Grieco and McDevitt (2017)

Mayur Choudhary

May 24, 2021

Overview

Grieco and McDevitt (2017) estimate the production function of Dialysis care giving centres. The key trade-off faced by the producer is in choosing the quality and quantity. The authors assume a negative marginal rate of transformation between the two output choices, and based on certain assumptions regarding the timing and evolution process of productivity estimate the production function of the dialysis centres. First, I mention all the files that are required to get the results.

Files

1. *finalSample.csv*: This file is the only data file required to run the code.

2. *main.R*: This file is the main function file. This file calls all the functions required to estimate the coefficients in Tables (5),(6),(7),(8) in the paper. Also following the estimation, this file also calls the functions to compute the Bootstrapped standard errors for the estimation.

3. *model.R*: This is the model function that takes in the pre-processed data from the main function to replicate the results in Table 5 and the parametric and non-parametric cases in Table 7 of the original paper. This function calls *firstStage.R* to recover the coefficient α_q and Φ_t. Then it calls *secondStage.R* to compute (β_k, β_l) and hence recover the production function.

4. *modelTab6.R*: This function takes in the data from the main function, and computes the estimates for Table 6. This is similar to the *model.R* file and recovers the production function estimates for the different cases such as not instrumenting for quality,

computing quality from infection rate without controlling for patient characteristics etc.

5. *modelTab8.R*: This function takes in the data from the main function and estimates the production function where the rate of transformation or α depends on the factors of production and would vary across centres. This function outputs the coefficients $(\alpha_q, \alpha_{qk}, \alpha_{ql}, \beta_k, \beta_l)$.

6. *firstStage.R*: This function is called from the model related functions and it calls the *locReg.R* function. This function estimates the α_q coefficient, the trade-off parameter between quantity and quality.

7. *locReg.R*: This function is called from the *firstStage.R* function. It computes the LOWESS (locally weighted scatterplot smoothing) estimates of each point to compute $\hat{E}(z|i_{jt}, h_{jt}, k_{jt}, l_{jt}, x_{jt})$ where $z \in \{y, q, \Phi\}$ the quantity, quality and the recovered $\hat{\Phi}_t$.

8. *secondStage.R*: This function is again called from the model related functions and it is called after the estimation of α_q. This function uses the estimated $\hat{\Phi}_t$ and $\hat{\alpha}_q$ to estimate (β_k, β_l) using GMM. In order to use GMM this function calls *gmmFun.R* function that computes the GMM objective function for the model with two moment conditions.

9. *secondStageTab6.R*: This function is exactly the same as *secondStage* function. However, here I suppress the no quality case and this is specifically for estimates in Table 6.

10. *gmmfun.R*: This function takes the lagged and current values of $(\hat{\Phi}_t, k_t, l_t)$ and $\hat{\alpha}_q$ along with $\boldsymbol{\beta} = (\beta_k, \beta_l)'$ and weight matrix $\boldsymbol{W}$ to compute the GMM objective function.

11. *gmmFunPara.R*: This function is exactly similar to the function *gmmFun*(), however it does not compute the $g(\cdot)$ function according to the type and adds dummy for each type δ_p. This is to estimate the results in column 2 of Table 7 in the paper.

There is a separate folder *Output 24 05 21* that contains the *.csv* files for all the output tables (Table 7 results are in the Table 5 files). There is one file of estimates, one file for the bootstrap sample estimates and one file for the computed standard errors. These are contained in two subfolders marking the number of bootstrap samples, *n.boot* $\in \{10, 100\}$. To get the results simply change the working directory and run the *main*() function file.

Estimation Process

In this section, I briefly review the idea behind the estimation in Grieco and McDevitt (2017). The authors point out that the dialysis centres face a trade-off between treating higher number of patients y and providing a higher quality of care q because increasing q may entail cleaning the equipment to prevent the risk of infection. The dialysis centre solves the following maximization,

$$\max_{y,q} E[\pi(\tilde{y}, \tilde{q}, l, k)]$$

subject to

$$T(y, q) \leq F(k, l, \omega)$$

and

$$\tilde{y} = y + \varepsilon^y$$

$$\tilde{q} = q + \varepsilon^q$$

In the above, the realized quantity and quality $(\tilde{y}, \tilde{q})$ are stochastic. π is the profit of the centre which depends on the realization of quantity, quality, and the factors of production used. The centre can only choose the planned quantity and quality through the transformation function $T(y, q)$ and a production possibility frontier $F(k, l, \omega)$ where ω is the productivity of the centre. The remaining two constraints denote the stochastic nature of quantity and quality $(\tilde{y}, \tilde{q})$. The authors assume a Cobb-Douglas production function and a linear Transformation function,

$$T(y, q) = y + \alpha q$$

and

$$F(k, l, \omega) = \beta_k k + \beta_l l + \omega$$

The goal is to estimate

$$\mathbf{\Theta} = (\alpha_q, \beta_k, \beta_l)'$$

Timing Assumption

The authors point out that the centre makes the output choice after observing the productivity ω_t^q which is unobserved to the econometrician. Therefore straightforward least squares estimation of production function would be biased. Therefore, along the lines of Olley and Pakes (1996) and Ackerberg, Caves and Frazer (2015) the authors make a timing assumption about the productivity and human capital investment decision. Since, the centres have more control over changing the staff than adding more stations the authors observe in the data that investment has very little variation. The change in the paper from Olley and Pakes (1996) is that the authors focus on the hiring decision and invert that to arrive at the productivity instead of the investment decision. The authors assume that the final productivity is ω_t^h and it evolves as

$$\omega_t^h = \omega_t^q + \varepsilon^q$$

The (k_t, l_t) is determined in the previous period, therefore the next period capital and labor will be determined after observing the productivity ω_t^h, which evolves stochastically after the quality decision q_t is made conditional on $(l_t, k_t, x_t, \omega_t^q)$. Here, x_t are incentive shifters and include variables such as indicators for whether the centre is for profit etc. The hiring decision is observed by the econometrician and therefore we can invert it to compute,

$$\omega_t^h = h^{-1}(i_t, h_t, k_t, l_t, x_t)$$

The above assumption implies that the choice of quality q_t is independent of the shock to the final productivity ω_t^h given by ε_t^ω. The authors subset the data to focus on cases with $i_t = 0$, this drops only a few observations as during the sample there are very few firms with non-zero investment.

Productivity Process

The authors assume that the productivity evolves according to a Markov process

$$\omega_t^q = g(\omega_{t-1}) + \xi_t$$

Measurement Error

The quality measure used in the paper is the residual infection rate after controlling for patient characteristics such as age, degree of the disease etc. However, there may be other unobserved shocks to this measurement such as patient specific characteristics etc. which may imply a measurement error while measuring quality using the residual infection rate. This error in variables issue is likely to lead to the attenuation bias in the estimates of coefficient on quality, α_q. The authors address this issue by using the residuals of death ratio variable after controlling for patient characteristics as an instrumental variable.

Estimation Process: Main Specification

First, we estimate the coefficient on quality α_q. Local nonsatiation implies that the budget constraint binds, and therefore we have

$$\tilde{y}_t + \alpha_q \tilde{q}_t = \beta_k k_t + \beta_l l_t + \omega_t^q$$

Substituting for the stochastic components as well as for the productivity term $\omega_t^q = \omega_t^h - \varepsilon_t^\omega$

$$y_t = -\alpha_q q_t + \beta_k k_t + \beta_l l_t + \omega_t^h - \alpha_q \varepsilon_t^q - \varepsilon_t^y - \varepsilon_t^\omega$$

Using the timing assumption,

$$\omega_t^h = h^{-1}(h_t, i_t, k_t, l_t, x_t)$$

And adding the subscript j for each centre j,

$$y_{jt} = -\alpha_q q_{jt} + \underbrace{\beta_k k_{jt} + \beta_l l_{jt} + h^{-1}(h_{jt}, i_{jt}, k_{jt}, l_{jt}, x_{jt})}_{=\Phi_t(h_{jt}, i_{jt}, k_{jt}, l_{jt}, x_{jt})} - \underbrace{(\alpha_q \varepsilon_{jt}^q + \varepsilon_{jt}^y + \varepsilon_{jt}^\omega)}_{=-\varepsilon_{jt}}$$

$$y_{jt} = -\alpha_q q_{jt} + \Phi_t(h_{jt}, i_{jt}, k_{jt}, l_{jt}, x_{jt}) + \varepsilon_{jt}$$

Our assumptions imply that

$$\varepsilon_{jt} \perp\!\!\!\perp q_{jt}$$

Taking expectation conditional on $(h_{jt}, i_{jt}, k_{jt}, l_{jt}, x_{jt})$ implies that Φ_t remains a constant and we can compute $\hat{E}[y_{jt}|h_{jt}, i_{jt}, k_{jt}, l_{jt}, x_{jt}]$ and $\hat{E}[q_{jt}|h_{jt}, i_{jt}, k_{jt}, l_{jt}, x_{jt}]$ using local linear regressions based on Robinson (1988). This is done using the function *locReg()* in the code. We use Instrumental variables (residual death ratio), also demeaned using local linear regression as above to estimate $\hat{\alpha}_q$.
Once we have $\hat{\alpha}_q$ we can recover $\hat{\Phi}_{jt}$

$$\hat{\Phi}_{jt} = y_{jt} + \hat{\alpha}_q q_{jt} - \varepsilon_{jt}$$

and therefore,

$$\hat{\omega}_{jt} = \hat{\Phi}_{jt} - \beta_k k_{jt} - \beta_l l_{jt}$$

Then using the Markov Assumption,

$$\hat{\omega}_{jt} = g(\omega_{jt-1} + \xi_{jt} = g\left(\hat{\Phi}_{jt-1} - \beta_k k_{jt-1} - \beta_l l_{jt-1}\right) + \xi_{jt}$$

We use polynomial of order 4 to estimate the polynomial function $g(\cdot)$. In the paper we use several variations in the specification of this function for results in Table 6 of the paper. Resubsituting for the productivity,

$$y_{jt} + \hat{\alpha}_q q_{jt} = g\left(\hat{\Phi}_{jt-1} - \beta_k k_{jt-1} - \beta_l l_{jt-1}\right) + \beta_k k_{jt} + \beta_l l_{jt} + \underbrace{\varepsilon_{jt} + \xi_{jt}}_{=\eta_{jt}}$$

Finally we use the moment conditions and GMM to estimate $\boldsymbol{\beta} = (\beta_k, \beta_l)'$

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{jt} (k_{jt}\eta_{jt} \quad l_{jt}\eta_{jt}) \ \boldsymbol{W} \ \begin{pmatrix} k_{jt}\eta_{jt} \\ l_{jt}\eta_{jt} \end{pmatrix}$$

In the estimation I use a simple weight matrix $\boldsymbol{W} = \boldsymbol{I}_2$ and identity matrix of order 2.

Results

Table 5

Below I reproduce the results from Table 5. I use only 100 bootstraps samples as the code takes a long time to run otherwise.[1] The standard errors are computed using the bootstrapping procedure and is run through the *main* function file. The estimation results are in variable $result.tab5$ and standard errors in $tab5.boot.se$

[1]The results in this file use even fewer $n.boot = \mathbf{10}$ samples as results for table 6 still took 5 hrs to compile on my machine. I attach the .csv files for $n = 100$ separately for your perusal. The file format is *tab6_boot_se_n.csv* for table 6 SE results with $n \in \{10, 100\}$ bootstrap samples.

	With Quality			Without Quality		
	(1)	(2)	(3)	(4)	(5)	(6)
	Model	OLS	FE	Model	OLS	FE
Quality, $-\alpha_q$	-0.0157	-0.0026	-0.0019			
	(0.0032)	(0.0005)	(0.0005)			
Capital, β_k	0.4717	0.4574	0.5530	0.4335	0.4604	0.5568
	(0.0615)	(0.0257)	(0.2736)	(0.0634)	(0.0256)	(0.2817)
Labour, β_l	0.3178	0.6884	0.1982	0.3708	0.6865	0.1974
	(0.0593)	(0.0216)	(0.0241)	(0.0635)	(0.0216)	(0.0238)

Table 1: Table (5)

The results are in line with the estimates in the original paper for the main specification. However, the Fixed effects estimates are slighly overstated for β_k. The issue is with the subsetting of the data. In a particular case of subsetting sequence (removing entries with NAs), I realize that if I use the full sample of data the OLS estimates are not necessarily much lower than the estimate of α_q in the main model and are of the same order. I think this is somewhat curious. Moreover, my estimates are in line with the authors if I do not remove NAs related to the recovered $\hat{\Phi}_t$ before subsetting the data to account for lags in the second stage estimation. This implies that the OLS and FE estimates should be interpreted with caution as they seem to be quite unstable. I include a note to this effect in the code of *secondStage.R* file.

Table 6

This is simply the main specification estimate for 3 separate cases,

1. **Column 1**: When quality is not instrumented for, i.e. ignoring the measurement error in quality.

2. **Column 2**: When patient characteristics are not controlled for and we simply use the infection rate as the quality measure.

3. **Column 3**: We ignore the competition indicator in the incentive shifter variables x_{jt}. That is competition is not considered as distinctive feature that is exogenous to the model.

The results are presented in Table(6) and they seem to be in line with the estimates in the original paper. Not instrumenting for quality leads to attenuation bias in $\hat{\alpha}_q$ and not including the competition indicator does not affect the estimates by much. The results are in $result.tab6$ and standard errors in $tab6.boot.se$. The standard error for β_k seem to be off by some margin than the original table. This is due to the very low number of bootstrap samples and higher variance in the estimation error for β_k.

	(1) No quality instrument	(2) No patient char. controls	(3) No competition
Quality, -α_q	-0.0119 (0.0008)	-0.0122 (0.0034)	-0.0164 (0.0042)
Capital β_k	0.4713 (0.0677)	0.4938 (0.0745)	0.4643 (0.0605)
Labour β_l	0.3177 (0.0491)	0.3088 (0.0433)	0.2884 (0.0402)

Table 2: Table (6)

Table 7

The two differences are coming from the specification of $g(\cdot)$ function,

$$g(\cdot) = \delta_p + \tilde{g}(\omega_{jt-1})$$

for the parametric case where p is for each type of centre in Column(3) of Table 7 and the other,

$$g(\cdot) = g_p(\omega_{jt-1})$$

for the non-parametric case in Column(2) of Table 7. Below, I show the results of the estimation. The results are in the variable *result.tab5* and standard errors in *tab5.boot.se*

	(1) Baseline	(2) Separable	(3) Nonpara.
Quality, $-\alpha_q$	-0.0157	-0.0157	-0.0157
	(0.0032)	(0.0032)	(0.0032)
Capital β_k	0.4717	0.4621	0.4488
	(0.0615)	(0.0459)	(0.0528)
Labour β_l	0.3178	0.3110	0.3095
	(0.0593)	(0.0610)	(0.0563)

Table 3: Table (7)

Table 8

In this case we modify the transformation function to allow for the differences in the marginal rate of transformation based on the capital-labor ratio of centres, i.e. the slope will no longer be homogeneous and would vary across centres.

$$T(y_{jt}, k_{jt}) = y_{jt} + \underbrace{(\alpha_q + \alpha_{qk} k_{jt} + \alpha_{ql} l_{jt})}_{=\alpha_{qjt}} q_{jt} = y_{jt} + \alpha_q(k_{jt}, l_{jt}) \times q_{jt}$$

The estimates are in the variable *result.tab8* and standard errors in *tab8.boot.se*

	(1) Model	(2) OLS	(3) FE
Quality, $-\alpha_q$	0.0053	0.0005	-0.0115
	(0.0249)	(0.0075)	(0.0047)
Quality $\times$ Capital, α_{qk}	-0.0476	0.0042	0.0059
	(0.0091)	(0.0047)	(0.0028)
Quality $\times$ Labour, α_{ql}	0.0482	-0.0063	-0.0025
	(0.0108)	(0.0027)	(0.0020)
Capital β_k	0.4241	0.4598	0.5182
	(0.0988)	(0.0226)	(0.2920)
Labour β_l	0.4647	0.6854	0.1956
	(0.0695)	(0.0107)	(0.0177)

Table 4: Table (8)

References

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer. 2015. "Identification properties of recent production function estimators." *Econometrica*, 83(6): 2411–2451.

Grieco, Paul LE, and Ryan C McDevitt. 2017. "Productivity and quality in health care: Evidence from the dialysis industry." *The Review of Economic Studies*, 84(3): 1071–1105.

Olley, G Steven, and Ariel Pakes. 1996. "The dynamics of productivity in the telecommunications equipment industry." *Econometrica*, 64(6): 1263.

Robinson, Peter M. 1988. "Root-N-consistent semiparametric regression." *Econometrica: Journal of the Econometric Society*, 931–954.