xtvfreg

Varying Fixed Effects Panel Regression

A Step-by-Step Guide to Modeling Variance in Panel Data

Tim F. Liao
University of Illinois Urbana-Champaign
&
Irma Mooi-reci
University of Melbourne

What Problem Does xtvfreg Solve?

The Problem

Traditional panel models assume constant variance (homoscedasticity)

But inequality is often heterogeneous:
- Wage dispersion varies by education
- Income volatility differs by sector
- Risk heterogeneity across groups

Standard models miss these variance patterns

The xtvfreg Solution

Models BOTH mean AND variance:

Mean equation:
- Who has higher/lower outcomes?

Variance equation:
- Who has more/less dispersion?

Captures complete inequality picture

1 Step 1: Prepare Your Data

Load Your Dataset

```
webuse nlswork, clear
```

Decompose Variables into Between/Within Components

```
* Between (individual means)
egen mage = mean(age), by(idcode)
egen mtenure = mean(tenure), by(idcode)

* Within (deviations from means)
gen dage = age - mage
gen dtenure = tenure - mtenure
```

Declare Panel Structure (Required)

```
xtset idcode year
```

2 Step 2: Run the Command

Basic Command Structure

```
xtvfreg ln_wage, \
  groupvar(south) \
  panelid(idcode) \
  meanvars(collgrad mage mtenure dage dtenure) \
  varvars(collgrad mage mtenure dage dtenure)
```

What Each Part Does

$$\hat{y}_{it} = \mathrm{x}'_{1i}\boldsymbol{\beta}_1 + +\bar{\mathbf{x}}'_{2i}\boldsymbol{\beta}_2 + (\mathbf{x}_{2it} - \bar{\mathbf{x}}_{2i})'\boldsymbol{\beta}_3 + \bar{u}_i\alpha_1 + (u_{it} - \bar{u}_i)\,\alpha_2$$

ln_wage → Dependent variable (log wage)

groupvar(south) → Estimate separate models for south=0 and south=1

panelid(idcode) → Panel identifier (individual ID)

$$e_{it} = v_i + d_{it}$$

meanvars(...) → Predictors for mean (expected wage)

varvars(...) → Predictors for variance (wage dispersion)

$$\widehat{\log\sigma^2_{it}} = \mathrm{z}'_{1i}\boldsymbol{\theta}_1 + \bar{\mathbf{z}}'_{2i}\boldsymbol{\theta}_2 + (\mathbf{z}_{2it} - \bar{\mathbf{z}}_{2i})'\boldsymbol{\theta}_3 + \bar{u}_i\alpha_1 + (u_{it} - \bar{u}_i)\,\alpha_2$$

3 Step 3: Understanding the Output - Iteration Log

What You See First: Convergence Progress

```
Iteration History:
──────────────────────────────────────────────────────
Iter     Log-likelihood      Change          Criterion
──────────────────────────────────────────────────────
1          -1234.5678         0.123456         0.000001
2          -1234.4444         0.012340         0.000001
3          -1234.4321         0.001230         0.000001
──────────────────────────────────────────────────────
Converged in 3 iterations
```

What This Tells You

✓ **Converged in 3 iterations**
→ Algorithm successfully found optimal solution

✓ **Change < Criterion**
→ Results are stable and reliable

3 Step 3: Understanding the Output - Mean Equation

Mean Equation Results

	Coef.	Std.Err.	z	P>|z|
collgrad	0.1523	0.0234	6.51	0.000
mage	0.0089	0.0012	7.42	0.000
mtenure	0.0234	0.0045	5.20	0.000
dage	0.0156	0.0023	6.78	0.000
dtenure	0.0098	0.0019	5.16	0.000
_cons	1.2345	0.0567	21.78	0.000

How to Read Mean Equation

collgrad = 0.1523 → College grads earn 15.2% higher wages
mage = 0.0089 → Each year older (between-person) adds 0.89% to wages
dage = 0.0156 → Each year older (within-person) adds 1.56% to wages

3 Step 3: Understanding the Output - Variance Equation

Variance Equation Results (Log Scale)

	Coef.	Std.Err.	z	P>|z|
collgrad	0.2341	0.0456	5.13	0.000
mage	-0.0234	0.0089	-2.63	0.009
mtenure	-0.0123	0.0067	-1.84	0.066
dage	0.0456	0.0123	3.71	0.000
dtenure	0.0234	0.0098	2.39	0.017
_cons	-1.5678	0.1234	-12.71	0.000

How to Read Variance Equation (CRITICAL)

Coefficients are on LOG scale → Exponentiate for interpretation

collgrad = 0.2341 → exp(0.2341) = 1.26 → 26% MORE variance among college grads
mage = -0.0234 → exp(-0.0234) = 0.98 → 2% LESS variance per year of age

3 Step 3: Variance Decomposition

Output Shows How Variance is Partitioned

```
Variance Decomposition:
──────────────────────────────────────────────────

Total variance of ln_wage:              0.234567
  Variance explained by mean model:     0.098765 (42.1%)
  Variance explained by variance model: 0.067890 (28.9%)
  Unexplained variance:                 0.067912 (29.0%)
```

What This Means

42.1%

Mean Model

Systematic differences between groups

28.9%

Variance Model

Modeled inequality patterns

29.0%

Unexplained

Individual effects + residual

4 Step 4: Interpretation Framework

Mean Equation → Between-Group Inequality

Effect: ADDITIVE

Example: β_education = 0.15 means college grads earn 15% more than non-grads

Interpretation: Shifts the AVERAGE level of the outcome
Increases gaps BETWEEN groups.

Variance Equation → Within-Group Inequality

Effect: MULTIPLICATIVE (on log scale)

Example: γ_education = 0.20 means exp(0.20) = 1.22, so college grads have 22% MORE wage dispersion

Interpretation: Changes the SPREAD of the outcome
Increases inequality WITHIN groups

5 Step 5: Working with Stored Results

xtvfreg stores results for each group

mean_0, mean_1, ... → Mean equation coefficients
var_0, var_1, ... → Variance equation coefficients

View Specific Results

```
estimates replay mean_1    * Mean equation for group 1
estimates replay var_0     * Variance equation for group 0
```

Create Publication Tables

```
* Side-by-side comparison
esttab mean_*, se star(* 0.10 ** 0.05 *** 0.01)

* Export to Word
etable, replay export(results.docx, replace)
```

Real Example: Regional Wage Inequality

Research Question: Does education affect wage levels AND wage inequality differently in the South vs. Non-South?

Hypothetical Findings

Non-South Region

Mean Equation:
College: $\beta = 0.18$ ($p < 0.001$)
→ 18% wage premium

Variance Equation:
College: $\gamma = 0.15$ ($p = 0.023$)
→ $\exp(0.15) = 1.16$
→ 16% more dispersion

South Region

Mean Equation:
College: $\beta = 0.21$ ($p < 0.001$)
→ 21% wage premium

Variance Equation:
College: $\gamma = 0.28$ ($p < 0.001$)
→ $\exp(0.28) = 1.32$
→ 32% more dispersion

Tips and Best Practices

Use between/within decomposition

Separate time-invariant (between) from time-varying (within) effects for clearer interpretation

Check convergence

Most models converge in <10 iterations. If not, check for collinearity or sparse groups

Exponentiate variance coefficients

Remember: variance equation uses log link. Always exp(β) for interpretation!

Compare variance decomposition

Look at % explained by mean vs. variance model across groups to assess inequality sources

Use 'combined' option

See side-by-side results to easily spot differences across groups

Theory guides variable choice

Mean and variance variables can differ. Let theory guide which affects location vs. dispersion

Reference & Resources

Based on Published Research

Mooi-Reci, I., and T. F. Liao. 2025. "Unemployment: A Hidden Source of Wage Inequality?" European Sociological Review 41(3): 382-394.

https://doi.org/10.1093/esr/jcae052

Stata Help

```
Type: help xtvfreg
```

Author

Tim F. Liao
University of Illinois Urbana-Champaign
tfliao@illinois.edu

xtvfreg reveals the complete inequality picture by modeling both mean differences (who has more) and variance differences (who has more dispersion)

Thank You

Questions?

xtvfreg: Uncovering hidden sources of inequality