



Trabalho de Conclusão de Curso

**Controle Estatístico de Processo Utilizando
Modelos de Vetores Autorregressivos (VAR)**

Renan Faraon Cintra

16 de dezembro de 2018

Renan Faraon Cintra

**Controle Estatístico de Processo Utilizando Modelos de
Vetores Autorregressivos (VAR)**

Trabalho de Conclusão apresentado à comissão de Graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

Orientadores:
Prof. Dr. Danilo Marcondes Filho
Prof. Dr. Márcio Valk

Porto Alegre
Dezembro de 2018

Renan Faraon Cintra

**Controle Estatístico de Processo Utilizando Modelos de
Vetores Autorregressivos (VAR)**

Este Trabalho foi julgado adequado para obtenção dos créditos da disciplina Trabalho de Conclusão de Curso em Estatística e aprovado em sua forma final pela Orientadores e pela Banca Examinadora.

Orientadores:

Prof. Dr. Danilo Marcondes Filho, UFRGS
Doutor pela Universidade Federal do Rio Grande do Sul, Porto Alegre, Brasil

Prof. Dr. Márcio Valk, UFRGS
Doutor pela Universidade Estadual de Campinas, Campinas, Brasil

Banca Examinadora:

Profa. Dra. Liane Werner, UFRGS
Doutora pela Universidade Federal do Rio Grande do Sul, Porto Alegre, RS

Porto Alegre
Dezembro de 2018

Resumo

Processo em batelada é um tema corrente na área de controle estatístico de processo (CEP), sendo um dos procedimentos mais empregados na indústria. Por apresentar uma estrutura de dados peculiar em *three-way*, (N bateladas $\times K$ variáveis $\times T$ instantes), há uma literatura específica voltada ao desenvolvimento de cartas de controle para esse tipo de processo. Nesse sentido, grande parte dos estudos aplica técnicas estatísticas multivariadas com objetivo de reduzir a sua estrutura a um formato *two-way*. Mais recentemente, abordagens que utilizam modelos de séries temporais multivariados têm emergido no contexto de CEP. Inserido nessa temática, o presente trabalho tem como objetivo estudar o desempenho das cartas T^2 de Hotelling e W da Variância Generalizada num estudo de caso simulado bivariado de processos em batelada gerados sob um estrutura de Vetores Autorregressivos (VAR) de *lag* 1. Com a modelagem via VAR, pretendemos captar a estrutura temporal do processo e, posteriormente, analisar os resíduos gerados a fim de identificar descon- troles. Dois cenários de descon- troles no processo são testados: *i*) sob a correlação dos resíduos ; *ii*) sob a estrutura de autocorrelação e correlação cruzada das variáveis. Os resultados evidenciam um bom desempenho de ambas as cartas utilizadas, apesar de a carta W ter se sobressaído a T^2 .

Palavras-Chave: Controle Estatístico de Processo, Modelos de Vetores Autorregressivos, Cartas de Controle, Carta de Controle de Hotelling, Carta de Controle da Variância Generalizada, Processo em batelada.

Abstract

Batch process is a current topic in statistical control process(SCP) area, being one of the most used procedures in the industry. Due to its peculiar *three-way* data structure, (N batch \times K variables \times T instants), there is a specific literature focused on the development of control charts for this type of process. In this sense, most of the studies apply multivariate statistical techniques with the aim of reducing its structure to a *two-way* format. More recently, approaches using multivariate time series models have emerged in the context of SCP. The objective of this paper is to study the performance of the T^2 Hotelling Chart and W Generalized Variance Chart in a bivariate simulated case study of batch processes generated under an Autoregressive Vectors (VAR) structure of *lag* 1. With VAR modeling, we intend to capture the temporal structure of the process and, later, to analyze the generated residuals in order to identify disturbances. Two disturbance scenarios in the process are tested: *i*) under the residuals' correlation; *ii*) under the variables structure of autocorrelation and cross-correlation. The results presented a good performance of both used charts, although W has shown better performance than T^2 .

Keywords: Statistical Process Control, Vector Autoregressive Model, Control Charts, Hotelling Chart, Generalized Variance Chart, Batch process.

Sumário

1	Introdução	9
2	Objetivos	11
2.1	Objetivo Geral	11
2.2	Objetivos Específicos	11
3	Processos em Batelada	12
4	Cartas de Controle Multivariada	14
4.1	Carta de Controle de Hotelling	14
4.2	Carta de Controle de Variância Generalizada	15
5	Modelos Vetores Autoregressivos (VAR)	16
6	Metodologia - Controle Estatístico de Processo Multivariado via VAR	19
7	Estudo de Caso Simulado	22
7.1	Simulação: Descontroles Testados - Cenários	23
8	Resultados	25
8.1	Cenário 1 - Alteração da Estrutura de Covariância dos Resíduos	25
8.2	Cenário 2 - Alteração da Estrutura de B_1	27
8.2.1	Alteração da Estrutura de B_1 — Autocorrelação de X_t defasado de lag 1, para diferentes ρ fixos	27
8.2.2	Alteração da Estrutura de B_1 — Correlação Cruzada entre X_t e Y_t defasado de lag 1, para diferentes ρ fixos	28
9	Considerações Finais	30
	Referências Bibliográficas	31

Lista de Figuras

Figura 3.1: Abordagens para estrutura de dados de processo em batelada. Adaptado de Choi et al. (2008)	13
---	----

Lista de Tabelas

Tabela 8.1:	Percentual médio e desvio padrão de batealdas perturbadas identificadas — Cartas de T_e^2 de Hotelling e da Variância Generalizada W_e para \mathbf{B}_1^0 e \mathbf{B}_1 , variando ρ . Número de repetições: 200 repetições.	26
Tabela 8.2:	Percentual médio e desvio padrão de batealdas perturbadas identificadas — Cartas de T_e^2 de Hotelling e da Variância Generalizada W_e para $\mathbf{B}_{11,\rho}$, variando a	27
Tabela 8.3:	Percentual médio e desvio padrão de batealdas perturbadas identificadas — Cartas de T_e^2 de Hotelling e da Variância Generalizada W_e para $\mathbf{B}_{21,\rho}$, variando a	28

1 Introdução

O controle estatístico de processo (CEP) engloba um conjunto de técnicas para analisar e avaliar a qualidade de processos industriais, sendo as cartas de controle sua principal ferramenta. As cartas de controle são amplamente utilizadas para monitoramento de variáveis (de processo e de qualidade do produto acabado) em uma ampla gama de indústrias. Parte dessas indústrias conduzem seus processos por bateladas.

O CEP voltado para processos que ocorrem por bateladas passou a ganhar atenção especial no início da década de 1990. De lá pra cá, cartas de controle mais sofisticadas vêm sendo desenvolvidas e adaptadas para atender as peculiaridades presentes nos dados dos processos desta natureza. Estas abordagens buscam um monitoramento mais preciso do estado do processo, permitindo identificação de eventos atípicos em tempo real.

Sendo um dos processos mais amplamente difundidos nas indústrias de transformação, especialmente na produção química, bioquímica e farmacêutica, o processo em batelada caracteriza-se por ser recarregável, de modo que há intermitências em sua produção. Nesse sentido, no contexto de Controle Estatístico de Qualidade, processos em batelada apresentam uma peculiaridade em sua estrutura de dados, pois, além das variáveis em cada instante que compõem o processo, evidencia-se variabilidade entre bateladas.

Dessa forma, esses processos apresentam uma estrutura de dados *three-way* (N bateladas $\times K$ variáveis $\times T$ instantes de tempo). As principais abordagens de controle de tais processos envolvem a aplicação de técnicas estatísticas multivariadas (tais como, Componentes Principais, Redes Neurais, Regressão Multivariada,...) numa estrutura de dados *two-way*, que por sua vez possui duas formas clássicas. O arranjo ($N \times KT$) prioriza na análise o comportamento dinâmico das variáveis, isto é, as autocorrelações e correlações cruzadas presentes nos dados. O trabalho precursor desta abordagem é o de [Nomikos e MacGregor \(1995\)](#). [Wold et al. \(1998\)](#) apresentam uma abordagem alternativa usando o arranjo ($NT \times K$). Este arranjo prioriza a análise no domínio das variáveis e permite diagnosticar distúrbios no processo, identificando as variáveis responsáveis. [Camacho et al. \(2009\)](#) apresentam uma discussão abrangente destas abordagens.

Um dos principais recursos estatísticos utilizados no controle de qualidade para este tipo de processos é o de Análise de Componentes Principais (ACP). [Camacho et al. \(2009\)](#) desenvolveram uma análise para verificar a influência das transformações de *three-way* para o *two-way* no desempenho do monitoramento da qualidade de bateladas. Dentre suas conclusões, estão a de que modelos mais parcimoniosos

alcançam bons desempenhos e que a identificação das mudanças na estrutura de correlação é crucial para modelagem.

Trabalhos mais recentes propõem o uso de modelos de séries temporais no contexto do Controle Estatístico Multivariado de Processo, mais especificamente, o Modelo de Vetores Autoregressivos (VAR, do inglês Vector Autoregressive Model). Entretanto, os trabalhos estão restritos ao monitoramento de processos contínuos, isto é, cuja estrutura natural de dados é *two-way* do tipo $N \times K$. Nesta estrutura, não há dimensão temporal, pois não há repetição de medições dentro de cada uma das n amostras. Estes trabalhos basicamente utilizam a modelagem VAR das amostras numa fase preliminar para eliminar o efeito de correlação serial nos dados. Numa etapa seguinte cartas de controle tradicionais são configuradas para monitorar os resíduos resultantes do ajuste do modelo VAR. Uma boa descrição de tais abordagens pode ser encontrada em [Vanhatalo e Kulahci \(2015\)](#) e [Pan e Jarrett \(2012\)](#).

Este trabalho propõe uma abordagem baseada no trabalho de [Choi et al. \(2008\)](#) e apresenta três contribuições em relação a abordagem destes autores. A primeira está no método utilizado para estimar os coeficientes do modelo VAR na fase I do controle (fase de construção das cartas de controle). A segunda está na utilização da carta de controle da Variância Generalizada em conjunto com a tradicional carta de Hotelling na fase II do controle (monitoramento de amostras futuras). Finalmente, o presente trabalho traz como terceira contribuição um estudo simulado amplo permitindo uma avaliação consistente da abordagem proposta.

Além desta introdução, este estudo está organizado da seguinte forma: no capítulo 2 os objetivos gerais e específicos são introduzidos; no capítulo 3, há uma explicação mais detalhada sobre processos em batelada e a abordagem a ser utilizada; no capítulo 4, apresenta-se a descrição acerca das cartas de controle utilizadas - Hotelling e Variância Generalizada; posteriormente, no capítulo 5, o modelo de Vetores Autorregressivos (VAR) é exposto; a abordagem proposta no contexto de Controle Estatístico de Processos usando VAR é explicado no capítulo 6; no capítulo 7 explicita-se o estudo de caso proposto, com os cenários de descontrole que serão vistos; o capítulo 8 descreve os resultados obtidos; e, por fim, tem-se as considerações e discussões finais.

2 Objetivos

2.1 Objetivo Geral

Estudar uma abordagem para controle de processos em bateladas utilizando modelos de Vetores Autorregressivos (VAR).

2.2 Objetivos Específicos

- Propor uma abordagem de controle utilizando a estatística da Variância Generalizada para monitorar os resíduos de um modelo VAR.
- Através de um estudo simulado, com diversos descontroles impostos no processo, comparar o desempenho desta abordagem em relação à abordagem baseada na estatística de Hotelling para monitorar os resíduos do VAR (esta última proposta por [Choi et al. \(2008\)](#)).

3 Processos em Batelada

O trabalho ora proposto tem como motivação inicial os processos em batelada, no contexto industrial de monitoramento e controle estatístico de qualidade. Especificamente, no que se refere ao meio industrial, a dinâmica de um processo em batelada apresenta uma dinâmica peculiar, que pode ser explicada a partir de três fases: (i) um conjunto de matérias-primas é colocada em um recipiente; (ii) as matérias-primas sofrem uma série de transformações decorrentes de determinados estímulos cujas características podem ser monitoradas, como temperatura, pressão e concentração; e (iii), por último, tem-se o produto final, que é analisado segundo as especificações desejadas pelo planejamento, devendo-se retornar ao ponto (i) para continuidade do processo (Marcondes Filho, 2001).

Assim, como afirmam Ge et al. (2013), a modelagem estatística do processo em batelada deve levar em conta a variabilidade entre as bateladas. Isso porque, a cada reinicialização do processo, deve-se atingir o ponto de operação novamente, o que significa que o processo está sujeito a uma variação em torno do ponto inicial “ótimo” de operação. Essa especificidade dos processos em batelada faz com que sua estrutura de dados seja no formato *three-way* (N bateladas \times K variáveis \times T instantes de tempo).

Na literatura, como afirmam Choi et al. (2008), destacam-se duas abordagens de arranjo dos dados com objetivo a contemplar a forma matricial *two-way*, principalmente tendo em vista a técnicas de estatística multivariada. Um primeiro arranjo seria no formato ($N \times KT$), em que a direção das bateladas é preservado (Nomikos e MacGregor, 1994); o segundo, trata-se de manter a direção das variáveis, de modo a estruturar os dados ($NT \times K$) (Wold et al., 1998).

A Figura 3.1 demonstra, primeiro, o arranjo para o método que preserva a direção da bateladas, e que está assinalado com a letra a , e, o segundo, referenciado com b , em que os dados estão empilhados em colunas com as variáveis medidas. Para o primeiro arranjo, privilegia-se a análise do comportamento dinâmico das variáveis, notadamente as autocorrelações e correlações cruzadas dos dados. Por outro lado, o segundo arranjo possibilita diagnosticar distúrbios no processo, identificando as variáveis responsáveis.

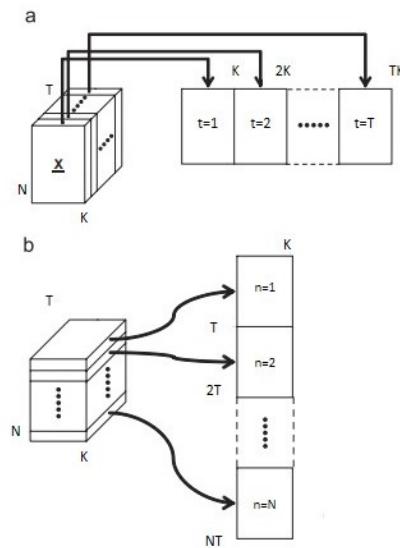


Figura 3.1: Abordagens para estrutura de dados de processo em batelada. Adaptado de Choi et al. (2008)

Ainda, a Figura 3.1 ressalta que as principais abordagens de controle de processo em batelada envolvem a aplicação de técnicas multivariadas de redução de dados — sobretudo, Análise de Componente Principais (ACP) —, em desdobramentos da estrutura *three-way* numa matriz de dados *two-way*. A ACP foca exclusivamente em caracterizar a estrutura de correlação das variáveis de processo. Adicionalmente, a grande maioria destas proposições priorizam o desdobramento descrito na Figura 3.1 a e a técnica ACP para redução dos dados. O arranjo $N \times KT$ (*batch-wise*) prioriza na análise o comportamento dinâmico das variáveis, isto é, as autocorrelações presentes nos dados.

Nesse sentido, as abordagens tradicionais para controle de processos em bateladas utilizam basicamente duas cartas de controle: (i) A carta de controle baseada na clássica estatística de Hotelling (T^2) monitora os escores gerados da projeção dos dados — via ACP no arranjo $(N \times KT)$ — num plano de dimensão muito menor em relação ao número K de variáveis originais. A carta T^2 permite a cada nova batelada o monitoramento conjunto das trajetórias médias das K variáveis; (ii) a carta de controle Q , que monitora os resíduos do modelo ACP, isto é, detecta algum evento atípico que tenha alterado a estrutura de correlação padrão dos dados modelados via ACP.

Especificamente para este trabalho, baseamo-nos na metodologia apresentada por Choi et al. (2008), que segue os seguintes passos: inicialmente, utiliza-se o modelo de Vetores Autorregressivos (VAR) para captar a estrutura temporal do processo, e, posteriormente, analisar os resíduos gerados a partir da modelagem de organizadas segundo a Figura 3.1 b, a fim de identificar os descontroles. Entretanto, neste trabalho, propomos uma carta de controle baseada na estatística da Variância Generalizada (W) em conjunto com a carta T^2 proposta por esses autores.

A abordagem levará em conta um estudo de caso simulado, com VAR(1) - isto é, com defasagem igual a 1 - bivariado. Na detecção dos descontroles, serão usadas as metodologias das cartas de T^2 de Hotelling e W da Variância Generalizada.

4 Cartas de Controle Multivariada

Neste capítulo, apresentaremos as duas cartas de controle utilizadas ao longo deste trabalho: T^2 de Hotelling e W da Variância Generalizada. Cabe notar que a primeira é uma das abordagens mais corrente no contexto de CEP, sendo largamente utilizada na literatura e na indústria. Por sua vez, a pertinência de se utilizar a segunda carta assenta-se na necessidade de se monitorar a estrutura de covariância dos resíduos dada a aplicação de uma modelagem multivariada de séries temporais, descrita nas seções seguintes. As adaptações ao contexto específico deste estudo serão apresentadas no capítulo 6.

4.1 Carta de Controle de Hotelling

A carta de controle baseada na estatística T^2 de Hotelling (Montgomery, 2007) é uma das mais empregadas para o monitoramento de vetor de médias. Sendo um desdobramento direto da carta univariada \bar{x} de Shewart, para a carta de Hotelling, parte-se de um processo com K variáveis de interesse distribuídas segundo uma Normal K variada, com vetor de médias $\boldsymbol{\mu}_0$ e matriz de covariância $\boldsymbol{\Sigma}_0$. Levando-se em conta o contexto de controle estatístico de qualidade, tem-se que, na fase I do processo de controle, $\boldsymbol{\mu}_0$ e $\boldsymbol{\Sigma}_0$ são estimados através de amostras históricas independentes obtidas do processo sob controle estatístico, isto é, com estrutura de médias e covariâncias não contaminadas por nenhum evento atípico presente no processo. Considerando amostras de tamanho unitário, na fase II de controle, T^2 possui distribuição F . Tem-se assim:

$$T^2 = (\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \sim \frac{K(n+1)(n-1)}{n(n-K)} F_{K, n-K} \quad (4.1)$$

onde $\bar{\mathbf{x}}$ e \mathbf{S} representam respectivamente o vetor de médias a matriz de covariâncias amostrais obtidos a partir de n amostras preliminares das K variáveis sob análise. Para cada nova amostra do processo o escore T^2 é obtido através de (4.1). Escore acima do percentil α indica que a nova observação pode ter sido gerada sob um processo com $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ ou $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$. Em termos práticos, podemos entender que a amostra foi obtida durante o processo em que as variáveis (ou um subconjunto delas) apresentaram valores discrepantes em relação as suas médias ou apresentaram covariâncias distantes do esperado no processo sob controle.

Especificamente no segundo caso, mudanças apenas na estrutura de covariância das variáveis no processo podem não ser detectadas pela carta de controle T^2 . A

estatística de Hotelling computa a distância multivariada entre vetores de observações e o vetor estimado de médias, dado uma estrutura de covariância fixa, isto é, não compara diretamente diferenças entre matrizes de covariância.

4.2 Carta de Controle de Variância Generalizada

Para monitoramento explícito da estrutura de covariância do processo utiliza-se a carta de controle baseada na *Variância Generalizada*. Num processo sob controle, considere n amostras preliminares de tamanho $m > 1$. Na fase II, novas amostras são monitoradas através da carta de controle W (Montgomery, 2007). Temos:

$$W = -Km + Km \ln(m) - m \ln\left(\frac{\det(\mathbf{V})}{\det(\mathbf{S})}\right) + \text{tr}(\mathbf{S}^{-1}\mathbf{V}) \sim \chi_{K(K+1)/2}^2, \quad (4.2)$$

onde \mathbf{S} representa a matriz de covariâncias amostrais obtidos a partir de n amostras preliminares, $\mathbf{V} = (m-1)\mathbf{S}_{new}$ representa a matriz de covariâncias da nova observação (i -ésima amostra) de tamanho $m > 1$. Dessa forma, para cada nova amostra do processo o escore W é obtido através de (4.2). Escore acima do percentil α indica que a nova observação pode ter sido gerada sob um processo com $\Sigma \neq \Sigma_0$. De maneira prática, entende-se que a amostra sinaliza uma estrutura de covariância significativamente diferente da estrutura descrita na matriz \mathbf{S} . Uma descrição ampla incluindo exemplos de aplicação de Cartas de Controle Multivariadas pode ser encontrada em Johnson e Wichern (2007) e Montgomery (2007).

5 Modelos Vetores Autorregressivos (VAR)

Uma direção de pesquisa em CEP que vem crescendo é aplicação da teoria de séries temporais multivariadas baseadas em sua maioria nos modelos Vetores Autorregressivos (VAR), [Pan e Jarrett \(2012\)](#). Como uma extensão natural do modelo autorregressivo univariado, o VAR, proposto por [Sims \(1980\)](#), é especialmente útil para descrever conjuntamente o comportamento dinâmico de séries temporais. Esse modelo permite expressar um modelo autorregressivo de ordem L por um vetor de K variáveis endógenas, \mathbf{Z}_t , que estão conectadas entre si por meio de uma matriz \mathbf{A} , conforme a equação

$$\mathbf{AZ}_t = \mathbf{B}_0 + \mathbf{B}_1\mathbf{Z}_{t-1} + \mathbf{B}_2\mathbf{Z}_{t-2} + \cdots + \mathbf{B}_k\mathbf{Z}_{t-L} + \boldsymbol{\varepsilon}_t, \quad (5.1)$$

em que \mathbf{B}_0 é uma matriz $n \times 1$ de constantes, \mathbf{B}_j , $j = 1, \dots, p$ são matrizes $n \times n$ de coeficientes do modelo, $\boldsymbol{\varepsilon}_t$ é um vetor $n \times 1$ com as características de um ruído branco, o que significa que $\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$, $\text{Var}[\boldsymbol{\varepsilon}_t] = \boldsymbol{\Sigma}$ e $\text{Cov}[\boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_{t+h}] = \mathbf{0}$, para todo $h = 1, 2, 3, \dots$. A equação (5.1) é conhecida como a forma estrutural do modelo VAR (SVAR) e expressa as relações entre as variáveis endógenas ([da Silveira Bueno, 2008](#)). Note que nessa definição (5.1) a matriz $\boldsymbol{\Sigma}$ é diagonal, ou seja, não existe correlação contemporânea ou temporal entre os componentes do vetor de choques $\boldsymbol{\varepsilon}$. As inter-relações entre os choques são capturadas indiretamente pela matriz \mathbf{A} .

A título de simplificação, considere o modelo autorregressivo bivariado estrutural SVAR(1) com o vetor de variáveis $\mathbf{Z}_t = [X_t, Y_t]'$. A equação (5.1) pode ser então escrita da seguinte forma:

$$\begin{bmatrix} 1 & a_{12} \\ a_{21} & 1 \end{bmatrix} \begin{bmatrix} X_t \\ Y_t \end{bmatrix} = \begin{bmatrix} b_{0;1} \\ b_{0;2} \end{bmatrix} + \begin{bmatrix} b_{1;11} & b_{1;12} \\ b_{1;21} & b_{1;22} \end{bmatrix} \begin{bmatrix} X_{t-1} \\ Y_{t-1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{1,t} \\ \boldsymbol{\varepsilon}_{2,t} \end{bmatrix}, \quad (5.2)$$

em que

$$\boldsymbol{\Sigma} = \mathbb{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

Na expressão (5.2) podemos escrever as equações individualmente, da forma

$$X_t = b_{0;1} - a_{12}Y_t + b_{1;11}X_{t-1} + b_{1;12}Y_{t-1} + \boldsymbol{\varepsilon}_{1,t}; \quad (5.3)$$

$$Y_t = b_{0;2} - a_{21}X_t + b_{1;21}X_{t-1} + b_{1;22}Y_{t-1} + \boldsymbol{\varepsilon}_{2,t},$$

onde a_{12} e a_{21} denotam a correlação contemporânea entre X_t e Y_t ; $b_{1;11}$ e $b_{1;22}$ são coeficientes da autocorrelação das variáveis X_t e Y_t , respectivamente, com defasagem 1; $b_{1;12}$ e $b_{1;21}$, coeficientes de correlação cruzada de X_t e Y_t com defasagem 1.

Na expressão (5.3) fica explícito que as variáveis são mutuamente influenciadas uma pela outra. Dessa forma, ambas as variáveis contemporâneas X_t e Y_t são individualmente correlacionadas com os erros $\mathcal{E}_{1,t}$ e $\mathcal{E}_{2,t}$. Essa endogeneidade impossibilita uma estimação não viesada dos parâmetros diretamente da forma estrutural. Esse problema pode ser resolvido escrevendo-se o modelo VAR em sua forma reduzida:

$$\begin{aligned}\mathbf{Z}_t &= \mathbf{A}^{-1}\mathbf{B}_0 + \sum_{j=1}^L \mathbf{A}^{-1}\mathbf{B}_j\mathbf{Z}_{t-j} + \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t \\ &= \boldsymbol{\Phi}_0 + \sum_{j=1}^L \boldsymbol{\Phi}_j\mathbf{Z}_{t-j} + \mathbf{e}_t,\end{aligned}\quad (5.4)$$

em que $\boldsymbol{\Phi}_j = \mathbf{A}^{-1}\mathbf{B}_j$, para $j = 0, 1, \dots, L$ e $\mathbf{e}_t = \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t$. Assim, na forma reduzida, todas as variáveis do lado direito da equação são pré-determinadas, possibilitando estimar diretamente o modelo através de mínimo quadrados ordinários, por exemplo.

Uma característica importante dos processos VAR é sua estabilidade. Para que se tenha média, variâncias e estruturas de covariâncias invariantes no tempo é necessário verificar se

$$\det(\mathbf{I}_K - \boldsymbol{\Phi}_1 z - \dots - \boldsymbol{\Phi}_L z^L) = 0, \quad (5.5)$$

para $|z| \leq 1$, em que \mathbf{I}_K é a matriz identidade $K \times K$. Se a equação acima admite uma solução (raíz unitária) para $|z| = 1$ então uma ou várias variáveis do processo VAR(L) são integradas de ordem 1, I(1). Nesse caso é possível que as variáveis cointegrem e o modelo pode ser melhor analisado no contexto de modelos de correção de erros (VECM).

As inovações da forma reduzida do modelo VAR, descrito na equação (5.4), são compostas por choques estruturais $\mathbf{e}_t = \mathbf{A}^{-1}\boldsymbol{\varepsilon}_t$. A ocorrência de um choque estrutural $\varepsilon_{j,t}$ potencialmente pode causar um choque em todos os termos de erros $e_{j,t}$, criando assim um movimento contemporâneo em todas as variáveis endógenas. Conseqüentemente, a matriz de variâncias e covariâncias do modelo VAR na forma reduzida, que é dada por

$$\begin{aligned}\boldsymbol{\Gamma} &= \mathbb{E}(\mathbf{e}_t\mathbf{e}_t') = \mathbb{E}[\mathbf{A}^{-1}\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'(\mathbf{A}^{-1})'] \\ &= \mathbf{A}^{-1}\boldsymbol{\Sigma}(\mathbf{A}^{-1})',\end{aligned}\quad (5.6)$$

pode ter elementos diferentes de zero fora da diagonal, isto é, correlações diferentes de zero entre os termos de erro. Se consideramos um modelo VAR(1)

$$\mathbf{Z}_t = \boldsymbol{\Phi}_0 + \boldsymbol{\Phi}_1\mathbf{Z}_{t-1} + \mathbf{e}_t, \quad (5.7)$$

este será estacionário se os autovalores da matriz de coeficientes $\boldsymbol{\Phi}_1$ forem menores do que 1 em valor absoluto (Reinsel, 2003). Assumindo estacionariedade, a média do processo (5.7) é dada por

$$\mathbb{E}[\mathbf{Z}_t] = (\mathbf{I}_K - \boldsymbol{\Phi}_1)^{-1}\boldsymbol{\Phi}_0. \quad (5.8)$$

A matriz de variância e covariâncias (contemporânea) do processo (5.7) pode ser escrita como

$$\boldsymbol{\Gamma} = \boldsymbol{\Phi}_1'\boldsymbol{\Gamma}\boldsymbol{\Phi}_1 + \boldsymbol{\Sigma}, \quad (5.9)$$

em que Φ_1' é a transposta da matriz Φ_1 .

A forma explícita pode ser escrita da seguinte forma:

$$\text{Vec}\Gamma = [\mathbf{I}_{p^2} - \Phi_1 \otimes \Phi_1']^{-1} \text{Vec}\Sigma \quad (5.10)$$

onde \otimes é o produto de Kronecker e Vec é o operador que transforma a matriz em um vetor.

Essa forma deixa explícita a relação entre a correlação contemporânea do processo e a correlação nos resíduos. Se no momento de estimar o processo VAR, este for contemporaneamente correlacionado, a matriz de covariâncias dos resíduos terá elementos não nulos fora da diagonal principal.

6 Metodologia - Controle Estatístico de Processo Multivariado via VAR

Neste capítulo, apresentamos uma abordagem para Controle Multivariado de Processos em Bateladas através do uso combinado da modelagem via VAR e da carta de controle T^2 e da carta baseada na Variância Generalizada (carta W). Adicionalmente, descrevemos a carta de controle T^2 proposta por (Choi et al., 2008) neste contexto. A título de simplificação vamos considerar n bateladas de um processo com $K = 2$, X_t e Y_t . Supomos então n realizações de um processo estocástico bivariado, amostradas em T instantes de tempo igualmente espaçados. Suponha ainda que estas séries caracterizam o processo apenas com a presença de eventos típicos, ou seja, apenas causas comuns de variabilidade presentes.

Na fase I de controle, numa etapa preliminar, obtemos a “batelada média” através do cálculo da média das séries temporais amostrais de X e Y , a partir de um conjunto de bateladas de referência geradas de um processo sob controle. Na etapa seguinte, ajustamos um modelo VAR nas séries \bar{X}_t e \bar{Y}_t . Os vetores de resíduos estimados para cada batelada são arrajados e uma maneira sugerida por Choi et al. (2008), a qual prioriza o controle on-line das bateladas futuras. Em cada batelada temos então vetores $\hat{\mathbf{e}}_{x,b} = [\hat{e}_{x,L+1}^b, \dots, \hat{e}_{x_t}^b, \dots, \hat{e}_{x,T}^b]'$ e $\hat{\mathbf{e}}_{y,b} = [\hat{e}_{y,L+1}^b, \dots, \hat{e}_{y_t}^b, \dots, \hat{e}_{y,T}^b]'$, ambos vetores de dimensão $(T - L) \times 1$, para $b = 1, 2, 3, \dots, n$ bateladas e L o número de defasagens usadas no VAR ajustado. Considerando n bateladas, empilham-se os resíduos estimados numa matriz \mathbf{E} , de dimensão $(T - L)n \times 2$. De forma explícita, temos:

$$\mathbf{E} = \begin{bmatrix} \hat{\mathbf{e}}_{x,1} & \hat{\mathbf{e}}_{y,1} \\ \vdots & \vdots \\ \hat{\mathbf{e}}_{x,b} & \hat{\mathbf{e}}_{y,b} \\ \vdots & \vdots \\ \hat{\mathbf{e}}_{x,n} & \hat{\mathbf{e}}_{y,n} \end{bmatrix} = \begin{bmatrix} \mathbf{E}_1 \\ \vdots \\ \mathbf{E}_b \\ \vdots \\ \mathbf{E}_n \end{bmatrix}. \quad (6.1)$$

O modelo VAR(L) ajustado “a batelada média” captura a dinâmica das variáveis X e Y , incluindo suas autocorrelações e correlações cruzadas significativas de lag até L .

Na fase II o controle on-line de uma nova batelada em progresso é feito através da estatística T^2 e W com vetor de médias e matriz de covariâncias obtidas da matriz de resíduos \mathbf{E}_{n+1} . A cada instante, o resíduo gerado é dado por:

$$\hat{\mathbf{e}}_t = [\hat{\epsilon}_{x,t}^{n+1}, \hat{\epsilon}_{y,t}^{n+1}] = \mathbf{Z}_t - (\hat{\mathbf{B}}_0 - \hat{\mathbf{B}}_1 \mathbf{Z}_{t-1} - \dots - \hat{\mathbf{B}}_L \mathbf{Z}_{t-L}), t = L + 1, \dots, T, \quad (6.2)$$

onde $\hat{\mathbf{B}}_0$ é um vetor estimado de interceptos e $\hat{\mathbf{B}}_l$ é a matriz de autocorrelação e correlação cruzada de lag l , estimados a partir do ajuste do modelo VAR à batelada média do processo sob controle, para $l = 1$ até L .

Reescrevemos então as eqs. (4.1) e (4.2) da seguinte forma:

$$T_e^2 = (\hat{\mathbf{e}}_t - \bar{\mathbf{e}})' \mathbf{S}_e^{-1} (\hat{\mathbf{e}}_t - \bar{\mathbf{e}}) \sim \frac{K(n(T-L) + 1)(n(T-L) - 1)}{(n(T-L))(n(T-L) - K)} F_{K, n(T-L)-K} \quad (6.3)$$

e

$$W_e = -K(T-L) + K(T-L) \ln(T-L) - (T-L) \ln \left(\frac{\det(\mathbf{V})}{\det(\mathbf{S}_e)} \right) + \text{tr}(\mathbf{S}_e^{-1} \mathbf{V}) \sim \chi_{K(K+1)/2}^2, \quad (6.4)$$

onde $\bar{\mathbf{e}}$ e \mathbf{S}_e representam respectivamente o vetor de médias e a matriz de covariâncias dos vetores linha da matriz \mathbf{E} . $\mathbf{V} = (T-L)\mathbf{S}_e^{(n+1)}$, onde $\mathbf{S}_e^{(n+1)}$ representa a matriz de covariâncias dos resíduos de uma nova batelada, arranjados numa matriz \mathbf{E}_{n+1} , isto é, com a mesma estrutura da matriz \mathbf{E}_b , descrita na equação 6.1.

Em resumo, a cada instante de uma batelada, novos resíduos são obtidos e os escores T_e^2 são calculados. Já o escore W_e é obtido após a batelada completada. Dessa forma, para cada nova batelada, obtém-se $(T-L)$ escores T_e^2 e 1 escore W_e . Mais especificamente, dado uma nova batelada sob análise, escores T_e^2 com valores superiores ao percentil α da distribuição F indicam que, no atual instante da batelada em curso, as variáveis de processo apresentaram estrutura de correlação (autocorrelação ou correlação cruzada) significativamente diferente daquela prevista pelo modelo VAR ajustado. O escore W_e com valor superior ao percentil α da distribuição χ^2 indica que estrutura de covariância dos resíduos desta batelada é significativamente diferente da estrutura de covariância da matriz \mathbf{E} , representada por \mathbf{S}_e .

Além de permitir o controle on-line, o arranjo dos resíduos na matriz \mathbf{E}_{n+1} traz outra característica importante, pois prioriza a identificação de variáveis mais afetadas por descontroles apontados pela carta. Isto é possível visto que cada coluna de \mathbf{E}_{n+1} representa o conjunto dos resíduos, em todas as bateladas históricas, de uma variável em particular. Este arranjo possibilita desdobrar as estatísticas de Hotelling e da Variância Generalizada e avaliar a influência de cada variável separadamente nos escores T_e^2 e W_e . Embora a fase de diagnóstico não seja explorada neste trabalho, no controle multivariado sabemos que a identificação das variáveis causadoras dos descontroles é tão importante quanto à sua precisa detecção pela carta.

O trabalho aqui proposto se diferencia da abordagem de Choi et al. (2008) em alguns pontos. Na fase I estes autores utilizam uma regressão PLS (do inglês, Partial Least Squares) no conjunto de todas as n séries temporais amostradas. Posteriormente, os coeficientes são estimados através de estatísticas próprias da PLS. Esses coeficientes estimados são usados para obter as estimativas dos coeficientes de um VAR. Diferentemente desses autores, o presente trabalho propõe o ajuste de um modelo VAR para uma “batelada média”, obtida a partir dos dados históricos de um processo sob controle. Enfatiza-se aqui que os coeficientes do modelo VAR são estimados utilizando-se a função de máxima verossimilhança do próprio modelo VAR.

Adicionalmente, na fase II, o presente trabalho propõe o uso conjunto das cartas de controle de Hotelling e da Variância Generalizada, o que não é contemplado no trabalho de [Choi et al. \(2008\)](#). Estes empregam unicamente a carta de controle de Hotelling em escores gerados por redução de dados através de Componentes Principais.

Este trabalho se propõe investigar a sensibilidade da abordagem proposta diante de diferentes descontroles que serão impostos na estrutura de correlação e autocorrelação das variáveis descritas no modelo VAR.

7 Estudo de Caso Simulado

Neste capítulo, será apresentada a metodologia para avaliar o desempenho das cartas de controle descritas no capítulo 6. Um processo em bateladas com duas variáveis será simulado através de um modelo VAR, descrito no capítulo 5. Usando essa estruturação, diferentes cenários de descontroles serão impostos através de modificações nos parâmetros do VAR. Os parágrafos seguintes descrevem detalhadamente a estrutura de simulação.

Considere um processo em batelada com duas variáveis sob monitoramento. Suponha que o mecanismo gerador dos dados seja descrito por um modelo VAR(1) estacionário [Eq: (5.2)] com $a_{12} = a_{21} = 0$, vetor de interceptos $\mathbf{B}_0 = \mathbf{0}$ e matriz de covariância dos resíduos $\Sigma = \Sigma_0$.

Mais explicitamente, o processo em questão é descrito por VAR(1) com a seguinte estrutura:

$$\mathbf{Z}_t = \mathbf{B}_1 \mathbf{Z}_{t-1} + \boldsymbol{\varepsilon}_t, \quad (7.1)$$

No processo sob controle, adotamos as correlações e autocorrelações de lag 1 com os seguintes coeficientes:

$$\mathbf{B}_1 = \begin{bmatrix} -0,3 & 0,4 \\ 0,4 & 0,5 \end{bmatrix}, \quad (7.2)$$

onde -0,3 e 0,5 representam as autocorrelações de lag 1 de X_t e Y_t respectivamente; e 0,4 representa a correlação cruzada defasada de lag 1 entre X_t e Y_t e Y_t e X_t , respectivamente.

Na prática, na fase 1 do controle, precisamos estimar os coeficientes da matriz \mathbf{B}_1 , o vetor de médias dos resíduos e a matriz de covariância Σ_0 dos resíduos a partir de dados de bateladas históricas obtidas num processo sob controle. Entretanto, assumimos neste trabalho que as leis físicas do processo estão descritas pelo modelo VAR proposto na equação (7.1). Dessa forma, cada batelada gerada sob essas leis tem a estrutura de correlação descrita em (7.2), média dos resíduos igual a zero e covariância dos resíduos igual a Σ_0 . As cartas de controle T^2 e W propostas nas equações (6.3) e (6.4) são reescritas da seguinte forma:

$$T_e^2 = (\mathbf{e}_t)' \Sigma_{0e}^{-1} (\mathbf{e}_t) \sim \frac{K(n(T-1)+1)(n(T-1)-1)}{(n(T-1))(n(T-1)-K)} F_{K, n(T-1)-K} \quad (7.3)$$

e

$$W_e = -K(T-1) + K(T-1) \ln(T-1) - (T-1) \ln\left(\frac{\det(\mathbf{V})}{\det(\boldsymbol{\Sigma}_{\mathbf{0}_e})}\right) + \text{tr}(\boldsymbol{\Sigma}_{\mathbf{0}_e}^{-1} \mathbf{V}) \sim \chi_{K(K+1)/2}^2, \quad (7.4)$$

onde $\mathbf{V} = (T-1)\mathbf{S}_e^{(n+1)}$ e $\mathbf{S}_e^{(n+1)}$ representa a matriz de covariâncias dos resíduos de uma nova batelada, arranjados numa matriz \mathbf{E}_{n+1} (conforme descrito no capítulo 6), isto é, com a mesma estrutura da matriz \mathbf{E}_b , descrita na equação 6.1. Os escores T_e^2 são obtidos a partir dos resíduos calculados através da equação, reescritos como:

$$\mathbf{e}_t = [e_{x,t}^{n+1}, e_{y,t}^{n+1}] = \mathbf{Z}_t - \mathbf{B}_1 \mathbf{Z}_{t-1} \quad (7.5)$$

Note que os limites da carta de controle T_e^2 e W_e na fase 1 dependem apenas do número n de bateladas, da quantidade T de instantes e do número K de variáveis em análise. Neste trabalho consideraremos um processo em bateladas com medições disponíveis em $T = 50$ instantes e $n = 500$ bateladas de referência na fase 1. Dessa forma, em cada batelada, dispomos de $T - 1$ resíduos [gerados via eq.(7.5)], visto que o sistema gerador é dado por um VAR(1). Conforme descrito no início da seção, $K = 2$ variáveis sob análise.

Os limites de controle para as cartas T_e^2 e W_e são obtidos utilizando probabilidade de alarme falso de $\alpha = 0,05$ (percentil 5% na calda superior das distribuições F e χ^2 , respectivamente).

Na fase II, em cada cenário de descontrole, 500 novas bateladas foram simuladas, os resíduos foram calculados de acordo com a eq.(6.1) e os escores T_e^2 e W_e foram comparados aos limites de controle obtidos via eqs.(6.3) e (6.4). Cada cenário foi replicado 200 vezes. A seção seguinte descreve os cenários de descontrole que serão investigados neste trabalho.

7.1 Simulação: Descontroles Testados - Cenários

Neste trabalho, dois cenários serão analisados: (i) descontroles serão impostos sobre a matriz de covariâncias dos resíduos $\boldsymbol{\Sigma}_0$, para matriz \mathbf{B}_1 [eq.(7.2)] fixa; (ii) um coeficiente de autocorrelação e um coeficiente correlação cruzada da matriz \mathbf{B}_1 serão alterados, fixando a estrutura de covariância dos resíduos em três níveis.

Assim, tem-se:

- para o primeiro caso:

$$\boldsymbol{\Sigma}_{\text{faseII}} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad (7.6)$$

onde $\rho = -0,95, -0,90, \dots, 0,90, 0,95$, isto é, percorrendo a amplitude total do coeficiente de correlação entre os resíduos de X e Y , utilizando intervalos de 0,05. Além disso, neste primeiro caso, utilizamos nas simulações a seguinte matriz:

$$\mathbf{B}_1^0 = \begin{bmatrix} -0,3 & 0,0 \\ 0,0 & 0,5 \end{bmatrix}. \quad (7.7)$$

Observa-se que a matriz \mathbf{B}_1^0 representa um processo VAR onde as variáveis X_t e Y_t não apresentam correlação cruzada defasada de lag=1. Note que para $\rho = 0$ em eq.(7.6) tem-se um processo VAR gerando duas variáveis não correlacionadas. Adicionalmente, para $\rho \neq 0$ na eq.(7.6) temos um processo onde X_t e Y_t apresentam apenas correlação contemporânea, isto é, de lag=0.

- para o segundo caso:

$$\mathbf{B}_{1, faseII} = \begin{bmatrix} a & 0,4 \\ 0,4 & 0,5 \end{bmatrix}, \mathbf{B}_{1, faseII} = \begin{bmatrix} -0,3 & a \\ 0,4 & 0,5 \end{bmatrix}, \quad (7.8)$$

onde $a = 0,95, -0,90, \dots, 0,90, 0,95$, percorrendo a amplitude de valores para os coeficientes da matriz \mathbf{B}_1 , utilizando intervalos de 0,05. Cada um dos cenários descritos na eq. (7.8) será investigado fixando a matriz de covariância dos resíduos em três níveis:

$$\Sigma_0 = \begin{bmatrix} 1 & 0,0 \\ 0,0 & 1 \end{bmatrix}, \Sigma_0 = \begin{bmatrix} 1 & 0,3 \\ 0,3 & 1 \end{bmatrix}, \Sigma_0 = \begin{bmatrix} 1 & 0,6 \\ 0,6 & 1 \end{bmatrix}, \quad (7.9)$$

As simulações foram realizadas através do software *R*¹, a partir das bibliotecas *tsDyn*², *qcc*³ e *MSQC*⁴.

¹R Core Team (2018), versão 3.5.0.

²Fabio Di Narzo et al. (2009).

³Scrucca (2004).

⁴Santos-Fernández (2013).

8 Resultados

8.1 Cenário 1 - Alteração da Estrutura de Covariância dos Resíduos

A Tabela 8.1 apresenta os resultados das simulações para o primeiro cenário descrito no capítulo anterior. Podemos verificar o desempenho das cartas propostas diante de perturbações na estrutura de covariância dos resíduos. Especificamente, utilizam-se duas matrizes: \mathbf{B}_1^0 (7.7), definida originalmente pela ausência de correlação contemporânea, e \mathbf{B}_1 , exposta em (7.2). Para ambos os casos, o processo está sob controle quando $\rho = 0$ (representado na linha marcada em cinza).

Tabela 8.1: Percentual médio e desvio padrão de batealdas perturbadas identificadas — Cartas de T_e^2 de Hotelling e da Variância Generalizada W_e para \mathbf{B}_1^0 e \mathbf{B}_1 , variando ρ . Número de repetições: 200 repetições.

ρ	\mathbf{B}_1^0				\mathbf{B}_1			
	T_e^2		W_e		T_e^2		W_e	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
-0,95	8,09	0,17	100,00	0,00	8,10	0,18	100,00	0,00
-0,90	7,81	0,16	100,00	0,00	7,83	0,18	100,00	0,00
-0,85	7,58	0,15	100,00	0,00	7,58	0,18	100,00	0,00
-0,80	7,35	0,17	100,00	0,00	7,32	0,16	100,00	0,00
-0,75	7,08	0,16	100,00	0,00	7,08	0,17	100,00	0,00
-0,70	6,83	0,15	100,00	0,02	6,84	0,17	100,00	0,00
-0,65	6,58	0,14	99,96	0,09	6,57	0,16	99,95	0,11
-0,60	6,36	0,16	99,56	0,32	6,37	0,15	99,55	0,29
-0,55	6,17	0,15	97,90	0,60	6,14	0,14	97,77	0,61
-0,50	5,96	0,15	93,02	1,11	5,97	0,15	92,99	1,18
-0,45	5,78	0,15	84,10	1,64	5,75	0,14	84,35	1,44
-0,40	5,59	0,15	71,65	1,93	5,61	0,14	71,63	1,92
-0,35	5,48	0,13	57,32	2,32	5,47	0,14	57,26	2,22
-0,30	5,36	0,14	43,06	2,13	5,34	0,14	42,97	2,15
-0,25	5,24	0,15	30,10	2,14	5,22	0,13	30,28	2,05
-0,20	5,15	0,14	20,27	1,77	5,16	0,15	20,49	1,81
-0,15	5,06	0,14	13,45	1,64	5,09	0,15	13,34	1,37
-0,10	5,04	0,14	9,04	1,22	5,04	0,15	8,92	1,31
-0,05	4,99	0,14	6,52	1,07	5,02	0,13	6,52	1,16
0,00	5,00	0,14	5,95	1,09	5,00	0,14	5,74	1,01
0,05	5,00	0,14	6,53	1,07	5,01	0,13	6,48	0,97
0,10	5,04	0,14	9,03	1,37	5,03	0,13	8,98	1,16
0,15	5,07	0,14	13,41	1,48	5,09	0,13	13,48	1,53
0,20	5,13	0,15	20,53	1,82	5,13	0,14	20,26	1,78
0,25	5,23	0,13	30,31	2,22	5,25	0,14	30,33	2,28
0,30	5,34	0,14	42,81	2,11	5,34	0,13	42,86	2,15
0,35	5,45	0,13	57,06	2,42	5,48	0,16	57,16	2,25
0,40	5,61	0,15	71,75	2,00	5,60	0,15	71,59	2,09
0,45	5,76	0,14	84,11	1,69	5,77	0,16	84,15	1,68
0,50	5,94	0,13	92,89	1,03	5,96	0,17	93,08	1,08
0,55	6,16	0,16	97,87	0,63	6,15	0,15	97,79	0,62
0,60	6,37	0,15	99,57	0,30	6,37	0,16	99,55	0,31
0,65	6,59	0,17	99,97	0,08	6,59	0,16	99,95	0,10
0,70	6,82	0,17	100,00	0,01	6,84	0,15	100,00	0,01
0,75	7,06	0,16	100,00	0,00	7,07	0,16	100,00	0,00
0,80	7,32	0,16	100,00	0,00	7,34	0,15	100,00	0,00
0,85	7,59	0,18	100,00	0,00	7,58	0,17	100,00	0,00
0,90	7,81	0,17	100,00	0,00	7,83	0,17	100,00	0,00
0,95	8,08	0,17	100,00	0,00	8,08	0,17	100,00	0,00

Observamos maior sensibilidade na detecção dos descontroles da carta W_e em relação à carta T_e^2 em toda a faixa de descontroles, tanto para \mathbf{B}_1^0 quanto para \mathbf{B}_1 . Na medida em que descontroles em ρ vão ficando mais distantes do valor do proceso sob controle (isto é, com $\rho = 0$), observamos que a sensibilidade da carta W_e aumenta substancialmente.

Adicionalmente, os resultados mostram um bom desempenho da carta W_e , tanto num processo onde apenas correlação contemporânea está presente (processo gerador seguindo \mathbf{B}_1^0), quanto num processo onde, além da correlação contemporânea, correlações defasadas cruzadas estão presentes (processo gerador dos dados segue \mathbf{B}_1). Em outras palavras, observamos que, se uma nova amostra sob monitoramento não é bem descrita pelo modelo VAR ajustado num processo sob controle (isto é, covariância dos resíduos diferente de zero) a carta W_e apresenta boa sensibilidade para detectar tais alterações, independente da estrutura de covariância das variáveis do processo.

O desempenho superior da carta W_e sobre a carta T_e^2 era esperado e se justifica pelo fato de a carta W_e monitorar diretamente a estrutura de covariância dos resíduos. Dessa forma, visto que estamos impondo alterações sobre a covariância dos resíduos, não modificando sua média, é esperado que a carta T_e^2 apresente sensibilidade baixa neste caso.

8.2 Cenário 2 - Alteração da Estrutura de B_1

Para apresentar os resultados relativos aos descontroles em B_{11} no que se refere às diferentes estruturas de covariâncias dos resíduos, optou-se por, primeiro, expor a comparação referente às mudanças na autocorrelação defasada e, posteriormente, para a correlação cruzada defasada entre X_t e Y_t . Cabe salientar que alguns descontroles descritos na seção anterior não serão apresentados por não gerarem processos VAR(1) estacionários.

8.2.1 Alteração da Estrutura de B_1 — Autocorrelação de X_t defasado de lag 1, para diferentes ρ fixos

Tabela 8.2: Percentual médio e desvio padrão de batealdas perturbadas identificadas — Cartas de T_e^2 de Hotelling e da Variância Generalizada W_e para $B_{11,\rho}$, variando a . Número de repetições: 200.

a	$\rho = 0,0$				$\rho = 0,3$				$\rho = 0,6$			
	T_e^2		W_e		T_e^2		W_e		T_e^2		W_e	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
-0,85	26,11	0,79	86,09	1,41	25,21	0,74	84,79	1,46	28,05	0,76	89,82	1,27
-0,80	15,96	0,43	71,35	1,98	15,47	0,44	70,08	2,22	17,50	0,45	77,92	1,79
-0,75	11,24	0,31	53,31	2,11	10,99	0,29	51,91	2,35	12,35	0,31	60,79	2,15
-0,70	8,75	0,22	35,74	2,14	8,66	0,24	34,37	2,28	9,55	0,24	41,95	2,25
-0,65	7,33	0,20	22,03	1,95	7,28	0,19	21,54	1,90	7,90	0,20	26,60	2,04
-0,60	6,46	0,16	13,61	1,56	6,43	0,16	13,30	1,49	6,82	0,16	16,13	1,56
-0,55	5,89	0,16	9,09	1,47	5,89	0,16	9,05	1,35	6,13	0,15	10,09	1,30
-0,50	5,49	0,15	6,90	1,09	5,50	0,16	7,19	1,13	5,67	0,16	7,48	1,26
-0,45	5,27	0,15	6,24	1,09	5,25	0,14	6,26	1,10	5,36	0,16	6,20	1,21
-0,40	5,10	0,15	5,86	1,07	5,10	0,15	5,82	1,06	5,15	0,14	5,87	1,01
-0,35	5,01	0,13	5,75	0,96	5,01	0,14	5,77	1,05	5,04	0,14	5,64	0,96
-0,30	4,99	0,14	5,84	1,05	5,00	0,14	5,68	1,00	5,01	0,14	5,72	1,01
-0,25	5,04	0,15	5,85	1,01	5,02	0,14	5,84	1,10	5,02	0,13	5,74	1,02
-0,20	5,09	0,14	5,93	1,10	5,08	0,12	5,83	1,10	5,13	0,14	5,82	0,97
-0,15	5,21	0,13	5,95	0,99	5,24	0,14	6,02	1,16	5,34	0,14	6,08	1,02
-0,10	5,38	0,14	6,26	1,03	5,42	0,14	6,36	1,09	5,59	0,16	6,50	1,16
-0,05	5,60	0,15	6,75	1,08	5,67	0,16	6,88	1,17	5,99	0,17	7,51	1,24
0,00	5,90	0,14	7,42	1,16	6,01	0,16	8,04	1,26	6,54	0,16	10,22	1,41
0,05	6,26	0,15	9,12	1,32	6,47	0,16	10,08	1,43	7,19	0,18	14,47	1,48
0,10	6,71	0,16	11,73	1,43	6,99	0,17	13,31	1,49	8,07	0,18	21,50	1,97
0,15	7,30	0,18	15,54	1,77	7,72	0,19	18,44	1,88	9,25	0,22	31,72	2,29
0,20	8,02	0,19	21,23	1,78	8,65	0,20	26,16	2,03	10,74	0,24	44,55	2,16
0,25	8,96	0,19	28,42	1,81	9,84	0,23	35,83	2,01	12,66	0,31	58,36	2,42
0,30	10,15	0,24	37,59	2,10	11,40	0,27	46,93	1,98	15,13	0,31	71,99	1,88
0,35	11,70	0,28	47,99	2,26	13,40	0,29	58,77	2,17	18,23	0,34	82,57	1,78
0,40	13,75	0,29	58,53	2,28	16,13	0,37	70,59	2,00	22,19	0,42	90,21	1,43
0,45	16,65	0,38	69,18	2,11	19,69	0,39	79,85	1,70	27,05	0,49	94,85	1,01
0,50	20,52	0,48	77,84	1,91	24,46	0,53	87,38	1,41	33,13	0,57	97,70	0,63
0,55	26,04	0,59	85,07	1,60	30,99	0,68	92,46	1,23	40,51	0,72	98,97	0,49
0,60	34,03	0,76	90,32	1,32	39,91	0,75	95,85	0,92	49,72	0,83	99,61	0,27
0,65	45,63	1,04	94,51	0,91	51,44	1,12	98,02	0,59	60,43	0,96	99,87	0,15

Na Tabela 8.2 observamos o bom desempenho da carta W_e nas três estruturas de covariância dos resíduos simulados no processo sob controle. Em oposição ao que imaginávamos, diferentes estruturas de correlação nos resíduos do processo sob controle não influenciaram no desempenho da carta W_e na detecção de descontroles na autocorrelação de lag 1 da variável X_t . Adicionalmente, nos três casos, a carta W_e apresentou boa sensibilidade na detecção dos descontroles.

Observamos também o desempenho superior da carta W_e sobre a carta T_e^2 nos três casos simulados. Este resultado, embora não esperado, pode ser justificado

em parte pelo fato de que foram realizadas alterações nos coeficientes da matriz \mathbf{B}_{11} , que por sua vez representam correlações. Dessa forma, descontroles nesses coeficientes acabaram perturbando a estrutura inicial de correlação dos resíduos, refletindo diretamente na carta de controle W_e .

Outra razão para tal resultado está no fato de que os dados foram simulados utilizando uma estrutura VAR com matriz de interceptos $\mathbf{B}_0 = 0$. Dessa forma, possíveis alterações que poderiam ser refletidas nas médias dos resíduos, diante de tal fato, podem ter sido confundidas com alterações na estrutura de covariância dos resíduos. Tendo em vista que o objetivo deste trabalho é investigar o potencial do conjunto de cartas de controle propostas na detecção rápida de descontroles (e não de avaliar o potencial no diagnóstico de descontroles), os resultados evidenciam a boa sensibilidade da abordagem proposta. Em contrapartida, pensando exclusivamente no diagnóstico de descontroles, isto é, na identificação das variáveis causadoras, identificamos a importância da utilização de um modelo VAR incluindo os interceptos.

8.2.2 Alteração da Estrutura de B_1 — Correlação Cruzada entre X_t e Y_t defasado de lag 1, para diferentes ρ fixos

Tabela 8.3: Percentual médio e desvio padrão de batealdas perturbadas identificadas — Cartas de T_e^2 de Hotelling e da Variância Generalizada W_e para $\mathbf{B}_{21,\rho}$, variando a . Número de repetições: 200.

a	$\rho = 0,0$				$\rho = 0,3$				$\rho = 0,6$			
	T_e^2		W_e		T_e^2		W_e		T_e^2		W_e	
	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão	Média	Desvio Padrão
-0,40	12,78	0,25	77,14	1,91	14,29	0,27	86,68	1,56	19,00	0,29	98,26	0,60
-0,35	11,86	0,23	68,84	2,03	13,20	0,24	79,87	1,77	17,51	0,27	96,53	0,92
-0,30	10,98	0,22	59,59	2,20	12,17	0,24	71,51	2,02	16,03	0,26	93,15	1,15
-0,25	10,13	0,19	49,52	2,34	11,18	0,24	61,34	2,23	14,57	0,26	87,38	1,59
-0,20	9,38	0,22	39,49	2,34	10,25	0,20	50,50	2,20	13,21	0,24	79,05	1,71
-0,15	8,67	0,21	31,15	2,17	9,42	0,21	40,08	2,28	11,97	0,23	68,11	2,07
-0,10	7,99	0,20	23,19	1,78	8,62	0,20	30,15	1,95	10,72	0,24	55,48	2,21
-0,05	7,44	0,17	17,30	1,70	7,95	0,19	22,38	1,91	9,65	0,22	42,13	2,10
0,00	6,93	0,18	12,79	1,56	7,35	0,18	16,24	1,80	8,65	0,19	30,18	1,99
0,05	6,47	0,16	9,74	1,43	6,78	0,17	11,76	1,60	7,80	0,18	20,50	1,82
0,10	6,09	0,15	7,86	1,27	6,32	0,16	8,79	1,32	7,05	0,17	13,55	1,64
0,15	5,76	0,14	6,66	1,01	5,92	0,16	7,11	1,21	6,41	0,16	9,37	1,16
0,20	5,47	0,14	5,99	1,12	5,58	0,15	6,10	1,15	5,92	0,16	7,03	1,12
0,25	5,26	0,14	5,69	0,97	5,32	0,14	5,85	1,07	5,50	0,14	6,04	1,03
0,30	5,13	0,15	5,58	1,11	5,15	0,14	5,80	1,04	5,21	0,14	5,81	1,06
0,35	5,03	0,13	5,64	1,05	5,03	0,14	5,80	1,03	5,05	0,15	5,70	0,94
0,40	5,00	0,13	5,71	0,96	5,01	0,13	5,80	1,01	5,00	0,12	5,75	1,03
0,45	5,04	0,14	5,64	1,07	5,04	0,14	5,64	1,07	5,04	0,14	5,70	1,15
0,50	5,13	0,13	5,71	1,06	5,17	0,14	5,68	0,94	5,25	0,14	5,75	0,99
0,55	5,30	0,13	5,75	0,97	5,40	0,14	5,98	0,99	5,60	0,14	6,18	1,11
0,60	5,59	0,15	6,13	1,06	5,70	0,14	6,35	1,05	6,10	0,16	7,72	1,20
0,65	5,93	0,16	6,96	1,23	6,13	0,14	7,51	1,27	6,81	0,17	10,94	1,22
0,70	6,43	0,18	8,80	1,23	6,73	0,16	10,36	1,37	7,77	0,17	17,35	1,70
0,75	7,03	0,17	12,02	1,44	7,50	0,19	15,16	1,54	9,01	0,21	28,16	2,04
0,80	7,83	0,20	17,50	1,57	8,46	0,21	22,48	1,76	10,55	0,22	42,02	2,21
0,85	8,79	0,20	25,15	1,83	9,63	0,22	33,05	2,23	12,37	0,27	57,29	2,22
0,90	9,91	0,19	35,27	1,89	11,11	0,26	45,33	2,31	14,58	0,31	71,54	1,98
0,95	11,30	0,24	46,47	2,23	12,80	0,28	57,90	1,96	17,09	0,32	82,59	1,72

Assim como observado na seção anterior, a Tabela 8.3 reforça o bom desempenho de detecção da carta W_e . Novamente, observamos a maior sensibilidade da carta W_e em relação à carta T_e^2 . Similarmente aos resultados discutidos na Tabela 8.2, a baixa sensibilidade da carta T_e^2 se deve ao fato de que novamente foram impostas perturbações numa estrutura de correlação (neste caso, uma correlação cruzada de lag 1), e também pelo fato de estarmos considerando o modelo VAR sem interceptos. Entretanto, novamente, considerando que o foco deste trabalho está na detecção de descontroles, os resultados ilustram a boa sensibilidade considerada o conjunto de cartas de controle.

Os resultados apresentados nesta seção evidenciam o bom desempenho da carta de controle W_e , proposta neste trabalho, na detecção de descontroles variados, tanto na estrutura temporal das variáveis, quanto na estrutura de correlação dos resíduos do modelo VAR. Mesmo no cenário 2 apresentado, em oposição ao que era esperado, verificamos o desempenho superior da carta de controle W_e em relação à carta T_e^2 , proposta por (Choi et al., 2008).

Acreditamos que a carta T_e^2 pode ser bastante útil diante da modelagem VAR considerado intercepto não-nulo e a inclusão de descontroles na trajetória média das variáveis nas bateladas (além dos descontroles impostos nos coeficientes da matriz \mathbf{B}_1 e Σ , os quais representam correlações).

9 Considerações Finais

Neste trabalho investigamos a sensibilidade das cartas T_e^2 de Hotelling e da Variância Generalizada W_e num estudo de caso simulado de processo em batelada sob uma estrutura de Vetor Autorregressivos (VAR). Utilizando os resíduos do modelo VAR, arranjados conforme [Choi et al. \(2008\)](#), propusemos o uso da carta W_e e investigamos o seu desempenho em conjunto com a carta T_e^2 (esta última proposta por estes autores).

Através de um processo em batelada simulado com duas variáveis sob análise, descontroles de diferentes graus foram impostos na estrutura de autocorrelação e de correlação cruzada defasada entre as variáveis, bem como na estrutura de correlação de resíduos. Os resultados evidenciaram o desempenho superior da carta W_e em relação a carta T_e^2 , proposta na literatura.

Referências Bibliográficas

- Camacho, J., Picó, J., e Ferrer, A. (2009). The best approaches in the on-line monitoring of batch processes based on pca: Does the modelling structure matter? *Analytica chimica acta*, 642(1-2):59–68.
- Choi, S. W., Morris, J., e Lee, I.-B. (2008). Dynamic model-based batch process monitoring. *Chemical Engineering Science*, 63(3):622–636.
- da Silveira Bueno, R. D. L. (2008). *Econometria de séries temporais*. Cengage Learning.
- Fabio Di Narzo, A., Aznarte, J. L., e Stigler, M. (2009). *tsDyn: Time series analysis based on dynamical systems theory*. R package version 0.7.
- Ge, Z., Song, Z., e Gao, F. (2013). Review of recent research on data-based process monitoring. *Industrial & Engineering Chemistry Research*, 52(10):3543–3562.
- Johnson, R. A. e Wichern, D. W. (2007). Applied multivariate statistical analysis.
- Marcondes Filho, D. (2001). Monitoramento de processos em bateladas através de cartas de controle multivariadas utilizando análise de componentes principais multidirecionais.
- Montgomery, D. C. (2007). *Introduction to statistical quality control*. John Wiley & Sons.
- Nomikos, P. e MacGregor, J. F. (1994). Monitoring batch processes using multiway principal component analysis. *AIChE Journal*, 40(8):1361–1375.
- Nomikos, P. e MacGregor, J. F. (1995). Multivariate spc charts for monitoring batch processes. *Technometrics*, 37(1):41–59.
- Pan, X. e Jarrett, J. E. (2012). Why and how to use vector autoregressive models for quality control: the guideline and procedures. *Quality & Quantity*, 46(3):935–948.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reinsel, G. C. (2003). *Elements of multivariate time series analysis*. Springer Science & Business Media.
- Santos-Fernández, E. (2013). *Multivariate Statistical Quality Control Using R*, volume 14. Springer.

- Scrucca, L. (2004). qcc: an r package for quality control charting and statistical process control. *R News*, 4/1:11–17.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48.
- Vanhatalo, E. e Kulahci, M. (2015). The effect of autocorrelation on the hotelling t2 control chart. *Quality and Reliability Engineering International*, 31(8):1779–1796.
- Wold, S., Kettaneh, N., Fridén, H., e Holmberg, A. (1998). Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemometrics and intelligent laboratory systems*, 44(1-2):331–340.