Panel Threshold regression models with endogeneous threshold variable

Chien-Ho Wang [1] Eric S. Lin [2]

[1]National Taipei University [2]National Tsinghua University

Panel data meeting, 2010

Outline

Threshold models under different types of data

- Cross section data

$$y_i = \begin{cases} \beta_1 x_i + u_i & \text{if } \ z_i \leq \gamma \\ \beta_2 x_i + u_i & \text{if } \ z_i > \gamma \end{cases}$$

Threshold models under different types of data

- Cross section data

$$y_i = \begin{cases} \beta_1 x_i + u_i & \text{if } z_i \leq \gamma \\ \beta_2 x_i + u_i & \text{if } z_i > \gamma \end{cases}$$

- Time series data

$$y_t = \begin{cases} \beta_1 y_{t-1} + u_t & \text{if } y_{t-1} \leq \gamma \\ \beta_2 y_{t-1} + u_t & \text{if } y_{t-1} > \gamma \end{cases}$$

Threshold models under different types of data

- Cross section data

$$y_i = \begin{cases} \beta_1 x_i + u_i & \text{if } z_i \leq \gamma \\ \beta_2 x_i + u_i & \text{if } z_i > \gamma \end{cases}$$

- Time series data

$$y_t = \begin{cases} \beta_1 y_{t-1} + u_t & \text{if } y_{t-1} \leq \gamma \\ \beta_2 y_{t-1} + u_t & \text{if } y_{t-1} > \gamma \end{cases}$$

- Panel data

$$y_{it} = \begin{cases} \beta_1 x_{it} + u_{it} & \text{if } z_{it} \leq \gamma \\ \beta_2 x_{it} + u_{it} & \text{if } z_{it} > \gamma \end{cases}$$

Possible estimation problems

If we want to obtain consistent estimators of $\hat{\beta}_1$ and $\hat{\beta}_2$, the threshold value γ must be estimated consistency in advance.

Possible estimation problems

If we want to obtain consistent estimators of $\hat{\beta}_1$ and $\hat{\beta}_2$, the threshold value γ must be estimated consistency in advance.

Under z_{it} is correlated with u_{it}, We cannot obtain consistent estimator for γ.

Threshold model with endogenous threshold variable

Consider the threshold model

$$y_i = \begin{cases} \beta_1 x_i + u_i & \text{if } q_i \leq \gamma \\ \beta_2 x_i + u_i & \text{if } q_i > \gamma \end{cases} \tag{1}$$

If we want to obtain consistent estimators of β_1 and β_2, q_i cannot be correlated with u_i. If q_i is correlated with u_i, we need new method to estimate γ, β_1 and β_2.

Bias-Correction Estimator for endogenous threshold models

Kourtellos, Stengos and Tan (2007) consider a threshold model with endogenous threshold variables like Equation(2).

$$y_i = \begin{cases} x_i\beta_1 + u_i & \text{if } q_i \leq \gamma \\ x_i\beta_2 + u_i & \text{if } q_i > \gamma \end{cases} \tag{2}$$

z_i are exogenous variables. q_i is an endogenous variable. The threshold equation is

$$q_i = z_i\pi + \upsilon_i. \tag{3}$$

Define the indicator variable

$$I_i = \begin{cases} 1 & \text{iff } \upsilon_i \leq \gamma - z_i\pi \\ 0 & \text{iff } \upsilon_i > \gamma - z_i\pi \end{cases}$$

The joint distribution between u_i and υ_i is defined as

$$\begin{pmatrix} u_i \\ \upsilon_i \end{pmatrix} | x_i, z_i \sim N\left(0 \quad \begin{pmatrix} \sigma_u^2 & \sigma_{u\upsilon} \\ \sigma_{u\upsilon} & 1 \end{pmatrix} \right)$$

Use the relationship between u_i and ϵ_i.

$$\begin{pmatrix} \epsilon_i \\ \upsilon_i \end{pmatrix} = \begin{pmatrix} 1 & -\sigma_{u\upsilon} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u_i \\ \upsilon_i \end{pmatrix}$$

Let $\kappa_1 = \sigma_{u\upsilon} = \rho_1 \sigma_u$, $\kappa_2 = \sigma_{u\upsilon} = \rho_2 \sigma_u$, and define

$$u_i = \kappa_1 \upsilon_i + \epsilon_i = \kappa_1 \lambda_{1i}(\gamma - z_i \pi) + e_i$$

where $\lambda_{1i}(\gamma - z_i\pi) = -\frac{\phi(\gamma - z_i\pi)}{\Phi(\gamma - z_i\pi)}$ (Inverse Mills bias correction item). We may get conditional expectations for each of the regimes.

$$E(y|x_1, z_1, \upsilon_i \leq \gamma - z_i\pi) = x_i\beta_1 + \kappa_1\lambda_{1i}(\gamma - z_i\pi)$$

$$E(y|x_2, z_2, \upsilon_i > \gamma - z_i\pi) = x_i\beta_2 + \kappa_2\lambda_{2i}(\gamma - z_i\pi)$$

When two regimes have the same error structure, THRET model can be estimate by

$$y_i = x_i\beta_2 + x_i(\gamma)\varphi + \kappa\lambda_i(\gamma - z_i\pi) + e_i, \qquad (4)$$

where $x_i(\gamma) = x_i I(q_i \leq \gamma)$ and $\varphi = \beta_2 - \beta_1$.
This estimation method looks like sample-selection model. The main difference about these two model is THRET model using all data.

Estimation of Threshold model with endogenous threshold variable

Kourtellos, Stengos and Tan (2007) use three steps to obtain consistent estimator.

1. Estimate the parameter vector π in Equation (3) by least square.
2. Estimate the threshold value $\hat{\gamma}$ by minimizing a concentrated two stage least square criterion using $\hat{\pi}$ from first stage.

$$S_n(\gamma) = \sum_{i=1}^{n} (y_i - x_i\beta_1 - x_i(\gamma)\varphi - \kappa\lambda_i(\gamma - z_i\hat{\pi}))^2.$$

3. Estimate the lease square estimates of the slop parameters based on the split samples implied by $\hat{\gamma}$.

PTHET models

We consider three kinds of panel threhold models with endogeneous threshold variable

1.

$$y_{it} = x_{it} I(q_{it} \leq \gamma)\beta_1 + x_{it} I(q_{it} > \gamma)\beta_2 + e_{it} \quad (5)$$

$$q_{it} = z_{it}\pi + u_{it}, \quad (6)$$

PTHET models

We consider three kinds of panel threhold models with endogeneous threshold variable

1.
$$y_{it} = x_{it} I(q_{it} \leq \gamma)\beta_1 + x_{it} I(q_{it} > \gamma)\beta_2 + e_{it} \quad (5)$$
$$q_{it} = z_{it}\pi + u_{it}, \quad (6)$$

2.
$$y_{it} = x_{it} I(q_{it} \leq \gamma)\beta_1 + x_{it} I(q_{it} > \gamma)\beta_2 + \eta_i + e_{it} \quad (7)$$
$$q_{it} = z_{it}\pi + u_{it}, \quad (8)$$

PTHET models

We consider three kinds of panel threhold models with endogeneous threshold variable

1.
$$y_{it} = x_{it} I(q_{it} \leq \gamma)\beta_1 + x_{it} I(q_{it} > \gamma)\beta_2 + e_{it} \quad (5)$$
$$q_{it} = z_{it}\pi + u_{it}, \quad (6)$$

2.
$$y_{it} = x_{it} I(q_{it} \leq \gamma)\beta_1 + x_{it} I(q_{it} > \gamma)\beta_2 + \eta_i + e_{it} \quad (7)$$
$$q_{it} = z_{it}\pi + u_{it}, \quad (8)$$

3.
$$y_{it} = x_{it} I(q_{it} \leq \gamma)\beta_1 + x_{it} I(q_{it} > \gamma)\beta_2 + \eta_i + e_{it} \quad (9)$$

$$q_{it} = z_{it}\pi + c_i + u_{it}, \tag{10}$$

PTHET model with fixed effects

Consider a panel threshold model with endogenous threshold variable and fixed effect:

$$y_{it} = x_{it} I(q_{it} \leq \gamma)\beta_1 + x_{it} I(q_{it} > \gamma)\beta_2 + \eta_i + e_{it} \tag{11}$$

$$q_{it} = z_{it}\pi + u_{it}. \tag{12}$$

Under $n \to \infty$ and T fixed, we may derive consistent estimators for β_1 and β_2.

Basic Assumptions

1. Assumption 1: $\{y_{it}, x_{it}, q_{it}, \Delta e_{it}\}$ is strictly stationary, ergodic.
2. Assumption 2: $\{y_{it}, x_{it}, q_{it}, e_{it} : 1 \leq i \leq n, 1 \leq t \leq T\}$ are from balanced panel data
3. Assumption 3: $u_{it}|z_{it} \sim \mathcal{N}(0, 1)$
4. Assumption 4': The joint distribution between Δe_{it} and u_{it} is defined as:

$$\begin{bmatrix} \Delta e_{it} \\ u_{it} \end{bmatrix} | x_{it}, z_{it} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma_e^2 & \gamma_j \\ \gamma_j & 1 \end{bmatrix} \right),$$

   where γ_j is covariance between Δe_{it} and u_{it}, $\gamma_j = \gamma_1$ when $q_{it} \leq \theta$ and $\gamma_j = \gamma_2$ when $q_{it} > \theta$.
5. Assumption 5: $n \to \infty$ and T is fixed.

Estimation of panel threshold model with endogenous threshold variable

- Using first difference transformation to eliminate fixed effect η_i in Equation (11).
- Estimate the parameter vector π in Equation (12) by least square.
- Estimate the threshold value $\hat{\gamma}$ by minimizing a concentrated two stage least square criterion using $\hat{\pi}$ from first stage.

$$S_n(\gamma) = \sum_{i=1}^{n}\sum_{t=1}^{T}(\Delta y_{it} - (x_{it} I(q_{it} \leq \gamma) - x_{it-1} I(q_{it-1} \leq \gamma))\beta_1$$

$$-(x_{it} I(q_{it} > \gamma) - x_{it-1} I(q_{it-1} > \gamma))\beta_2 - \kappa\lambda_i(\gamma - z_{it}\hat{\pi}))^2.$$

- Estimate the lease square estimates of the slop parameters based on the split samples implied by $\hat{\gamma}$.

Why not use fixed effect transformation? Because fixed effect transformation will generate heteroskedasticity under T small.

$$E(u_{it} - \bar{u}_i)(u_{is} - \bar{u}_i) = -\frac{\sigma_u^2}{T}.$$

PTHET with double fixed effects

If a panel threshold model with endogenous threshold variable and fixed effect:

$$y_{it} = x_{it} I(q_{it} \leq \gamma)\beta_1 + x_{it} I(q_{it} > \gamma)\beta_2 + \eta_i + e_{it} \quad (13)$$

$$q_{it} = z_{it}\pi + c_i + u_{it}, \quad (14)$$

we use the first difference transformation to cancel out fixed effect η_i in main equation.

For the threshold equation, we cannot use first difference transformation to elminiate fixed effect. The least square dummy variable method is used to obtain $\hat{\pi}$ and $\hat{c}_i$. The bias-correction estimator γ can be estimated under $\hat{\pi}$ and $\hat{c}_i$

Estimation of panel threshold model with endogenous threshold variable

- Using first difference transformation to eliminate fixed effect η_i in Equation (13).
- Estimate the parameter vector π and c_i in Equation (14) by least square dummy variable method.
- Estimate the threshold value $\hat{\gamma}$ by minimizing a concentrated two stage least square criterion using $\hat{\pi}$ and $\hat{c}_i$ from first stage.

$$S_n(\gamma) = \sum_{i=1}^{n} \sum_{t=1}^{T} [\Delta y_{it} - (x_{it} I(q_{it} \leq \gamma) - x_{it-1} I(q_{it-1} \leq \gamma))\beta_1$$

$$-(x_{it} I(q_{it} > \gamma) - x_{it-1} I(q_{it-1} > \gamma))\beta_2 - \kappa\lambda_i(\gamma - \hat{c}_i - z_{it}\hat{\pi})]^2.$$

- Estimate the lease square estimates of the slop parameters based on the split samples implied by $\hat{\gamma}$.

The data generation process considered is based on Equation (11), where the threshold value $\gamma = 2$. We fixed $\beta_2 = 1$ and $\beta_1 - \beta_2 = (0.05, 0.5)$. For threshold equation, we consider without fixed effect case:

$$q_{it} = 2 + 3z_{1it} + 3z_{2it} + u_{it}. \tag{15}$$

where x_{it}, z_{1it}, and z_{2it} are identical and mutually independent standard normal distributed. The fixed effects c_i and η_i follow $i.i.d.\ \mathcal{N}(0, 1)$.

For controlling the endogeneity, we generate $(\epsilon_{it}, u_{it}) \sim i.i.d.$ $\mathcal{N}(0, 1)$, and $e_{it} = \sigma_e^2(\kappa_o \epsilon_{it} + (1 - \kappa_o)u_{it})/\sqrt{\kappa_o^2 + (1 - \kappa_o)^2}$. So that the correlation between e_{it} and threshold variable q_{it} is given by $\rho = \kappa_o/\sqrt{\kappa_o^2 + (1 - \kappa_o)^2}$. We set $\kappa_o = (0.05, 0.5, 0.9)$ to simulate strong and weak correlations between e_{it} and u_{it}.

Table: Finite-sample Performance for Threshold Value Estimation ($\beta_1 - \beta_2 = 0.05; \gamma = 2$)

	Method	$T = 10$			$T = 20$		
$\kappa_0 = 0.05$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	-0.665	-0.036	24961.78	-0.628	0.265	25832.25
	FE	-11.345	-11.16	181688.4	-12.43	-12.257	210775.7
$N = 100$	pthet	-0.546	0.282	24626.43	-0.289	0.732	22597.33
	FE	-12.382	-12.185	209352.7	-13.285	-13.066	236098.5
$N = 200$	pthet	-0.265	0.968	23640.77	0.622	1.508	14370.53
	FE	-13.232	-13.232	234145.7	-14.115	-13.92	261835
$N = 500$	pthet	0.804	1.726	13200.37	1.57	1.961	5116.832
	FE	-14.396	-14.209	270869	-15.181	-15.034	297085.8

	Method	$T = 10$			$T = 20$		
$\kappa_0 = 0.5$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	1.76	1.843	4180.392	1.868	1.855	2629.118
	FE	-11.388	-11.183	181967.1	-12.502	-12.336	212850.7
$N = 100$	pthet	1.984	1.982	2716.145	1.986	2.057	1914.692
	FE	-12.347	-12.188	208403	-13.316	-13.139	236969.3
$N = 200$	pthet	2.096	2.055	2163.062	1.994	2.013	1368.487
	FE	-13.213	-12.977	233551.9	-14.142	-13.948	26267.85
$N = 500$	pthet	1.892	1.946	1272.687	2.017	2.018	831.415
	FE	-14.432	-14.236	272146.5	-15.204	-15.011	297784.2

Table: Finite-sample Performance for Threshold Value Estimation ($\beta_1 - \beta_2 = 0.05$) (continue)

	Method	$T = 10$			$T = 20$		
$\kappa_0 = 0.9$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	1.968	1.952	2903.662	1.94	1.97	1950.399
	FE	-11.404	-11.191	182333	-12.457	-12.226	211744.9
$N = 100$	pthet	2.087	2.061	2039.976	1.987	1.992	1346.099
	FE	-12.407	-12.201	210181.9	-13.177	-13.003	132646.9
$N = 200$	pthet	2.053	2.07	1388.162	1.98	1.991	810.065
	FE	-13.233	-13.022	233997.3	-14.21	-14.029	264870.4
$N = 500$	pthet	2.032	2.01	768.197	2.015	2.034	488.923
	FE	-14.43	-14.213	272078.7	-15.15	-14.985	296049.4

Table: Finite-sample Performance for Threshold Value Estimation ($\beta_1 - \beta_2 = 0.5; \gamma = 2$)

	Method	$T = 10$			$T = 20$		
$\kappa_0 = 0.05$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	1.981	1.986	166.392	2.001	1.996	22.96
	FE	-11.51	-11.381	185594.5	-12.452	-12.28	211379.6
$N = 100$	pthet	2.001	1.997	26.769	1.996	1.996	4.78
	FE	-12.348	-12.196	208522.5	-13.35	-13.218	237957.2
$N = 200$	pthet	1.997	1.998	5.771	1.998	1.998	1.294
	FE	-13.246	-13.081	234529.1	-14.171	-13.946	263563.2
$N = 500$	pthet	1.998	1.999	0.954	1.999	1.999	0.225
	FE	-14.31	-14.088	268120.9	-15.23	-15.059	298604.3

	Method	$T = 10$			$T = 20$		
$\kappa = 0.5$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	1.956	1.974	588.074	2.01	2.005	146.177
	FE	-11.397	-11.204	182358.2	-12.519	-12.303	213241.6
$N = 100$	pthet	1.987	1.991	147.534	2	1.996	37.577
	FE	-12.403	-12.22	209992.5	-13.288	-13.061	235957.1
$N = 200$	pthet	1.994	1.995	39.786	1.997	1.999	9.798
	FE	-13.218	-13.032	233865.7	-14.061	-13.888	259966.5
$N = 500$	pthet	1.997	1.999	6.236	2	1.999	1.536
	FE	-14.353	-14.207	269418.1	-15.198	-14.978	297707.8

Table: Finite-sample Performance for Threshold Value Estimation ($\beta_1 - \beta_2 = 0.5$)(continue)

	Method	$T = 10$			$T = 20$		
$\kappa_0 = 0.9$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	2.007	1.978	586.007	2.017	2.002	152.444
	FE	-11.449	-11.245	183448.1	-12.462	-12.259	211568.1
$N = 100$	pthet	2.005	2.001	163.956	2.002	1.994	35.211
	FE	-12.281	-12.117	206305.2	-13.229	-13.084	236487.9
$N = 200$	pthet	1.997	1.999	46.364	1.997	1.998	1.177
	FE	-13.332	-13.1	237440	-14.134	-13.936	262362.6
$N = 500$	pthet	1.997	1.999	6.466	2	1.999	0.147
	FE	-14.31	-14.072	268027.8	-15.18	-14.979	297017.2

Table: Finite-sample Performance for Main Regression Coefficient ($\beta_1 - \beta_2 = 0.05$)

	Method	$T = 10$			$T = 20$		
$\kappa_0 = 0.05$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	0.066	0.072	210.707	0.062	0.059	296.961
	FE	1.026	1.029	955.946	1.026	1.026	956.9
$N = 100$	pthet	0.038	0.064	363.922	0.046	0.059	745.058
	FE	1.025	1.023	953.157	1.025	1.025	952.595
$N = 200$	pthet	0.006	0.06	6352.047	0.045	0.054	48.168
	FE	1.026	1.026	953.414	1.025	1.025	951.868
$N = 500$	pthet	0.061	0.054	89.671	0.05	0.051	0.66
	FE	1.024	1.024	949.718	1.024	1.023	949.313

	Method	$T = 10$			$T = 20$		
$\kappa_0 = 0.5$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	0.037	0.039	6.476	0.044	0.045	2.826
	FE	1.025	1.024	953.712	1.024	1.022	951.684
$N = 100$	pthet	0.043	0.043	2.752	0.044	0.045	1.167
	FE	1.023	1.023	948.686	1.024	1.024	951.048
$N = 200$	pthet	0.045	0.047	1.326	0.044	0.044	0.621
	FE	1.025	1.026	951.615	1.025	1.024	950.783
$N = 500$	pthet	0.045	0.044	0.518	0.046	0.047	0.257
	FE	1.024	1.025	949.627	1.025	1.025	951.217

Table: Finite-sample Performance for Main Regression Coefficient ($\beta_1 - \beta_2 = 0.05$)(continue)

	Method	$T = 10$			$T = 20$		
$\kappa_0 = 0.9$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	0.046	0.045	5.483	0.044	0.044	2.309
	FE	1.025	1.025	953.695	1.027	1.028	959.098
$N = 100$	pthet	0.045	0.043	2.51	0.043	0.042	1.236
	FE	1.024	1.025	949.796	1.024	1.025	951.172
$N = 200$	pthet	0.045	0.043	1.291	0.045	0.044	0.553
	FE	1.025	1.024	951.191	1.026	1.028	953.335
$N = 500$	pthet	0.046	0.045	0.486	0.046	0.046	0.234
	FE	1.025	1.026	951.089	1.025	1.025	950.986

Table: Finite-sample Performance for Main Regression Coefficient ($\beta_1 - \beta_2 = 0.5$)

	Method		$T = 10$			$T = 20$	
$\kappa_0 = 0.05$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	0.507	0.511	4.431	0.504	0.506	1.961
	FE	1.25	1.251	567.927	1.25	1.25	566.405
$N = 100$	pthet	0.504	0.504	2.154	0.502	0.502	1.085
	FE	1.25	1.249	565.006	1.248	1.248	561.43
$N = 200$	pthet	0.502	0.502	1.067	0.501	0.501	0.528
	FE	1.25	1.25	563.493	1.249	1.25	562.564
$N = 500$	pthet	0.5	0.5	0.447	0.499	0.499	0.214
	FE	1.25	1.25	562.583	1.251	1.251	564.662

	Method		$T = 10$			$T = 20$	
$\kappa_0 = 0.5$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	0.485	0.489	6.411	0.489	0.49	2.69
	FE	1.251	1.252	568.459	1.249	1.249	564.645
$N = 100$	pthet	0.492	0.495	2.632	0.495	0.496	1.273
	FE	1.25	1.25	564.915	1.251	1.252	566.119
$N = 200$	pthet	0.498	0.498	1.219	0.498	0.498	0.554
	FE	1.25	1.25	564.07	1.25	1.25	563.196
$N = 500$	pthet	0.498	0.498	0.49	0.499	0.499	0.213
	FE	1.25	1.249	562.307	1.249	1.249	561.724

Table: Finite-sample Performance for Main Regression Coefficient ($\beta_1 - \beta_2 = 0.5$)(continue)

	Method	$T = 10$			$T = 20$		
$\kappa_0 = 0.9$		mean	median	MSE	mean	median	MSE
$N = 50$	pthet	0.481	0.483	6.197	0.487	0.487	3.039
	FE	1.249	1.25	565.868	1.252	1.25	569.449
$N = 100$	pthet	0.488	0.488	2.978	0.494	0.494	1.221
	FE	1.247	1.247	560.682	1.252	1.251	567.521
$N = 200$	pthet	0.495	0.495	1.374	0.498	0.498	0.536
	FE	1.251	1.25	564.748	1.251	1.25	565.302
$N = 500$	pthet	0.497	0.497	0.487	0.499	0.5	0.215
	FE	1.25	1.25	563.459	1.25	1.251	563.469

Future study and possible problems

- Panel threshold unit root case
- Panel smooth transition threshold model