

予測方法の評価、選択について

講師: 長倉大輔

† この資料は私のゼミにおいて使用するために作成した資料です。ゼミのWEBページ上で公開しており、自由に参照して頂いて構いません。ただし、内容について、一応検証してありますが、もし間違いがあった場合でもそれによって生じるいかなる損害、不利益について責任を負いかねますのでご了承ください。間違いは発見次第、継続的に直していますが、まだ存在する可能性があります。間違いを見つけた場合 nagakura@z7.keio.jp までご連絡いただくと大変嬉しく思います。

予測法、その評価、比較について

本日は、予測および予測の評価、比較に用いられる統計的手法についていくつか紹介する。

予測精度の改善を目的とした**予測平均**という手法を紹介する。また予測評価の際、その評価が統計的に有意かどうかの検定として代表的な検定として **Diebold – Mariano (DM) 検定**を紹介する。

今日の予定

1. 予測 – サンプル内予測とサンプル外予測
2. 予測評価の指標
3. 予測評価の有意性 – Diebold Mariano 検定
4. 予測平均について

予測法、その評価、比較について

■ 予測について

ここでは**予測**の定義として、「現在までに観測された変数を用いて将来の値を推定すること」とする。

ファイナンス変数やマクロ経済変数の精度のよい予測は、リスク管理から政策決定に至るまで様々なレベルで重要である。

予測法、その評価、比較について

■ 予測

過去の観測されたベクトル \mathbf{x}_t の値を用いて、スカラー変数 y_{t+1} の値を**予測するモデル**を

$$y_{t+1} = g(\mathbf{x}_t, \varepsilon_{t+1}; \theta)$$

とする。ここで g はある関数、 ε_{t+1} はモデルに含まれる攪乱項、 θ はモデルに含まれるパラメーターベクトルを表す。

\mathbf{x}_t には y_t の過去の値を含んでもよい。

予測法、その評価、比較について

例として、例えば関数 g が線形であれば、予測モデルは線形回帰モデル

$$y_{t+1} = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_K x_{K,t} + \varepsilon_{t+1}$$

になる。ここで $\mathbf{x}_t = [x_{1,t}, \dots, x_{K,t}]^T$, $\boldsymbol{\theta} = [\beta_0, \dots, \beta_K]^T$ である。またさらに、 \mathbf{x}_t が過去の p 個の値のみを含んでいるとして、AR(p) モデル

$$y_{t+1} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-2} + \dots + \beta_p y_{t+1-p} + \varepsilon_{t+1}$$

$\mathbf{x}_t = [y_t, \dots, y_{t+1-p}]^T$, $\boldsymbol{\theta} = [\beta_0, \dots, \beta_K]^T$ などもこの定式化は含んでいる。

予測法、その評価、比較について

さらに予測モデルとしては、 \mathbf{x}_t として y_t の過去の値と他の変数の過去の値を含むような VAR モデル

$$\begin{aligned} y_{t+1} = & \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \dots + \beta_p y_{t+1-p} \\ & + \alpha_1 z_t + \alpha_1 z_{t-1} + \dots + \alpha_p z_{t+1-p} \\ & + \dots + \varepsilon_{t+1} \end{aligned}$$

も予測モデルの一例である。

予測法、その評価、比較について

また g の関数形が非線形なものとしては非線形回帰モデル

$$y_{t+1} = g(\mathbf{x}_t; \boldsymbol{\theta}) + \varepsilon_{t+1}$$

や非線形 AR(p) モデル、

$$y_{t+1} = g(y_t, y_{t-1}, \dots, y_{t+1-p}, \varepsilon_{t+1}; \boldsymbol{\theta})$$

など、様々なものが考えられる。

予測法、その評価、比較について

これらの予測モデルを基に、例えば線形モデルの場合、 \mathbf{x}_t の条件付き期待値 ($E(\varepsilon_{t+1}|\mathbf{x}_t) = 0$ として)

$$E[y_{t+1} | \mathbf{x}_t] = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_K x_{K,t}$$

を予測値として計算する。

この場合、一般的には

$$E[y_{t+1} | \mathbf{x}_t] = E[g(\mathbf{x}_t, \varepsilon_{t+1}; \boldsymbol{\theta}) | \mathbf{x}_t] = f(\mathbf{x}_t; \boldsymbol{\theta})$$

のように \mathbf{x}_t についての関数形は $g(\cdot)$ とは異なる関数形になる。

予測法、その評価、比較について

また例えば真のモデルが線形でなくても**予測モデルとして線形モデルを用いる**という事も考えられる。

以後は y_{t+1} の予測値を \hat{y}_{t+1} と表記する。予測値は過去の値の関数になる、つまり

$$\hat{y}_{t+1} = f(\mathbf{x}_t; \boldsymbol{\theta}_t)$$

となる。

予測法、その評価、比較について

t 時点までのデータに基づいて $t + h$ 時点の値を予測する事を **h 期先予測**という。

以後、 y_t の t 時点における h 期先予測を $\hat{y}_{t+h|t}$ と表記し、

$$\hat{y}_{t+h|t} = f^{(h)}(\mathbf{x}_t; \boldsymbol{\theta})$$

とする。ここで $f^{(h)}(\cdot)$ は h 期先予測の時の予測値を計算する関数である(h が異なれば異なりうる事に注意)。

予測法、その評価、比較について

- サンプル内予測 (In-sample-forecast)とサンプル外予測 (Out-of-sample forecast)

予測を評価する際に、**サンプル内予測**による評価と**サンプル外予測**による評価がある。

サンプル内予測とサンプル外予測の違いは主に**未知パラメーター θ** を推定するのに用いる**データの違い**である

以下では1期先予測の場合を考える (h 期先予測でも考え方は同じ)。

予測法、その評価、比較について

$\{y_{t+1}, \mathbf{x}_t\}_{t=1}^T$ が観測されたとする。

サンプル内予測ではまず未知パラメータ θ を**観測されたデータ全てを用いて推定する**。 T 時点までの全てのデータを用いて推定した θ を $\hat{\theta}_T$ とする。

サンプル内予測ではこの $\hat{\theta}_T$ を用いて予測値を

$$\hat{y}_t = f(\mathbf{x}_{t-1}; \hat{\theta}_T)$$

$t = j, \dots, T$ と計算していく。ここで j は予測値を計算できる最初の時点(例えばAR(p)モデルであれば $j = p + 1$)である。サンプル内予測による評価ではこの予測値を用いて評価をしていく。

予測法、その評価、比較について

サンプル外予測のやり方は実はいくつかあるが、それらに共通するのは**パラメーターの推定に用いるデータと予測の評価に用いるデータを分ける**ということである。

まず1つ目を説明する。

- サンプル外予測 1

S 時点までのデータ $\{y_{t+1}, \mathbf{x}_t\}_{t=1}^S$ を用いて θ を推定する。この推定値を $\hat{\theta}_S$ とする。この $\hat{\theta}_S$ を用いて次の期の予測値を

$$\hat{y}_{S+2} = f(\mathbf{x}_{S+1}; \hat{\theta}_S)$$

と計算する。

予測法、その評価、比較について

以降の予測では常に $\hat{\theta}_S$ を用いる。即ち

$$\hat{y}_{S+3} = f(\mathbf{x}_{S+2}; \hat{\theta}_S)$$

$$\hat{y}_{S+4} = f(\mathbf{x}_{S+3}; \hat{\theta}_S)$$

$$\hat{y}_{S+5} = f(\mathbf{x}_{S+4}; \hat{\theta}_S)$$

⋮

$$\hat{y}_{T+1} = f(\mathbf{x}_T; \hat{\theta}_S)$$

のように予測値を計算していく。要は **S 時点までのデータのみから推定したパラメータ値を用いて予測値を計算する**ということ。

予測法、その評価、比較について

- サンプル外予測 2

2つめとしては、推定値を逐次更新していくやり方がある。即ち

$$\hat{y}_{S+3} = f(\mathbf{x}_{S+2}; \hat{\boldsymbol{\theta}}_{S+1})$$

$$\hat{y}_{S+4} = f(\mathbf{x}_{S+3}; \hat{\boldsymbol{\theta}}_{S+2})$$

⋮

$$\hat{y}_{T+1} = f(\mathbf{x}_T; \hat{\boldsymbol{\theta}}_{T-1})$$

のように予測値を計算していく。予測をする1期前までのデータ全てを使って推定したパラメータ値を用いて予測値を計算するということ。

予測法、その評価、比較について

- サンプル外予測 3

3つめとして**ローリングウィンドウ**によって予測値を計算する方法もよく用いられる。

これは1番目と2番目のやり方を合わせたようなもので未知パラメーターを推定するデータの数を固定し(これを**ウィンドウサイズ**という)、かつ更新もしていく。

$\hat{\theta}_{k,n}$ を k 時点から n 時点までのデータを用いて推定した未知パラメーターの推定値とする。

予測法、その評価、比較について

この時、ウィンドウサイズ S のローリングウィンドウによるサンプル外予測値は

$$\hat{y}_{S+2} = f(\mathbf{x}_{S+1}; \hat{\boldsymbol{\theta}}_{1,S})$$

$$\hat{y}_{S+3} = f(\mathbf{x}_{S+2}; \hat{\boldsymbol{\theta}}_{2,S+1})$$

⋮

$$\hat{y}_{T+1} = f(\mathbf{x}_T; \hat{\boldsymbol{\theta}}_{T-S,T-1})$$

のように計算される。**パラメータの推定に直前の S 個のデータのみ使用すること。**

予測法、その評価、比較について

■ 予測誤差

予測誤差は実際の値と予測した値の差として定義される。すなわち 時点 t の予測誤差 e_t は

$$e_t = \hat{y}_t - y_t$$

と定義される。ここで \hat{y}_t はサンプル内予測もしくはサンプル外予測によって計算された予測値である。

予測法、その評価、比較について

競合するいくつかのモデルがある場合は、それらのモデルについて全て予測誤差を計算する。

それぞれのモデルからの予測誤差に対して、何らかの基準を用いてどちらの予測が優れているかを比較する必要がある。

この基準は分析する目的によって異なる。

以下でよく用いられる基準をいくつか紹介する

予測法、その評価、比較について

■ いくつかの予測評価の基準

予測の良さを評価する基準として代表的なものをいくつか挙げる。以下では $\{e_t\}_{t=j}^T$ という $n = T - j + 1$ 個の**予測誤差**が得られているとする。

□ 平均二乗誤差 (Mean Squared Error; RMSE)

$$MSE = n^{-1} \sum_{t=j}^T e_t^2 ,$$

□ 平均二乗平方根誤差 (Root Mean Squared Error; RMSE)

$$RMSE = \sqrt{n^{-1} \sum_{t=j}^T e_t^2}$$

予測法、その評価、比較について

- 平均絶対誤差 (Mean Absolute Error; MAE)

$$MAE = n^{-1} \sum_{t=j}^T |e_t|$$

- 平均絶対パーセント誤差 (Mean Absolute Percentage Error; MAPE)

$$MAPE = n^{-1} \sum_{t=j}^T \frac{|e_t|}{Y_t}$$

予測法、その評価、比較について

- Theil 不等式係数 (Theil Inequality Coefficient)

$$U = \frac{\sqrt{\frac{1}{n} \sum_{t=j}^T e_t^2}}{\sqrt{\frac{1}{n} \sum_{t=j}^T \hat{y}_t^2} + \sqrt{\frac{1}{n} \sum_{t=j}^T y_t^2}}$$

これは実際には後述の理由で**近年ではまず使われない**。

予測法、その評価、比較について

これらの基準はどれも小さければ小さいほどよいと解釈される。

これらの基準のうち MSE や RMSE、MAE、MAPE については直観的に非常に理解しやすい形をしているのでその意味について説明の必要はあまりないと思われる。

しかしながら U についてはその性質について若干の説明がいると思われるので以下に説明する。

予測法、その評価、比較について

まず、三角不等式によって常に

$$\sqrt{n^{-1} \sum_{t=j}^T (\hat{y}_t - y_t)^2} \leq \sqrt{n^{-1} \sum_{t=j}^T \hat{y}_t^2} + \sqrt{n^{-1} \sum_{t=j}^T y_t^2}$$

であるので、 U は 1 以上の値を取らない事がわかる。さらに全ての t について $\hat{y}_t = y_t$ であれば分子が 0 になるので $U = 0$ となる。これは予測値と実際に値が完全に一致するという事なので $U = 0$ となる場合が一番良い予測ということがわかる。

予測法、その評価、比較について

これに対して $U = 1$ となるのは、

- (1) 全ての t について $y_t = 0$ の場合
- (2) 全ての t について $\hat{y}_t = 0$ の場合
- (3) 全ての t について $\hat{y}_t = -y_t$ の場合

の3つがある。

予測法、その評価、比較について

U が 0 に近いほどよい予測という解釈であれば、上限の $U = 1$ は予測が最悪の場合を示していると考えられるが、(1) は現実のデータではありえないし、(2) はモデルによってはこれが最適な予測になりえる。(3) も場合によっては最悪の予測というわけではない。

よって U は 1 に近いからと言って必ずしも悪い予測なのかどうかの判断ができないため、予測の評価の基準としてはあまり有用ではなく、実際に用いられることは近年ではまずない。

予測法、その評価、比較について

これらの基準を用いて、競合する M 個の予測モデルの予測誤差

$$\{e_t^{(1)}\}_{t=j}^T, \{e_t^{(2)}\}_{t=j}^T, \dots, \{e_t^{(M)}\}_{t=j}^T$$

が与えられた時に(ここで $\{e_t^{(m)}\}_{t=j}^T$ は、 m 番目のモデルの予測誤差)、これらの予測誤差の RMSE、MAE、MAPE などを計算しそれらのうちで一番小さいものが良い予測モデルだと考える事ができる。

予測法、その評価、比較について

■ 予測評価の有意性

しかしながら、これらの基準による評価はあくまでも**観測されたデータに基づいたものである事に注意する必要がある。**

これらの基準は母集団のある値を推定していると考えられるが、確率的な変動(データの偏り)によって、予測誤差も変動するため、これらの基準による評価も、標本の偏りによって実際の予測力をきちんと反映していない可能性がある。****

予測法、その評価、比較について

例えば i 番目の予測モデルの RMSE (これは $\sqrt{E(e_t^{(i)2})}$ の推定値と考えられる) が j 番目の予測モデルの RMSE よりも小さい場合に、それが単に標本の偏りによって生じたのかそれとも実際に $\sqrt{E(e_t^{(i)2})} < \sqrt{E(e_t^{(j)2})}$ という事実を反映しての結果なのかには注意が必要である。

この場合、実際にある予測モデルが他の予測モデルよりも予測力が高いかどうかを統計的に**検定**する必要がある。

予測法、その評価、比較について

このような問題意識から提案されたものがこれから説明する **Diebold – Mariano 検定** である。以下 DM 検定と呼ぶ。

DM 検定においては **2つの予測値の予測力が等しいというのを帰無仮説としている**。この帰無仮説が棄却されるのであれば、片方の予測値の予測力はもう方の方予測力よりも有意に高いという事ができる。

予測法、その評価、比較について

■ DM検定

2つの予測誤差の系列 $\{e'_t\}_{t=j}^T$ と $\{e_t\}_{t=j}^T$ が与えられているとしよう。

損失関数 を $L(e)$ とし **損失差** を

$$d_t = L(e_t) - L(e'_t)$$

とする。例えば $L(e) = e^2$ や $L(e) = |e|$ など。前者がMSE、後者はMAEに対応する。

予測法、その評価、比較について

この損失差の期待値が0である時、2つの予測値の予測力は同じと考える(どちらかの損失が平均的に大きい場合は損失差の期待値は0ではないので)

つまり、検定したい帰無仮説と対立仮説は

$$H_0: E(d_t) = 0 \quad \text{および} \quad H_1: E(d_t) \neq 0$$

となる。

予測法、その評価、比較について

この帰無仮説に対して、Diebold and Mariano (1995) は以下の検定統計量を提案した。

$$DM_T = \frac{\bar{d}_T}{se(\bar{d}_T)}$$

ここで $\bar{d}_T = (T - j - 1)^{-1} \sum_{t=j}^T d_t$ と定義され、 $sd(\bar{d}_T)$ は \bar{d} の標準誤差である。

予測法、その評価、比較について

Diebold and Mariano (1995) は d_t について緩い仮定のもとで DM_T は帰無仮説の下で、漸近的に

$$DM_T \sim N(0, 1)$$

となる事を示した。Diebold and Mariano (1995) は $se(\bar{d}_T)$ に HAC (Heteroskedasticity and autocorrelation consistent) 推定量を用いる事を提案している

DM_T は非常に簡単に計算できるので、広く用いられるようになった。また有限標本における DM 検定のパフォーマンスの改善のためいくつか改良がなされている (Harvey, Leybourne, Newbold (1997) などがよく知られている。

予測法、その評価、比較について

■ 予測平均(Forecasting Average)

しばしば、ある1つの変数 y_t に対して、**様々な予測**(異なるモデルを用いたり、異なったデータ、情報を用いたりした予測) が利用可能な場合がある。このような場合、従来のアプローチはその予測の中から(何らかの基準で)最良の予測を選び、それを用いて予測を行うというものであった。

それに対して、**予測平均**(Forecast Averaging や Forecasting Combination と呼ばれる)とは利用可能な予測を(有効に使用できるもの)**全てを組み合わせ**て**予測を行うというアプローチ**であり、予測を組み合わせることにより最良の予測を単独で使用する場合よりも予測精度がよくなることが実証的に知られている。

予測法、その評価、比較について

y_t の i 番目の予測(値)を $\hat{y}_t^{(i)}$ とし、 N 個の予測、 $i=1, \dots, N$ が使用可能であるとする。この時、予測平均は通常これらの予測の加重平均

$$\hat{y}_t^c = \sum_{i=1}^N \omega_{i,t} \hat{y}_t^{(i)}$$

を最終的な予測とすることである。

より一般的には、必ずしも加重平均でない N 個の予測の関数が考えられるが、現時点では実際の応用上はほとんど上記の加重平均の形しか用いられていない。

予測法、その評価、比較について

加重平均の重み $\omega_{i,t}$ に下付き文字 t がついていることに注意。これは**それぞれの予測に掛かる重みを予測点に依存して変更すること**を許容するということである。

以下の例のいくつかでは実際に予測点に応じて重みが変わるものがある。

予測法、その評価、比較について

■ 予測平均の例 (8つ)

1. 単純平均 (Simple Mean)

N 個の予測の平均をとったもの。

2. 調整平均 (Trimmed Mean)

予測値を大きい順に並べて上下何%かを除いた残りの予測で平均を取ったもの(異常値に頑強)。

3. 中央値

予測の**中央値**をとったもの。

予測法、その評価、比較について

4. 最小二乗加重 (Least Square Weights)

N 個の予測 $\hat{y}_t^{(i)}$, $i=1, \dots, N$ の実際の値への線形回帰の係数を最小二乗法で推定し、その推定した係数を重みとして使用する。最小二乗法で推定できるだけの実際の値と予測(値)が必要。

5. 平均二乗誤差加重 (Mean Square Error Weights)

それぞれの予測のMSEを計算して、 ω_i を

$$\omega_i = \frac{1 / MSE_i^k}{\sum_{j=1}^N 1 / MSE_j^k}$$

によって決定する。ここで MSE_i は i 番目の予測のMSE、 k は分析者が決定する調整パラメーター。

予測法、その評価、比較について

6. 最小二乗誤差順位 (MSE Ranks)

先ほど同様 MSE を計算し、小さい方から順に順位をつける(MSEが小さいほど順位が小さい)。 r_i を i 番目の予測の順位だとするとその加重を

$$\omega_i = \frac{1/r_i}{\sum_{j=1}^N 1/j}$$

とする(順位の逆数の比)。

予測法、その評価、比較について

7. 平滑化AIC加重 (Smoothed AIC Weights)

それぞれの予測に用いられたモデルのAICを計算し、 i 番目の予測の重みを

$$\omega_i = \frac{\exp(-0.5AIC_i)}{\sum_{j=1}^N \exp(-0.5AIC_j)}$$

によって計算する。ここで AIC_j は j 番目の予測に用いられたモデルのAICの値。全てのモデルのAICが計算できることが前提となっている。

予測法、その評価、比較について

8. 近似ベイズモデル平均加重 (Approximate Bayesian Model Averaging)

それぞれの予測に用いられたモデルのBICを計算し、 i 番目の予測の重みを

$$\omega_i = \frac{\exp(-0.5BIC_i)}{\sum_{j=1}^N \exp(-0.5BIC_j)}$$

によって計算する。ここで BIC_j は j 番目の予測に用いられたモデルのBICの値。全てのモデルのBICが計算できることが前提となっている。

予測法、その評価、比較について

これら以外にも様々な加重法が提唱されているが、現状ではそれらの理論的な性質についての研究はあまり進んでいないようである(シミュレーションによる比較だけで終わっている)。

おそらくそれらの性質は真のデータ育成過程や予測モデルの推定法、予測の損失関数の形状などにも依存するであろうから、どんな状況でも一番よい方法というのは存在しないかもしれない。理論的な発展が待たれるところである。

本日のまとめ

1. よく使われる予測の評価基準の紹介
2. ある予測が他の予測より良いかは統計的に検定する必要があるが、そのための統計量としてDM 検定統計量を説明。
3. 予測精度の改善のため、予測を組み合わせる予測平均という方法の紹介。