

# Econometria II

## 2019

### Lista para Modelo Dinâmico

Victor Gomes

Entrega: 5 de Julho, 2019

**Instruções** O problema deve ser enviado por email (para victorgomes@unb.br) na data informada acima. Por favor, envie os arquivos em formato .zip.

Todas as submissões devem ser em formato eletrônico e devem incluir suas respostas ao conjunto de pergunta bem como os códigos. Por favor, apresente seus resultados em separado e com tabelas e figuras objetivas. Por favor, documente (comente) o seu código e forneça um arquivo de comando central tal que seja possível replicar inteiramente a sua solução.

A colaboração é recomendada. Entretanto, todo aluno deve implementar seu próprio código e submeter a sua resposta. No arquivo enviado deve ser claramente descrito quem trabalhou em conjunto.

**Descrição** Esta lista tem dois problemas: um teórico e outro empírico.

O problema teórico está dividido em 3 partes, seguindo a ideia de construir um modelo para análise estrutural. Na primeira seção é preciso construir o modelo do paper de John Rust. Na segunda seção é preciso adicionar uma estrutura estocástica para permitir se ajustar o modelo aos dados. Na seção três é preciso adicionar hipóteses sobre funções e distribuições para reduzir o custo computacional da estimação.

A primeira parte do problema empírico pergunta para aplicar o método de Rust de Nested Fixed Point no seu modelo. Na segunda parte é preciso aplicar o método CCP Hotz-Miller ao seu modelo.

Alguns itens consomem mais tempo. Eles estão marcados com (\*). O conselho é que sejam solucionados por último.

# 1 Teoria

## 1.1 Modelo Econômico

Harold Zurcher gerencia uma frota de ônibus que é sujeita a todo tipo de problema quando esta na rua. A milhagem (quilometragem) acumulada de um ônibus  $x_t$  é a variável de estado do problema. O desgaste do ônibus afeta o custo operacional esperado  $c(x_t; \theta_1)$  que depende da milhagem e um vetor de parâmetros não conhecido  $\theta_1 = \{\theta_{11}, \dots, \theta_{1n}\}$ .

Assuma que os custos dos ônibus vem de dois componentes: manutenção regular e despesas operacionais  $m(\cdot)$  e o custo  $f(\cdot)$  de substituir o motor no caso de falha (que é um evento estocástico que ocorre com alguma probabilidade).

- a) Escreva o custo como uma combinação destes dois componentes. Indique claramente quais elementos são função de  $x_t$ . Argumente informalmente que  $\frac{\partial c}{\partial x_t} > 0$ . Esta hipótese é necessária para a solução do modelo. Ela é uma boa hipótese? Você pode fazer um argumento para explicar por que  $\frac{\partial c}{\partial x_t}$  poderia ser negativa (pelo menos para alguns valores de  $x_t$ )?
- b) Rust não possui dados suficientes para estimar os vários componentes da função custo, então ele simplesmente estimata  $c(x_t, \theta_1)$ . Que dados ele precisaria e que estratégia ele poderia empregar se fosse querer identificar separadamente  $m(\cdot)$  e  $f(\cdot)$ ? Descreva qualquer hipótese que você precisa usar com os dados para fazer funcionar a sua solução.

No começo de cada período  $t$  Harold Zurcher observa  $x_t$  e decide quando pagar o custo de manutenção  $c(x_t, \theta)$  ou substituir o motor, que instantaneamente leva a medida de uso para zero  $x_t = 0$ . Então, o benefício de um único período é dado por

$$u(x_t, i_t, \theta) = \begin{cases} -c(x_t, \theta) & \text{if } i_t = 0 \\ -R - c(0, \theta) & \text{if } i_t = 1 \end{cases}$$

tal que  $R$  é o custo esperado de substituição do motor e  $i_t$  é a decisão binária de investimento.

Faça  $F(x_{t+1} | x_t, i_t, \theta)$  representar a função de distribuição cumulativa do processo estocástico que governa a evolução da variável de estado.

- c) Escreva a equação de Bellman para este problema.

Se pode obter uma solução analítica para a equação de Bellman em c) assumindo que a distância percorrida em cada período é distribuída exponencialmente com o parâmetro  $\theta_2$ , que é independente da milhagem rodada no período anterior:

$$F(x_{t+1} - x_t) = 1 - \exp(-\theta_2(x_{t+1} - x_t))$$

Fazendo isso implica em uma função política estacionária

$$i(x_t, \theta) = \begin{cases} 1 & \text{if } x_t \geq \gamma(\theta) \\ 0 & \text{if } x_t < \gamma(\theta) \end{cases}$$

tal que  $\gamma(\theta)$  é a solução única para

$$R(1 - \beta) = \int_0^{\gamma(\theta)} [1 - \beta \exp(-\theta_2(1 - \beta)y)] \frac{\partial c(y, \theta_1)}{\partial y} dy$$

Caso esteja interessado em detalhes deste modelo, ele foi derivado por Rust em um paper anterior (“Stationary Equilibrium in a Market for Durable Assets”, Econometrica 1985).

- d) Este modelo sofre de um problema de “sobre-previsão.” Explique o que significa. Se pode usar este modelo para estimação? Por que ou por que não?

## 1.2 Modelo Estocástico

Para estimar o comportamento que vemos nos dados é preciso adicionar termos de erro ao modelo. Uma forma de fazer isso é assumindo que os agentes desviam da solução ótima do modelo:

$$i_t = i(x_t, \theta) + \omega_t$$

Rust opta por uma estratégia diferente. Ele assume que os payoffs recebem choques aleatórios  $\varepsilon_{i,t}$ :

$$u(x, i, \theta) = \begin{cases} -c(x_t, \theta_1) + \varepsilon_{0,t} & \text{if } i_t = 0 \\ -R - c(0, \theta_1) + \varepsilon_{1,t} & \text{if } i_t = 1 \end{cases}$$

$\varepsilon_{i,t}$  é observado por Harold Zurcher, mas não pelo econometrista.

- e) Interprete  $\omega_t$  e  $\varepsilon_t$ . Que tipo de erro eles representam? Discuta as vantagens e desvantagens de cada escolha de modelagem. Por que você acha que Rust optou pela segunda opção?
- f) Escreva a nova função-valor  $V(x_t, \varepsilon_t)$

A solução para este problema é dada por uma regra de decisão estacionária  $i_t = i(x_t, \varepsilon_t; \theta)$ , que especifica a decisão ótima do agente quando a variável de estado é  $(x_t, \varepsilon_t)$ . Neste ponto, o modelo estatístico é completamente especificado mas ainda é muito complicado de estimar – especialmente com dados limitados.

### 1.3 Simplificando as hipóteses para estimação

Rust fez as hipóteses simplificadoras  $P(x_{t+1}; \varepsilon_{t+1} | x_t, \varepsilon_t) = P_1(x_{t+1} | x_t) P_2(\varepsilon_{t+1})$  (ele chama isso de “Conditional Independence Assumption”).

g) Defina o valor esperado condicional de  $V(x_t)$  sobre  $\varepsilon$  como

$$EV(x_t) = \int V(x_t; \varepsilon_t) P_2(\varepsilon_t)$$

Mostre que

$$EV(x_t) = \mathbb{E}_t \left[ \max \left\{ -c(x_t, \theta_1) + \varepsilon_{0,t} + \beta \int EV(x_{t+1}) P(dx_{t+1} | x_t); \right. \right. \\ \left. \left. -R - c(0, \theta) + \varepsilon_{1,t} + \beta \int EV(x_{t+1}) P(dx_{t+1} | 0) \right\} \right] \quad (1)$$

h) (\*) Prove que a equação (1) é um mapeamento de contração (contraction mapping).

i) Explique o tradeoff resultante da hipótese de indenpendência condicional (CI). Use um exemplo concreto para explicar o que é descartado (ou deixado de lado). O que esta hipótese “compra” (i.e. ela entrega) para a estimação? Essa hipótese é razoável?

Embora equação 1 é uma contração, ainda é difícil computar sem fazer hipóteses distribucionais adicionais sobre  $\varepsilon$ . Rust assume que o choque  $\varepsilon_{i,t}$  segue uma distribuição valor extremo Tipo I normalizada. Então

$$\varepsilon_{i,t,i} \sim EV1 \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{\pi^2}{6} & 0 \\ 0 & \frac{\pi^2}{6} \end{bmatrix} \right)$$

j) Por que Rust escolheu uma distribuição padrão para o termo de erro ao invés de adicionar um parâmetro para a sua variância?

A hipótese de distribuição valor extremo nos permite derivar formas funcionais para  $EV(x_t)$ , e  $P(1 | x_t, \theta)$ .

- k) Discuta a racionalidade por trás da hipótese da distribuição de  $\varepsilon_t$ . Existem outras alternativas? Escolha uma alternativa e discuta os prós e contras em relação à distribuição valor extremo Tipo I.

A última simplificação que Rust faz é discretizar o espaço de  $x_t$  dividindo a milhagem em 90 intervalos de 5000 milhas. Ele assume que o processo  $x_t$  por avançar com pelo menos dois incrementos no tempo, que essencialmente reduz a distribuição de  $(x_{t+1} - x_t)$  como uma multinomial com parâmetros  $\theta_3 = \{\theta_{30}, \theta_{31}\}$ .

- l) Escreva  $P(x_{t+1} | x_t, i_t, \theta)$ .

Aqui já se está em posição de estimar o modelo. As únicas duas coisas que não foram especificadas são  $\beta$  e a forma da função custo. Aqui podemos colocar  $\beta = 0.9999$ .<sup>1</sup> Finalmente, se pode assumir que a função custo é linear:

$$c(x_t, \theta_1) = 0.001\theta_1 x_t$$

- m) Observe que não foi incluída uma constante nesta especificação de função custo. O que aconteceria se isto fosse feito? Poderíamos estimar uma constante separadamente? Discuta como isso impacta o resultado final de sua estimativa.

## 2 Estimação

### 2.1 Dados

Primeiro é necessário construir a base de dados. A base consiste em oito arquivos ASCII, contendo dados da leitura mensal do odômetro dos 162 ônibus do frota (Madison Metropolitan Bus Company). Estes dados são da operação dos ônibus entre Dezembro de 1974 e Maio de 1985. Cada arquivo corresponde a um modelo/categoria de ônibus da frota da Madison Metro:

1. g870.ASC 36x15 matriz do modelo Grumman 870
2. rt50.ASC 60x4 matriz do modelo Chance RT50
3. t8h203.ASC 81x48 matriz do GMC T8H203

---

<sup>1</sup> $\beta$  e  $R$  são em geral altamente colineares e portanto de difícil distinção. Para um tratamento rigoroso desta questão veja a seção 3.5 do capítulo escrito por Rust no Handbook of Econometrics (1994). Outra referência útil é o paper de Magnac e Thesmar (Econometrica, 2002).

4. a452372.ASC 137x18 matriz do GMC A4523, modelo ano 1972
5. a452374.ASC 137x10 matriz do GMC A4523, modelo ano 1974
6. a530872.ASC 137x18 matriz do GMC A5308, modelo ano 1972
7. a530874.ASC 137x12 matriz do GMC A5308, modelo ano 1974
8. a530875.ASC 128x37 matriz do GMC A5308, modelo ano 1975

Os dados em cada arquivo estão vetorizados em uma única coluna: e.g. g870.ASC é um vetor 540x1 consistindo nas colunas de uma matriz 110x4 que foi empilhada. Comece importando os dados e reorganizando cada vetor em uma matriz. As primeiras 11 “linhas” de cada matriz consiste no metadado:

Row	Field	Sample Entry
1	Bus ID	5297
2	Month purchased	8
3	Year purchased	75
4	Month of 1st engine replacement	4
5	Year of 1st engine replacement	79
6	Odometer reading at 1st engine replacement	153400
7	Month of 2nd engine replacement	0
8	Year of 2nd engine replacement	0
9	Odometer reading at 2nd engine replacement	0
10	Month odometer data begins	9
11	Year odometer data begins	75
12+	Monthly odometer reading	4235

Você deve organizar os dados em objetos que são fáceis de manipular para realizar a sua estimação. Em particular, *observe que os dados do odômetro NÃO são zerados quando o motor é substituído. Você deve realizar esse ajuste.*

**Importante:** Todos os dados estão nos arquivos em anexo, mas Rust roda o principal procedimento para um conjunto restrito de ônibus. Esta restrição é devido a aparente heterogeneidade na matriz de transição do espaço de estado entre os ônibus. Os arquivos que devem ser usados para comparar com os resultados de Rust são 1, 2, 3 e 8.

## 2.2 Computação I: NFXP de Rust

Dada as observações  $(\{x_0, i_0\}, \{x_1, i_1\}, \dots, \{x_T, i_T\})$  podemos escrever uma função verossimilhança

$$\ell(x_1, \dots, x_T, i_1, \dots, i_T | x_0, i_0, \theta) = \prod_{t=1}^T P(i_t | x_t, \theta) \cdot P(x_t | x_{t-1}, i_{t-1}, \theta) \quad (2)$$

Você deve ter fórmulas para  $P(i_t | x_t, \theta)$  e  $P(x_t | x_{t-1}, i_{t-1}, \theta)$  da Parte 1 da lista.

Rust usa o seguinte procedimento:

**Passo 1** Estime  $\hat{\theta}_3$  usando

$$\ell^1(x_1, \dots, x_T, i_1, \dots, i_T | x_0, i_0, \theta) = \prod_{t=1}^T P(x_t | x_{t-1}, i_{t-1}, \theta_3)$$

**Passo 2** Use as estimativas  $\hat{\theta}_3$  como valores consistentes e a estimativa  $(\hat{\theta}_1, \hat{R})$  usando

$$\ell^2(x_1, \dots, x_T, i_1, \dots, i_T | \theta) = \prod_{t=1}^T P(i_t | x_t, \theta)$$

Para implementar este método você precisará calcular  $EV(x_t)$  para cada valor de  $\theta$  considerado o algoritmo de máxima verossimilhança. Portanto, o seu código deve ser estruturado para ter um algoritmo de ponto fixo “interno” que calcula  $EV(x_t)$  (usando a expressão derivada na parte **k**) e um algoritmo de busca “externo” que maximiza a função de verossimilhança. Um bom ponto de partida pode ser o *payoff* do período ou mesmo o valor de  $EV(x_t)$  computado na iteração anterior da estimação.

**Passo 3** Usando  $(\hat{\theta}_1, \hat{R}, \hat{\theta}_3)$  como um ponto inicial, use a função de verossimilhança completa para obter uma estimativa eficiente dos parâmetros.

- n)** Estime  $(\theta_1, R, \theta_3)$  usando a metodologia de Rust de maximização da função de verossimilhança na equação 2 (ou uma função log-verossimilhança equivalente).<sup>2</sup> Você deve comparar os números que estimou com os obtidos por Rust no seu paper (a tabela relevante é Table IX).

---

<sup>2</sup>Você pode cair em um problema de estimação porque sua linguagem de programação aproxima números bem pequenos para zero. Este é um problema quando estiver calculando (como você verá)  $\log(\exp(x_1) + \exp(x_2))$  onde tanto  $x_1$  ou  $x_2$  (ou ambos) são números negativos bem grandes. Nesse caso, se sua linguagem de programação soluciona  $\exp(x_1) = 0$ , então  $\log(\exp(x_1) + \exp(x_2)) = x_2$ , que é errado. Você pode encontrar uma descrição detalhada deste problema em <https://lips.cs.princeton.edu/computing-log-sum-exp/>.

- o) Explique como você pode calcular os erros padrão para estes coeficientes.
- p) (\*) Calcule o erro padrão usando a metodologia que você descreveu no item o).
- q) (\*) Rust estimou diferentes versões do seu modelo, variando (i) a função custo, (ii) o parâmetro  $\beta$  e (iii) o número de valores possíveis para a variável de estado  $x_t$ . Ele relata alguns resultados para todas parametrizações diferentes do modelo. Baseado apenas no que ele apresenta no paper, argumente quais destas três hipóteses possuem maior consequência. Descreva e motive uma forma de relaxar a hipótese (cabe a você desenvolver isto e fique a vontade para usar uma das expansões apresentadas por Rust). Refaça a estimação com esta hipótese relaxada. Comente sobre qualquer diferença relevante nos resultados.

## 2.3 Computação II: método CCP

O método de estimação montado por Rust é “time-consuming” porque requer que você compute um ponto fixo para cada “guess” de  $\theta$ . Hotz e Miller propõem um método alternativo.

- r) Estime  $\hat{F}(x_{t+1}|x_t, i_t)$  a partir dos dados. Você pode usar qualquer método (parametric, non-parametric, etc.). Justifique sua escolha.
- s) Estime  $\hat{P}(i_t = 1|x_t)$  a partir dos dados. Novamente justifique sua escolha de método.

Agora faça

$$\tilde{V}(i_t, x_t; \theta) \equiv \mathbb{E}_{\varepsilon_{i_t, t}} [V(i_t, x_t, \varepsilon_t; \theta)]$$

Uma vez que o termo de erro é média zero, temos

$$\tilde{V}(i_t, x_t; \theta) \equiv -c(x_t, i_t; \theta_1) + \beta \mathbb{E}_{\varepsilon_{i_{t+1}, t+1}} [u(x_{t+1}, i_{t+1}; \theta) + \varepsilon_{i_{t+1}, t+1} + \beta \mathbb{E}_{\varepsilon_{i_{t+2}, t+2}} [\dots]] \quad (3)$$

tal que as variáveis são distribuídas como

$$\begin{aligned} x_{t+1} &\sim \hat{F}(\cdot | x_t, i_t) \\ i_{t+1} &\sim \hat{P}(\cdot | x_{t+1}) \\ x_{t+2} &\sim \hat{F}(\cdot | x_{t+1}, i_{t+1}) \end{aligned}$$

---

A solução consiste essencialmente em tirar  $\exp(x_1)$  da soma do log:

$$\begin{aligned} \log(\exp(x_1) + \exp(x_2)) &= \log(\exp(x_1) \cdot (1 + \exp(x_2 - x_1))) \\ &= x_1 + \log((1 + \exp(x_2 - x_1))) \end{aligned}$$



e as expectativas são condicional, significando que elas são tomadas mantendo a história fixa (e conhecida). As propriedades da distribuição valor extremo Tipo I nos dizem que

$$\mathbb{E} [\varepsilon_{i_t,t} | i_t, x_t] = \gamma - \log (P (i_t | x_t))$$

tal que  $\gamma$  é a constante de Euler.

- t) Use um argumento recursivo para escrever  $\tilde{V} (i_t, x_t; \theta)$  como uma soma infinitamente descontada (infinite discounted sum).

Você deve ser capaz de calcular a expressão que você derivou para  $\tilde{V} (i_t, x_t; \theta)$  usando simulação numérica. Faça  $\{(x_t^s, i_t^s)\}_{t=1, s=1}^{T,S}$  serem os valores simulados de  $x_t$  e  $i_t$ . Então você pode calcular  $\tilde{V}^s (i_t^s, x_t^s; \theta)$  e obter uma estimativa consistente de

$$\tilde{V}^{sim} (i, x; \theta) = \frac{1}{S} \sum_{s=1}^S \tilde{V}^s (i_t^s, x_t^s; \theta)$$

Seu  $\tilde{V}^{sim}$  deve ser uma função de  $\theta$  que é essencialmente calculado para cada “guess.” Observe que este passo deve fazer o estimador mais rápido (uma vez que você não precisa computar a iteração de ponto fixo toda vez que tiver um novo “guess” de  $\theta$ . Entretanto, o quão rápido será o seu estimador depende do quão “objetivo” (esperto) será a sua escolha de simulação. Em particular tenha em mente o seguinte: (i) é mais rápido sortear um número de uma matriz existente do que simular um número; e (ii) seus sorteios de  $(x_t^s, i_t^s)$  não dependem dos parâmetros que você está estimando.

Para concluir o procedimento, relembre do começo que usando a hipótese valor extremo T1

$$\tilde{P} (i_t = 1 | x_t; \theta) = \frac{\exp (\tilde{V} (x_t, i_t = 1; \theta))}{\exp (\tilde{V} (x_t, i_t = 0; \theta)) + \exp (\tilde{V} (x_t, i_t = 1; \theta))}$$

Isto fornece um  $\tilde{P}$  simulado para cada “guess” de  $\theta$  e agora o que falta fazer é encontrar a condição de momento para a estimação (basicamente qualquer coisa que é baseado em  $\|\tilde{P} - \hat{P}\| = 0$  irá funcionar, embora você pode pensar se irá querer uma matriz de pesos para tornar a estimação mais eficiente).

- u) Estime  $\theta_1$  e  $R$  usando a metodologia Hotz-Miller. Compare os resultados com aqueles obtidos usando o método de Rust. Comente as diferenças.
- v) Faça um gráfico de como as estimativas Hotz-Miller mudam a medida que você varia  $T$  e  $S$ . Para quais níveis de  $T$  e  $S$  suas estimativas ficam estáveis? Por que novos aumentos no valor de  $T$  não mudam suas estimativas?