D Econometric issues and sensitivity testing

Available BLS data for the four-year period 1994-95 to 1997-98 provides a basis for controlling for firm-specific and dynamic factors using panel regression methods in the basic and augmented modes reported in chapter 4. This appendix outlines the choice of panel regression estimator adopted in work to date and some contemporary panel data estimation techniques.

D.1 Panel estimation and choice of estimator

There has been an increasing use of sophisticated panel methods by empirical researchers to account for various biases and other disturbances in the regression analysis. A panel model can be represented as:

$$\dot{y} = \beta_0 + \beta x_{it} + \mu_i + \lambda_t + m_{it} + \upsilon_{it} \quad i = 1,..., N \qquad t = 1,..., T$$

where β_0 is unidentified technical change, x_{it} are independent variables. The latter four terms are related to the components of the error term in the regression. The assumptions of correlation between the errors and the independent variables determine the appropriate estimator for panel regressions. The first term (μ_i) represents firm-specific effects. For example, efficiency or comparative advantage known to the firm but not otherwise specified that affect the input choices of the firm. The second term (λ_t) represents time-specific effects such as business cycles and trend growth that affect all firms in the sample. The third term (m_{it}) represents net errors in measurement of variables (eg, quality of labour, utilization of capital and decline of ICT prices not accounted for in the data). The final term (υ_{it}) represents pure random error and does not affect input choices.

The means of controlling for unobservable omitted variables depends upon whether a fixed effects or random effects model is used. A fixed effects (within-firm) estimator is typically chosen when unobservable firm-specific effects are assumed to be *correlated* with the observed explanatory variable(s). In this set up, the within transformation of the fixed effects estimator eliminates firm-specific effects and thereby lessens possible simultaneity bias. A random effects estimator is typically chosen when unobservable firm-specific effects are assumed to be *uncorrelated* with the observed explanatory variables.

A choice was made between fixed or random effects panel estimators in the current study on the basis of formal statistical tests (ie the Hausman specification test and Breusch Pagan LM test) and the sign and magnitude of parameters that could be independently validated. In particular, the coefficient on capital intensity was compared to the capital share in value added as reported in the Australian national accounts adjusted for owner operator. A parameter value similar to independent estimates of the sectoral capital share reported in Australian national accounts indicated that constant returns to scale was not rejected by the data, and it has provided some overall confidence in model specification and estimates.

Another reason for preferring the random effects over fixed effects estimators is that the explanatory power of variables, particularly ICT variables, is less likely to be diminished by the presence of a relatively large set of dummy variables included in the fixed effects. In addition, random effects models by taking into account within and between firm variability, provide estimates with greater precision than the fixed effect models.

D.2 Sources of bias and recent development

Productivity estimates obtained using regression analysis may be subject to a number of biases. Key sources of bias include, endogeneity, which leads to simultaneous equation bias; and omitted variable bias. In addition, input and output measures are subject to measurement error and leads and lags can exist among variables (including concerns relating to the valuation of flows).

There have also been a number of recent developments in panel data estimation techniques which are discussed in section D.1. To give a flavour of these developments and an indication of the sensitivity of our results to different estimators, some key econometric issues and sensitivity tests for basic growth model are applied for eight sectors.

However, biases can also be pertinent to models estimated in growth rates, where current year productivity *growth* is influenced by the initial level of productivity (or other state variables). In our growth regressions, we have attempted to incorporate some dynamics and correct for bias by the adoption of the partial adjustment model mentioned above. However, there could remain residual specification problems due to simultaneity and mis-specified dynamics, and measurement error.

To correct for any outstanding specification problems, we are investigating dynamic panel analysis. The method we have used so far is referred to as 'first differenced Generalised Method of Moments (GMM)' estimation (Arellano and Bond 1991) and is applied using algorithms in STATA 7. Box D.1 provides an outline of dynamic panel analysis and methods currently suggested in the literature on growth regressions. The first differenced GMM technique was applied to eight sectors of the basic growth model and results are reported in table D.1.

The relation between duration of computer use and labour productivity growth is sensitive to model specification, although the differing levels of statistical significance on the items has implications for interpretation of model results.

Box D.1 **An outline of dynamic panel data models**

Traditional panel techniques provide a means of including firm-specific factors. However, they do not correct the bias introduced by the presence of correlation between lagged dependent variables and firm specific-effects.

There have been a number of developments to address this source of bias, for example:

- a first difference transformation which imposes independence between the lag dependent variable and unobserved firm-specific effects (Anderson and Hsiao 1981). The instrumental variable method used leads to consistent but not necessarily efficient estimates of parameters. The method is a restricted form of the 'growth model' adopted in this study (see equation in box 4.2). It was applied to the data with very similar results to the unrestricted model.
- the first differenced Generalised Method of Moments (GMM) uses a different instrumental variable method to the first difference transformation (dot point 1) and was found to have no singularities and much smaller variances than characterised the earlier method (Arellano and Bond 1991). However, the method is highly sensitive to the persistence of data and initial conditions (the state variables). Application of the method to short panels (such as the ABLS) can result in biased parameter estimates. The method was applied as a sensitivity test in this study (see table D.1).
- the system GMM estimator introduces additional moment conditions to eliminate the identification problem characterising the first differenced GMM estimator (Arellano and Bover 1995, Blundell and Bond 1998). The additional conditions are also supposed to lessen the effect of measurement errors (including those owing to lack of price index data) and endogenous explanatory variables. The method is specifically designed for panel data sets such as the ABLS. We are considering the application of this method for sensitivity testing.

These modelling approaches further the contribution of the error correction model (ECM) which takes into account the fact that a change in labour productivity consists of a change associated with a movement along a long-run path plus part of the deviation of the firm from its 'equilibrium'.

Source: Anderson and Hsiao (1981), Arellano and Bond (1991), Arellano and Bover 1995, Blundell and Bond (1998), Griliches and Mairesse (1998) and Bond (2002).

Table D.1 **Sensitivity test for determinants of labour productivity**

Unweighted panel

	Manufacturing		*Construction*		*Wholesale trade*		*Retail trade*		*Accommodation, cafes & restaurants*		*Transport & storage*		*Property and business services*		*Cultural and recreational services*	
	GLS	1st GMM	GLS	1st GMM	GLS	1st GMM	GLS	1st GMM	GLS	1st GMM	GLS	1st GMM	GLS	1st GMM	GLS	1st GMM
Intercept	1.49***	0.17***	2.41***	0.30***	1.82***	0.11***	1.68***	0.12***	1.33***	0.324***	1.53***	0.14**	1.76***	0.31***	0.52***	0.08
Log(y-1)	-0.42***		-0.64***		-0.48***		-0.48***		-0.39***		-0.42***		-0.45***			
Change log(y_{-1}) **a**		-0.21***		-0.30***		-0.16***		-0.29***		-0.321***		-0.18*		-0.20***	-0.22***	-0.20***
Change log(k) (γ_k)	0.40***	0.38***	0.27***	0.001#	0.42***	0.47***	0.39***	0.48***	0.33***	0.052	0.39***	0.44***	0.33***	0.31***	0.30***	0.011
Size ('00)	0.05*	-0.80***	0.21*	-1.57***	0.02	-0.48**	-0.23***	-0.14***	-0.20#	-0.845***	0.09	-0.02	-0.04	-1.00**	-0.23#	-3.64***
ICT_d1(< 2yrs) **b**	0.12***	0.14*	-0.10#	-0.07	-0.01	-0.12	0.12**	0.03	0.07	0.102	0.13#	-0.13	-0.03	0.05	-0.15	-0.24
ICT_d2(short)	0.22***		0.32***		0.28***		0.20***		0.18#		0.17#		0.14#		0.54**	
ICT_d3 (2-5yrs)	0.13***	0.08	0.06	-0.22	0.14*	0.24	0.12***	-0.08	0.19***	0.490#	0.10#	-0.10	-0.01	0.32	0.32*	1.80***
ICT_d4(> 5yrs)	0.05*		-0.13#		-0.03		0.05#		0.11#		0.01		-0.15		0.21#	
Net Access	0.06***	0.02	0.06	-0.03	0.08***	0.004	0.07*	0.05	0.05	-0.118	0.07	0.11	0.13***	0.02#	0.16#	-0.22
Model summary																
No.of firms **c**	1790	1249	289	151	738	511	549	331	207	116	192	104	652	363	92	40
Wald χ^2 **d**	1783.7	243	656.4	31.68	1738.8	115.2	1242.7	302.6	326.1	92.86	415.7	67.3	1329.7	163.2	91.7	482.1
Sargan test (*p value)* **e**		1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00

*** significant at 1% level, ** 5% level, * 10% level, # relevant the t-statistic is more than or equal to one. **a** Instead of log level of labour productivity, this estimator takes the lag of the dependent variable without any options left for analysts to use appropriate instruments. In this case the change in labour productivity is the dependent variable. **b** Because of the multicolinearity problem the estimates for all ICT variables are not estimated (STATA 7, the GMM estimator requires stricter orthogonality requirements than the traditional panel estimators). **c** The number of observations in the panel model refers to the number of firms for which complete data were available for the years 1995-96 to 1997-98. As the model is in first differences, data were not available for 1994-95 (the base year). The number of observations in the dynamic models refers to the number of firms operating in 1997-98 for which complete data on lagged and instrumental variables were available for the years 1995-96 and 1996-97. **d** The Wald χ^2 test is applied in dynamic models to test the overall significance of model parameters. The critical value at (6,1249) for these models is 1.64 (5% level). **e** The Sargan test is applied to models that test the appropriateness of instruments applied. A high 'p' value indicates that the instruments are not over identified. However, the Sargan test quite often accepts a misspecified model, particularly in small samples with mismeasured variables.

Source: STATA 7 estimates based on ABS (Business Longitudinal Survey Confidentialised Unit Record File, Cat. No. 8141.0.30.001