

2º Lista de exercícios

Docente: Prof. Dr. Daniel Domingues dos Santos

Monitor: Matheus Mascioli Berlinger

Email: mbmascioli@gmail.com

1

Sobre dados em painéis, responda:

- a) Suponha que você tenha dados de diversos indivíduos ao longo do tempo, ou seja, observe os indivíduos $i = 1, 2, 3, \dots, N$ durante o período $t = 1, 2, 3, \dots, N$. Em cada linha do banco de dados temos um indivíduo em um determinado ponto do tempo. Quais as hipóteses necessárias para que a estimação por MQO seja consistente? Qual ganho você terá se estimar utilizando dados em painel?

Lembre-se de que uma das hipóteses importantes do modelo clássico de regressão linear é que não haja correlação entre os regressores e o termo de erro.

Se $E(\epsilon_{it}|x_{it}) = 0$, então o estimador de β por MQO (pooled) é não-viesado mas *ineficiente* se ϵ_{it} não é *iid* (independente e identicamente distribuído), o que é comum em dados em painel.

Se $E(\epsilon_{it}|x_{it}) \neq 0$ temos uma violação de uma das hipóteses principais do modelo clássico de regressão linear – ou seja, que o termo de erro não está correlacionado com os regressores. Como sabemos, nessa situação, as estimativas de MQO não são apenas tendenciosas, mas também inconsistentes.

Vantagens dos dados em painel (Econometria Básica 5ed - Gujarati):

1. Uma vez que os dados em painel se relacionam a indivíduos, empresas, Estados, países etc., com o tempo, tende a haver heterogeneidade nessas unidades. As técnicas de estimação dos dados em painel podem levar em consideração a heterogeneidade explicitamente, permitindo variáveis específicas ao sujeito, como mostraremos rapidamente. Usamos o termo sujeito em sentido genérico, para incluir microunidades como indivíduos, empresas, Estados e países.
2. Combinando séries temporais com observações de corte transversal, os dados em painel oferecem “dados mais informativos, maior variabilidade, menos colinearidade entre variáveis, mais graus de liberdade e mais eficiência”.

3. Estudando repetidas observações em corte transversal, os dados em painel são mais adequados para examinar a dinâmica da mudança. Períodos de desemprego, rotatividade no emprego e mobilidade da mão de obra são analisados de maneira mais apropriada com dados em painel.
4. Os dados em painel podem detectar e medir melhor os efeitos que simplesmente não podem ser observados em um corte transversal puro ou em uma série temporal pura. Por exemplo, os efeitos das leis de salário mínimo sobre o emprego e ganhos poderão ser estudados mais adequadamente se incluirmos ondas sucessivas de aumentos de salários nos salários mínimos estadual e/ou federal.
5. Dados em painel permitem estudar modelos de comportamento mais complicados. Por exemplo, fenômenos como economias de escala e mudança tecnológica podem ser mais bem conduzidos pelos dados em painel do que apenas pelo corte transversal ou pelas séries temporais.
6. Ao disponibilizar os dados referentes a milhares de unidades, os dados em painel podem minimizar o viés que poderia resultar se estivéssemos trabalhando com um agregado de indivíduos ou empresas.

- a) Qual problema podemos enfrentar com os dados de painel que o MQO não seja mais a melhor escolha de estimação?

É provável que os termos de erro para um modelo usando dados de painel exibam certos tipos de dependência, que devem ser levados em conta quando estimamos tal modelo.

2

A literatura médica indica claramente que indivíduos melhor nutridos são mais produtivos na execução de tarefas. Isso significa que a qualidade da dieta pode afetar a produtividade no trabalho, e portanto, salários comandados pelos agentes. Por outro lado, é fácil admitir que pessoas mais produtivas e com maiores salários podem ter acesso a dietas melhores. Neste exercício você está encarregado de mensurar o impacto do consumo de calorias sobre os salários. Considere o seguinte modelo:

$$\text{salario}_{it} = \alpha_0 + \alpha_1 \text{calorias}_{it} + \epsilon_{it}$$

enquanto, por hipótese, leva tempo para que a maior renda tenha impacto sobre o consumo de calorias do indivíduo, ou seja:

$$\text{calorias}_{it} = \beta_0 + \beta_1 \text{salario}_{it-1} + \beta_2 z_{it} + \eta_{it}$$

Suponha ainda que:

- $cov(calorias_{is}, \epsilon_{it}) = 0 \quad \forall s < t$;
- $cov(z_{it}, \epsilon_{it}) = 0$;
- $cov(z_{it}, \eta_{it}) = 0$;
- $cov(\epsilon_{it}, \eta_{it}) = 0$;

- a) Mostre que se $cov(\epsilon_{it}; \epsilon_{it-1}) \neq 0$, o estimador MQO de α_1 é inconsistente. Apresente pelo menos uma razão para que essa hipótese sobre a covariância do distúrbio seja verdadeira.

$$\begin{aligned}
\hat{\alpha}_1 &= \frac{Cov(salario_{it}, calorias_{it})}{Var(calorias_{it})} \\
&= \frac{Cov(\alpha_0 + \alpha_1 calorias_{it} + \epsilon_{it}, calorias_{it})}{Var(calorias_{it})} \\
&= \underbrace{\frac{Cov(\alpha_0, calorias_{it})}{Var(calorias_{it})}}_0 + \alpha_1 \frac{Cov(calorias_{it}, calorias_{it})}{Var(calorias_{it})} + \frac{Cov(\epsilon_{it}, calorias_{it})}{Var(calorias_{it})} \\
&= \alpha_1 + \frac{Cov(\epsilon_{it}, \beta_0 + \beta_1 salario_{it-1} + \beta_2 z_{it} + \eta_{it})}{Var(calorias_{it})} \\
&= \alpha_1 + \underbrace{\frac{Cov(\epsilon_{it}, \beta_0)}{Var(calorias_{it})}}_0 + \beta_1 \frac{Cov(\epsilon_{it}, salario_{it-1})}{Var(calorias_{it})} + \beta_2 \underbrace{\frac{Cov(\epsilon_{it}, z_{it})}{Var(calorias_{it})}}_0 + \underbrace{\frac{Cov(\epsilon_{it}, \eta_{it})}{Var(calorias_{it})}}_0 \\
&= \alpha_1 + \beta_1 \frac{Cov(\epsilon_{it}, \alpha_0 + \alpha_1 calorias_{it-1} + \epsilon_{it-1})}{Var(calorias_{it})} \\
&= \alpha_1 + \beta_1 \underbrace{\frac{Cov(\epsilon_{it}, \alpha_0)}{Var(calorias_{it})}}_0 + \beta_1 \alpha_1 \underbrace{\frac{Cov(\epsilon_{it}, calorias_{it-1})}{Var(calorias_{it})}}_0 + \beta_1 \frac{Cov(\epsilon_{it}, \epsilon_{it-1})}{Var(calorias_{it})} \\
&= \alpha_1 + \beta_1 \underbrace{\frac{Cov(\epsilon_{it}, \epsilon_{it-1})}{Var(calorias_{it})}}_{\neq 0} \\
&\neq \alpha_1
\end{aligned}$$

Um forte argumento que defende a hipótese da covariância do distúrbio ser diferente de zero é que fatores não observados (como a aptidão) que afetam o salário de um indivíduo em $t - 1$ também afetarão o salário dessa pessoa em t .

- b) Com $\beta_2 = 0$, mostre que o estimador de variáveis instrumentais disponível para α_1 não é consistente se apenas uma cross-section de dados está disponível (incluindo nestes os dados defasados de salários). Justique brevemente sua resposta.
- c) Com $\beta_2 \neq 0$, mostre como você poderia testar se, no modelo original, a medida de nutrição é realmente uma variável endógena. Justique brevemente a sua resposta.
- d) Novamente com $\beta_2 = 0$, apresente o estimador consistente de α_1 que você poderia propor (mostre sua consistência e as condições necessárias para tanto) caso um painel de dados estivesse disponível. Justique brevemente a sua resposta.

3

Suponha que o interesse do pesquisador seja medir se educação tem diferentes componentes que trazem retorno em termos de salário para aqueles que investem nessa acumulação de capital humano. O pesquisador se propõe a examinar se os empregadores usam nível educacional na hora de estabelecer ofertas de trabalho por dois motivos:

- educação faz com que o indivíduo seja mais produtivo (com o que ele aprende na escola/faculdade);
- acumular anos de educação sinaliza que o indivíduo tem atributos de dedicação e perseverança que também são associados à características de produtividade do trabalhador.

O pesquisador consegue organizar um banco de dados com as seguintes informações:

- logaritmo salário de entrada no trabalho e os estabelecidos depois de cada renovação contratual entre 2001 e 2010 - $\ln(w_{it})$;
- nível de educação do indivíduo (estas credenciais são constantes desde que o indivíduo entra na rma) - que é observável com exatidão pelo empregador (diplomas, históricos escolares, etc., são verificáveis) - $educ$;
- uma medida de produtividade intrínseca de cada indivíduo (invariante no tempo), que é observado pelo pesquisador (econometrista), mas nunca está a disposição do empregador (e até por isso ele usa educação para inferir esta habilidade) - $prod$;
- Indicadores de raça e sexo de cada trabalhador - x .

O pesquisador então propõe o seguinte modelo para analisar se realmente a sinalização tem efeito nos retornos econômicos da educação:

$$\ln(w_{it}) = \beta_0 + \beta_1 educ_i + \beta_2 prod_i + \beta_3 t + \beta_4 t * educ_i + \beta_5 t * prod_i + \delta' x_i + \theta_i + \epsilon_{it}$$

onde t é a tendência.

Responda:

- a) Qual método de estimação você utilizaria para estimar o modelo? Discuta a adequação do mesmo ao problema enfrentado pelo pesquisador e em particular as hipóteses que garantem a consistência do seu estimador.
- b) Tendo este banco de dados em mãos, proponha testes de hipótese que examina a seguinte armação: “o efeito de sinalização embutido no retorno para educação deve diminuir a medida que os empregadores aprendem sobre a verdadeira produtividade dos trabalhadores, ao mesmo tempo que o retorno para esta última aumenta”.