PHD ASSISTANCE RESEARCH LAB · DOWNLOADABLE WALKTHROUGH

Panel Regression and Time-Series Analysis in Stata and R for Finance Doctorates: A Downloadable Walkthrough for Qatar Scholars

Paired Stata and R code for setup, stationarity, estimation, diagnostics and reporting. Version 1, November 2026.

This econometrics walkthrough gives paired Stata and R code for the five stages of a finance panel or time-series analysis: data setup, stationarity, estimation, diagnostics and reporting. Use panel data analysis in Stata and R when you follow many firms over a few years. Use time-series models when you follow one series, such as an index or exchange rate, over a long period.

At a glance

- **Many firms, few years:** panel models (fixed or random effects) with firm-clustered errors.
- **One series, many periods:** unit-root tests first, then ARIMA or VAR.
- **Short panels:** unit-root tests have little power, so focus on serial correlation and dynamics instead.
- **Small markets like Qatar:** firms move together, so test for cross-sectional dependence.
- **Everything in a script:** do-file or R script from raw data to final table.

Panel or time series: which applies to you?

Pick the model family before you open the software. Panel data analysis Stata workflows start with xtset and R workflows with plm's pdata.frame (Stage 1), while time-series analysis R workflows use the forecast, urca and vars packages (Stages 2 to 4).

	Panel regression	Time-series analysis
Data	Many firms × several years (e.g. QSE-listed firms, 2010–2023)	One series × many periods (e.g. an index, exchange rate, oil price)
Main concern	Unobserved firm differences	Dependence on its own past; stationarity
	Pooled OLS, fixed effects, random effects	AR/ARIMA, VAR

	Panel regression	Time-series analysis
Core models		
Key tests	FE vs RE (Mundlak/Hausman), serial correlation, cross-sectional dependence	Unit root (ADF), lag selection, residual autocorrelation

Dynamic panels, where last year's value of the outcome enters the model, combine both ideas and need their own estimators [1].

Stage 1: Set up the data

Task	Stata	R
Declare panel	`xtset firm_id year`	`pdata <- pdata.frame(d, index = c("firm_id","year")) (plm)`
Check balance	`xtdescribe`	`pdim(pdata)`
Within/between variation	`xtsum roa lev size`	`summary(pdata$lev) + firm means`
Declare time series	`tsset month`	`ts(x, start = c(2010,1), frequency = 12)`

Stage 2: Check stationarity (time series)

Non-stationary series can produce spurious regressions. Test each long series with an augmented Dickey–Fuller test before modelling [2]:

Stata	R
`dfuller oil, lags(4) trend`	`urca::ur.df(oil, type = "trend", lags = 4)`

If the series has a unit root, difference it or test for cointegration. For panels with only a handful of years, panel unit-root tests have little power and are rarely the main concern. Spend the effort on serial correlation and model choice instead.

Stage 3: Estimate

Panel models

Model	Stata	R
Pooled OLS	`regress roa lev size growth, vce(cluster firm_id)`	`feols(roa ~ lev + size + growth, cluster = ~firm_id, data = d) (fixest)`
Fixed effects (firm + year)	`xtreg roa lev size growth i.year, fe vce(cluster firm_id)`	`fixest feols() with firm and year fixed effects (code below)`
Random effects		

Model	Stata	R
	`xtreg roa lev size growth i.year, re vce(cluster firm_id)`	`plm(roa ~ lev + size + growth + factor(year), data = pdata, model = "random")`

Fixed effects in R with fixest:

```
library(fixest)
fe <- feols(roa ~ lev + size + growth | firm_id + year,
            cluster = ~firm_id, data = d)
```

The plm diagnostics in Stage 4 need a plm model, so fit the same fixed effects model with plm as well:

```
library(plm); library(lmtest)
fe_p <- plm(roa ~ lev + size + growth, data = pdata,
            model = "within", effect = "twoways")
```

Time-series models

Model	Stata	R
ARIMA	`arima ret, arima(1,0,1)`	`forecast::auto.arima(ret)`
VAR lag choice	`varsoc ret oil, maxlag(8)`	`vars::VARselect(cbind(ret, oil), lag.max = 8)`
VAR	`var ret oil, lags(1/2)`	`vars::VAR(cbind(ret, oil), p = 2)`

Stage 4: Run diagnostics

Check	Why	Stata	R
FE vs RE	Is the RE assumption plausible?	`Hausman (hausman fe re) without clustering, or the Mundlak test with clustering [3][4]`	`phtest(roa ~ lev + size + growth + factor(year), data = pdata, method = "aux", vcov = function(x) vcovHC(x, method = "arellano"))` (cluster-robust)
Serial correlation in panels	Inflated t-statistics	`xtserial roa lev size growth (ssc install; [5])`	`pwartest(fe_p)`
Cross-sectional dependence	Firms in a small market move together	`xtcsd, pesaran abs after xtreg, fe (ssc install) [6]`	`pcdtest(fe_p, test = "cd")`
Robust errors	Correct inference	`vce(cluster firm_id); Driscoll–Kraay via xtscc (ssc`	`lmtest::coeftest(fe_p, vcov = vcovHC(fe_p, cluster = "group"))`

Check	Why	Stata	R
		`install) when dependence is strong`	
VAR residuals	Leftover autocorrelation	`varlmar`	`vars::serial.test(var_model)`

Clustered standard errors are the default fix for within-firm correlation in finance panels [7][8].

Stage 5: Report

- One table with pooled OLS, FE and RE side by side, showing the number of firms, observations, fixed effects and clustering level.
- State which tests drove your model choice and show their results.
- Interpret coefficients in economic terms: size and direction, not just stars.
- Keep the do-file or R script that produces every table.

Questions to answer before choosing a model

- **Data structure:** many firms and few years, or one long series?
- **Variable behaviour:** are your long series stationary?
- **Market context:** do firms in your sample move together?
- **Supervisor expectation:** which tests does your department expect to see reported?

If a model does not match your data structure, change the model before you run it.

Future research directions for Qatar data

- **Short panels and small samples:** estimators and inference that stay reliable with few firms and years.
- **Structural breaks:** testing whether relationships changed around oil-price falls or COVID-19.
- **Machine learning and causal inference:** combining panel methods with newer designs in reproducible workflows. See our recent trends in finance data analysis.

Slide outline for the PPT (12 slides)

The regression analysis PDF follows the same five stages as the PPT:

1. Title. 2. Panel vs time series (table). 3. Data setup (code). 4. Stationarity (code). 5. Pooled OLS. 6. Fixed effects. 7. Random effects. 8. Choosing FE vs RE (Mundlak). 9. Panel diagnostics. 10. Time-series models (ARIMA, VAR). 11. Reporting checklist. 12. Next steps + contact.

Next step

See these methods applied in A Worked Panel-Data Analysis Report for a Finance Doctoral Thesis (Qatar, Annotated).

For finance doctoral scholars in Qatar

PhD Assistance's econometrics team helps finance doctoral scholars choose, run, test and report panel and time-series models in Stata and R. Our finance data analysis and econometrics support guides Qatar doctoral scholars through model selection, Stata and R coding and results interpretation.

- Download the PDF · Download the slides (PPTX)
- Get econometrics support

References

1. Baltagi, B. H. (2021). *Econometric analysis of panel data* (6th ed.). Springer. https://doi.org/10.1007/978-3-030-53953-5
2. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*(366), 427–431. https://doi.org/10.1080/01621459.1979.10482531
3. Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*(6), 1251–1271. https://doi.org/10.2307/1913827
4. Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, *46*(1), 69–85. https://doi.org/10.2307/1913646
5. Drukker, D. M. (2003). Testing for serial correlation in linear panel-data models. *The Stata Journal*, *3*(2), 168–177. https://doi.org/10.1177/1536867X0300300206
6. Pesaran, M. H. (2021). General diagnostic tests for cross-sectional dependence in panels. *Empirical Economics*, *60*, 13–50. https://doi.org/10.1007/s00181-020-01875-7
7. Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies*, *22*(1), 435–480. https://doi.org/10.1093/rfs/hhn053
8. Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317–372. https://doi.org/10.3368/jhr.50.2.317
9. Croissant, Y., & Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, *27*(2). https://doi.org/10.18637/jss.v027.i02

© 2026 PhD Assistance (Guires Solutions Private Limited). This walkthrough is for learning; adapt the code to your own data and check results with your supervisor. Online version: https://www.phdassistance.com/insights/pdf-ppt/panel-regression-time-series-stata-r-walkthrough/