Econometria e Séries Temporais
- Aula 17 -

Prof. Mestre. Omar Barroso

Instituto Brasileiro de Educação, Pesquisa e Desenvolvimento

Métodos de Avaliação de Previsão: Funções de Perda

- Uma função de perda pode ser definida: “como um método formal para **negociar** potenciais erros de previsão de diferentes sinais e magnitudes” (Elliot e Timmermann (2006), pp. 13).
- A função de perda quantifica o custo de usar uma previsão imperfeita, $\hat{f}$, dada a relação das variáveis e possivelmente várias outras variáveis de controle.
- Para definir uma função de perda, precisamos considerar as compensações entre diferentes erros de previsão
- As funções de perda influenciam quais modelos serão selecionados e como esses modelos serão estimados
- No trabalho empírico, geralmente selecionamos funções de perda de um conjunto de [funções] predefinidas.

Funções de Perda

- Formalmente, uma função de perda L é definida como:
- L : YxYxZ $\rightarrow l$
- Onde $l \in$ R e Y é o conjunto de todos os resultados possíveis de Y RV e f.
- Z é o conjunto de informações que contém todas as variáveis condicionantes.
- Assumimos que a função de perda está tomando seu valor mínimo para a previsão que é igual à realização.
- Normalmente usamos uma função de perda que tem uma forma razoável. Tais formas definem as propriedades da função de perda.

Funções de Perda

- A perda não é decrescente (ou não crescente) à medida que o erro de previsão aumenta (ou diminui).
- As funções de perda não são negativas.
- As funções de perda são zero para erro de previsão zero.
- Propriedades adicionais incluem homogeneidade, simetria, diferenciabilidade e limitação.

Funções de Perda

- A perda de erro quadrado é definida como:
- $L(y - f) = a(y - f)2, a > 0$
- A função de perda satisfaz as três propriedades mencionadas acima.
- É simétrica, homogênea e diferençável.
- Não é limitada da forma acima. Portanto, erros maiores terão um custo maior atribuído

Perda Linear Por Partes (PLPP)

- PLPP é definida como:
- onde $0 < \alpha < 1$
- A função de perda satisfaz as três propriedades mencionadas acima.
- Não é simétrica, homogênea e diferenciável (em todos os lugares, exceto zero).
- Perda de erro absoluto é um caso especial da função de perda linear por partes.

$$L(y-f) = \begin{cases} -a(1-\alpha)(y-f), & \text{if } y-f \leqslant, a>0 \\ a(y-f)\alpha, & \text{if } y-f>0, a>0 \end{cases}$$

Perda Assimétrica Por Partes (PAPP)

- n é geralmente definido como 2 e exp1 = 0
- Usamos L1, L2 que satisfazem as propriedades mencionadas acima.
- Para L1 = (1 − a)|y − f |p, L2 = (a)|y − f |p com p ∈ N e 0 < a < 1 definimos uma família de funções que abrangem as mencionadas acima.
- Perda de erro absoluto é um caso especial.

$$L(y-f) = \begin{cases} L_1(y-f) & \text{if } y-f \leqslant \exp_1 \\ L_2(y-f), & \text{if } \exp_1 < y-f \leqslant \exp_2 \\ ...., & .... \\ L_n(y-f), & \text{if } \exp_{n-1} < y-f \leqslant \exp_n \end{cases}$$

Teste de DM (Diebold e Mariano, 1995)

- Como podemos usar a função de perda para avaliar previsões?
- Em 1995, Diebold e Mariano propuseram um teste de precisão preditiva.
- O teste foi proposto como uma resposta a inúmeras comparações de previsões que falharam em considerar a incerteza da amostragem.
- O teste é simples e depende de duas suposições principais...

Teste de DM (Diebold e Mariano, 1995)

- Por exemplo, o seu cliente é solicitado a tomar uma decisão com base em dois previsores, A e B, que fornecem o seguinte {fA} e {fB } para uma medida de interesse (taxa de juros, inflação, títulos do tesouro, etc...).
- Observe que não há menção aos modelos usados para criar as previsões.
- Assumindo que a função de perda quadrada, o caminho a seguir é avaliar as perdas para todas as previsões e então calcular a média.
- Assumindo que os erros quadráticos médios da amostra para A são 1,9 e para B são 2,1. Dado esse resultado, A vence
- Podemos concluir que A fornece previsões melhores?

Teste de DM (Diebold e Mariano, 1995)

- O que sabemos é que para esta amostra em particular MSE(A) < MSE(B)
- Não temos nenhuma indicação se isso é verdade para a população.
- Mesmo que MSE(A) = MSE(B) seja válido para a população, pode haver casos (amostras) em que A ou B vencem.
- Como em qualquer medida estimada, precisamos levar em conta a variabilidade da amostra e conduzir um teste
- O teste de Diebold e Mariano (1995) avalia o nulo da perda esperada igual
- O teste é válido sob condições gerais.

Teste de DM (Diebold e Mariano, 1995)

- Considere uma função de perda arbitrária L. Assumindo duas séries de erros de previsão,
- {eAt } t=1 -> t= T e {eBt }t=1 -> t= T
- Podemos calcular o diferencial de perda como:
- {dABt = L(eAt) − L(eBt)} t=1 -> t= T
- A principal suposição do teste é que dABt é covariância estacionária:

$$\begin{cases} E(d_{ABt}) = \mu & \forall t \\ Cov(d_{ABt}, d_{ABt-\tau}) = \gamma(\tau), & \forall t \\ 0 < var(d_{ABt}) = \sigma^2 < \infty & \end{cases}$$

Teste de DM (Diebold e Mariano, 1995)

- Usando o diferencial de perda, o teste formula o nulo de precisão preditiva igual:
- E(dABt) = 0
- Sob a suposição mencionada acima: $DM_{AB} = \frac{\bar{d}_{AB}}{\hat{\sigma}_{\bar{d}_{AB}}} \rightarrow N(0,1)$
- Aonde, $\bar{d}_{AB} = 1/T \sum_{t=1}^{T} d_{ABt}$ é a média diferençável.
- E $\hat{\sigma}_{\bar{d}_{AB}}$ é o estimador consistente do desvio padrão.

Teste de DM (Diebold e Mariano, 1995)

- Dado que o diferencial de perda pode ser correlacionado em série, o erro padrão no denominador deve ser robusto.
- No artigo original, os autores usaram um estimador consistente do espectro diferencial na frequência zero para calcular os erros padrão.
- O teste pode ser implementado por meio de uma análise de regressão (diferencial de perda de regressão em uma constante).
- Ao fazer isso, podemos implementar erros padrão robustos de **heterocedasticidade e autocorrelação** (HAC).
- Além disso, sob tal abordagem, o teste pode ser estendido para um teste condicional.

Quando Utilizar o teste de DM?

- Na realidade, a principal suposição do modelo nunca é mantida precisamente. No entanto, ela serve como uma aproximação útil...
- A suposição é baseada no fato de que as previsões são ótimas. Portanto, o diferencial não é previsível.
- Se as previsões forem subótimas, a suposição pode ainda ser mantida, mesmo para o caso em que há componentes não estacionários.

Quando Utilizar o teste de DM?

- O conjunto de informações tende a ser compartilhado entre as previsões, de modo que as não estacionariedades do erro de previsão podem desaparecer do diferencial de perdas.
- Há um vasto número de ferramentas para avaliar a suposição. Portanto, sabemos quando o teste será válido

Quando Utilizar o teste de DM?

- Falamos sobre duas suposições principais e discutimos apenas uma. A outra suposição se refere ao que testamos: previsões, não modelos
- O teste DM foi criado para comparar previsões em um ambiente sem modelo: Desconhecido Modelos, temos apenas os erros de previsão disponíveis
- Modelos cujas previsões são fortemente afetadas pelo erro de estimativa de parâmetros, tendem a ser considerados inferiores em relação aos modelos que são menos fortemente afetados pelo erro de estimativa de parâmetros.

Tamanho do teste de DM

- Para avaliar o tamanho do teste, os autores incluíram uma variedade de especificações de correlação contemporânea de erro de previsão, correlação serial de erro de previsão e distribuição de erro de previsão.
- Para os erros de previsão, eles calculam os casos com correlação contemporânea (ρ) igual a 0, 0,5 ou 0,9.
- Para a correlação serial, eles impõem um processo MA(1) com um θ igual a 0, 0,5 ou 0,9.

Tamanho do teste de DM

- Eles consideram as variáveis aleatórias gaussianas e t padronizadas com 6 df para as inovações.
- As avaliações são realizadas no nível $\alpha = 0,1$.
- Os tamanhos das amostras são definidos igualmente para 8,16, 32, 64, 128, 256, 512 observações
- Os cálculos de tamanho são avaliados em pelo menos 5000 replicações.

Teste de West (1996)

- West (1996) critica a implementação do teste DM para avaliar modelos
- O autor sugere que, na maioria dos casos, as previsões usadas em economia e finanças são baseadas em modelos paramétricos:
- $\hat{f} = \hat{y}(x_t, \hat{\beta})$

Teste de West (1996)

- Então o teste DM é útil para avaliar o modelo de previsão $\hat{y}$ para uma sequência específica de variáveis preditoras e parâmetro estimado, em vez de todos os valores das variáveis preditoras e no "parâmetro pseudo-verdadeiro“.
- Em outras palavras, os testes DM são úteis para avaliar o desempenho preditivo histórico e não para aprender sobre modelos.
- Uma grande parte da literatura faz exatamente isso. Ao assumir que a incerteza da estimativa do parâmetro será pequena, o diferencial de perda será aproximadamente estacionário.

Teste de West (1996)

- West (1996) fornece os resultados que permitem testar a seguinte hipótese nula:

$$E\left(f(Y_t, \hat{Y}(X_t, \hat{\beta}), Z_t)\right) = 0$$

- Seja TR o número de observações usadas para estimar o $\hat{\beta}$. Isso permite que TP = T − TR − τ de observações de amostra.
- O Teste de West renderiza:

$$\sqrt{\frac{1}{T_P}}\left(\sum_{t=T_R}^{T-\tau} l_{t+\tau}(\hat{\beta}) - E[l_{t+\tau}(\beta^*)]\right) \to^d N(0, \Omega)$$

were $\Omega = S_y(0) + 2PFV_\beta F' + \Pi(FBS'_{y\tau}(0) + S_{y\tau}(0)B'F')$

Teste de West (1996)

- Este resultado estabelece que podemos (assintoticamente) conduzir inferência usando distribuição normal se tivermos a estimativa correta de Ω.
- O primeiro termo em Ω é a variância de longo prazo da perda sob parâmetros conhecidos.
- O segundo termo reflete a estimativa da variância da perda.

Teste de West (1996)

- O terceiro termo reflete a covariância entre o primeiro e o segundo termo.
- Se não houver estimativa, este resultado se reduz ao resultado de Diebold e Mariano (1995).
- Se limTP /TR = 0, o erro de estimativa é $\hat{\beta}$ pode ser ignorado
- No entanto, a estrutura de West não pode ser aplicada a testes de MSE igual quando os modelos são “aninhados”.

Teste de West (1996)

- O terceiro termo reflete a covariância entre o primeiro e o segundo termo.
- Se não houver estimativa, este resultado se reduz ao resultado de Diebold e Mariano (1995).
- Se limTP /TR = 0, o erro de estimativa é $\hat{\beta}$ pode ser ignorado
- No entanto, a estrutura de West não pode ser aplicada a testes de MSE igual quando os modelos são “aninhados”.

Teste de Giacomoni e White (2006)

- O teste GW baseia-se no resultado de DM e West para propor uma abordagem fundamentalmente diferente, mas altamente relevante para testar métodos de previsão alternativos...
- Ele retém o efeito que os erros de estimativa têm nas previsões. Para fazer isso, os parâmetros devem ser estimados com uma janela fixa ou móvel.
- Por fim, o teste pergunta se os dois métodos de previsão estão produzindo a mesma qualidade de previsões em vez de um modelo ser melhor.
- Sob a estrutura GW, as previsões são consideradas como observações do desempenho dos métodos.

Teste de Giacomoni e White (2006)

- Considere dois modelos que no tempo t são usados para gerar previsões um passo à frente
- Como estamos usando um conjunto de dados específico para estimar os modelos, as previsões desses dois modelos são uma função dos parâmetros e dos dados: $f_{A,t+1|t}(\hat{\beta}_{At}),\ f_{B,t+1|t}(\hat{\beta}_{Bt})$
- Podemos construir a série de perdas para qualquer função de perda L:

$$L\left(f_{A,t+1|t}(\hat{\beta}_{At})\right),\ L\left(f_{B,t+1|t}(\hat{\beta}_{Bt})\right)$$

Teste de Giacomoni e White (2006)

- A hipótese nula do teste é definida como: $E[L\left(f_{A,t+1|t}(\hat{\beta}_{At})\right) - L\left(f_{B,t+1|t}(\hat{\beta}_{Bt})\right)] = 0$
- A hipótese equivalente do teste de West (1996) é: $E[L\left(f_{A,t+1|t}(\beta^*_{At})\right) - L\left(f_{B,t+1|t}(\beta^*_{Bt})\right)] = 0$
- A principal diferença é que GW avalia as perdas esperadas incluindo o $\hat{\beta}t$ estimado em vez dos limites dos estimadores
- Por exemplo, considere um modelo restrito contra um modelo irrestrito maior. Qual seria o resultado de ambos os testes?

Teste de Giacomoni e White (2006)

- GW permite comparar métodos em vez de modelos (como a abordagem de West)
- Modelos não aninhados podem ser comparados.
- Todos os métodos estimados são permitidos
- Além disso, Giacomini e White (2006) introduzem testes condicionais de precisão preditiva:

$$E(L\left(f_{A,t+1|t}(\hat{\beta}_{At})\right) - L\left(f_{B,t+1|t}(\hat{\beta}_{Bt})\right)|Z_t] = 0$$

Teste de Giacomoni e White (2006)

- Em outras palavras, o teste GW é capaz de testar diferenças no desempenho médio, mas também diferenças no desempenho previsto que são previsíveis dado um conjunto de informações Zt
- Por exemplo, considere um modelo de previsão de retorno de ações que tem melhor desempenho durante recessões.
- Então podemos definir as informações de condicionamento como uma variável indicadora que tem um valor de um se houver uma recessão e um valor de zero caso contrário...

$$E(L\left(f_{A,t+1|t}(\hat{\beta}_{At})\right) - L\left(f_{B,t+1|t}(\hat{\beta}_{Bt})\right)|Z_t] = 0$$

Teste de Giacomoni e White (2006)

- Em outras palavras, o teste GW é capaz de testar diferenças no desempenho médio, mas também diferenças no desempenho previsto que são previsíveis dado um conjunto de informações Zt
- Por exemplo, considere um modelo de previsão de retorno de ações que tem melhor desempenho durante recessões.
- Então podemos definir as informações de condicionamento como uma variável indicadora que tem um valor de um se houver uma recessão e um valor de zero caso contrário...

$$E(L\left(f_{A,t+1|t}(\hat{\beta}_{At})\right) - L\left(f_{B,t+1|t}(\hat{\beta}_{Bt})\right)|Z_t] = 0$$

Teste de Giacomoni e White (2006)

- Considerando a hipótese nula:
- Para implementar o teste, precisamos selecionar funções dos dados (funções de Zt) disponíveis no momento da construção das previsões. Chame essas funções de ht.

$$E(L\left(f_{A,t+1|t}(\hat{\beta}_{At})\right) - L\left(f_{B,t+1|t}(\hat{\beta}_{Bt})\right)|Z_t] = 0$$

Teste de Giacomoni e White (2006)

- Se assumirmos que ht é um vetor (qx1), podemos testar o nulo usando uma forma quadrática GMM padrão:
- Onde W é a matriz de peso ótima para o problema GMM
- Em condições moderadas, os autores mostram que T → χ_q^2.

$$T = T_P \left(T_p^{-1} \sum_{T_R}^{T-1} h_t \Delta L_{t+1} \right)' W^{-1} \left(T_p^{-1} \sum_{T_R}^{T-1} h_t \Delta L_{t+1} \right)$$

Teste de Giacomoni e White (2006)

- Segundo os autores, para $\tau = 1$ e $Z_t \neq \{\emptyset, \Omega\}$, podemos usar a amostra variância.
- Para $\tau > 1$ e $Z_t \neq \{\emptyset, \Omega\}$, devemos considerar possíveis autocorrelações até lag $\tau - 1$.
- Para $\tau \geq 1$ e $Z_t = \{\emptyset, \Omega\}$, podemos usar um estimador HAC.

Teste de Giacomoni e White (2006)

- Os autores avaliam o tamanho sob a seguinte configuração:

$$Y_t = c + CPI_t + \epsilon_t, \epsilon_t \; N(0, \sigma^2)$$
$$f_{At} = CPI_t + u_{At}$$
$$f_{Bt} = \delta + \gamma CPI_t + u_{Bt}$$

- Note que fAt é especificado incorretamente, pois omite a interceptação. Para valores pequenos de c, o viés do primeiro modelo pode ser correspondido pela variância aumentada do segundo modelo maior
- O tamanho nominal é definido como 5%

Quadratic Loss

	Out-of-Sample			
Estimation	25	75	125	150
25	0.053	0.037	0.035	0.024
75	0.062	0.048	0.04	0.037
125	0.073	0.054	0.044	0.042
150	0.061	0.056	0.048	0.046

Fonte: Giacomoni e White (2006)

Teste de Giacomoni e White (2006)

- Os autores avaliam o tamanho sob a seguinte configuração:
- Sob o primeiro cenário, eles consideram vários valores para μ, mas correlação zero:

$$\Delta L_{t+1} = \mu(1-\rho) + \rho \Delta L_t + \epsilon_{t+1}, \epsilon_{t+1} \text{ iid} N(0,1)$$

- Sob o segundo cenário, eles consideram valores zero para μ, mas vários valores para correlação.

Referências

Diebold, F. X., and Mariano, R. S. (1995), Comparing Predictive Accuracy, Journal of Business and Economic Statistics, 13, 130-141

Diebold, F. X., and Mariano, R. S. (2015), Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold–Mariano Tests, Journal of Business and Economic Statistics, 33, 1-9

Elliott, G., and Timmermann, A. (2016), Economic Forecasting, Princeton University Press. (Ch 2, Ch16, Ch17)

Giacomini, F., and White, H. (2006), Tests of Conditional Predictive Ability, Econometrica 74, 1545–78

West, K. D. (1996), Asymptotic Inference About Predictive Ability, Econometrica, 64, 1067–1084.