

This README file contains descriptions and instructions concerning the data files (Excel) and program files (Matlab) used in the estimation exercises in “Do Data Revisions Matter for DSGE Estimation?”

## **Data**

**Output.xlsx** – This file contains quarterly data, separated into worksheets, on nominal GDP (Nominal GNP-GDP), real GDP (Real GNP-GDP), and the price index (Price Index) defined as the ratio of the former to the latter. The fourth and fifth worksheets compute the log of real GDP and the log of the price index from the raw data as described in the text. The last two worksheets identify the vintages in which revisions to Nominal GNP-GDP and Real GNP-GDP occurred as well as the size of each revision.

**Consumption.xlsx** – This file contains quarterly data, separated into worksheets, on nondurables consumption (PCE-Nondurables), services consumption (PCE-Services), government consumption expenditures and gross investment (GCE), and the price index. The fifth worksheet computes the log of real consumption from the raw data as described in the text. The last three worksheets identify the vintages in which revisions to PCE-Nondurables, PCE-Services, and GCE occurred as well as the size of each revision.

**Investment.xlsx** – This file contains quarterly data, separated into worksheets, on gross private domestic investment (GDI), durables consumption (PCE-Durables qtr), and the price index. Monthly durables consumption data (PCE-Durables mth) are converted to quarterly data by taking 3-month averages. The fifth worksheet computes the log of real investment from the raw data as described in the text. The last two worksheets identify the vintages in which revisions to GDI and PCE-Durables qtr occurred as well as the size of each revision.

**Wage.xlsx** – This file contains quarterly data, separated into worksheets, on hourly compensation in the nonfarm business sector (Hourly Comp NFB) and the price index. The third worksheet computes the log of the real wage from the raw data as described in the text. The last worksheet identifies the vintages in which revisions to Hourly Comp NFB occurred as well as the size of each revision.

**Labor Productivity.xlsx** – This file contains quarterly data on output per hour in the nonfarm business sector (Output per Hour NFB). The second worksheet computes the log of productivity from the raw data as described in the text. The last worksheet identifies the vintages in which revisions to Output per Hour NFB occurred as well as the size of each revision.

**Interest Rate.xlsx** – This file contains monthly data on the effective federal funds rate (Fed Funds Rate mth). The data are converted to a quarterly series by taking 3-month averages in the second worksheet (Fed Funds Rate qtr). The last worksheet identifies the vintages in which revisions to Fed Funds Rate qtr occurred as well as the size of each revision. Although federal funds rate data are never revised, I include this worksheet in the file for completeness.

**Profit.xlsx** – This file contains quarterly data on corporate profits after taxes with inventory and capital consumption adjustments (Corp Profits w IVACC). It also contains quarterly data on the price index. The third worksheet computes the log of real profits from the raw data as described in the text. The last worksheet identifies the vintages in which revisions to Corp Profits w IVACC occurred as well as the size of each revision.

Money Growth.xlsx – This file contains monthly data on the M2 money supply (M2 mth) converted to quarterly data (M2 qtr) by taking 3-month averages. The third worksheet computes the quarterly growth rate of M2 from the raw data as described in the text. The last worksheet identifies the vintages in which revisions to M2 qtr occurred as well as the size of each revision.

Dataset.xlsx – This file collects all of the transformed data from the above files and collects them into worksheets called output, consumption, price level, investment, real wage, labor productivity, fed funds rate, real profits, and money growth. This is the file that is called to the Matlab programs for DSGE estimation. It contains all of the quarterly vintages for each variable needed to estimate the model.

revision matrix.xlsx – This file contains the revision matrix described in the appendix of the paper. It is constructed using information from the individual revision logs for each of the raw data series described above.

Estimation Results.xlsx – This file contains the model-based impulse response functions ( $\Phi$ ) as well as the parameter and standard error estimates (estimates) for each vintage obtained by running the Matlab programs described below. It also contains model-based impulse response functions ( $\Phi_s$ ) and parameter estimates (estimates\_s) for the case in which the sample size is allowed to increase with each vintage date (section 5 of the manuscript). Finally, it also contains the fixed sample (rtmultfs) and variable sample (rtmultvs) real-time policy multipliers described in section 5 and illustrated in Fig. 4 of the text.

## Program Codes

estimate\_VAR.m – This script estimates the structural VAR and implied impulse response functions (with confidence bands) from Christiano, Eichenbaum, and Evans (2005, *JPE*). The script allows the user to choose which data vintage to use for estimation (tau, line 8). It also allows the user to estimate the VAR using a fixed sample period, or a sample period that grows in accordance with the vintage date. The VAR-based impulse response functions along with its corresponding covariance matrix are needed to estimate the structural parameters of the DSGE model.

subordinate codes: recursiveness.m, triang\_fact.m, impulse.m, irfbands\_guassian.m, vec.m, duplication.m, rows.m, vech.m, cols.m, multirandn.m, jbfill.m

estimate\_Var\_loop.m – This script estimates the structural VAR and implied impulse response functions (with confidence bands) from Christiano, Eichenbaum, and Evans (2005, *JPE*). In contrast to the previous script, this one obtains estimates for every quarterly vintage from Feb. 1997 through Nov. 2015 by performing estimation in a loop. Again, it permits the user to estimate the VAR using a fixed sample period, or a sample period that grows in accordance with the vintage date. At the end of the script (lines 161-169 or 172-181), the user has the option of saving the impulse response estimates ( $\Phi_{\text{Hat}}$ ), the variance of each of those estimates ( $V$ ), the lower (LB) and upper bound (UB) of the 90% confidence intervals, as well as the full sample covariance matrix of the impulse response functions ( $W$ ). This information is need in the estimation of the DSGE parameters later on.

subordinate codes: recursiveness.m, triang\_fact.m, impulse.m, irfbands\_guassian.m, vec.m, duplication.m, rows.m, vech.m, cols.m, multirandn.m

`irf_criterion.m` – This script is later called to the main routines to estimate the structural parameters of the DSGE model described in Christiano, Eichenbaum, and Evans (2005, *JPE*). It first specifies the calibrated and estimated parameters. It then sets up the various coefficient matrices. It computes the rational expectations equilibrium to the log-linearized structure, and then obtains the various impulse response functions. The script also calls `PhiHat` and `V` saved from a run of `estimation_VAR_loop.m`. This information is used to construct the DSGE estimation criterion, which is the weighted distance between model and VAR-based impulse response functions (line 126). This criterion ( $J$ ) is the output of `irf_criterion.m` which is minimized by a call to a different script described below.

subordinate codes: `e.m`, `klein.m`, `reorder.m`, `qzswitch.m`

`estimate_params.m` – This script obtains benchmark estimates of the DSGE model as seen in Fig. 3 of the text. It loads `PhiHat` and `V`, feeds them through to `irf_criterion.m`, and then minimizes `irf_criterion.m` over choice of the DSGE model parameters. The script only obtains estimates for one vintage of data at a time. The researcher must record the parameter estimates for each vintage ( $\tau$ , line 12). The estimated parameters along with the minimization criterion ( $J$ ) are recorded for each vintage in the Excel file called `Estimation Results.xlsx` under the worksheet entitled `estimates`. The estimated model-based impulse response functions are also recorded in `Estimation Results.xlsx` for each vintage under the worksheet entitled `Phi`.

subordinate codes: `irf_criterion.m`, `e.m`, `klein.m`, `reorder.m`, `qzswitch.m`

`estimate_stderrors.m` – This script loads the estimated DSGE parameter values recorded in `Estimation Results.xlsx` (for all vintages) as well as `PhiHat`, `V`, and `W` saved from a run of `estimate_VAR_loop.m`. It then computes the standard errors of the parameter estimates for each vintage using the asymptotic delta method by looping through all 76 quarterly vintages. Once computed (it takes some time!) the standard errors are recorded by the researcher alongside the parameter estimates in `Estimation Results.xlsx`.

subordinate codes: `grad_hess.m`, `deriv12_irf_criterion.m`, `numgrad.m`

`param_graph.m` – This script loads the parameter estimates and standard errors recorded in `Estimation Results.xlsx`. It uses them to produce Fig. 3 in the text along with Tables 4 and 5. The commands here do not need to be modified by the user. There are no subordinate codes for `param_graph.m`.

`revision_graph.m` – This script loads the real-time data set used for DSGE estimation across vintages called `Dataset.xlsx`. It then computes the revisions to real investment, money growth, and the real wage (as defined by the model) for the fourth quarter of 1995 as described in section 3 of the text. That is, the calculations track changes in a single data point across vintage time. The script then graphs these revisions to reproduce Fig. 1 and Fig. 2 in the text. The commands here do not need to be modified by the user. There are no subordinate codes for `revision_graph.m`.

## How to obtain the results from section 5 of the manuscript

To obtain estimates of the DSGE model while allowing the sample size to change with the vintage, the user should first return to `estimate.VAR.m`. Make sure you comment out (%) lines 12-23 and uncomment lines 25-36. Then rerun the VAR estimation for the six vintages reported in Table 6 of the text (Feb. 1997, Aug. 1999, Nov. 2003, May 2009, May 2013, Nov. 2015). For each one of these vintages, record and save the VAR impulse response functions (e.g., `PhiHat_99q3` in line 196) the variances of these response functions (e.g., `V_99q3` in line 197), the covariance matrix of the response functions (e.g., `W_99q3` in line 198) and the lower and upper bounds of the 90% confidence intervals (e.g., `LB_99q3` and `UB_99q3` in lines 199-200). Make sure you comment/uncomment the correct lines for all vintage estimations.

Next, open up the script file `estimate_params_samplesize.m`. Uncomment/comment out the appropriate lines depending on what vintage you are estimating. So for example, if you are estimating the Aug. 1999 vintage using the sample period 1965:Q3-1999:Q1, you would uncomment lines 10-16. You can then obtain parameter estimates by minimizing `irf_criterion.m` (line 77) as before and compute standard errors using the asymptotic delta method in lines 85-99.

Parameter estimates and standard errors are then recorded in the Excel file `Estimation Results.xlsx` under the worksheet entitled `estimates_s`. The implied model-based impulse response functions evaluated at the parameter estimates are recorded in the worksheet entitled `Phi_s`.

To obtain estimates of the real-time policy multipliers as shown in Fig. 4 of the text, you will first need to compute the DSGE parameter estimates for ALL vintages assuming (i) the benchmark fixed sample period and (ii) a sample that grows in accordance with vintage date. The fixed sample parameter estimates have already been obtained and are recorded in `Estimation results.xlsx` (`estimates` worksheet). The variable sample parameter estimates are recorded in `Estimation results.xlsx` (`estimates_s` worksheet). These are obtained in two steps. First, use `estimate_VAR_loop.m` to obtain the VAR-based impulse response functions across all vintages while allowing the sample period to vary (so uncomment lines 35-46 and comment out lines 22-33). Record and save estimates of the impulse response functions (`PhiHat_s`), the variances of these response functions (`V_s`), the full covariance matrix of the response functions (`W_s`), and the upper and lower bounds for the 90% confidence bands (`LB_s` and `UB_s`). You will need to uncomment lines 172-181 for this part and comment out lines 161-169. Second, return to `estimate_params.m` and uncomment lines 17-26 but comment out lines 8-15. For each vintage (`tau=1`, line 22), re-estimate the DSGE parameters and record the results in `Estimation Results.xlsx` (`estimates_s`).

To reproduce Fig. 4 in the text, just run the script `rtmultiplier.m`. This file loads the necessary impulse response functions described in section 5 of the text. They are recorded in the worksheets entitled `rtmultfs` and `rtmultvs` in `Estimation Results.xlsx`. These impulse responses are the ones implied by the model for real output to a 50-basis-point increase in the fed funds rate. Of course, the response functions differ over vintage time as the parameter estimates change, and they also differ depending on whether the parameters are estimated using a fixed sample period or a variable sample period. The script `rtmultiplier.m` computes the real-time multipliers as described in the text and graphs the results.