

パネル二項ロジットモデルの個別効果検定

内田 智之*

長倉 大輔†

平成 29 年 2 月 8 日

要旨

本稿ではパネル二項ロジットモデルにおける個別効果の有無の検定方法を提案する。一般的に、個別効果を含んだパネル非線形モデルにおいて、時点数 T を固定し、個別主体数 N が大きくなっていく漸近論のもとでは通常の推定方法は一致性をもたないことが知られている。そのような場合でも一致性をもつ、より高度な推定方法は存在するが、それらの方法は通常非常に複雑であり、推定にかかるコストが高い。よってそのような方法に進む前に実際に個別効果が存在するかどうかを検定することは有意義であろう。本稿では Nagakura (2017) の方法をパネル二項ロジットモデルに適用することにより、パネル二項ロジットモデルにおける個別効果の有無の検定を提案する。シミュレーションによって、この検定の小標本での実際のサイズは十分正確であり、その検出力も良好であることを確認する。また応用例としてドイツの医療データへの応用を紹介する。

Key Words: パネルデータ; 二項ロジットモデル; LM 検定; シミュレーション.

*慶應義塾大学 経済学部

†慶應義塾大学 経済学部

1 はじめに

パネルデータとは横断面データと時系列データの2つのデータの特徴を併せ持ち、個体番号と時点という2つのインディックスにより特徴づけられるデータである。¹ 通常、時点数 T に対して個体数 N が大きいという特徴がある。パネルデータを用いた分析には、個別効果という各個体に固有な効果を考慮して分析できるという長所があり、このような分析は横断面データではできないものである。

横断面データと同様、パネルデータの分析においても線形のモデルが最もよく用いられる。線形のパネルデータモデルの場合、通常、個別効果は個体間の切片項の違いとしてモデル化される。この時、個々の個別効果(切片項)を推定すべき未知パラメーターとして扱った場合を固定効果モデル、確率的に変動するものとしてその期待値と分散によって特徴づけた場合を変量効果モデルと呼ぶ。線形モデルの場合、固定効果モデルにおいてはその個別効果および説明変数の係数は最小二乗ダミー変数推定によって一致推定することができる。また変量効果モデルにおいては個別効果の平均、分散、および説明変数の係数を一般化最小二乗法により一致推定できることが知られている。これらのモデルにおいて、個別効果の有無の検定は固定効果モデルにおいては F 検定を用いて、また変量効果モデルにおいては Breusch and Pagan (1980) の LM 検定を用いるのが一般的である。

これに対して、一般に非線形のパネルデータモデルにおいて固定効果モデルでは、時点数 T を固定し個体数 N が大きくなっていく漸近論のもとで説明変数の係数を一致推定するには線形パネルデータモデルの場合より複雑で計算コストのかかる推定法を使用する必要がある、また変量効果モデルにおいても、多くの場合、積分計算をする必要が出てくるため、想定される分布によってはやはり推定量を構築するのが難しい。また線形パネルデータモデルの場合と異なり、非線形パネルデータモデルの場合には個別効果の検定において一般的となっている検定法は存在しないのが現状である。

本稿では非線形パネルデータモデルの代表的なものであるパネル二項ロジットモデルにおいて、個別効果の有無の検定方法を提案する。パネル二項ロジットモデルにおいて個別効果を含んだモデルを考えた場合、いわゆる“incidental parameter”問題によって通常の最尤法による推定は一致性をもたないことが知られている。そのような場合でも一致性を保持するより高度な推定方法は存在するが、それらの方法は非常に複雑であり、推定にかかる計算コストが高い。よって、そのような推定に進む前に本稿の提案する検定法によって個別効果が実際に存在するかどうかを検定することは有意義であろう。本稿では Nagakura (2017) の方法をパネル二項ロジットモデルに適用することにより、パネル二項ロジットモデルにおける個別効果の有無の検定を提案する。シミュレーションによって、提案した検定法はの小標本における実際のサイズは実用上十分に正確であり、またその検出力も良好であることを確認する。また実際のデータへの応用例としてドイツの医療データへの応用を紹介する。

本稿の構成は以下ようになる。まず次章で具体的にパネル二項ロジットモデルの設定を行い、個別効果なしを帰無仮説、変量効果モデルを対立仮説とする LM 検定を導出する。3章でシミュレーションによって提案した LM 検定統計量の小標本での実際のサイズと検出力を確認する。4章で提案した LM 検定統計量の実際のデータへの応用例を紹介する。補論では、本文で省略された計算を述べる。

¹実際には2次元のインディックスにより特徴づけられるデータであれば必ずしも個体と時点である必要はない。

2 パネル二項ロジットモデルと個別効果検定

この節ではまず本稿で扱うパネル二項ロジットモデルについて説明し、その個別効果の有無を検定するための Lagrange Multiplier (LM) 検定統計量を Nagakura (2017) の方法に従い、導出する。

2.1 パネル二項ロジットモデル

N 個の個体を T 時点にわたり観測したデータがあるとする。 y_{it} を個体 i の個体の時点 t における観測値とし、1 もしくは 0 をとる従属変数とする。また $\mathbf{x}_{it} = [x_{it1}, \dots, x_{itK}]'$ を (定数項を含めない) K 次の説明変数ベクトルとする。本稿では説明変数 \mathbf{x}_{it} という条件付で y_{it} が 1 となる確率が以下のように与えられるパネル二項ロジットモデルを考える。

$$P(y_{it} = 1) = F_{it} = \frac{\exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it})}{1 + \exp(\alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + \varepsilon_{it})} \quad (1)$$

ここで α_i は個別効果で、個体間では異なるが時間を通じては一定の変数である。通常の二項ロジットモデルとの違いは、個体に依存した項である α_i が含まれている点である。個別効果がなければこの項は個別主体間で同一の値をとる。

本稿では α_i を確率変数とみなし、以下を仮定する。²

$$\alpha_i \sim N(\alpha, \sigma^2) \quad (2)$$

また α_i 個体間で独立であることも仮定する。このモデルにおいて、分散 σ^2 が 0 であれば、全ての個体で $\alpha_i = \alpha$ で一定となり、個別効果は存在しない。すなわち、上記のモデルにおいて個別効果がないという帰無仮説は

$$H_0 : \sigma^2 = 0, \quad \forall i \quad (3)$$

という制約条件と等しい。

本稿では上記の帰無仮説の検定のために Nagakura (2017) で提案された非線形パネルデータモデルにおける個別効果の有無の LM 検定を用いる。LM 検定では帰無仮説のもとでのパラメータの推定をする必要があるので、ここでそれらのパラメータの推定方法を確認する。最尤推定法を用いて各パラメータを推定する³。帰無仮説のもとで推定するパラメータは $\boldsymbol{\psi} = [\alpha, \boldsymbol{\beta}]'$ である。ここで α は個別効果の平均 $\boldsymbol{\beta}$ は係数ベクトルである。制約条件下の尤度関数は

$$L(\boldsymbol{\psi}) = \prod_{i=1}^N \prod_{t=1}^T f_{it}^0, \quad (4)$$

と表せる。ここで

$$f_{it}^0 = F_{it}^{y_{it}} (1 - F_{it})^{1-y_{it}} \quad (5)$$

であり、 F_{it} は (1) 式で与えられる。ここで、(5) 式の指数部分にある y_{it} はこれが 1、0 を示す質的従属変数であることを利用した指示関数の役割をしている。これにより y_{it} が 1、0 のどちらを示しても 1 つの式でその確率関数を示すことができる。次に (4) の対数をとった対数尤度関数を各パラメータのスコア関数 (対数尤度関数のパラメータについての 1 次偏導関数) を求める。 α についてのスコア関数は

$$\frac{\partial \log L}{\partial \alpha} = \sum_{i=1}^N \sum_{t=1}^T \frac{\partial \log f_{it}^0}{\partial \alpha} \quad (6)$$

²Nagakura (2017) では正規性の仮定は緩められる。

³線形モデルで用いられる最小二乗法による推定が非線形モデルでは適当でない理由については、Davidson and MacKinnon (1993) の Chapter8 を参照されたい

である。ここで $\partial \log f_{it}^0 / \partial \alpha$ は

$$\begin{aligned} \frac{\partial \log f_{it}^0}{\partial \alpha} &= y_{it} \frac{\partial \log F_{it}}{\partial \alpha} + (1 - y_{it}) \frac{\partial \log(1 - F_{it})}{\partial \alpha} \\ &= y_{it}(1 - F_{it}) - (1 - y_{it})F_{it} \\ &= y_{it} - F_{it} \end{aligned} \quad (7)$$

と求められる。同様に、 β についてのスコア関数は

$$\frac{\partial \log L}{\partial \beta} = \sum_{i=1}^N \frac{\partial \log f_{it}^0}{\partial \beta} = \sum_{i=1}^N (y_{it} - F_{it}) \mathbf{x}_{it} \quad (8)$$

となる。(6) 式に (7) 式および (8) 式を代入したものをそれぞれ 0 とおいた方程式を解くことによってこれらのパラメータを推定することができる。

2.2 ラグランジュ乗数検定 (LM 検定)

ここでは個別効果の有無を検定する LM 検定を導出する。LM 検定を定義するには対立仮説のもとでのパネル二項ロジットモデルにおける確率的個別効果 α_i の分散 σ^2 についてのスコア関数を帰無仮説の下で評価する必要がある。Nagakura (2017) の結果を用いると、これは

$$\frac{\partial \log L}{\partial \sigma^2} = \sum_{i=1}^N v_{i,\sigma}(\psi),$$

で与えられる。ここで、 $v_{i,\sigma}(\psi)$ は

$$v_{i,\sigma}(\psi) = \frac{1}{2} \left[\sum_{t=1}^T F_{it}(F_{it} - 1) + \left(\sum_{t=1}^T (y_{it} - F_{it}) \right)^2 \right] \quad (9)$$

である。また、 $\mathbf{z}_{it} = [1, \mathbf{x}'_{it}]'$ および

$$v_{i,\psi}(\psi) = \sum_{t=1}^T (y_{it} - F_{it}) \mathbf{z}_{it} \quad (10)$$

を定義すると、前項のスコア関数とまとめてスコアベクトルを、

$$s_{0,N}(\psi) = \sum_{i=1}^N [v'_{i,\psi}(\psi), v'_{i,\sigma}(\psi)]' \quad (11)$$

と表すことができる。これらの結果を用いて、LM 統計量は

$$LM_N = N^{-1} s'_{0,N}(\tilde{\psi}_N) \mathbf{IM}_N^{-1}(\tilde{\psi}_N) s_{0,N}(\tilde{\psi}_N) \quad (12)$$

と定義される。ここで、 $\tilde{\psi}_N$ は帰無仮説の下でのパラメータの最尤推定値、 $\mathbf{IM}_N(\psi)$ は情報行列で

$$\mathbf{IM}_N(\psi) = E_0 \left(\frac{1}{N} \sum_{i=1}^N \begin{bmatrix} v_{i,\psi}(\psi) v'_{i,\psi}(\psi) & v_{i,\psi}(\psi) v_{i,\sigma}(\psi) \\ v'_{i,\psi}(\psi) v_{i,\sigma}(\psi) & v_{i,\sigma}^2(\psi) \end{bmatrix} \right) \quad (13)$$

と定義される。ここで $E_0(\cdot)$ は帰無仮説の下での期待値である。 $\mathbf{IM}_N(\psi)$ の具体的な形は補論で述べる。

この LM 統計量、 LM_N は帰無仮説のもとで自由度 1 のカイ二乗分布に従う。よって例えば有意水準 5% で個別効果の有無を検定する場合は、カイ二乗分布の上側 5% 点と検定統計量を比較し、検定統計量のほうが大きければ個別効果がないという帰無仮説を棄却することになる。

3 シミュレーション

検定統計量が実際にカイ二乗分布に従っているかを、シミュレーションを用いて検討する。統計ソフト R (version 3.3.1) を用いて行った。パネルデータの設定として個別効果の次元は 1 とした。また、 T は 3、5、10、 N は 50、100、200 と変化させ、それぞれの影響も確認する。パラメータの設定としては帰無仮説である $\sigma_i^2 = 0$ および、対立仮説として、0.01、0.10、0.50 の場合をそれぞれ仮定したデータを生成させた。手順としては、(1) 説明変数を 1 セット生成し、(2) 同時に各 σ_i^2 の条件に従って個別効果 α_i も生成、(3) 以上から出した数値をパネルロジットモデルに代入し、(4) 一様分布からの乱数と比較することで (5) それぞれの説明変数に対応する従属変数が 1 と 0 のどちらになるかを棄却法⁴により判定した。ここでできたデータセットに対して、 LM 統計量を計算する。この作業を書く条件毎に 3000 回行うことで、 LM 統計量の分布図が作成される。これにより今回導いた LM 統計量が理論通りの分布になっていることを確認する。

また手順 (2) に際しては標準正規分布を生成することで個別効果を作成したが、必ずしも現実の分析対象の誤差分布が正規分布に従うとは言えないことを考慮して、カイ 2 乗分布や指数分布といった歪んだ分布を用いた場合の結果も示す。

実際にここまでの理論が正しいならば、 LM 統計量の分布は自由度 1 のカイ二乗分布に従う。そこで自由度 1 のカイ二乗分布の上側 1%、5% 点を上回る LM 統計量の割合を確認する。

3.1 個別効果のないデータが従う分布

個別効果 α_i を標準正規分布の乱数を用いて生成した場合の結果を記述する。 T は観測時間、 N は観測主体の数を表し、pu5 は、検定統計量のうちカイ二乗分布の上側 5% 点を上回る割合であり、pu1 は上側 1% 点を上回る割合である。

表 1: 帰無仮説下の分布

T	N	pu5	pu1
3	50	0.0510	0.0093
3	100	0.0453	0.0090
3	200	0.0480	0.0127
5	50	0.0450	0.0080
5	100	0.0587	0.0097
5	200	0.0537	0.0100
10	50	0.0470	0.0097
10	100	0.0420	0.0077
10	200	0.0483	0.0063

結果を見ると、カイ二乗分布の各点を上回る LM 統計量の割合は概ね理論通りである 5%、1% に従っているといえる。このことから、帰無仮説下においては理論通りの結果が得られた。

3.2 個別効果がある場合

次に対立仮説下の LM 統計量を計算する。これは検出力の計算を行うことになる。検出力とは、1 から「第二種の誤り」を引いたものであり、対立仮説が真であるときに帰無仮説を採択しない

⁴棄却法については、アクチュアリー会によるテキストモデリング (2015) を参考にした

確率を表す。ここでは対立仮設が「個別効果が存在する」ということなので、個別効果の分散 σ_i^2 の値を変えながら LM 検定統計量を計算する。表の各意味は先ほどと同じであるが pu1,pu5 がそのまま検出力になる。

表 2: 対立仮設毎の検出力

		$\sigma_i^2=0.01$		$\sigma_i^2=0.1$		$\sigma_i^2=0.5$	
T	N	pu5	pu1	pu5	pu1	pu5	pu1
3	50	0.0520	0.0107	0.1780	0.0770	0.1780	0.0770
3	100	0.0453	0.0103	0.3130	0.1407	0.3130	0.1407
3	200	0.0527	0.0113	0.5447	0.3220	0.5447	0.3220
5	50	0.0427	0.0107	0.4620	0.2763	0.4620	0.2763
5	100	0.0490	0.0107	0.7227	0.5340	0.7227	0.5340
5	200	0.0550	0.0130	0.9410	0.8423	0.9410	0.8423
10	50	0.0470	0.0080	0.8980	0.7990	0.8980	0.7990
10	100	0.0567	0.0107	0.9943	0.9847	0.9943	0.9847
10	200	0.0593	0.0160	1.0000	1.0000	1.0000	1.0000

この結果から分散が大きくなると検出力が大きくなっていることがわかる。

3.3 カイ二乗分布を仮定した場合の検定結果

次に、個別効果の分散に正規分布以外の確率分布を仮定した場合の結果を示す。まずは、自由度2のカイ二乗分布を仮定した。ここで自由度を小さくした理由は、歪んだ分布を仮定して、その結果を見るためである。シミュレーションでは、個別効果の誤差分散としてカイ二乗分布を正規化したものを用いて誤算分散の変化尺度とした。

表 3: カイ二乗に従う分散の計算結果

		$\sigma_i^2=0.00$		$\sigma_i^2=0.01$		$\sigma_i^2=0.10$		$\sigma_i^2=0.50$	
T	N	pu5	pu1	pu5	pu1	pu5	pu1	pu5	pu1
3	50	0.0543	0.0100	0.0430	0.0063	0.0540	0.0143	0.1527	0.0527
3	100	0.0540	0.0063	0.0470	0.0100	0.0607	0.0127	0.2460	0.0993
3	200	0.0507	0.0097	0.0507	0.0093	0.0740	0.0217	0.4480	0.2363
5	50	0.0510	0.0100	0.0423	0.0100	0.0753	0.0237	0.3617	0.1983
5	100	0.0500	0.0120	0.0467	0.0110	0.1040	0.0333	0.6130	0.4127
5	200	0.0520	0.0100	0.0553	0.0120	0.1370	0.0433	0.8793	0.7303
10	50	0.0460	0.0103	0.0443	0.0117	0.1690	0.0720	0.8223	0.6980
10	100	0.0513	0.0093	0.0523	0.0137	0.2460	0.1227	0.9813	0.9453
10	200	0.0553	0.0097	0.0613	0.0150	0.4227	0.2457	0.9997	0.9993

結果を見ると、帰無仮説、対立仮設のどちらの場合でも正規分布を仮定した場合と概ね変わらない結果が得られた。

3.4 指数分布を仮定した場合の検定結果

個別効果の分散を、指数分布から生成した場合の計算結果を示す。ここではパラメータ1の指数分布を仮定した。

表 4: 指数分布に従う分散の計算結果

		$\sigma^2=0.00$		$\sigma^2=0.01$		$\sigma^2=0.10$		$\sigma^2=0.50$	
T	N	pu5	pu1	pu5	pu1	pu5	pu1	pu5	pu1
3	50	0.0423	0.0083	0.0517	0.0090	0.0553	0.0103	0.1480	0.0617
3	100	0.0500	0.0080	0.0530	0.0110	0.0630	0.0150	0.2460	0.1063
3	200	0.0490	0.0070	0.0507	0.0097	0.0787	0.0200	0.4413	0.2427
5	50	0.0460	0.0117	0.0507	0.0090	0.0730	0.0220	0.3587	0.1977
5	100	0.0480	0.0080	0.0447	0.0073	0.0903	0.0310	0.6217	0.4187
5	200	0.0513	0.0110	0.0487	0.0100	0.1250	0.0410	0.8783	0.7507
10	50	0.0447	0.0097	0.0537	0.0117	0.1550	0.0703	0.8317	0.7070
10	100	0.0463	0.0087	0.0600	0.0117	0.2670	0.1263	0.9730	0.9330
10	200	0.0540	0.0070	0.0580	0.0127	0.4117	0.2330	0.9990	0.9980

結果を見ると、カイ二乗分布よりは精度が落ちている。これはパラメータ1の指数分布がカイ二乗分布よりも、その確率密度が歪んでいるためであると思われる。

4 応用例

ここでは本稿で提案した検定統計量を実際のデータに応用する。データは Journal of Applied Econometrics の Archive よりダウンロードできるドイツの診療データを用いる。オリジナルのデータではかなり1個人に関して入手できるデータにばらつきがある、いわゆるアンバランスドパネルデータであるため、それぞれの個人で全てのデータがそろそろように観測年を考慮、いわゆるバランスドパネルデータにするために、観測年1984年から1986年の3年間とし、全ての観測年のデータがそろそろ個人のデータのみを抽出した。結果として2371人の個人について上記の3年間のデータで分析を行う。具体的な使用データは

- (1) DOCTOR : 過去3ヵ月以内に医者に診て貰った場合は1、そうでない場合は0を取るデータ
- (2) HOSPITAL : 当該年度の昨年中に病院を訪れた場合は1、そうでない場合は0を取るデータ
- (3) AGE : 年齢
- (4) EDU : 就学年数
- (5) HHNINC : 家計の名目上の総収入
- (6) PUBLIC : 公的保険に加入しているなら1していないなら0をとるダミー変数
- (7) MARRIED : 婚姻していたら1していなかったら0をとるダミー変数。
- (8) FEMALE : 女性なら1男性なら0を取るダミー変数
- (9) BLUEC : ブルーカラーとして雇用されていれば1そうでなければ0を取るダミー変数
- (10) WHITEC : ホワイトカラーとして雇用されていれば1そうでなければ0を取るダミー変数
- (11) SELF : 自営業者であれば1そうでなければ0をとるダミー変数
- (12) BEAMT : 公務員であれば1そうでなければ0を取るダミー変数
- (13) HSAT : 自分の健康状態に満足しているかどうかを10段階で評価したもの

を用いる。これらの変数に対して、(I) DOCTOR を被説明変数、(3) から (13) までの変数を説明

変数としたパネル二項ロジットモデルに対して、個別効果の有無の検定 および (II) HOSPITAL を被説明変数、(3) から (13) までの変数を説明変数としたパネル二項ロジットモデルに対して、個別効果の有無の検定を行う。結果は以下のようになった。

表 5: ドイツにおける診断傾向の個別効果検定

DOCTOR	HOSPITAL
553.306	113.755

5 おわりに

本稿では二項ロジットモデルへのパネルデータを適用させる際に必要な個別効果検定の方法を明示的にすることが目的であった。検定統計量自体は比較的記述しやすい計算結果になり、諸条件での棄却力もある程度保たれているため、実用性は十分かと思われる。一方で現実的なパネルデータを見ると、従属変数が二値的なものよりも、いわゆる「多項ロジットモデル」や「順序ロジットモデル」を適応することが妥当であるようなものが多いことも事実である。そのために本稿における結果にとどまらず、「パネル多項ロジットモデル」や「順序ロジットモデル」への個別効果検定方法を明示的にすることがより実務に有益であると考えている。そのために今後はそれらの計算も行っていきたい。

6 補論

この補論では、LM 検定統計量の定義に用いられた情報行列の具体的な形を用いる。(13) 式を情報行列 $\mathbf{IM}_N(\boldsymbol{\psi})$ は

$$\mathbf{IM}_N(\boldsymbol{\psi}) = \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} E_0 \left[v_{i,\boldsymbol{\psi}}(\boldsymbol{\psi}) v'_{i,\boldsymbol{\psi}}(\boldsymbol{\psi}) \right] & E_0 \left[v_{i,\boldsymbol{\psi}}(\boldsymbol{\psi}) v_{i,\sigma}(\boldsymbol{\psi}) \right] \\ E_0 \left[v'_{i,\boldsymbol{\psi}}(\boldsymbol{\psi}) v_{i,\sigma}(\boldsymbol{\psi}) \right] & E_0 \left[v_{i,\sigma}^2(\boldsymbol{\psi}) \right] \end{bmatrix}$$

と書き直される。以下では $E_0 \left[v_{i,\boldsymbol{\psi}}(\boldsymbol{\psi}) v'_{i,\boldsymbol{\psi}}(\boldsymbol{\psi}) \right]$, $E_0 \left[v_{i,\boldsymbol{\psi}}(\boldsymbol{\psi}) v_{i,\sigma}(\boldsymbol{\psi}) \right]$, および $E_0 \left[v_{i,\sigma}^2(\boldsymbol{\psi}) \right]$ の具体的な形を求める。まず、 $E_0 \left[v_{i,\boldsymbol{\psi}}(\boldsymbol{\psi}) v'_{i,\boldsymbol{\psi}}(\boldsymbol{\psi}) \right]$ は

$$\begin{aligned} E_0 \left[v_{i,\boldsymbol{\psi}}(\boldsymbol{\psi}) v'_{i,\boldsymbol{\psi}}(\boldsymbol{\psi}) \right] &= \sum_{t=1}^T E_0 \left[(y_{it} - F_{it})^2 \mathbf{z}_{it} \mathbf{z}'_{it} \right] \\ &= \sum_{t=1}^T F_{it} (1 - F_{it}) \mathbf{z}_{it} \mathbf{z}'_{it} \end{aligned} \quad (14)$$

次に、

$$\begin{aligned} F_{it}(F_{it} - 1) + (y_{it} - F_{it})^2 &= (0 - F_{it})(1 - F_{it}) + (y_{it} - F_{it})^2 \\ &= (y_{it} - F_{it})(1 - y_{it} - F_{it}) + (y_{it} - F_{it})^2 \\ &= (y_{it} - F_{it})((1 - y_{it} - F_{it}) + (y_{it} - F_{it})) \\ &= (y_{it} - F_{it})(1 - 2F_{it}) \end{aligned} \quad (15)$$

を利用して

$$\begin{aligned}
E_0[v_{i,\psi}(\psi)v'_{i,\sigma}(\psi)] &= \sum_{t=1}^T E_0[(y_{it} - F_{it})\mathbf{z}_{it} \frac{1}{2}[F_{it}(F_{it} - 1) + (y_{it} - F_{it})^2]] \\
&= \frac{1}{2} \sum_{t=1}^T E_0[(y_{it} - F_{it})(y_{it} - F_{it})(1 - 2F_{it})\mathbf{z}_{it}] \\
&= \frac{1}{2} \sum_{t=1}^T (1 - F_{it})F_{it}(1 - 2F_{it})\mathbf{z}_{it},
\end{aligned} \tag{16}$$

および

$$\begin{aligned}
E_0[v_{i,\sigma}(\psi)v'_{i,\sigma}(\psi)] &= \sum_{t=1}^T E_0[\frac{1}{4}[(y_{it} - F_{it})(1 - 2F_{it})]^2] \\
&= \frac{1}{4} \sum_{t=1}^T E_0[(y_{it} - F_{it})^2](1 - 2F_{it})^2 \\
&= \frac{1}{4} \sum_{t=1}^T F_{it}(1 - F_{it})(1 - 2F_{it})^2
\end{aligned} \tag{17}$$

と求まる。

7 ソースコード

以下に使用したコードを記述する。

```

#Y0: panel data set
#N: number of time
#T: number of individual
#k: number of explanatory variable of individual i at time t

pbltest<-function(Y0,N,T,k){
  Y1=matrix(Y0[,1],nrow=N*T)
  X=as.matrix(Y0[,-1],nrow=N*T,ncol=k)
  X2=cbind(1,X)
  dataA=data.frame(Y=Y1,X)
  logit1=glm(Y~.,family=binomial(link="logit"),data=dataA)
  b=coef(logit1)
  z=t(b%*%t(X2))
  F=exp(z)/(1+exp(z))
  Y2=t(Y1)-t(F)
  Y3<-vector(length=N)
  for(i in 1:N){
    W=sum(Y2[((i-1)*T+1):(i*T)])
    Y3[i]=W
  }
  S1=t(X2)%*%t(Y2)
  S2=sum(Y3*Y3)+sum(F*(F-1))
  S0=t(t(c(S1,S2)))
  W2=matrix(0,ncol(X2),ncol(X2))
  for(i in 1:(N*T)){
    W3=X2[i,]
    W4=W3%*%t(W3)
    W5=F[i]*(1-F[i])*W4
    W2<-W2+W5
  }
  IMa=W2
  IMb=t((F*(1-F)*(1-2*F)))%*%t(t(X2))
  FZ=F*(1-F)
  FZ5=matrix(0,ncol=1,nrow=1)
  for(i in 1:N){
    FZ2=FZ[((i-1)*T+1):(i*T)]
    FZ4=FZ2%*%t(FZ2)

```

```

    FZ5<-FZ5+sum(FZ4[upper.tri(FZ4)]) # Correct
  }
  IMc=sum(F*(1-F)*(1-2*F)*(1-2*F))+4*FZ5
  IM=cbind(rbind(IMa,IMb),rbind(t(IMb),IMc ))
  LM=t(S0)%%solve(IM)%%      S0
  LM<-c(LM)
  return(LM)
}

genPBL=function(a,omg2,b,X,N){
# X= [x_{11}';...;x_{1T}';x_{21}';...;x_{2T}';...;X_{N1};...;X_{NT}]
# X: NT*(L-1) matrix of explanatory variables
  Lb=length(b); L=Lb+1; T=nrow(X)/N; Xb= X%*%b;
  aind=a+sqrt(omg2)*rnorm(N);
  aall=aind%x%matrix(1,T,1);
  P=exp(aall+Xb)/(1+exp(aall+Xb)); # NT vector of probabilities
  U=runif(N*T);
  y=as.integer(U<P); #N*T vector of responses (dependent variables)
  Y=t(matrix(y,T,N)); #N*T matrix of responses (dependent variables)
  return(list(v=y,M=Y))
}

# Experiment

LMsim=function(N,T){
  Sim=3000; #Number of iteration
  # parameter values
  a=0; omg2=0; b=c(1);
  Lx=length(b);
  # N: number of individual
  # T: number of time periods
  # Lx: length of explanatory vector of individual i at time t
  #      (expect for constant term)
  RltST=matrix(0,Sim,1)
  sim=1;
  while(sim<Sim+1){
# Generate Sampel from usual ranodm coefficient model
# explanatory variable
  X=matrix(rnorm(N*T*Lx),N*T,Lx);
# dependent variable
  Y=genPBL(a,omg2,b,X,N);
  y=Y$v;
  Y0=cbind(y,X); Su=pbltest(Y0,N,T,Lx); RltST[sim,1]=Su;
  sim=sim+1
}
  return(RltST)
}

Sim=3000

#T=3
sim11<-LMsim(50,3)
pu511=sum(sim11[,1]>3.84)/Sim; pu111=sum(sim11[,1]>6.63)/Sim;
sim12<-LMsim(100,3)
pu512=sum(sim12[,1]>3.84)/Sim; pu112=sum(sim12[,1]>6.63)/Sim;
sim13<-LMsim(200,3)
pu513=sum(sim13[,1]>3.84)/Sim; pu113=sum(sim13[,1]>6.63)/Sim;

#T=5
sim21<-LMsim(50,5)
pu521=sum(sim21[,1]>3.84)/Sim; pu121=sum(sim21[,1]>6.63)/Sim;
sim22<-LMsim(100,5)
pu522=sum(sim22[,1]>3.84)/Sim; pu122=sum(sim22[,1]>6.63)/Sim;

```

```

sim23<-LMsim(200,5)
pu523=sum(sim23[,1]>3.84)/Sim;pu123=sum(sim23[,1]>6.63)/Sim;

#T=10
sim31<-LMsim(50,10)
pu531=sum(sim31[,1]>3.84)/Sim;pu131=sum(sim31[,1]>6.63)/Sim;
sim32<-LMsim(100,10)
pu532=sum(sim32[,1]>3.84)/Sim;pu132=sum(sim32[,1]>6.63)/Sim;
sim33<-LMsim(200,10)
pu533=sum(sim33[,1]>3.84)/Sim;pu133=sum(sim33[,1]>6.63)/Sim;

T=as.integer(c(3,3,3,5,5,5,10,10,10))
N=as.integer(c(50,100,200,50,100,200,50,100,200))
pu5=c(pu511,pu512,pu513,pu521,pu522,pu523,pu531,pu532,pu533)
pu1=c(pu111,pu112,pu113,pu121,pu122,pu123,pu131,pu132,pu133)

ResultData<-data.frame(T,N,pu5,pu1)

```

参考文献

- Breusch, T.S. and Pagan, A. R. (1980) The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics, *Review of Economic Studies*, 47(1), 239-253.
- Nagakura, D. (2017) Testing for individual effects for nonlinear panel data models, mimeo.
- Davidson, R., and MacKinnon, J. G. (1993) *Estimation and Information in Econometrics*. New York Oxford: Oxford university press

シミュレーション・日本アクチュアリー会