

# 労働力調査とローテーション・サンプリング

加納 悟<sup>†</sup>

(受付 2003年4月7日; 改訂 2003年10月7日)

## 要 旨

本稿では、わが国の労働力調査を例に取り、ローテーション・サンプリングによって得られた調査結果を分析するための時系列モデルを提示する。それは、失業の発生メカニズムがプロビットモデルで、その潜在変数が線型状態空間モデルとして記述される、非線型時系列モデルである。それゆえ近似に依らない限り通常のカルマンフィルターのアルゴリズムを用いた最尤法で推定することはできない。本稿では、モデルの推定はギブス・サンプリング(Gibbs Sampling)を用いて行なわれる。残念ながら、わが国の労働力調査結果の集計表にはローテーション構造に関する情報が欠落しており、モデルの有効性のチェックは現実のデータではなく擬似データによらざるを得ない。最後に、いくつかの重要と思われる応用とモデルの拡張について述べ結論とする。

キーワード：労働力調査、ローテーション・サンプリング、プロビットモデル、ギブス・サンプリング、状態空間表現。

## 1. はじめに

一般に官庁統計の調査結果は集計表のかたちで公表される。そこに見られる情報は地域や年齢などの様々な特性の組み合わせに対応する調査結果の集計値である。通常統計調査のサンプリングには調査時点ごとに層別の無作為抽出が用いられるが、なかには一度抽出された特定の個体が繰り返し調査対象となる場合もある。わが国では労働力調査や家計調査において繰り返し調査が採用されており、諸外国においてもほとんどの労働力調査において繰り返し調査が採用されている。このような標本の抽出方法はローテーション・サンプリングと呼ばれており、調査コストの観点からは明らかに望ましい調査方法である。反面、各時点の調査結果には相関が存在し、このような特殊な標本調査方法を考慮に入れた利用が必要になる。本稿では、わが国の労働力調査を例に取り、ローテーション・サンプリングによって得られた調査結果から失業率の全国平均やその変化などの情報を得る方法について検討を加える。それによって、調査結果の公表の際、調査回数情報が追加されることにより、研究者にとってより多様な分析が可能となることを指摘する。

オーストラリアなど諸外国においては、労働力調査をはじめとし官庁統計の調査結果の統計分析が盛んである。これに対し、わが国の統計調査結果では、プライバシー保護の観点から利用には目的外申請を義務付けており、その利用が簡単ではない。その結果、統計調査方法や調査結果の利用に関する研究は諸外国と比べかなり遅れている。

本論文の目的の1つは、わが国の労働力調査結果を分析するための時系列モデルを提示する

---

<sup>†</sup>一橋大学 経済研究所：〒186-8601 東京都国立市中 2-1; kanoh@ier.hit-u.ac.jp

ことにある。本論文の構成は以下のとおりである。第 2 節において、わが国の労働力調査におけるサンプリングの特徴について簡単にまとめる。続く第 3 節では、先行研究の紹介としてオーストラリアの労働力調査の時系列モデルを紹介する。第 4 節では、わが国の労働力調査において採用されている独特のサンプリング構造を考慮に入れた時系列モデルを示す。それは、失業のメカニズムをプロビットモデルで記述し、失業のマイクロデータをも分析しうる非線型時系列モデルである。例えばある地域における高齢者の失業率の推定など、とくに少数個のサンプルデータの分析の際、非線型性の問題は重要となる。このような場合、このモデルでは近似に依らない限り通常のカルマンフィルターのアルゴリズムを用いた最尤法で推定することはできない。それゆえ、モデルの推定はギブス・サンプリング(Gibbs Sampling)を用いて行なわれる。第 5 節では擬似データの作成とそれを用いたモデルの適合を行なう。残念ながら、わが国の労働力調査結果の集計表にはローテーション構造に関する情報が欠落しており、現実のデータを用いた分析は不可能である。それゆえ、本稿では擬似データによってローテーション構造を持つサンプリングの長所・短所について一般的に論じる。第 6 節はモデルの拡張や応用例について述べ結語とする。

## 2. 労働力調査とローテーション構造

わが国の、労働力調査は総務省により毎月行なわれており、その目的は、「わが国の 15 歳以上人口について、月々の就業状態・就業時間・産業・職業などの就業状況、失業・休職の状況などの実態とその変化を把握することにより、景気判断や雇用対策等の基礎資料を得ること」(労働力調査年報より)である。そこでは全国 11 地域から国勢調査の調査区を第 1 次抽出単位、抽出された調査区内の住戸を第 2 次抽出単位とした層化 2 段抽出法により抽出し、抽出された住戸内の 15 歳以上の世帯員を対象とする。1 調査区内の世帯数は約 50 である。その結果、得られる標本の大きさは、平成 11 年段階で調査区約 2,900、調査世帯約 40,000 世帯、15 歳以上の構成員計約 100,000 人となっている。わが国の 15 歳以上の人口は約 1 億人であることから、労働力調査における個人の抽出率は約 0.1%(0.001)とみなされる。また労働力調査の標本設計に当たっては(1)前月差や前年同月差などの時系列変化の精度を高める(2)10 地域別の結果を公表する(標本抽出は北海道、東北、南関東、北関東・甲信、北陸、東海、近畿、中国、四国、九州、沖縄の 11 地域別に行なわれるが、公表時には沖縄を九州に含め 10 地域とする)(3)調査員の負担から 1 調査区の抽出世帯は約 15 とする、などが考慮されている(総務庁(1999)を参照)。

わが国の労働力調査の標本は以下のような特徴的な抽出方法により得られている。まず標本調査区は 4 ヶ月継続して調査が行なわれる。その調査区は翌年同月に再び調査される。また標本調査区全体のうち 1/4 は毎月他の調査区に入れ替えられる。一方、抽出された住戸は 2 ヶ月継続して調査される。2 ヶ月後には同一標本調査区内の他の住戸に入れ替えられるが、翌年同月には再び調査される。すなわち、各住戸は計 4 回調査される。結局、以上の組み合わせにより 8 つのグループが形成され、それらが一部ずつ入れ替えられながら調査される。各時点においては調査のタイミングについて、1 年目 1 回目、1 年目 2 回目、2 年目 1 回目、2 年目 2 回目と異なる 4 タイプが、そしてそれらがある調査区の前半に行なわれるか後半に行なわれるかによって、計 8 通りの住戸が存在することになる。図 1 はこのような調査形式をまとめたものである。

このように同一の住戸を繰り返し抽出することの利点はいくつか考えられる。第 1 に、コストが小さくて済むことである。一般に、最初の調査のコストは高く、それゆえ同じ住戸を続けて調査すれば調査費用は大きく軽減される。たとえば途上国における調査では標本を得ることが難しく、一度得た標本はできうる限り続けて利用したいことが多い。第 2 に、時点間の変化

グループ	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	5月	6月
1	1	2	1	2	1	2	1						3	4	3	4		
2	3	4	3	4	3	4	3						1	2	1	2		
3	4		1	2	1	2		1	2						3	4	3	4
4	2	3	4	3	4	3	4							1	2	1	2	
5	3	4		1	2	1	2											
6	1	2	3	4	3	4	3											
7	4	3	4		1	2	1	2										
8	2	1	2	3	4	3	4											

図 1. わが国の労働力調査のローテーション構造．図中の数字は通算の調査回数を表している．

をより小さな誤差で推定しうることが予想される．異なる個体の時間的变化は同一の個体の時間的变化に個体としての標本誤差を加えたものになってしまう．とはいえ，母集団の性質は絶えず変化してゆくであろうから，それを反映するには新たな標本を採る必要がある．また，最初に選ばれた住戸が転出したり，回答しなくなるなどその質も悪化してゆくであろう．

わが国の労働力調査では特定の住戸が続けて2ヶ月選ばれ，その後10ヶ月をおいて再び2ヶ月選択される．McLaren and Steel(1997)ではこのようなローテーションのパターンは2-10-2(4)と表記されている(4)は住戸が計4回選択されることを表している．労働力調査で用いられているローテーションのパターンはこの他にもいろいろあり，各国ごとに異なるローテーションのパターンが採用されている．米国のCPS(Current Population Survey)ではわが国の労働力調査と似たローテーションの構造を採用している．各住戸は4ヶ月間続けて選択された後，8ヶ月間標本からはずされ，その後再び4ヶ月間標本に含まれる．よって4-8-4(8)と表記される．この他にも6-6-6(12)や1-1-1(6)，1-2-1(m)などがあり，英国の四半期の労働力調査は1-2-1(5)と考えられる．いずれも1年離れた時点で同一住戸を調査することにより失業率の前年同月差はより少ない誤差で推定可能となるであろう．これに対しオーストラリアでは各住戸が8ヶ月続けて調査され，翌年の繰り返しはない．カナダでも同様に，6ヶ月の継続調査のみが行なわれている．

ローテーション・サンプリングにより得られた標本データの利用には注意が必要である．まず，バイアスの存在である．ローテーション・サンプリングにおいては一般に第1回目の調査結果が他と比べて異なる値をとることが多い．例えば，Bailar(1975)は，様々な調査におけるローテーション・バイアスの存在を紹介し，同時に米国のCurrent Population Survey(CPS)においてこのようなバイアスが推定に対して与える影響について分析している．バイアスが生じる原因としては調査票の誤解や調査項目に対して正しく応えることに対する抵抗感が考えられる．我が国の労働力調査の調査票においては，一時期あいまいな表現の質問が含まれていた．このような時期にはバイアスが生じていたとしても不思議ではない．

第2に，観測される標本間に相関が存在することである．そのため母集団の平均やトレンドなどのパラメータの推定には複数の時点の情報を用いることが望ましい．もし標本観測値が推定すべきパラメータの1次式で表現されるのであれば，標本間の相関を推定することにより，より効率的な推定量を得ることが可能であろう．すなわち一般化最小二乗法(GLS)により最小分散不偏推定量を得ることが原理的に可能である．ただし，現実には標本間の相関が時間的に一定であるとは考えにくい．さらにこのような方法では時点間の相関が生じるメカニズムに関してはふれず，単に安定的な相関があると仮定することになる．

以下ではデータの生成されるメカニズムを時系列モデルを用いて定式化する．とくに4節ではわが国の標本観測値の生成メカニズムを状態空間表現により記述することにより，ローテーション・サンプリングによって得られた標本から失業率を推定する方法を提案したい．

### 3. オーストラリアの労働力調査の時系列モデル

オーストラリアの労働力調査では以下のような標本設計がなされている。センサスにおける調査区から標本調査区を層別抽出により選び、8つのローテーション・グループ(RG)に分ける。各調査区から一定のルールで住戸が系統的に選ばれる。各月ごとにRG中の1つのグループの住戸が同一グループの他の住戸に置き換えられる。残りの住戸は再び調査される。このパターンが概ね5年程度継続される。

このようにオーストラリアの労働力調査では各個体は8ヶ月継続して調査され、その後再び調査されることはない。上述の表記法によれば、8-0-0(8)と表すことができる。このようなローテーション構造においては、当然同一の個体(住戸)が調査される8ヶ月間の時系列データ(集計値)には高い相関が見られる。またRG内のデータはすべて同一の調査区に属するため、やはりある程度の相関が予想される。

Bell and Carolan(1998)はこのような相関を考慮するため、オーストラリアの労働力調査結果を時系列モデルを用いて分析し、失業率や労働力率のトレンドを推定している。以下では、時系列モデルを用いた分析の先行研究の紹介として彼らの時系列モデルの要点をまとめる。

図2に与えられる、8つのグループを考える。観測値は各グループにおける集計値 $y_t^j$ ( $j = 1, 2, \dots, 8$ )であり、 $y_t^j$ は各RGに共通のトレンド要因 $\beta_t$ と季節要因 $S_t$ 、そしてグループ固有の要因 $e_t^j$ によって構成されるとする。まずトレンド要因 $\beta_t$ はその変化 $\Delta\beta_t = (\beta_t - \beta_{t-1})$ が比較的スムーズに変化するものとし、

$$\Delta\beta_t = \Delta\beta_{t-1} + \epsilon_{\beta t}$$

あるいは

$$\beta_t = 2\beta_{t-1} - \beta_{t-2} + \epsilon_{\beta t}$$

と表されるとする。また季節要因 $S_t$ は

$$S_t = -(S_{t-1} + S_{t-2} + \dots + S_{t-11}) + \epsilon_{S t}$$

と表現されるものとする。ここで $\epsilon_{\beta t}$ 、 $\epsilon_{S t}$ はそれぞれ $N(0, \sigma_\beta^2)$ 、 $N(0, \sigma_S^2)$ に従う確率変数である。また $e_t^j$ は時点 $t$ において第 $j$ 回目の調査を受けるようなグループの集計値 $y_t^j$ の標本誤差を表すものである。より具体的に、 $e_t^j$ は3つの要因に分解される。それらは、調査区要因 $B_t^j$ と個体要因 $P_t^j$ とそしてランダムな要因 $U_t^j$ である。調査区要因とは標本が同一調査区にあることの影響であり、個体要因とは同一個体が抽出されることの影響を表す。これらはローテーション・サンプリングを行なうがゆえに考慮されなければならない要因である。 $e_t^j$ は3つの要

グループ	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	1月	2月	3月	4月	5月	6月
1	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2
2	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1
3	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
4	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7
5	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6
6	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5
7	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4
8	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3

図2. オーストラリアの労働力調査のローテーション構造。図中の数字は通算の調査回数を表している。

因の合計として

$$e_t^j = B_t^j + P_t^j + U_t^j$$

と表される．さらにこれら 3 つの要因のそれぞれは以下のようなメカニズムで変動するものとする．まず， $B_t^j$  は

$$B_t^j = \begin{cases} \theta_B B_{t-1}^{j-1} + \{(1 - \theta_B^2)(1 - k_2^2)(1 - k_1^2)\}^{1/2} \epsilon_{Bt}^j & j = 2, 3, \dots, 8 \\ \theta_B B_{t-1}^8 + \{(1 - \theta_B^2)(1 - k_2^2)(1 - k_1^2)\}^{1/2} \epsilon_{Bt}^j & j = 1 \end{cases}$$

と 1 次の自己回帰モデルで表現され，月次データの相関が安定的に減衰するものと想定する．ここで  $\epsilon_{Bt}^j$  は  $N(0, \sigma_B^2)$  に従う確率変数である．オーストラリアの労働力調査では同一の調査区から 5 年間抽出が行なわれるため，各グループ内において調査区の変更の影響を考えなくともよい．よって，調査区の影響を単純に相関で記述することが可能となり， $B_t^j$  は同一調査区内の個体からなる集計値が持つ相関を表す要因と解釈される．また， $P_t^j$  は，

$$P_t^j = \begin{cases} \theta_P P_{t-1}^{j-1} + \{(1 - \theta_P^2)(1 - k_2^2)\}^{1/2} k_1 \epsilon_{Pt}^j & j = 2, 3, \dots, 8 \\ (1 - k_2^2)^{1/2} k_1 \epsilon_{Pt}^j & j = 1 \end{cases}$$

と表される．ここで  $\epsilon_{Pt}^j$  は  $N(0, \sigma_P^2)$  に従う確率変数である．同一個体が選ばれる 8 ヶ月間には相関が存在するが，隣り合う 8 ヶ月間同士には相関がない．また， $U_t^j$  については，

$$U_t^j = k_2 \epsilon_{Ut}^j \quad j = 1, 2, \dots, 8$$

とする．ここで  $\epsilon_{Ut}^j$  は  $N(0, \sigma_U^2)$  に従う確率変数である．また， $\theta_B, \theta_P, k_1, k_2$  は未知のパラメータで，Bell and Carolan(1998)ではそれらの値は集計値データの自己相関からあらかじめ推定される．

以上の想定のもとで，結局観測値  $y_t^j$  は以上の要因の和で表される．すなわち，

$$y_t^j = \beta_t + S_t + e_t^j = \beta_t + S_t + B_t^j + P_t^j + U_t^j$$

である．実際の推定には  $\beta_t, S_t, B_t^j, P_t^j$  を未知の状態と考えた線型状態空間表現により，カルマン・フィルターのアルゴリズムに従って行なわれる．その結果，オーストラリアの労働力調査において  $\theta_B, \theta_P$  の値は 1 に近く， $e_t^j$  の分散は約 60% が  $P_t^j$  によって，約 10% が  $B_t^j$  によって占められると同論文は報告している．残り 30% はランダム要因である．

以上，Bell and Carolan(1998)では時系列モデルを用いて失業率の値そのものを複数の要因に分解することを考えている．以下では，個人の失業の決定要因をモデル化することを考える．

#### 4. 労働力調査結果の時系列分析

本節ではわが国の労働力調査に特有のローテーション構造 2-10-2(4) に対するモデル化を試みる．Bell and Carolan(1998)のオーストラリアの労働力調査 8-0-0(8)では標本誤差の各成分が AR モデルにより記述された．その場合，相関は安定的にかつずばやく減衰することを想定している．しかし，現実のデータに見られる相関はなかなか減衰しない．むしろランダム・ウォークとして捉えたほうがよいと思われる．また，Bell and Carolan(1998)の時系列モデルは集計値を分析の対象とするものであるが，ここでは個々の標本の特性を記述することを考える．本来，“失業”や“労働力人口”等の特性は質的変量であり，パイナリデータとして記述されるものである．そこで失業のメカニズムをプロビットモデルを用いて定式化する．一般に，ある特性を満たす個体に限って分析を行なうとすれば，そのデータ数は限られたものになる．

例えば、農村に住む 60 歳以上の男性の失業や転職を分析しようという場合である。このように標本数が大きくない場合には、質的データ特有の分析が必要になる。このような考え方は時系列モデルを用い小地域におけるトレンド推定を行なった Pfeffermann et al. (1998) と同様である。

具体的に、労働力調査の標本が調査対象に該当する特性(例えば失業、就業、転職など)を保有しているか否かは、以下のメカニズムによって決まると考える。まず、図 1 のグループ  $j$  ( $j = 1, 2, \dots, 8$ ) に属する個人  $i$  が調査特性を保有しているのであれば  $y_{it}^j = 1$ 、そうでなければ、 $y_{it}^j = 0$  とする。またその個人  $i$  に対し、潜在変数  $y_{it}^{j*}$  が存在し、 $y_{it}^{j*} \geq 0$  のとき、 $y_{it}^j = 1$  となるものとする。さらに、 $y_{it}^{j*}$  は、個人  $i$  の個体的特徴  $\alpha_{it}^j$  とトレンド要因  $\beta_t$ 、そしてグループ要因  $g_t^j$  によって構成され、

$$y_{it}^{j*} = \beta_t + \alpha_{it}^j + g_t^j$$

と表されるものとする。ここでグループ要因  $g_t^j$  とは 8 個のそれぞれのグループに特有な要因であり、各グループに属する個体が特定の調査区から続けて抽出されていることの影響や、同一個体の調査回数に応じて生ずるローテーション・バイアスを表すものである。我が国の労働力調査においては、同一調査区から 4 回続けて同じ住戸を選ぶという工夫がなされている。その場合、同一調査区に属する個体には何らかの共通性が見られるであろうから、調査区効果が 4 ヶ月続けて存在することを考慮する。よってグループ要因  $g_t^j$  はバイアスと調査区効果の 2 つの和と考えればよい。その結果、同じ調査区の前半の調査結果が後半の調査結果かということと、各個体(住戸)の調査回数との組み合わせの違いで 8 つのグループに分類され、それらが図 1 の各グループに対応する。

以下では、季節性要因は考慮されていない。季節性要因はオーストラリアの労働力調査と同様に定式化することは可能である。しかし現実には、まず失業率の推定値を求めた後に、曜日効果やうらう年効果など様々な要素を考慮に入れ、X12-ARIMA 等の移動平均法によって季節調整値が計算されるようである。Bell and Carolan (1998) はローテーション構造を考慮した場合と X-11 により従来のように季節調整を行なった場合のトレンド推定値の比較を行なっている。わが国の労働力調査の季節調整に関しては別途議論すべき重要な課題である。以下では議論の焦点をローテーション構造に絞るため、観測値として季節調整済みの系列を用いることとし、あえて季節性要因を考慮しないことにする。

このとき、 $y_{it}^j = 1$  となる個体が抽出される確率は

$$\Pr(y_{it}^j = 1) = \Pr(y_{it}^{j*} \geq 0) = \Pr(\beta_t + \alpha_{it}^j + g_t^j \geq 0)$$

と表現される。個票レベルにおいては  $y_{it}^j$  は観察可能であるが、一般に官庁統計における個票は公開されておらず、結果は集計値として公表されるのが通常である。ここでは、各グループにおいて  $\frac{\sum y_{it}^j}{N_j} = \bar{Y}_t^j$  とし、 $\bar{Y}_t^j$  が観察可能であるとしよう。ただし個票レベルでデータが得られる場合にも、ここでのモデルはそのまま適用可能である。また  $N_j$  は各グループ  $j$  における標本数を表し、簡単化のため時間に関しては変化しないものとする。

一方、 $y_{it}^{j*}$  の構成要因である  $\beta_t$ 、 $\alpha_{it}^j$ 、 $g_t^j$  は、それぞれ以下のメカニズムに従って時系列的に変動するものと仮定する。まずトレンド要因  $\beta_t$  はオーストラリアの労働力調査におけると同様、

$$\beta_t = 2\beta_{t-1} - \beta_{t-2} + \epsilon_{\beta t}$$

とする。ここで  $\epsilon_{\beta t}$  は  $N(0, \sigma_{\beta}^2)$  に従う確率変数である。またグループ要因  $g_t^j$  の構成要素のう

ち調査回数に応じて生ずるローテーション・バイアス  $b_{rt}$  ( $r = 1, \dots, 4$ ) はランダム・ウォーク

$$b_{rt} = b_{rt-1} + \epsilon_{bt} ; \quad \epsilon_{bt} \sim N(0, \sigma_b^2)$$

に従うものとする。このとき、ローテーション・バイアスの各グループに対する影響  $b_t^j$  は

$$b_t^j = \begin{cases} G_{1t}b_{1t} + G_{2t}b_{2t} + G_{3t}b_{3t} + G_{4t}b_{4t} & j = 1, 5 \\ G_{3t}b_{1t} + G_{4t}b_{2t} + G_{1t}b_{3t} + G_{2t}b_{4t} & j = 2, 6 \\ G_{1t-1}b_{1t} + G_{2t-1}b_{2t} + G_{3t-1}b_{3t} + G_{4t-1}b_{4t} & j = 3, 7 \\ G_{3t-1}b_{1t} + G_{4t-1}b_{2t} + G_{1t-1}b_{3t} + G_{2t-1}b_{4t} & j = 4, 8 \end{cases}$$

と表される。ここで  $G_{kt}$  ( $k = 1, \dots, 4$ ) は以下のように定義されるダミー変数ベクトル  $G_k$  の第  $t$  成分である。

$$\begin{aligned} G_1 &= [\overbrace{101010101010}^{12} | \overbrace{000000000000}^{12} | \overbrace{101010101010}^{12} | \dots]' \\ G_2 &= [\overbrace{010101010101}^{12} | \overbrace{000000000000}^{12} | \overbrace{010101010101}^{12} | \dots]' \\ G_3 &= [\overbrace{000000000000}^{12} | \overbrace{101010101010}^{12} | \overbrace{000000000000}^{12} | \dots]' \\ G_4 &= [\overbrace{000000000000}^{12} | \overbrace{010101010101}^{12} | \overbrace{000000000000}^{12} | \dots]' \end{aligned}$$

ここで、 $\sum_{k=1}^4 G_{kt} = 1$  (for all  $t$ ) であり、また、 $G_{10} = G_{20} = G_{30} = 0$ 、 $G_{40} = 1$  とする。さらに、

$$\begin{aligned} G_5 &= [\overbrace{100010001000}^{12} | \overbrace{000000000000}^{12} | \overbrace{100010001000}^{12} | \dots]' \\ G_6 &= [\overbrace{010001000100}^{12} | \overbrace{000000000000}^{12} | \overbrace{010001000100}^{12} | \dots]' \\ G_7 &= [\overbrace{001000100010}^{12} | \overbrace{000000000000}^{12} | \overbrace{001000100010}^{12} | \dots]' \\ G_8 &= [\overbrace{000100010001}^{12} | \overbrace{000000000000}^{12} | \overbrace{000100010001}^{12} | \dots]' \\ G_9 &= [\overbrace{000000000000}^{12} | \overbrace{100010001000}^{12} | \overbrace{000000000000}^{12} | \dots]' \\ G_{10} &= [\overbrace{000000000000}^{12} | \overbrace{010001000100}^{12} | \overbrace{000000000000}^{12} | \dots]' \\ G_{11} &= [\overbrace{000000000000}^{12} | \overbrace{001000100010}^{12} | \overbrace{000000000000}^{12} | \dots]' \\ G_{12} &= [\overbrace{000000000000}^{12} | \overbrace{000100010001}^{12} | \overbrace{000000000000}^{12} | \dots]' \end{aligned}$$

とする。 $G_k$  ( $k = 5, \dots, 12$ ) は新たな調査区の開始時点を示すためのダミー変数からなるベクトルである。このとき、各グループにおける調査区効果  $d_t^s$  は

$$d_t^1 = G_{5t}\epsilon_{dt} + G_{6t}d_{t-1}^1 + G_{7t}d_{t-2}^1 + G_{8t}d_{t-3}^1 + G_{9t}d_{t-12}^1 + G_{10t}d_{t-13}^1 + G_{11t}d_{t-14}^1 + G_{12t}d_{t-15}^1$$

$$d_t^2 = G_{5t-1}\epsilon_{dt} + G_{6t-1}d_{t-1}^2 + G_{7t-1}d_{t-2}^2 + G_{8t-1}d_{t-3}^2 + G_{9t-1}d_{t-12}^2 + G_{10t-1}d_{t-13}^2 + G_{11t-1}d_{t-14}^2 + G_{12t-1}d_{t-15}^2$$

$$d_t^3 = G_{5t-2}\epsilon_{dt} + G_{6t-2}d_{t-1}^3 + G_{7t-2}d_{t-2}^3 + G_{8t-2}d_{t-3}^3 + G_{9t-2}d_{t-12}^3 + G_{10t-2}d_{t-13}^3 + G_{11t-2}d_{t-14}^3 + G_{12t-2}d_{t-15}^3$$

$$d_t^4 = G_{5t-3}\epsilon_{dt} + G_{6t-3}d_{t-1}^4 + G_{7t-3}d_{t-2}^4 + G_{8t-3}d_{t-3}^4 + G_{9t-3}d_{t-12}^4 + G_{10t-3}d_{t-13}^4 + G_{11t-3}d_{t-14}^4 + G_{12t-3}d_{t-15}^4$$

と 4 種類に分類される．ここで  $\sum_{k=5}^{12} G_{kt} = 1$  (for all  $t$ ) が成り立つ．さらに  $G_{9,-3} = G_{10,-2} = G_{11,-1} = G_{12,0} = 1$  で，その他は  $G_{kt} = 0$  ( $k = 5, \dots, 8$ ;  $t = -3, -2, -1, 0$ ) とする．ここで  $\epsilon_{dt}$  は時点  $t$  において開始される新たな調査区の効果を表し， $\epsilon_{dt} \sim N(0, \sigma_d^2)$  とする．

これに対し，各グループに属する個体の変動  $\alpha_{it}^a$  は

$$\alpha_{it}^1 = G_{1t}\epsilon_{it} + G_{2t}\alpha_{it-1}^1 + G_{3t}\alpha_{it-12}^1 + G_{4t}\alpha_{it-13}^1$$

$$\alpha_{it}^2 = G_{3t}\epsilon_{it} + G_{4t}\alpha_{it-1}^2 + G_{1t}\alpha_{it-12}^2 + G_{2t}\alpha_{it-13}^2$$

$$\alpha_{it}^3 = G_{1t-1}\epsilon_{it} + G_{2t-1}\alpha_{it-1}^3 + G_{3t-1}\alpha_{it-12}^3 + G_{4t-1}\alpha_{it-13}^3$$

$$\alpha_{it}^4 = G_{3t-1}\epsilon_{it} + G_{4t-1}\alpha_{it-1}^4 + G_{1t-1}\alpha_{it-12}^4 + G_{2t-1}\alpha_{it-13}^4$$

とやはり 4 種類に分類される．ここで， $\epsilon_{it}$  は時点  $t$  におけるサンプリングによる個体抽出の変動効果を表す．各グループに対する影響は  $d_{tr}^a$  と  $\alpha_{it}^a$  の組み合わせにより 8 通りに分類される．

以上をまとめ，ローテーション・サンプリングによる観測値およびその背後にある様々な要因の変動をまとめて記述する．そのため，以下のように記号を定義する．8 つのグループからの観測値，潜在変数，そして誤差項のベクトルをそれぞれ

$$\bar{Y}_t = [\bar{Y}_t^1, \bar{Y}_t^2, \bar{Y}_t^3, \bar{Y}_t^4, \bar{Y}_t^5, \bar{Y}_t^6, \bar{Y}_t^7, \bar{Y}_t^8]'$$

$$y_{it}^* = [y_{it}^{1*}, y_{it}^{2*}, y_{it}^{3*}, y_{it}^{4*}, y_{it}^{5*}, y_{it}^{6*}, y_{it}^{7*}, y_{it}^{8*}]'$$

$$v_t = [v_{1t}, v_{2t}, v_{3t}, v_{4t}, v_{5t}, v_{6t}, v_{7t}, v_{8t}]'; \quad v_t \sim N(0, \sigma^2 I_8)$$

とする．ここで， $I_n$  は  $n$  次元の単位行列である．またローテーション・バイアス，調査区効果，個体効果をそれぞれ

$$b_t = [b_{1t}, b_{2t}, b_{3t}, b_{4t}]'$$

$$d_t = [d_{1t}, d_{1t-1}, \dots, d_{1t-14} | d_{2t}, d_{2t-1}, \dots, d_{2t-14} | \dots | d_{4t}, d_{4t-1}, \dots, d_{4t-14}]'$$

$$\alpha_t = [\alpha_{1t}, \alpha_{1t-1}, \dots, \alpha_{1t-12} | \alpha_{2t}, \alpha_{2t-1}, \dots, \alpha_{2t-12} | \dots | \alpha_{4t}, \alpha_{4t-1}, \dots, \alpha_{4t-12}]'$$

とし，変動要因をまとめて

$$\theta_t = [\beta_t, \beta_{t-1}, b_t, d_t, \alpha_t]'$$

と表記する．また，



$$A_{jt} = \begin{matrix} & \overbrace{\hspace{10em}}^{13} \\ \left[ \begin{array}{cccccc} & g_{jt} & & & & \\ 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \vdots \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right] \end{matrix}$$

とする．ここで  $A_{jt} (j = 1, \dots, 4)$  は各グループに対する個体変動の影響を表す  $13 \times 13$  行列であり，その第 1 行はそれぞれ

$$\begin{aligned} g_{1t} &= \overbrace{[G_{2t}, 0, \dots, 0, G_{3t}, G_{4t}]}^{13}' & g_{2t} &= \overbrace{[G_{4t}, 0, \dots, 0, G_{1t}, G_{2t}]}^{13}' \\ g_{3t} &= \overbrace{[G_{2t-1}, 0, \dots, 0, G_{3t-1}, G_{4t-1}]}^{13}' & g_{4t} &= \overbrace{[G_{4t-1}, 0, \dots, 0, G_{1t-1}, G_{2t-1}]}^{13}' \end{aligned}$$

となっている．同様に，

$$D_{jt} = \begin{matrix} & \overbrace{\hspace{10em}}^{15} \\ \left[ \begin{array}{cccccc} & h_{jt} & & & & \\ 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \vdots \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right] \end{matrix}$$

とおく．ここで  $D_{jt} (j = 1, \dots, 4)$  は調査区効果を表す  $15 \times 15$  行列であり，その第 1 行はそれぞれ

$$\begin{aligned} h_{1t} &= \overbrace{[G_{6t}, G_{7t}, G_{8t}, 0, \dots, 0, G_{9t}, G_{10t}, G_{11t}, G_{12t}]}^{15}' \\ h_{2t} &= \overbrace{[G_{6t-1}, G_{7t-1}, G_{8t-1}, 0, \dots, 0, G_{9t-1}, G_{10t-1}, G_{11t-1}, G_{12t-1}]}^{15}' \\ h_{3t} &= \overbrace{[G_{6t-2}, G_{7t-2}, G_{8t-2}, 0, \dots, 0, G_{9t-2}, G_{10t-2}, G_{11t-2}, G_{12t-2}]}^{15}' \\ h_{4t} &= \overbrace{[G_{6t-3}, G_{7t-3}, G_{8t-3}, 0, \dots, 0, G_{9t-3}, G_{10t-3}, G_{11t-3}, G_{12t-3}]}^{15}' \end{aligned}$$

となっている．さらに表記の簡便化のため，以下のようなベクトルを定義する．

$$\begin{aligned} G_1 &= \overbrace{[G_{1t}, 0, \dots, 0]}^{13}' & G_3 &= \overbrace{[G_{3t}, 0, \dots, 0]}^{13}' \\ G_1^- &= \overbrace{[G_{1t-1}, 0, \dots, 0]}^{13}' & G_3^- &= \overbrace{[G_{3t-1}, 0, \dots, 0]}^{13}' \\ G_5 &= \overbrace{[G_{5t}, 0, \dots, 0]}^{15}' & G_6 &= \overbrace{[G_{6t}, 0, \dots, 0]}^{15}' \end{aligned}$$

$$G_7 = \overbrace{[G_{7t}, 0, \dots, 0]}'^{15} \quad G_8 = \overbrace{[G_{8t}, 0, \dots, 0]}'^{15}$$

$$\iota_1 = \overbrace{[1, 0, \dots, 0]}^{15} \quad \iota_2 = \overbrace{[1, 0, \dots, 0]}^{12}$$

以上の準備のもとで、ローテーション・サンプリングの時系列モデルは以下のようにまとめられる。

$$(4.1) \quad \Pr(y_{it}^j = 1) = \Pr(y_{it}^{j*} \geq 0), \quad \bar{Y}_t^j = \sum_i y_{it}^j / N_j$$

$$(4.2) \quad \mathbf{y}_{it}^* = \begin{bmatrix} 1 & 0 & G_{1t} & G_{2t} & G_{3t} & G_{4t} & \iota_1 & & \iota_2 & \mathbf{0} \\ 1 & 0 & G_{3t} & G_{4t} & G_{1t} & G_{2t} & \iota_1 & \mathbf{0} & \iota_2 & \\ 1 & 0 & G_{1t-1} & G_{2t-1} & G_{3t-1} & G_{4t-1} & \iota_1 & & \iota_2 & \\ 1 & 0 & G_{3t-1} & G_{4t-1} & G_{1t-1} & G_{2t-1} & \iota_1 & & \mathbf{0} & \iota_2 \\ 1 & 0 & G_{1t} & G_{2t} & G_{3t} & G_{4t} & & \iota_1 & \iota_2 & \mathbf{0} \\ 1 & 0 & G_{3t} & G_{4t} & G_{1t} & G_{2t} & & \iota_1 & \iota_2 & \\ 1 & 0 & G_{1t-1} & G_{2t-1} & G_{3t-1} & G_{4t-1} & \mathbf{0} & \iota_1 & \iota_2 & \\ 1 & 0 & G_{3t-1} & G_{4t-1} & G_{1t-1} & G_{2t-1} & & \iota_1 & \mathbf{0} & \iota_2 \end{bmatrix} \theta_t + \mathbf{v}_t$$

$$= H\theta_t + \mathbf{v}_t; \quad \mathbf{v}_t \sim N(0, R)$$

ただし、 $\mathbf{y}_{it}^* = [y_{it}^{1*}, y_{it}^{2*}, y_{it}^{3*}, y_{it}^{4*}, y_{it}^{5*}, y_{it}^{6*}, y_{it}^{7*}, y_{it}^{8*}]'$ 、 $R = \sigma^2 I_8$  で  $H$  は  $8 \times 118$  の行列、 $\theta_t$  は 118 次元のベクトルである。ここで潜在変数  $\mathbf{y}_{it}^*$  は状態変数の線型方程式 (4.2) として表現されるが、(4.2) より

$$(4.2)' \quad \bar{Y}_t^* = H\theta_t + \bar{\mathbf{v}}_t, \quad \bar{\mathbf{v}}_t \sim N(0, \bar{R})$$

が成り立つ。ここで、 $\bar{Y}_t^* = [\bar{Y}_t^{1*}, \bar{Y}_t^{2*}, \bar{Y}_t^{3*}, \bar{Y}_t^{4*}, \bar{Y}_t^{5*}, \bar{Y}_t^{6*}, \bar{Y}_t^{7*}, \bar{Y}_t^{8*}]'$ 、 $\bar{Y}_t^{j*} = \sum_i y_{it}^{j*} / N_j$

$$\bar{R} = \sigma^2 \begin{bmatrix} 1/N_1 & & & & & & & & & \\ & 1/N_2 & & & & & & & & \\ & & \ddots & & & & & & & \\ & & & & & & & & & \\ & & & & & & & & & \\ & & & & & & 1/N_7 & & & \\ & & & & & & & & & 1/N_8 \end{bmatrix}$$

である。そこで (4.1) より  $\bar{Y}_t^*$  を発生し、(4.2)' を観測方程式とみなす。また状態方程式は

$$(4.3) \quad \theta_t = \begin{bmatrix} 2 & -1 & & & & & & & & \\ 1 & 0 & & & & & & & & \mathbf{0} \\ & & I_4 & & & & & & & \\ & & & D_{1t} & & & & & & \\ & & & & D_{2t} & & & & & \\ & & & & & D_{3t} & & & & \\ & & & & & & D_{4t} & & & \\ & & & & & & & A_{1t} & & \\ & & & & & & & & A_{2t} & \\ & & & & & & & & & A_{3t} \\ & & & & & & & & & & A_{4t} \\ \mathbf{0} & & & & & & & & & & & \end{bmatrix} \theta_{t-1} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & G_5 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & G_6 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & G_7 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & G_8 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & G_1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & G_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & G_1^- \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & G_3^- \end{bmatrix} \begin{bmatrix} \epsilon_{\beta t} \\ \epsilon_{bt} \\ \epsilon_{dt} \\ \epsilon_{it} \end{bmatrix}$$

$$= F_t \theta_{t-1} + J_t \epsilon_t ; \quad \epsilon_{\beta t} \sim N(0, \sigma_\beta^2), \quad \epsilon_{it} \sim N(0, \sigma_\alpha^2)$$

となる．ここで， $F_t$  は  $118 \times 118$  の行列， $J_t$  は  $118 \times 4$  の行列である．

以上のシステムの推定は，観測値を得るメカニズム(4.1)が非線型であるため，通常のカルマン・フィルターのアルゴリズムを用いた最尤法で推定することはできない．Kim and Nelson (1999)は状態空間モデルにおいて Gibbs Sampling を用いてベイズ法により状態を推定するアルゴリズムを提案している．以下の方法はこれを拡張したものである．特に以下にあるように(4.1)を利用して発生した  $\tilde{Y}_t^*$  を(4.2)の  $y_{it}^*$  に置換えパラメータ推定を行なう．

具体的には，推定すべきパラメーター・ベクトルを

$$\begin{aligned} \Theta &= [\beta_1, \beta_2, \dots, \beta_T, b_{11}, \dots, b_{1T}, \dots, b_{41}, \dots, b_{4T}, d_{11}, \dots, d_{1T}, \dots, d_{41}, \dots, d_{4T}, \\ &\quad \alpha_{11}, \dots, \alpha_{1T}, \dots, \alpha_{41}, \dots, \alpha_{4T}, \sigma_\beta^2, \sigma_b^2, \sigma_d^2, \sigma_\alpha^2, \sigma^2] \\ &= [\tilde{\beta}_T, \tilde{\gamma}_T, \tilde{\sigma}^2] = [\tilde{\theta}_T, \tilde{\sigma}^2] \end{aligned}$$

と設定し，以下のステップ(a)(b)を繰り返す．

- (a)  $g_1(\tilde{\theta}_T | \tilde{\sigma}^{2(k)}, \tilde{Y}^T)$  から， $\tilde{\theta}_T^{(k+1)}$  を生成する．
- (b)  $g_2(\tilde{\sigma}^2 | \tilde{\theta}_T^{(k+1)}, \tilde{Y}^T)$  から， $\tilde{\sigma}^{2(k+1)}$  を生成する．

ここで， $\tilde{Y}^T$  は  $T$  時点までのすべての観測値を表し，

$$\begin{aligned} \tilde{\beta}_T &= [\beta_1, \beta_2, \dots, \beta_T] \\ \tilde{\gamma}_T &= [b_{11}, \dots, b_{1T}, \dots, b_{41}, \dots, b_{4T}, d_{11}, \dots, d_{1T}, \dots, d_{41}, \dots, d_{4T}, \\ &\quad \alpha_{11}, \dots, \alpha_{1T}, \dots, \alpha_{41}, \dots, \alpha_{4T}] \\ \tilde{\theta}_T &= [\tilde{\beta}_T, \tilde{\gamma}_T] \\ \tilde{\sigma}^2 &= [\sigma_\beta^2, \sigma_b^2, \sigma_d^2, \sigma_\alpha^2, \sigma^2] \end{aligned}$$

である． $k$  は  $k = 1, 2, \dots, M_1, M_1 + 1, \dots, M_1 + M_2$  とし， $M_1 + M_2$  回の繰り返しの結果のうち，最初の  $M_1$  回の結果は捨て，残りの  $M_2$  回のサンプリングの結果から  $\Theta$  の平均値を計算しそれを推定値とする．すなわち，

$$E[g(\Theta | \tilde{Y}^T)] = \frac{1}{M_2} \sum_{k=M_1+1}^{M_1+M_2} g(\Theta^{(k)} | \tilde{Y}^T)$$

を推定値とする．ステップ(a)(b)は具体的に以下の手順から構成される．

ステップ(a)

$$\theta_{t|t} = E(\theta_t | \tilde{Y}^t), \quad \Sigma_{t|t} = V(\theta_t | \tilde{Y}^t), \quad \theta_{t+1|t} = E(\theta_{t+1} | \tilde{Y}^t), \quad \Sigma_{t+1|t} = V(\theta_{t+1} | \tilde{Y}^t)$$

とする． $\tilde{\theta}_T$  は，

$$(4.4) \quad P(\tilde{\theta}_T | \tilde{Y}^T) = P(\theta_T | \tilde{Y}^T) \prod_{t=1}^{T-1} P(\theta_t | \theta_{t+1}, \tilde{Y}^t)$$

によって生成する． $(\theta_T | \tilde{Y}^T)$ ， $(\theta_t | \theta_{t+1}, \tilde{Y}^t)$  は，

$$(4.5) \quad (\theta_T | \tilde{Y}^T) \sim N(\theta_{T|T}, \Sigma_{T|T})$$

$$(4.6) \quad (\theta_t | \theta_{t+1}, \tilde{Y}^t) \sim N(\theta_{t|T}, \Sigma_{t|T})$$

に従う．ここで，

$$\begin{aligned}\theta_{T|T} &= E(\theta_T|\tilde{Y}^T) \\ \Sigma_{T|T} &= V(\theta_T|\tilde{Y}^T) \\ \theta_{t|T} &= E(\theta_t|\theta_{t+1}, \tilde{Y}^T) = E(\theta_t|\theta_{t+1}) \\ \Sigma_{t|T} &= V(\theta_t|\theta_{t+1}, \tilde{Y}^T) = V(\theta_t|\theta_{t+1})\end{aligned}$$

である．(4.4) は  $\tilde{\theta}_T$  のサンプリングを，まず  $P(\theta_T|\tilde{Y}^T)$  から  $\theta_T$  を生成し，その後  $t = T - 1$  から  $t = 1$  まで逐次的に  $P(\theta_t|\theta_{t+1}, \tilde{Y}^T)$  から  $\theta_t$  を生成することを示している．(4.5) における  $\theta_{T|T}$ ， $\Sigma_{T|T}$  は基本的にカルマン・フィルターのアルゴリズムにより得られる．与えられた初期値  $\theta_{0|0} = E(\theta_0)$ ， $\Sigma_{0|0} = V(\theta_0)$  および  $\hat{\sigma}_0^2$  に対し，以下の更新アルゴリズムにより逐次  $\theta_{t|t}$ ， $\Sigma_{t|t}$  を計算する．

- (i)  $\theta_{t|t-1} = F_t\theta_{t-1|t-1}$   
 $\Sigma_{t|t-1} = F_t\Sigma_{t-1|t-1}F_t' + Q_t$  ;  $Q_t = V(J_t\epsilon_t) = J_t V(\epsilon_t)J_t'$
- (ii) グループ  $j$  ( $j = 1, \dots, 8$ ) の観測値  $\tilde{Y}_t^j$  に対し， $N(0, \sigma^2)[v_{it}^j \geq -H_{t(j)}\theta_{t|t-1}]$  より， $N_j\tilde{Y}_t^j$  個の  $v_{it}$  を発生させる．ここで， $H_{t(j)}$  は行列  $H_t$  の第  $j$  行を表す．同様に  $N(0, \sigma^2)[v_{it}^j \leq -H_{t(j)}\theta_{t|t-1}]$  より  $N_j(1 - \tilde{Y}_t^j)$  個の乱数を発生させる．
- (iii) 合計  $N_j$  個の  $v_{it}$  に対し， $y_{it}^{j*} = v_{it}^j + H_{t(j)}\theta_{t|t-1}$  とし，それらの  $y_{it}^{j*}$  の平均  $\bar{Y}_t^{j*}$  を求め， $\bar{Y}_t^* = [\bar{Y}_t^{1*}, \bar{Y}_t^{2*}, \bar{Y}_t^{3*}, \dots, \bar{Y}_t^{8*}]'$  とする．
- (iv)  $\eta_{t|t-1} = \mathbf{y}_{it}^* - H_t\theta_{t|t-1}$   
 $f_{t|t-1} = H_{t-1}\Sigma_{t-1|t-1}H_{t-1}' + \bar{R}$   
 $\theta_{t|t} = \theta_{t|t-1} + K_t\eta_{t|t-1}$   
 $\Sigma_{t|t} = \Sigma_{t|t-1} - K_tH_t\Sigma_{t|t-1}$   
とする．ここで， $K_t = \Sigma_{t|t-1}H_t'f_{t|t-1}^{-1}$  である．  
以上の手順に従い， $\theta_{1|1}, \dots, \theta_{T|T}$ ， $\Sigma_{1|1}, \dots, \Sigma_{T|T}$  を得る．
- (v) 続いて平滑化を行なう．  
 $\theta_{t|T} = \theta_{t|t} + \Sigma_{t|t}F_t'(F_t\Sigma_{t|t}F_t' + Q_t)^{-1}(\theta_{t+1|T} - F_t\theta_{t|t})$   
 $\Sigma_{t|T} = \Sigma_{t|t} - \Sigma_{t|t}F_t'(F_t\Sigma_{t|t}F_t' + Q_t)^{-1}F_t\Sigma_{t|t}$   
により， $\theta_{T|T}, \theta_{T-1|T}, \dots, \theta_{1|T}$  を得た後，(4.6) 式より  $(\theta_t|\theta_{t+1}, \tilde{Y}^t)$  を発生する．これが平滑化後の状態の推定値となる．

ステップ (b)  $\hat{\sigma}^2$  の成分  $\hat{\sigma}_i^2$  は以下の事後分布から生成される．

事前分布  $\sigma_i^2|\theta_i \sim IG(\frac{\nu_0}{2}, \frac{\delta_0}{2})$  とし，(a) で得られた推定値を用い，

事後分布  $\sigma_i^2|\theta_i, \tilde{Y}^t \sim IG(\frac{\nu_1}{2}, \frac{\delta_1}{2})$  を計算する．ここに

$$\nu_1 = \nu_0 + T$$

である．また識別可能性のため  $\sigma^2 = 1$  とし， $\sigma_\beta^2, \sigma_b^2, \sigma_d^2, \sigma_\alpha^2$  に対しては，事後分布のパラメータを

$$\delta_1 = \delta_0 + \sum (\theta_{(i)t+1|T} - F_{(i)t}\theta_{t|T})^2$$

として発生させる．ここで  $\theta_{(i)}$ ， $F_{(i)}$  は  $\sigma_i^2$  に対応する適当な行を示す．

5. 擬似データによる失業率の推定

労働力調査結果は毎月総務省統計局から公表されており、そこには就業状態別人口や年齢別就業者数、失業率などの情報がクロス集計結果として与えられているが、ローテーションに対応した調査回数別の数値は与えられておらず、それゆえ、現実のデータを用いて上記モデルの有効性を確かめることはできない。実際、公表される失業率もローテーション構造を考慮した推定が行なわれてはいない。

そこで以下では完全失業率の時系列データをもとに、グループ別失業率の擬似データを作成し、試験的に前節のモデルの適合を行なう。擬似データの作成方法は以下のようになされた。1981年から2002年まで(22年間)の完全失業率の季節調整済み月次系列に対し、調査回数に応じて0.6%、-0.4%、-0.2%、0.0%を平均とした-0.125%~0.125%の一様乱数を加える。バイアスの平均は0となっている。このようにして生成された調査回数別失業率の推移は図3のとおりである。8グループの集計値データに対し、比較のため(1)ケースA: Bell and Carolan (1998)におけると同様、失業率の値  $\bar{Y}_i^j$  そのものを複数の要因に分解することを考える。ここでは潜在変数は考えなくても良い(2)ケースB:(4.1)~(4.3)で記述されるモデルを用い変動要因を推定する、の2ケースの推定を行なった。ケースBにおいては  $N = 200$  とし、 $N = 50$  についても参考として分析を行なった。以上の結果は図4~図6に見られるとおりである。

ここでのモデルにおいてはいくつかの単純化がなされている。たとえば、調査区効果は捨象されている。またローテーション・バイアスはランダム・ウォークではなく定数として扱われている。

図4-a~4-gはケースAにおける推定結果である。Gibbs Samplingの繰り返し回数は  $M_1=2000$ ,  $M_2=3000$  とした。図4-aはトレンド、4-bはバイアスの推定値を表す。トレンドはスムーズに変化している。バイアスの推定値は多少真の値と異なっている。また図4-c~4-fは  $M_2=3000$  回の繰り返しにおけるパラメータや状態の推定値のヒストグラムを描いたものである。ヒストグラムはいずれも単峰であり、また  $\sigma_\beta^2$  の推定値は定常な動きを示している。これらはいずれも繰り返しの結果、推定値が収束していることを反映している。図4-gは失業率の予測値(平滑値)である。予測値は概ねデータの動きを追っている。

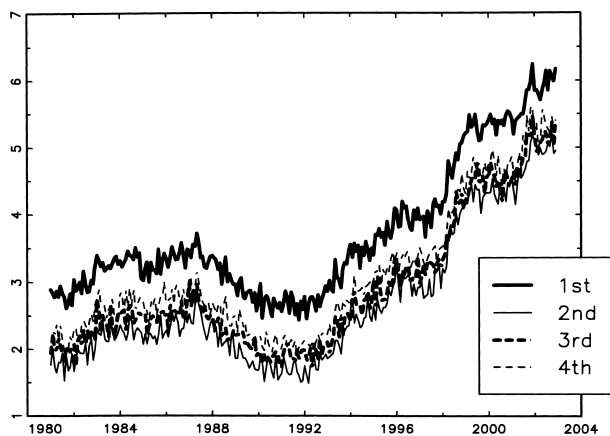


図3. 調査回数別失業率。

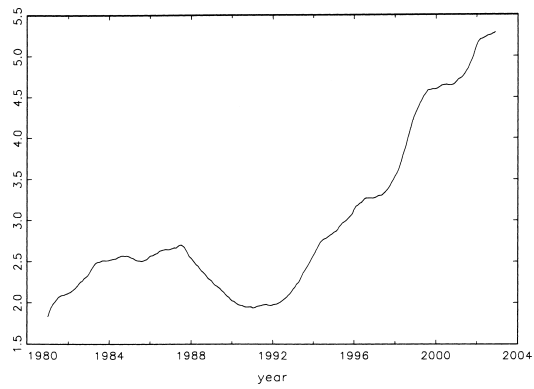


図 4-a.  $\beta_t$  の推定値 :  $t = 1981$  年 1 月 ~ 2002 年 12 月の時系列プロット .

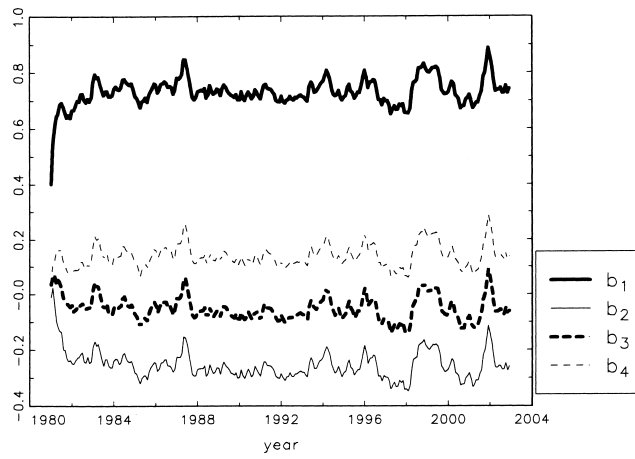


図 4-b.  $b_1, b_2, b_3, b_4$  の推定値 :  $t = 1981$  年 1 月 ~ 2002 年 12 月の時系列プロット .

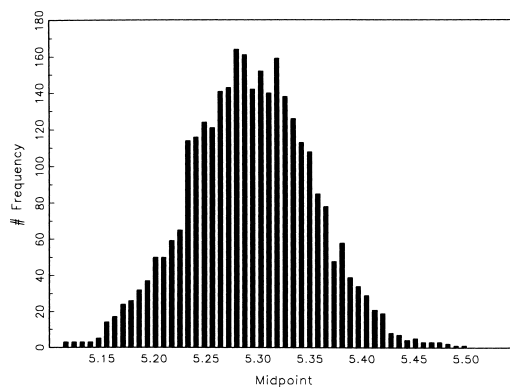


図 4-c.  $\beta_t$  の推定値 :  $t = 2002$  年 12 月のヒストグラム .

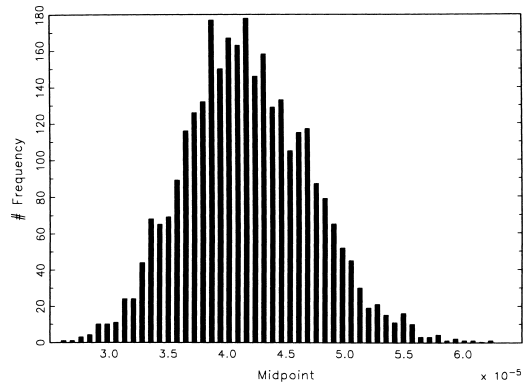


図 4-d.  $\sigma_{\beta}^2$  の推定値 : ヒストグラム .

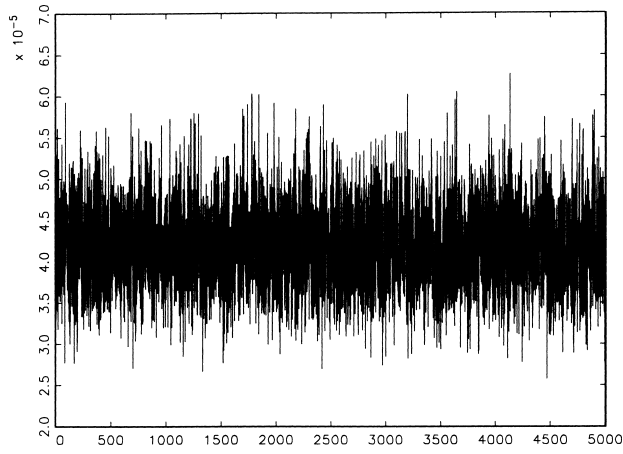


図 4-e.  $\sigma_{\beta}^2$  の推定値 : iteration 回数でのプロット .

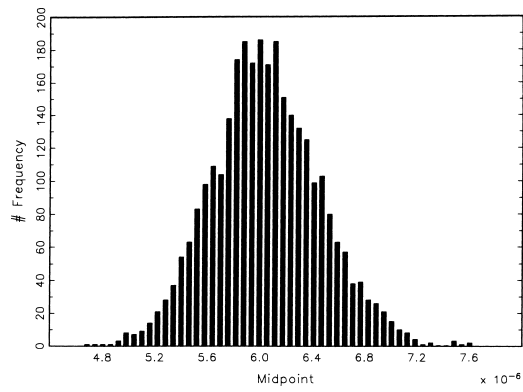


図 4-f.  $\sigma_{\alpha}^2$  の推定値 : ヒストグラム .

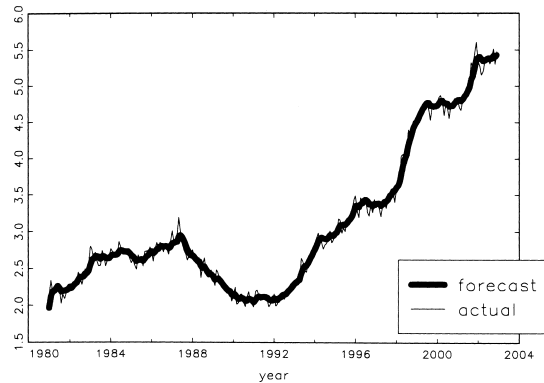


図 4-g. 失業率の予測値 :  $t = 1981$  年 1 月 ~ 2002 年 12 月の時系列プロット .

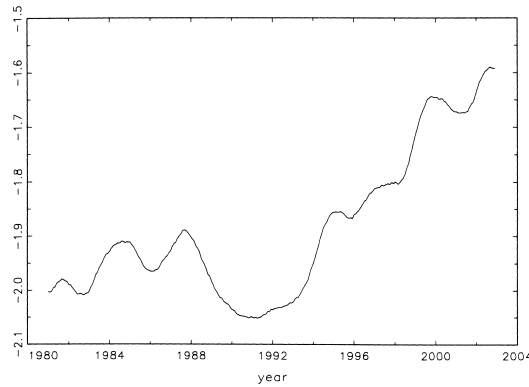


図 5-a.  $\beta_t$  の推定値 :  $t = 1981$  年 1 月 ~ 2002 年 12 月の時系列プロット .

図 5-a ~ 5-g はケース B (標本数  $N = 200$ ) における推定結果である。トレンドの推定値はケース A とやや異なるものの、よりスムーズに変化している。5-b のバイアスの推定値は図 4-b と比べ変動が大きいが、バイアスの真の大きさの順序どおり推定されている。ただし、プロビットモデルで推定される(4.2)の右辺は、個人の失業を規定する潜在変数の平均値であって集計値としての失業率ではない。それゆえ、バイアスの推定値は必ずしも現実の失業率に対して実験的に加えられたバイアスに対応しないことに注意したい。ケース A との違いは、図 5-g の失業率の予測値(平滑値)である。推定値系列はケース A の結果と比べて、若干データの動きから離れている。これは、バイナリーなデータから得られる情報が必然的に小さいことを示唆している。ただし、図 5-c ~ 5-f の繰り返しにおけるパラメータや状態の推定値のヒストグラムは単峰であり、推定値が収束していることを示している。

また図 6-a ~ 6-d はケース B において、 $N=50$  とした場合のトレンドの推定値、 $\sigma_\beta^2$  の推定値及び失業率の予測値である。プロビットによる定式化の場合、 $N=50$  では Gibbs Sampling による推定結果が安定的ではないことがわかる。これは擬似データ数 ( $N = 50$ ) の問題であると同時にその質の問題でもある。データ数  $N$  が小さいと、失業率が低いとき、擬似データの値



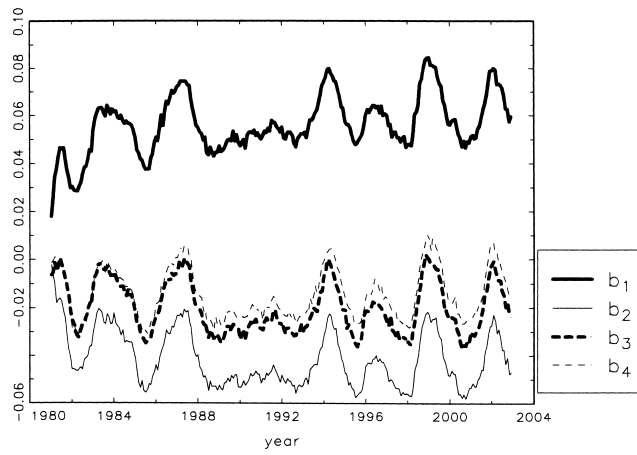


図 5-b.  $b_1$ ,  $b_2$ ,  $b_3$ ,  $b_4$  の推定値：時系列プロット。

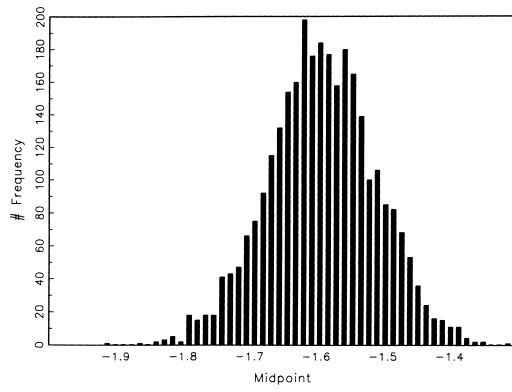


図 5-c.  $\beta_t$  の推定値： $t = 2002$  年 12 月のヒストグラム。

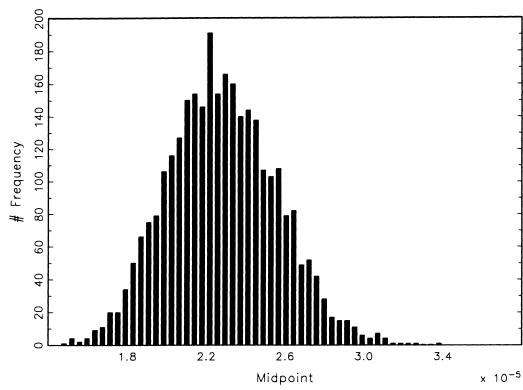


図 5-d.  $\sigma_\beta^2$  の推定値：ヒストグラム。

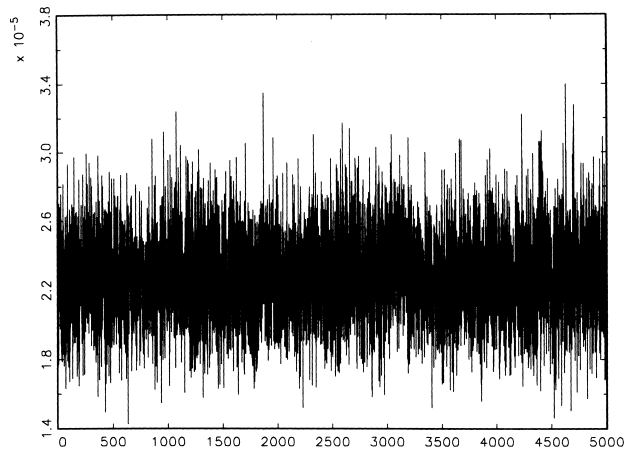


図 5-e.  $\sigma_\beta^2$  の推定値 : iteration 回数でのプロット .

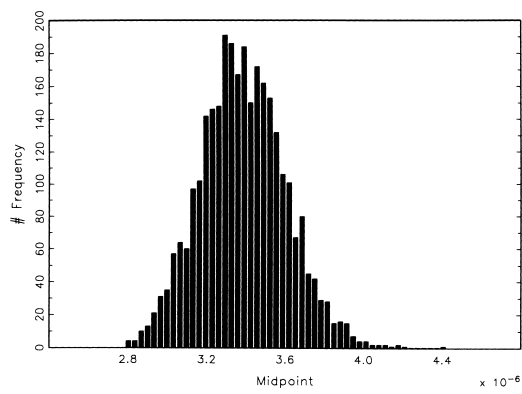


図 5-f.  $\sigma_\alpha^2$  の推定値 : ヒストグラム .

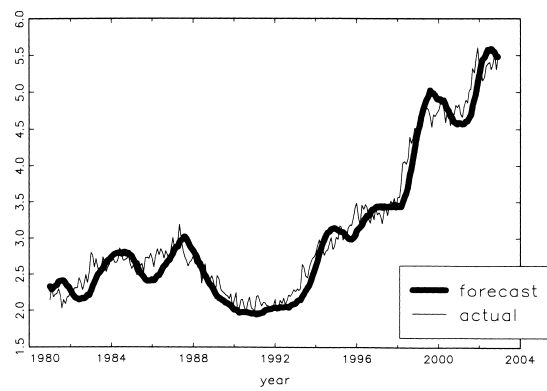


図 5-g. 失業率の予測値 :  $t = 1981$  年 1 月 ~ 2002 年 12 月の時系列プロット .

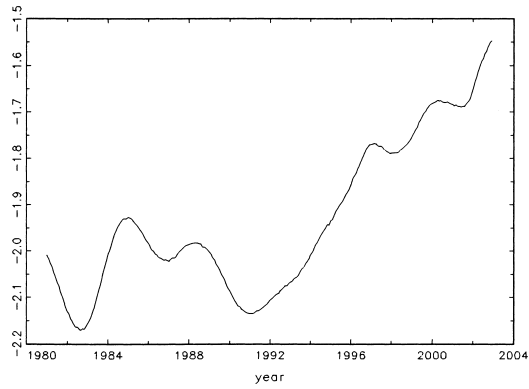


図 6-a.  $\beta_t$  の推定値 :  $t = 1981$  年 1 月 ~ 2002 年 12 月の時系列プロット .

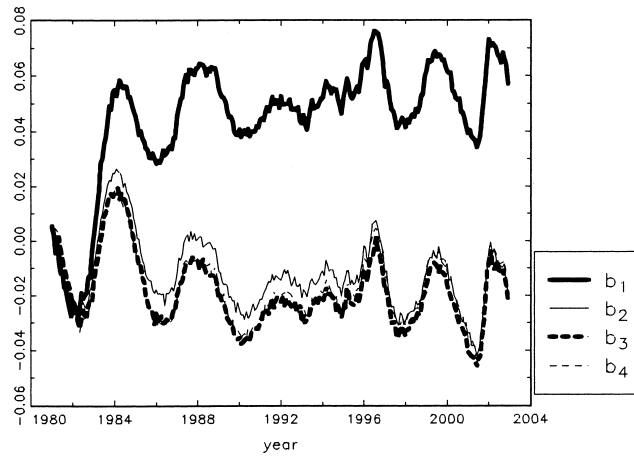


図 6-b.  $b_1, b_2, b_3, b_4$  の推定値 : 時系列プロット .

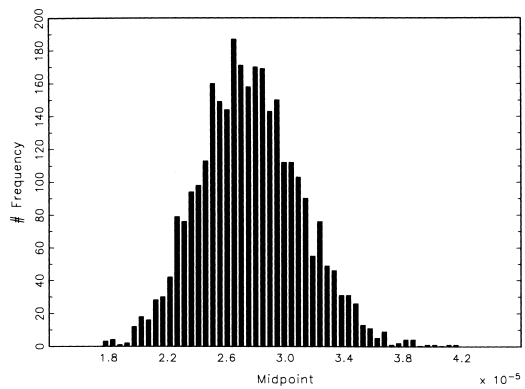


図 6-c.  $\sigma_\beta^2$  の推定値 : ヒストグラム .

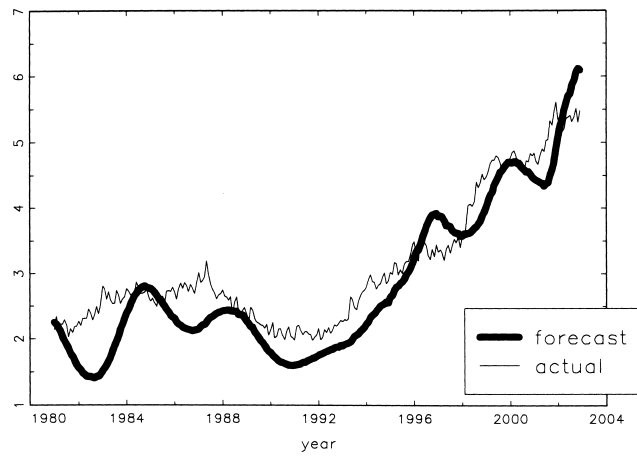


図 6-d. 失業率の予測値 :  $t = 1981$  年 1 月 ~ 2002 年 12 月の時系列プロット .

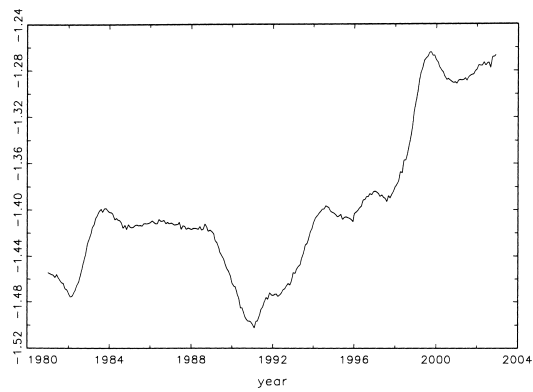


図 7-a.  $\beta_t$  の推定値 :  $t = 1981$  年 1 月 ~ 2002 年 12 月の時系列プロット .

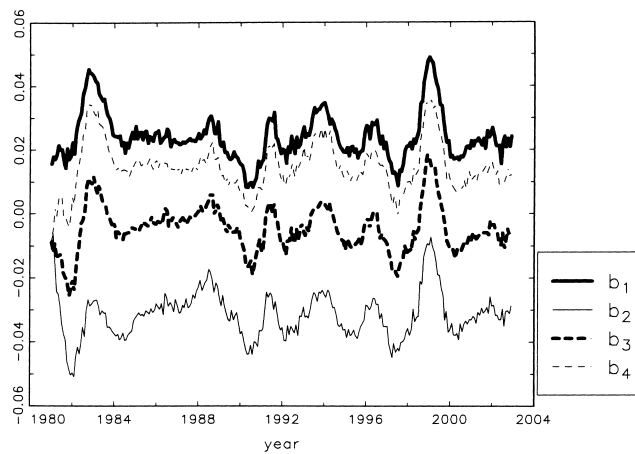


図 7-b.  $b_1, b_2, b_3, b_4$  の推定値 : 時系列プロット .

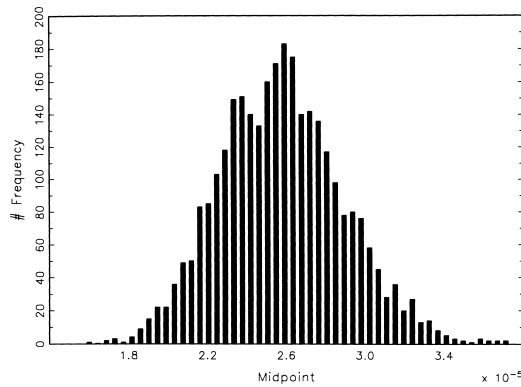


図 7-c.  $\sigma_{\beta}^2$  の推定値：ヒストグラム。

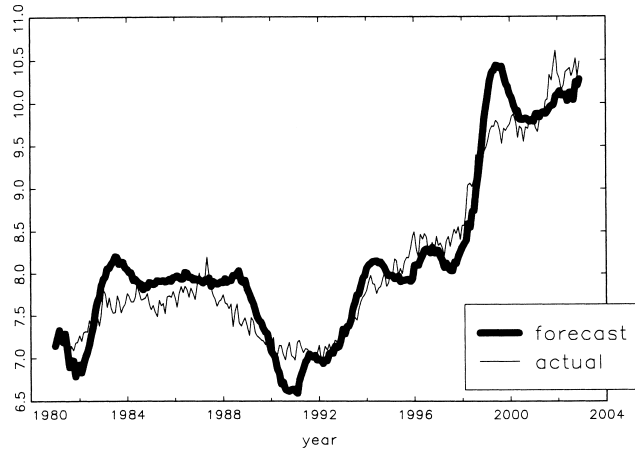


図 7-d. 失業率の予測値： $t = 1981$  年 1 月 ~  $2002$  年 12 月の時系列プロット。

は多くの 0 を続けて発生することになる．すなわちほとんど生起していない現象を分析することになる．その結果推定値が大きく振れてしまう．しかし，ある地域における高齢者の失業率の推定などを行なうのは失業率が高いからであり，現実に 0 が続く事はないはずである．そこで，現実の失業率の値に 5% を加えることにより新たな擬似データを作成し， $N = 50$  として同様の実験を行なってみた．図 7-a ~ 7-d はその結果であるが，予測値には改善が見られよりデータの動きを追っている事がわかる．

以上のように (4.1) ~ (4.3) で記述されるモデルはシミュレートされたデータに対しローテーションバイアスやトレンドなどを一定の精度で推定することに成功していることがわかる．小さなデータ数の現実データに対する分析が待たれるところである．

## 6. 結語

本稿では，わが国の労働力調査を念頭におき，ローテーション構造をもつサンプリングにより得られた観測値を分析するための時系列モデルについて論じた．公表データにおいては，

ローテーションに関する情報が得られないため、モデルを試験的に適合することに主眼がおかれた。もしわが国の労働力調査結果の集計表にローテーションに関する情報が附加されるならば、本稿で紹介したモデルの応用範囲は広いと思われる。以下にこのようなモデルの有用性について述べ結語としたい。

現在は標本の単純平均として失業率が計算されている。すでに述べたようにローテーション構造を考慮していない推定は誤差分散が大きいはずである。本モデルを用いることにより、少なくともより有効な推定がなされうるはずである。

季節性要因の除去の問題も重要な実際の問題である。すでに述べたように、現実には曜日効果やうろう年効果などを考慮に入れたより細やかな季節調整が好まれ、失業率の推定値を求めた後、X12-ARIMA 等によって季節調整値が計算されている。本モデルの状態変数として季節性要因を加えることにより、ローテーション構造が季節調整に与える影響を調べることは重要な課題であろう。

さらに本モデルを用いて調査方式の変更の影響を調べることも可能である。わが国の労働力調査の調査方式はたびたび変更されてきている。さらに IT 化の進行に伴い、今後インターネットによる調査等も近々導入されるであろう。そのような場合に新たな調査方式のインパクトを把握することは重要な課題であろう。Bell (1998) は状態変数のなかにステップ関数を導入することにより、電話によるインタビュー方式が労働力調査結果に与える影響を分析している。本モデルを用いて、わが国における調査方式の変更の影響を調べることは是非試みるべき実験であろう。

しかし、モデルのより重要な応用は、ローテーション構造の決定であろう。わが国で用いられている 2-10-2(4)、オーストラリアの労働力調査で用いられている 8-0-0(8) というローテーション構造は果たして妥当なものであろうか。どのようなローテーション構造の調査を行なうべきか、またそもそもローテーション標本調査を行なうべきかは各国の現実のデータの振る舞いに依存するはずである。上のモデルで言えば、 $\sigma_\alpha^2$  や  $\sigma_b^2$  は標本調査にローテーションを導入するがゆえに生ずる誤差である。 $\beta_t$  やその変化など、興味あるパラメータの変動の大きさとこれらの要因の相対的な大小関係によって、どのようなローテーション構造の調査を行なうべきか、そもそもローテーション標本調査を行なうべきか否か等が決定される。この点は、今後検討すべき重要課題である。

わが国の官庁統計を用いた実証研究はごく初歩的な分析例を除きあまり見られず、この点では諸外国に比べ大きく遅れている。それはわが国の官庁統計の一般利用が難しいことに起因する。今後はより多くの研究者に利用されるような調査結果の公表を考えるべきであろう。それによって研究者からのフィードバックがなされ、調査方法の改善も期待される。中長期的な視点に立ち、より質の高い統計の蓄積を目指すべきであろう。

## 謝 辞

一橋大学経済学研究科の経済統計ワークショップ、東京大学経済学研究科の応用統計ワークショップの参加者、および 2 人の匿名レフリーから貴重なコメントをいただいた。記して感謝致したい。なお、本研究は統計数理研究所共同研究プログラム(14-共研-2024)、および文部科学省科学研究費補助金(13553001, 14208023)の援助を受けている。

## 参 考 文 献

- Bailar, B. A. (1975). The effect of rotation group bias on estimates from panel surveys, *J. Amer. Statist. Assoc.*, **70**, 23–30.
- Bell, P. A. (1998). Trend estimation from small areas from a continuing survey with controlled sample overlap, Australian Bureau of Statistics, Working Paper, No. 98/1.
- Bell, P. A. and Carolan, M. (1998). Using state space models and composite estimation to measure the effects of telephone interviewing on labour force estimates, Australian Bureau of Statistics, Working Paper, No. 98/2.
- McLaren, C. H. and Steel, D. G. (1997). The effect of different rotation patterns on the sampling variance of seasonal and trend filters, *Proceedings of the Survey Research Methods Section, American Statistical Association*, 790–795.
- Kim, C-J. and Nelson C. R. (1999). *State-space Models with Regime Switching*, The MIT Press, Cambridge, Massachusetts.
- Pfeffermann, D., Feder, M. and Signorelli, D. (1998). Estimation of autocorrelations of survey errors with application to trend estimation in small areas, *J. Bus. Econom. Statist.*, **16**, 339–348.
- 総務庁統計局 (1999). 『労働力調査標本設計の解説』.

## Labor Force Survey and Rotation Sampling

Satoru Kanoh

(Institute of Economic Research, Hitotsubashi University)

The labor force survey in Japan has a unique characteristic in its sampling scheme called the rotation sampling. This paper proposes a time-series model for analyzing the result of this survey taking this sampling scheme into consideration. In the time series model, individual unemployment is expressed by the probit model and its latent variable is related to trend, group effect and individual effect. Because of the model's non-linearity, the maximum likelihood method using the Kalman Filtering algorithm is not applicable for estimating the model's parameters. In this paper, the estimation is carried out by the Gibbs sampling technique. Unfortunately, the survey result does not include information regarding the rotation groups and, for this reason, an experiment to validate the model must rely on simulated data. Finally, a couple of important applications and extensions of the model are given to conclude this paper.

---

Key words: Labor force survey, rotation sampling, probit model, Gibbs sampling, state space presentation.