READ ME

The following files provide the needed replication code for vom Lehn and Winberry (2021), "The Investment Network, Sectoral Comovement, and the Changing U.S. Business Cycle." Please cite our paper if you use these codes or have found them helpful.

The following describes what each folder of these replication files provides in terms of data and code and where the tables/figures shown in the paper are generated from. The code and data use Matlab, Stata, or Excel.

The code and data are set up such that there is no need to run files in any particular sequence to obtain the final results, with the exception of code/data obtained from the work of other researchers. To be able to run all code, the following data/code should be obtained from the following locations (drawn from the following research projects):

1. For constructing the rental services matrix (place in Raw Data folder): "McDaniel_tax_02_01_2017_orig.xlsx": http://www.caramcdaniel.com/researchpapers, 15 countries link; Citation: McDaniel, C. (2007): "Average tax rates on consumption, investment, labor and capital in the OECD 1950-2003," Manuscript, Arizona State University, 19602004.

2. For converting historical SIC data to NAICS (place in Converting SIC to NAICS folder) "SIC_NAICS_concordance.dta": http://faculty.tuck.dartmouth.edu/teresa-fort/data/, SIC-NAICS concordance file; Citation: Fort, Teresa C., and Shawn Klimek," The Effects of Industry Classification Changes on US Employment Composition," CES Working Paper 18-28, 2018.

Aside from these additional data/code files, because the output of intermediate code steps is saved alongside the original code (and in some cases is constructed by copy/pasting output from Stata or Excel to Excel or Matlab), there is no need to run all code in the replication files to produce specific results. However, were one to start from scratch, the general sequence of steps to follow would be:

1. Gather all the Raw Data contained in the "Raw Data" folder.
2. Run the .do files "*loadPEQbridge*.do" and "*SIC_NAICS_BEA_allsec.do*" in the "Converting SIC to NAICS" folder to convert historical data into consistent sectoral definitions.
3. Run the .m file "buildinvmat.m" in the "Investment Network Construction" folder to produce the investment network.
4. Run the .m files "*LoadIOMat_37sec.m*" and "*LoadIOMat_30sec.m"* in the "Input Output and Production Parameters" folder to produce datasets containing moments used for model calibration.
5. Run the .do files "*DATA_BUILD_bea_to_stata_37.do*" and "*DATA_BUILD_bea_to_stata_30.do"* in the "TFP Construction" folder to produce time series of TFP (this in particular draws from a single Excel file -- "37 Sector Data" or "30 Sector data" -- in which raw data and data outputs from prior steps has been manually compiled).
6. Run the .m file "empirical_results.m" in the "Empirical Analysis" folder to produce the empirical results presented in the paper (this draws from a single .mat file -- "*beadat_37sec.mat" or "beadat_30sec.mat"* -- in which raw data and data outputs from prior steps has been manually compiled).
7. Run the .do file "*DATA_BUILD_excel.do"* which converts the NBER CES database into an Excel file, which is then manually transplanted into the .mat file "nberces_naics.mat." Then, run the .m file "nberces_analysis.m" to produce the results in Appendix H using the NBER CES database.

8. Run the .m file "GenerateModelResults" in the "Model Analysis" folder to produce all model results.
9. Run the .m file "Main_EFP_vLW_nonstationary_HW.m" in the "MMTT Algorithm" folder to produce results for a transition path with trend changes in model parameters.
10. Run the .m files in the folder "Graphics" to produce the final remaining graphics presented in the paper.

The following provides more details on the folders included with the replication packet and the files contained therein.

Raw Data

This folder contains the raw data (primarily drawn from the BEA) used throughout the paper. All data is either in Excel workbooks or in .txt file formats.

Converting SIC to NAICS

This folder contains code and data needed for the conversion of historical BEA data measuring sectors using SIC codes to current NAICS sector definitions. This conversion is accomplished using data taken from Fort and Klimek (2016). We use this crosswalk to convert historical employment data, measured in SIC codes, to the NAICS codes in order to obtain a full time series of employment at the 37 sector level; to convert historical GDP by Industry data (including compensation in each sector) to obtain a full time series of the labor share in each sector at the 37 sector level; and to convert the 1987 and 1992 bridge files for equipment investment from SIC to NAICS codes.

The following Stata code files accomplish the following tasks:

- *loadPEQbridge.do*: This code imports the raw bridge files from the BEA (see Appendix A.2 for details about bridge files), converts them to NAICS codes if needed (only required for 1987 and 1992), and then exports reformatted Excel files containing the updated bridge file data.
- *SICbridgeconv.do*: This code assigns SIC codes to the Input Output codes used in the 1987 and 1992 bridge files, a necessary precursor to converting the bridge data to NAICS codes.
- *SIC_NAICS_BEA_allsec.do*: This code uses the SIC to NAICS concordance from Fort and Klimek (2016) to convert historical sectoral, employment, compensation and value added data to NAICS codes.

This folder also contains self-employment scaling factors used to adjust sectoral labor share measures in the model calibration (*Self Employment Scaling Factors*.xlsx), the final converted time series of labor shares by sector (*Full Converted Labor Share Time Series*.xlsx), and the final converted time series of employment by sector (*Full Converted Employment Time Series.xlsx*). The self-employment scaling factors are constructed in the Excel file using raw data files on self-employed and full-time part time employment by sector; the final converted time series of labor shares and employment are taken from the Stata output of *SIC_NAICS_BEA_allsec.do.*

Investment Network Construction

This folder contains the code used to construct the time series of the investment network used in the paper as well as code used to construct alternative networks (as described in Appendix A.3):

The following core Matlab code files accomplish the following tasks:

- *buildinvmat.m*: Constructs the bridge files and investment networks for equipment, non-residential structures, intellectual property and residential investment. Combines these to form

the investment network as used in the paper, as well as other networks (alternative sector coverage, adding maintenance, etc.)
- *buildrkmat.m*: Constructs a time series of the capital rental services payments made by each sector in each year on each type of capital asset.
- *buildrentmat.m:* Given the constructed time series of sectoral capital rental services payments made across assets each year, constructs a capital rental services matrix using the same bridge files as in the baseline investment network.

The folder also contains the following data files (in .mat format) that contain various definitions of the investment network:
- *invmatdat_all.mat*: includes all objects created in the code, *buildinvmat.m*, including bridge files for structures, intellectual property, equipment and residential investment.
- *invmatdat_37.mat*: contains the 37 sector investment network as used throughout vom Lehn and Winberry (2021). Also contains this network with maintenance added to it (invmat_maint), as well as separate investment networks for equipment, structures, residential, and intellectual property investment.
- *invmatdat_41.mat*: contains the investment network for a 41 sector partition of the economy with additional detail in the mining and real estate sectors.
- *invmatdat_govt.mat*: contains the 43 sector investment network, which adds the federal and state/local government sectors to the 41-sector network described above.
- *invmatdat_noRE.mat*: contains the 37 sector investment network with the real estate sector omitted (making it a 36 sector network).
- *invmatdat_30.mat*: contains a 30 sector investment network where all non-durable manufacturing sectors in the 37 sector partition have been condensed into a single sector.
- *rentmatdat_all.mat*: includes all objected created in the code, *buildrentmat.m*.
- *rkdat.mat*: includes all time series of capital rental services expenditures as constructed in *buildrk.m*
- *rentmatdat_37.mat*: contains the capital rental services matrix for the 37 sector partition.
- *rentmatdat_41.mat:* contains the capital rental services matrix for the 41 sector partition.
- *rentmatdat_govt.mat*: contains the capital rental services matrix for the 43 sector partition.

<u>Empirical Analysis</u>
This folder contains code, data and graphics files to replicate the empirical analyses in Section 2 and Appendices B, F and H.

The following core Matlab code files accomplish the following tasks:
- *empirical_results.m*: Takes core data on employment, value added, investment, gross output, and TFP for each sector and constructs the empirical moments reported in Tables 2, 3, 4, 5, 8, 9, B.1, D.1, F.1, F.2, H.1, H.2, H.3, H.4, H.5, H.7, and H.8 and Figures 2, 7, B.1, B.2, D.5, F.2, G.1, G.2, H.1, and H.2. Note the code can be run with the option of detrending via first differences or the HP filter (on line 10); results for certain tables/figures require specifying one of these options. To run the principal components analysis on the 30 sector version of the data (Table D.1), line 9 in the code needs to be uncommented (loading the 30 sector data instead of the 37 sector data). In this .m file, the graphics for Figure D.5 and H.1 are created as well.
- *produce_correlogram.m*: This code produces the correlogram figures in the paper (Figures 2, B.1, B.2, and F.2).
- *nberces_analysis.m*: Does the analysis of NBER-CES manufacturing data, constructing the empirical moments reported in Table H.6.

The following Stata do files accomplish the following tasks:

- *DATA_BUILD_excel.do*: Loads the NBER-CES database and converts it into an Excel file.

The following datasets may be of interest to other researchers because they compile in a single place a range of data on sectoral production:

- *beadat_37sec.mat*: Contains employment, investment expenditures (real and nominal), value added (real and nominal), nominal gross output, and TFP (measured using value added or gross output) for each of the 37 sectors studied in vom Lehn and Winberry (2021) for each year from 1948-2018. Also contains the share of total intermediates and investment produced by each sector for each year.
- *beadat_30sec.mat*: Contains employment, investment expenditures (real and nominal), value added (real and nominal), nominal gross output, and TFP (measured using value added or gross output) for the 30 sector partition used in vom Lehn and Winberry (2021), where non-durable manufacturing is collapsed into a single sector.

Input Output and Production Parameters

This folder contains code (*LoadIOMat_37sec.m*) which takes raw BEA data, converts sectors into the consistent 37 sector definitions used in vom Lehn and Winberry (2021), and constructs data moments needed for the calibration of the model. In particular, this code generates depreciation rates by sector and year and the input-output matrix (and nominal final uses, value added and gross output) by sector and year. It also loads in capital shares for each year, as constructed in the "Converting SIC to NAICS" folder.

A similar file (*LoadIOMat_30sec.m*) is also included to do the same tasks for the 30 sector partition of the model used in Appendices D and G.

To construct a depreciation rate for the real estate sector, the folder also includes an Excel file (*Constructing Residential Assets Depreciation Rate.xlsx*) computing an implied depreciation rate of residential assets, drawing from BEA data on the stock and measured depreciation of these assets.

TFP Construction

This folder contains data and code used to produce measured TFP for each sector in the model.

The following Stata code files accomplish the following tasks:

- *DATA_BUILD_bea_to_stata_37.do*: Loads in investment, employment, gross output, intermediates and production cost share to construct TFP by sector for each year from 1948 to 2018. The capital stock in each sector is constructed using the perpetual inventory method, starting from initial valuations of the capital stock in each sector in the year 1948. TFP is constructed using both gross output and value added.
- *DATA_BUILD_bea_to_stata_30.do*: Performs the same tasks for the 30 sector partition of the economy.

The Excel files "37 Sector Data" and "30 Sector Data" contain all the necessary inputs to the construction of TFP. The Stata files "TFP_37.dta" and "TFP_30.dta" contain the output of the TFP construction process.

Model Analysis

This folder contains the code necessary to compute the log-linearized solution to the model presented in vom Lehn and Winberry (2021). Dynare (version 4.5.7) is used to solve the model.

Running the Matlab script "GenerateModelResults.m" will produce all model results in the paper and output to .mat files for constructing graphics (as documented elsewhere in this document) or to the .mat file "finaltables.mat" where all model output for the tables of the paper can be found. This overarching Matlab script calls various other scripts (many starting with the prefix "SolveSimulateModel…m"), which invoke Dynare .mod files and other functions contained in the folder (such as hpfilter.m for filtering or arma_mlear.m for estimating persistence parameters).

Because there are so many model simulations, including 2nd order simulations which take longer to run, it may be advisable to comment out parts of the script "GenerateModelResults.m" to focus on certain results.

The script "GenerateModelResults.m" provides significant detail about which results are generated by which segments of code and should be consulted for specific questions about where to find code/data for specific model results.

MMTT Algorithm

This folder contains code necessary to run the perfect foresight structural change exercise described in Appendix G.3. Much of the code is adapted from Maliar, Maliar, Tsener and Taylor (2020), "A tractable framework for analyzing a class of nonstationary Markov models" (MMTT).

The core Matlab file that calls all other Matlab files to solve the perfect foresight model is "*Main_EFP_vLW_nonstationary_HW.m*." This code first constructs trends in each of the model parameters over time and then solves the model using the MMTT framework. Then, results for Table G.2 are produced.

Graphics

This folder contains code and output used to produce many figures in the paper.

The following Matlab code files accomplish the following tasks:

- *produce_heatmaps.m*: Produces all the heatmaps in the paper (Figures 1, 3, 4, A.1, A.2, A.3).
- *produce_bar_charts.m*: Produces the bar charts reporting parameter values used in calibration (Figures D.1, D.2, D.3, D.4, and D.6) and reports numerical elasticities from simulation (Figures 5, F.1, F.4, and F.5).
- *produce_irfs.m*: Produces the impulse response figure (Figure 6).
- *model_data_comparisons.m*: Produces figures comparing model and data outcomes – the rolling window correlation of productivity and output (Figure 7), the distribution of Domar weights (Figure F.3), the time series of aggregate fluctuations in model and data (Figures G.1, G.2), and the comparison of covariance changes in the model and in data (Figure H.2).
- *kernelplots.m*: Produces the kernel density estimates from 10,000 calibration simulations plotted in Figure F.6.