ここで、どの推定結果を採用すべきかについて、以下の検定を通じて判断することができる。

1. 固定効果推定がプーリング推定に対して正当化されるか: F 検定（F Test）
2. ランダム効果推定がプーリング推定に対して正当化されるか: ラグランジュ乗数検定（Lagrange Multiplier (LM) Test）
3. 固定効果推定がランダム効果推定に対して正当化されるか: ハウスマン検定（Hausman Test）

推定法選択のための検定

```
pFtest(grun.w, grun.p)
plmtest(grun.p, effect = "individual", type = "bp")
phtest(grun.w, grun.r)
```

出力結果

```
> pFtest(grun.w, grun.p)

F test for individual effects

data:  inv ~ value + capital
F = 49.1766, df1 = 9, df2 = 188, p-value < 2.2e-16
alternative hypothesis: significant effects

> plmtest(grun.p, effect = "individual", type = "bp")

Lagrange Multiplier Test - (Breusch-Pagan)

data:  inv ~ value + capital
chisq = 798.1615, df = 1, p-value < 2.2e-16
alternative hypothesis: significant effects

> phtest(grun.w, grun.r)

Hausman Test

data:  inv ~ value + capital
chisq = 2.3304, df = 2, p-value = 0.3119
alternative hypothesis: one model is inconsistent
```

最初の F 検定の結果から、固定効果推定はプーリング推定に対して正当化され、二番目のラグラ

ンジュ乗数検定の結果からランダム効果推定はプーリング推定に対して正当化される。3 番目のハウスマン検定から固定効果推定は正当化されないため、ランダム効果推定が選択されることとなる。

パネルデータ分析の実際においては、プールド OLS が、固定効果モデルやランダム効果モデルに比べて、データによりよく当てはまる事例は少ない。そのため、パネルデータの分析において、この 3 つのモデルから最終的なモデルを選択する場合、実際には固定効果モデルかランダム効果モデルかを選択することとなる。

5.7　固定効果モデルとランダム効果モデルの選択について

前節でみたように、固定効果モデルかランダム効果モデルかの選択は、両モデルのデータへの当てはまりを通じて決定することが可能であるが、両モデルの特徴は大きく異なり、どちらのモデルを選択するかによって推定可能な変数も異なる点に注意が必要である。

稲葉（2002）の議論を参照すると、以下のように要約できる。まず、ランダム効果モデルについては、その仮定が満たされる限り、個体内の変動と個体間の変動の両方の情報を用いて推定を行い、不偏かつ有効な推定量を得ることから、統計学的にはこれを用いることが望ましいとされる。しかし、ランダム効果モデルにおいては、モデルでは説明されない個人に固有の誤差は、すべての説明変数に対して独立であるという強い仮定をおいている。このことは、従属変数の個体間の差異を説明するあらゆる要因がモデルに含まれていることを意味しており、社会調査データを用いた分析においては該当しないケースがほとんどである。この仮定が満たされないままランダム効果モデルによる推定を行うと、Between 推定量が不偏ではなくなるため、最終的に推定される GLS 推定量にもバイアスが生じる。したがって、社会調査データを用いた分析においては、実際にはランダム効果モデルの適用が限定される場合が多い。

一方、固定効果モデルでは、個体の説明変数の変化（偏差）によって、その個体の従属変数の変化（偏差）を説明している。じつは、これは時間に対して不変な個体に固有の要因、例えば、人種や性別あるいは幼少時の家庭環境等の影響を除去した上で、説明変数の効果（回帰係数）を推定していることを意味する（Wooldridge 2003）。このとき重要な点は、この「時間に対して不変な個体に固有の要因」には、調査によって観測される変数のみならず、性格や容姿、IQ、特定の物事に対する好みや考え方等のように、必ずしも調査項目として捉えられない要因も含まれるということである。したがって、モデルでは説明されない個体に固有の誤差については、それが時間によって変化しない要因によるものであれば、この影響はすでに除去されたものとして係数の推定を行うことが可能となる。これはランダム効果モデルにおける誤差の仮定と比べて、かなり緩いものであり、データへの当てはまりとしても良好な場合が多い。

モデルでは説明されない個体に固有の誤差のことを非観察要因（unobserved factor）ともいうが、現実の世界においては、しばしばこのような非観察要因は従属変数に重要な影響を与えており、回帰分析においてこのような要因を変数として含むことができない場合には係数の推定値にバイアスがかかることが知られている（＝欠落変数バイアス（omitted variable bias））。固定効果モデルでは、時間によって変化しないと仮定できる限り、このような非観察要因に起因するバイアス

を回避した形で回帰係数を推定することができる。性格や志向の違いなどによって人の行動パターンは千差万別であるため、そうした影響を排除した上で、諸要因のもつ直接的な効果を推定できることは、固定効果モデルの大きな強みである。

一方で、固定効果モデルでは、観察を通じて変化する要因のみしか説明変数に加えることはできない。しばしば、人種や性別などの時間によって変化しない変数と従属変数との関係が、分析者の研究対象となることがある。そのような場合に、固定効果モデルでは答えを得ることができない。また、固定効果モデルでは、観察期間を通じて全く変化が生じなかった個体については、分析から除外されることから、変化が生じることがごく希な事象の分析については、安定的な推定に必要なサンプル数を得ることができない恐れがあるため不向きである。さらに、固定効果モデルでは個体内の偏差（平均からの差）は、平均値の大小にかかわらずその重みは等しいことを仮定している。これは、例えば、テストで 50 点から 10 点上昇することと、90 点から 10 点上昇することを同列に扱うことを意味している。同様に、回帰係数で表される各説明変数の効果は、従属変数を増加させる場合にも減少させる場合にも等しく作用することを仮定している。これらの仮定が分析にあたって妥当であるのかについては、十分に検討される必要がある。

なお、固定効果モデルでは、これらの制約を部分的に緩和する方法として、以下のような方法がある。まず、時間不変の説明変数については、時間によって変化する説明変数との間に交互作用効果を取ることで、部分的にモデルに含めることができる。この場合、例えば、ある時間不変変数の値（例えば、人種）によって説明変数（スキルアップ研修の受講の有無）が従属変数（所得）に与える効果がどの程度異なるのか、といったことを推定できるようになる。ただし、時間不変変数の主効果については、モデルでは推定することはできない。次に、個体内の平均値によって異なる偏差の重みの問題については、従属変数の個体内平均値のばらつきが大きいことが問題であるならば、従属変数の個体内平均値によって標本を分割し、できるだけ同質な集団の中で固定効果モデルを実行するという方法が考えられる。また、説明変数の個体内平均値のばらつきが大きいことが問題であるならば、その説明変数について個体内平均値と個体内偏差との交互作用項をモデルに含め、個体内平均の水準によって、説明変数の効果が異なるモデルを設定するといった方法が検討される。各説明変数の効果について、正の効果と負の効果を別々に推定する方法については有田（2013）を参照されたい。

5.8 固定効果モデルとランダム効果モデルのハイブリッドモデル

二者択一のように思われる固定効果モデルとランダム効果モデルであるが、この両者の特性を併せもつハイブリッドモデルを推定することが可能である（Allison 2009）ハイブリッドモデルにおいては、時間不変の変数と時間によって変化する変数の両方をモデルに含めることができる。しかも、時間不変の変数についてはランダム効果係数を得て、時間によって変化する変数については固定効果係数を得ることができる。このとき得られる固定効果係数は、同じサンプルと説明変数を用いて固定効果モデルを行った場合と全く同じ値と標準誤差を得る 。そのため、ハイブリッドモデルは、時間不変変数を含むことが可能な固定効果モデルと解釈することも可能である。

ハイブリッドモデルの分析の手順は以下である（Allison 2009）。

[1] 時間によって変化する変数 X について、個体内平均との差である個体内偏差をとって説明変数として用いる（→固定効果モデルと同じ）

[2] 従属変数 Y については、偏差を取らず、そのままの値を用いる

[3] 時間不変の変数である Z を説明変数に含める

[4] 時間によって変化する変数 X について、個体内平均値を算出して、該当する各データレコードに貼り付け、これを説明変数に含める

[5] OLS の代わりに、ランダム効果モデルを用いて推定を行う

この手順で分析を行うと、[1] については固定効果係数を、[3] と [4] についてはランダム効果係数を得る。このとき、[1] で得られる係数については、モデルに含まれる時間不変変数の影響を受けない（時間不変変数の specification に対して不変）という特徴がみられる。また、[4] で得られる係数については、すでに [1] の固定効果係数を得ているためそれ自体にはあまり意味はないが、この変数を含めることは、以下の 2 つの点で重要である。まず第 1 に、[3] の変数についてより良い推定値を得るために必要である。[4] の変数がない場合、モデルでは時間によって変化する説明変数 X を統制せずに、Z の推定値を得ることとなってしまう。第 2 に、[1] と [4] の各変数について、同一変数間でその係数を比較することでランダム効果モデルの仮定についての検証を行うことができる。もし、ランダム効果モデルの仮定が正しければ、[1] と [4] から得られる係数は同一変数については同じとなる。したがって、例えば、Wald 検定などの手法を用いれば、時間によって変化する変数 X の一つ一つについて、ランダム効果モデルの仮定が当てはまるのかを検証することができる。このようなモデルの利用についても是非検討すべきであろう。

Allison (2009) においては、ここで紹介したような線形回帰モデルのみならず、ロジスティック回帰モデル、ポワソン回帰モデル、イベントヒストリーモデル、共分散構造モデルなどにおける固定効果モデルならびにハイブリッドモデルについて解説されているため、関心のある読者はそちらを参照されたい。

5.9 出生児縦断調査への応用例

次に示したのは、Fukuda (2014) による出生児縦断調査への固定効果モデルならびにハイブリッドモデルの適用例である。

ここで示す例の従属変数は第 1 回から第 3 回調査において得られている夫の家事参加頻度（調理、食後の片付け、部屋の掃除、洗濯、ゴミ出し、日常の買い物の 6 項目）と育児参加頻度（食事の世話、入浴、おむつの取り替えや排泄の後始末の 3 項目）より算出された因子得点である。各因子得点は、第 1 回から第 3 回までの回答サンプルをプールしたデータを用いて、家事参加項目と育児参加項目のそれぞれに対して主成分分析を適用することで得ている*2 。したがって、同一個人であっても調査回によって因子得点が異なる。Fukuda (2014) の分析では、これらの因子得点に対して、固定効果モデルを適用して夫の家事参加・育児参加に対する個人別の固定効果を推定しているが、ここではさらに時間によって変化しない変数を含めたハイブリッドモデルを適用した結果についても示すこととする。

説明変数は時間によって変化する変数として、夫妻の家事・育児参加についての因子得点、夫妻の調査当年の収入、妻の就業形態、夫の就業形態を使用する。時間によって変化しない変数として、夫の最終学歴、第 1 子の性別、結婚年、妻の第 1 子出生年齢を用いる。また、従属変数の時間による変化を捕捉する変数（時間変数）として調査回を用いる。

以下の表に、分析の結果を示した。分析の結果、時間によって変化する変数については、Allison (2009) にあるように固定効果モデルとハイブリッドモデルで係数ならびに標準誤差（非表示）の値が完全に一致していることから、ハイブリッドモデルにおいても成功裏に固定効果係数が推定されていることが分かる。

ハイブリッドモデルの使用によるメリットのひとつとして、時間によって変化しない変数のランダム効果を推定できることにある。この分析によって、夫の最終学歴が高く、妻の出産年齢が高い場合に夫の家事参加頻度が高いことや結婚年が最近になるほど夫の家事参加頻度が低い傾向があることなどが明らかになった。

*2 分析では第 1 回から第 3 回のすべての回について母親から回答を得ており、モデルで使用するすべての変数に欠損値がないケースに限定した、バランスドパネルを用いている。

夫の家事参加・育児参加についての固定効果モデルとハイブリッドモデルの推定結果

		夫の家事頻度(因子得点)		夫の育事頻度(因子得点)	
		ハイブリッドモデル	固定効果モデル	ハイブリッドモデル	固定効果モデル
		b	b	b	b
時間変化変数	夫妻の家事・育児参加				
	夫の育児参加頻度(因子得点)	0.207 ***	0.207 ***	–	–
	夫の家事参加頻度(因子得点)	–	–	0.298 ***	0.298 ***
	妻の育児参加頻度(因子得点)	−0.027 ***	−0.027 ***	0.015 **	0.015 **
	妻の家事参加頻度(因子得点)	−0.015 ***	−0.015 ***	0.030 ***	0.030 ***
	夫妻の収入				
	夫の収入	0.000	0.000	−0.030 *	−0.030 *
	妻の収入	−0.017 ***	−0.017 ***	−0.008	−0.008
	(妻の収入 / 夫妻の収入の合計)*10	0.052 ***	0.052 ***	0.016	0.016
	妻の就業形態 (基準: 無職・学生)				
	自営業・家族従業者	−0.025	−0.025	−0.066	−0.066
	正規雇用	0.127 ***	0.127 ***	0.007	0.007
	非正規雇用	0.061 **	0.061 **	0.000	0.000
	夫の就業形態(基準:正規雇用、自営業・家族従業者)				
	無職・学生・非正規雇用	0.148 ***	0.148 ***	0.144 ***	0.144 ***
	妻・妊娠	0.197 ***	0.197 ***	0.075 ***	0.075 ***
	妻・出産	−0.071 ***	−0.071 ***	0.059 ***	0.059 ***
	親と同居 (基準: 非同居)	−0.380 ***	−0.380 ***	0.013	0.013
	保育士・保育ママ・ベビーシッターの利用	0.226 ***	0.226 ***	0.126 ***	0.126 ***
時間変数	調査回 (基準: 第1回)				
	第2回	−0.127 ***	−0.127 ***	0.210 ***	0.210 ***
	第3回	−0.184 ***	−0.184 ***	0.265 ***	0.265 ***
時間不変変数	夫の最終学歴(基準:高校)				
	中学・中卒資格の専門学校	−0.080 ***		−0.034	
	高卒資格の専門学校・短大・高専	0.042 **		−0.026	
	大学・大学院	0.063 ***		−0.104 ***	
	第1子の性別:女児(基準:男児)	0.024 *		−0.026 **	
	結婚年(基準:1995年)	−0.017 ***		0.005	
	妻の第1子出生年齢 (基準: 25～29歳)				
	24歳未満	−0.044 **		0.023	
	30～34歳	0.010		−0.050 ***	
	35歳以上	0.067 **		−0.052 *	
時間変化変数の個人内平均値	夫妻の家事・育児参加				
	夫の育児参加頻度(因子得点)	0.546 ***		–	
	夫の家事参加頻度(因子得点)	–		0.500 ***	
	妻の育児参加頻度(因子得点)	−0.153 ***		0.084 ***	
	妻の家事参加頻度(因子得点)	0.053 ***		−0.081 ***	
	夫妻の収入				
	ln(夫の収入)	−0.005		−0.009	
	ln(妻の収入)	−0.004		0.011	
	(妻の収入 / 夫妻の収入の合計)*10	−0.007		0.019	
	妻の就業形態 (基準: 無職・学生)				
	自営業・家族従業者	−0.025		−0.010	
	正規雇用	0.337 ***		−0.054	
	非正規雇用	0.103 **		−0.012	
	夫の就業形態(基準:正規雇用、自営業・家族従業者)				
	無職・学生・非正規雇用	0.048		−0.097 *	
	妻・妊娠	−0.002		0.135 ***	
	妻・出産	−0.074		0.111	
	親と同居 (基準: 非同居)	−0.468 ***		0.092 ***	
	保育士・保育ママ・ベビーシッターの利用	0.130 ***		0.092 **	
	定数	0.192 *	0.073	−0.131	−0.018
	人-年数	33597	33597	33597	33597
	サンプル数	11199	11199	11199	11199
	個人効果と説明変数の相関係数	–	0.238	–	0.188
	個人内 R2	0.112	0.112	0.136	0.136
	個人間 R2	0.356	0.293	0.310	0.288
	全体 R2	0.308	0.247	0.258	0.233
	個人効果によって説明されるyの分散の割合	0.621	0.700	0.478	0.576
	自由度	38	11214	38	11214

*: p<.10, **: p<.05, ***: p<.01

さらに、ハイブリッドモデルにおいては、時間によって変化する変数について、ランダム効果モデルの仮定が当てはまるのかどうかを変数毎に検定することができる。ランダム効果モデルでは、時間によって変化する変数について、個人内平均からの偏差を用いて推定した固定効果係数と個人内平均値を用いて推定した係数とが同一であることを仮定している。そのため、Stataでは以下のコマンドを用いてWald検定を行い、この仮定の検定を行うことができる。

```
/* 各変数の Wald 検定 */
* 変数名_dm は個人内平均からの偏差を取った変数、変数名_m は個人内平均値を表す変数 *

foreach var in h_kaji_fac w_kaji_fac w_ikuji_fac preg ybro_now gpcore c_care w_job2
w_job3 w_job4 f_part ln_h_inc2_f ln_w_inc2_f rw_incw2_f {
test 'var'_dm 'var'_m
}

/* 時間によって変化する変数全体で仮定がみたされているか否かの Wald 検定 */

test (h_kaji_fac_dm = h_kaji_fac_m) (w_ikuji_fac_dm = w_ikuji_fac_m) (w_kaji_fac_dm =
w_kaji_fac_m) (preg_dm = preg_m) (ybro_now_dm = ybro_now_m) (gpcore_dm = gpcore_m)
(c_care_dm = c_care_m) (w_job2_dm = w_job2_m) (w_job3_dm = w_job3_m) (w_job4_dm =
w_job4_m) (f_part_dm = f_part_m) (ln_h_inc2_f_dm = ln_h_inc2_f_m) (ln_w_inc2_f_dm =
ln_w_inc2_f_m) (rw_incw2_f_dm = rw_incw2_f_m)
```

上記のコマンドを実行した結果を表したのが以下の表である。

偏差変数の係数と平均変数の係数が同じであるか否かに関するWald検定の結果

	夫の家事参加モデル		夫の育児参加モデル	
	カイ2乗値	P値	カイ2乗値	P値
夫の育児参加頻度(因子得点)	-	-	5664.84	0.000
夫の家事参加頻度(因子得点)	5664.84	0.000	-	-
妻の育児参加頻度(因子得点)	262.16	0.000	83.27	0.000
妻の家事参加頻度(因子得点)	36.17	0.000	96.41	0.000
夫の収入	0.08	0.961	3.95	0.138
妻の収入	8.76	0.013	2.71	0.258
(妻の収入 / 夫妻の収入の合計)*10	31.29	0.000	3.89	0.143
妻・自営業・家族従業者	0.78	0.679	2.65	0.266
妻・正規雇用	94.29	0.000	2.06	0.357
妻・非正規雇用	10.95	0.004	0.08	0.963
夫・無職・学生・非正規雇用	24.43	0.000	18.66	0.000
妻・妊娠	240.05	0.000	32.52	0.000
妻・出産	26.19	0.000	14.56	0.001
親と同居 (基準: 非同居)	803.01	0.000	19.99	0.000
保育士・保育ママ・ベビーシッターの利用	234.28	0.000	53.65	0.000
全変数	1339.21	0.000	480.18	0.000

この検定においては、偏差変数と平均変数の係数は同一であるという帰無仮説を検定している。モデルの選択上重要であるのは、時間によって変化する変数全体で帰無仮説を棄却できるかどうかであるが (Allison 2009)、ここでは夫の家事参加・育児参加ともに帰無仮説が棄却されているた

め、ランダム効果モデルではなく固定効果モデルが選択される。この検定方法によるモデルの選択は一般的に使われるハウスマン検定の代替であるともいえる。次に、各変数の検定結果をみると、夫の家事参加モデルでは、夫の収入と妻が自営業・家族従業者であるか否かが、育児参加モデルでは、夫妻の収入に関する変数と妻の就業に関する変数において 10 %水準で帰無仮説が棄却できず、ランダム効果係数による解釈を許容する結果となっている。ただし、推定結果にあるように、これらの変数の係数はいずれもほぼ 0 であるため、係数が 0 であるということを追認しているに過ぎない。

このようにハイブリッドモデルでは、時間によって変化しない変数の効果を固定効果係数と同時に推定することで、通常の固定効果モデルよりも多くの情報を得ることができるという点に特徴がある。稲葉（2002）が指摘するように、社会学的な分析においては、時間によって変化しない変数の水準による従属変数の相違が分析の関心となることも多いため、このようなモデルの利用も併せて検討されるべきであろう。また、時間によって変化する変数について、変数毎にランダム効果の妥当性を検討できるという点についても、従来のハウスマン検定にはない利点であると言える。

参考文献

Allison,P. D. (2009) *Fixed Effects Regression Models, Series: Quantitative Applications in the Social Sciences*, Sage Publications Inc.

有田伸（2013）「変化の向き・経路と非変化時の状態を区別したパネルデータ分析: 従事上の地位変化がもたらす所得変化を事例として」『理論と方法』vol. 28: 69-85.

Balatagi, B. H. (2005) *Econometric Analysis of Panel Data*: John Wiley and Sons, Ltd.

Croissant, Y. and G. Millo (2008) “Panel Data Econometrics in R: The plm Package”, *Journal of Statistical Software*, Vol. 27, No. 2.

Fukuda, S. (2014) “Gender Equality and Transition to Second Birth in Japan”, 金子隆一（編）『厚生労働科学研究費補助金「縦断および横断調査によるライフコース事象の経時変化分析と施策への対応に関する研究（平成 24-政策-一般-004)」平成 25 年度総括研究報告書』

樋口美雄，太田清，新保一成 (2006) 『入門パネルデータによる経済分析』, 日本評論社, 東京.

稲葉昭英（2002)「研究動向　Pooled Time Series モデル」,『家族社会学研究』, 第 14 巻, 第 1 号, pp. 5-10.

北村行伸 (2005) 『パネルデータ分析』, 岩波書店, 東京.

Wooldridge, J. M. (2003) *Introductory Econometrics: A Modern Approach*, 2nd Edition, Thomson.

第6章

ダイナミックパネル

6.1 はじめに

経済現象は基本的には経済主体がダイナミックな枠組みの中で、最適化行動を行った結果であるという認識から、最近の経済学は、異時点間の資源分配の最適化を分析の中心にして、投資、消費、雇用、金融政策、財政政策などの議論が組み立てられている。パネルデータを用いる最大のメリットの一つに、同一経済主体の異時点間の変動、すなわち動学的最適化をデータとして捉え、それを実証的に検証できるということがある。個別経済主体の初期値を知りダイナミックな変動過程（運動方程式）を知ることができれば、将来の変動や政策反応を予測できることになる。これがパネルデータを経済学者が利用したがる大きな理由になっている[*1]。

パネルデータの動学的側面については、Balestra and Nerlove (1966) など 1960 年代より意識されてきたことではあるが、1980 年代の時系列分析の発展を受けて、本格的に進展してきた。とりわけ動学的最適化にマッチした形で誕生してきた一般化積率法 (GMM) が Arellano and Bond (1991) によってパネルデータ分析に導入されて以来、急速な発展を遂げている。

本章ではダイナミック・パネルデータ分析の主要な結果をサーベイしているが、限られた紙幅では限定的なものにならざるを得ない。ダイナミック・パネルデータ分析の理論的側面について、さらに知りたい方は Arellano (2003) が包括的な参考文献となっているので参照されたい。

また、本書では全体としてクロスセクション方向の N が大きく、時系列方向の T が短いミクロ・パネルデータを扱っており、N も T も大きいマクロ・パネルデータについては扱ってこなかった。この分野も急速に研究が進んでおり、実証上の応用も増えている。それらの研究をフォローするのは不可能に近いが、Smith and Fuerter (2004) が現在のところ最も包括的なサーベイになっている。

さらに、生存時間解析（サバイバル分析）あるいはデュレーション・モデルとして知られている動学分析は医学、生物学を中心とした自然科学の分野で広く応用されているし、政治学、社会学の分野を中心に社会科学の分野でも最近利用されるようになってきた。これについては、日本語で読

[*1] 本章は北村（2005）『パネルデータ分析』（岩波書店）の第 4 章に基づいている。詳細については北村（2005）第 4 章を参照されたい。実証分析は新たに『21 世紀出生児縦断調査』を用いて行っている。

める教科書も良質のサーベイ論文もあるので、本章では明示的には扱わない*2。ランダム係数モデル (the random coefficient model) もダイナミック・パネルデータ分析に含まれるべき問題であるが、本章では取り扱うことができない。関心のある方は、Hsiao (2003, 第6章)、Maddala, Li, Trost and Joutz (1997)、Hsiao and Pesaran (2004) 等を参照されたい。

6.2 ダイナミック・パネルデータの考え方

一般にパネルデータでダイナミックな関係とは、被説明変数のラグ項が説明変数に入っていることをさす。すなわち、

$$y_{it} = \gamma y_{it-1} + \mathbf{x}'_{it}\beta + \varepsilon_{it} \quad i = 1, 2, ..., N; t = 1, 2, ...T \tag{6.1}$$

ここで、γ はスカラー、x'_{it} は $1 \times K$ 行列、β は $K \times 1$ 行列。ε_{it} は一元配置誤差構成要素モデルに従っているとする。

$$\varepsilon_{it} = \mu_i + u_{it} \tag{6.2}$$

ここで、$\mu_i \sim iidN(0, \sigma^2_\mu)$ は固定効果を表しており, $u_{it} \sim iidN(0, \sigma^2_u)$ は誤差項を表し、相互に独立である。

ダイナミック・パネル推定を巡る大きな問題はラグ被説明変数が誤差項 ε_{it} と相関していること、そしてデータがクロスセクション方向 (N) には大きいが、時系列方向 (T) には小さいということである*3。これは誤差項 u_{it} が系列相関していない場合にも当てはまる。

より一般的には、Maddala (2001) はダイナミック・パネルデータ分析は次の二つのモデルに分類することが重要であることを示した。

[1] 系列相関モデル

$$y_{it} = \mathbf{x}'_{it}\beta + \mu_i + w_{it} \tag{6.3}$$
$$w_{it} = \rho w_{it-1} + u_{it} \quad |\rho| < 1 \tag{6.4}$$

[2] 状態依存モデル

$$y_{it} = \gamma y_{it-1} + \mathbf{x}'_{it}\beta + \mu_i + u_{it} \tag{6.5}$$

*2 日本語で読める教科書として大橋・浜田 (1995)、中村 (2001) を挙げておきたい。英語の最新の教科書として Singer and Willett (2003、第9-11章) がある。また、山口 (2001-2) はイベントヒストリー分析のサーベイとして有益である。サバイバル分析の手法を用いた最近の興味深い実証研究として阿倍・小黒 (2004) を挙げておく。

*3 時系列が短いという問題に対しては逆に時間軸は長くなくてもよいと考えることもできる。むしろ経済主体のダイナミックな調整パラメータは時間と共に変化する可能性が高いので、それが一定とみなされる期間（例えば5年）ぐらいに限定したほうがいいとも言える。調整スピードが速い場合には1年以内に調整が終わり、前年の実績（ラグ変数）はほとんど説明力をもたないケースもある。

[1] はダイナミックな要素が誤差項にあるモデルであり、[2] はダイナミックな要素が被説明変数自体にあるモデルである。パネルデータ分析では後者を扱うことが多いが、前者についても研究されているので、以下では [1] について解説した後、[2] の問題に入って行きたい。

系列相関モデル

系列相関が問題になるのは、それがモデルの特定化において重要な説明変数を落としている可能性を示唆しているからである。しかし、実際のデータではある程度の系列相関が見出されるのはむしろ当たり前であって、その問題をいかに軽減するかというのがここでの問題である。

パネルデータ分析において系列相関の問題を最初に取り上げたのは Lillard and Willis (1978) である。彼らは誤差項が AR(1) に従うケースをライフサイクル所得に当てはめ、マルコフ連鎖に従うと仮定した所得階層移動と対比する形て検討している。Lillard and Weiss (1979) では、アメリカにおける科学者の 1960-70 年における所得のダイナミックな変動と同時点内での所得変動に関して分析している。ここでも所得変動の誤差に系列相関がある場合を検討している。

Baltagi and Li (1991a)、Wansbeek (1992) は (6.4) 式のような系列相関問題に対して、the Paris-Winsten (PW) transformation を用いて、系列相関を取り除いた後で最小二乗法で推定することを提案している。

代替的な推定方法として Maddala(2001)、Nerlove((2002、第 7 章) は一般化最小二乗法を提案している。(6.3) 式を LSDV 推定し、誤差項の推定値 $\hat{w}_{it}$ を得る。さらにそれを用いて (6.4) 式を OLS 推定し $\hat{\rho}$ を得る。これらのパラメータを用いて (6.3) 式を次のように変換する。

$$
\begin{aligned}
y_{it}^{*} &= \mathbf{x}_{it}^{*\prime}\beta + \mu_i^{*} + u_{it} \qquad (6.6)\\
y_{it}^{*} &= y_{it} - \hat{\rho} y_{it-1}\\
\mathbf{x}_{it}^{*} &= \mathbf{x}_{it} - \hat{\rho}\mathbf{x}_{it-1}\\
\mu_i^{*} &= \mu_i(1-\hat{\rho})
\end{aligned}
$$

上式を再び LSDV 推定すれば、β および μ_i の一致推定を得ることができるというものである。

Bhargava, Franizini and Narendranathan (1982) は系列相関検定の定番である Durbin-Watson Statistic をパネルデータ用に一般化した。誤差項の系列相関を次のように定義すると、$w_{it} = \rho w_{it-1} + u_{it}$、帰無仮説は $H_0 : \rho = 0$ で対立仮説が $H_1 : |\rho| < 1$ として、次の検定量を求めて判断することを提案している。

$$
d = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}(\hat{w}_{wit} - \hat{w}_{wit-1})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{w}_{wit}^2} \qquad (6.7)
$$

ここで $\hat{u}_{wit}$ はウイズイン誤差である。誤差項が AR(1) でなく AR(n) になっても、この統計量を対応させることは容易である。

Baltagi and Li (1995) は系列相関と固定効果を同時検定するためのラグランジェ乗数検定を 3 つ提案している。AR(1) の系列相関がゼロでランダム効果を検定するための統計量、AR(1) の系

列相関がゼロであり固定効果であることを検定するための統計量、ランダム効果の下で AR(1) の系列相関がゼロかどうかを鑑定するための統計量、である。これらの統計量は誤差項が AR(1) であれ MA(1) であれ同じになることが示されている*4。

状態依存モデルの初期の研究

ダイナミック・パネル推定の研究は Balestra and Nerlove(1966) に始まる。彼らは [1] パネルデータの特色を生かさないプーリング OLS 推定は被説明変数のラグ項の係数は高くなるバイアスをもつ、[2] 固定効果推定は被説明変数のラグ項の係数を低くするバイアスをもつ、[3] ダイナミック・モデルでは一般化最小二乗法は一致推定でも有効推定でもなくなる、という点をはじめて指摘し、その後のダイナミック・パネルデータ分析の方向性を決定づけた。[1] についてはデータをプールするのかパネルデータとして使うのかという議論に発展した。Maddala (1971a, 1971b) では尤度比検定を用いることを提案している。[2] については Nickell (1981) によって厳密に証明された。[3] については代替的な推定方法として最尤法、操作変数法、一般化積率法 (GMM) など様々な推定方法の開発へと進展していった。この点については次節以後で詳しく解説する。

Maddala (2001) は最尤法推定を初期値に依存した条件付尤度関数と初期値には依存しない条件無し尤度関数に分けることが重要であると指摘している。条件無し尤度関数を仮定するためには、初期条件が定常状態あるいは均衡にあるような長期データであることを仮定していることになる。また条件付尤度関数では初期値が決定的に重要であり、かつ計算が大変であると論じている*5。

Sevestre and Trognon (1996, pp.130-33) は、Maddala (1971a) が系列相関を取り除くために変換する時に用いたウェイト λ にちなんで $\lambda - class$ と名付けられたウイズイン推定、一般化最小二乗法、プーリング最小二乗法、ビトウィーン推定に関して、$\beta = 0$ の場合のラグ項の係数 $\gamma(\lambda)$ のバイアスを推定した。バイアスの大きさの順序は次のようになった。

$$p \lim_{within} \hat{\gamma}(0) < \gamma < p \lim_{GLS} \hat{\gamma}(\lambda) < p \lim_{poolingOLS} \hat{\gamma}(1) < p \lim_{between} \hat{\gamma}(\infty) \tag{6.8}$$

一般化最小二乗法 (GLS) では系列相関の問題を取り除いたはずであるが、ラグ項がモデル変換後の誤差項と相関しているためにバイアスが残っている (Ridder and Wansbeek (1990, pp.566-571))。

*4 詳しくは Baltagi and Li (1995) か Baltagi (2001,pp.90-95) を参照されたい。

*5 ダイナミック・パネル分析において初期値をどこに取るかは実証研究上、極めて重要な問題である。例えば、経済成長を扱うときに、19 世紀には世界経済の中心にあったイギリスが、英国病と呼ばれる低成長に陥っていた 1970 年と、戦後高度成長期を経て奇跡の復興を成し遂げていた 1970 年の日本と、植民地から独立してようやく国造りを始めた 1970 年のアフリカ諸国を並べて考えた時に、それらの国を集めたパネルデータが 1970 年から始まっているからといって 1970 年の経済をそれぞれの国の初期値として扱っていいとは決して言えないだろう。医学や生物学では、このような初期値の問題を考えるために、それぞれの主体の初期値が等しいと考えられる状態でデータを揃え、それからどのような時間経過で変化が起こるかを捉えようするアプローチである生存時間解析 (survival analysis) が用いられている。経済問題にもこのアプローチが使われるようにはなってきたが、医学のようにある病原体に感染した時間から死亡するまでの経過のように、きれいに初期状態が確定できないことが多く、現在のところ応用は限定されているし、実際に使われているケースでも初期値の扱いは恣意的なものが多い。

Trognon(1978) は外生変数が次のような AR(1) 過程に従い、$x_{it} = \delta x_{it-1} + w_{it}$ $w_{it} \sim N(0, \sigma_w^2)$、初期値 y_{i0} を固定と仮定すると、最尤法推定は端点解になり、一致推定にはならないことを示した。また時系列 T が短く、被説明変数のラグ項の係数 γ が負であったり、かなり小さい場合に最尤法推定は一致推定とならない可能性が高いことも示した。この研究は次節で説明する Anderson and Hsiao (1981,1982) に引き継がれた。

6.3 最尤法推定と操作変数法推定

Anderson and Hsiao (1981,1982) はダイナミックなパネルデータの推定問題を検討している。とりわけ、初期値に関して様々な仮定をおいて、推定結果の違いを比較している。彼らは次のようなモデルを考えている。

$$w_{it} = \gamma w_{it-1} + \rho' \mathbf{z}_i + \beta' \mathbf{x}_{it} + \mu_i + u_{it} \quad i = 1, 2, ..N. \ \ t = 1, 2, ...T \tag{6.9}$$

$$y_{it} = w_{it} + \eta_i \tag{6.10}$$

$$\mu_i = (1-\gamma)\eta_i \quad E(\eta_i) = 0 \quad Var(\eta_i) = \sigma_\eta^2 = \sigma_\mu^2/(1-\gamma)^2 \tag{6.11}$$

ここで $\mathbf{z}_i$ は時間とともに変動しない外生変数である。

最尤法と一般化最小二乗法は初期値に依存しているので、初期値に関しては次の 4 つの仮定をおく。

(仮定 1) y_{i0} は固定。この場合は任意の初期値から始まり、$(\mu_i + \rho' \mathbf{z}_i)/(1-\gamma) + \beta' \sum_{j=0} \mathbf{x}_{it-j}\gamma^j$ に収束すると想定されている。また固定効果 μ_i は平均ゼロで分散一定の変数がランダムに選択されたものであると想定されている。データをいつから開始するかということが y_{i0} とは無関係に任意に決定されている場合には、y_{i0} を固定と仮定することには問題がある。すなわち μ_i と y_{i0} が無相関であり、かつ y_{i1}、y_{i1}... に影響を与えるような固定効果 μ_i を想定することは難しいのである。

(仮定 2) y_{i0} はランダムであり、μ_i と u_{it} から独立であり、次のような関係に従っている。$y_{i0} = \bar{y}_0 + \varepsilon_i$、ここで $\bar{y}_0$ は 0 期における全体の平均であり、ε_i は iid に従う。これはさらに 2 つのサブグループに分かれる。

(仮定 2a) y_{i0} と μ_i は独立しており、初期値の違いは時間とともに消滅する。

(仮定 2b)y_{i0} と μ_i は独立しておらず、$cov(y_{i0}, \mu_i) = \varphi\sigma_{y_0}^2$ である。この場合、初期値の違いが将来の y_{it} に影響を与え、長期的には $[\varphi\varepsilon_i/(1-\gamma)] = \lim_{t\to\infty} E[y_{it} - \rho' \mathbf{z}_i/(1-\gamma) - \beta' \sum_{j=0}^{t-1} \mathbf{x}_{it-j}\gamma^j | \varepsilon_i]$ に収束する。

(仮定 3) w_{i0} は固定であり、$y_{it} = w_{it} + \eta_i$ かつ $\mu_i = (1-\gamma)\eta_i$ なので、y_{it} と μ_i は相関している。y_{i0} は任意の値から始まっても $\eta_i + \rho' \mathbf{z}_i/(1-\gamma) + \beta' \sum_{j=0}^{t-1} \mathbf{x}_{it-j}\gamma^j$ に収束する。データの開始時期と確率過程の開始時期は一致している必然性はない。

(仮定 4) w_{i0} はランダムであり、(3) と同様の関係に従っている。ここでは w_{it} の性質によって、4 つのサブグループに分かれる。

(仮定 4a)w_{i0} はランダムで、平均は全体共通 θ_w で、分散は $\sigma_u^2/(1-\gamma^2)$ であり、均一である。

(仮定 4b)w_{i0} はランダムで、平均は全体共通 θ_w で、分散は任意の $\sigma^2_{w_0}$ となる。
(仮定 4c)w_{i0} はランダムで、平均は θ_{i0} で、分散は $\sigma^2_u/(1-\gamma^2)$ である。
(仮定 4d)w_{i0} はランダムで、平均は θ_{i0} で、分散は任意の σ^2_{w0} となる。

Anderson and Hsiao (1981,1982) はこのような仮定の下で、合計 8 通りの推定を行っている。初期値の仮定に応じて尤度関数が違ってくるが、ここでは全てについて厳密な展開をすることはしないで、一般型で表現しておく[*6]。

$$L(\gamma,\rho,\beta,\gamma,\eta,\sigma^2_u,\sigma^2_w,\sigma^2_\mu) = (2\pi)^{-NT/2}|v|^{-N/2} \times \exp\{-\frac{1}{2\sigma^2}\sum_i\sum_t(y_{it}-\gamma y_{it-1}-\rho'\mathbf{z}_i-\beta'\mathbf{x}_{it})^2\} \tag{6.12}$$

ここで v は誤差分散共分散行列を表わしており、誤差構成要素によって表現が違ってくる。

一般にはこの式を未知パラメータに関して最大化すれば、最尤法推定が得られる。最尤法推定の一致性は初期値と標本数 N と時系列数 T に依存している。ただし、この最適化は $\sigma^2_\mu=0$ の場合には端点解となる。この点は Anderson and Hsiao (1981) で検討されている。

最尤法推定では初期値の仮定によって尤度関数の形状が変わり、また推定結果も変動することがわかった。ここでは代替的に初期値に依存しない推定法である操作変数法について論じたい。

固定効果推定であれランダム効果推定であれ、上の (6.9)(6.10) 式から一階の階差をとれば時間とともに変動しない z_i と μ_i は消去されてしまう。すなわち、

$$y_{it}-y_{it-1} = (x_{it}-x_{it-1})'\beta+\gamma(y_{it-1}-y_{it-2})+(u_{it}-u_{it-1}) \tag{6.13}$$

このモデルはラグ被説明変数の階差が誤差項 u_{it} の階差と相関しているという意味では問題が残っているが[*7]、操作変数法を用いて推定することで内生性バイアスを取り除くことができる。すなわち、有効ではないが一致推定を得ることができる。

操作変数に応じて β と γ は次のように推計される[*8]。

$(y_{it-2}-y_{it-3})$ を操作変数として使った場合は次のようになる。

$$\begin{pmatrix}\gamma_{iv}\\ \beta_{iv}\end{pmatrix} = \left[\sum_{i=1}^N\sum_{t=3}^T\begin{pmatrix}(y_{i,t-1}-y_{i,t-2})(y_{it-2}-y_{it-3}) & (y_{it-2}-y_{it-3})(\mathbf{x}_{it}-\mathbf{x}_{it-1})') \\ (\mathbf{x}_{it}-\mathbf{x}_{it-1})(y_{it-2}-y_{it-3}) & (\mathbf{x}_{it}-\mathbf{x}_{it-1})(\mathbf{x}_{it}-\mathbf{x}_{it-1})'\end{pmatrix}\right]^{-1} \times\left[\sum_{i=1}^N\sum_{t=3}^T\begin{pmatrix}y_{it-2}-y_{it-3}\\ \mathbf{x}_{it}-\mathbf{x}_{it-1}\end{pmatrix}(y_{it}-y_{it-1})\right] \tag{6.14}$$

y_{it-2} を操作変数として使った場合には次のように書ける。

[*6] 詳細については Anderson and Hsiao (1982) および Hisao (2003、第 4 章) を参照されたい。
[*7] 具体的には y_{it-1} と u_{it-1} は (1) 式より明らかに相関している。
[*8] 以下の式の展開は Hsiao (2003, pp.85-86) を参照。

$$\begin{pmatrix} \gamma_{iv} \\ \tilde{\beta}_{iv} \end{pmatrix} = \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \begin{pmatrix} y_{it-2}(y_{i,t-1} - y_{i,t-2}) & y_{it-2}(\mathbf{x}_{it} - \mathbf{x}_{it-1})') \\ (\mathbf{x}_{it} - \mathbf{x}_{it-1})y_{it-2} & (\mathbf{x}_{it} - \mathbf{x}_{it-1})(\mathbf{x}_{it} - \mathbf{x}_{it-1})' \end{pmatrix} \right]^{-1} \times \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \begin{pmatrix} y_{it-2} \\ \mathbf{x}_{it} - \mathbf{x}_{it-1} \end{pmatrix} (y_{it} - y_{it-1}) \right] \tag{6.15}$$

実証上、ラグ変数の水準であれば最小期間は 2 期間ですむが、階差であれば最低 3 期間は必要になる。また 3 期間以上のデータであれば、最適操作変数としては $(y_{i,t-1} - y_{i,t-2})$ と $(y_{it-2} - y_{it-3})$ あるいは y_{it-2} の相関がどれぐらいあるかで判断すれば良い[*9]。

第 2 段階として推計した $\hat{\beta}$ と $\hat{\gamma}$ を (6.9) 式に代入し，ρ を最小二乗法で推定する。

$$\bar{y}_{it} - \hat{\gamma}\bar{y}_{it-1} - \hat{\beta}'\bar{\mathbf{x}}_{it} = \rho'\mathbf{z}_i + \mu_i + \bar{u}_{it} \quad i = 1, ..N \tag{6.16}$$

ここで、$\bar{y}_i = \sum_{t=1}^{T} y_{it}/T$, $\bar{\mathbf{x}}_i = \sum_{t=1}^{T} \mathbf{x}_{it}/T$, $\bar{u}_i = \sum_{t=1}^{T} u_{it}/T$。

第 3 段階として分散共分散行列 σ_u^2、σ_μ^2 を推定する。

$$\sigma_u^2 = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} [(y_{it} - y_{it-1}) - \hat{\gamma}(y_{it-1} - y_{it-2}) - \hat{\beta}'(\mathbf{x}_{it} - \mathbf{x}_{it-1})]^2}{2N(T-1)} \tag{6.17}$$

$$\sigma_\mu^2 = \frac{\sum_{i=1}^{N} (\bar{y}_i - \hat{\gamma}\bar{y}_{i,-1} - \hat{\rho}'\mathbf{z}_i - \hat{\beta}'\bar{\mathbf{x}}_i)}{N} - \frac{1}{T}\hat{\sigma}_u^2 \tag{6.18}$$

これらの推定が一致性を持つことは、初期値には依存していない。N か T あるいは両方が無限に近づくと、操作変数法パラメータ γ、β、σ_u^2 は一致推定となる。ρ、σ_μ^2 が一致推定になるのは N が無限になる場合のみであり、N が固定されている場合には、T が無限になろうと一致推定にはならない。

Anderson and Hsiao (1982) は N 固定で T が無限、T 固定で N が無限の場合に分けた上で、上述の 8 通りについて最尤法推定法、固定効果推定法、操作変数法推定を行った。推計結果をまとめると次のようになる。最尤法でラグ項のパラメータ γ が一致推定にならないのは（仮定 3）で T が固定、N が無限の場合のみである。時間とともに変動しない変数の係数 ρ は N 固定、T 無限の場合には常に一致推定にはならない。また（仮定 3）で T が固定、N が無限の場合にも一致推定にはならない。操作変数法では時間とともに変動しない変数の係数 ρ については最尤法と同じで N 固定、T 無限の場合には一致推定にはならない。また、ラグ項のパラメータ γ は全ての場合で一推定になる。固定効果推定法では ρ は推定できないが、ラグ項のパラメータ γ が一致推定にならないのは N 固定、T 無限の全ての場合と (仮定 3) で T が固定、N が無限の場合である。ここでの結果は

[*9] Arellano (1989) は操作変数としてはラグ変数の水準 y_{it-2} や y_{it-3} を用いた方がラグの階差 $(y_{it-2} - y_{it-3})$ を用いるより望ましいとしている。

ダイナミック・モデル推定は誤差項の条件や初期値の仮定というより、TやNの仮定により強く依存していることを物語っている。

Hsiao, Pesaran and Tahmiscioglu (2002) や Fujiki, Hsiao and Shen (2002) では代替的に Chamberlain (1982,1984) を嚆矢とする推定方法である最小距離推定法 (Minimum Distance Estimation: MDE) を用いることを主張している*10。基本的な考え方は、誤差項の階差2次式を最小化するようにパラメータ (β,γ) を決定するということである。すなわち、

$$\min[\sum_{i=1}^{N}\triangle\mathbf{u}_i^{*\prime}\Omega^{-1}\triangle\mathbf{u}_i^{*}] \tag{6.19}$$

ここで Ω は $\triangle\mathbf{u}_i^*$ の共分散行列、$\triangle\mathbf{u}_i^*=[\triangle y_{i1}-\beta\triangle x_{i1}-\gamma\triangle y_{i0},\triangle y_{i2}-\beta\triangle x_{i2}-\gamma\triangle y_{i1},...]$

この方法は誤差分布が正規分布に従うことを要求していないので、有効推定ではないが、N が大きければ漸近的に一致推定となる。しかも計算ははるかに簡単になる。

6.4　一般化積率法推定

操作変数法で推計するアプローチに対して、Arellano and Bond (1991)、Ahn and Schmidt (1995) らは操作変数法は重要な情報を用いていないので、有効でないと論じている。例えば、一階の階差モデルを想定すると、2期ラグをとった y の水準は誤差項の階差とは無相関であることを示すことができる*11。

$$E[y_{is},(u_{it}-u_{i,t-1})]=0,\ \ s=0,1,...t-2,\ \ t=2,....T \tag{6.20}$$

Arellano and Bond (1991) はこのような情報も利用した一般化積率法 (GMM) を用いることを提唱している。

$$\frac{1}{n}\sum_{i=1}^{n}y_{is}[(y_{it}-y_{i,t-1})-(y_{i,t-1}-y_{i,t-2})'\gamma-(x_{it}-x_{i,t-1})'\beta]=0 \tag{6.21}$$

$$s=0,....,t-2,\ \ t=2,....,T$$

被説明変数のラグ項の階差を次々に取っていくと、それに対応して使える操作変数は $(y_{i1},y_{i2},y_{i3},...y_{iT-2})$ と増えていく。これを行列で表現すると次のようになる*12。

$$W_i=\begin{bmatrix}[y_{i1}] & 0 & ... & 0\\ 0 & [y_{i1},y_{i2}] & ... & 0\\ 0 & 0 & ... & 0\\ 0 & 0 & ... & [y_{i1},...y_{iT-2}]\end{bmatrix} \tag{6.22}$$

*10 MDEの詳細についてはChamberlain (1982,1984)、Lee(2002, 第3章) を参照。
*11 すなわち直交条件が成立する。これはHoltz-Eakin(1988)、Holtz-Eakin, Newey and Rosen(1988) によって指摘された。
*12 以下はBaltagi (2001, p.132) を参照。

操作変数が誤差ラグ項と無相関である条件 (6.20) は次のようにベクトル形式で書き直せる。

$$E(W_i' \bigtriangleup u_i) = 0 \tag{6.23}$$

ここで (6.1) 式のようなダイナミック・モデルを考えてみよう。

$$\begin{aligned}(y_{it} - y_{it-1}) &= (y_{it-1} - y_{it-2})'\gamma + (x_{it} - x_{it-1})'\beta + (u_{it} - u_{it-1}) \\ \bigtriangleup y_{it} &= \bigtriangleup y_{it-1}'\gamma + \bigtriangleup x_{it}'\beta + \bigtriangleup u_{it} \qquad i = 1, 2...N\end{aligned} \tag{6.24}$$

この式に対して操作変数行列 W_i を掛けてパラメータを推定するのが Arellano and Bond (1991) の一般化積率法 (GMM) である。(6.24) 式を次のように書き換えて、未知のパラメータについて積率法を当てはめるのである。

$$W' \bigtriangleup y_{it} = W' \bigtriangleup y_{it-1}'\gamma + W' \bigtriangleup x_{it}'\beta + W' \bigtriangleup u_{it} \tag{6.25}$$

γ、β について解くと次のように表せる。

$$\hat{\gamma}_{GMM} = [(\bigtriangleup y_{it-1})'W\hat{V}_N^{-1}W'(\bigtriangleup y_{it-1})]^{-1}[(\bigtriangleup y_{it-1})'W\hat{V}_N^{-1}W'(\bigtriangleup y_{it})] \tag{6.26}$$

$$\hat{\beta}_{GMM} = [(\bigtriangleup x_{it-1})'W\hat{V}_N^{-1}W'(\bigtriangleup x_{it-1})]^{-1}[(\bigtriangleup x_{it-1})'W\hat{V}_N^{-1}W'(\bigtriangleup y_{it})] \tag{6.27}$$

ここで $V_N = \sum_{i=1}^{N} W_i'(\bigtriangleup u_i)(\bigtriangleup u_i)'W_i$ である[*13]。

説明変数 x_{it} が厳密に外生変数であれば、$E(x_{it}u_{is}) = 0, \forall t, s = 1, 2, ..., T$ であり、x_{it} が μ_i と相関している場合には、(6.24) 式のようなダイナミック・モデルに関して全ての x_{it} が操作変数になり得る。もし x_{it} が外生変数ではなく先決変数 (predetermined) であり、$E(x_{it}u_{is}) \neq 0\ for\ s < t$、$E(x_{it}u_{is}) = 0\ for\ s \geq t$ の場合は、操作変数になり得るのは $(x_{i1}, x_{i2}, ..., x_{is-1})$ までである。先決変数と外生変数が混在しているような場合でも、操作変数行列 W_i を適切に書き換えればよい。

Arellano and Bond (1991) の GMM 推定で重要な仮定は誤差項に系列相関がないということである。この点が確保されていなければ、操作変数として使えないものが出てくるし、その結果として GMM 推定は一致推定でなくなりバイアスをもつことになる。そこで次のような系列相関検定が Arellano and Bond (1991) で提案されている。

j 次の誤差ラグ項の自己相関係数の平均を次のように定義する[*14]。

$$r_j = \frac{1}{T-3-j}\sum_{t=4+j}^{T} r_{tj} \tag{6.28}$$

[*13] 実際、Arellano and Bond (1991, p.279) では (6.25) 式を均一分散を仮定した一般化最小二乗法で推計した one-step 推定と初期値や誤差項の分布に制約を課さずに GMM で推計した two-step 推定を用いている。誤差項が iid に従う場合には漸近的に等しくなる。

[*14] Arellano (2003, pp.121-23) を参照。

ここで $r_{tj} = E(\triangle\ u_{it}\ \triangle\ u_{it-j})$ である。帰無仮説 $H_0 : r_j = 0$ として検定量は次のように定義する。

$$m_j = \frac{\hat{r}_j}{SE(\hat{r}_j)} \tag{6.29}$$

ここで $\hat{r}_j$ は標本から得られた $\triangle\ \hat{u}_{it}$ と $\hat{r}_{tj} = N^{-1}\sum_{i=1}^{N} \triangle\ \hat{u}_{it}\ \triangle\ \hat{u}_{it-j}$ に基づいて計算されたものを使う。

Arellano and Bond (1991) では操作変数に関する Sargan (1958) の過剰識別制約テストも導入している。

$$s = \triangle\ \hat{u}'W[\sum_{i=1}^{N} W_i'(\triangle\ \hat{u}_i)(\triangle\ \hat{u}_i)'W_i]^{-1}W'(\triangle\ \hat{u}) \sim \chi^2_{p-k-1} \tag{6.30}$$

ここで p は行列 W における行数、$\triangle\ \hat{u}$ は (6.25) 式における推定誤差を表わしている。Arellano and Bond (1991) では誤差ラグ構造が1階と2階の場合を考慮している。

Ahn and Schmidt(1995) は y の水準からだけではなく、y と誤差項の階差 $(u_{it} - u_{it-1})$ との間からも重要な情報 (ここでは直交条件) が得られることを示している。これは次のように表せる。

$$E(y_{is}\ \triangle\ u_{it}) = 0 \quad t = 2, ...T,\ s = 0, 1, ...t-2 \tag{6.31}$$

さらに、次のような直交条件も利用できることを示した。

$$E(u_{iT}\ \triangle\ u_{it}) = 0 \quad t = 2, ...T-1 \tag{6.32}$$

上の2種類の直交条件を合わせると $T(T-1)/2 + (T-2)$ の制約が加わることになる。(6.32) 式は γ が1に近いか、σ^2_μ/σ^2_u が大きい場合には、有効な情報（パラメータの分散が小さい）を提供してくれることがモンテカルロ実験からわかった。

Ahn and Schmidt(1995) によれば、(6.31)(6.32) 式は通常の誤差項が満たすべき仮定より緩やかな仮定である。すなわち次のように書くことができる。

(仮定1)全ての i と t に対して、$cov(u_{it}, y_{i0})$ は等しい。ちなみに、従来の仮定では $cov(u_{it}, y_{i0}) = 0$ である。

(仮定2)全ての i と t に対して、$cov(u_{it}, \mu_i)$ は等しい。ちなみに、従来の仮定では $cov(u_{it}, \mu_i) = 0$ である。

(仮定3) 全ての i と $t \neq s$ に対して、$cov(u_{it}, u_{is})$ は等しい。ちなみに、従来の仮定では $cov(u_{it}, u_{is}) = 0$ である。

この（仮定1-3）に基づく GMM 推定は先に論じた Chamberlain (1982,1984) の最小距離推定 (Minimum Distance Estomator) と漸近的には等しくなり、漸近的に有効推定となることが示されている。

Blundell and Bond (1998) はこれまで GMM 推定の問題点と指摘されてきた操作変数の弱相関問題と、これは GMM 推定固有の問題ではないが Anderson and Hsiao (1981,1982) 以来、ダイナミック・パネル・モデルの本源的な課題とされてきた初期値問題とを取り上げて、それを解決する目的で従来の GMM をシステム GMM に拡張した。

まず、操作変数の弱相関問題であるが、Arellano and Bond の GMM 推定は γ が 1 に近づくか、固定効果 μ_i の分散が大きいときにはバイアスが大きくなることが知られている。Blundell and Bond (1998) はこの問題を明らかにするために T=3 と仮定し、直交条件を $E(y_{i1} \bigtriangleup u_{i3}) = 0$ に限定した。これによってパラメータ γ は適度識別される。この場合、GMM 推定は次のような操作変数法推定になる。

$$\bigtriangleup y_{i2} = \pi y_{i1} + \mu_i + u_{i2} \quad i = 1, 2, ...N \tag{6.33}$$

ここで γ が 1 に近づくか、固定効果 μ_i の分散が大きければ、パラメータ π は 0 に近づいても不思議はない。この場合、y_{i1} は $\bigtriangleup y_{i2}$ とは弱相関になる。このことは次の式からも明らかである。すなわち $E(y_{i1}\mu_i) > 0$ で、$\sigma_\mu^2 = var(\mu_i)$、$\sigma_u^2 = var(u_{it})$ の時、パラメータ π の確率極限は次のように表せる。

$$p \lim \hat{\pi} = (\gamma - 1)\frac{k}{(\sigma_\mu^2/\sigma_u^2) + k} \qquad k = \frac{(1-\gamma)}{(1+\gamma)} \tag{6.34}$$

Blundell and Bond (1998) は GMM 推定のバイアスはかなりの程度、Nelson amd Startz (1990a,b) や Staiger and Stock (1997) によって論じられた操作変数の弱相関に起因していることを示したのである。

初期値の問題に関しては次のように考えている。すなわち、Ahn and Schumidt (1995) らが主張した直交条件に加えて、T-3 本の直交条件を導入する。

$$E(u_{it} \bigtriangleup y_{it-1}) = 0 \quad t = 4, 5, ...T \tag{6.35}$$

ところで、$\bigtriangleup y_{i2}$ は観察可能であるから、次の直交条件も導入可能である。

$$E(u_{i3} \bigtriangleup y_{i2}) = 0 \tag{6.36}$$

この条件は初期値 y_{i1} がどのように発生したかに依存している。すなわち初期値を次のように定義する[*15]。

[*15] これは y_{i0} がそれ以前の y_{i0-t} と全て等しくなる（定常状態）と仮定した場合に得られる値であるが、すでに何度も論じてきたように、このような初期値を経済変数で実際に求めることは極めて難しい。また、初期値が個人の個体属性である固定効果と関連していると考えるのはモデルの構造上自然であるが、現実の固定効果というのは初期値の帰結として出てきたものも多く（例えば親の経済状態を反映した学歴）、初期値と固定効果の関係についてはさらに深い議論が必要である。

$$y_{i1} = \frac{\mu_i}{1-\gamma} + u_{i1} \tag{6.37}$$

$t = 2$ 以後の関係式は与えられているので、y_{it} の全ての流列が確定する。ところで、(6.36) 式は次のように書き換えることができる。

$$E[(\mu_i + u_{i3})(u_{i2} + (\gamma - 1)u_{i1})] = 0 \tag{6.38}$$

であり、これはさらに次のような十分条件に書き換えることができる。

$$E(u_{i1}\mu_i) = E(u_{i1}u_{i3}) = 0 \quad i = 1, 2, ...N \tag{6.39}$$

この条件が意味しているのは、初期値 y_{i0} からの乖離 u_{i1} が初期値のレベル $\mu_i/(1-\gamma)$ とは無相関であることである。

Blundell and Bond (1998) では γ が 1 に近いか、σ_μ^2/σ_u^2 が大きい場合でも、(6.35)(6.36) 式を用いることで GMM 推定のバイアスを劇的に改善できることを示した。そのために GMM 推定において操作変数行列 $\mathbf{Z}_i^+$ を次のように定義したものを使ったシステム GMM を提案した。

$$\mathbf{Z}_i^+ = \begin{bmatrix} \mathbf{Z}_i & 0 & 0 & ... & 0 \\ 0 & \triangle\, y_{i2} & 0 & ... & 0 \\ 0 & 0 & \triangle\, y_{i3} & ... & 0 \\ . & . & . & ... & 0 \\ 0 & 0 & 0 & ... & \triangle\, y_{iT-1} \end{bmatrix} \tag{6.40}$$

ここで $\mathbf{Z}_i$ は (T-2)×m 行列で従来の GMM 推定で用いられた操作変数行列である。

$$\mathbf{Z}_i = \begin{bmatrix} y_{i1} & 0 & 0 & ... & 0 & ... & 0 \\ 0 & y_{i1} & y_{i2} & ... & 0 & ... & 0 \\ . & . & . & ... & . & ... & . \\ 0 & 0 & 0 & ... & y_{i1} & ... & y_{iT-2} \end{bmatrix} \tag{6.41}$$

このように従来の GMM 推定を拡張することによって、$\sigma_\mu^2/\sigma_u^2 = 1, T = 4$ の場合の従来の GMM 推定とシステム GMM 推定の分散比率をとると、$\gamma = 0$ では 1.75、$\gamma = 0.5$ では 3.26、$\gamma = 0.9$ では 55.40 と γ の値が上昇するに従って、バイアスの差が拡大していくことが明らかになった。逆に言えば、システム GMM を用いることで、従来 GMM 推定が破綻すると言われてきた、γ が 1 に近いか、σ_μ^2/σ_u^2 が大きい場合でも、かなりバイアスを抑えることができることを示したのである。