README file to accompany "Vacancy Chains," by Elsby, Gottfries, Michaels, and Ratner

External Data

Restricted-access BLS microdata

The paper uses restricted-access BLS microdata from the Quarterly Census of Employment and Wages (QCEW) and the Job Openings and Labor Turnover Survey (JOLTS). These are used to produce Figures 1, 2, and 3, in the main text, and Figures B1, B2, and B3 in online appendix B. In addition, the replacement share and implied chain length series in Figure 3B are used to compute the chain length data entries in Table 3 and Figure 9D in the main text, and in Table G and Figure H1I in the online appendices.

Access to these microdata was restricted to a subset of forty states that approved access onsite at the BLS for this project. (The samples exclude data for Florida, Illinois, Massachusetts, Michigan, Mississippi, New Hampshire, New York, Oregon, Pennsylvania, Wisconsin, and Wyoming.)

An application to access these microdata at BLS must be filed through the Standard Application Process (SAP) portal. See the SAP user guide for details. The BLS' Restricted Data Access (RDA) page also offers further guidance. For any questions, contact a RDA administrator at BLS.

We provide copies of the code used to create the relevant final results and we describe that code below.

Compustat microdata

The paper uses Compustat microdata for two analyses of firm-level average labor productivity: The estimates of the autocorrelation of average labor productivity reported in Table 2; and the estimates of the dispersion of average labor productivity growth conditional on employment adjustment reported in section 5. Details are provided in Online Appendix F.

We access Compustat data through a subscription to Wharton Research Data Services (WRDS). For questions regarding a subscription, contact WRDS.

We provide copies of the code used to create the relevant final results and we describe that code below.

Publicly-available aggregate data

The final column of Table 3, panel A, uses standard, publicly available data from the Bureau of Labor Statistics' CPS, JOLTS, and LPC series, as well as Fallick and Fleischman's (2004) estimates of the E-to-E rate, and Davis, Faberman and Haltiwanger's (2012) synthetic JOLTS-BED quit and hires rates.

The raw data are provided in the red tabs in the spreadsheet `Table-3-data.xlsx` located in the folder `Output`. The file describes how each of the series was accessed.

The latter data are also used to compute the empirical descriptive impulse responses reported in Figure 9 and Online Appendix H.

Footnote 30 of the paper reports the average monthly job-leaver inflow rate into unemployment as a fraction of employment and quits. These also are inferred from publicly available BLS data from the CPS and JOLTS.

The raw data are provided in the red tabs in the spreadsheet `Footnote-30.xlsx` located in the folder `Output`. The file describes how each of the series was accessed.

Figure D in online appendix D uses publicly available Business Dynamics Statistics (BDS) data from the Census Bureau. These are taken from the 2020 vintage BDS U.S. firm age-by-firm size dataset. The data are provided in the BDS tab in the spreadsheet `Chains.xlsx` located in the folder `Output`.

Software Used

MATLAB (version 2023b/2024a). Stata (version 18.0). SAS (version SAS9.4M8).

Description of Code and Output

We split our description of our code and output into empirical analyses and model simulations as follows:

Empirical analyses: Restricted-access BLS microdata (Figures 1, 2, 3, B1, B2, B3)

The programs used to produce figures based on BLS microdata are listed below. We first report the programs that pertain to QCEW data followed by the programs that pertain to JOLTS data. The QCEW programs are housed in a folder `QCEW programs`, and the JOLTS programs are housed in a folder `JOLTS programs`. Throughout this document, all folders are assumed to be located in the `/home` directory.

QCEW programs

`Compile_QuartData.sas`: This pulls together state-level QCEW files and assembles a complete panel dataset for all states in our sample, `QuartData.sas7bdat`. All of our analysis of the QCEW is based on the latter dataset.

`naics2dig.sas`: This program is referenced by `Compile_QuartData.sas` and assigns a supersector (NAICS) code to each establishment. This is for the sake of completeness; we do not use the sector information in what follows.

`GrowthDist.sas`: This produces the histogram of quarterly employment growth displayed in **panel A of Figure 1**. The series is named `means`. (Series with names of the form `means`x'` are based on alternative definitions of inaction described in the paper. For instance, for the histogram named `means11`, establishments with growth rates in the max{1, 1%} window are treated as if their net growth were zero.)

`growthbins_smaller.sas`: This program is referenced by `GrowthDist.sas` and specifies the net employment growth bins used in the histogram.

The next four QCEW-related programs produce inaction rates.

The first two (i) – (ii) produce inaction rates by horizon. For any given quarter, the programs in (i) – (ii) look up to 16 quarters ahead to calculate inaction rates by horizon. Each then rolls forward to the next quarter and repeats. Thus, for each calendar quarter, there is a profile of inaction rates by horizon.

The third (iii) produces inaction rates by establishment size class. Again, these estimates are made for each calendar quarter.

Each program (i) – (iii) calculates inaction rates for each of the three definitions of inaction described in the paper. The names given to the final series all begin with `inactive`, and the suffix refers to the definition of inaction on which it is based. For instance `inactive11` refers to the max{1,1%} window around zero net change. If the suffix is omitted, inaction is defined as exact zero net change.

The fourth program (iv) averages the results in (i) – (iii) over calendar time to produce a single profile of inaction rates over horizon (in the case of (i) and (ii)) or size class (in the case of (iii)).

(i) `inactdecay.sas`: This produces the inaction rates by horizon that underlie what is shown in **panel A of Figure 2** and **panel A of Figure B2**. The sample for Figure 2 is the universe of establishments. The sample for Panel A of Figure B2 consists of establishments with initial size greater than 20 (these results are saved to the dataset with suffix "`over20`"). For each sample, we compute and store both

establishment- and employment-weighted inaction rates (the latter will have suffix "wg"). Note, though, that panel A of Figure B2 displays only the employment-weighted results.

(ii) inactdecay_everytau.sas: This produces the inaction rates by horizon that underlie what is shown in **panel C of Figure B2**. These profiles are employment-weighted. (This file also produces establishment-weighted profiles of inaction rates as well as establishment- and employment-weighted profiles for the subsample of employers with initial employment greater than than 20.)

(iii) inactdecay_bysizegrp.sas: This produces the inaction rates by size class that underlie what is shown in **panel E of Figure B2**. The figure reports, specifically, the establishment-weighted inaction rates (within class).

(iv) inactdecay_averages.do: This program pulls from the SAS datasets made by programs (i)—(iii) and produces average inaction rates over calendar time. These average profiles of inaction rates are then saved to various csv files.

The Excel file QCEW.xlsx compiles these results and constructs the relevant figures. This file is located in the folder Output.

<u>JOLTS programs</u>

altntilde.do: This produces a panel dataset of establishment-level employment that is consistent with the reported gross flows. See Davis et al. (2013). This dataset, named JOLTS_ntilde.dta, is referenced by all of the programs below.

GrowthDist.do: This produces the histograms of quarterly employment growth presented in **panel B of Figure 1** and **Figure B1**. The former is based on the full JOLTS sample. The latter figure has two parts. Panel A of Figure B1 is based on a sample of establishments whose initial employment exceeds 20 (these results are saved to the file with suffix "g20"). Panel B is based on a sample of establishments with positive gross hires (these results are saved to the file with suffix "posh"). Note that, in each sample, we produce histograms for each of the definitions of inaction considered in the paper. Every series name starts with dist, and the suffix refers to the definition of inaction on which is based. The results reported in the figures take inaction to be exact zero net change.

The next two JOLTS-related programs produce profiles of inaction rates by horizon. In each program, the output is a collection of series whose names all begin with inactalt, and the suffix refers to the definition of inaction on which it is based.

- inactdecay.do: This produces the inaction rates by horizon shown in **panel B of Figure 2** and **panels B and F of Figure B2**. The sample for the former is the

universe of establishments. Both establishment- and employment-weighted inaction rates are stored (the former dataset is marked by "estab", the latter by "emp"). The sample for Panel B of Figure B2 consists of establishments whose initial employment exceeds 20, whereas the sample for panel F consists of establishments with positive gross hires over the given horizon. These results are employment-weighted. (The program also produces establishment-weighted results for each of the samples from the latter panels.)

- inactdecay_everytau.do: This produces the profiles of inaction rates by horizon shown in **panel D of Figure B2.** These profiles are employment-weighted. (This file also produces establishment-weighted profiles of inaction rates as well as establishment- and employment-weighted profiles for the subsample of employers with initial employment greater than than 20.)

cumulflows_all_unwt.do: This produces two figures.

- The first is **panel A of Figure 3,** which shows the cumulative turnover rates at horizons one through 24 months conditional on net inaction. The results reported in this figure treat net inaction as exact zero net change, but the program also produces estimates conditional on each of the alternative definitions of net inaction described in the paper. The quits series is denoted by averquits`x', where x refers to the definition of inaction on which it is based (see above). The hires series is denoted by averhires`x', and the all-separations series is avertotsep`x'.

- In addition, this program produces **panel B of Figure B3**. Again, the results displayed in the panel treat net inaction as exact zero net change, but the program considers other definitions. The quits series is denoted by szadj_averquits`x', where x refers to the definition of inaction on which it is based. The series for hires and total separations are named analogously.

cumulflows_all_wt.do: This produces **panel C of Figure B3**. The structure of the dataset and the names of the series follow the same pattern in the .do files described above.

rephires.do: This produces **panel B of Figure 3**, which reports the time series of the replacment hire share. See the main text for details of its definition and construction. The final series is named repquitshare.

The Excel file JOLTS.xlsx compiles these results and constructs the relevant figures. This file is located in the folder Output.

Empirical analyses: Compustat microdata (Table 2 autocorrelation, Tables F1 and F2)

The Stata `.do` file that produces results from Compustat is `compmoments.do`. This file is located in the folder `Compustat programs`.

Part (A) of the code constructs the sample. This removes outliers and applies the filters in Bloom (2009).

Part (B) of the code computes the autocorrelation of log average product.

Part (C) reports the standard deviation of average product among firms that adjust and, separately, do not adjust employment.

Parts (B) and (C) produce the moments reported in the paper. These results are saved to the Excel file, `Compustat-results.xlsx`, which is located in the folder `Output`. There are two tabs. The first reports the autocorrelation estimates in **Figure F1**. The second tab reports the standard deviations in **Figure F2**.

Empirical analyses: Publicly-available aggregate data (Table 3, Figure 9, Footnote 30, Figure D, Appendix H)

The empirical relative standard deviations reported in the final column of **Table 3** are calculated in `Table-3-data.xlsx`. The calculations reported in footnote 30 of the main text are provided in `Footnote-30.xlsx`. Both files are located in the `Output` folder.

Please note that, for the remainder of this document, we will assume that all data files read into programs *and* outputed from programs are housed in `Output`.

The main execution file for the empirical impulse response functions shown in **Figure 9** and **Figure H1** is the MATLAB file `irfregs_main.m`. This file reads in the data file `Table-3-data.xlsx` and selects the specification to be run. The execution file calls the program `irfregs_runols.m` to calculate the least squares estimates and the program `irfregs_makeirfs.m` to compute the impulse responses. These three files are located in the folder `IRF programs`. The results are saved in the Excel file, `IRFs.xlsx`. There is one tab for each figure.

In addition, `irfregs_main.m`. produces impulse responses for the sensitivity analysis described in **Footnote 44**. These IRFs are also stored in `IRFs.xlsx`. There are three tabs associated with this footnote: `Footnote 44 - Six lags`, which presents results based on more than 4 lags of output per worker and the unemployment rate; `Footnote 44 - First diff APL`, which is based on a specification where output per worker enters in first differences rather than log levels; and `Footnote 44 - Lags of JF`

rate, which is based on a specification including (four) lags of the job finding rate in the first and second stage regressions.

The empirical firm-size distribution in **Figure D** in online appendix D is computed in the BDS tab in the Excel file Chains.xlsx. Specifically, this computes the firm-size distribution among firms whose age is under 20 years. This is done for each year 1993-2016. Figure D displays the average of the distribution over these years. (The model-generated size distribution is also restricted to firms under 20 years of age.)

Model simulations (Tables 1, 3, part of 4, and C; Figures 4, 6, 7, 8, D, and G)

The model simulations for these Tables and Figures are generated by running the MATLAB file run_all_scripts.m.

All results are automatically saved in the Excel file Chains.xlsx located in Output. The tabs in this spreadsheet are labeled according to the respective Table and/or Figure that the results correspond to. The MATLAB code populates the cells marked in yellow (overriding their current content). The remainder of the spreadsheet then automatically uses this MATLAB output to produce the relevant Tables and Figures in the paper.

run_all_scripts.m runs the following files:

main.m

This file calibrates the model for the baseline case in Table 1 and all the alternatives reported in Table 4. The file saves the parameters for each calibration into a subfolder: parameters_22 for the baseline calibration (which targets a 22.5% net inaction rate); parameters_OJS for the $C = 0$ model; and parameters_Y (for Y $\in$ {10,17,27,38}) for the robustness exercises in Table 4 which target alternative net inaction rates.

The parameters and equilibrium objects for the baseline steady state are saved in param_ini; analogous outcomes for the steady state with 1% lower aggregate productivity are saved in param_end.

generate_graphs.m

This file calculates all the remaining equilibrium objects and moments. It takes as inputs the parameters and equilibrium objects in the initial (param_ini) and new steady state (param_end) created in main.m. The output is saved in Chains.xlsx.

transition.m

This file solves the transition path for the baseline calibrated model, as well as for all the configurations in Table 4 (except the $C = 0$ case). It takes as inputs the parameters and equilibrium objects in the initial (`param_ini`) and new steady state (`param_end`) created in `main.m`. The file further takes as input a guess for the path of the job-finding rate λ (currently set to the equilibrium path). The output is saved in `Chains.xlsx`.

`transition_OJS.m`

This file solves the transition path for the $C = 0$ model. It takes as inputs the parameters and equilibrium objects in the initial (`param_ini`) and new steady state (`param_end`) created in `main.m`. The file further takes as input a guess for the path of the job-finding rate λ (currently set to the equilibrium path). The output is saved in `Chains.xlsx`.

The next set of files computes model outcomes reported in the Online Appendices:

`main_SA.m`

This file calibrates the alternative model with offer matching described in Appendix C. The file does not take any other files as inputs. It calculates the moments in Table C and saves these in `Chains.xlsx`.

`Firm_age.m`

This file simulates the model-implied firm-size distribution reported in Figure D. It takes as inputs the parameters and equilibrium objects in the initial (`param_ini`) and new steady state (`param_end`) created in `main.m`. The output is saved in `Chains.xlsx`.

`analytic.m`

This file solves for the analytical impulse response function stated in Proposition 3. It takes as inputs the parameters and equilibrium objects in the initial (`param_ini`) and new steady state (`param_end`) created in `main.m`. The output is saved in `Chains.xlsx`, and the results are depicted in Figure G in the Appendix of the paper.

These files in turn use a number of auxiliary files which are self explanatory. The file `model.m` takes as inputs all the parameters (in a structure array) of the model and calculates equilibrium objects (which are also saved as a structure array). Using the output from the model, one can then calculate firm marginal value (denoted J in the paper) using `J_fun.m`, the quit rate (denoted δ in the paper) using `delta_fun.m`, unemployment rate (U/L in the paper) using `u_fun.m`, and so on.

Note: To distinguish more saliently C from c, this code refers to C as `K` and c as `c`. In addition, although not used in the paper, this code also allows for an exogenous separation rate, referred to as `zeta`. Finally, this code refers to the searcher distribution, Γ in the paper, as `q`. (The latter notation is a legacy from the notation used in our companion paper, Elsby and Gottfries 2022.)

Model simulations (Tables 2, remainder of 4, F2, and G; Figures 5, 9, E, and H2)

The model simulations of firm-level outcomes shown in **Tables 2, the remainder of 4, and F2** as well as **Figures 5 and E** are generated by three main execution MATLAB files.

The first file, `simfirms_main_225.m`, executes the baseline calibration and saves results to the Excel file, `Chains.xlsx`. This file will pull the relevant parameter vector from the folder `parameters_22`.

The other two files, `simfirms_main_altC.m` and `simfirms_main_ojs.m`, execute the alternative calibrations reported in Table 4 of the paper. The calibrations are distinguished by the value of the expansion cost, C. The former file ("`altC`") is used to solve the model under alternative, but strictly positive, values of C. This file will pull the relevant parameter vector from the folder `parameters_Y` (for `Y` $\in$ {10,17,27,38}), where `Y` denotes the 4-quarter inaction rate associated with each value of C. The latter file ("`ojs`") is used to solve the model under $C = 0$. This pulls the parameter vector from the folder `parameters_OJS`. The moments reported in Table 4 based on these alternative calibrations are also saved to `Chains.xlsx`. Other output from the alternative calibrations (e.g., the analogue of Figure 5) is saved to the Excel file, `ExtraChains.xlsx`. Each tab stores results for a single figure or table.

Each of these main execution programs references up to 11 files, each of which is described below. However, since the form of the solution under $C > 0$ is different than under $C = 0$, the contents of the files differ between these cases and some files are not needed for the $C = 0$ calibration. Therefore, we have divided the MATLAB programs between two folders. Each folder includes the appropriate version of the files described below. The first folder is `firm_sims_posC`. This collects all files pertaining to calibrations where $C > 0$, including the main execution programs `simfirms_main_225.m` and `simfirms_main_altC.m`. The second folder is `firm_sims_OJS`. This collects all files pertaining to the calibration where $C = 0$, including the main execution program `simfirms_main_ojs.m`.

`solve_main.m`

This file solves the steady state equilibrium. This key outputs are (i) the labor demand policy rules; (ii) the steady state distributions of marginal products; and (iii) equilibrium turnover.

- For (i), `solve_main.m` pulls from three files:

`solve_boundaries.m:` This solves for the boundaries of the natural wastage, replacement, and expansion regions. Note that this file is only used for the $C > 0$ calibrations. (The boundaries are solved straightforwardly within `solve_main.m` for the $C = 0$ model.)

`solve_initialize.m:` This generates an initial guess for the lower boundary (the minimum of the natural wastage region) that is used as an input into `solve_boundaries.m`. This file is only used for the $C > 0$ calibrations.

`solve_nwfun.m:` This solves a version of the natural wastage problem that is found in Elsby and Gottfries (2022). For the $C > 0$ model, this solution helps guide the formation of the initial guess and is, therefore, an input into `solve_initialize.m`.

- For (ii), `solve_main.m` pulls from three files:

`solve_ssdensity.m:` This computes the employment-weighted distribution of marginal products.

`solve_ssdensity_estab.m:` This computes the employer-weighted distribution of marginal products.

`legendre.m:` This calculates the nodes and weights for Legendre quadrature. It is used to make grids for the `ssdensity` programs above.

- For (iii), `solve_main.m` pulls from one file:

`solve_policy.m:` This computes the equilibrium quit and hiring rates using results from the preceding programs.

Each main execution program then simulates employment flows for 5 million firms. The following files calculate moments based on the model-generated panel.

`solve_hockey.m:` This computes the "hockey-stick" figure of gross flows from Davis et al. (2012) that is shown in **Figure 5** of the paper.

`solve_apl.m:` This computes two sets of moments. The first is the standard deviation and autocorrelation of log annual average product in the full (simulated) sample. These

results are reported in **Tables 2 and 4** of the paper. The second is the standard deviation of annual average product growth among net-employment adjusters and, separately, non-adjusters. These results are reported in **Table F2**.

solve_poach.m: This computes the Bagger-Lentz (2019) poaching rank outcomes reported in **Figure E**.

Finally, the results from the simulation of the stochastic model shown in **Figures 9 and H2** and **Table G** are generated by the MATLAB file modelirfs_main.m. The latter in turn references four programs described below. All of these programs are located in the folder, IRF programs.

irfregs_runols.m: This estimates equations (127) and (128) on the model-generated time series.

irfregs_makeirfs.m: This computes impulse responses based on the estimates in irfregs_runols.m.

zirfregs_runols.m: This estimates a version of (128) that uses the simulated TFP series as regressors rather than the average product residuals (which are derived from estimation of (127)).

zirfregs_makeirfs.m: This computes impulse responses based on estimates in zirfregs_runols.m.

Two Excel files are inputs into modelirfs_main.m. The first is Chains.xlsx, which includes the *theoretical* impulse responses. The second is IRFs.xlsx, which includes the *empirical* impulse response of average product that is used to calibrate the stochastic TFP process. The results of the stochastic model simulations are also saved to IRFs.xlsx.

References

Bagger, Jesper, and Rasmus Lentz. 2019. "An Empirical Model of Wage Dispersion with Sorting." *Review of Economic Studies* 86(1): 153-90.

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica* 77(3): 623-685.

Davis, Steven J., R. Jason Faberman and John C. Haltiwanger. 2012. "Labor market flows in the cross section and over time." *Journal of Monetary Economics* 59(1): 1-18.

Davis, Steven J., R. Jason Faberman and John C. Haltiwanger. 2013. “The establishment-level behavior of vacancies and hiring.” *Quarterly Journal of Economics* 128(2): 581-622.

Elsby, Michael W. L. and Axel Gottfries. 2022. “Firm Dynamics, On-the-Job Search and Labor Market Fluctuations.” *Review of Economic Studies* 89(3): 1370–1419.

Fallick, Bruce, and Charles A. Fleischman. 2004. “Employer-to-employer flows in the US labor market: The complete picture of gross worker flows.” Finance and Economics Discussion Series 2004-34, Federal Reserve Board.