

Guide to Supplemental Material for: *Productivity and Quality in Health Care: Evidence from the Dialysis Industry*

Paul L. E. Grieco* Ryan C. McDevitt†

This document describes the supplemental material needed to reproduce the calculations appearing in the main text. All the data and code discussed below is included in the supplemental directory.

1 Data and Output Files

The data and output files contained in the directory are:

- **prodEstData.csv**: Contains the center-level data on dialysis centers scrapped from CMS reports available at <http://projects.propublica.org/dialysis/>. The reports were systematically downloaded in the summer of 2011 and converted from PDF into text files.
- **finalSample.csv**: Is produced by **GenSumStats.m** and contains a cleaned version of the data (dropping observations with missing variables) to be used in the STATA estimation for table 4.
- **bootRes.mat**: Is produced by **runBoot.m** and contains the primary estimation results (Section 6.1) as well as results from the robustness checks in Section 6.2.
- **bootFlexRes.mat**: Is produced by **runFlexBoot.m** and contains the estimation results for the flexible frontier model (Section 6.3).

2 MATLAB code files

These files were run to produce the output files and tables for the draft using MATLAB version r2016a on a Mac Pro with 12-cores (parallel and optimization toolboxes are required). Full estimation of the model with bootstrapped standard errors takes roughly 8 hours to compute; estimation of the baseline model can be run by calling the **runBoot.m** script from this directory. The following is a description of the MATLAB code files in alphabetical order.

*The Pennsylvania State University, Department of Economics, paul.grieco@psu.edu

†Duke University, The Fuqua School of Business, ryan.mcdevitt@duke.edu

- **backEnvelopeCalcs.m**: Based on computed results, constructs several back of the envelope calculations (e.g., the distribution of elasticities and the cost of infection calculations from the conclusion. A subset of these calculations appear in the body of the paper.
- **cleanCols.m**: Based on the specified specification, removes observations with missing data. This may vary based on which variables are included in the specification.
- **convertToFull.m**: Expands the vector of estimates $\hat{\Phi}$ from the length of observations included in the nonparametric estimation (i.e., those with non-zero hiring) to the full set of observations, using NaN to pad the dropped variables.
- **dec2bigbase.m**: Convert an integer to base B vector, used in **genBasis.m** to construct sieves.
- **GenSumStats.m**: Computes summary statistics, and formatted Tables for Tables 1 and 3 in the main paper. Additional tables also included. Also produces **finalSample.csv** for STATA regression analysis appearing in Section 5.2.
- **getQuality.m**: Computes the quality proxy as described in Section 5.1, also computes the proxy based on the death ratio to be used as an instrument.
- **GMMObj.m**: GMM Objective function for second stage estimation of β parameters when the productivity process is non-parametric (either baseline case or non-parametric robustness check).
- **GMMObjPara.m**: GMM Objective function for second stage estimation of β parameters when the productivity process is $\omega_{jt} = \delta_p + g(\omega_{jt-1}) + \xi_{jt}$.
- **locLin.m**: Locally linear non-parametric estimator used to construct,

$$\hat{E}[y|h_{jt}, i_{jt}, k_{jt}, \ell_{jt}, x_{jt}]$$

and other expectations for Robinson estimator.

- **mainEst.m**: Runs estimator for a single dataset (either full dataset or bootstrap dataset based on arguments) based on specification described by the structure **op**. Also computes ols and fixed effects estimators and returns all results in the matrix **coeff_table**.
- **opFirst.m**: Computes first stage estimate of α_q using sieve approximation for expectations. Legacy code which is not used, use **opFistLL.m** instead.
- **opFirstLL.m**: Computes first-stage estimate of α_q following partially linear estimation procedure describe in Section 5.3.

- `opSecond.m`: Computes second stage estimator of β following procedure described in Section 5.3, calls objective function is `GMMObj.m` or `GMMObjPara.m` based on specification controls.
- `printCoeffTable.m`: Use stored results in `BootRes.mat` to print Table 5.
- `printCoeffTable_flex.m`: Use stored results in `BootFlexRes.mat` to print Table 8.
- `printCoeffTable_gspec.m`: Use stored results in `BootRes.mat` to print Table 7 and additional un-included table on average productivity by type.
- `printCoeffTable_robust.m`: Use stored results in `BootRes.mat` to print Table 6.
- `runBoot.m` Main script for primary estimation and bootstrap inference. Coefficient estimates estimated by calling `runMainEst`. Parallel bootstrap in main body. Output saved to `bootRes.mat`.
- `runFlexBoot.m` Main script for flexible frontier estimation and bootstrap inference. Coefficient estimates estimated by calling `runFlexEst`. Parallel bootstrap in main body. Output saved to `bootFlexRes.mat`.
- `runFlexEst.m` : Computes estimates of the flexible frontier estimation without performing bootstrap inference. Also used as initial call in `runFlexBoot.m` for consistent setup. Essentially a wrapper script for `mainEst.m` setting the appropriate specification flags.
- `runMainEst.m` : Computes estimates of the primary estimation without performing bootstrap inference. Also used as initial call in `runBoot.m` for consistent setup. Essentially a wrapper script for `mainEst.m` setting the appropriate specification flags.
- `setupDialysis.m` : Reads in data from `prodEstData.csv` and sets up data structures for the variables and the specification flags.

3 STATA Files

The following files are used for calculations in Table 2 and the regression analysis presented in Section 5.2 and Table 4.

- `sumStats.do` : Reads in data from `finalSample.csv` and generates the summary statistics in Table 2.
- `regressions.do` : Reads in data from `finalSample.csv` and generates the regressions in Table 4.