README file

Our project repository contains the `113270` folder, which contains the replication package from the paper “Consumption Inequality and Partial Insurance” by Richard Blundell, Luigi Pistaferri, and Ian Preston (2008). It also contains the following files:

1. `adjust-AER.Rmd`
2. `cexall_new_AER.Rmd`
3. `cleaning.Rmd`
4. `cleaning2.R`
5. `compare-predictions.Rmd`
6. `crossvalidation.Rmd`
7. `data-stats.Rmd`
8. `gbm-sim.Rmd`
9. `gbm.Rmd`
10. `impute_AER.Rmd`
11. `imputed-PSID.csv`
12. `libraries.R`
13. `montecarlo.Rmd`
14. `RMSE_out.csv`
15. `neuralnetwork.Rmd`
16. `regressions.Rmd`
17. `results.Rmd`

- The `libraries.R` file contains the packages required to run all of the files in our repository. Please run it prior to running any of the other files mentioned above.
- The files `adjust-AER.Rmd`, `impute-AER.Rmd`, `cexall_new_AER.Rmd` are translated versions of the Stata .do files `adjust_AER.do`, `impute_AER.do`, and `cexall_new_AER.do` from the Blundell et. al replication package in folder `113270`.
- The `cleaning.Rmd` file contains the cleaning, dummy variable conversion, and significant feature selection process. The `cleaning2.R` file contains the R script for this, which we will call in other R markdown files to ensure that our CEX and PSID files are properly cleaned prior to use.
- The `compare-predictions.Rmd` file contains the procedure to run each of our three models on the PSID data, and then compare the results.
- The `data-stats.Rmd` file contains the information on the datasets we used, along with the code used to create the tables on them.

- The `gbm.Rmd` and `regressions.Rmd` files contain our code to build and train the Gradient Boosting Machines model, our basic multiple linear regression model, and the imputation-based model from the Blundell et. al (2008) paper.
- The `gbm-sim.Rmd` file contains the code to
- The `gbm-montecarlo.Rmd` file contains the code to run our 500 bootstrap Monte Carlo Simulations for our linear regression model, paper imputation-based model, and GBM model, and then calculate the RMSE across simulations for each.
- The `results.Rmd` file contains the code to create the graphs, plots, and charts shown in our results section, along with any associated findings.
- The `neural-network.Rmd` and `montecarlo.Rmd` files contain the code we used to build and train a feedforward neural network model and then run the 500 bootstrapped simulations on our existing three models in addition to the neural network. However, due to limitations on computational power and processing time, we were unable to include the results of these attempts in our paper, and our findings as described in our paper do not include anything resulting from these. However, we decided to include them in our documentation for transparency, along with demonstrating potential routes for which our paper could potentially be expanded on in the future.
- The `imputed-PSID.csv` file contains the output of the `compare-predictions.Rmd` file when we ran our code. We used the data in this file to create Figure 2 in our paper.
- The `RMSE_out.csv` file contains the output of the `gbm-sim.Rmd` file. We used this data to create Figure 1, Table 2, and Table 3 as shown in our paper.