

Consistent estimation for the full-fledged fixed effects
zero-inflated Poisson model
(本格的な固定効果ゼロ過剰ポアソンモデルの一致推定)

北沢良継(九州産業大学)

Yoshitsugu Kitazawa (Kyushu Sangyo University)

E-mail: kitazawa@ip.kyusan-u.ac.jp

日本経済学会2014年度秋季大会

(2014 Autumn Meeting , Japanese Economic Association)

2014年10月11日, 12日, 西南学院大学

(Seinan Gakuin University, 2014/10/11-12)

論文のダウンロード (paper download):

<http://www.ip.kyusan-u.ac.jp/keizai-kiyo/dp66.pdf>

概要

- 本論文は**本格的な固定効果ゼロ過剰ポアソンモデル**の一致推定のために使われる変換を提唱する。**ここで用いられる本格的な固定効果ゼロ過剰ポアソンモデル**のゼロは、ロジット部とポアソン部の両方から発生する可能性があって、固定効果はロジット部とポアソン部の両方に存在する。
- **妥当な積率条件**が、提唱された変換に基づいて構築される。
- 積率条件を使ったGMMとEL推定量の小標本特性がモンテカルロ実験で調査される。
- **キーワード:** 固定効果ゼロ過剰ポアソン(ZIP)モデル; ポアソンの部分の先決説明変数; 積率条件; GMM; EL; モンテカルロ実験

1 序

- Lambert (1992) によって提唱された洗練されたゼロ過剰ポアソンモデル (ZIP (zero-inflated Poisson) モデル) はゼロ値が過剰に存在する計数 (カウント) データを取り扱うモデルの1つである。
- ZIPモデルを使った実証分析は計量経済分析の文献によくある: Gurmu and Trivedi (1996) (レジャー・ボート旅行とボートの所有者の属性との関係について), Crépon & Duguet (1997) and Hu & Jefferson (2009) (特許とR&Dの関係について) など。

ZIP モデル(簡単な例)

非負整数値

- 計数(カウント)値従属変数: y_i (例. 0, 1, 2, 3, 4, 5 ...)
- 説明変数: w_i, x_i
- $i = 1, \dots, N$ ($N \rightarrow \infty$)
- $y_i = 0$ with probability $1 - p_i$
- $y_i \sim Pois(q_i)$ with probability p_i
- ロジット確率 $p_i = \frac{\exp(\gamma + \delta w_i)}{1 + \exp(\gamma + \delta w_i)}$
- ポアソン平均 $q_i = \exp(\alpha + \beta x_i)$
- パラメータ $\gamma, \delta, \alpha, \beta$ は通常、最尤法で一致推定される。

通常、連続変数

初期の固定効果を有する ZIP モデル

- Majo (2010) 及び Majo & Van Soest (2011) が考えた固定効果ZIPモデルはポアソン部にゼロ切断ポアソンモデル(truncated-at-zero Poisson model)を仮定している。これの意味することはゼロ値計数(カウント)変数はロジット部からしか出てこないということである。
- Gilles (2012) and Gilles & Kim (2013) は、また、固定効果ZIPモデルを考えたが、彼らのモデルはロジット部に固定効果を入れていない。

本論文で議論される固定効果ZIPモデル

- Majo (2010) 及び Majo & Van Soest (2011) と Gilles (2012) 及び Gilles & Kim (2013)とは違って、**本論文で議論される固定効果ZIPモデル**では、**ポアソン部からのゼロ値計数の発生**がありえて、**ロジット部には固定効果が備わっている**。
- したがって、本論文で議論される ZIP モデルは比較的完璧といえる。

本論文で議論される固定効果ZIPモデル の推定法

- この ZIP モデルに対する**妥当な積率条件**は、ポアソン部の説明変数の違った定式化に基づいて構築される。
- それから、興味あるパラメータが、Hansen (1982) によって提唱された**GMM (一般化積率法)**と Owen (1988, 1990, 1991, 2001) によって提唱され、Qin & Lawless (1994) によって発展させられた**EL (経験尤度) 法**を使って一致推定される。

2 モデルと積率条件

- この論文で議論される固定効果 ZIP モデルはゼロ値 計数値変数の2つの潜在的な発生源(ロジット確率とポアソン密度)を有し、ロジット部とポアソン部の両方に固定効果を具備している。
- 固定効果 ZIP モデルは陰伏形式で記述され、その攪乱項の平均と分散が指定される。それから、攪乱項とその二乗が過去の攪乱項と固定効果を使って構成する任意の変換と無相関であるということを前提として、興味あるパラメータを一致推定するための積率条件が、若干強い仮定と緩和された仮定の下で構築される。

2 モデルと積率条件

- 若干強い仮定の下では、ロジット確率とポアソン平均の両方に中の説明変数は、若干外生である。一方、緩和された仮定の下では、ロジット確率の中の説明変数は若干外生で、ポアソン平均の中の説明変数は先決である。
- 本論文で導入される若干外生のニュアンスは、ある任意の時点の計数値従属変数は、計数値変数の発生直後の時点の説明変数には影響を与えないが、その後の説明変数にはなんらかの影響を与える可能性がある、ということである。

2.1 固定効果ZIPモデル

- 固定効果ZIPモデルは次の2つのゼロ値計数值従属変数の発生源を持つ:
- $y_{it} = 0$ with probability $1 - p_{it}$
- $y_{it} \sim Pois(q_{it})$ with probability p_{it}
- 下添字 i ($i = 1, \dots, N$) 及び t ($t = 1, \dots, T$) は、それぞれ、個別主体(individual)と時点を表す。
- $N \rightarrow \infty$ であるが T は固定であると仮定する。

2.1 固定効果ZIPモデル

- 二値過程を発生させる**ロジット確率**

$$p_{it} = \frac{\exp(\psi_i + \delta w_{it})}{1 + \exp(\psi_i + \delta w_{it})}$$

- **ポアソン過程**を発生させる平均

$$q_{it} = \exp(\phi_i + \beta x_{it})$$

- ψ_i と ϕ_i : 固定効果
- w_{it} と x_{it} : (連続) 説明変数
- 固定効果ZIPモデルの**陰伏形式**(Implicit form)

$$y_{it} = p_{it} q_{it} + v_{it}$$

- v_{it} : 攪乱項 (若干強い仮定についてのものと緩和された仮定についてのもの)

2.1 若干強い仮定と積率条件

- 若干強い仮定:

- $E[v_{it} | \psi_i, w_i^{t+1}, \eta_i, x_i^{t+1}, v_i^{t-1}] = 0$

- $E[v_{it}^2 | \psi_i, w_i^{t+1}, \eta_i, x_i^{t+1}, v_i^{t-1}]$
 $= p_{it}q_{it}(1 + (1 - p_{it})q_{it})$

- ここで、 $w_i^{t+1} = (w_{i1}, \dots, w_{i,t+1})$,

$$x_i^{t+1} = (x_{i1}, \dots, x_{i,t+1}),$$

$$v_i^{t-1} = (v_{i0}, \dots, v_{i,t-1}) \text{ (ただし、} v_{i0} \text{ は空集合)}$$

2.1 若干強い仮定と積率条件

- 若干強い仮定の下での条件付積率制約
- $E[\Phi_{it}(\delta, \beta) | \psi_i, w_i^t, \eta_i, x_i^t, v_i^{t-2}] = 0$
- $$\begin{aligned} \Phi_{it}(\delta, \beta) = & (\tanh(\delta \Delta w_{it}/2) - 1) \exp(-\beta \Delta x_{it}) (y_{it}^2 - y_{it}) \\ & + (\tanh(\delta \Delta w_{it}/2) + 1) \exp(\beta \Delta x_{it}) (y_{i,t-1}^2 - y_{i,t-1}) \\ & - 2 \tanh(\delta \Delta w_{it}/2) y_{it} y_{i,t-1} \end{aligned}$$
- 本論文では以上の変換を“PHI 変換”と呼ぶ。

2.2 緩和された過程と積率条件

- 緩和された仮定:

- $E[v_{it} | \psi_i, w_i^{t+1}, \eta_i, x_i^t, v_i^{t-1}] = 0$

- $E[v_{it}^2 | \psi_i, w_i^{t+1}, \eta_i, x_i^t, v_i^{t-1}]$
 $= p_{it}q_{it}(1 + (1 - p_{it})q_{it})$

2.2 緩和された仮定と積率条件

- 緩和された仮定の下での条件付積率制約
- $E[\Psi_{it}(\delta, \beta) | \psi_i, w_i^t, \eta_i, x_i^{t-1}, v_i^{t-2}] = 0,$
- $\Psi_{it}(\delta, \beta) = (\tanh(\delta \Delta w_{it}/2) - 1) \exp(-2\beta \Delta x_{it}) (y_{it}^2 - y_{it})$
 $+ (\tanh(\delta \Delta w_{it}/2) + 1) (y_{i,t-1}^2 - y_{i,t-1})$
 $- 2 \tanh(\delta \Delta w_{it}/2) \exp(-\beta \Delta x_{it}) y_{it} y_{i,t-1}$
- 本論文では以上の変換を“PSI 変換”と呼ぶ。

3 推定法

- PHI変換とPSI変換に基づいた無条件積率条件を使った2つの推定量: GMM 推定量、及び、EL 推定量。
- **GMM推定量**は、積率条件の標本版ベクトルと重み付け行列から構成される**二次形式**を最小化することによって得られる。
- GMM推定量に取って代わる推定量としての**EL 推定量**はインプライド確率によって重み付けされた積率条件の標本版ベクトルの制約のもとでインプライド確率を使って構築された**対数尤度**を最大化することによって得られる。
- 多くの研究によって明らかになったことは、EL 推定量はGMM 推定量よりも小標本でより**よく振る舞う**ということである。(例 Newey & Smith, 2004; Anatolyev, 2005; Ramalho, 2005).

3.1 GMM 推定量

目的関数

$$\hat{\theta}_{\text{GMM}} = \arg \min_{\theta} \bar{g}(\theta)' (\bar{\Omega}(\hat{\theta}_1))^{-1} \bar{g}(\theta).$$

無条件積率条件の経験対応 (m by 1)

$$\bar{g}(\theta) = (1/N) \sum_{i=1}^N g_i(\theta) = 0,$$

重み付け行列の逆行列
(m by m)

$$\bar{\Omega}(\hat{\theta}_1) = (1/N) \sum_{i=1}^N g_i(\hat{\theta}_1) g_i(\hat{\theta}_1)',$$

- 興味あるパラメータのベクトル: $\theta = [\delta, \beta]$
- ベクトルの1ステップ推定値: $\hat{\theta}_1$

無条件積率条件 $E[g_i(\theta)] = 0$,
(m by 1), は条件付き積率条件
に基づいて構築される。

3.2 EL 推定量

目的関数

$$\min_{\theta, \pi_1, \dots, \pi_N} -(1/N) \sum_{i=1}^N ((-\ln(1/N)) - (-\ln \pi_i)),$$

subject to

$$\sum_{i=1}^N \pi_i = 1, \quad \sum_{i=1}^N \pi_i g_i(\theta) = 0,$$



$$\hat{\theta}_{\text{EL}} = \arg \min_{\theta} (\max_{\lambda} (1/N) \sum_{i=1}^N \ln(1 - \lambda' g_i(\theta))),$$

双対問題に対する変換によって、推定されるパラメータ数は $2+N$ から $2+m$ へと減少する。(ただし $N > m$.)

$g_i(\theta)$ を構成する変数の実現確率: π_i

ラグランジエ乗数 (m by 1): λ

GMM推定量とEL推定量の漸近分布

- Qin & Lawless (1994) が示したのは、EL 推定量 $\hat{\theta}_{EL}$ は GMM 推定量 $\hat{\theta}_{GMM}$ と同じ漸近分布を持つということである。すなわち、

$$N^{1/2}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N(0, (D(\theta_0)'(\Omega(\theta_0))^{-1}D(\theta_0))^{-1}),$$

- ここで、

$$D(\theta_0) = (\partial E[g_i(\theta)] / \partial \theta') \big|_{\theta=\theta_0} \quad \Omega(\theta_0) = E[g_i(\theta_0)g_i(\theta_0)'].$$

- θ_0 : θ の真の値

4 モンテカルロ

- PHI変換及びPSI変換に基づいたGMM推定量とEL推定量の有限標本特性がモンテカルロ実験で調査される。
- モンテカルロ実験はR Core Team (2013)によって開発されたプログラム言語R (version 3.0.2)を使って実施される。[GMM推定及びEL推定：パッケージ“gmm”「開発者 Chaussé (2010)」, ML推定：パッケージ“pscl”「開発者 Zeileis et al. (2008).」]

4.1 データ発生過程

- DGP (固定効果ZIPモデル)

$$y_{it} = y_{it}^p y_{it}^q,$$

$$y_{it}^p \sim \text{Bin}(1, p_{it}),$$

$$p_{it} = \exp(\psi_i + \delta w_{it}) / (1 + \exp(\psi_i + \delta w_{it})),$$

$$w_{it} = \alpha w_{i,t-1} + \iota \psi_i + \zeta_{it},$$

$$w_{i1} = (1/(1 - \alpha)) \iota \psi_i + (1/(1 - \alpha^2)^{(1/2)}) \zeta_{i1},$$
$$\psi_i \sim N(0, \sigma_\psi^2); \zeta_{it} \sim N(0, \sigma_\zeta^2),$$

$$y_{it}^q \sim \text{Pois}(q_{it}),$$

$$q_{it} = \exp(\eta_i + \beta x_{it}),$$

$$x_{it} = \rho x_{i,t-1} + \tau \eta_i + \varepsilon_{it},$$

$$x_{i1} = (1/(1 - \rho)) \tau \eta_i + (1/(1 - \rho^2)^{(1/2)}) \varepsilon_{i1},$$
$$\eta_i \sim N(0, \sigma_\eta^2); \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2).$$

横断面サイズ:

N=1000, 5000, 10000

時点数:

T= 4, 8

モンテカルロ複製数:

10000.

値が以下のパラメータに

セットされる: $\delta, \alpha, \iota, \sigma_\psi^2,$

$\sigma_\zeta^2, \beta, \rho, \tau, \sigma_\eta^2, \sigma_\varepsilon^2.$

4.2 調査される推定量

- PHI変換とPSI変換に基づいた無条件積率条件を使ったGMM推定量とEL推定量

PHI変換に基づいた無条件積率条件

$$E[\Phi_{it}(\delta, \beta) \Delta w_{it}] = 0, \quad \text{for } t = 2, \dots, T,$$

$$E[\Phi_{it}(\delta, \beta) \Delta x_{it}] = 0, \quad \text{for } t = 2, \dots, T.$$

PSI変換に基づいた無条件積率条件

$$E[\Psi_{it}(\delta, \beta) \Delta w_{it}] = 0, \quad \text{for } t = 2, \dots, T,$$

$$E[\Psi_{it}(\delta, \beta) x_{is}] = 0, \text{ for } s = 1, \dots, t - 1; t = 2, \dots, T.$$

- 比較対照として、(一致性のない)プールされた最尤推定量(“ML(POOL)”推定量)が使われる。この推定量は固定効果を考慮していないので、固有のバイアスがある。

4.3 結果

- $T=4$ と $T=8$ に対して調査された推定量のモンテカルロ実験の結果が Table 1 と Table 2 にそれぞれ示されている。
- Figure 1 と Figure 2 は、それぞれ、 $T=4$ のときの δ と β の GMM 推定量 and EL 推定量の箱髭図である。一方、Figure 3 と 4 は、 $T=8$ のときのそれである。

固定効果ZIPモデルの モンテカルロ実験の結果, $T=8$

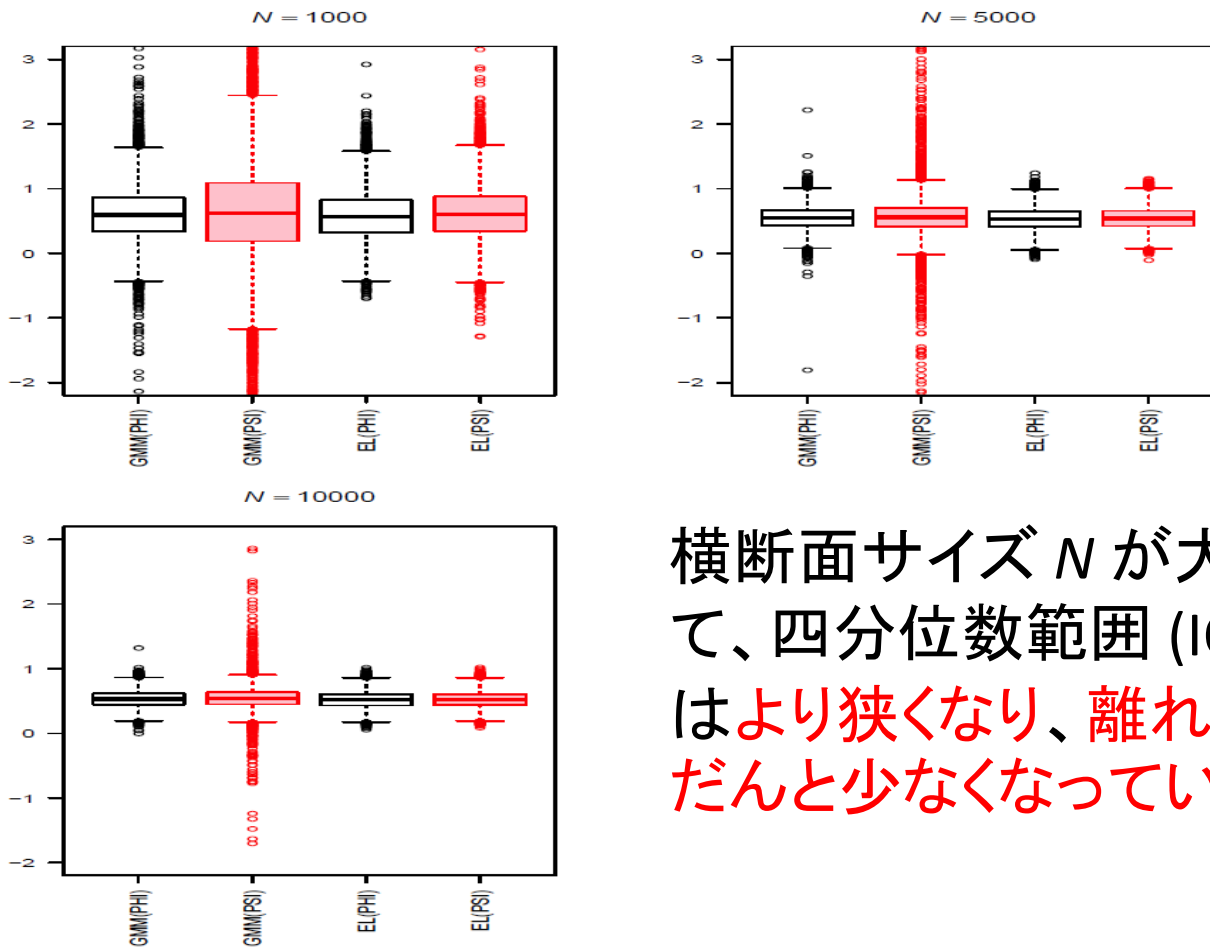
Table 2: Monte Carlo results for the fixed effects ZIP model, $T = 8$

		$N = 1000$		$N = 5000$		$N = 10000$	
		bias	rmse	bias	rmse	bias	rmse
GMM(PHI)	δ	0.085	2.633	0.046	0.212	0.032	0.131
	β	0.001	0.073	0.001	0.035	0.000	0.026
GMM(PSI)	δ	1.982	106.559	0.043	3.052	0.061	1.761
	β	-0.114	0.222	-0.032	0.080	-0.017	0.051
EL(PHI)	δ	0.082	0.395	0.027	0.177	0.018	0.130
	β	0.003	0.072	0.001	0.036	0.000	0.027
EL(PSI)	δ	0.119	0.440	0.040	0.178	0.025	0.129
	β	0.015	0.116	0.005	0.055	0.003	0.040
ML(POOL)	δ	0.342	0.346	0.341	0.342	0.341	0.342
	β	0.476	0.479	0.477	0.478	0.477	0.477

GMM 推定量と **EL** 推定量の**バイアス**の絶対値と **rmse** は、**横断面サイズが増えるにつれて、縮小していくが**、これは、両推定量の**一致性**を反映している。一方、**一致性のない ML(POOL) 推定量**の**かなりの上方バイアス**は**横断面サイズが増えても変化しない**。

δ についての GMM 推定量と EL 推定量の モンテカルロ実験の結果の箱髭図, $T=8$

Figure 3: Monte Carlo boxplots of the GMM and EL estimates for δ , $T = 8$



横断面サイズ N が大きくなるにつれて、四分位数範囲 (IQR) と髭の範囲はより狭くなり、離れた外れ値はだんだんと少なくなっていく。

ELはGMMより優れている

- (緩和された課程に基づいた) PSI 変換を用いるとき、**EL 推定量の小標本特性は GMM 推定量の小標本特性を圧倒的に凌駕する。** PSI 変換を用いるとき、GMM推定量の小標本特性は極端に悪い。これは、EL(PSI) 推定量の性能と GMM(PSI) 推定量の性能を比較すればわかる。
- より小さいバイアスとrmse、より狭い IQR と鬚の範囲、そして、あまり離れていない外れ値が EL 推定量に認められる。

EL が GMM よりも優れている理由 (1)

- GMM(PSI) 推定量は、Bound et al. (1995) and Staiger & Stock (1997) によって指摘された弱い操作変数の問題を被っているかもしれない。
- すなわち、PSI変換に基づいた積率条件の中の説明変数 x_{it} のラグ付きレベル(lagged level)はPSI変換に対する弱い操作変数であり得る(のではないか)。
- EL 推定量は以上の問題を解決し得るだろう。

EL が GMM よりも優れている理由 (2)

- GMM(PSI) (2ステップ) 推定量は**高次のバイアス**に苦しめられているかもしれない。高次のバイアスとは GMM 推定量に特有なもので Newey & Smith (2004) によって示された。高次のバイアスによって、GMM(PSI) (2ステップ) 推定量は貧弱な小標本特性を持つのではないか。時点数 T が増えるにつれて、GMM 推定量は PSI 変換に対してどんどんたくさんの操作変数を使うようになるということから判断して、高次のバイアスに苦しめられているのは十分考えられることである。
- EL 推定量は以上の問題を解決し得るだろう。

固定効果ZIPモデルを 一致推定するために廃棄された標本

- $(y_{it}, y_{i,t-1}) = (0,0), (0,1), (1,0)$ となる観測値は、GMM 推定量や EL 推定量を使った識別に**何の貢献もしない** (the PHI 変換と PSI 変換を見よ)。
- ここでのDGPでは以上の従属変数の組み合わせは、モンテカルロの各複製に対して**約70%**に達する。それらの組み合わせは、推定の際、**廃棄される**。
- したがって、**かなりの大きさの標本サイズ**が、GMM 推定量と EL 推定量の**正確度**と**精度**を高めるために必要とされるだろう。それは、モンテカルロ実験の結果に反映されている。

5 結論

- **固定効果ZIPモデル**において興味のあるパラメータを一致推定する二つのタイプの積率条件が提唱された。この固定効果ZIPモデルでは、ゼロ値の計数値がポアソン部とロジット部の両方から発生し、固定効果が両部に備え付けられている:
- **ロジット部とポアソン部の両方**において説明変数が**若干外生**である場合の**積率条件**、及び、**ロジット部**で説明変数が**若干外生**であり、**ポアソン部**で**先決**である場合の**積率条件**。
- モンテカルロ実験が示したことは、**たくさんの個別主体(individuals)**が、正確度と精度が高い GMM 推定値と EL 推定値を得るために必要であるということである。
- これは、推定に貢献する標本サイズの実質的な減少によって引き起こされるものと思われる。標本サイズの実質的な減少は、**ゼロ値計数値従属変数の大量発生**によるものである。