# Network Contagion NVI Computation Readme

August 27, 2018

This document is meant to provide important additional information about the files and programs contained san git repository for Collin and Fernando's NVI project. The portion of this project that uses Matlab simulations resides on the RAN, and has its own separate readme file.

## Updating the data

The Stata do file NtwkCntagn.do will call nearly ever script that needs to be run to completely update the data - and every script that needs to be run every time any data is updated. What follows is a step-by-step instruction of how to properly update the data using the NtwkCntagn.do file.

### If you're doing this for the first time

1. Clone the git bare repository on the san at /san/RDS/Work/cmf/b1cdj02/Fernando/Network_Contagion.git into some other san location of your choice.
2. Sign up for a WRDS account.
3. Create a .pgpass file in your home directory of the WRDS cloud server at wrds-cloud.wharton.upenn.edu.
4. Change the WRDS_meta.py and batch_pull.sh to match your WRDS username/password and your own WRDS cloud directory structure.
5. Make sure you have access to the KMV default probability database. Email the research library to request access.
6. Update RawY9C.stcmd to path to your new directory.

### Step-By-Step

Unless otherwise indicated, these are all done within the NtwkCntagn.do file.

1. If you haven't done it yet this quarter, run ffunds.do in Stata locally on your bank machine.
2. If you're doing a full update of everything set the global wipe_temp to 1. (Would not recommend doing this on the first run - makes code run for 2hours rather than 15minutes)
3. Change the name of this runthrough's log file by changing global runthrough tag.
4. Decide whether to change any of the NVI settings (globals gamma_benchmark, gammas, delta_fixed, snapshot_date, bank_sample)
5. If you're extending the endpoint of the series, change global charts_end.
6. If you're updating permco_cusip.csv, and you have already updated WRDS_meta.py and batch_pull.sh with your account info, set changed_WRDS_files to 1

7. If you're only outputting simulation data to use on the RAN, change global output_simulation_data to 1. This will skip running most scripts in the directory, so make sure you've run the do file with output_simulation_data = 0 first, to update things.

8. Run NtwkCntagn.do!

### Some Things to Do Occassionally

1. See if FI has a new RSSID-Permco match posted online. If it does, change entity_permco_date accordingly. (Note that Sean Hundtofte, used to be the economist in charge of this data and it is unclear when someone will update it again)

2. See if there's a new version of Moody's KMV available. Could email the Research Library, or poke around the data dictionary. If there is, we'll have to decide whether to switch to the new version (can be quite different).

3. If you're extending the end of the series, you'll have to change Analysis_Y9C to deal with any variable changes. This is a major task. If we ever decide to extend the series, I would suggest checking with FI to see whether there's some way to overhaul the code in Analysis_Y9C. FI's reg_data_nc dataset qbhc_nc_clean.dta in theory does many of the same things we do, and would be maintained by the FI reg_data RA (who has better knowledge of everything than we do). However, we would need to put some work into making sure things would be consistent between the two. This would be a time-consuming task. FI's make_clean program is very, very long. If you decide to go the route of updating Analysis_Y9C using the unprocessed data, compairing the make_clean dataset to Analysis_Y9C in addition to using MDRM should expidite the update process

4. If you're extending the end of the series beyond 2016Q4, you'll also need new a few other data files which must be updated manually:

   (a) input/rssd_hh_match_all. dta – BHC parent structure. Ask the reg_data FI RA for it (nicely).

   (b) SIFMA aggregate dealer data. Ask Fernando about this - got it from the Research Library last time.

5. Make sure that the matching in Match_RSSID_MKMVID is going okay. I would periodically check to be sure that most of the big banks (e.g. the ones in the bank_sample global) aren't being dropped during the sample. You could plot the KMV default probabiliy of those KMVIDs and make sure there aren't any strange breaks.

## Directory of Scripts

| File Name | Language | Where to run | Description |
|---|---|---|---|
| NtwkCntagn.do | Stata | San cluster | Main script for updating data and producing charts for paper. Sets global for use in other programs. |
| ffunds.do | Stata | Locally | Pulls flow of funds data from haver (hence must be run locally) and compiles balance sheet info on aggregate sector nodes. Run about once a quarter. |
| rssd_hh_gen | Stata | San cluster | Generates input/rssd_hh_match_all. dta. Completely regenerates the file and therefore is rather inefficient. Run overnight, independently from the other code. |
| Update_Data.do | Stata | San cluster | Primary script to update data. Performs ODBC data loads and executes WRDS_meta.py, if needed. |
| WRDS_meta.py | Python | San cluster | Interfaces with WRDS cloud server to pull Permco-CUSIP match. IMPORTANT: Must be customized with individual WRDS account info. |

| File Name | Language | Where to run | Description |
|---|---|---|---|
| WRDS_query.py | Python | WRDS Cloud | Helper file to execute WRDS cloud batch job for Permco-CUSIP match. |
| batch_pull.sh | Shell | WRDS Cloud | Helper file to execute WRDS cloud batch job for Permco-CUSIP match. |
| Match_RSSID_MKMVID.do | Stata | San cluster | Matches BHC FR-Y9C RSSID numbers to corresponding MKMVID in KMV default probability database (match goes RSSID – Permco – CUSIP – MKMVID). Will pull FI's permco-CUSIP link from web, to fill in gaps from WRDS pull. |
| CallDeposits.do | Stata | San cluster | Compile data on each BHC's amount of FDIC-insured deposits. This will be matched with FR-Y9C data. |
| drd_matching.do | Stata | San cluster | Use Moody's database of firm default events to find bankruptcies of firms in our FR-Y9C sample. Bankrupt firms will be excluded from NVI. |
| Analysis_Y9C.do | Stata | San cluster | Clean Raw KMV default probability data, process Fr-Y9C data to produce %in/%out numbers for each BHC. |
| compile_agg_sector.do | Stata | San cluster | Create data on aggregate sector nodes to include in NVI - particularly the sector aggregate probability of default |
| Model_series_processing | Stata | San cluster | Take in processed data to produce fields in interest from Glasserman - like the NVI, contagion index, and robustness exercises on the two. |
| Plots_Paper.do | Stata | San cluster | Produce tables and figures for use in paper |
| Plots_Appendix.do | Stata | San cluster | Archive of some old scripts for producing past figures and tables. |
| process_FOCUS.py | Python | San cluster | Process SIFMA files in input/ produce data on Top 1-10, 11-25 dealers that can be used to produce aggregate sector nodes. |
| RawY9C.stcmd | Stattransfer | San cluster | Convert sas7bdat file of non-classified FR-Y9C data from FI directory into a workable stata format. |
| format_sim_data.mat | Matlab | Locally | Convert excel file of simulation-ready NVI data into matlab database, for use in RAN matlab simulations. |
| sumary_stats.do | Stata | San cluster | Produce summary statistics tables of balance sheet breakdowns, for use in paper. |