

# ECO310 - Tutorial 3

## Estimating Production Functions

Francis Guiton

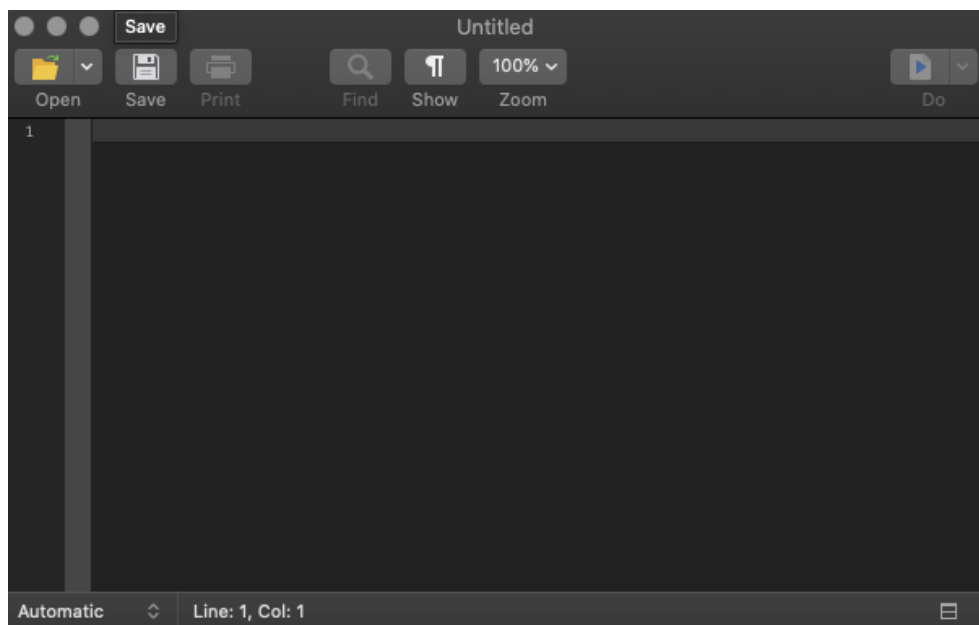
January 25, 2021

In this tutorial, we will consider the Blundell-Bond (2000) panel dataset, which is available on Quercus under **blundell\_bond\_2000\_production\_function.dta**. At any point, if we are unsure about how to use a certain command, we simply type *help command\_name* in the Command Window of Stata.

# 1 Creating a Do-File

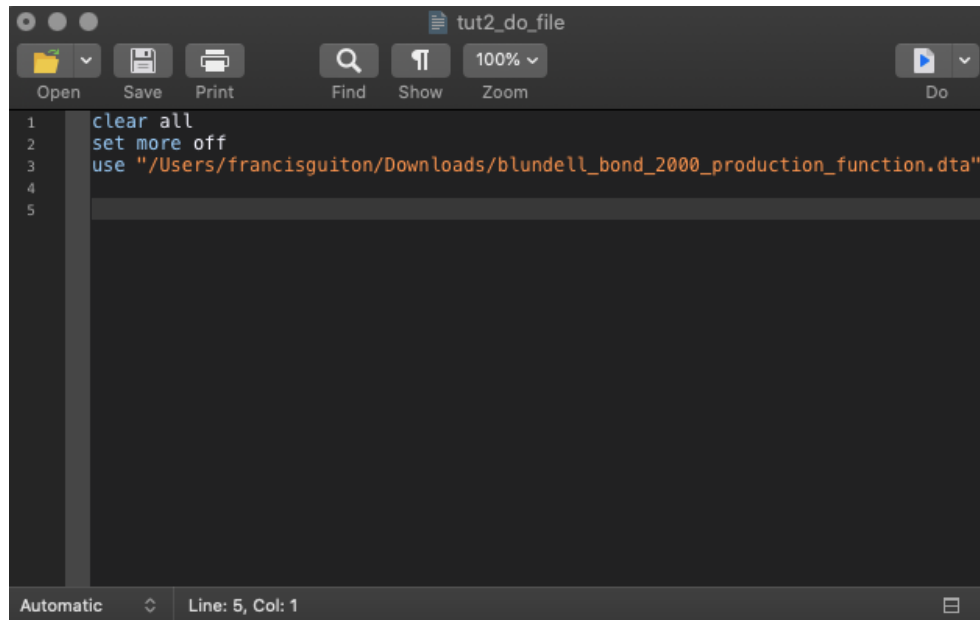
In last week's tutorial, we typed all of our commands in Stata's "Command Window" in succession. This week, we will create a Do-File in which we will write all of the commands we want Stata to execute, and we will save the file to our computer for future reference.

In order to create a Do-File, in Stata we navigate to *File* → *New* → *Do – file*.



In our Do-File window, every line of code that we type will be a command that is executed by Stata. In order to execute a specific line, we highlight our line of code and click on the top right icon *Do*. Alternatively, we can run our whole Do-file by clicking on *Do* without highlighting any lines.

We begin our Do-File by clearing Stata's memory of data with the command *clear all*. In the next line, we can type *set more off* to disable any requests for user interaction when we run the whole Do-File. Next, we will want to load our dataset with the command *use*.



With our Do-File created and our dataset loaded, we will now estimate the production function using various commands that we will input into our Do-File. Note that the dataset we are using is in panel format, with the data being index by *id* and *year*. Therefore, we must tell Stata that our dataset is a panel using the command *xtset*.

```
. xtset id year
    panel variable:  id (strongly balanced)
    time variable:  year, 1982 to 1989
                delta:  1 unit
```

## 2 Ordinary Least Squares

First, similarly to last week's tutorial, we will begin by creating the logarithms of *sales*, *capital*, and *labor*. The logarithm of sales will be our (dependent) variable of interest in our analysis, while the logarithms of capital and labor will be the main explanatory variables.

### Pooled OLS

In order to investigate the relationship between these variables, we run a simple OLS regression of (log) sales against (log) labor and (log) capital.

. reg ln_sales ln_labor ln_capital						
Source	SS	df	MS	Number of obs = 4,072		
Model	15942.9273	2	7971.46365	F(2, 4069) = 63804.90		
Residual	508.360451	4,069	.124934984	Prob > F = 0.0000		
Total	16451.2878	4,071	4.04109255	R-squared = 0.9691		
				Adj R-squared = 0.9691		
				Root MSE = .35346		
ln_sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ln_labor	.560581	.0096412	58.14	0.000	.541679	.5794829
ln_capital	.4298586	.0079525	54.05	0.000	.4142675	.4454498
_cons	3.005052	.0293099	102.53	0.000	2.947588	3.062515

As expected, we obtain positive (and statistically significant) parameter estimates for both labor and capital.

## Time Fixed Effects

In order to control for any time effect in our dataset, we will include year fixed effects in our previous regression using *i.year*. This effectively creates year dummy variables that will control for any time trend in our data. Importantly, year 1982 in the regression output below is omitted. This is because Stata by default uses 1982 as the "reference" or "base" year, against which each dummy variable is compared. Therefore, the parameter estimate for each year dummy is to be interpreted as a relative estimate with respect to year 1982.

. reg ln_sales ln_labor ln_capital i.year						
Source	SS	df	MS	Number of obs	=	4,072
Model	15946.3907	9	1771.82119	F(9, 4062)	=	14254.66
Residual	504.897075	4,062	.124297655	Prob > F	=	0.0000
				R-squared	=	0.9693
				Adj R-squared	=	0.9692
Total	16451.2878	4,071	4.04109255	Root MSE	=	.35256
ln_sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
ln_labor	.5578836	.0098286	56.76	0.000	.5386142	.577153
ln_capital	.4322828	.0081396	53.11	0.000	.4163247	.4482409
year						
1983	-.0568626	.022107	-2.57	0.010	-.1002045	-.0135206
1984	-.050041	.0221342	-2.26	0.024	-.0934362	-.0066458
1985	-.0875714	.0221985	-3.94	0.000	-.1310926	-.0440503
1986	-.092866	.0222691	-4.17	0.000	-.1365256	-.0492063
1987	-.0580931	.0223043	-2.60	0.009	-.1018218	-.0143644
1988	-.0211632	.0223277	-0.95	0.343	-.0649378	.0226114
1989	-.0382923	.0224365	-1.71	0.088	-.0822802	.0056957
_cons	3.046843	.0315266	96.64	0.000	2.985033	3.108652

## Clustered Standard Errors

We notice that our data indexes observations by the variable *id*. This leads us to believe that observations within an *id* group are related, and this correlation should therefore be controlled for. We allow for clusters at the *id* level by adding the option *vce(cluster id)* to our previous regression.

```
. reg ln_sales ln_labor ln_capital i.year, vce(cluster id)
```

Linear regression		Number of obs	=	4,072
		F(9, 508)	=	2507.63
		Prob > F	=	0.0000
		R-squared	=	0.9693
		Root MSE	=	.35256
(Std. Err. adjusted for 509 clusters in id)				

ln_sales	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
ln_labor	.5578836	.0308763	18.07	0.000	.4972227	.6185445
ln_capital	.4322828	.0274846	15.73	0.000	.3782853	.4862803
year						
1983	-.0568626	.0083657	-6.80	0.000	-.0732982	-.0404269
1984	-.050041	.0110933	-4.51	0.000	-.0718355	-.0282465
1985	-.0875714	.0135255	-6.47	0.000	-.1141442	-.0609987
1986	-.092866	.016461	-5.64	0.000	-.125206	-.0605259
1987	-.0580931	.0174944	-3.32	0.001	-.0924634	-.0237228
1988	-.0211632	.0185846	-1.14	0.255	-.0576754	.015349
1989	-.0382923	.020265	-1.89	0.059	-.0781058	.0015213
_cons	3.046843	.0915369	33.29	0.000	2.867005	3.22668

## Lagged Explanatory Variables

We hypothesize that our dependent variable *ln\_sales* may depend on past values of the labor and capital inputs. That is, the effects of increasing labor and capital on output may occur with a delay. Therefore, we include lagged values of *ln\_labor* and *ln\_capital* as explanatory variables, as well as the lagged value of the dependent variable *ln\_sales*.

```
. reg ln_sales ln_labor ln_capital L.ln_sales L.ln_labor L.ln_capital i.year, vce(cluster id)
```

Linear regression

Number of obs	=	3,563
F(11, 508)	=	75113.93
Prob > F	=	0.0000
R-squared	=	0.9949
Root MSE	=	.1426

(Std. Err. adjusted for 509 clusters in id)

ln_sales	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
ln_labor	.4789653	.0288963	16.58	0.000	.4221944 .5357362
ln_capital	.2348098	.0352859	6.65	0.000	.1654854 .3041341
ln_sales L1.	.9216418	.0105283	87.54	0.000	.9009575 .9423261
ln_labor L1.	-.4233997	.0305806	-13.85	0.000	-.4834797 -.3633198
ln_capital L1.	-.2120621	.0347141	-6.11	0.000	-.280263 -.1438613
year					
1984	.0661134	.0096619	6.84	0.000	.0471313 .0850956
1985	.0127151	.0101712	1.25	0.212	-.0072678 .032698
1986	.034839	.0102827	3.39	0.001	.0146373 .0550408
1987	.0767564	.0086928	8.83	0.000	.0596782 .0938346
1988	.0779253	.0094209	8.27	0.000	.0594167 .096434
1989	.0356445	.0097659	3.65	0.000	.0164579 .0548311
_cons	.246604	.0322971	7.64	0.000	.1831516 .3100564

### 3 Panel Regression

As we are considering a panel dataset (and we have already declared to Stata that our data is a panel), we will now perform various panel regressions to investigate the relationship between *ln\_sales* and the two inputs *ln\_labor* and *ln\_capital*.

#### Fixed Effects (Within-Group Estimator)

We begin by performing a fixed effects panel regression using the command *xtreg*, with option *fe*. That is, we believe that there is group-specific unobserved heterogeneity in the error term of our model, and we therefore want to control for this (fixed) effect.



```
. xtreg ln_sales ln_labor ln_capital L.ln_sales L.ln_labor L.ln_capital i.year, fe vce(cluster id)
```

Fixed-effects (within) regression

Group variable: id

Number of obs = 3,563

Number of groups = 509

R-sq:

within = 0.7825

between = 0.9879

overall = 0.9847

Obs per group:

min = 7

avg = 7.0

max = 7

F(11,508) = 345.86

Prob > F = 0.0000

corr(u\_i, Xb) = 0.7191

(Std. Err. adjusted for 509 clusters in id)

ln_sales	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
ln_labor	.4880013	.0299207	16.31	0.000	.4292178	.5467848
ln_capital	.1765454	.0340853	5.18	0.000	.1095799	.2435109
ln_sales L1.	.4039344	.0293221	13.78	0.000	.3463269	.4615418
ln_labor L1.	-.0231194	.03455	-0.67	0.504	-.0909977	.044759
ln_capital L1.	-.1305487	.0252882	-5.16	0.000	-.1802311	-.0808663
year						
1984	.0564054	.0074048	7.62	0.000	.0418576	.0709533
1985	.0271379	.0092935	2.92	0.004	.0088795	.0453963
1986	.0494812	.0116301	4.25	0.000	.0266322	.0723302
1987	.1033078	.0116548	8.86	0.000	.0804104	.1262053
1988	.1310847	.01305	10.04	0.000	.1054461	.1567234
1989	.1174383	.0150838	7.79	0.000	.087804	.1470726
_cons	2.625541	.1591593	16.50	0.000	2.31285	2.938233
sigma_u	.31731619					
sigma_e	.12076713					
rho	.87347826	(fraction of variance due to u_i)				

## Difference GMM

We suspect that there is endogeneity in our model. That is, we believe that the explanatory variables are correlated with the error term. We will therefore use an instrumental variable approach to deal with this endogeneity problem.

First, we install the *xtabond2* command - the user-created Stata command for the Arellano–Bond dynamic panel data estimators. We type *ssc install xtabond2* in the Command Window of Stata.

```
. ssc install xtabond2
checking xtabond2 consistency and verifying not already installed...
installing into /Users/francisguiton/Library/Application Support/Stata/ado/plus/...
installation complete.
```

We will now run the Difference GMM estimation, which effectively considers the regression equation in first differences, and instruments the endogenous variables with their higher lagged values (in this case, higher than the first lag). We type *xtabond2* followed by our usual dependent and independent variables, and add the option *gmm()*. We include in the parentheses the set of endogenous variables in our model, and the lowest lag of these variables that is not correlated with the error term. We add the option *iv()* in which we list the strictly exogenous variables, and *robust* which tells Stata that we wish to have robust standard errors. Finally, we include *noleveleq* to specify that we wish to implement the Difference GMM estimator (and not the System GMM estimator which we will explore in the next section):

```
xtabond2 ln_sales ln_labor ln_capital L.ln_sales L.ln_labor L.ln_capital i.year,  
gmm(ln_sales ln_capital ln_labor, lag(2 .)) iv(i.year) robust noleveleq
```

Dynamic panel-data estimation, one-step difference GMM						
Group variable: <b>id</b>	Number of obs	=	<b>3054</b>			
Time variable : <b>year</b>	Number of groups	=	<b>509</b>			
Number of instruments = <b>69</b>	Obs per group: min	=	<b>6</b>			
Wald chi2(0) = <b>.</b>	avg	=	<b>6.00</b>			
Prob > chi2 = <b>.</b>	max	=	<b>6</b>			
<b>ln_sales</b>	<b>Coef.</b>	<b>Robust Std. Err.</b>	<b>z</b>	<b>P&gt; z </b>	<b>[95% Conf. Interval]</b>	
<b>ln_labor</b>	<b>.5128697</b>	<b>.0892432</b>	<b>5.75</b>	<b>0.000</b>	<b>.3379563</b>	<b>.6877831</b>
<b>ln_capital</b>	<b>.1318291</b>	<b>.1179029</b>	<b>1.12</b>	<b>0.264</b>	<b>-.0992562</b>	<b>.3629145</b>
<b>ln_sales</b>						
<b>L1.</b>	<b>.3264209</b>	<b>.0521606</b>	<b>6.26</b>	<b>0.000</b>	<b>.2241881</b>	<b>.4286538</b>
<b>ln_labor</b>						
<b>L1.</b>	<b>.0726061</b>	<b>.0927269</b>	<b>0.78</b>	<b>0.434</b>	<b>-.1091353</b>	<b>.2543475</b>
<b>ln_capital</b>						
<b>L1.</b>	<b>-.2066298</b>	<b>.0949589</b>	<b>-2.18</b>	<b>0.030</b>	<b>-.3927458</b>	<b>-.0205137</b>
<b>year</b>						
<b>1983</b>	<b>-.1552384</b>	<b>.0224468</b>	<b>-6.92</b>	<b>0.000</b>	<b>-.1992334</b>	<b>-.1112435</b>
<b>1984</b>	<b>-.0876288</b>	<b>.0172895</b>	<b>-5.07</b>	<b>0.000</b>	<b>-.1215155</b>	<b>-.0537421</b>
<b>1985</b>	<b>-.0987342</b>	<b>.0126165</b>	<b>-7.83</b>	<b>0.000</b>	<b>-.1234621</b>	<b>-.0740063</b>
<b>1986</b>	<b>-.0639394</b>	<b>.00849</b>	<b>-7.53</b>	<b>0.000</b>	<b>-.0805794</b>	<b>-.0472994</b>
<b>1988</b>	<b>.0376223</b>	<b>.00671</b>	<b>5.61</b>	<b>0.000</b>	<b>.024471</b>	<b>.0507736</b>
<b>1989</b>	<b>.0347582</b>	<b>.0108629</b>	<b>3.20</b>	<b>0.001</b>	<b>.0134673</b>	<b>.0560491</b>
Instruments for first differences equation						
Standard						
D.(1982b.year 1983.year 1984.year 1985.year 1986.year 1987.year 1988.year 1989.year)						
GMM-type (missing=0, separate instruments for each period unless collapsed)						
L(2/7).(ln_sales ln_capital ln_labor)						
Arellano-Bond test for AR(1) in first differences: z = <b>-6.21</b> Pr > z = <b>0.000</b>						
Arellano-Bond test for AR(2) in first differences: z = <b>-1.36</b> Pr > z = <b>0.173</b>						
Sargan test of overid. restrictions: chi2(58) = <b>213.56</b> Prob > chi2 = <b>0.000</b>						
(Not robust, but not weakened by many instruments.)						
Hansen test of overid. restrictions: chi2(58) = <b>98.39</b> Prob > chi2 = <b>0.001</b>						
(Robust, but weakened by many instruments.)						
Difference-in-Hansen tests of exogeneity of instrument subsets:						
iv(1982b.year 1983.year 1984.year 1985.year 1986.year 1987.year 1988.year 1989.year)						
Hansen test excluding group: chi2(52) = <b>78.58</b> Prob > chi2 = <b>0.010</b>						
Difference (null H = exogenous): chi2(6) = <b>19.81</b> Prob > chi2 = <b>0.003</b>						

The Sargan test of overidentifying restrictions presents the null hypothesis that the instruments are exogenous as a group. According to the above output, we reject this null hypothesis, which therefore entails that we may have a poor set of instruments.

## System GMM

With the same command *xtabond2*, we can implement the System GMM estimator. This estimator is the augmented version of the previous Difference GMM estimator. Broadly speaking, the System GMM estimator considers both the first-differenced regression equation and the regression equation in levels, and therefore produces two sets of instruments for the endogenous variables. In this case, we exclude the option *noleveleq* to tell Stata we wish to implement System GMM, and we add the option *h(1)* to specify the one-step estimator to be used - here we consider 2SLS. Finally, we will consider the third lags of the endogenous variables (and earlier) as instruments for the equation in first differences, and therefore include *lag(3 .)*.

```
xtabond2 ln_sales ln_labor ln_capital L.ln_sales L.ln_labor L.ln_capital i.year,  
gmm(ln_sales ln_capital ln_labor, lag(3 .)) iv(i.year) robust h(1)
```

Dynamic panel-data estimation, one-step system GMM						
Group variable: <b>id</b>	Number of obs		=	3563		
Time variable : <b>year</b>	Number of groups		=	509		
Number of instruments = 67	Obs per group: min		=	7		
Wald chi2(11) = <b>1.08e+06</b>	avg		=	7.00		
Prob > chi2 = <b>0.000</b>	max		=	7		
ln_sales	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ln_labor	.4650059	.1128788	4.12	0.000	.2437675	.6862443
ln_capital	.4070006	.1522005	2.67	0.007	.1086931	.7053082
ln_sales L1.	.6087156	.0997699	6.10	0.000	.4131701	.8042611
ln_labor L1.	-.2814321	.1208803	-2.33	0.020	-.5183532	-.044511
ln_capital L1.	-.2168042	.1194903	-1.81	0.070	-.4510008	.0173925
year						
1983	-.0609928	.0158425	-3.85	0.000	-.0920436	-.0299421
1984	-.0187287	.0169831	-1.10	0.270	-.0520149	.0145574
1985	-.0671562	.0129478	-5.19	0.000	-.0925333	-.041779
1986	-.0537623	.0122459	-4.39	0.000	-.0777637	-.0297608
1987	-.0144227	.0094953	-1.52	0.129	-.0330332	.0041878
1989	-.0398253	.0095249	-4.18	0.000	-.0584936	-.0211569
_cons	1.144956	.3606024	3.18	0.001	.4381879	1.851723
Instruments for first differences equation						
Standard						
D.(1982b.year 1983.year 1984.year 1985.year 1986.year 1987.year 1988.year 1989.year)						
GMM-type (missing=0, separate instruments for each period unless collapsed)						
L(3/7).(ln_sales ln_capital ln_labor)						
Instruments for levels equation						
Standard						
1982b.year 1983.year 1984.year 1985.year 1986.year 1987.year 1988.year 1989.year						
_cons						
GMM-type (missing=0, separate instruments for each period unless collapsed)						
DL2.(ln_sales ln_capital ln_labor)						

```

Arellano-Bond test for AR(1) in first differences: z = -6.69 Pr > z = 0.000
Arellano-Bond test for AR(2) in first differences: z = -0.45 Pr > z = 0.654

Sargan test of overid. restrictions: chi2(55) = 58.04 Prob > chi2 = 0.364
(Not robust, but not weakened by many instruments.)
Hansen test of overid. restrictions: chi2(55) = 75.48 Prob > chi2 = 0.035
(Robust, but weakened by many instruments.)

Difference-in-Hansen tests of exogeneity of instrument subsets:
GMM instruments for levels
Hansen test excluding group: chi2(40) = 46.20 Prob > chi2 = 0.232
Difference (null H = exogenous): chi2(15) = 29.28 Prob > chi2 = 0.015
iv(1982b.year 1983.year 1984.year 1985.year 1986.year 1987.year 1988.year 1989.year)
Hansen test excluding group: chi2(49) = 64.06 Prob > chi2 = 0.073
Difference (null H = exogenous): chi2(6) = 11.42 Prob > chi2 = 0.076

```