RedForm_Design1.m,
MC_TE_Script_Lasso_RedForm_Design2.m,
MC_TE_Script_Lasso_RedForm_Design22.m,
MC_TE_Script_Lasso_RedForm_Design3.m,
MC_TE_Script_Lasso_RedForm_Design4.m,
MC_TE_Script_Lasso_RedForm_Design44.m,
MC_TE_Script_Lasso_RedForm_Design5.m,
MC_TE_Script_Lasso_RedForm_Design6.m,
MC_TE_Script_Lasso_RedForm_Design7.m,
MC_TE_Script_Lasso_RedForm_Design72.m,
MC_TE_Script_Lasso_RedForm_Design722.m,
MC_TE_Script_Lasso_RedForm_Design8.m, and
MC_TE_Script_Lasso_RedForm_Design666.m.
2. ***MC_TE_Script_Lasso_RedForm_Design1001.m***: Simulation for design 1001. Calls the file ***MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m***.
3. ***MC_TE_Script_Lasso_RedForm_Design1.m***: Simulation for design 1. Calls the file ***MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m***.
4. ***MC_TE_Script_Lasso_RedForm_Design2.m***: Simulation for design 2. Calls the file ***MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m***.
5. ***MC_TE_Script_Lasso_RedForm_Design22.m***: Simulation for design 22. Calls the file ***MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m***.
6. ***MC_TE_Script_Lasso_RedForm_Design3.m***: Simulation for design 3. Calls the file ***MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m***.

7. *MC_TE_Script_Lasso_RedForm_Design4.m*: Simulation for design 4. Calls the file *MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m*.
8. *MC_TE_Script_Lasso_RedForm_Design44.m*: Simulation for design 44. Calls the file *MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m*.
9. *MC_TE_Script_Lasso_RedForm_Design5.m*: Simulation for design 5. Calls the file *MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m*.
10. *MC_TE_Script_Lasso_RedForm_Design6.m*: Simulation for design 6. Calls the file *MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m*.
11. *MC_TE_Script_Lasso_RedForm_Design7.m*: Simulation for design 7. Calls the file *MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m*.
12. *MC_TE_Script_Lasso_RedForm_Design72.m*: Simulation for design 72. Calls the file *MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m*.
13. *MC_TE_Script_Lasso_RedForm_Design722.m*: Simulation for design 722. Calls the file *MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m*.
14. *MC_TE_Script_Lasso_RedForm_Design8.m*: Simulation for design 8. Calls the file *MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m*.
15. *MC_TE_Script_Lasso_RedForm_Design666.m*: Simulation for design 666. Calls the file *MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m*.
16. *MC_TE_FixedDesign_Heteroskedastic_Lasso_RedForm.m*: Wrapper to call estimation files for simulations. This function calls the files *MC_TE_GetCoef_RedForm.m*, *MC_TE_Design_New.m*, *MC_TE_LassoHeteroskedastic_unpenalized.m*, *MC_TE_SimulateLambda.m*, *MC_TE_GetSupport.m*, *Heteroskedastic_se.m*, *MC_TE_PostEstimator.m*, *MC_TE_SqrtLassoHeteroskedastic_unpenalized.m*, *MC_TE_SimulateLambdaSqrtLASSO.m*, and *MC_TE_LassoHeteroskedastic_SplitSample*.
17. *MC_TE_Script_Infeasible_RedForm_Design22.m*: This file gives the simulation results for the oracle estimator for Design 22 reported in the paper. The file calls *MC_TE_FixedDesign_Heteroskedastic_Infeasible_RedForm.m*.
18. *MC_TE_Script_Infeasible_RedForm_Design44.m*: This file gives the simulation results for the oracle estimator for Design 44 reported in the paper. The file calls *MC_TE_FixedDesign_Heteroskedastic_Infeasible_RedForm.m*.
19. *MC_TE_Script_Infeasible_RedForm_Design666.m*: This file gives the simulation results for the oracle estimator for Design 666 reported in the paper. The file calls *MC_TE_FixedDesign_Heteroskedastic_Infeasible_RedForm.m*.
20. *MC_TE_FixedDesign_Heteroskedastic_Infeasible_RedForm.m*: Wrapper to call estimation files for simulations used in obtaining results for the oracle estimator used in the main paper. This function calls the files *MC_TE_GetCoef_RedForm.m*, *MC_TE_Design_New.m*, *MC_TE_LassoHeteroskedastic_unpenalized.m*, *MC_TE_SimulateLambda.m*, *MC_TE_GetSupport.m*, *Heteroskedastic_se.m*, *MC_TE_PostEstimator.m*, *MC_TE_SqrtLassoHeteroskedastic_unpenalized.m*,

MC_TE_SimulateLambdaSqrtLASSO.m, and
MC_TE_LassoHeteroskedastic_SplitSample.

21. *MC_TE_Script_Lasso_RedFormPlusRidge_Design22.m*: This file gives the simulation results for the estimator that includes a ridge fit among the variables to be selected over for Design 22. The file calls
MC_TE_FixedDesign_Heteroskedastic_Lasso_RedFormPlusRidge.m.
22. *MC_TE_Script_Lasso_RedFormPlusRidge_Design44.m*: This file gives the simulation results for the estimator that includes a ridge fit among the variables to be selected over for Design 44. The file calls
MC_TE_FixedDesign_Heteroskedastic_Lasso_RedFormPlusRidge.m.
23. *MC_TE_Script_Lasso_RedFormPlusRidge_Design666.m*: This file gives the simulation results for the estimator that includes a ridge fit among the variables to be selected over for Design 666. The file calls
MC_TE_FixedDesign_Heteroskedastic_Lasso_RedFormPlusRidge.m.
24. *MC_TE_Script_Lasso_RedFormPlusRidge_Design1001.m*: This file gives the simulation results for the estimator that includes a ridge fit among the variables to be selected over for Design 1001. The file calls
MC_TE_FixedDesign_Heteroskedastic_Lasso_RedFormPlusRidge.m.
25. *MC_TE_FixedDesign_Heteroskedastic_Lasso_RedFormPlusRidge.m*: Wrapper to call estimation files for simulations used in obtaining results for the oracle estimator used in the main paper. This function calls the files
MC_TE_GetCoef_RedForm.m, *MC_TE_Design_New.m*, *CVridgeK.m*,
MC_TE_LassoHeteroskedastic_unpenalized.m, *MC_TE_SimulateLambda.m*,
MC_TE_GetSupport.m, *Heteroskedastic_se.m*, *MC_TE_PostEstimator.m*,
MC_TE_SqrtLassoHeteroskedastic_unpenalized.m,
MC_TE_SimulateLambdaSqrtLASSO.m, and
MC_TE_LassoHeteroskedastic_SplitSample.
26. *RunLassoSims.m*: Wrapper to run simulations for different designs. All designs are reported in the supplementary appendix as Design 1a, 2a, etc. This function calls the files *MC_TE_Script_Lasso_Design1001.m*, *MC_TE_Script_Lasso_Design1.m*,
MC_TE_Script_Lasso_Design2.m, *MC_TE_Script_Lasso_Design22.m*,
MC_TE_Script_Lasso_Design3.m, *MC_TE_Script_Lasso_Design4.m*,
MC_TE_Script_Lasso_Design44.m, *MC_TE_Script_Lasso_Design5.m*,
MC_TE_Script_Lasso_Design6.m, *MC_TE_Script_Lasso_Design7.m*,
MC_TE_Script_Lasso_Design72.m, *MC_TE_Script_Lasso_Design722.m*, and
MC_TE_Script_Lasso_Design8.m.
27. *MC_TE_Script_Lasso_Design1001.m*: Simulation for design 1001a. Calls the file
MC_TE_FixedDesign_Heteroskedastic_Lasso.m.
28. *MC_TE_Script_Lasso_Design1.m*: Simulation for design 1a. Calls the file
MC_TE_FixedDesign_Heteroskedastic_Lasso.m.

29. **MC_TE_Script_Lasso_Design2.m**: Simulation for design 2a. Calls the file **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**.
30. **MC_TE_Script_Lasso_Design22.m**: Simulation for design 22a. Calls the file **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**.
31. **MC_TE_Script_Lasso_RedForm_Design3.m**: Simulation for design 3a. Calls the file **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**.
32. **MC_TE_Script_Lasso_Design4.m**: Simulation for design 4a. Calls the file **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**.
33. **MC_TE_Script_Lasso_Design44.m**: Simulation for design 44a. Calls the file **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**.
34. **MC_TE_Script_Lasso_Design5.m**: Simulation for design 5a. Calls the file **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**.
35. **MC_TE_Script_Lasso_Design6.m**: Simulation for design 6a. Calls the file **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**.
36. **MC_TE_Script_Lasso_Design7.m**: Simulation for design 7a. Calls the file **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**.
37. **MC_TE_Script_Lasso_Design72.m**: Simulation for design 72a. Calls the file **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**.
38. **MC_TE_Script_Lasso_Design722.m**: Simulation for design 722a. Calls the file **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**.
39. **MC_TE_Script_Lasso_Design8.m**: Simulation for design 8a. Calls the file **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**.
40. **MC_TE_FixedDesign_Heteroskedastic_Lasso.m**: Wrapper to call estimation files for simulations. This function calls the files **MC_TE_GetCoef.m**, **MC_TE_Design_New.m**, **MC_TE_LassoHeteroskedastic_unpenalized.m**, **MC_TE_SimulateLambda.m**, **MC_TE_GetSupport.m**, **Heteroskedastic_se.m**, **MC_TE_PostEstimator.m**, **MC_TE_SqrtLassoHeteroskedastic_unpenalized.m**, **MC_TE_SimulateLambdaSqrtLASSO.m**, and **MC_TE_LassoHeteroskedastic_SplitSample**.
41. **MC_TE_GetCoef_RedForm.m**: Wrapper with coefficient structures for different simulation designs 1, 2, 22, etc.
42. **MC_TE_GetCoef.m**: Wrapper with coefficient structures for different simulation designs 1a, 2a, 22a, etc.
43. **MC_TE_Design_New.m**: Wrapper that takes coefficient structure and generates simulation data.
44. **CVridgeK.m**: Computes K-fold CV for ridge regression.
45. **MC_TE_LassoHeteroskedastic_unpenalized.m**: Lasso estimator using penalty weights to accommodate heteroskedasticity.
46. **MC_TE_SimulateLambda.m**: Function to simulate penalty level for Lasso based on uniformly dominating the score as in Belloni and Chernozhukov (2011b).
47. **MC_TE_GetSupport.m**: Select the variables with non-zero coefficient estimates.

48. **Heteroskedastic_se.m**: Jack-knife standard error estimator.
49. **MC_TE_PostEstimator.m**: Post-Lasso estimator.
50. **MC_TE_SqrtLassoHeteroskedastic_unpenalized.m**: Square-root Lasso estimator. Not reported in paper.
51. **MC_TE_SimulateLambdaSqrtLASSO.m**: Function to simulate penalty level for square-root Lasso based on uniformly dominating the score as in Belloni and Chernozhukov (2011b).
52. **MC_TE_LassoHeteroskedastic_SplitSample**: Split-sample Lasso estimator. Not reported in paper.

2. Files for empirical example:

1. **LevittExampleREStudTable.m**: This file produces the results for Table 2 of the main paper. The file calls one dataset, **levitt_ex.dat**, and the auxiliary files **recode.m**, **dummy.m**, **cluster_se.m**, **acfcomp2.m**, **findNonCollinear.m**, and **LassoShooting2.m**.
2. **LevittExamplePLMPaperLevels.m**: Produces results in supplementary appendix based on level (fixed effects) specification allowing for a combination of trigonometric and polynomial trend terms. This file calls two datasets, **levitt_ex.dat** and **levitt2008.txt**, and the auxiliary files **recode.m**, **dummy.m**, **cluster_se.m**, and **LassoShooting.m**.
3. **LevittExamplePLMPaperLevelsPoly.m**: Produces results in supplementary appendix based on level (fixed effects) specification allowing for polynomial trend terms. This file calls two datasets, **levitt_ex.dat** and **levitt2008.txt**, and the auxiliary files **recode.m**, **dummy.m**, **cluster_se.m**, and **LassoShooting.m**.
4. **levitt_ex.dat**: Data set used in the empirical example. Refer to **LevittExampleREStudTable.m** for column identities.
5. **levitt2008.txt**: Data set used in the empirical results provided in the supplementary appendix. See the files **LevittExamplePLMPaperLevels.m** and **LevittExamplePLMPaperLevelsPoly.m** for variable identities.
6. **recode.m**: Changes categorical (index) variable to be consecutive integers ranging from 1 to the total number of categories.
7. **dummy.m**: Creates a matrix of indicator variables.
8. **cluster_se.m**: Calculates clustered standard errors.
9. **acfcomp2.m**: Calculates autocorrelations.
10. **findNonCollinear.m**: Finds columns in a matrix that are not perfectly collinear.
11. **LassoShooting.m**: Lasso estimator.
12. **LassoShooting2.m**: Lasso estimator allowing for penalty loading appropriate for heteroskedastic, non-Gaussian data.