

従業上の地位別結果の推計に関する検討状況

1 背景

2013年1月に、労働力調査基礎調査票の「従業上の地位」に関する項目を変更し、「常雇の人」を「常雇の人（無期の契約）」と「常雇の人（有期の契約）」の2区分に分割した。（臨時・日雇の区分は変更していない）

この結果、一般常雇（調査票で「常雇の人」と回答した者）の人数が対前年同月で+288万人、臨時雇は▲227万人となり、これまでの調査事項と単純比較が困難となった。

※ この内容は、ホームページ上で「労働力調査の結果を見る際のポイント」として注意喚起を行うなど、利用者への周知を図っている。

調査票変更後の新たな区分について2013年1月以降の時系列データは作成、公表しているところであるが、従来の系列（常雇・臨時・日雇の3区分。以下「旧系列」という。）へのニーズもあり、調査票変更後のデータから推計を行うこととなった。

これを踏まえ、2014年12月に開催した第2回雇用失業統計研究会及び2015年3月に開催した第3回雇用失業統計研究会では、2013年の旧系列に相当する結果を推計した。

第3回研究会において、2013年推計結果が了承されたが、平均比率は時間の経過に伴い変化するため、2014年推計では更なる推計の工夫が必要との指摘があった。

これらの指摘に対応するため、第4回雇用失業統計研究会において、「従業上の地位別結果の推計に関する分科会」を立ち上げ、分科会の場において推計方法を検討することとされた。

以上の背景の下、分科会を平成28年1月12日に開催した。このほか、推計方法に関し、数次にわたり打合せを実施したところである。

2 推計方法について

第4回雇用失業統計研究会で提示した方法（以下の(1)のほか、(2)及び(3)の推計を行った。

また、分科会では、(4)の方法についても検討し、推計を試みることとされた。

(1) 回帰式を用いた断層の推定（Level Shift）

雇用者数の時系列に対して、スパイク・レベルシフトを検出して、これを取り除く方法。第4回雇用失業統計研究会で提示したもの。

雇用者の一定数が、調査票変更前後で回答傾向が変わるとの考えに基づく方法であり、雇用者数そのものが大幅に変わると、適切でない推計となる。

(2) 遷移行列を用いた従業上の地位別結果の推計（Transition Matrix）

労働力調査が2か月連続世帯を対象とする特性を踏まえ、前月×今月の「従業上の地位」の変化を確率的に表し（遷移行列の作成）、確率の時系列データに対してスパイク・レベルシフトを検出して、これを取り除く方法。

雇用者の一定割合が、調査票変更前後で回答傾向が変わるとの考えに基づく方法であり、雇用者数そのものが変動しても問題ない。

(3) 多項ロジットモデル (Multinomial Logistic)

ある月の「従業上の地位」は、男女、年齢、産業、職業など、いくつかの要素で確率的に決定されると考え、調査票変更前のモデル(パラメータ)を推計し、調査票変更後のデータに適用して推計する方法。

雇用者の男女、年齢構成などが変化した場合でも、それらの変化を取り込んだ形で推計することができる方法であり、また、個別データに確率が付与されるので、男女・年齢階級別等の推計も可能な手法。

(4) 2か月間の従業上の地位を被説明変数とした多項ロジットモデル
(applying Multinomial Logistic to Transition Matrix)

前月から当月に「従業上の地位」が変化する確率が、男女、年齢、産業、職業など、いくつかの要素で確率的に決定されると考える方法。

3 推計結果

2の(2)から(4)の方法を適用した結果は、それぞれ、

遷移行列を用いた従業上の地位別結果の推計 別紙1 参照
多項ロジットモデル 別紙2 参照
2か月間の従業上の地位を被説明変数とした多項ロジットモデル . . . 別紙3 参照

のとおりとなっている。なお、比較のため、レベルシフトによる方法も掲載している。いずれの方法も、単純に適用した場合にはレベルシフトによる方法と乖離がある。推計方法の当てはまりの良さを評価し、乖離の補正を行うと、レベルシフトによる方法と同程度の数値となる。

4 検討結果(事務局案)

以上のとおり、どの推計方法でもそれなりの時系列結果を出すことができていると考えられるが、この中でも多項ロジットモデルは、男女、年齢等の結果が、同じ推計結果から作成できる点で優れていると考える。

※ 他の方法の場合、時系列モデルでスパイク・レベルシフトを検出して取り除くため、例えば男女別の結果数値を得ようと思えば、男女別の時系列データを作成し、スパイク・レベルシフトを検出して推計する手順が必要になる。

遷移行列を用いた従業上の地位別結果の推計

1. 遷移行列の定義

連続する 2 か月間調査対象となった標本における従業上の地位についての移動を個別データより集計する。集計には、2 か月目のサンプルのみに付与されている「1/2 集計用乗率(H)」を使用する。1/2 集計用乗率は、2 か月目サンプルの集計用乗率のみで推計人口にベンチマークされた乗率である。また、2 か月目のみの転入者や、2 か月間のうちどちらかが従業上の地位が不詳となっているサンプルについて除外する。行列の縦方向の成分（前月の従業上の地位）の合計が 1 になるように行列を規格化することで、前月と今月の従業上の地位についての遷移行列ができる¹。

$$a_{i,j}^t = \frac{h_{i,j}^t}{\sum_i h_{i,j}^t} \quad \begin{array}{l} i: \text{今月の従業上の地位} \\ j: \text{前月の従業上の地位} \end{array} \quad \dots (1)$$

この行列を先月の就業者以外を含めた各成分の就業者数を縦に並べたベクトルに作用させることで、今月の従業上の地位別就業者数を導出することが期待される。

$$x_i^{t+1} = \sum_{j=1}^5 a_{i,j}^{t+1} x_j^t \quad \dots (2)$$

2. 遷移行列の各成分について調査票変更による影響の確認

遷移行列の各成分を時系列で表示し、遷移行列の成分について、調査票の変更による断層がどのように発生しているかを確認する。

一般常雇を「有期契約」と「無期契約」に分けたことにより、2012年12月に「臨時雇」と回答した者が、翌1月は「一般常雇（有期）」へ回答を移した影響から、2013年1月に急激な変化（spike:スパイク）が発生していることが確認できる。

また、2月以降は同じ調査票による回答なのでスパイクは発生しないが、水準が変わっていることがわかる。

レベルシフトの要因は、1月の急激な就業者数の増減により、2月以降の遷移行列の分母が拡大または縮小したことと考えられる。このため、調査票の変更による影響が直接あった「一般常雇」と「臨時雇」間の移動のみでなく、他の成分間についてもレベルシフトが生じている。また、(1)は前月の従業上の地位別就業者を分母としていることから、2013年1月の遷移行列にはレベルシフトは存在しないことがわかる。

¹ 数式におけるラベルを 1:一般常雇、2:臨時雇、3:日雇、4:その他(役員、自営業主等)、5:就業者以外 とする。

3. ダミー変数を用いた断層の除去²

調査票が変更された影響は2013年1月のスパイクと2月以降のレベルシフトによって現れることがわかった。ここでは遷移行列の時系列について、季節調整で用いられる異常値除去の方法により、その影響の大きさを定量的に測り、調整した時系列を作成する³。

2013年1月のスパイクは、以下の加法的外れ値 (additive outliers) を入れることで、そのパラメーターにより水準差以外の違いを表すことができる。

$$AO_t = \begin{cases} 0 & t \neq 2013.1 \\ 1 & t = 2013.1 \end{cases} \quad \dots (3)$$

また、2013年2月前後の水準の違いは、レベルシフト (level shift) を回帰式に入れることで、そのパラメーターにより前後の水準差を定量的に表すことができる。

$$LS_t = \begin{cases} 0 & t \leq 2013.1 \\ 1 & t \geq 2013.2 \end{cases} \quad \dots (4)$$

これらを踏まえ、次の回帰式を考える。

$$(1 - B^{12})(a_t - LS_t \times \beta_L - AO_t \times \beta_A) = u_t$$

a : 観測値

β_L : 水準差を表すパラメーター

β_A : スパイクを表すパラメーター

u : 残差項

B : 階差作用素 ($Bx_t = x_{t-1}$)

$\dots (5)$

ここで、労働力調査の標本交代による誤差を取り除くため、前年との階差をとっている⁴。

4. 遷移行列による雇用者数の計算

ダミー変数により調整した遷移行列を用いて、(2)式により、前月の地位別雇用者数に遷移行列を乗ずることで、当月の雇用者数を推計する。初期値は2012年12月とする。また、毎月の「就業者以外」の人口は調査票変更の影響を受けていないとするため、上式での計算の後に既公表値に置き換える。同様に、「一般常雇」「臨時雇」「日雇」「その他」の合計も、就業者数の合計を公表値に合うようにする⁵。

² 遷移行列を調整する方法のほか、その元となっている実数値を直接調整する方法も考えられる。実数値について、ダミー変数を入れて調整し、調整後の実数値から(1)式を用いて遷移確率を計算する。

³ 「就業者以外」は調査票の変更による影響はないと考えられるため、調整は行わない。

⁴ 当初は前月との階差も入れていたが、1/2乗率は2か月目サンプルのみを対象としているため、標本交代による前月サンプルとの関係性はなく、前月階差は不要であるとした。また、この関係式は季節調整プログラム X12-ARIMA において用いられる RegARIMA モデル ($d=0, D=1$) に対応しているため、それを用いて計算する。

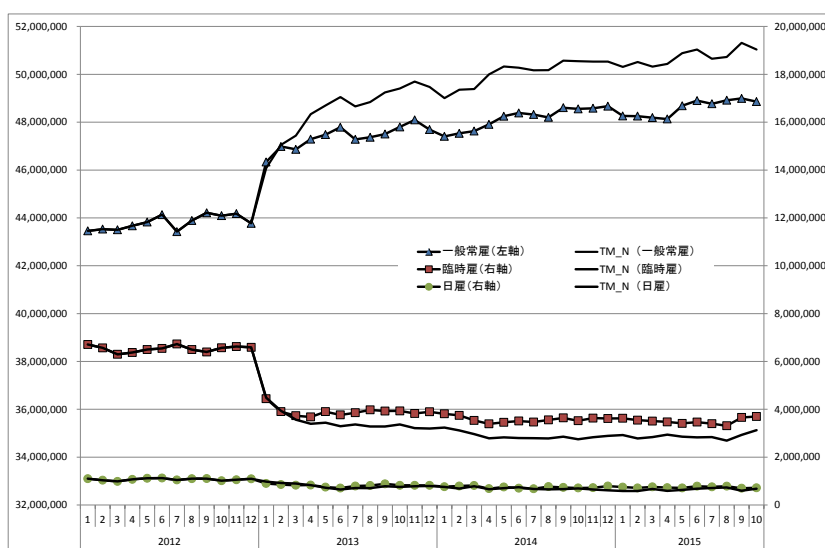
⁵ 2013年1月のスパイクと2013年2月以降のレベルシフトのそれぞれについて、遷移行列の縦成分の合

5. 検証

推計方法の当てはまりを評価するため、調整前の遷移行列を用いて(2)式で時系列に展開する。当てはまりが良ければ、公表値を再現できることが期待されるが、調整前の推計値と公表値と比較すると、一般常雇は上方に乖離し、臨時雇は下方に乖離した。これは、遷移行列を用いた(2)式による推計方法に問題があることを示唆している。

例えば、ある月で一般常雇が公表値より大きくなった場合、次月の推計はさらにその値を前月値として用いるため、当てはまりの悪い月の影響が後々に続くことになる。

図表 1：調整前遷移行列による結果 (TM_N)



6. 補正

このバイアスを解消するため、遷移行列の調整とは別の補正が必要となる。例えば、調整のない推計値を公表値と一致させるための補正項を(2)式に追加する方法が考えられる。

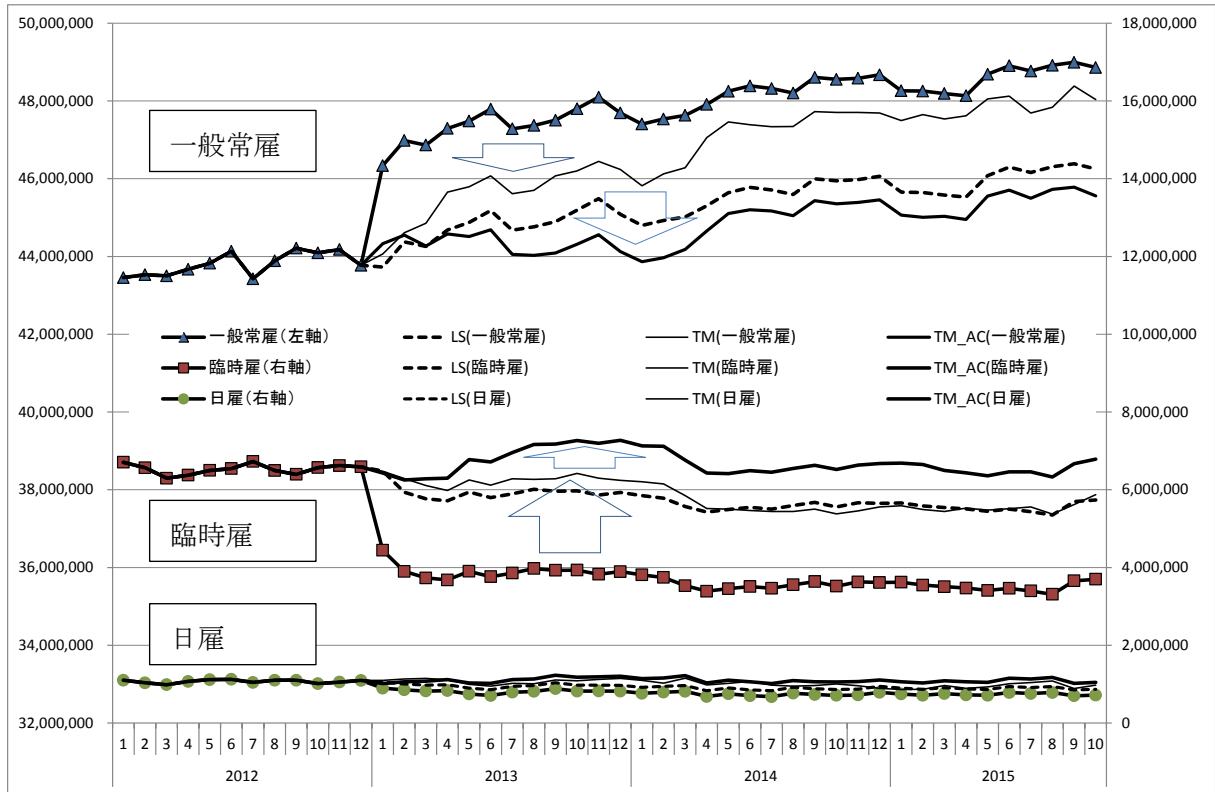
$$\textcircled{1} \quad \text{乗法補正} \quad \overline{x_i^{t+1,c}} = c_i^{t+1} \times \sum_{j=1}^5 a_{i,j}^{t+1} \times x_j^{t,c} \quad c_i^{t+1} = \frac{P_i^{t+1}}{\sum_{j=1}^5 \hat{a}_{i,j}^{t+1} \times P_j^t} \quad \dots (6)$$

$$\textcircled{2} \quad \text{加法補正} \quad \overline{x_i^{t+1,d}} = d_i^{t+1} + \sum_{j=1}^5 a_{i,j}^{t+1} \times x_j^{t,d} \quad d_i^{t+1} = P_i^{t+1} - \sum_{j=1}^5 \hat{a}_{i,j}^{t+1} \times P_j^t \quad \dots (7)$$

各調整項の \hat{a} は調整前の遷移行列、 P は従業上の地位別就業者数の公表値である。右辺の遷移行列を調整前のものにおきかえる ($a \rightarrow \hat{a}$) と、 $x_j^{t,(c,d)} = P_j^t$ で $x_i^{t+1,(c,d)} = P_j^{t+1}$ となり、公表値と一致する。

計が 1 になる規格化条件を満たすため、各推計値の合計を 0 にする必要はあるが、これにより、15 歳以上人口が推計値と公表値で一致するため、遷移行列の規格化と同等の効果があると考えられるため、水準調整後の遷移行列の再規格化は行わない。

図表 2：遷移行列による推計結果 (TM_AC)



LS . . . レベルシフト
 TM . . . 遷移行列による推計結果 (補正無し)
 TM_AC . . . 加法補正による推計結果

多項ロジットモデルによる従業上の地位別結果の推計

1. 多項ロジットモデル

多項ロジットモデルは、3つ以上の選択肢に対し、各サンプルの属性情報と実際の選択との関係を定量的に表すモデルである。また、その関係性から選択を確率的に予測することができる。サンプル*i*が選択肢*j*を選ぶ確率は以下で表される。

$$p_{i,j} = \frac{\exp\left[\sum_k \beta_{j,k} x_{i,k}\right]}{\sum_j \exp\left[\sum_k \beta_{j,k} x_{i,k}\right]} \quad \dots (1)$$

x はサンプルが属性 k に含まれる場合は 1 を、含まれない場合は 0 となるダミー変数であり、 β は属性 k に含まれるサンプルが選択肢 j を選ぶ割合をあらわすパラメーターである。

2. 推計に使用するデータ

本推計では、1月から12月の1年分のクロスセクションデータを用いてロジットモデルの推計を行った。また、各レコードには集計用乗率をウエイトとして付与した上で、最尤法により説明変数のパラメーター β を求めた¹。

3. 被説明変数

調査票の変更による回答状況の変化は、主に「臨時雇」と「一般常雇（有期）」との間で起こった。しかし、他の選択肢間への影響も取り入れるため、従業上の地位のすべての選択肢を被説明変数に含めることとする。本推計では、被説明変数として、従業上の地位の選択肢のうち、「一般常雇（無期+有期）」「臨時雇」「日雇」の3つと、役員、自営業主、家族従業員、内職などをまとめた「その他」の合計4通りを設定する²。

4. 説明変数

説明変数は、基礎調査票から得られる情報をもとに、性別ごとの年齢10歳階級（2×6通り）、続柄（11通り）、子の年齢階級別有無（5通り）、就業状態（4通り）、勤め先の経営組織（4通り）、従業員数（10通り）、産業（20通り）、職業（11通り）を設

¹ 5節では2012年の個別データ1年分を用いてパラメーターを求め、6節では、5節で求めたパラメーターを用いて2013年1月以降の各月の個別データの属性情報から選択確率を推計した。7節では、2014年の個別データからパラメーターを求め、2012年1月以降の各月の個別データの属性情報から選択確率を推計した。

² ここでは選択肢間に類似性や順序がないとする。

定した³。

5. 2012年データによるパラメーター推定

推定値 β は、各選択肢に対する属性の寄与を表している。ただし、各選択肢の絶対水準を求める必要はなく、特定の基準に対する相対的な選択確率が求めれば、確率の合計を1にすることで各選択肢の選択確率を求めることができる。以下の推定では、「臨時雇」を基準とした。各選択の相対確率への寄与度は下式により計算できる⁴。

$$p_{i,j,2} = \frac{P_{i,j}}{P_{i,j=2}} = \exp\left(\hat{\alpha}_j + \sum_k \hat{\beta}_{j,k} x_{i,k}\right) \quad \dots (2)$$

6. 2013年以降の推計

前項で得られたパラメーターを用いて、2013年以降の各サンプルに対して従業上の地位の選択確率 $\hat{p}_{i,j,2}$ を計算し、その確率と集計用乗率 Y_i を乗じることにより、2012年以前の属性情報と従業上の地位との関係式に基づく、2013年1月以降の推計値を計算する。

$$\text{選択確率: } \hat{p}_{i,j,2} = \exp\left(\hat{\alpha}_j + \sum_k \hat{\beta}_{j,k} x_{i,k}\right) \quad \dots (3)$$

$$\text{推計乗率: } \hat{Y}_{i,j} = \frac{\hat{p}_{i,j,2}}{\sum_l \hat{p}_{i,l,2}} \times Y_i \quad \dots (4)$$

計算結果について、ML（ロジットモデルによる推計結果）とLS（レベルシフトによる水準調整のみの結果⁵）を比較すると、徐々に両者の差が拡大した。MLとLSの乖離の要因として、①調査票の変更による影響と②就業構造の変化による影響の2つが考えられる。

7. 検証

モデルの当てはまりを検証するため、2014年の個票データを用いたロジットモデルによる2013年の分類を考える。両年の調査票に違いがないことから、ロジットモデルによる推計値と原数値の差は、②の就業構造要因の差と考えることができる（①はロジットモデルと原数値で変化はない）⁶。推計の結果、2013年の一般常雇はロジットモデルの方が実測値より雇用者数が多くなった。これは、ロジットモデルでは2013年においても一般常雇と計測したのに対し、実際は雇用情勢により同じ属性においても一

³ これらからステップワイズ法により変数の選定を行う。

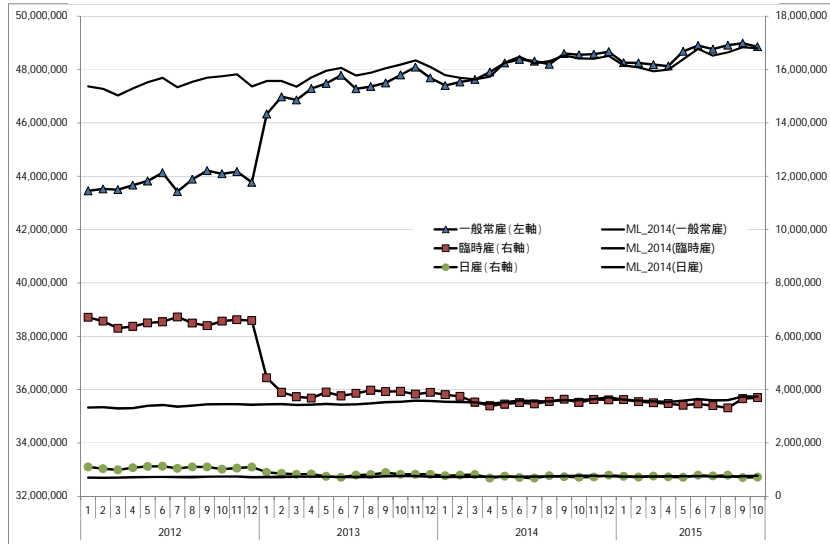
⁴ 数式における被説明変数のラベルを1:一般常雇、2:臨時雇、3:日雇、4:その他とする。

⁵ 原系列を、LS2013.1をダミー変数としてRegARIMAモデルにより異常値を除去した結果である。このため、2013年1月以降は固定されたパラメーターにより水準調整される。

⁶ ただし、この差は2014年の属性情報を基準とした各年の雇用者数であることに注意が必要である。

一般常雇以外に属していたことを示している。また、一般常雇と臨時雇の2014年基準推計値と実測値を見ると、2014年4月以降は推計値と実測値が概ね重なることから、この時点以降の就業構造要因は解消されていると推測される。

図表 1：2014年を基準にした推計結果 (ML_2014)

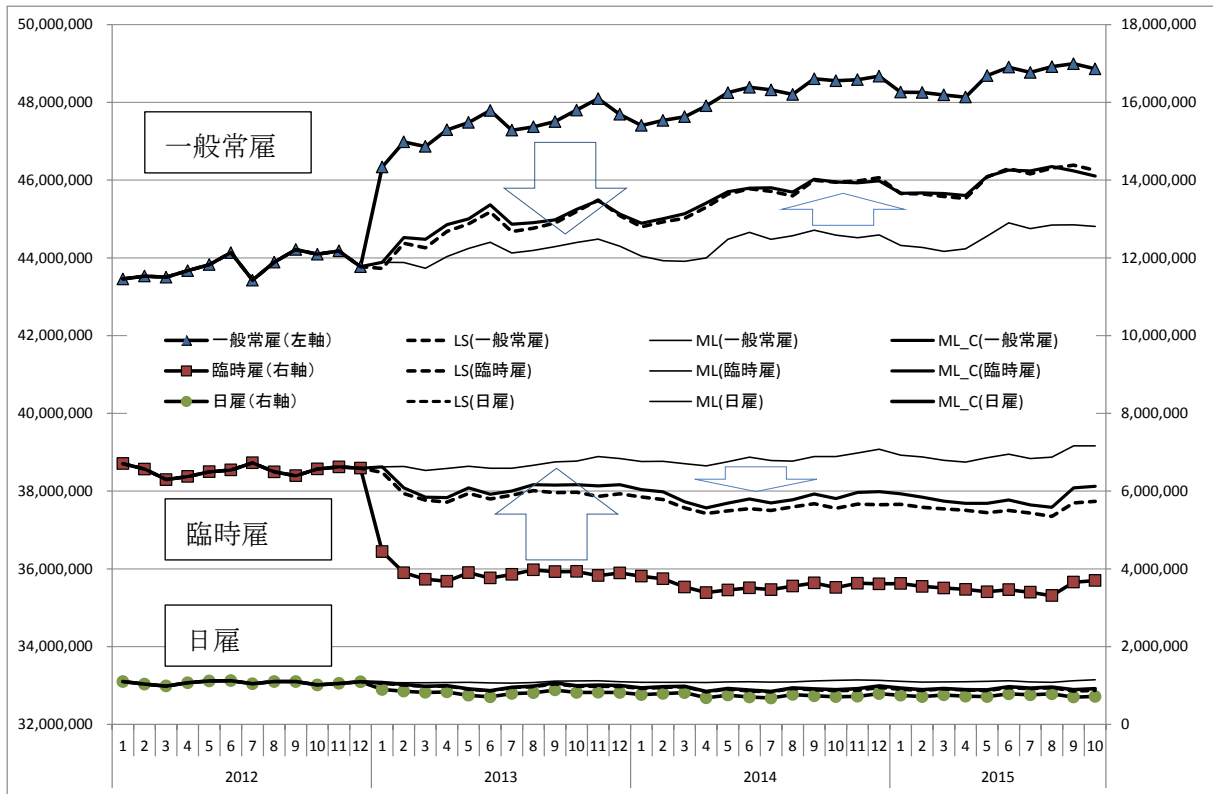


8. 補正

7節の就業構造要因分を用いて6節の計算結果を補正する。就業構造要因の補正を、実測値と2014年基準の推計値との差の累積分を各月の補正值として下式で計算する。

$$\begin{aligned}
 \tilde{Y}_t &= \hat{Y}_t + \sum_{T=2013.2}^t (\delta_{T-1} - \delta_T) = \hat{Y}_t + (\delta_{2013.1} - \delta_t) \\
 &= \underbrace{\hat{Y}_t}_{\text{2012年基準の系列}} + \underbrace{(Y_t - Y_{2013.1})}_{\text{原系列の増減 (調査票の影響+雇用情勢の影響)}} - \underbrace{(\bar{Y}_t - \bar{Y}_{2013.1})}_{\text{2014基準ロジットモデル (調査票の影響のみ)}}
 \end{aligned}
 \left(\begin{array}{l} \delta_t = \bar{Y}_t - Y_t \\ \bar{Y}_t: \text{2014年基準推計系列} \\ Y_t: \text{原系列} \\ \tilde{Y}_t: \text{補正後系列} \\ \hat{Y}_t: \text{2012年基準推計系列} \end{array} \right) \dots (5)$$

図表 2：補正後推計結果(ML_C)



LS . . . レベルシフト
 ML . . . 多項ロジットモデルによる推計結果 (補正なし)
 ML_C . . . 多項ロジットモデルによる推計結果 (補正あり)

2 か月間の従業上の地位を被説明変数とした
多項ロジットモデルによる従業上の地位別結果の推計

はじめに

分科会において、同一サンプルにおける2 か月間の従業上の地位を被説明変数とする多項ロジットモデルによる推計の提案があった。本資料ではその推計手順を説明する。

1. 被説明変数

従業上の地位の選択肢を、「一般常雇（無期+有期）」「臨時雇」「日雇」の3つと、役員、自営業主、家族従業員、内職などをまとめた「その他」の合計4通りにまとめ、各サンプルの前月と今月の選択を組み合わせた「4×4=16」通りの被説明変数を設定する¹。

2. 2012年データによるパラメーターの推計

2 か月間共に「臨時雇」を選択した場合(22)を基準とした相対確率を求める²。

$$p_{i,j,22} = \frac{P_{i,j}}{P_{i,j=22}} = \exp\left(\hat{\alpha}_j + \sum_k \hat{\beta}_{j,k} x_{i,k}\right) \quad \dots (1)$$

3. 2013年以降の1/2乗率の推計

前項で得られたパラメーターを用いて、2013年以降の各サンプル(i)に対して従業上の地位の予測確率 $\hat{p}_{i,j,22}$ を計算し、その確率と1/2集計用乗率 H_i を乗じることにより、2012年以前の属性情報と2 か月間の従業上の地位との関係式に基づく、2013年1月以降の推計値を計算する³。

$$\text{予測確率: } \hat{p}_{i,j,22} = \exp\left(\hat{\alpha}_j + \sum_k \hat{\beta}_{j,k} x_{i,k}\right) \quad \dots (2)$$

$$\text{推計 1/2 乗率: } \hat{H}_{i,j} = \frac{\hat{p}_{i,j,22}}{\sum_{l=1}^{44} \hat{p}_{i,l,22}} \times H_i \quad \dots (3)$$

$$\text{調整後 1/2 乗率: } \hat{H}_j = \sum_i \hat{H}_{i,j} \quad j: 11 \sim 44 \quad \dots (4)$$

¹ 推計に使用するデータと説明変数は別紙2と同じである。

² 数式における被説明変数のラベルは、従業上の地位の選択肢を1:一般常雇、2:臨時雇、3:日雇、4:その他とし、今月と前月の選択肢を並べた11~44とする(前月「臨時雇」今月「一般常雇」の場合「12」となる)。

³ 別紙2で示した補正は行っていない。

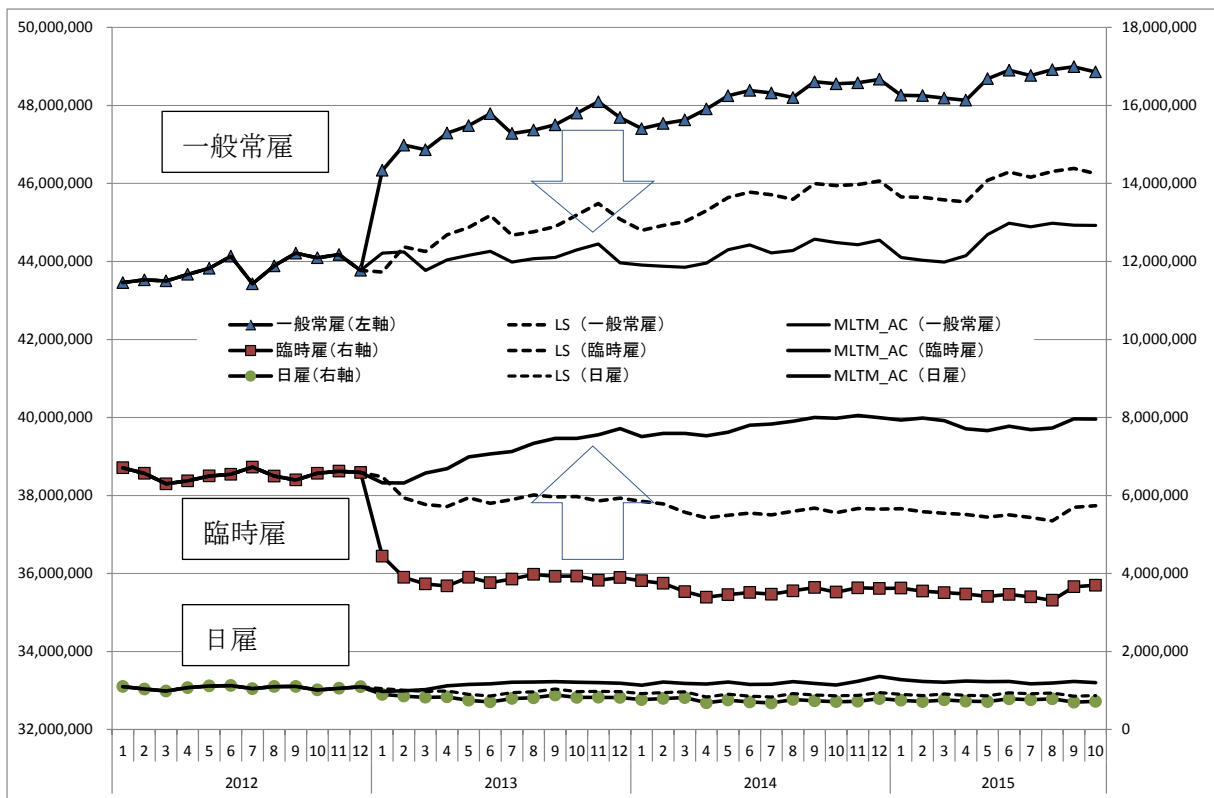
4. 遷移行列の作成

前節で作成した 1/2 乗率から遷移行列を作成する⁴。

$$\hat{A}_{i,j}^t = \frac{\hat{H}_{i,j}^t}{\sum_i \hat{H}_{i,j}^t} \quad \dots (5)$$

この行列を先月の就業者以外を含めた各成分の就業者数を縦に並べたベクトルに作用させることで、今月の従業上の地位別就業者数を導出する。以下は、別紙1で示した方法により推計を行う。

図表 1：計算結果 (MLTM_AC)



LS	・・・レベルシフト
MLTM_AC	・・・多項ロジットモデル (1/2 乗率) + 遷移行列 (加法補正)

⁴ 多項ロジットモデルでは前月と当月の選択肢を結合した被説明変数を設定したため、遷移行列の成分を表す際は (j=11→i=1, j=1) のような添字の読み替えが必要である。