新たな「外れ値」処理手法の詳細

１．新たな「外れ値」処理手法のポイント

・CI を構成する個別系列について、その変動を「共通循環変動」と「系列固有変動」に分解。「系列固有変動」を「外れ値」処理の対象とすることによって、体系全体に対する共通ショックが「外れ値」として処理されることを防ぐことが可能。

具体的には、

－「共通循環変動」とは、「外れ値」処理なし CI を算出する過程で得られる「『外れ値』処理なし基準化変化率」の中央値とし、先行、一致、遅行指数毎に定義（下表「（１）」を参照）。

－「系列固有変動」とは、
系列の「基準化変化率」が「共通循環変動」から乖離した部分と定義（詳細は、下表「（２－１）『外れ値』処理対象を選定」を参照）。

・この「系列固有変動」から算出される対称変化率を用い、当該系列の「外れ値」処理の是非を決定。

２．処理手法の詳細（比較のため、右欄に旧手法を併記している。）

新たな手法 （平成 23 年 9 月分から使用）	計算ステップ		旧手法 （平成 23 年 8 月分まで使用）
（対称変化率） $r_i^j(t) = 200\times\dfrac{y_i^j(t)-y_i^j(t-1)}{y_i^j(t)+y_i^j(t-1)}$ （トレンド） $\mu_i^j(t) = \dfrac{\sum_{\tau=t-59}^{t-s} r_i^j(\tau)}{60-s}$ （四分位範囲基準変化率） $z_i^j(t) = \dfrac{r_i^j(t)-\mu_i^j(t)}{Q3_i^j - Q1_i^j}$ ・・・① （「共通循環変動」） $ZC^j(t) =$ ①の中央値	**（１）事前処理** **（「外れ値処理」なし四分位範囲基準化変化率等を算出）** 個別系列の「外れ値処理」なし -対称変化率（注1） -トレンド（注2） -四分位範囲基準化変化率（注3） を算出。 （この際、「四分位範囲（注4）」も算出。） -「外れ値」処理なし四分位範囲基準化変化率を用い、先行、一致、遅行指数毎に「共通循環変動」を算出。	―	―
（「系列固有変動」） $z_i^j(t)' = z_i^j(t) - ZC^j(t)$ （「共通循環変動」を除いた対称変化率） $r_i^j(t)' = z_i^j(t)'\times(Q3_i^j - Q1_i^j) + \mu_i^j(t)$・・② （「共通循環変動」を表す対称変化率） $r_i^j(t)^{共通} = ZC^j(t)\times(Q3_i^j - Q1_i^j)$ ・・③ $\left(\begin{array}{rl} & r_i^j(t) = z_i^j(t)\times(Q3_i^j - Q1_i^j) + \mu_i^j(t) \\ - & r_i^j(t)' = z_i^j(t)'\times(Q3_i^j - Q1_i^j) + \mu_i^j(t) \\ \hline & r_i^j(t) - r_i^j(t)' = (z_i^j(t) - z_i^j(t)')\times(Q3_i^j - Q1_i^j) \\ & r_i^j(t)^{共通} = ZC^j(t)'\times(Q3_i^j - Q1_i^j) \end{array}\right)$	**（２）「外れ値」処理** **（２－１）「外れ値」処理対象を選定** 「系列固有変動」を、「『外れ値』処理なし四分位範囲基準化変化率」が「共通循環変動」から乖離した部分として、差分を算出。 また、これを用い、「『共通循環変動』を除いた」対称変化率※を算出。<u>※対称変化率を「外れ値」処理対象とする。</u> ②式と変形した①式を用い③を算出。	個別系列データから -対称変化率 を算出。	（対称変化率） $r_i^j(t) = 200\times\dfrac{y_i^j(t)-y_i^j(t-1)}{y_i^j(t)+y_i^j(t-1)}$

<table>
<tr>
<td>（「外れ値」処理の実施）
$$\psi_1(r_i^j(t)')=\begin{cases}-k'(Q3_i^j{}'-Q1_i^j{}') : r_i^j(t)'<-k'(Q3_i^j{}'-Q1_i^j{}')\\ r_i^j(t)' : -k'(Q3_i^j{}'-Q1_i^j{}')\le r_i^j(t)'\le k'(Q3_i^j{}'-Q1_i^j{}')\\ k'(Q3_i^j{}'-Q1_i^j{}') : k'(Q3_i^j{}'-Q1_i^j{}')<r_i^j(t)'\end{cases}$$
$Q3_i^j{}'-Q1_i^j{}'$: $r_i^j(t)'$の四分位範囲(1985.01－2010.12)
k':1985.01－2010.12の間、一致指数の採用系列$r_i^C(t)'$の５％相当分を「外れ値」として算出する値。
$$\psi_2(r_i^j(t)')=\psi_1(r_i^j(t)')+r_i^j(t)^{共通}$$</td>
<td colspan="2">（２－２）「外れ値」処理の実施</td>
<td>（「外れ値」処理の実施）
$$\psi_2(r_i^j(t))=\begin{cases}-k(Q3_i^j-Q1_i^j) : r_i^j(t)<-k(Q3_i^j-Q1_i^j)\\ r_i^j(t) : -k(Q3_i^j-Q1_i^j)\le r_i^j(t)\le k(Q3_i^j-Q1_i^j)\\ k(Q3_i^j-Q1_i^j) : k(Q3_i^j-Q1_i^j)<r_i^j(t)\end{cases}$$
$Q3_i^j-Q1_i^j$: $r_i^j(t)$の四分位範囲(1980.01－2010.12)
k:1980.01－2010.12の間、一致指数の採用系列の$r_i^j(t)$の５％相当分を「外れ値」として算出する値。</td>
</tr>
<tr>
<td></td>
<td>閾値として定数ｋ’（注５）を与え、
$r_i^j(t)'$
の「外れ値」を処理する。

系列固有変動のみを「外れ値」処理した対称変化率を算出。</td>
<td>閾値として定数ｋを与え、
$r_i^j(t)$
の「外れ値」を処理する。</td>
<td></td>
</tr>
<tr>
<td>（個別系列のトレンド）
$$\mu_i^j(t)=\frac{\sum_{\tau=t-59}^{t-s}\psi_2(r_i^j(\tau)')}{60-s}$$</td>
<td colspan="2">旧手法に同じ
（３）個別系列のトレンド算出</td>
<td>（個別系列のトレンド）
$$\mu_i^j(t)=\frac{\sum_{\tau=t-59}^{t-s}\psi_2(r_i^j(\tau))}{60-s}$$</td>
</tr>
<tr>
<td></td>
<td>「外れ値」処理後
対称変化率を用い、
60か月平均をとることで、
個別系列のトレンド（注２）を算出。</td>
<td>「外れ値」処理後
対称変化率を用い、
60か月平均をとることで、
個別系列のトレンドを算出。</td>
<td></td>
</tr>
<tr>
<td>（個別系列の四分位範囲基準化変化率）
$$z_i^j(t)=\frac{\psi_2(r_i^j(t)')-\mu_i^j(t)}{Q3_i^j-Q1_i^j}$$</td>
<td colspan="2">（４）個別系列の四分位範囲基準化変化率の算出
旧手法に同じ</td>
<td>（個別系列の四分位範囲基準化変化率）
$$z_i^j(t)=\frac{\psi_2(r_i^j(t))-\mu_i^j(t)}{Q3_i^j-Q1_i^j}$$</td>
</tr>
<tr>
<td></td>
<td>（２－２）及び（３）から算出された
－対称変化率
－トレンド
－四分位範囲（注４）
を用い、
四分位範囲基準化変化率（注３）を算出。</td>
<td>（２－２）及び（３）から算出された
－対称変化率
－トレンド
－四分位範囲
を用い、
四分位範囲基準化変化率を算出。</td>
<td></td>
</tr>
<tr>
<td>（CI一致指数のトレンド）
$$\overline{\mu}^C(t)=\frac{1}{n^c}\times\sum_{i=1}^{n^c}\mu_i^c(t)$$</td>
<td colspan="2">旧手法に同じ
（５）CI一致指数のトレンドの算出</td>
<td>（CI一致指数のトレンド）
$$\overline{\mu}^C(t)=\frac{1}{n^c}\times\sum_{i=1}^{n^c}\mu_i^c(t)$$</td>
</tr>
<tr>
<td></td>
<td>「外れ値」処理後
－トレンド
を用い、CI一致指数のトレンドを算出。</td>
<td>「外れ値」処理後
－トレンド
を用い、CI一致指数のトレンドを算出。</td>
<td></td>
</tr>
<tr>
<td>（合成四分位範囲基準化変化率）
$$\overline{Z}^j(t)=\frac{1}{n^j-n_b^j(t)}\times\sum_{i\in N_F^j(t)}z_i^j(t)$$</td>
<td colspan="2">旧手法に同じ
（６）合成四分位範囲基準化変化率の算出</td>
<td>（合成四分位範囲基準化変化率）
$$\overline{Z}^j(t)=\frac{1}{n^j-n_b^j(t)}\times\sum_{i\in N_F^j(t)}z_i^j(t)$$</td>
</tr>
<tr>
<td></td>
<td>（４）で求めた四分位範囲基準化変化率を用い、合成四分位範囲基準化変化率（注６）を算出。</td>
<td>（４）で求めた四分位範囲基準化変化率を用い、合成四分位範囲基準化変化率を算出。</td>
<td></td>
</tr>
<tr>
<td>（合成四分位範囲）
$$\overline{Q3-Q1}^j=\frac{1}{n^j}\times\sum_{i=1}^{n^j}(Q3_i^j-Q1_i^j)$$</td>
<td colspan="2">旧手法に同じ
（７）合成四分位範囲の算出</td>
<td>（合成四分位範囲）
$$\overline{Q3-Q1}^j=\frac{1}{n^j}\times\sum_{i=1}^{n^j}(Q3_i^j-Q1_i^j)$$</td>
</tr>
<tr>
<td></td>
<td>（１）における対称変化率の四分位範囲（注４）を用い、合成四分位範囲を算出。</td>
<td>（１）における対称変化率の四分位範囲を用い、合成四分位範囲を算出。</td>
<td></td>
</tr>
<tr>
<td>（合成変化率）
$$V^j(t)=\overline{\mu}^c(t)+\overline{Q3-Q1}^j\times\overline{Z}^j(t)$$</td>
<td colspan="2">旧手法に同じ
（８）合成変化率の算出</td>
<td>（合成変化率）
$$V^j(t)=\overline{\mu}^c(t)+\overline{Q3-Q1}^j\times\overline{Z}^j(t)$$</td>
</tr>
<tr>
<td></td>
<td>（５）（６）（７）を用い、
合成変化率（注７）を算出。</td>
<td>（５）（６）（７）を用い、
合成変化率を算出。</td>
<td></td>
</tr>
<tr>
<td>（CI）
$$I^j(t)=I^j(t-1)\times\frac{200+V(t)}{200-V(t)}$$
$$CI^j(t)=\frac{I^j(t)}{I^j}\times 100$$</td>
<td colspan="2">旧手法に同じ
（９）CIの算出</td>
<td>（CI）
$$I^j(t)=I^j(t-1)\times\frac{200+V(t)}{200-V(t)}$$
$$CI^j(t)=\frac{I^j(t)}{I^j}\times 100$$</td>
</tr>
<tr>
<td></td>
<td>（８）を用い、CIを算出</td>
<td>（８）を用い、CIを算出</td>
<td></td>
</tr>
</table>

注１：個別系列が「0」または負の値を取る場合、またはその内容が比率になっている場合には、以下の通りとする。

$$r_i^j(t)=y_i^j(t)-y_i^j(t-1)$$

なお、景気動向と逆に動く（景気がよくなると値が小さくなる）逆サイクルの系列については、$r_i^j(t)$を求めた上で、符号を入れ替える。

j:指数を意味。CI先行指数ならば、$j=L$となる。一致指数はC、遅行指数はLag。
i:各指数における系列の番号。$i=1,\cdot\cdot\cdot,n_j$となる。
t:時点。

注２：四半期系列等、直近 s 期間においてデータが欠落している場合は、データを欠落させたまま平均値を算出する。このため、個別系列のトレンドは、全系列について昭和 60 年１月分以降の全期間について存在する。

注３：データが欠落している期間、その期間に対応する四分位範囲基準化変化率は存在しない。なお、（１）における四分位範囲基準化変化率は、昭和 55 年１月から存在する。（４）における四分位範囲基準化変化率は、昭和 60 年１月から存在する。

注４：四分位範囲は、昭和 55 年１月分から直近の 12 月分までの対称変化率（(１）において算出されたもの）のデータから計算されている。データ期間の更新頻度は年 1 回、3 月分速報時に行われ、その更新の幅は、直近１年間となる。系列によっては、季節調整替等により、データが遡及改訂されるがある。その場合は、3 月分速報時を待たずに、その都度更新される。

なお、（２－２）で用いている「四分位範囲」は、「外れ値」処理のためのものであり、それ以外の（１）（４）（７）において用いている四分位範囲とは異なることに留意。

$Q3_i^j : r_i^j(t)$の四分位範囲における第３四分位

$Q1_i^j : r_i^j(t)$の四分位範囲における第１四分位

$Q3_i^{j'} : r_i^j(t)'$の四分位範囲における第３四分位

$Q1_i^{j'} : r_i^j(t)'$の四分位範囲における第１四分位

注５：（２－２）で用いている閾値は、毎年３月分速報時に合わせ、昭和 60 年 1 月分から直近の 12 月までの一致系列の「系列固有変動」のデータから５％の「外れ値」を算出するような値を算出し、変更すべきか否かについて検討を行う。

注６：欠落項が存在する場合その欠落項に対応する基準化変化率は存在しないため、欠落していない系列のみで平均値を計算する。昭和 60 年１月から全期間において存在する。

$n_b^j(t)$：t時点の欠落項のある系列数
$N_F^j(t)$：t時点の欠落項のない系列の集合

注７：（８）で用いるトレンドは、CI 先行指数、CI 遅行指数であっても、（５）で算出した CI 一致指数のトレンドを用いることに留意。昭和 60 年１月から全期間において存在する。

注８：なお、上記手法は、統計学におけるウィンザー化平均に類似した手法となっている。