

4. 機械受注統計調査データを用いた分析

本節では、第2節で紹介した各手法を内閣府の「機械受注統計調査」のデータへ適用する。機械受注統計調査は、「機械等製造業者の受注した設備用機械類について、毎月の受注実績を調査したものであり、調査対象は主要機械等製造業者、調査時点は毎月月末である」(内閣府『機械受注統計調査報告』より)。1987年4月実績以降、調査対象企業数は280社となっている。主要な調査項目には、発注部門別、受注機種別の受注額がある。

2006年4月から2014年3月までの期間を分析対象期間とする。この期間の無回答発生状況は、無回答が生じた場合でも1~3社にとどまる月が多く、無回答率は概して低い。機械受注統計調査では、無回答への対応としてLOCFを用いている。

機械受注統計調査の調査項目である「受注額合計」を、興味の対象とする。当該統計調査はパネルデータであることから、前回調査以前のデータが補助変数として利用できる。前回調査の発注者の経済部門(「製造業」、「非製造業」、「官公需」、「外需」等)及び受注機種(「原動機」、「重電機」、「電子・通信機械」、「産業機械」等)、前回調査以前の「受注額合計」の調査客体ごと平均値及び変動係数を、前回調査以前のデータからの補助変数として用いる。このほか、調査客体企業の所在都道府県も補助変数として利用可能である。

LOCF、層化平均値代入法、層別合計伸び率による代入、回帰代入法、確率的回帰代入法、最近傍マッチング代入法、傾向スコアマッチング代入法、IPW法、多重代入法、及びHeckmanの選択モデルによる尤度法を機械受注統計調査に適用する。このうち「層別合計伸び率による代入」以外は、第2節で説明した。「層別合計伸び率による代入」は、まずデータを補助変数によって層化し、前月調査も受注額が観測されている調査客体に限って層ごとに前月及び当月調査の受注額総計を求める。そして両者の値から算出される層別受注額総計の伸び率を、各層に含まれる当月無回答企業の前月値に乗じた値を、当月受注額の代入値とする。これは、個別企業の受注額伸び率という変数について、層ごとの加重平均値を代入値としているので、層化平均値代入法の亜種である。回帰代入法、確率的回帰代入法、多重代入法、及び選択モデルによる尤度法については、受注額の対数値に処理を適用した。各手法の詳細は表4-0の通りである。

表4-0 機械受注統計調査に適用した各処理法の詳細

層化平均値代入法:

前回調査以前の「受注額合計」の調査客体ごと平均値及び変動係数のそれぞれについて4分位に分割し、合計16層に層化。各層で観測値の平均値を代入。

<p>層別合計伸び率による代入:</p> <p>前回調査以前の「受注額合計」の調査客体ごと平均値及び変動係数のそれぞれについて4分位に分割し、合計16層に層化。各層で前月及び当月で受注額を回答している調査客体にそれぞれの月の限り受注額総計を求め、その伸び率を当月無回答客体の前月値に乗じた値を代入値とする。</p>
<p>回帰代入法・確率的回帰代入法:</p> <p>前回調査の発注者の経済部門ダミー及び受注機種ダミー、前回調査以前の「受注額合計」の調査客体ごと平均値及び変動係数を補助変数とする回帰モデル。</p>
<p>最近傍マッチング代入法:</p> <p>前回調査の発注者の経済部門ダミー及び受注機種ダミー、前回調査以前の「受注額合計」の調査客体ごと平均値及び変動係数を補助変数とし、マハラノビス距離による1対1マッチング。</p>
<p>傾向スコアマッチング代入法:</p> <p>次式の2項ロジットモデルにより回答傾向スコアを推定し、その推定値に関して最近傍マッチング代入法を実行。</p> $\text{logit} \frac{\text{回答傾向スコア}_{it}}{1 - \text{回答傾向スコア}_{it}} = \beta' \begin{pmatrix} (\text{受注額 within 平均})_{i,t-1} \\ (\text{受注額 within 変動係数})_{i,t-1} \\ (\text{発注部門ダミー})_{i,t} \\ (\text{受注機種ダミー})_{i,t} \\ (\text{所在都道府県ダミー})_i \\ (\text{調査月ダミー})_t \end{pmatrix}$
<p>IPW 法:</p> <p>上記傾向スコアマッチング代入法における回答傾向スコアの逆数をウェイトとする。</p>
<p>多重代入法:</p> <ol style="list-style-type: none"> 1. 回帰モデル(受注額)$_{it} (\text{補助変数})_{it} \sim N(\beta'(\text{補助変数})_{it}, \sigma^2)$を完全ケース分析により推定し推定値$\hat{\beta}$及び$\hat{\sigma}^2$を得る。 2. 以下の乱数を得る。 $\hat{\sigma}^2 = \frac{\hat{\sigma}^2(\text{完全ケースの数} - \text{パラメータの数})}{\text{自由度}(\text{完全ケースの数} - \text{パラメータの数})}$ $\tilde{\sigma}^2 \sim \chi^2(\text{自由度}(\text{完全ケースの数} - \text{パラメータの数}))$ $\tilde{\beta} \tilde{\sigma}^2 \sim N(\hat{\beta}, \tilde{\sigma}^2(X_0'X_0)^{-1})$ <p>ただし、X_0は完全ケースの補助変数データ行列</p> 3. 以下の乱数を代入値とする。 $(\text{受注額})_{it} \sim N(X_M \tilde{\beta}, \tilde{\sigma}^2 I)$ <p>ただし、X_Mは欠測が生じたレコードの補助変数データ行列</p> 4. 2と3のステップを10回繰り返し、10個の疑似完全データを作成する。

$$\text{ただし、(補助変数)}_{it} = \begin{pmatrix} (\text{受注額 within 平均})_{i,t-1} \\ (\text{受注額 within 変動係数})_{i,t-1} \\ (\text{発注部門ダミー})_{i,t} \\ (\text{受注機種ダミー})_{i,t} \\ (\text{所在都道府県ダミー})_i \end{pmatrix}$$

選択モデルによる尤度法:

下記の選択モデルを最尤推定し、調査時点毎に受注額 between 平均の推定値 $\hat{\mu}_t$ に調査対象企業数を乗じた値を受注額総計の推定値 $\hat{\tau}_t$ とする。

$$\left\{ \begin{array}{l} (\text{受注額})_{it} = (\text{受注額 between 平均}\mu)_t + (\text{誤差項}\varepsilon)_{it} \\ (\text{観測指標})_{it} = 1 \left[\gamma' \begin{pmatrix} (\text{受注額 within 平均})_{i,t-1} \\ (\text{受注額 within 変動係数})_{i,t-1} \\ (\text{発注部門ダミー})_{i,t} \\ (\text{受注機種ダミー})_{i,t} \\ (\text{所在都道府県ダミー})_i \end{pmatrix} + (\text{誤差項}\eta)_{it} \right] \\ (\varepsilon, \eta) \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1 \\ \rho\sigma_1 & 1 \end{pmatrix} \right) \end{array} \right.$$

手法ごとの推定結果を図4 - 1に示す。手法ごとに調査客体約 280 社の受注額総計の推定値が得られるが、実際には無回答率が小さい(0~3%)ことから、手法間で推定値に大きな差は出ない。そこで、手法ごとの推定値の代わりに、図4 - 1では、手法ごとの「補完率」を示す。補完率は次式で定義される。

$$\text{補完率} = \frac{\text{推定値} - \text{観測値総計}}{\text{観測値総計}}$$

図4 - 1上段左端のグラフはLOCF、層化平均値代入法、及び層別合計伸び率による代入について、同上段中央のグラフは回帰代入法、確率的回帰代入法、最近傍マッチング代入法、及び傾向スコアマッチング代入法について、同上段右端のグラフはIPW法、及び多重代入法について、同下段のグラフは選択モデルについて、補完率の時系列を示している。すべての手法の結果を1つのグラフに表示すると、判別できなくなるので、このように分けて示している。補完率の値が0の月は、全ての調査対象企業が回答した月である。選択モデル以外の手法では、多くの月で補完率が0.1%未満であり、高い月でも1~数%である点が共通している。

次に、機械受注統計調査の観測データを完全データとみなして、そのデータに欠

測を確率的に発生させることで、上記各手法のパフォーマンスを比較するシミュレーションを行う。 欠測を発生させるモデルは、MNAR と MAR の2通りを用いる。MNAR のモデルとしては、企業*i*が回答する確率を企業*i*の受注額と企業*i*の所在地に依存させ、MAR のモデルとしては、企業*i*が回答する確率を企業*i*の所在地のみに依存させる。また、それぞれの欠測データメカニズムに対して、無回答企業の割合の期待値(以下「欠測率」)が10%の場合と20%の場合を考える。

欠測データメカニズムと欠測率の組合せ各々に対して、繰り返し互いに独立に不完全データを生成させ、それぞれの不完全データに各手法を適用する。繰り返しの試行回数は100とした。各手法について、試行ごとの推定値及び真の値(完全データとみなしている観測データの受注額総計)から RRMSE(相対平方平均自乗誤差)を算出する。RRMSEは次式で定義される。

$$RRMSE \equiv \frac{\sqrt{\frac{1}{100} \sum_{k=1}^{100} (\text{第}k\text{試行の推定値} - \text{真値})^2}}{\text{真値}}$$

RRMSEは、推定バイアス及び推定量の分散の増加関数である(MSE(平均自乗誤差)は推定バイアスの自乗と推定量の分散の和に等しい)ため、RRMSEの値が小さい手法ほどパフォーマンスがよいといえる。 RRMSEでパフォーマンスを測る場合、確率的回帰代入法は回帰代入法に必ず劣後することが分かっているので、確率的回帰代入法を試す必要はないが、参考として含めた。

シミュレーションの結果を、図4-2-1~4に示す。LOCFと他の手法の比較を容易にするため、LOCFと比較対象となる他の手法のRRMSEを1つのグラフに重ねて表示している。LOCFのRRMSEは、2008年後半から2009年にかけて高まっている。これは、リーマンショックのようなマクロショックに対してLOCFがぜい弱なためである。それでも、LOCFは、層別合計伸び率による代入、回帰代入法、及び最近傍マッチング代入法とともに最もパフォーマンスの良い手法に属している。

図4 - 1 推定結果

縦軸は補完率 $[= (\text{推定値} - \text{観測値総計}) \div \text{観測値総計}]$

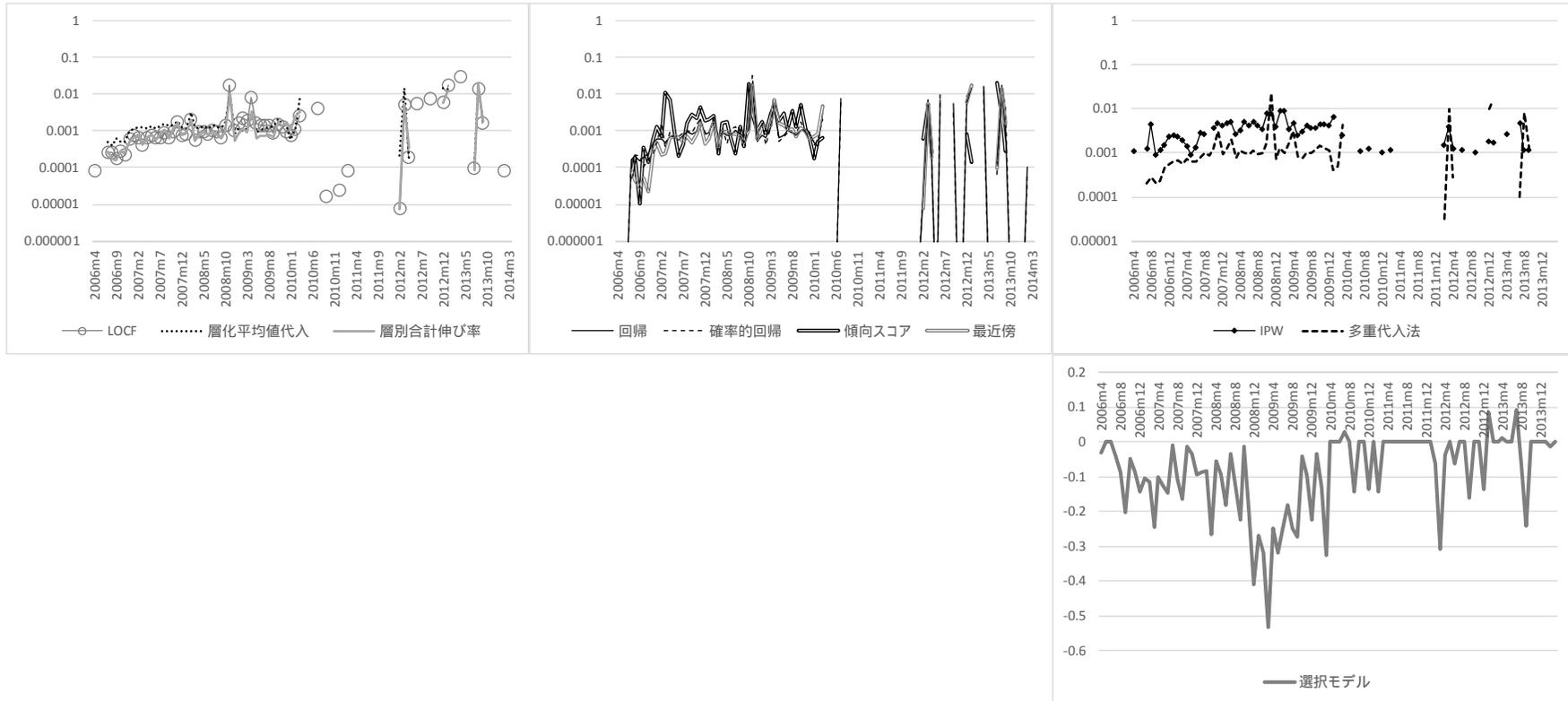


図4 - 2 - 1 シミュレーション結果(MNAR、欠測率 10%)：縦軸は RRMSE

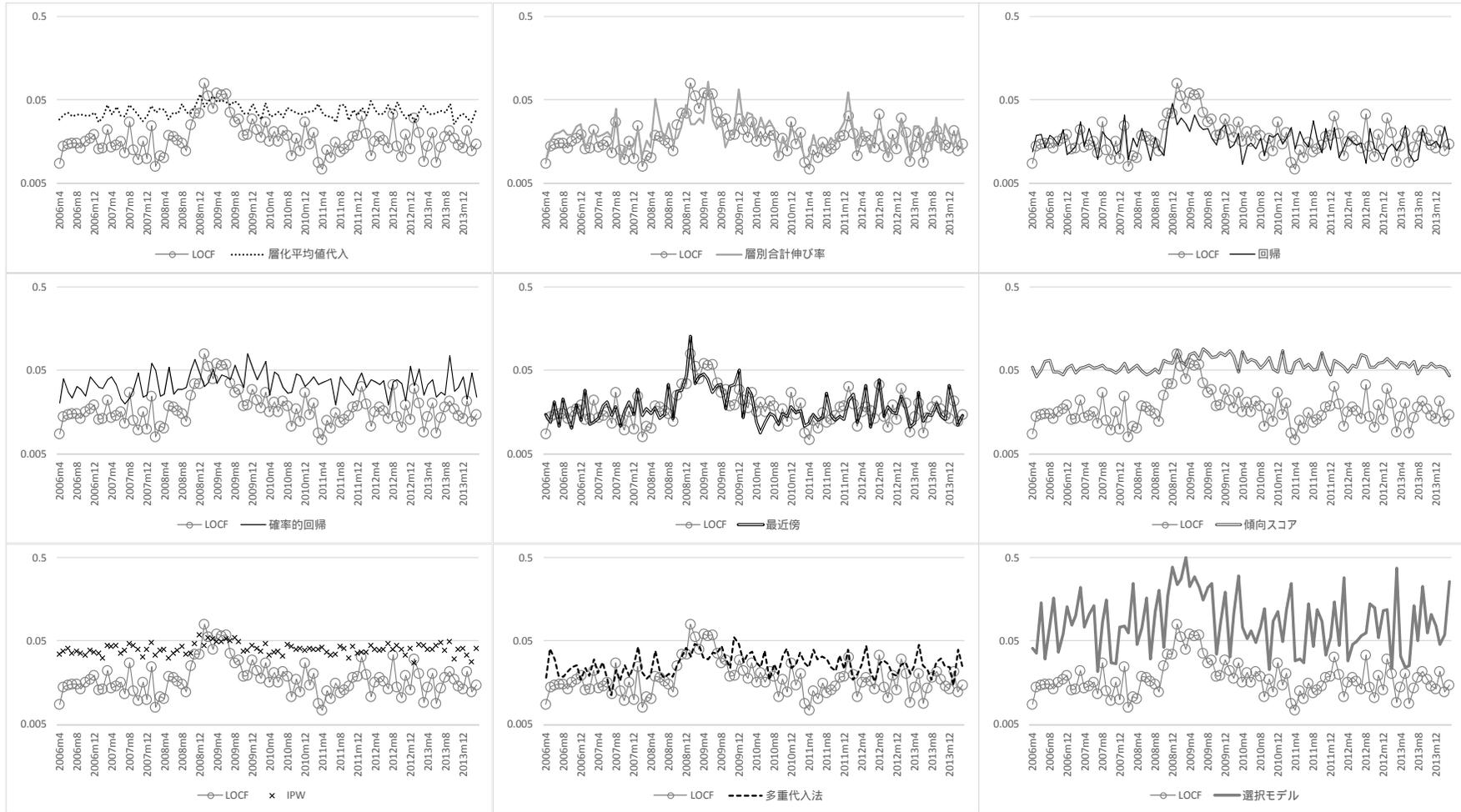


図4 - 2 - 2 シミュレーション結果(MNAR、欠測率 20%)：縦軸は RRMSE

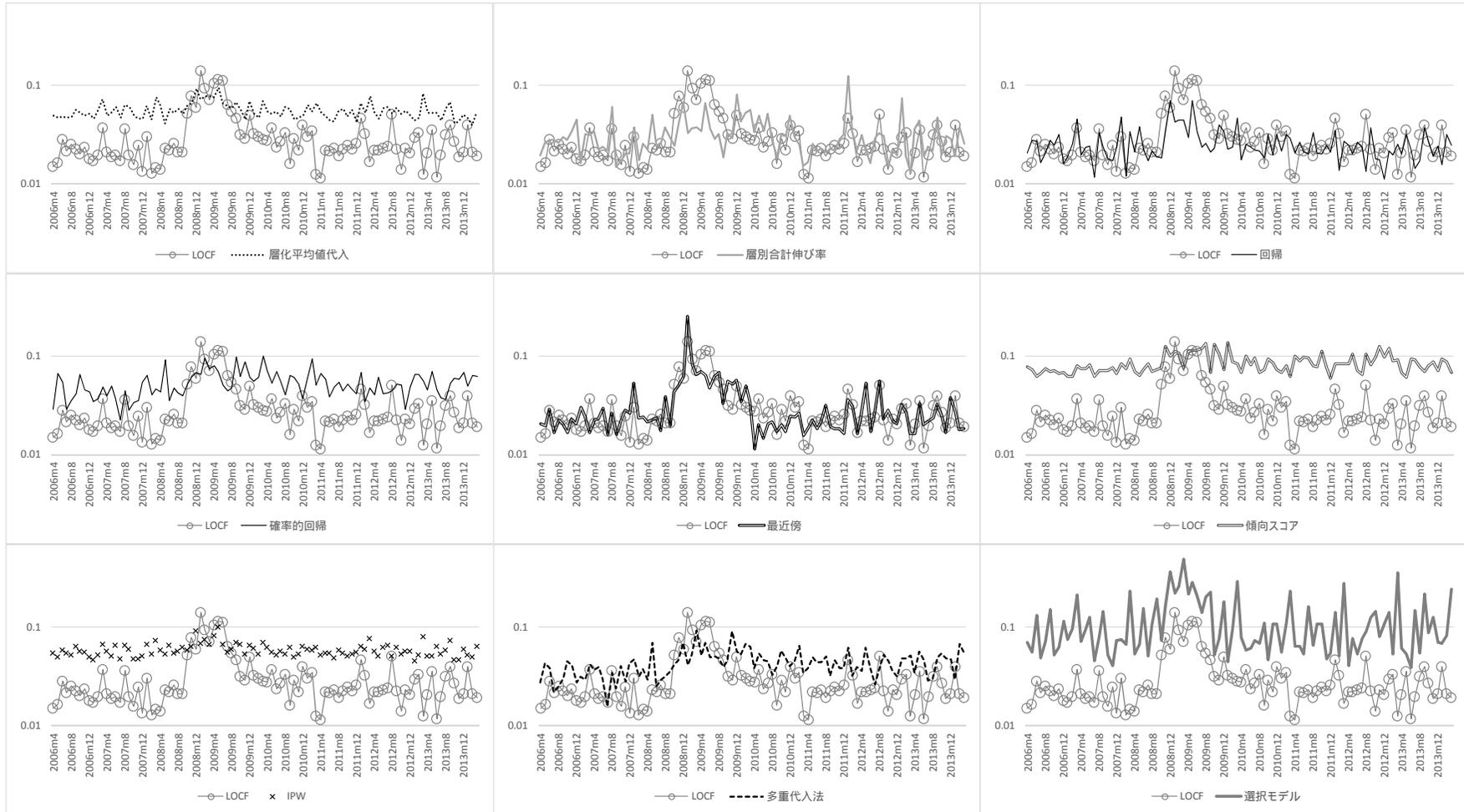


図4 - 2 - 3 シミュレーション結果(MAR、欠測率 10%)：縦軸はRRMSE



図4 - 2 - 4 シミュレーション結果(MAR、欠測率 20%)：縦軸は RRMSE

