Session 14: Panel Data

Essential reading: Chapter 11 in Brooks.

Dr Artur Semeyutin

BHE0013 Applied Econometrics

Huddersfield Business School
Spring 2025

The Nature of Panel Data

- Panel data, also known as longitudinal data, have both time series and cross-sectional dimensions.
- They arise when we measure the same collection of people or objects over a period of time.
- Econometrically, the setup is

$$y_{it} = \alpha + \beta x_{it} + u_{it}$$

  where y_{it} is the dependent variable, α is the intercept term, β is a $k \times 1$vector of parameters to be estimated on the explanatory variables, x_{it}; $t = 1, \dots , T$; $i = 1, \dots , N$.
- The simplest way to deal with this data would be to estimate a single, pooled regression on all the observations together.
- But pooling the data assumes that there is no heterogeneity – i.e. the same relationship holds for all the data.

The Advantages of using Panel Data

- There are a number of advantages from using a full panel technique when a panel of data is available.
- We can address a broader range of issues and tackle more complex problems with panel data than would be possible with pure time series or pure cross-sectional data alone.
- It is often of interest to examine how variables, or the relationships between them, change dynamically (over time).
- By structuring the model in an appropriate way, we can remove the impact of certain forms of omitted variables bias in regression results.

Seemingly Unrelated Regression (SUR)

- One approach to making more full use of the structure of the data would be to use the SUR framework initially proposed by Zellner (1962). This has been used widely in finance where the requirement is to model several closely related variables over time.
- A SUR is so-called because the dependent variables may seem unrelated across the equations at first sight, but a more careful consideration would allow us to conclude that they are in fact related after all.
- Under the SUR approach, one would allow for the contemporaneous relationships between the error terms in the equations by using a generalised least squares (GLS) technique.

Seemingly Unrelated Regression (SUR) (Cont'd)

- The idea behind SUR is essentially to transform the model so that the error terms become uncorrelated. If the correlations between the error terms in the individual equations had been zero in the first place, then SUR on the system of equations would have been equivalent to running separate OLS regressions on each equation.

Fixed and Random Effects Panel Estimators

- The applicability of the SUR technique is limited because it can only be employed when the number of time series observations per cross-sectional unit is at least as large as the total number of such units, N.
- A second problem with SUR is that the number of parameters to be estimated in total is very large, and the variance-covariance matrix of the errors also has to be estimated. For these reasons, the more flexible full panel data approach is much more commonly used.
- There are two main classes of panel techniques: the *fixed effects estimator* and *the random effects estimator*.

Fixed Effects Models

- The fixed effects model for some variable *y it* may be written

$$y_{it} = \alpha + \beta x_{it} + \mu_i + v_{it}$$

- We can think of μ_i as encapsulating all of the variables that affect y_{it} cross-sectionally but do not vary over time – for example, the sector that a firm operates in, a person's gender, or the country where a bank has its headquarters, etc. Thus we would capture the heterogeneity that is encapsulated in μ_i by a method that allows for different intercepts for each cross sectional unit.
- This model could be estimated using dummy variables, which would be termed the least squares dummy variable (LSDV) approach.
- The LSDV model may be written

$$y_{it} = \beta x_{it} + \mu_1 D1_i + \mu_2 D2_i + \mu_3 D3_i + \cdots + \mu_N DN_i + v_{it}$$

Fixed Effects Models (Cont'd)

where $D1_i$ is a dummy variable that takes the value 1 for all observations on the first entity (e.g., the first firm) in the sample and zero otherwise, $D2_i$ is a dummy variable that takes the value 1 for all observations on the second entity (e.g., the second firm) and zero otherwise, and so on.

- The LSDV can be seen as just a standard regression model and therefore it can be estimated using OLS.
- Now the model given by the equation above has $N + k$ parameters to estimate. In order to avoid the necessity to estimate so many dummy variable parameters, a transformation, known as the within transformation, is used to simplify matters.

The Within Transformation

- The within transformation involves subtracting the time-mean of each entity away from the values of the variable.
- So define $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ as the time-mean of the observations for cross-sectional unit i, and similarly calculate the means of all of the explanatory variables.
- Then we can subtract the time-means from each variable to obtain a regression containing demeaned variables only.
- Note that such a regression does not require an intercept term since now the dependent variable will have zero mean by construction.
- The model containing the demeaned variables is

$$y_{it} - \overline{y}_i = \beta(x_{it} - \overline{x}_i) + u_{it} - \overline{u}_i$$

- We could write this as $\ddot{y}_{it} = \beta\ddot{x}_{it} + \ddot{u}_{it}$

The Within Transformation (Cont'd)

where the double dots above the variables denote the demeaned values.

- This model can be estimated using OLS, but we need to make a degrees of freedom correction.

The Between Estimator

- An alternative to this demeaning would be to simply run a cross-sectional regression on the time-averaged values of the variables, which is known as the *between estimator*.
- An advantage of running the regression on average values (the between estimator) over running it on the demeaned values (the within estimator) is that the process of averaging is likely to reduce the effect of measurement error in the variables on the estimation process.
- A further possibility is that instead, the first difference operator could be applied so that the model becomes one for explaining the change in y_{it} rather than its level. When differences are taken, any variables that do not change over time will again cancel out.
- Differencing and the within transformation will produce identical estimates in situations where there are only two time periods.

Time Fixed Effects Models

- It is also possible to have a time-fixed effects model rather than an entity-fixed effects model.
- We would use such a model where we think that the average value of y_{it} changes over time but not cross-sectionally.
- Hence with time-fixed effects, the intercepts would be allowed to vary over time but would be assumed to be the same across entities at each given point in time. We could write a time-fixed effects model as

$$y_{it} = \alpha + \beta x_{it} + \lambda_t + v_{it}$$

  where λ_t is a time-varying intercept that captures all of the variables that affect y and that vary over time but are constant cross-sectionally.

Time Fixed Effects Models (Cont'd)

- An example would be where the regulatory environment or tax rate changes part-way through a sample period. In such circumstances, this change of environment may well influence y, but in the same way for all firms.

- Time-variation in the intercept terms can be allowed for in exactly the same way as with entity fixed effects. That is, a least squares dummy variable model could be estimated

$$y_{it} = \beta x_{it} + \lambda_1 D1_t + \lambda_2 D2_t + \lambda_3 D3_t + \cdots + \lambda_T DT_t + v_{it}$$

  where D_{1t}, for example, denotes a dummy variable that takes the value 1 for the first time period and zero elsewhere, and so on.

Time Fixed Effects Models (Cont'd)

- The only difference is that now, the dummy variables capture time variation rather than cross-sectional variation. Similarly, in order to avoid estimating a model containing all T dummies, a within transformation can be conducted to subtract away the cross-sectional averages from each observation
- Finally, it is possible to allow for both entity fixed effects and time fixed effects within the same model. Such a model would be termed a two-way error component model, and the LSDV equivalent model would contain both cross-sectional and time dummies

Investigating Banking Competition with a Fixed Effects Model

- The UK banking sector is relatively concentrated and apparently extremely profitable.
- It has been argued that competitive forces are not sufficiently strong and that there are barriers to entry into the market.
- A study by Matthews, Murinde and Zhao (2007) investigates competitive conditions in UK banking between 1980 and 2004 using the Panzar-Rosse approach.
- The model posits that if the market is contestable, entry to and exit from the market will be easy (even if the concentration of market share among firms is high), so that prices will be set equal to marginal costs.
- The technique used to examine this conjecture is to derive testable restrictions upon the firm's reduced form revenue equation.

Methodology

- The empirical investigation consists of deriving an index (the Panzar-Rosse H-statistic) of the sum of the elasticities of revenues to factor costs (input prices).
- If this lies between 0 and 1, we have monopolistic competition or a partially contestable equilibrium, whereas $H < 0$ would imply a monopoly and $H = 1$ would imply perfect competition or perfect contestability.
- The key point is that if the market is characterised by perfect competition, an increase in input prices will not affect the output of firms, while it will under monopolistic competition. The model Matthews *et al*. investigate is given by

$$
\begin{aligned}
lnREV_{it} &= \alpha_0 + \alpha_1 lnPL_{it} + \alpha_2 lnPK_{it} + \alpha_3 lnPF_{it} \\
&\quad + \beta_1 lnRISKASS_{it} + \beta_2 lnASSET_{it} + \beta_3 lnBR_{it} \\
&\quad + \gamma_1 GROWTH_t + \mu_i + v_{it}
\end{aligned}
$$

Methodology (Cont'd)

where REV_{it} is the ratio of bank revenue to total assets for firm i at time t, PL is personnel expenses to employees (the unit price of labour); PK is the ratio of capital assets to fixed assets (the unit price of capital); and PF is the ratio of annual interest expenses to total loanable funds (the unit price of funds).

- The model also includes several variables that capture time-varying bank-specific effects on revenues and costs, and these are: *RISKASS*, the ratio of provisions to total assets; *ASSET* is bank size, as measured by total assets; *BR* is the ratio of the bank's number of branches to the total number of branches for all banks.
- Finally, $GROWTH_t$ is the rate of growth of *GDP*, which obviously varies over time but is constant across banks at a given point in time; ν *i* is a bank-specific fixed effects and *v it* is an idiosyncratic disturbance term. The contestability parameter, *H* is given as $\alpha_1 + \alpha_2 + \alpha_3$

Methodology (Cont'd)

- Unfortunately, the Panzar-Rosse approach is only valid when applied to a banking market in long-run equilibrium. Hence the authors also conduct a test for this, which centres on the regression

$$\begin{aligned} lnROA_{it} &= \alpha'_0 + \alpha'_1 lnPL_{it} + \alpha'_2 lnPK_{it} + \alpha'_3 lnPF_{it} \\ &\quad + \beta'_1 lnRISKASS_{it} + \beta'_2 lnASSET_{it} + \beta'_3 lnBR_{it} \\ &\quad + \gamma'_1 GROWTH_t + \eta_i + w_{it} \end{aligned}$$

- The explanatory variables for the equilibrium test regression are identical to those of the contestability regression but the dependent variable is now the log of the return on assets (ln *ROA*).
- Equilibrium is argued to exist in the market if $\alpha'_1 + \alpha'_2 + \alpha_3$'
- Matthews *et al.* employ a fixed effects panel data model which allows for differing intercepts across the banks, but assumes that these effects are fixed over time.

Methodology (Cont'd)

- The fixed effects approach is a sensible one given the data analysed here since there is an unusually large number of years (25) compared with the number of banks (12), resulting in a total of 219 bank-years (observations).
- The data employed in the study are obtained from banks' annual reports and the Annual Abstract of Banking Statistics from the British Bankers Association. The analysis is conducted for the whole sample period, 1980-2004, and for two sub-samples, 1980–1991 and 1992–2004.

Results from Test of Banking Market Equilibrium by Matthews et al .

Variable	1980–2004	1980–1991	1992–2004
Intercept	0.0230***	0.1034*	0.0252
	(3.24)	(1.87)	(2.60)
lnPL	−0.0002	0.0059	0.0002
	(0.27)	(1.24)	(0.37)
lnPK	−0.0014*	−0.0020	−0.0016*
	(1.89)	(1.21)	(1.81)
lnPF	−0.0009	−0.0034	0.0005
	(1.03)	(1.01)	(0.49)
lnRISKASS	−0.6471***	−0.5514***	−0.8343***
	(13.56)	(8.53)	(5.91)
lnASSET	−0.0016***	−0.0068**	−0.0016**
	(2.69)	(2.07)	(2.07)
lnBR	−0.0012*	0.0017	−0.0025
	(1.91)	(0.97)	(1.55)
GROWTH	0.0007***	0.0004	0.0006*
	(4.19)	(1.54)	(1.71)
R^2 within	0.5898	0.6159	0.4706
$\mathrm{H}_0 : \eta_i = 0$	$F(11, 200) = 7.78^{***}$	$F(9, 66) = 1.50$	$F(11, 117) = 11.28^{***}$
$\mathrm{H}_0 : E = 0$	$F(1, 200) = 3.20^{*}$	$F(1, 66) = 0.01$	$F(1, 117) = 0.28$

Notes: *t*-ratios in parentheses; *, ** and *** denote significance at the 10%, 5% and 1% levels respectively.
Source: Matthews *et al.* (2007). Reprinted with the permission of Elsevier.

Analysis of Equilibrium Test Results

- The null hypothesis that the bank fixed effects are jointly zero (H_0: $\eta_i = 0$) is rejected at the 1% significance level for the full sample and for the second sub-sample but not at all for the first sub-sample.
- Overall, however, this indicates the usefulness of the fixed effects panel model that allows for bank heterogeneity.
- The main focus of interest in the table on the previous slide is the equilibrium test, and this shows slight evidence of disequilibrium (E is significantly different from zero at the 10% level) for the whole sample, but not for either of the individual sub-samples.
- Thus the conclusion is that the market appears to be sufficiently in a state of equilibrium that it is valid to continue to investigate the extent of competition using the Panzar-Rosse methodology. The results of this are presented on the following slide.

Results from Test of Banking Market Competition by Matthews et al .

18pt Variable	1980–2004	1980–1991	1992–2004
Intercept	−3.083	1.1033**	−0.5455
	(1.60)	(2.06)	(1.57)
lnPL	−0.0098	0.164***	−0.0164
	(0.54)	(3.57)	(0.64)
lnPK	0.0025	0.0026	−0.0289
	(0.13)	(0.16)	(0.91)
lnPF	0.5788***	0.6119***	0.5096***
	(23.12)	(18.97)	(12.72)
lnRISKASS	2.9886**	1.4147**	5.8986
	(2.30)	(2.26)	(1.17)
lnASSET	−0.0551***	−0.0963***	−0.0676**
	(3.34)	(2.89)	(2.52)
lnBR	0.0461***	0.00094	0.0809
	(2.70)	(0.57)	(1.43)
GROWTH	−0.0082*	−0.0027	−0.0121
	(1.91)	(1.17)	(1.00)
R^2 within	0.9209	0.9181	0.8165
$H_0 : \eta_i = 0$	$F(11, 200) = 23.94$***	$F(9, 66) = 21.97$***	$F(11, 117) = 11.95$***
$H_0 : H = 0$	$F(1, 200) = 229.46$***	$F(1, 66) = 205.89$***	$F(1, 117) = 71.25$***
$H_1 : H = 1$	$F(1, 200) = 128.99$***	$F(1, 66) = 16.59$***	$F(1, 117) = 94.76$***
H	0.5715	0.7785	0.4643

Notes: *t*-ratios in parentheses; *, ** and ***, denote significance at the 10%, 5% and 1% levels respectively. The final set of asterisks in the table was added by the present author.
Source: Matthews *et al.* (2007). Reprinted with the permission of Elsevier.

Analysis of Competition Test Results

- The value of the contestability parameter, H, which is the sum of the input elasticities, falls in value from 0.78 in the first sub-sample to 0.46 in the second, suggesting that the degree of competition in UK retail banking weakened over the period.
- However, the results in the two rows above that show that the null hypotheses that $H = 0$ and $H = 1$ can both be rejected at the 1% significance level for both sub-samples, showing that the market is best characterised by monopolistic competition.
- As for the equilibrium regressions, the null hypothesis that the fixed effects dummies (μ_i) are jointly zero is strongly rejected, vindicating the use of the fixed effects panel approach and suggesting that the base levels of the dependent variables differ.

Analysis of Competition Test Results (Cont'd)

- Finally, the additional bank control variables all appear to have intuitively appealing signs. For example, the risk assets variable has a positive sign, so that higher risks lead to higher revenue per unit of total assets; the asset variable has a negative sign, and is statistically significant at the 5% level or below in all three periods, suggesting that smaller banks are more profitable.

The Random Effects Model

- An alternative to the fixed effects model described above is the random effects model, which is sometimes also known as the error components model.
- As with fixed effects, the random effects approach proposes different intercept terms for each entity and again these intercepts are constant over time, with the relationships between the explanatory and explained variables assumed to be the same both cross-sectionally and temporally.
- However, the difference is that under the random effects model, the intercepts for each cross-sectional unit are assumed to arise from a common intercept α (which is the same for all cross-sectional units and over time), plus a random variable ϵ_i that varies cross-sectionally but is constant over time.

The Random Effects Model (Cont'd)

- ϵ_i measures the random deviation of each entity's intercept term from the "global" intercept term α. We can write the random effects panel model as

$$y_{it} = \alpha + \beta x_{it} + \omega_{it}, \quad \omega_{it} = \epsilon_i + v_{it}$$

How the Random Effects Model Works

- Unlike the fixed effects model, there are no dummy variables to capture the heterogeneity (variation) in the cross-sectional dimension.

- Instead, this occurs via the ϵ_i terms.

- Note that this framework requires the assumptions that the new cross-sectional error term, ϵ_i, has zero mean, is independent of the individual observation error term v_{it}, has constant variance, and is independent of the explanatory variables.

- The parameters (α and the β vector) are estimated consistently but inefficiently by OLS, and the conventional formulae would have to be modified as a result of the cross-correlations between error terms for a given cross-sectional unit at different points in time.

How the Random Effects Model Works (Cont'd)

- Instead, a generalised least squares (GLS) procedure is usually used. The transformation involved in this GLS procedure is to subtract a weighted mean of the y_{it} over time (i.e. part of the mean rather than the whole mean, as was the case for fixed effects estimation).

Quasi-Demeaning the Data

- Define the 'quasi-demeaned' data as $y_{it}^* = y_{it} - \theta \overline{y}_i$ and similarly for x_{it},
- ρ will be a function of the variance of the observation error term, σ_v^2, and of the variance of the entity-specific error term, σ_ϵ^2:

$$\theta = 1 - \frac{\sigma_v}{\sqrt{T\sigma_\epsilon^2 + \sigma_v^2}}$$

- This transformation will be precisely that required to ensure that there are no cross-correlations in the error terms, but fortunately it should automatically be implemented by standard software packages.
- Just as for the fixed effects model, with random effects, it is also conceptually no more difficult to allow for time variation than it is to allow for cross-sectional variation.

Quasi-Demeaning the Data (Cont'd)

- In the case of time-variation, a time period-specific error term is included and again, a two-way model could be envisaged to allow the intercepts to vary both cross-sectionally and over time.

Fixed or Random Effects?

- It is often said that the random effects model is more appropriate when the entities in the sample can be thought of as having been randomly selected from the population, but a fixed effect model is more plausible when the entities in the sample effectively constitute the entire population.
- More technically, the transformation involved in the GLS procedure under the random effects approach will not remove the explanatory variables that do not vary over time, and hence their impact can be enumerated.
- Also, since there are fewer parameters to be estimated with the random effects model (no dummy variables or within transform to perform), and therefore degrees of freedom are saved, the random effects model should produce more efficient estimation than the fixed effects approach.

Fixed or Random Effects? (Cont'd)

- However, the random effects approach has a major drawback which arises from the fact that it is valid only when the composite error term ω_{it} is uncorrelated with all of the explanatory variables.
- This assumption is more stringent than the corresponding one in the fixed effects case, because with random effects we thus require both ϵ_{it} and v_{it} to be independent of all of the x_{it}.
- This can also be viewed as a consideration of whether any unobserved omitted variables (that were allowed for by having different intercepts for each entity) are uncorrelated with the included explanatory variables. If they are uncorrelated, a random effects approach can be used; otherwise the fixed effects model is preferable.
- A test for whether this assumption is valid for the random effects estimator is based on a slightly more complex version of the Hausman test.

Fixed or Random Effects? (Cont'd)

- If the assumption does not hold, the parameter estimates will be biased and inconsistent.
- To see how this arises, suppose that we have only one explanatory variable, x_{2it} that varies positively with y_{it}, and also with the error term, ω_{it}. The estimator will ascribe all of any increase in y to x when in reality some of it arises from the error term, resulting in biased coefficients.

Credit Stability of Banks in Central and Eastern Europe: A Random Effects Analysis

- Foreign participants in the banking sector may improve competition and efficiency to the benefit of the economy that they enter.
- They may have a stabilising effect on credit provision since they will probably be better diversified than domestic banks and will therefore be more able to continue to lend when the host economy is performing poorly.
- But on the other hand, it is also argued that foreign banks may alter the credit supply to suit their own aims rather than that of the host economy.
- They may act more pro-cyclically than local banks, since they have alternative markets to withdraw their credit supply to when host market activity falls.
- Moreover, worsening conditions in the home country may force the repatriation of funds to support a weakened parent bank.

The Data

- There may also be differences in policies for credit provision dependent upon the nature of the formation of the subsidiary abroad – i.e. whether the subsidiary's existence results from a take-over of a domestic bank or from the formation of an entirely new startup operation (a 'greenfield investment').
- A study by de Haas and van Lelyveld (2006) employs a panel regression using a sample of around 250 banks from ten Central and East European countries.
- They examine whether domestic and foreign banks react differently to changes in home or host economic activity and banking crises.
- The data cover the period 1993–2000 and are obtained from BankScope.

The Model

- The core model is a random effects panel regression of the form:

$$
\begin{aligned}
gr_{it} \quad = \quad & \alpha + \beta_1 Takeover_{it} + \beta_2 Greenfield_i + \beta_3 Crisis_{it} \\
& + \beta_4 Macro_{it} + \ \beta_5 Contr_{it} + (\mu_i + \epsilon_{it})
\end{aligned}
$$

where the dependent variable, gr_{it}, is the percentage growth in the credit of bank i in year t; *Takeover* is a dummy variable taking the value one for foreign banks resulting from a takeover and zero otherwise; *Greenfield* is a dummy taking the value one if bank is the result of a foreign firm making a new banking investment rather than taking over an existing one; *crisis* is a dummy variable taking the value one if the host country for bank i was subject to a banking disaster in year t.

The Model (Cont'd)

- *Macro* is a vector of variables capturing the macroeconomic conditions in the home country (the lending rate and the change in GDP for the home and host countries, the host country inflation rate, and the differences in the home and host country GDP growth rates and the differences in the home and host country lending rates).

- *Contr* is a vector of bank-specific control variables that may affect the dependent variable irrespective of whether it is a foreign or domestic bank.

- These are: *weakness parent bank*, defined as loan loss provisions made by the parent bank; *solvency* is the ratio of equity to total assets; *liquidity* is the ratio of liquid assets / total assets; *size* is the ratio of total bank assets to total banking assets in the given country; *profitability* is return on assets and *efficiency* is net interest margin.

The Model (Cont'd)

- α and the β's are parameters (or vectors of parameters in the cases of β_4 and β_5), μ_i is the unobserved random effect that varies across banks but not over time, and ϵ_{it} is an idiosyncratic error term.

Estimation Options

- de Haas and van Lelyveld discuss the various techniques that could be employed to estimate such a model.
- OLS is considered to be inappropriate since it does not allow for differences in average credit market growth rates at the bank level.
- A model allowing for entity-specific effects (i.e. a fixed effects model that effectively allowed for a different intercept for each bank) is ruled out on the grounds that there are many more banks than time periods and thus too many parameters would be required to be estimated.
- They also argue that these bank-specific effects are not of interest to the problem at hand, which leads them to select the random effects panel model.
- This essentially allows for a different error structure for each bank. A Hausman test is conducted, and shows that the random effects model is valid since the bank-specific effects μ_i are found "in most cases not to be significantly correlated with the explanatory variables."

Results

30pt Explanatory variables	Full sample I	Full sample II	Domestic banks	Foreign banks I	Foreign banks II
Takeover	-11.58	-5.65			
	(1.26)	(0.29)			
Greenfield	14.99	29.59		12.39	8.11
	(1.29)	(1.55)		(0.88)	(0.65)
Crisis	-19.79^{***}	-14.42^{***}	-19.36^{***}	0.31	-4.13
	(4.30)	(2.93)	(3.43)	(0.03)	(0.33)
Host – home ΔGDP	8.08^{***}			8.86^{***}	
	(4.18)			(4.11)	
Host ΔGDP		6.68^{***}	6.74^{***}		8.64^{***}
		(7.39)	(6.98)		(2.93)
Home ΔGDP		-6.04^{*}			-8.62^{***}
		(1.89)			(2.78)
Host – home lending rate	1.12^{**}			0.85	
	(1.97)			(0.88)	
Host lending rate		0.28	0.34		1.50
		(1.08)	(1.36)		(1.11)
Home lending rate		2.97^{***}			1.11
		(4.03)			(1.15)
Host inflation	-0.01	0.03	0.03	0.08	0.07
	(0.37)	(1.01)	(0.12)	(0.61)	(0.44)
Weakness parent bank	-0.19^{***}	-0.16^{***}		-0.23^{***}	-0.19^{***}
	(4.37)	(3.04)		(7.00)	(4.27)
Solvency	1.29^{***}	1.25^{***}	0.85^{***}	3.33^{***}	3.18^{***}
	(5.34)	(4.77)	(3.24)	(5.53)	(5.30)
Liquidity	-0.05^{**}	0.02	0.02	-0.53	-0.43
	(2.09)	(0.78)	(0.70)	(1.40)	(1.14)
Size	-34.65^{**}	-29.14	-21.93	-108.00	-136.19
	(1.96)	(1.56)	(1.16)	(0.54)	(0.72)
Profitability	1.09^{**}	1.09^{**}	1.21^{***}	2.16	0.91
	(2.18)	(2.14)	(2.81)	(0.75)	(0.29)
Interest margin	1.66^{***}	1.90^{***}	2.71^{***}	-3.42	-2.84
	(2.90)	(3.41)	(4.96)	(1.18)	(0.94)
Observations	1003	1003	770	233	233
No. of banks	247	247	184	82	82
Hausman test statistic	0.66	0.94	0.76	0.58	0.92
R^2	0.28	0.33	0.30	0.46	0.47

Notes: *t*-ratios in parentheses. Intercept and country dummy parameter estimates are not shown. Empty cells occur when a particular variable is not included in a regression.
Source: de Haas and van Lelyveld (2006). Reprinted with the permission of Elsevier.

Analysis of Results

- The main result is that during times of banking disasters, domestic banks significantly reduce their credit growth rates (i.e. the parameter estimate on the *crisis* variable is negative for domestic banks), while the parameter is close to zero and not significant for foreign banks.
- There is a significant negative relationship between home country GDP growth, but a positive relationship with host country GDP growth and credit change in the host country.
- This indicates that, as the authors expected, when foreign banks have fewer viable lending opportunities in their own countries and hence a lower opportunity cost for the loanable funds, they may switch their resources to the host country.
- Lending rates, both at home and in the host country, have little impact on credit market share growth.

Analysis of Results (Cont'd)

- Interestingly, the greenfield and takeover variables are not statistically significant (although the parameters are quite large in absolute value), indicating that the method of investment of a foreign bank in the host country is unimportant in determining its credit growth rate or that the importance of the method of investment varies widely across the sample leading to large standard errors.
- A weaker parent bank (with higher loss provisions) leads to a statistically significant contraction of credit in the host country as a result of the reduction in the supply of available funds.
- Overall, both home-related ('push') and host-related ('pull') factors are found to be important in explaining foreign bank credit growth.

Essential Reading

- Please read the text book chapter: Chris Brooks - Introductory Econometrics for Finance, 4th Edition (2019) Cambridge University Press, Chapter 11.
- Or read: Jeffrey Wooldridge - Introductory Econometrics, 7th Edition (2019) Cengage, Chapters 13 and 14.