

Projeto de Dissertação - Mestrado FGV-EESP

Ricardo Semião e Castro

2025-04-28

1 Introdução e Etapas

O objetivo geral da tese é estudar, para modelos de mudanças de regimes (RS), a capacidade de corretamente identificarem regimes, em diferentes processos geradores de dados, via simulações de Monte Carlo.

O projeto está em fase inicial, mas as etapas principais são descritas abaixo.

Etapla 1: Definir os processos geradores de regimes (PGR) a serem considerados.

- A lista inicial inclui Markov Switching, Thresholds (e self-exciting thresholds), Structural Breaks, Clustering (não-supervisionado), smooth transition, e modelos baseados em random forests.
- Para cada PGR, será necessário definir alguns hiperparâmetros, sendo o mais importante o número de regimes. Inicialmente, serão estudados casos com 2 e 3 regimes.

Etapla 2: Definir os processos geradores de dados (PGD) a serem considerados.

- O foco inicial será em modelos autorregressivos, univariados ou multivariados (AR ou VAR). A combinação PGR + PGD resultará no modelo final, como MS-AR e MS-VAR. Outros modelos, como ARMA/VARMA e modelos de volatilidade endógena, também podem ser considerados. Para simplificar, o estudo pode focar em apenas um PGD (ex.: AR).
- Para cada PGD, será necessário definir parâmetros como número de lags, variáveis, observações e coeficientes (estacionariedade, persistência alta ou baixa, relação entre variáveis). Farei uma análise informal sobre a relevância de cada, e darei mais atenção apenas ao que for importante.

Etapla 3: Definir o algoritmo de simulação. Em termos gerais:

Algorithm 1 Simulação de Modelos de Regime

```
1: for  $sim = 1$  to  $n_{sims}$  do
2:   for each  $pgd$  in  $pgds$  do
3:     for each  $pgr$  in  $pgrs$  do
4:       for each  $mod$  in  $modelos$  do
5:         Gerar  $n_{obs}$  dos dados de acordo com  $pgr + pgd$ 
6:         Estimar  $mod$  na porção  $1 : n_{train}$  dos dados gerados
7:         Salvar os dados, resultados do modelo, e previsão para  $(n_{train} + 1) : n_{obs}$ 
8:       end for
9:     end for
10:   end for
11: end for
```

Etapla 4: Estudar a performance de identificação dos regimes e definir métricas descritivas para a análise.

- Para cada pgd , avaliar como cada modelo mod identifica os regimes gerados por cada pgr .
- Comparar a variável indicadora de regime estimada com a verdadeira.

- Um ponto inovador do trabalho será definir métricas que descrevam a *natureza* dos regimes gerados, para entender melhor as diferenças entre os modelos.

Etapa 5: Estudar a performance de previsão dos modelos.

- Avaliar como cada modelo *mod* prevê os regimes e os dados gerados por cada *pgr*.
- Tentar estabelecer uma relação entre performance de identificação de regimes e de previsão. Quais características dos regimes “precisam ser capturadas” para estarem associadas a uma boa previsão, e quais não são relevantes?
- Existe evidência para um aproximador universal? Ou seja, existe um modelo que performa bem em todos os casos?
- Prescrever uma tabela padrão de estatísticas descritivas a ser feita em trabalhos de mudanças de regime, que adicione qualidade à análise dos resultados. Adicionalmente, criar um pacote em R para a geração dessas tabelas.

Etapa 6: Aplicar os conhecimentos adquiridos empiricamente.

- Definir uma variável de interesse, a princípio no tema de índices de sentimento ou incerteza.
- Buscar informações institucionais e empíricas sobre o contexto de mudança de regimes relacionado a essa variável.
- Estimar diferentes modelos de RS e comparar os resultados com base no conhecimento obtido. A ideia de um aproximador universal será especialmente relevante.

Outros:

- Realizar uma revisão de literatura para motivar o trabalho e construir um conjunto de conhecimento técnico sobre o funcionamento e a performance de cada modelo. Isso ajudará a identificar lacunas e formular hipóteses.
- Concluir o trabalho com uma análise final.
- Criar um código 100% reproduzível e disponibilizá-lo [no meu GitHub](#).

2 Métricas de Comparação de Regimes

Estudarei um conjunto amplo de métricas descritivas para identificar as mais relevantes na comparação de regimes. Algumas ideias iniciais:

Estatísticas básicas:

- Momentos da(s) variável(is) de interesse:
 - Média, variância, covariância, etc.
 - ACF, PACF, CCF.
- Duração média.
- Qualidade do fit: R^2 , MSE, critérios de informação.
- Força de causalidade de Granger.
- Métricas específicas de modelo: probabilidades de transição.

Específicas de tema:

- Relação com regimes institucionais:
 - Dominância fiscal ou monetária, crescimento ou recessão, etc.
 - Correlação (ou mais geral) com variáveis de interesse.

Estatísticas específicas:

- Métricas baseadas nas observações de cada regime:
 - Distâncias dos centroides: índice Davies–Bouldin, silhueta.
 - Distâncias de distribuição das variáveis: Earth mover's.
- Relevância de variáveis em cada regime: pesos de PCA, % de explicação da variância, etc.
- Métricas baseadas nos parâmetros de cada regime:
 - Testes de diferença entre parâmetros.
 - Comparação de ACFs e IRFs implicadas pelos parâmetros.
 - * $\Delta_{r_1}^{r_2} P = \Phi^{r_1} - \Phi^{r_2}$, possivelmente calculado linha/coluna a linha/coluna.
 - * $\Delta_{r_1}^{r_2} \rho_i(l; \Phi^r)$, acumulado alternativo: $\sum_{l=1}^H \Delta_{r_1}^{r_2} \rho_i(l; \Phi^r)$.
 - * $\Delta_{r_1}^{r_2} IRF_{ij}(h; \Phi^r)$, acumulado, descontado, alternativo: $\sum_{h=1}^H \beta^h \Delta_{r_1}^{r_2} IRF_{ij}(h; \Phi^r)$.