Faculty Economics & Business, Amsterdam School of Economics
University of Amsterdam

Testing constant returns to scale for the Cobb-Douglas function

Assignment AE2 Week 4: Panel data methods

Vien Dinh, Derek Dirks & Jorrim Prins

11002115, 11029633 & 11038934

30-01-2020

Abstract

A Cobb-Douglas production function with auto-regressive properties is tested for constant returns to scale. Lagged regressors are used as instruments in a two-step GMM model to estimate the parameters and constant returns to scale are rejected.

1 Introduction

The Cobb-Douglas production function is one of the most commonly known production functions to this day. It has been used for a lot of research and its details have been discussed thoroughly. Griliches and Mairesse (1995) estimate a Cobb-Douglas function by OLS, find reasonable parameter estimates and explore the errors. Biases that still exists in these estimations are not believed to be solvable by OLS, therefore Blundell and Bond (2000) explore GMM options in a comparable model. This research follows on aforementioned papers by further exploring the Cobb-Douglas model with data on sales, employment and capital stocks. The objective is to closely approximate the specifications of the true model, such that a plausible test for the returns to scale can be performed.

The base model estimated in this paper leads to endogeneity problems, in the form of the lagged dependent variable $y_{i,t-1}$ and unobserved productivity. To mitigate this obstacle, this research turns to internal instrumental variable estimation with panel data techniques. The two-step GMM estimator is utilised, with first differences to account for the unobserved productivity.

2 Models and Techniques

The simple Cobb-Douglas formulation where logarithms of the variables are taken, with y_{it} log sales, l_{it} log employment and k_{it} log capital stock, looks as follows:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + \alpha_i + \lambda_t + v_{it} \tag{1}$$

α_i is a vector of individual specific effects. To solve problems caused by auto-regressive productivity shock (v_{it}), the model is rewritten and reparameterized to find (2), which allows for valid internal instruments. This model is then estimated by taking first differences (for removal of individual specific effects) and performing two-step GMM to account for endogeneity and/or predeterminedness of the regressors. The model uses Windmeijer (2005) corrected standard errors.

$$y_{it} = \pi_1 l_{it} + \pi_2 l_{i,t-1} + \pi_3 k_{it} + \pi_4 k_{i,t-1} + \pi_5 y_{i,t-1} + \alpha_i^* + \lambda_t + \varepsilon_{it} \tag{2}$$

This is subject to $-\pi_2 = -\pi_1\pi_5$ and $-\pi_4 = -\pi_3\pi_5$. The returns to scale parameter that is tested to be equal to 1 is calculated by adding β_l and β_k from (1) or $\frac{\pi_1+\pi_2}{1-\pi_5}$ and $\frac{\pi_3+\pi_4}{1-\pi_5}$ from (2).

Arellano and Bond (1991) tests are evaluated to rule out residual auto-correlation and make sure internal instruments may be used. Hansen tests determine the validity of instruments and test over-identifying instruments, whereas the Kleibergen and Paap (2006) LM statistic tests for under-identification. The most important test in this paper however is the Wald test for $H_0 : \mathit{Returns\ to\ Scale} = \beta_l + \beta_k = 1$.

3 Results

As benchmark, the OLS is estimated on this dataset. From the ordinary least squares estimation output in the appendix, the conclusion may be drawn that accounting for heterogeneity by using fixed effect (within) estimation significantly impacts the model parameters and standard errors. On the contrary, accounting for time effects does not affect the model and seems to be irrelevant, although we still include them in further models. The 4th column shows that exploiting clustered errors does impact the OLS model as it significantly increases the standard errors of the parameters.

GMM/IV regression is subsequently performed with second lags of y, l and k and third lags of l and k used as instruments. This results in a p-value of 0.5274 by the Kleinbergen-Paap test which indicates that stronger instruments would be preferred. More lags of the mentioned regressors are used and a stronger Kleinbergen-Paap statistic is the result (0.8858).

A more effective way to increase the number of instruments is recycling, which is the next procedure that is followed. As the parameter π_5 is strongly identified (0.4180778 with standard error of 0.090212), we collapse the internal instruments following from Δy_{t-1} to increase the p-value of the validity test by Hansen. The p-value is increased from 0.155 to 0.475 which implies valid instruments. Finally, a third recycled lagged instrument from l is added to better identify its parameters. The estimated model still satisfies Hansen and Arellano-Bond tests with p-values of 0.578 and 0.475 respectively. Performing a test on the non-linear hypothesis:

$$H_0 : Returns\ to\ Scale = \frac{\pi_1 + \pi_2 + \pi_3 + \pi_4}{1 - \pi_5} = 1$$

shows a Wald statistic with the value 8.28, meaning the hypothesis is rejected as the statistic is chi-squared distributed with one degree of freedom.

4 Conclusion

The Cobb-Douglas model that includes a productivity shock of auto-regressive nature cannot be consistently estimated by ordinary least squares regression. The model has to be rewritten and several lagged terms of the regressors have to be added as internal instruments to solve the endogeneity and predeterminedness problems. Using recycled instruments of second and third order lags, depending on the variable, yields a two-step GMM estimation model that can be used to test for constant returns to scale. This research concludes that the hypothesis of constant returns to scale has to be rejected in this model.

Table 1: Pooled OLS regression

π_1	0.78	0.79	0.80	0.80
	(0.04)	(0.04)	(0.04)	(0.11)
π_3	0.14	0.08	0.06	0.06
	(0.03)	(0.03)	(0.03)	(0.03)
π_5	0.82	0.35	0.35	0.35
	(0.02)	(0.03)	(0.03)	(0.05)
Fixed effects	-	yes	yes	yes
Time fixed effects	-	-	yes	yes
Clustered standard errors	-	-	-	yes

Note: Regression coefficient estimate with the standard error in parentheses.

Table 2: Instrumental variable estimates

π_1	-1.10	-1.10	-0.66	0.35	0.34	0.38
	(2.66)	(2.86)	(1.32)	(0.18)	(0.20)	(0.16)
π_3	0.08	0.08	0.22	-0.04	-0.05	-0.05
	(0.87)	(1.20)	(0.85)	(0.08)	(0.12)	(0.12)
π_5	0.64	0.64	0.89	0.42	0.39	0.38
	(0.40)	(0.25)	(0.33)	(0.09)	(0.08)	(0.07)
Robust & Cluster errors	-	yes	yes	yes	yes	yes
Extra lagged instr.	-	-	yes	-	-	-
Recycled instruments*	-	-	-	yes	yes	yes
Collapsed instruments	-	-	-	-	yes	yes
Recycled instruments**	-	-	-	-	-	yes
Hansen test	-	-	0.394	0.155	0.475	0.578
Arellano-Bond test***	-	-	-	0.439	0.441	0.475

Note: Regression coefficient estimate with the standard error in parentheses. The reported values of the tests are the p-values.

*Second lags as recycled instruments for l, k and y, third lags as recycled instruments for l

**Second lagged values as recycled instruments for l

***AR(2)

References

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, *58*(2), 277–297.

Blundell, R., & Bond, S. (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric reviews*, *19*(3), 321–340.

Griliches, Z., & Mairesse, J. (1995). *Production functions: the search for identification* (Tech. Rep.). National Bureau of Economic Research.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics*, *133*(1), 97–126.

Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step gmm estimators. *Journal of econometrics*, *126*(1), 25–51.