

Panel Data Threshold Regression with Interactive Fixed Effects

2026 UK Stata Conference

Jan Ditzen¹, Yiannis Karavias², Joakim Westerlund³

¹Free University of Bozen-Bolzano, Bozen, Italy

²Brunel University, UK

³Lund University, Sweden

September 4, 2026

Motivation

- In applied work, effects (coefficients) may differ above or below an unknown threshold.
- Example: Effects...
 - ▶ ... of a policy depends on the state of the economy.
 - ▶ ... of a medication depend on general health of person.
- Ignoring the "state" leads to biased coefficient estimates.
- In panel data: Units are exposed to common shocks, creating cross-sectional dependence; also often panels have large N but short T .

Motivation

Literature

- Panel (Time) Series
 - ▶ Hansen (1999) estimation of thresholds in fixed T, large N panel data but without cross-sectional dependence, Chudik et al. (2017) for dynamic models.
 - ▶ Ditzen et al. (2026) extend Hansen (1999) to panel models with interactive fixed effects, Barassi et al. (2026) with heterogenous thresholds.
- `xtthreshold` introduces estimation for thresholds in panel data with and without interactive fixed effects.

The threshold panel model

$$y_{it} = \beta' \mathbf{x}_{it} + \delta' \mathbf{h}_{it}(\gamma) + e_{it}$$
$$\mathbf{h}_{it}(\gamma) = \mathbf{S}' \mathbf{x}_{it} \mathbb{I}(z_{it} > \gamma).$$

- z_{it} is the scalar threshold variable and included in $\mathbf{x}_{it}$
 - $\mathbf{h}_{it}(\gamma)$ is the threshold function with threshold γ is unknown.
 - $\mathbf{S}$ selects the regressors whose coefficients change across regimes.
 - Below the threshold, selected slopes are β .
 - Above the threshold, selected slopes are $\beta + \delta$.
- ⇒ The threshold, γ , tells us where the regime changes
- ⇒ δ tells us how much the selected slopes change.

Interactive Fixed Effects

- Assume an error factor structure with unobserved factors:

$$\begin{aligned}e_{it} &= \boldsymbol{\mu}_i' \mathbf{f}_t + \varepsilon_{it}, \\ \mathbf{x}_{it} &= \boldsymbol{\Lambda}_i' \mathbf{f}_t + \mathbf{v}_{it}.\end{aligned}$$

- Common factors $\mathbf{f}_t$ affect all units, but with unit-specific loadings $\boldsymbol{\mu}_i$ and $\boldsymbol{\Lambda}_i$.
- Regressors can be correlated with the unobserved common factors, so simple within-demeaning does not solve the problem.
- Ditzén et al. (2026) combine threshold estimation with the Common Correlated Effects (CCE) (Pesaran, 2006).

CCE Estimator

- Idea: span space of latent factors using cross-sectional averages of X and y (Pesaran, 2006).
- CCE estimator popular, straightforward to implement and widely used. Robust to temporal correlation in observables, factors and errors (Kapetanios et al., 2011; Chudik et al., 2011; Chudik and Pesaran, 2015).
- Idea: partial out cross-section averages using annihilator matrix $Q_{\bar{H}}$:

$$\begin{aligned}Q_{\bar{H}} &= I_T - \bar{H}(\bar{H}'\bar{H})^{-1}\bar{H}' \\ \bar{H} &= [\bar{Y}, \bar{X}], \quad \tilde{X}_i = Q_{\bar{H}}X_i \\ \hat{\beta} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y}\end{aligned}$$

- For threshold estimation, use cross-section averages of X rather than y .

Threshold Estimation

- Create grid over (subset) of unique values of thresholding variable z_{it} .
- For each candidate threshold γ :
 - ① build $H_i(\gamma)$;
 - ② defactor using CCE;
 - ③ estimate the slope coefficients by least squares;
 - ④ calculate $SSR(\gamma)$.
- Estimate threshold $\hat{\gamma}$ using minima of SSRs:

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma_{NT}} SSR(\gamma).$$

$\Rightarrow$ Candidate set can contain up to NT values of z_{it} . A quantile grid reduces computation; a finer grid gives a closer approximation to the full search.

Inference

Three Questions

- 1 **Slope coefficients:** standard t and Wald inference is available for $\theta = (\beta', \delta')'$.
- 2 **Where is the threshold?** Use the LR statistic

$$LR(\gamma_0) = N(T - k - l) \frac{SSR(\gamma_0) - SSR(\hat{\gamma})}{SSR(\hat{\gamma})}$$

and invert it to obtain a confidence set for γ . Critical values at level ϕ from $c(\phi) = -2\log(1 - \sqrt{1 - \phi})$.

- 3 **Is there a threshold at all?** Under $H_0 : \delta = 0$, γ is unidentified, so standard Wald inference fails; use a sup-Wald test with a pairs bootstrap over cross-sectional units.

► Details on Asymptotics

Syntax

```
xtthreshold depvar indepvars1 | indepvars2 [ if ] , threshold(varlist)  
[ csa[(varlist)] gridlength(real) pooled options1 ]
```

- *indepvars1* are variables exposed to threshold, *indepvars2* are not affected by threshold.
- *csa[()]* specifies cross-section averages.
- *grid()* sets grid length.
- *pooled* pooled coefficients.
- *options1* include options to control type of inverter, type of constant and fixed effects.

Empirical Example

Effect of Inflation on Economic Growth (Ditzen et al., 2026)

- GDP a function of inflation (khanvar), government spending (govspend), trade openness (open), population growth (popgr) and gross capital formation (grosscap).
- Empirically likely that effect of inflation and government spending on GDP depends on level of inflation.
- Question: What is the level?
- Inflation acts as a threshold variable and inflation and government spending are regime switching.

Empirical Example

Effect of Inflation on Economic Growth (Ditzen et al., 2026)

```
. xtthreshold gdppcgr govspend khanvar | open popgr grosscap , ///  
>      threshold(khanvar) grid(400) pooled noreg ///  
>      csa(govspend khanvar open popgr grosscap)
```

Threshold Estimation (Ditzen, Karavias, Westerlund, 2026)
Threshold variable(s): khanvar

Threshold	Value	95% CI	
1	1.24	.0619	1.57

- Estimated threshold for inflation of 1.24%.
- `xtthreshold` automatically displays results from a CCE regression. Option `noreg` prevents this behaviour.

Empirical Example

Effect of Inflation on Economic Growth (Ditzen et al., 2026)

```
. xtthreshold gdpccgr govspend khanvar | open popgr grosscap , ///  
> threshold(khanvar) grid(400) pooled ///  
> csa(govspend khanvar open popgr grosscap)
```

Threshold Estimation (Ditzen, Karavias, Westerlund, 2026)

Threshold variable(s): khanvar

Threshold	Value	95% CI
1	1.24	.0619 1.57

Pooled Regression

gdpccgr	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
thresholded						
govspend_0	-.3833193	.0392082	-9.78	0.000	-.460166	-.3064726
govspend_1	.1258445	.0196636	6.40	0.000	.0873045	.1643846
khanvar_0	-1.651545	.1749324	-9.44	0.000	-1.994406	-1.308684
khanvar_1	.3586473	.0942674	3.80	0.000	.1738866	.543408
non_thresholded						
open	.0279824	.0059451	4.71	0.000	.0163302	.0396346
popgr	-.8679307	.0979326	-8.86	0.000	-1.059875	-.6759864
grosscap	.1033035	.0185232	5.58	0.000	.0669986	.1396083

Cross-Section Averages

with threshold: govspend khanvar

without threshold: open popgr grosscap

Individual fixed effects partialled out

- Variables with "_0" indicate variable larger than the threshold ($z_{it} > \gamma$), "_1" smaller ($z_{it} \leq \gamma$).
- Note: standard errors are standard OLS errors, not from the CCEP estimator!

Empirical Example

Post Estimation

- Test if threshold exists, i.e. $H_0 : \delta = 0$.
- `estat gtest` $\neq$ implements this test, with $\neq$ the value under the null.

```
. estat gtest 0
```

```
Test for Thresholds (Ditzen, Karavias, Westerlund 2026)
```

```
H_0: Threshold(s) at [0.00]
```

```
GTest    = 17.42
```

```
p-value = 0.00
```

```
. estat gtest 1
```

```
Test for Thresholds (Ditzen, Karavias, Westerlund 2026)
```

```
H_0: Threshold(s) at [1.00]
```

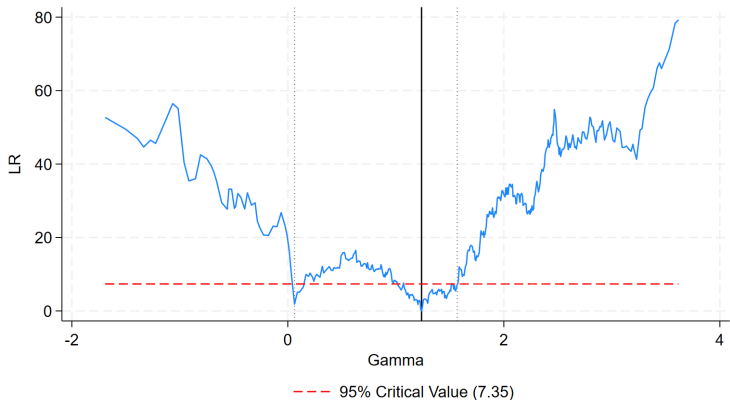
```
GTest    = 7.71
```

```
p-value = 0.04
```

Empirical Example

Post Estimation

estat graph lr and *estat graph ssr* plot the LR test and SSR across grid of γ

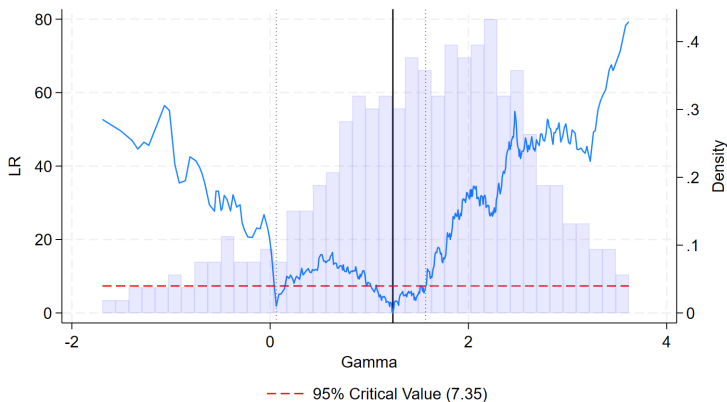


Estimated threshold: 1.24, 95%-CI: [0.06,1.57].

Empirical Example

Post Estimation

estat graph lr, histogram adds a histogram of γ in the background:



Estimated threshold: 1.24, 95%-CI: [0.06,1.57].

- Graph options for histogram can be set with `histogram(opt)`.
- *estat graph lr, kdensity* plots a kdensity plot. [► Example](#)

Empirical Example

Post Estimation

- *estat split* creates new variables for the regimes:

```
. estat split
Variables govspend khanvar split using estimated threshold.
```

- Splitted variables are saved in *r(varlist)*:

```
. qui estat split
. xtddce2 gdpccgr `r(varlist)' open popgr grosscap , ///
> pooled(all) cr(open popgr grosscap `r(varlist)') ///
> pooledvce(ols)
(Dynamic) Common Correlated Effects Estimator - Pooled

Panel Variable (i): id                Number of obs   =       3922
Time Variable (t): year                Number of groups =        74
Degrees of freedom per group:
  without cross-sectional averages     = 45
  with cross-sectional averages        = 38
Number of cross-sectional lags         0 to 0          F(599, 3323)    =    0.40
variables in mean group regression = 7              Prob > F        =    1.00
variables partialled out               = 592           R-squared       =    0.93
                                          R-squared (P)   =    0.05
                                          Root MSE       =    3.91
                                          CD Statistic    =   65.34
                                          p-value        =    0.0000
```

gdpccgr	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Pooled:						
govspend_0	-.3833193	.0392082	-9.78	0.000	-.460166	-.3064726
govspend_1	.1258445	.0196636	6.40	0.000	.0873045	.1643846
khanvar_0	-1.651545	.1749324	-9.44	0.000	-1.994406	-1.308684
khanvar_1	.3586473	.0942674	3.80	0.000	.1738866	.543408
open	.0279824	.0059451	4.71	0.000	.0163302	.0396346
popgr	-.8679307	.0979326	-8.86	0.000	-1.059875	-.6759864
grosscap	.1033035	.0185232	5.58	0.000	.0669986	.1396083

Pooled Variables: govspend_0 govspend_1 khanvar_0 khanvar_1 open popgr grosscap
Cross Sectional Averaged Variables: open popgr grosscap govspend_0 govspend_1 khanvar_0 khanvar_1
Heterogenous constant partialled out.
OLS standard errors for pooled coefficients.

Conclusion

- `xtthreshold` estimates and test threshold in panel data models with large N and T , and interactive fixed effects.
- Command flexible for models without interactive fixed effects, fixed effects, homo- and heterogenous coefficients.
- Post estimation commands allow inspection of estimated thresholds and creation of thresholded variables.
- Work in progress!!



<https://janditzen.github.io/xtthreshold/>



Ditzen et al. (2026)

References

- Barassi, M., Y. Karavias, and C. Zhu. 2026. Threshold Regression in Heterogeneous Panel Data with Interactive Fixed Effects. arXiv (2308.04057).
- Chudik, A., K. Mohaddes, M. H. Pesaran, and M. Raissi. 2017. Is there a debt-threshold effect on output growth? Review of Economics and Statistics 99(1): 135–150.
- Chudik, A., and M. H. Pesaran. 2015. Large Panel Data Models with Cross-Sectional Dependence: A Survey. In The Oxford Handbook Of Panel Data, ed. B. H. Baltagi, chap. 1, 2–45. Oxford: Oxford University Press.
- Chudik, A., M. H. Pesaran, and E. Tosetti. 2011. Weak and strong cross-section dependence and estimation of large panels. The Econometrics Journal 14(1): C45–C90.

References

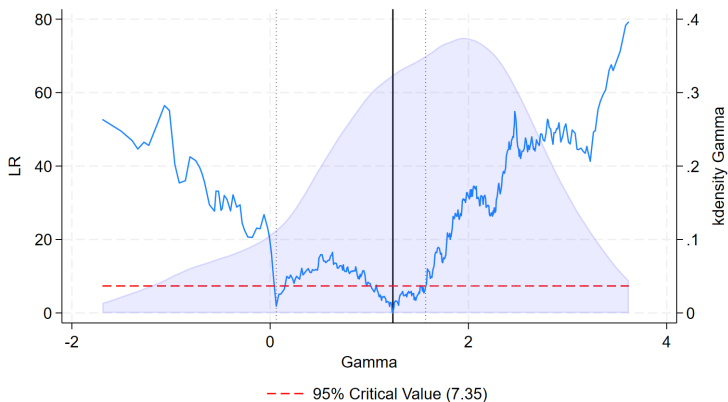
- Ditzen, J., Y. Karavias, and J. Westerlund. 2026. Threshold Regression for Fixed-T Panel Data with Interactive Fixed Effects. Oxford Bulletin of Economics and Statistics .
- Hansen, B. E. 1999. Threshold effects in non-dynamic panels : Estimation , testing , and inference 93.
- Kapetanios, G., M. H. Pesaran, and T. Yamagata. 2011. Panels with non-stationary multifactor error structures. Journal of Econometrics 160(2): 326–348.
- Pesaran, M. H. 2006. Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica 74(4): 967–1012.

- Asymptotics are for $N \rightarrow \infty$ with T **fixed**.
- Need $T > k + l$ and sufficient variation to identify both slopes and the threshold.
- $\hat{\theta}$ is $\sqrt{N}$ -consistent and asymptotically normal.
- $\hat{\gamma}$ is consistent; under shrinking threshold effects its convergence can be faster than $\sqrt{N}$.
- The error assumptions allow serial correlation and heteroskedasticity in the idiosyncratic component.
 $\Rightarrow$ The clean nuisance-parameter-free threshold distribution relies on a shrinking-threshold-effect assumption. Inference can be less accurate when the threshold effect is very large.
- Monte Carlo simulations in Ditzen et al. (2026) confirm small sample properties.

Empirical Example

Post Estimation

estat graph lr, kdensity adds a kernel density of γ in the background:



Estimated threshold: 1.24, 95%-CI: [0.06,1.57]. Kernel density shown as shaded area.