Econometria II

Relaxando as Premissas do Modelo MQO

Walberti Saith

Universidade Federal de Rondônia - UNIR

18 de Fevereiro de 2020

Hipóteses

Relembrando as premissas

Na Econometria I realizamos uma discussão aprofundada sobre o modelo de regressão linear clássico e mostramos como ele pode ser utilizado para lidar com problemas de inferência estatística, tais como estimação e teste de hipóteses, bem como os problemas de previsão. No entanto, como vimos, é importante destacar que esse modelo baseia-se nas diversas hipóteses simplificadoras listadas à seguir (Gujarati, 2000):

- Hipótese 1: O modelo de regressão é linear nos parâmetros.
- Hipótese 2: Os valores dos regressores, os $\boldsymbol{X}$, são fixos, ou valores de $\boldsymbol{X}$ são independentes do termo de erro. Isso significa que exigimos covariância zero entre $\boldsymbol{u_i}$ e cada variável $\boldsymbol{X}$.
- Hipótese 3: Para os $\boldsymbol{X}$ dados, o valor médio do erro $\boldsymbol{u_i}$ é zero.
- Hipótese 4: Para os $\boldsymbol{X}$ dados, a variância de $\boldsymbol{u_i}$ é constante ou homocedástica.

Relembrando as premissas

- Hipótese 5: Para os $\boldsymbol{X}$ dados, não há autocorrelação, nem correlação serial, entre os termos de erro.
- Hipótese 6: O número de observações $\boldsymbol{n}$ deve ser maior que o número de parâmetros a serem estimados.
- Hipótese 7: Deve haver variação suficiente nos valores das variáveis $\boldsymbol{X}$.
- Hipótese 8: Não há colinearidade exata entre as variáveis $\boldsymbol{X}$.
- Hipótese 9: O modelo está especificado corretamente, logo não há viés de especificação.
- Hipótese 10: O termo estocástico (de erro) $\boldsymbol{u_i}$ é distribuído normalmente.

Relaxamento da hipótese 8

O foco dessa aula será dado ao relaxamento da hipótese 8. Além disso, discutiremos as hipóteses 6 e 7. 🔔

Multicolineariedade

O que veremos?

Começaremos falando da premissa da multicolineariedade.

1. Qual a natureza da multicolinearidade?
2. A multicolinearidade é realmente um problema?
3. Quais são suas consequências práticas?
4. Como é detectada?
5. Que medidas podem ser tomadas para atenuar o problema da multicolinearidade?

Qual a natureza da multicolinearidade?

Originalmente, o termo multicolinearidade significava a existência de uma relação linear "perfeita" ou exata entre algumas ou todas as variáveis explicativas do modelo de regressão.

Hoje, no entanto, o termo multicolinearidade é usado em um sentido mais amplo, para incluir, além do caso de multicolinearidade perfeita, também o caso em que as variáveis explicativas (X) estão intercorrelacionadas, mas não perfeitamente.

Para entendermos a diferença entre a multicolinearidade *perfeita* e *menos que perfeita* e, além disso, ilustrar como o problema decorrente da multicolinearidade ocorre na prática, vamos usar como exemplo os dados que se encontram na Tabela 1.

Qual a natureza da multicolinearidade?

Y	X2	X3
1	1	2
2	2	0
3	3	4
4	4	6
5	5	8

Note que:

- $X_3 = 5X_2$
- X_3^* = à coluna X_3 somada a uma tabela de números aleatórios: 2, 0, 7, 9 e 2.

Correlação

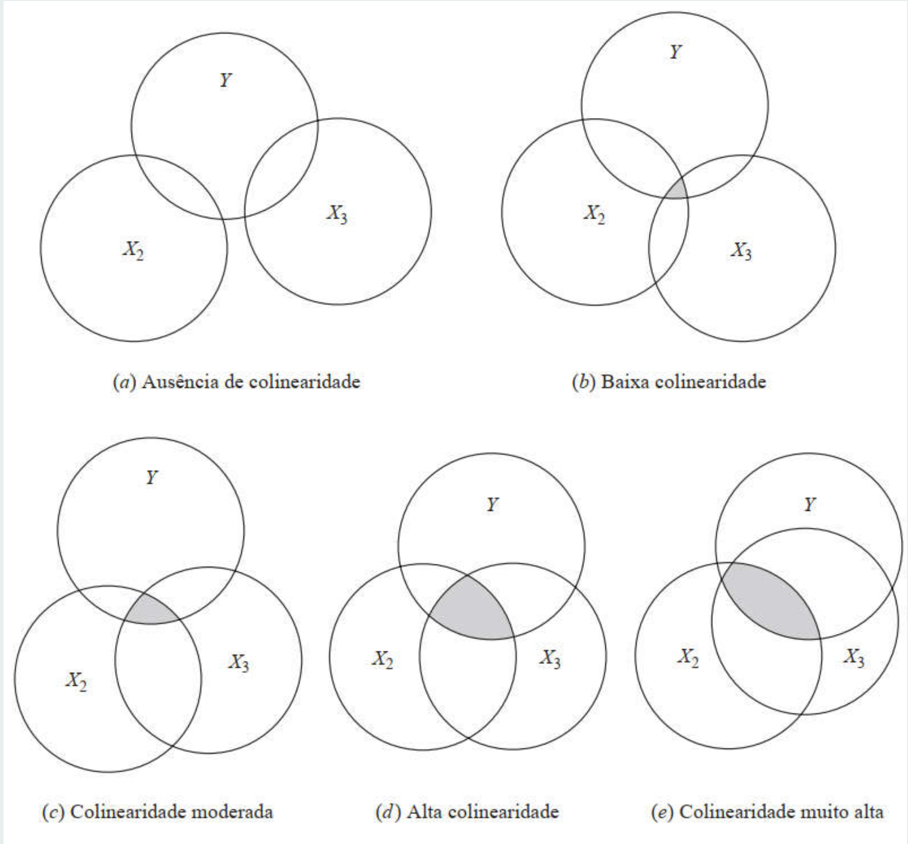
```
Coeficientes de Correlação, usando as observações 1 - 5
5% valor crítico (bicaudal) = 0.8783 para n = 5

           Y             X2             X3
      1.0000         0.9000         0.5063  Y
                     1.0000         0.5523  X2
                                    1.0000  X3
```

- Há uma colinearidade perfeita entre X_2 e X_3, pois o coeficiente de correlação r_{23} é igual à unidade.
- Não há colinearidade perfeita entre X_2 e X_3^*. Contudo, as duas variáveis estão altamente correlacionadas.

Como funciona ?

A multicolinearidade também pode ser descrita sucintamente através do diagrama de Venn.

Figura 1. Diagrama de Venn



(a) Ausência de colinearidade (b) Baixa colinearidade

(c) Colinearidade moderada (d) Alta colinearidade (e) Colinearidade muito alta

A multicolinearidade é realmente um problema?

Por que o modelo clássico de regressão linear pressupõe que não há multicolinearidade entre as variáveis explicativas X?

- Consequência 1 - Se a multicolinearidade for perfeita, os coeficientes de regressão das variáveis X serão indeterminados e seus erros padrão, infinitos.

- Consequência 2 - Se a multicolinearidade for menos que perfeita, os coeficientes de regressão, embora determinados, possuirão grandes erros padrão (em relação aos próprios coeficientes), o que significa que os coeficientes não podem ser estimados com grande precisão ou exatidão.

Fontes

Há várias fontes de multicolinearidade. Esse problema pode ser decorrente de algum ou alguns dos seguintes fatores:

1. O método de coleta de dados empregado.
2. Restrições ao modelo ou à população que está sendo amostrada.
3. Especificação do modelo.
4. Um modelo sobredeterminado.

Porque é importante?

Lembre-se do significado de β. Ele nos dá a variação do valor médio de $\boldsymbol{Y}$ quando $\boldsymbol{X_i}$ varia por uma unidade, mantendo $\boldsymbol{X_k}$ constante.

Agora, se existe colineariedade, não conseguimos separar a influência de cada $\boldsymbol{X_i}$. Isso significa que não há como distinguir as influências de $\boldsymbol{X_2}$ e $\boldsymbol{X_3}$ de uma forma separada na amostra dada.

O caso de multicolinearidade perfeita é uma situação patológica extrema. Em geral, não há relação linear exata entre as variáveis $\boldsymbol{X}$, principalmente, em dados envolvendo séries temporais econômicas.

Porque é importante?

Do ponto de vista prático a multicolinearidade esta presente em praticamente todas as regressões. Lembre-se de que, se as hipóteses do modelo clássico forem satisfeitas, os estimadores de MQO da regressão serão os melhores estimadores lineares não viesados.

Agora podemos mostrar que, mesmo se a multicolinearidade for muito alta, como no caso da quase multicolinearidade, os estimadores de MQO ainda conservarão a propriedade de melhores estimadores lineares não viesados.

Por que nos preocuparmos tanto com esse problema? De fato, Gujarati (2011) destaca que multicolinearidade, poucas observações e pequenas variâncias da variável independente são essencialmente o mesmo problema.

Porque é importante?

Perguntar "O que devo fazer com a multicolinearidade?" é como perguntar "O que devo fazer se não tenho muitas observações?". Não há resposta estatística para essa pergunta.

Para reforçar a importância do tamanho da amostra, Goldberger criou o termo micronumerosidade, para contrapor à polissílaba multicolinearidade. De acordo com Goldberger, a micronumerosidade exata surge quando n, o tamanho da mostra, é zero, caso em que qualquer tipo de estimação é impossível.

A quase micronumerosidade, como a quase multicolinearidade, surge quando o número de observações mal excede o número de parâmetros a serem estimados.

Quais são as consequências práticas

1. Embora sejam os melhores estimadores lineares não viesados, os estimadores de MQO têm grandes variâncias e covariâncias, tornando difícil uma estimação precisa.

2. Devido à consequência 1, os intervalos de confiança tendem a ser muito mais amplos, levando à aceitação imediata da "hipótese nula igual a zero" (isto é, o verdadeiro coeficiente populacional igual a zero).

3. Também, devido à consequência 1, a razão t de um ou mais coeficientes tende a ser estatisticamente insignificante.

4. Embora a razão t de um ou mais coeficientes seja estatisticamente insignificante, R^2, a medida geral da qualidade do ajustamento, pode ser muito alta.

Quais são as consequências práticas

Na verdade, à medida em que a correlação entre as variáveis independentes aumenta, as suas respectivas variâncias também aumentam, assim como a covariância entre elas. Essa propriedade é conhecida como Fator de Inflação de Variância (FIV).

Na presença de multicolineariedade, os erros padrão ficam extremamente sensíveis à variações na amostra.

Na próxima aula, veremos como detectar a multicolinearidade e que medidas podem ser tomadas para atenuar o problema.

Fator de Inflação da Variância - FIV

Como discutimos anteriormente, quando há o problema da multicolinearidade, os estimadores de MQO têm grandes variâncias e covariâncias, tornando difícil uma estimação precisa.

A velocidade com a qual as variâncias e covariâncias aumentam pode ser vista com o fator de inflação da variância (FIV), definido como:

$$FIV = \frac{1}{(1 - r_{23}^2)}$$

O FIV mostra como a variância de um estimador é inflada pela presença da multicolinearidade. Quando $\boldsymbol{r_{23}^2}$ aproxima-se de 1, o FIV aproxima-se do infinito. Se não houver colinearidade entre $\boldsymbol{X_2}$ e $\boldsymbol{X_3}$, o FIV será 1.

FIV continuação...

Os resultados discutidos podem ser facilmente estendidos ao modelo com $\boldsymbol{k}$ variáveis. Em tal modelo, a variância do $\boldsymbol{k}$-ésimo coeficiente, pode ser expressa como:

$$var(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2} FIV_j$$

Podemos notar que o inverso de FIV é chamado de tolerância (TOL). Ou seja:

$$TOL_j = \frac{1}{FIV_j} = (1 - R_j^2)$$

FIV continuação...

1. Quando $\boldsymbol{R_j^2 = 1}$ (colinearidade perfeita), $\boldsymbol{TOL_j = 0}$.
2. Quando $\boldsymbol{R_j^2 = 0}$ (não há colinearidade nenhuma), $\boldsymbol{TOL_j}$ é 1.

Para se ter uma ideia da rapidez com que as variâncias aumentam quando $\boldsymbol{r_{23}}$ aumenta, considere a Tabela 10.1 seguinte que destaca essas variâncias para valores selecionados de $\boldsymbol{r_{23}}$:

Exemplo de FIV

Valor de r_{23} (1)	FIV (2)	var $(\hat{\beta}_2)$ (3)*
0,00	1.00	$\frac{\sigma^2}{\sum x_{2i}^2} = A$
0,50	1,33	$1,33 \times A$
0,70	1,96	$1,96 \times A$
0,80	2,78	$2,78 \times A$
0,90	5,76	$5,26 \times A$
0,95	10,26	$10,26 \times A$
0,97	16,92	$16,92 \times A$
0,99	50,25	$50,25 \times A$
0,995	100,00	$100,00 \times A$
0,999	500,00	$500,00 \times A$

Fator de Inflação da Variância - FIV

Como discutimos anteriormente, quando há o problema da multicolinearidade, os estimadores de MQO têm grandes variâncias e covariâncias, tornando difícil uma estimação precisa.

A velocidade com a qual as variâncias e covariâncias aumentam pode ser vista com o fator de inflação da variância (FIV), definido como:

$$FIV = \frac{1}{(1 - r_{23}^2)}$$

O FIV mostra como a variância de um estimador é inflada pela presença da multicolinearidade. Quando r_{23}^2 aproxima-se de 1, o FIV aproxima-se do infinito. Se não houver colinearidade entre X_2 e X_3, o FIV será 1.

Fator de Inflação da Variância - FIV

Os resultados discutidos podem ser facilmente estendidos ao modelo com $\boldsymbol{k}$ variáveis. Em tal modelo, a variância do $\boldsymbol{k}$-ésimo coeficiente, pode ser expressa como:

$$\boldsymbol{var}(\hat{\beta}_j) = \frac{\boldsymbol{\sigma^2}}{\sum \boldsymbol{x_j^2}} \boldsymbol{FIV_j}$$

Podemos notar que o inverso de FIV é chamado de \textbf{tolerância (TOL)}. Ou seja:

$$\boldsymbol{TOL_j} = \frac{1}{\boldsymbol{FIV_j}} = (1 - \boldsymbol{R_j^2})$$

Fator de Inflação da Variância - FIV

1. Quando $\boldsymbol{R_j^2 = 1}$ (colinearidade perfeita), $\boldsymbol{TOL_j = 0}$.

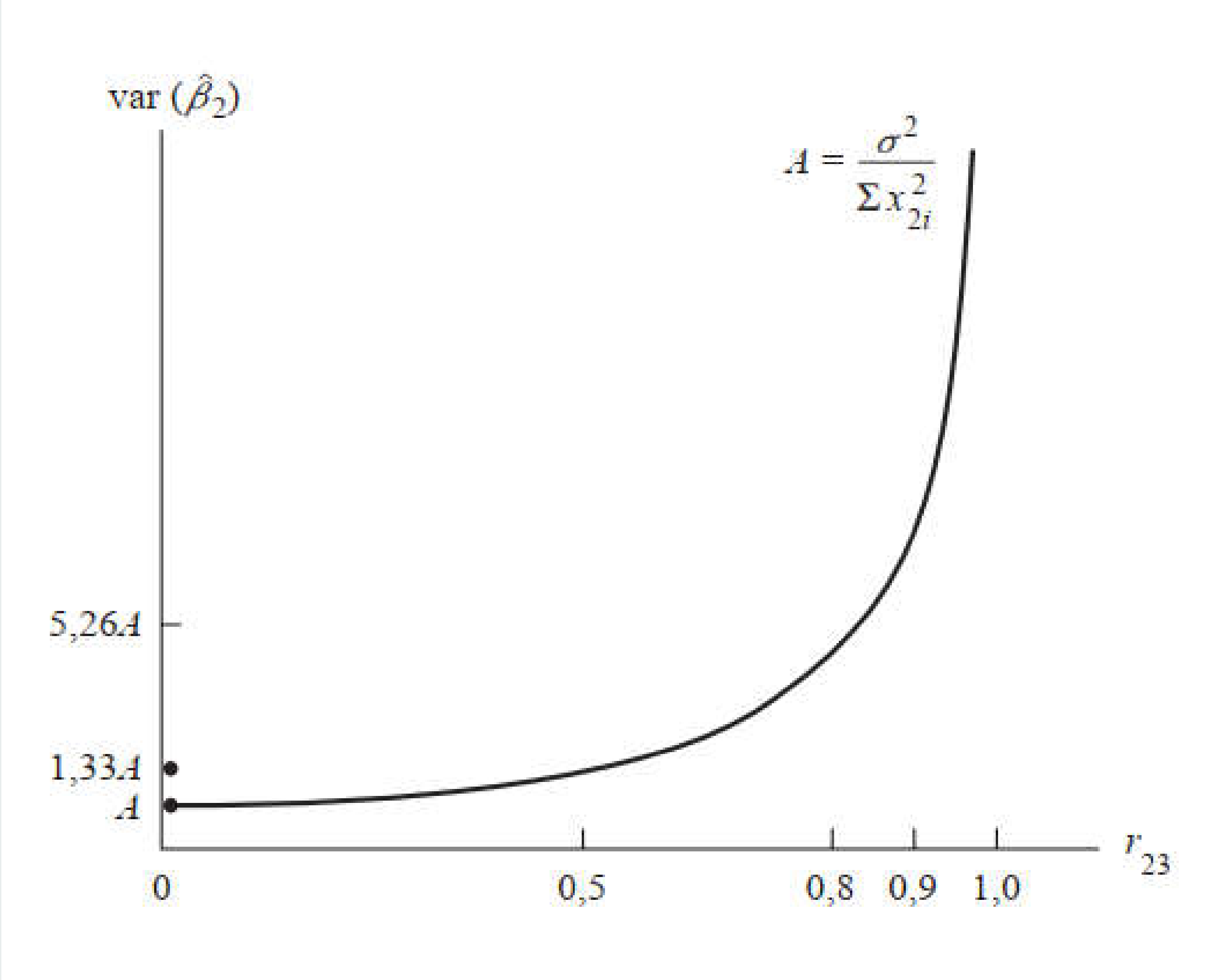
2. Quando $\boldsymbol{R_j^2 = 0}$ (não há colinearidade nenhuma), $\boldsymbol{TOL_j}$ é 1.

Para se ter uma ideia da rapidez com que as variâncias aumentam quando $\boldsymbol{r_{23}}$ aumenta, considere a Tabela 10.1 seguinte que destaca essas variâncias para valores selecionados de $\boldsymbol{r_{23}}$:

Tabela. Efeito de aumentos de r_{23} na var $(\hat{\beta}^2)$.

Valor de r_{23} (1)	FIV (2)	var $(\hat{\beta}_2)$ (3)*
0,00	1.00	$\frac{\sigma^2}{\sum x_{2i}^2} = A$
0,50	1,33	$1,33 \times A$
0,70	1,96	$1,96 \times A$
0,80	2,78	$2,78 \times A$
0,90	5,76	$5,26 \times A$
0,95	10,26	$10,26 \times A$
0,97	16,92	$16,92 \times A$
0,99	50,25	$50,25 \times A$
0,995	100,00	$100,00 \times A$
0,999	500,00	$500,00 \times A$

Figura 10.2 - Comportamento da var $\left(\hat{\beta}^2\right)$ como função de r_{23}.



Outras consequências.

Outras consequências práticas da multicolinearidade são: intervalos de confiança mais amplos e razões t insignificantes.

- Intervalos de confiança mais amplos

Dados os erros padrão grandes, os intervalos de confiança dos parâmetros populacionais relevantes tendem a ser maiores. Portanto, em casos de alta multicolinearidade, os dados da amostra podem ser compatíveis com um conjunto diverso de hipóteses. A probabilidade de aceitar uma hipótese falsa (erro tipo II) aumenta.

Observe a Tabela 10.2 abaixo. Por exemplo, quando $\boldsymbol{r_{23} = 0,95}$, o intervalo de confiança para β_2 é maior que quando $\boldsymbol{r_{23} = 0}$ por um fator de $\sqrt{\mathbf{10,26}}$, ou cerca de 3.

Tabela 10.2 - O efeito da colinearidade crescente no intervalo de confiança.

Valor de r_{23}	Intervalo de confiança de 95% para β_2
0,00	$\hat{\beta}_2 \pm 1{,}96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0,50	$\hat{\beta}_2 \pm 1{,}96\sqrt{(1{,}33)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0,95	$\hat{\beta}_2 \pm 1{,}96\sqrt{(10{,}26)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0,995	$\hat{\beta}_2 \pm 1{,}96\sqrt{(100)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0,999	$\hat{\beta}_2 \pm 1{,}96\sqrt{(500)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$

Outras consequências.

Razões *t* "insignificantes"

Lembre-se: para testar a hipótese nula que, por exemplo, $\beta_2 = 0$, usamos a razão t, isto é, e comparamos o valor de t estimado com o valor crítico de t na tabela t. Mas, como vimos, em casos de alta colinearidade, os erros padrão estimados aumentam acentuadamente, tornando os valores t menores. Em tais casos, aceita-se cada vez mais a hipótese nula de que o verdadeiro valor populacional relevante é zero.

Sensibilidade dos estimadores de MQO

Contanto que a multicolinearidade não seja perfeita, é possível estimar os coeficientes de regressão, mas as estimativas e seus erros padrão tornam-se muito sensíveis até mesmo à menor alteração nos dados.

Para comprovar isso, considere o exemplo descrito abaixo:

Sensibilidade dos estimadores de MQO

Tabela 10.3

Sensibilidade dos estimadores de MQO

Tabela 10.4

Outros Aspectos

- Todas essas alterações podem ser atribuídas a um aumento na multicolinearidade na Tabela 10.3, $\boldsymbol{r_{23} = 0,5523}$, enquanto na Tabela 10.4 é 0,8285.

- Da mesma forma, os erros padrão de $\boldsymbol{\beta_2}$ e $\boldsymbol{\beta_3}$ aumentam entre as duas regressões, um sintoma comum de colinearidade.

Detecção da Multicolineariedade

- Verificar a existência da multicolineariedade não é uma tarefa fácil, mas podemos estabelecer algumas formas de testar a presença desse fenômeno.

- O R^2 alto, mas poucas razões t significativas.

Detecção da Multicolinearidade

1. Altas correlações entre os pares de regressores - Um dos testes simples de se fazer antes de se iniciar uma regressão é analisar a matriz de correlação, justamente para verificar como as variáveis se relacionam.

2. Exame de correlações parciais.

3. Regressões auxiliares - Se o $\boldsymbol{R}^2$ obtido de uma regressão auxiliar for maior que o $\boldsymbol{R}^2$ geral, aquele obtido da regressão de $\boldsymbol{Y}$ contra todos os regressores.

4. Autovalores e índice condicional.

5. Tolerância e fator de inflação de variância - FIV.

6. Diagrama de dispersão.

Medidas corretivas

O que podemos fazer se a multicolinearidade for grave? Temos duas opções:

- Não fazer nada;
- Seguir alguns procedimentos;

Procedimentos

- Uma informação a *priori*;
- Combinando dados de corte transversal e de séries temporais;
- Exclusão de variável(is) e viés de especificação;
- Transformação de variáveis;
- Dados adicionais ou novos;
- Reduzindo a colinearidade em regressões polinomiais.